

TESIS DE MAESTRÍA
ADMINISTRACION DE NEGOCIOS (MBA)

Título:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Autor: Lic. Agustín María Orlandi
Director de Tesis: Dr. Lic. Roberto Vola-Luhrs
Codirector de Tesis: Mg. Ing. Agr. Enrique C. Bombelli

Buenos Aires - 2020

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

DEDICATORIA

Gracias a todos los que confiaron en mí. En esta investigación han colaborado e influido en el desarrollo de este trabajo. Mi familia, Flor, Enrique, Roberto, Mario, Gerardo, Federico, Kenny, Nicolás, Meggy. Viole, Sandra, Anahí, Daniela, Guillermo, Alejandro, Jean Paul y Beatriz.

Agustín María Orlandi.

ÍNDICE

LISTA DE TABLAS	6
LISTA DE FIGURAS	7
LISTA DE ABREVIACIONES	14
RESUMEN	15
PALABRAS CLAVE	15
KEYWORDS	16
CAPÍTULO 1: INTRODUCCIÓN	17
1.1 INTRODUCCIÓN AL TEMA DE TESIS	17
1.2 FUNDAMENTACIÓN Y ANTECEDENTES	19
1.3 DESCRIPCIÓN DEL PROBLEMA Y/O PREGUNTA SIN RESPUESTA ABORDADOS	24
1.4 PREGUNTA DE INVESTIGACIÓN	25
1.5 HIPÓTESIS	25
1.6 OBJETIVOS DE LA INVESTIGACIÓN	25
1.7 METODOLOGÍA	25
CAPÍTULO 2: MARCO TEÓRICO	29
2.1 CRITERIOS DE ÉXITO	29
2.1.1 ANÁLISIS FODA	29
2.1.2 RECURSOS FINANCIEROS ADMINISTRADOS POR LA DIRECCIÓN DE RECURSOS HUMANOS	30
2.1.3 RECURSOS NO FINANCIEROS ADMINISTRADOS POR LA DIRECCIÓN DE RECURSOS HUMANOS	31
2.2 SISTEMAS DE SOPORTE DE DECISIÓN (SSD)	31
2.3 GESTORES DE BASE DE DATOS (SGBD)	33
2.3.1 CONCLUSIÓN	36

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

2.4 CLASIFICACIÓN DE BASES DE DATOS (SGBD)	36
2.5 <i>BIG DATA</i>	38
2.5.1 OBJETIVO DEL <i>BIG DATA</i>	38
2.5.2 ¿QUÉ DATOS UTILIZA <i>BIG DATA</i> ?	39
2.5.3 CASOS DE EJEMPLIFICADORES, PORQUÉ <i>BIG DATA</i>	41
2.5.4 TÉCNICA DE SISTEMATIZACIÓN DE INFORMACIÓN DE <i>BIG DATA</i>	43
2.5.5 COMO SE IMPLEMENTA <i>BIG DATA</i>	43
2.5.6 <i>BIG DATA</i> APLICADO EN LA GESTIÓN DE RECURSOS HUMANOS	44
2.5.7 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)	46
2.6 BODEGA DE DATOS (<i>DATA WAREHOUSE</i>)	46
2.7 DEFINICIÓN DE ETL	47
2.8 MINERÍA DE DATOS (DATAMINING)	48
2.8.1 OBJETIVO DE LA MINERÍA DE DATOS	48
2.8.2 TIPOS DE MINERÍA DE DATOS	51
2.8.3 METODOLOGÍAS DE MINERÍA DE DATOS	65
2.9 ANALYTICS	69
2.9.1 OBJETIVO DE <i>PEOPLE ANALYTICS</i>	70
2.9.2 CASOS DE EJEMPLIFICADORES DEL USO DE <i>PEOPLE ANALYTICS</i>	72
2.9.3 FUNCIONES Y TAREAS DE <i>PEOPLE ANALYTICS</i>	74
CAPÍTULO 3: GESTIÓN DE RECURSOS HUMANOS Y SECTOR FINANCIERO DE LA REPÚBLICA ARGENTINA	76
3.1 LA MOTIVACIÓN.	77
3.2 EL ENTORNO LABORAL Y LOS TRABAJADORES	77
3.3 RELACIÓN ENTRE ACTIVIDADES DE FORMACIÓN Y DESEMPEÑO	81
3.4 RELACIÓN ENTRE AUSENTISMO Y CLIMA LABORAL	84

3.5 SECTOR FINANCIERO DE LA REPÚBLICA ARGENTINA _____	86
CAPÍTULO 4: DISEÑO DE LA SOLUCIÓN _____	87
4.1 DATA VISUALIZATION _____	88
4.1.1 ¿QUÉ ES LA VISUALIZACIÓN DE DATOS? _____	88
4.1.2 <i>ORACLE SQL DEVELOPER</i> _____	101
4.2 ANÁLISIS DE DATOS _____	102
4.3 PREPARACIÓN DE LOS DATOS _____	102
4.4 CREACIÓN DE LA BASE DE DATOS _____	103
4.5 CONSTRUCCIÓN DE ETL _____	103
4.5.1 PREPARAR LOS DATOS ETL _____	103
4.5.2 SELECCIÓN DE LA FUENTE DE DATOS _____	104
4.5.3 CONSTRUCCIÓN DEL <i>DATA WAREHOUSE</i> _____	104
4.6 IMPLEMENTACIÓN DE LA SOLUCIÓN _____	106
4.6.1 CREACIÓN DE TABLAS DIMENSIONALES <i>ORACLE SQL DEVELOPER</i> _____	106
4.6.2 PROCESAMIENTO DE DATOS EN <i>ORACLE ANALYTICS</i> _____	109
4.6.3 CREACIÓN DE SET DE DATOS EN <i>ORACLE ANALYTICS</i> _____	111
4.6.4 Generación de <i>FACT_ANALYTICS_RRHH</i> _____	123
4.6.5 IMPORTACIÓN DE DATOS EN LOS DIMENSIONALES _____	130
4.6.6 DISEÑO DE LAS PRUEBAS _____	158
CAPÍTULO 5: CONCLUSIONES _____	169
5.1 CONCLUSIONES _____	169
5.2 TRABAJO A FUTURO _____	171
CAPÍTULO 6: BIBLIOGRAFÍA _____	172

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

LISTA DE TABLAS

Tabla 1 Clasificación de los datos.....	43
Tabla 2 Tipología de los algoritmos de minería de datos	53
Tabla 3 Relación existente entre las tareas y técnicas de minería de datos	62

LISTA DE FIGURAS

Figura 1 Diagrama de Venn de Data Science de Drew Covey	20
Figura 2 Matriz Análisis FODA.....	29
Figura 3 Gestores de base de datos (SGBD).....	33
Figura 4 Estructura jerárquica usada en los SGBD.....	34
Figura 5 Modelo en red.....	34
Figura 6 Estructura relacional.....	35
Figura 7 Estructura multidimensional.....	35
Figura 8 Dimensiones del big data.....	40
Figura 9 “Cada minuto del día”	42
Figura 10 Variables de estudio.....	44
Figura 11 Asociación de diversas disciplinas de minería de datos	49
Figura 12 Knowledge Discovery in Database.....	50
Figura 13 Proceso de extracción de patrones mediante el uso de técnicas de minería de datos.....	57
Figura 14 Clustering.....	59
Figura 15 Regresión, ejemplos regresión lineal y regresión logística	60
Figura 16 Árbol de decisión, ejemplo entrevista laboral	61
Figura 17 “El modelo de madurez en analítica”	72
Figura 18 Ordenamiento de las necesidades	82
Figura 19 Interés estratégico	82
Figura 20 Rentabilidad.....	83
Figura 21 Oracle Analytics	90
Figura 22 Detalle de la descarga	91
Figura 23 Generación de cuenta.....	92
Figura 24 Comprobación de usuario	92

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

Figura 25 Verificación de correo electrónico	93
Figura 26 Confirmación generación de cuenta	93
Figura 27 Panel principal Oracle Analytics	94
Figura 28 Panel de aplicación	94
Figura 29 Descarga archivo .rar	95
Figura 30 Menu principal Oracle Analytics.....	96
Figura 31 Menu principal Data Visualization. Oracle Analytics.....	96
Figura 32 Menú generación de conexión. Oracle Analytics	97
Figura 33 Data flow. Oracle Analytics	97
Figura 34 Console. Oracle Analytics	98
Figura 35 Academy. Oracle Analytics	99
Figura 36 Look and feel. Oracle Analytics	99
Figura 37 Elaboración de la visualización. Oracle Analytics	100
Figura 38 Look and feel: Distribución y permisos	101
Figura 39 Etapas del proceso analítico	102
Figura 40 Construcción del data warehouse	105
Figura 41 Creación de la tabla DIM_EMPLEADOS.....	107
Figura 42 Validación de la tabla DIM_EMPLEADOS	107
Figura 43 Creación de la tabla DIM_PUESTOS	108
Figura 44 Validación de la tabla DIM_PUESTOS	108
Figura 45 Creación de la tabla DIM_SUCURSALES	109
Figura 46 Validación de la tabla DIM_SUCURSALES.....	109
Figura 47 Conexión de base de datos. Paso 1	110
Figura 48 Conexión de base de datos. Paso 2	110
Figura 49 Conexión de base de datos. Paso 3	110

Figura 50 Validación de creación de la conexión. Paso 1	111
Figura 51 Validación de creación de la conexión. Paso 2	111
Figura 52 Creación set de datos. Paso 1.....	111
Figura 53 Creación set de datos. Paso 2.....	112
Figura 54 Visualización de esquemas.....	112
Figura 55 Visualización de tablas de contenidos.....	113
Figura 56 Selección de atributos. Paso 1	113
Figura 57 Selección de atributos. Paso 2	114
Figura 58 Conexión contra base de datos	114
Figura 59 Confirmación de la conexión.....	115
Figura 60 Validación creación set de datos	115
Figura 61 Creación set de datos. Paso 1.....	115
Figura 62 Creación set de datos. Paso 2.....	116
Figura 63 Visualización de esquemas.....	116
Figura 64 Visualización de tablas de contenidos.....	117
Figura 65 Selección de atributos. Paso 1	117
Figura 66. Selección de atributos. Paso 2	118
Figura 67 Conexión contra base de datos	118
Figura 68 Confirmación de la conexión.....	119
Figura 69 Validación creación set de datos	119
Figura 70 Creación set de datos. Paso 1.....	119
Figura 71 Creación set de datos. Paso 2.....	120
Figura 72 Visualización de esquemas.....	120
Figura 73 Visualización de tablas de contenidos.....	121
Figura 74 Selección de atributos. Paso 1	121

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

Figura 75 Selección de atributos. Paso 2	122
Figura 76 Conexión contra base de datos	122
Figura 77 Confirmación de la conexión.....	123
Figura 78 Validación creación set de datos	123
Figura 79 Generación de data flow. Paso 1	123
Figura 80 Generación de data flow. Paso 2	124
Figura 81 Generación de data flow. Paso 3	124
Figura 82 Generación de data flow. Paso 4	124
Figura 83 Generación de data flow. Paso 5	125
Figura 84 Generación de data flow. Paso 6	125
Figura 85 Generación de data flow. Paso 7	125
Figura 86 Generación de data flow. Paso 8	126
Figura 87 Generación de data flow. Paso 9	126
Figura 88 Generación de data flow. Paso 10	126
Figura 89 Generación de data flow. Paso 11	127
Figura 90 Generación de data flow. Paso 12	127
Figura 91 Generación de data flow. Paso 13	127
Figura 92 Ejecución de data flow. Paso 1	128
Figura 93 Ejecución de data flow. Paso 2.....	128
Figura 94 Validación de la creación de la tabla en Oracle database.....	129
Figura 95 Validación de la creación del data set “FACT_ANALYTICS_RRHH” en Oracle Analytics	129
Figura 96 Validación de la creación del data flow.....	129
Figura 97 Importación de datos. Paso 1	130
Figura 98 Importación de datos. Paso 2.....	130

Figura 99 Importación de datos. Paso 3	130
Figura 100 Importación de datos. Paso 4	131
Figura 101 Importación de datos. Paso 5	131
Figura 102 Importación de datos. Paso 6	131
Figura 103 Validación importación datos de origen.....	132
Figura 104 Importación de datos. Paso 1	132
Figura 105 Importación de datos. Paso 2	133
Figura 106 Importación de datos. Paso 3	133
Figura 107 Importación de datos. Paso 4	133
Figura 108 Importación de datos. Paso 5	134
Figura 109 Importación de datos. Paso 6	134
Figura 110 Validación importación datos de origen.....	135
Figura 111 Importación de datos. Paso 1	135
Figura 112 Importación de datos. Paso 2	136
Figura 113 Importación de datos. Paso 3	136
Figura 114 Importación de datos. Paso 4	137
Figura 115 Importación de datos. Paso 5	137
Figura 116 Importación de datos. Paso 6	137
Figura 117 Validación importación datos de origen.....	138
Figura 118 Página home sección Data	138
Figura 119 Elaboración data flow. Paso 1	139
Figura 120 Elaboración data flow. Paso 2	139
Figura 121 Ejecución de data flow	139
Figura 122 Creación del proyecto.....	140
Figura 123 Página home Data Visualization.....	140

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

Figura 124 Selección del data set.....	141
Figura 125 Preparación del data set. Paso 1.....	141
Figura 126 Preparación del data set. Paso 2.....	141
Figura 127 Preparación del data set. Paso 3.....	142
Figura 128 Preparación del data set. Paso 4.....	143
Figura 129 Preparación del data set. Paso 5.....	143
Figura 130 Preparación del data set. Paso 6.....	144
Figura 131 Preparación del data set. Paso 7.....	144
Figura 132 Preparación del data set. Paso 8.....	144
Figura 133 Preparación del data set. Paso 9.....	145
Figura 134 Preparación del data set. Paso 10.....	145
Figura 135 Preparación del data set. Paso 11.....	146
Figura 136 Preparación del data set. Paso 12.....	146
Figura 137 Preparación del data set. Paso 13.....	147
Figura 138 Preparación del data set. Paso 14.....	147
Figura 139 Preparación del data set. Paso 15.....	147
Figura 140 Preparación del data set. Paso 16.....	148
Figura 141 Preparación del data set. Paso 17.....	148
Figura 142 Preparación del data set. Paso 18.....	149
Figura 143 Complementación del data set. Paso 1	150
Figura 144 Depuración de datos. Paso 1	150
Figura 145 Definición de tratamiento de datos.....	151
Figura 146 Complementación del data set. Paso 2	151
Figura 147 Complementación del data set. Paso 3	152
Figura 148 Depuración de datos. Paso 2.....	152

Figura 149 Definición de tratamiento de datos.....	153
Figura 150 Diagrama de datos las relaciones existentes. Paso 1	153
Figura 151 Diagrama de datos las relaciones existentes. Paso 2	154
Figura 152 Diagrama de datos las relaciones existentes. Paso 3	154
Figura 153 Diagrama de datos las relaciones existentes. Paso 4	155
Figura 154 Diagrama de datos las relaciones existentes. Paso 5	155
Figura 155 Diagrama de datos las relaciones existentes. Paso 6	156
Figura 156 Diagrama de datos las relaciones existentes. Paso 7	156
Figura 157 Diagrama de datos las relaciones existentes. Paso 8	157
Figura 158 Guardado de proyecto. Paso 1	157
Figura 159 Guardado de proyecto. Paso 2	158
Figura 160 Ausentismo 2016-2017 vs Clima laboral 2016-2017	159
Figura 161 Ausentismo 2016-2017 vs Clima laboral 2016-2017. Ordenado por clima laboral de mayor a menor.....	160
Figura 162 Ausentismo 2018-2019 vs Clima laboral 2018-2019	161
Figura 163 Ausentismo 2018-2019 vs Clima laboral 2018-2019. Ordenado por clima laboral de mayor a menor.....	161
Figura 164 Diferencia clima laboral 2016/2017 – clima laboral 2018/2019 vs ausentismo 2016/2017 – ausentismo 2018/2019.....	162
Figura 165 Horas de capacitación 2016 vs desempeño 2016	163
Figura 166 Horas de capacitación 2016 vs desempeño 2016. Ordenado por horas de capacitación de mayor a menor	164
Figura 167 Horas de capacitación 2017 vs desempeño 2017	164
Figura 168 Horas de capacitación 2017 vs desempeño 2017. Ordenado por horas de capacitación de mayor a menor	165
Figura 169 Horas de capacitación 2018 vs desempeño 2018	165

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 170 Horas de capacitación 2018 vs desempeño 2018. Ordenado por horas de capacitación de mayor a menor	166
Figura 171 Horas de capacitación 2019 vs desempeño 2019	166
Figura 172 Horas de capacitación 2019 vs desempeño 2019. Ordenado por horas de capacitación de mayor a menor	167
Figura 173 Horas de capacitación total anual 2016 a 2019 y desempeño total anual 2016 a 2019.....	167

LISTA DE ABREVIACIONES

BDD - Bases de datos distribuidos

CEO – Chief Executive Officer

DSS - Sistemas de soporte de decisión

ETL - Extract, Transform and Load (Extraer, Transformar y Cargar)

FODA - Fortalezas y debilidades, amenazas y oportunidades

GPS - Sistema de Posicionamiento Global

IA - Inteligencia artificial

IB O BI - inteligencia de negocios / business intelligence

KPI - Key Performance Indicator

RRHH – Recursos Humanos

SGBD - Gestores de base de datos

SGBD - Gestores de base de datos

RESUMEN

La utilización de la tecnología *big data* y analítica aplicada a la gestión de recursos humanos, *people analytics*, indica que la empresa desea obtener, procesar, analizar e interpretar información, con el propósito de mejorar las decisiones que se toman en el ámbito de la gestión del capital humano. Se está cambiando la naturaleza básica de la gestión del capital humano para que evolucione desde un rol transaccional a uno en donde su contribución estratégica se vuelve crucial.

El área de recursos humanos debe pasar de una etapa en donde las decisiones se toman a partir de la intuición, a otra en donde las decisiones surgen del proceso inteligente de evidencias: la etapa de big data, data mining y analítica aplicada a la gestión de recursos humanos. Esta investigación apunta de manera directa a brindar las herramientas, técnicas y enfoques que faciliten dicha transición, asegurando de este modo el rol estratégico del área. Actualmente, los avances tecnológicos son utilizados en todos los sectores de las empresas, es la oportunidad de mejorar la capacidad para tomar decisiones basadas en el procesamiento inteligente de la información.

PALABRAS CLAVE

ANALÍTICA, APRENDIZAJE, AUSENTISMO, BODEGA. CIENCIA DE DATOS, CLIMA, DATOS, DESEMPEÑO, DIMENSIÓN, FLUJO, FORMACIÓN, MACRODATOS, MINERÍA, SET, VISUALIZACIÓN.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

ABSTRACT

The use of big data and analytics applied to human resource management, people analytics, indicates that companies want to obtain, process, analyze and interpret information for the purpose of improving decisions made in the field of human capital management (HCM).

Furthermore, the basic nature of HCM is evolving from a transactional role to one where its strategic contribution is becoming crucial.

The field of human resources must advance from a stage where decisions are mostly based on intuition to a stage where decisions emerge from processing evidence intelligently: the stage of big data, data mining and analytics applied to HCM. This study is directed at providing tools, techniques and approaches that facilitate said transition, thus ensuring the strategic role of this field.

Currently, technological advances are being adopted in all areas of companies; therefore, there is an opportunity of improving the decision-making process in the field of HCM based on the intelligent processing of information.

KEYWORDS

ABSENTEEISM, ANALYTICS, BIG DATA, DATA, DATA SCIENCE, DIMENSION, DISPLAY, ENVIRONMENT, FLOW, LEARNING, MINING, PERFORMANCE, SET, TRAINING, WAREHOUSE.

CAPÍTULO 1: INTRODUCCIÓN

1.1 INTRODUCCIÓN AL TEMA DE TESIS

Los saltos en el almacenamiento y el poder computacional de la última década nos llevan a invertir y aprovechar la información de los trabajadores, se utiliza bases de datos y se aplican diferentes técnicas de extracción de conocimiento. Los tiempos cambian y también lo hacen los requisitos del negocio y las exigencias de los usuarios se han vuelto aún más complejas.

La gran cantidad de datos obtenidos día a día en las empresas proporcionan amplias oportunidades, pero estos mismos vienen con desafíos significativos en términos de gastos y de inversión de almacenamiento de la información. Las técnicas de extracción de información han sido empleadas para crear patrones, tendencias y algoritmos de predicción en diferentes contextos, obteniendo resultados prometedores que demuestran cómo determinadas características sociológicas, económicas y educativas pueden afectar el rendimiento laboral encontrando un indicador que permita identificar a los colaboradores con mayor riesgo de fallo, abandono, juicios laborales, rendimiento laboral, entre otros. Es importante poder predecir la posibilidad que esto suceda desde que ingresa alguien a una empresa y poder cambiar los factores que pudieran estar causando esta situación.

El principal problema que se presenta en las empresas es que no poseen la cultura de llevar a cabo una gestión del almacenamiento de datos de los colaboradores que formaron y forman parte de ella para poder llevar adelante una completa medición y análisis basado en los recursos humanos. Para ello será indispensable la utilización de un sistema de soporte de decisión que sea sustentable.

Cada vez que se decide medir algo en particular, un costo, un tiempo, un nivel de calidad, etc., es porque, previamente, se ha considerado que esa medición reportará algún beneficio. Toda medición siempre debe comenzar con la definición del propósito que persigue, por ejemplo, tiempo de respuesta para un incidente reportado.

Medir con objetividad los datos que proporcionan las acciones, no solamente para analizar los desvíos presupuestarios, sino también los efectos que causan las actividades de formación en el desempeño y el impacto del clima organizacional en el ausentismo, podrían ser consecuencia de la falta de atención en estos factores.

Sin embargo, en muchos casos, lo que parece más explícito no es precisamente el propósito, sino un conjunto de supuestos y afirmaciones que denominamos preocupaciones, en verdad son las dos caras de una misma moneda, aunque cada preocupación puede ser expresada como un propósito. Comenzar una

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

medición a partir de una preocupación puede llevar a algún tipo de error inicial que, seguramente, será costoso; en tanto comenzar dicho proceso a partir de la definición clara de uno o varios propósitos es el mejor camino para que medir sea un proceso que agregue valor. Una vez clarificado el propósito, la próxima tarea es plantear la siguiente: si el costo que se genera por el proceso de medición vale la pena.

La revolución de *analytics* está ganando mucho terreno. Las organizaciones han mejorado sus capacidades en la construcción de equipos encargados de hacer un uso más estratégico de la información. Los especialistas en negocios han tomado conciencia de que necesitan información para entender las razones por las cuales la gente quiere trabajar para ellos, cuál es su desempeño; porqué desean quedarse en el largo plazo, quién tendrá más posibilidades de ser exitoso en la organización, quiénes serán los líderes más adecuados para cada etapa que deba transitar la empresa y qué se necesita para prestar un mejor servicio e innovar en un entorno cambiante y cada vez más competitivo. Tomar decisiones subjetivas puede desmotivar a las personas ya que no comprenden el sustento de estas. Todo ello se puede obtener a través de *analytics* aplicado a la gestión de recursos humanos.

El trabajo en cuestión tiene como objetivo identificar los beneficios en la calidad de la toma de decisiones, en el área de recursos humanos a partir del uso de tecnologías con respecto a la gestión tradicional, en una entidad financiera.

Las herramientas tecnológicas a utilizar son: *big data*, *data mining* y analítica aplicada a la gestión de recursos humanos, *people analytics*. En la actualidad, el poco uso de estas tres en el área de recursos humanos se debe a que gran parte de las decisiones que se toman en relación con las personas, se hacen en base a suposiciones, datos subjetivos e incluso hasta la intuición. Hoy en día, el área de recursos humanos, gracias al inminente avance de la tecnología y una sociedad cada vez más digitalizada, tiene a disposición grandes volúmenes de información de sus colaboradores que es necesario analizar correctamente. Sin embargo, las empresas no han sabido sacarle provecho a estos datos.

La intuición, el instinto de los ejecutivos seguirá siendo esencial e irremplazable y hasta se podría afirmar que es lo que seguirá definiendo su valor como tales. Pero una gran ventaja de la economía digital es la posibilidad de validar las intuiciones contra los datos duros y concretos. El análisis de grandes volúmenes de datos no reemplaza a la intuición, la complementa (Bolo, 2017).

La disciplina de recursos humanos no ha alcanzado el final de su evolución. El siguiente paso será la fase de *analytics* para la toma de decisiones. Esto implica que, a partir de ahora, la toma de decisiones debe basarse más en las evidencias procesadas e interpretadas científicamente, que en aquellas suposiciones que surgen de un enfoque meramente intuitivo (Cravino, 2011).

Desde el año 2015 a 2016 se ha doblado la cantidad de organizaciones que están en capacidad de desarrollar modelos predictivos. Las mismas oscilan alrededor del 4% en el 2015 a 8% en el 2016. Mientras en el 2015 solo el 24% se sentía listo para emprender temas de analítica, en el 2016 dicho valor aumentó al 32%. (Deloitte University Press, 2016).

1.2 FUNDAMENTACIÓN Y ANTECEDENTES

En los últimos años, dada la presión de la competencia y la necesidad de tener sistemas más integrados, las organizaciones se han dedicado a crear equipos de analítica y desarrollar una oferta global en este tema.

La analítica de *big data* es donde las técnicas de análisis avanzado de datos operan en grandes volúmenes. Por lo tanto, el análisis de *big data* se trata, en realidad, de dos cosas: el *big data* y el análisis, y cómo los dos se han unido para crear una de las tendencias más profundas de la inteligencia de negocios (IB) en la actualidad (Russom, 2011).

Es importante definir qué significa *big data* (*macrodatos*) y en qué se diferencia del concepto de *analytics* y *data mining* (minería de datos o exploración de datos). La tecnología *big data* es donde las técnicas de analítica avanzada operan en grandes conjuntos de datos y surge el concepto o enfoque de *data science* (desarrolladores de técnicas analíticas y algoritmos matemáticos que permiten optimizar el valor de los datos); es en cierta forma un concepto que integra y da sentido a gran parte de las propuestas vinculadas al trabajo de *analytics* en las áreas del gerenciamiento de personas.

Al intentar rastrear el origen del término *data science*, es posible remontarse al año 1997 (Wu, 1997). Se caracterizó al trabajo estadístico como una trilogía de recolección de datos, modelado y análisis de los mismos y toma de decisión.

Se puede considerar al término *data science* como una disciplina autónoma, la cual engloba seis áreas (Cleveland, 2001). Investigaciones multidisciplinarias, modelos, métodos de datos, pedagogía, evaluación de herramientas y teoría.

Se entiende a la *data science* como una combinación de tres dimensiones: Conocimiento estadístico matemático: es el componente de contenido duro, que involucra conocimientos de estas ciencias, utilizadas de manera práctica. *Hacking skills*: comprende elementos que deberían no solo ser pensados como habilidades sino también como actitudes. ¿Qué cosas definen a un hacker? Seguramente no solo sus conocimientos. Tan importante son sus habilidades y, tal vez más aun, su actitud, lo cual comprende por una parte la constancia y perseverancia, e impulso que orienta a la búsqueda continua e incesante del

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

objetivo del trabajo o tarea en cada oportunidad. El trabajo en *data science* no permite llegar fácil o mágicamente a conclusiones, requiere tiempo, esfuerzo y dedicación. *Substantive Expertise*: el conocimiento experto, es el saber sobre el tema de estudio. Podría ser contar con el relevamiento del estado del arte sobre el asunto, el cual permitirá un mejor abordaje de la situación, pero también requiere que este sea utilizado de manera crítica, con la capacidad para distanciarse o cuestionarlo (Bodenheimer, 2017).

Figura 1 Diagrama de Venn de Data Science de Drew Covey



Fuente: (Kozyrkov, 2018)

La investigación tradicional es lo que se obtiene a través de la mezcla del conocimiento de la matemática y la estadística con conocimiento experto. Los *papers* tradicionales que se publican día a día representan esta combinación, ya que se fundamentan en el estudio y desarrollo de conocimiento mediante las herramientas clásicas. Estos modelos provienen de la ciencia dura.

La zona de peligro se presenta cuando se combinan las habilidades de *hacker* y el conocimiento especializado, ya que implicaría el espíritu de creatividad y perseverancia que se asocian al mundo *hacker*. Es una combinación de actitud y conocimiento carente de instrumentos específicos que permitan aprovechar y vincular ambos elementos.

Machine learning: se trata de la utilización y aprovechamiento de los datos mediante técnicas computacionales, matemáticas y estadísticas asociadas al fenómeno de *big data* buscando patrones significativos, pero no evidentes. La práctica parece haber convergido en la definición de “aprendizaje automático” (Escudero, 2019).

El *data science* involucra, estadística, métodos tradicionales de análisis; *data munging*, tomar datos y cambiar su formato; y la visualización, estrategias múltiples para facilitar su comprensión mediante imágenes de los que esconden los datos. Puede incluir el uso de herramientas gráficas.

Data mining puede entenderse como el proceso de descubrimiento automático de información útil en grandes repositorios de datos (Tan, 2016). También se podría pensar como el proceso de utilización de diversas herramientas que permiten encontrar patrones útiles y no evidentes en grandes conjuntos de datos.

Otro de los aspectos consiste en resaltar que de los datos almacenados surja información útil o relevante, por lo tanto, que no sea trivial, es decir, conocimiento nuevo, con capacidad de generar valor, sobre el objeto de estudio. Otro elemento para destacar es el concepto de patrones o información no evidente. Al implementar procesos de *data mining* se busca sorprender con conocimiento que otras herramientas no hubiesen podido brindar o señalar. Se lo puede comprender como un análisis que combina múltiples variables de grandes conjuntos de datos.

Analytics, people analytics o la analítica aplicada a la gente. Recoge información de diversas fuentes que permite: analizar y seleccionar candidatos que puedan tener un alto desempeño, identificar competencias de equipos comerciales exitosos, predecir riesgos reputacionales, analizar temas de cultura y compromiso e identificar planes de carrera de alto valor para desarrollar a los líderes del negocio.

Las empresas líderes incorporan el análisis en sus organizaciones resolviendo que se base en datos y definiendo lo que esperan lograr mediante el uso de *big data*. El CEO y el equipo de liderazgo deben describir cómo los análisis darán forma al desempeño del negocio, ya sea mejorando los productos y servicios existentes, optimizando los procesos internos, creando nuevos productos u ofertas de servicios o transformando los modelos de negocios. Las organizaciones de alto rendimiento a menudo construyen sus organizaciones en torno a los datos y se comprometen a crear nuevos datos.

El nombre de esta tendencia, analítica aplicada a las personas, *people analytics*, hace referencia al uso de la información relacionada con la gente. Dicha información se usa para tomar decisiones de negocio. Las áreas de foco pueden variar dependiendo de la industria y de los temas específicos del negocio.

A diferencia de sus orígenes, donde los profesionales de recursos humanos se centraban en el cumplimiento de los requerimientos exigidos por la normativa vigente en materia laboral poniendo en práctica un rol de apoyo (Vola-Luhrs, 2010), el trabajo de los profesionales de recursos humanos tiene una serie de frentes y responsabilidades que van más allá del lugar o posición que ocupen en el organigrama. La cruzada actual incluye, entre otras cosas, tomar más deberes para facilitar que la

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

organización cumpla sus objetivos. El área es hoy valorada en la toma de decisiones por y para el negocio, tanto por los resultados de los procesos implementados en gestión humana, como por la opinión experta y el asesoramiento profesional que brinda.

Sumar obligaciones e incumbencias al área de recursos humanos es un efecto de la evolución de la tecnología en la gestión humana. Sus primeras tareas, operativas y administrativas (administración y archivo de legajos de los empleados), agregaban poco valor a la organización, pero con el tiempo se amplió el alcance de la función. Los profesionales del área se hicieron cargo del reclutamiento y la selección del personal, sumaron la capacitación de estos, se hicieron cargo de la seguridad e higiene y se ocuparon de los asuntos de desarrollo del personal, entre otras muchas cosas más.

El aspecto cualitativo, la lógica y las razones de cada una de estas tareas fue mutando, de simplemente desarrollar actividades administrativas pasaron a ser socios estratégicos del negocio, es decir formar parte de la toma de decisiones necesarias para que la organización pueda crecer exitosamente. Reclutar personal ya no se trata simplemente de seleccionar personas y ocupar puestos, sino de escoger el tipo de empleado adecuado para que la estrategia del negocio suceda. La selección de los empleados se encuentra ligada al entendimiento de que las personas son acertadas para cumplir los objetivos que tiene la empresa y es por ello que el reclutamiento se convirtió en un trabajo para el presente y futuro del negocio.

Se comprende el término “socio” como un aliado estratégico que conoce el negocio de la compañía a la cual asiste y asesora, en términos relacionados a la administración de los recursos humanos, con impacto en las finanzas, niveles de producción, distribución de las operaciones y otras áreas críticas de la organización. Se trata de un rol superador, incluso, al de experto en brindar servicios de recursos humanos (Vola-Luhrs, 2010).

El aporte dinámico del colaborador será permanente, ya que el ámbito en el que deberá desempeñarse cambiará vertiginosamente, obligando a que el negocio cambie y si cambia el negocio, cambia, consecuentemente, el rol requerido y esperado. En este contexto, no podemos pensar sólo en funciones y tareas. Necesitamos que los colaboradores, por sí mismos, propongan cambios que impactarán en los resultados, vínculos y tecnología a aplicar. La persona adquiere mayor supremacía.

En el área de desarrollo de recursos humanos, el trabajo de gestión humana es poder garantizar que el negocio pueda, a futuro, hacer lo que se espera a partir de la gestión de personas. Actualmente desarrollo no solo es desarrollo de personas, sino que, principalmente, está en juego el desarrollo organizacional (Bodenheimer, 2017).

La analítica aplicada a los recursos humanos, *people analytics*, es importante porque la gestión debe evolucionar desde la presunción o el instinto hacia el análisis más preciso de los que pasó o está pasando, para poder predecir o influenciar sobre lo que pasará o debería pasar. El uso de datos, evidencias, para la toma de decisiones.

Como se mencionó durante la justificación del tema, una de las causas de la desmotivación de los trabajadores surge de la falta de equidad interna. Una manera efectiva de conocer más acerca de los motivos de la desmotivación es el uso de *big data* en recursos humanos lo que permitiría, en gran parte, que las decisiones que se tomen con relación a las personas, se hagan con menos suposiciones, menos datos subjetivos e intuición.

Cada uno de estos antecedentes, evidencia la oportunidad de usar la información de la gente (proveniente de recursos humanos, de otras fuentes, incluso fuera de la compañía) para tomar mejores decisiones de negocio.

Google, Twitter y muchas otras firmas tienen hoy equipos dedicados específicamente a la analítica. Hoy, dichos equipos se denominan a sí mismos como “equipos que escuchan a la gente”. Recogen datos de varias fuentes, incluyendo sistemas de recursos humanos, encuestas de compromiso de diversos portales y redes sociales. Analizan esta información para entender la cultura de la organización e identificar oportunidades para mejorar la retención, el desempeño o para diagnosticar debilidades gerenciales (Deloitte University Press, 2016).

Muchas compañías adquieren tecnología de recursos humanos en la nube, lo cual les permite ver su información de manera más integrada. Asimismo, están ingresando a recursos humanos profesionales con competencias en analítica (Bersin, 2015). “Las organizaciones están vinculando psicólogos organizacionales, estadísticos y otros profesionales con orientación cuantitativa. Ahora los profesionales que se orientan al manejo de información son muy valorados” (Deloitte University Press, 2016. 90).

El área de recursos humanos tiene la oportunidad de mostrar el valor y el retorno de la inversión que puede generar la analítica, justificando una mayor inversión en este proceso, lo cual, se convierte en un círculo virtuoso, en el cual una mayor inversión en analítica redundaría en medir el valor de la gente para la organización.

Analizar la información es solo una parte de la solución. El valor real es convertir estos hallazgos en cambios reales que impacten al negocio. La clave será invertir tanto en desarrollar competencias en analítica como en desarrollar competencias de interpretación y transformación, para asegurar cambios visibles, equitativos, estratégicos y sostenibles en el tiempo.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

1.3 DESCRIPCIÓN DEL PROBLEMA Y/O PREGUNTA SIN RESPUESTA

ABORDADOS

Las empresas en su gestión de personas deben enfrentar los desafíos de resolver problemas y encontrar explicaciones para aquello que sucede en su interior. Es necesario, en cada oportunidad, tener información, datos y hechos que permitan formular hipótesis y realizar diagnósticos certeros. Sin estos, resulta prácticamente imposible tomar decisiones certeras sobre las acciones a implementar. No efectuar estos análisis determina que los efectos sean del orden de la casualidad. Pero ¿Qué organización tiene la posibilidad de financiar o invertir en casualidades?

Frente al contexto Económico Financiero, las empresas deben mantenerse en estándares de calidad y financieramente hablando mantener la rentabilidad del negocio y así mismo lograr optimizar cada vez más los costos que genera.

El mercado financiero de esta manera les exige a las empresas contemplar pautas y estrategias de crecimiento económico, sin mencionar los lineamientos de dicha incorporación económica, por lo tanto, si la empresa si quiere estar a la altura de las circunstancias debe considerar diversidad de métodos.

El alcance de esta investigación es abordado en el marco de los recursos humanos, haciendo foco específicamente en los procesos de ausentismo, clima organizacional, desempeño y actividades de formación. Estas tecnologías son transversales a todas las áreas que componen las organizaciones incluyendo una gran cantidad de procesos que no están alcanzados en este trabajo tales como programas de beneficios, *payroll*, relaciones gremiales, seguridad e higiene, rotación, reclutamiento y selección de personal, entre otros.

A partir de esta investigación, se espera dar soporte en la calidad de la toma de decisiones basada en el desarrollo de modelos predictivos para la gestión de personas, los cuales constituyen herramientas útiles para las organizaciones al sustentar en datos objetivos que mejoran la calidad de la decisión y reducen la incertidumbre.

En esta tesis se pretende demostrar en qué medida la aplicación de *big data* y analítica para la gestión de recursos humanos mejora la calidad de la toma de decisiones con respecto a la asignación de recursos financieros y no financieros, por basarse en indicadores objetivos, administrados con métodos rigurosos que desestiman patrones subjetivos.

1.4 PREGUNTA DE INVESTIGACIÓN

¿Cómo impacta *big data* y analítica aplicada a las personas sobre la calidad de la toma de decisiones con respecto a la asignación de recursos financieros y no financieros con respecto a la gestión tradicional?

1.5 HIPÓTESIS

La aplicación de *big data* y analítica aplicada a la gestión de recursos humanos, *people analytics*, mejora sustancialmente la calidad de la toma de decisiones sobre la asignación de recursos financieros y no financieros con respecto a los métodos tradicionales de gestión según los resultados que arroja el análisis de las relaciones del ausentismo vs. clima laboral y desempeño vs. actividades de formación en la entidad financiera donde se desarrolla la investigación.

1.6 OBJETIVOS DE LA INVESTIGACIÓN

Objetivo principal

Identificar los beneficios en la calidad de la toma de decisiones a partir del uso de *big data* y analítica aplicada a la gestión de recursos humanos, *people analytics* con respecto a la gestión tradicional.

Objetivos secundarios

Optimizar la calidad de la toma de decisiones con respecto a la asignación y el uso eficiente de recursos financieros destinados al personal.

Optimizar la calidad de la toma de decisiones con respecto a la asignación y el uso eficiente de recursos no financieros destinados al personal.

1.7 METODOLOGÍA

Se propone profundizar el desarrollo de los conceptos de *big data*, *data mining* y *people analytics* y su utilidad, en las diferentes unidades de negocio administradas, para luego hacer foco en las relaciones del ausentismo vs. clima laboral y desempeño vs. actividades de formación.

Dentro del desarrollo metodológico se asegura la utilización de aplicaciones específicas de medición tales como: *performance analytics*, a partir de la gestión del desempeño; *learning analytics*, a partir de los programas de formación del personal y evaluación de la capacitación; *talent analytics*, estimación de potencial del personal, *engagement analytics*, a partir de la medición del clima laboral y medición del ausentismo.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Respecto a la bibliografía existente, se recurre a la tecnología de banco de datos para seleccionar y organizar lo requerido dentro del proceso de investigación.

Las fuentes de consulta serán: bibliotecas, sistemas de búsqueda de datos informatizados, base de datos de la entidad financiera, información personal de los colaboradores, informes de ausentismo, encuestas de clima organizacionales, evaluaciones de desempeño e historial de actividades de formación.

La investigación propone respaldar el accionar del área de recursos humanos en cuanto a las decisiones que toma sobre el personal a través de la vinculación de la medición de los indicadores de “medición de ausentismo vs. clima laboral” y “medición del desempeño vs. actividades de formación”.

La técnica de investigación lleva a descubrir patrones en el material relevado. El enfoque se efectúa mediante la tabulación y sistematización de la información disponible para luego aplicar la técnica de análisis de resultados.

Es un estudio de caso, ya que se trata de una herramienta para describir en forma exhaustiva, la ocurrencia de un evento ya sea un problema o un fenómeno en un determinado contexto. Al mismo tiempo constituye una investigación empírica que indaga un fenómeno contemporáneo dentro de un contexto real. Este método se utiliza en los sistemas de información para identificar descripciones o interrelaciones de procesos y en la realización de estudios comparativos.

La estructura por seguir es de tipo lineal, pudiendo existir alguna interacción en la fase de interpretación de los resultados y la verificación de las interpretaciones.

Permite la producción de documentación que caracteriza perfectamente al evento en estudio para ofrecer patrones sobre dicho evento.

La validación de la información que ofrecen los resultados obtenidos viene dada por el análisis cualitativo, a través de las diversas visiones del investigador acerca de los resultados emanados de los instrumentos de captura de datos, es decir, lo que se conoce como triangulación de la información y control de un par.

La triangulación hace referencia al procedimiento que permite organizar diferentes tipos de datos de manera que puedan ser constatados. Mientras que el control de un par se refiere a la realización de un análisis por parte de un investigador, de modo que se neutralicen los posibles sesgos que pudiera tener alguna persona. Esta técnica permite ampliar y consensuar las interpretaciones que se realizan.

Los estudios de caso pueden relevar información sobre los problemas de la implementación, es decir de las relaciones entre los sistemas nuevos y los recursos humanos.

Se investiga la herramienta *big data*, *data mining* y analítica aplicada a las personas, sometida a análisis, respecto a un proceso comparativo de información del personal en la empresa y su aporte a la dirección estratégica de los recursos humanos en los planes de gestión del ausentismo, clima laboral, desempeño de los colaboradores y actividades de formación. Se expone desde el entendimiento en profundidad diversos aspectos y dimensiones que proporciona la implementación de la herramienta planteada y las diferencias con el método tradicional.

Esta tarea se desarrolla a partir de la propuesta del análisis de la información recabada, se enumeran las etapas de la medición en la gestión de los recursos humanos. Se investiga a través de un proceso comparativo bibliografía, análisis y exposición de casos junto a sus resultados.

Se recopila información de la tecnología *big data* para “analítica aplicada a las personas” durante los últimos cuatro años. Se acompaña de una descripción de las principales métricas estandarizadas por esta tecnología respecto a un proceso comparativo, información propia de la entidad financiera y entrevistas a profesionales de recursos humanos y expertos. Con el mismo se intenta demostrar la relevancia y de qué manera se contribuye a la toma de decisiones a partir de los resultados de la relación entre indicadores tales como ausentismo y clima laboral, desempeño y actividades de formación. A diferencia de la gestión tradicional en la que las mediciones se elaboran de manera individual y sin relación de resultados

Lo que esta tesis se propone aportar, es determinar específicamente:

Si partir de la implementación de *big data*, *data mining* y *people analytics* (analítica aplicada a las personas), las decisiones del área de recursos humanos se toman basadas en datos objetivos, compartidos y relacionados entre sí, y si esto minimiza la subjetividad y favorece a la equidad interna respecto la asignación de recursos y la toma de decisiones.

Si el área de recursos humanos evoluciona desde un rol transaccional a uno en donde su contribución estratégica en la asignación y uso eficiente de recursos financieros y no financieros sea crucial.

Determinar, además, si el uso de la tecnología aplicada a las decisiones en el área de recursos humanos resulta de una mayor profesionalización para que quienes laboran en dicha área mejoren su capacidad para tomar decisiones basados en el procesamiento inteligente de la información.

Determinar si pueden ser derribados los paradigmas “a mejor clima laboral, menor ausentismo” y “a mayor cantidad de actividades de formación, mejor desempeño” para pasar de una etapa en donde las

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

decisiones se toman a partir de la intuición, a otra (la etapa de *people analytics*) donde las decisiones surgen del proceso inteligente de evidencias.

La disponibilidad de la información presentada sobre la herramienta y los resultados de las entrevistas realizadas a expertos, sobre la complejidad en la gestión de los recursos humanos y el uso de la tecnología, permitirá encontrar la base de sustentación para validar los objetivos mencionados.

Se pretende que la información generada en esta investigación sea de suma utilidad para aquellos que deseen implementar la herramienta *big data*, analítica aplicada a la gente, en las organizaciones que cuenten con los medios para ponerla en práctica y así generar valor agregado a las organizaciones desde un ejercicio superador de la gestión de los recursos humanos.

CAPÍTULO 2: MARCO TEÓRICO

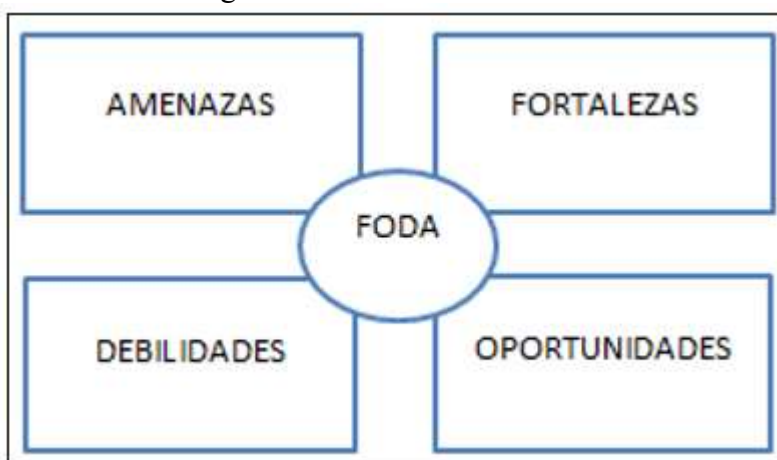
2.1 CRITERIOS DE ÉXITO

Al cumplir con los objetivos secundarios se podrá determinar con qué recursos cuenta la empresa, cuáles administra la Gerencia de Recursos Humanos y de ellos, cuáles necesitarán ser gestionados con mayor calidad de la toma de decisiones.

2.1.1 ANÁLISIS FODA

Para dar paso a la investigación se hace uso de la matriz FODA, es una herramienta de estudio y análisis para evidenciar la situación actual de una empresa, institución, proyecto o persona. Se hace un análisis de las características internas, fortalezas y debilidades, y externas, amenazas y oportunidades, en una matriz cuadrada. A continuación, se detallan los recursos con los que cuenta la Dirección de Recursos Humanos para administrar de manera equitativa y eficiente entre los colaboradores a los fines de optimizar el desempeño, la formación y el compromiso de los colaboradores. Una de las funciones de la Dirección, es planificar, organizar y controlarlos para influenciar decisiones en la organización.

Figura 2 Matriz Análisis FODA



Fuente: Elaboración propia

Fortalezas

Buen soporte comercial.

Amplio número de recursos.

Más de 1.000.000 clientes particulares y más de 30.000 empresas de todos los rubros en la República Argentina.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

Capacidad de adaptación a entornos cambiantes.

Capital humano calificado, en formación continua y conocedor del mercado financiero.

Estabilidad de los sueldos en el sector empresarial

Debilidades

Servicio costoso.

Calidad media.

Ausencias y licencia importantes de los empleados.

Alto índice de rotación.

Oportunidad

Buena imagen corporativa. Buen uso de la marca empleadora.

Mercado abierto, posibilidad de comercializar tanto presencialmente como de manera remota.

Buenas posibilidades de crecimiento.

Expansión de ventas en el mercado financiero para personas y pequeñas, medianas y grandes empresas.

Acceso a crédito y/o políticas de crédito fiscal en materia de formación del personal.

Amenazas

Aumento de competidores.

Posibilidad de recesión de la economía.

Directivas repentinas del gobierno.

2.1.2 RECURSOS FINANCIEROS ADMINISTRADOS POR LA DIRECCIÓN DE RECURSOS HUMANOS

Los recursos financieros son aquellos de los cuales las empresas obtienen los fondos que necesitan para financiar sus inversiones y capital. Son aquellos que tienen algún grado de liquidez. Aporte del dinero de los accionistas, utilidades, acceso programas de financiamiento y crédito fiscal gubernamental a través del Ministerio de Trabajo Empleo y Seguridad Social de la Nación para la gestión de la formación de las personas y bonos por alcance de objetivos.

Las formas en las que se canalizan son donaciones, no reembolsable, y créditos, reembolsables.

2.1.3 RECURSOS NO FINANCIEROS ADMINISTRADOS POR LA DIRECCIÓN DE RECURSOS HUMANOS

Los recursos no financieros son aquellos de los cuales las empresas obtienen y necesitan para administrar sus las actividades actuales, estos son generados por la propia empresa. Administración de actividades de formación, administración de reconocimientos con asignación de premiaciones no dinerarias, administración de eventos corporativos e institucionales con fines de integración, administración de actividades lúdicas, administración de beneficios corporativos como planes de medicina prepaga personales o familiares, membresías, acceso a actividades culturales y turismo.

Las formas en las que se canalizan son en especie, equipos, materiales, asistencia técnica, actividades de formación y tiempo (mano de obra).

2.2 SISTEMAS DE SOPORTE DE DECISIÓN (SSD)

Los sistemas de soporte de decisión son el registro de datos y modelos analíticos que interactúan entre si para brindar soporte necesario para la toma de decisiones y desarrollar mejores acciones de negocio tanto para la operación, como para la administración.

“Un SSD es un sistema interactivo basado en computadoras, que asiste a los tomadores de decisiones, utilizando datos y modelos. El objetivo es resolver problemas semiestructurados y no estructurados dentro de una organización. El concepto no es necesariamente nuevo, de hecho, se habla de ello desde 1960 especialmente sobre las áreas de finanzas y operaciones” (Morton, 1971).

“Los Sistemas de Soporte a la Decisión acoplan los recursos intelectuales de los individuos con las capacidades de las computadoras para mejorar la calidad de las decisiones. Es un sistema basado en computadora para los tomadores de decisión que tratan con problemas semiestructurados”

“Los sistemas de soporte a la decisión unen los recursos intelectuales del individuo con las capacidades de la computadora para mejorar la calidad de las decisiones” (Efraim, 2005).

La incorporación de lo que en inglés se denomina EPSS (*Electronic Performance Support Systems* o Sistemas Electrónicos de Soporte) ha cambiado la relación entre el trabajo y el aprendizaje, por ejemplo, uno está trabajando y tiene una duda, tal vez esta duda la tuvo alguien más y tal vez alguien pensó en una respuesta que se halla “embebida” en la computadora para brindarla en el momento que la necesitamos.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

En algunos casos, es la misma computadora quien descubre nuestra carencia y gentilmente nos ofrece la respuesta, aunque nosotros no la hayamos pedido (Cravino, Un trabajo feliz, 2003).

Los SSD se apoyan de técnicas avanzadas de análisis de datos para apoyar la toma de decisiones con información analizada y que permite generar modelos predictivos o bien encontrar patrones en los datos.

A continuación, se nombran algunos sistemas de apoyo:

- Bodega de datos (*Datawarehouse*)
- Minería de Datos (DataMining)
- Enterprise Resource Planning
- Tablero de Comando
- Customer Relationship Management
- Sistemas de Integración de aplicaciones empresariales
- Sistema de Flujo de Trabajo (Workflow)
- Sistemas para la toma de Decisiones en Grupo
- Sistemas de información Gerencial
- Sistemas de Información para Ejecutivos
- Sistemas de Inteligencia Artificial.
- Sistemas Expertos de Soporte de Decisión (Expert Systems, ES)
- Sistemas de Expertos basados en inteligencia Artificial
- Redes Neuronales.

Características de los sistemas de soporte de decisión:

Interactividad: el sistema computacional otorga la posibilidad de interactuar en forma amigable y con respuestas a tiempo real.

Mayor velocidad de procesamiento: se procesa, mediante el uso de la computadora, un volumen de datos, superior al que se puede procesar manualmente y a un costo más bajo.

Tipo de decisiones: apoya el proceso de toma de decisiones estructuradas y no estructuradas.

Frecuencia de uso: tiene una utilización frecuente por parte de la administración.

Variedad de usuarios: puede emplearse por usuarios de diferentes áreas funcionales.

Desarrollo: permite que el usuario desarrollo de manera directa modelos de decisión sin la participación operativa de profesionales en informática.

Interacción Ambiental: permite interactuar con información externa como parte de los modelos de decisión.

Comunicación inter-organizacional: facilita la comunicación de información relevante de los niveles altos a los niveles operativos y viceversa, a través de gráficas.

Simplicidad: simple y fácil de aprender y utilizar por el usuario final.

2.3 GESTORES DE BASE DE DATOS (SGBD)

Son programas que permiten el almacenamiento, modificación y extracción de la información en una base de datos dentro del nivel conceptual, dentro de los programas de software que manejan las bases de datos proporcionan herramientas para añadir, borrar, modificar y analizar los datos. Los usuarios pueden acceder a la información mediante consultas SQL (lenguaje de programación) y dependiendo el repositorio usar filtros o preguntas específicas de interrogación para la generación de informes. Las bases de datos han estado en uso desde los primeros días de las computadoras electrónicas. (Cood, 1970)

A diferencia de los sistemas modernos, que se pueden aplicar a datos y necesidades muy diferentes, la mayor parte de los sistemas originales estaban enfocados a bases de datos específicas y pensadas para ganar velocidad a costa de perder flexibilidad. Los gestores de base de datos (*DataBase Management System*) son un sistema que permite la creación, gestión y administración de bases de datos, así como la elección y manejo de las estructuras necesarios para el almacenamiento y búsqueda de la información del modo más eficiente posible (Marin, 2019).

En la actualidad, existen multitud de SGDB, en la mayoría relacionales.

Figura 3 Gestores de base de datos (SGBD)



Fuente: Elaboración propia

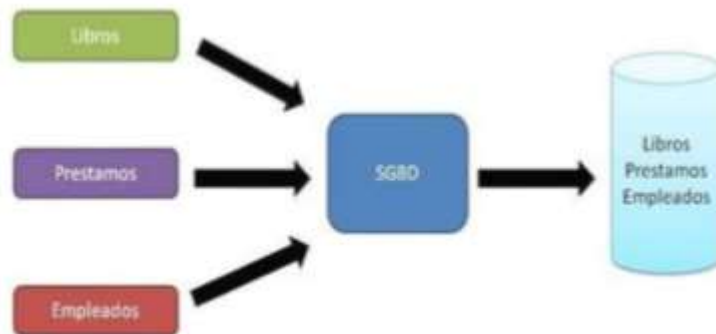
A continuación, se nombran algunos sistemas de gestores de bases de datos:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

- Sistemas SQL
- Sistemas orientados a objetos
- Sistemas NOSQL
- Sistemas XLM
- Lenguajes de modelación

La estructura jerárquica fue usada en los SGBD de los primeros mainframes

Figura 4 Estructura jerárquica usada en los SGBD

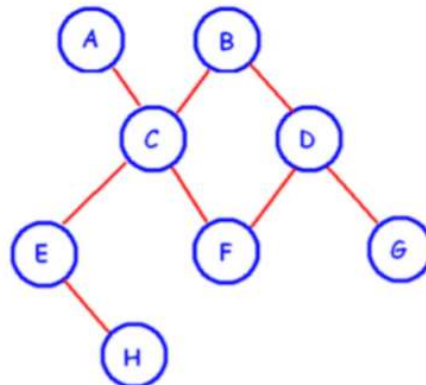


Fuente: Elaboración propia

Esta estructura es simple pero inflexible ya que las relaciones están confinadas al tipo 1: N.

En cambio, la estructura en red obtiene relaciones más complejas que las jerárquicas. Admite relaciones de cada registro con varios que se pueden seguir por distintos caminos. En otras palabras, el modelo permite relaciones N: N. El modelo en red está concebido como un modelo flexible de representar objetos y sus relaciones. Su cualidad distintiva es que el esquema visto como un conjunto de nodos conectados por arcos no tiene ninguna restricción.

Figura 5 Modelo en red



Fuente: Elaboración propia

El inventor de este modelo fue Charles Bachman, y el estándar fue publicado en 1969 por CODASYL. La estructura relacional que definió E. F. Codd es la más extendida hoy en día en IBM en 1972, esta estructura se usa en mainframes, computadoras medias y microcomputadoras. Almacena los datos en fila (tuplas) y columnas (atributos).

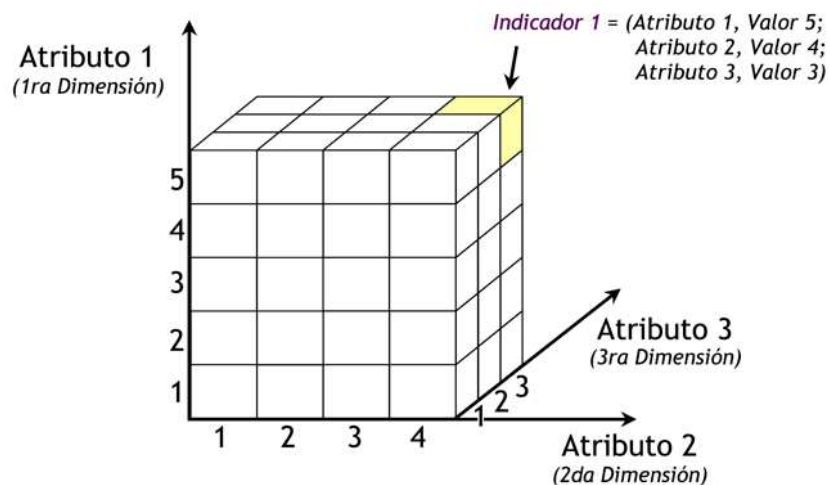
Figura 6 Estructura relacional



Fuente: Elaboración propia

La estructura multidimensional tiene parecidos a la del modelo relacional, pero en vez de dos dimensiones filas-columnas, tiene N dimensiones. Esta estructura ofrece el aspecto de una hoja de cálculo. Es fácil de mantener y entender ya que los registros se almacenan del mismo modo como se detalla la figura 7. Sus altas prestaciones han hecho de ella la base de datos más popular para el proceso analítico de transacciones en línea (OLAP – *online analytical processing*).

Figura 7 Estructura multidimensional



Fuente: Elaboración propia

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

2.3.1 CONCLUSIÓN

El uso de un estándar en los SGDB es uno de los conceptos clave ya que responde a las exigencias de evolución del sistema de información, ayudando a una mejor utilización de los recursos (humanos y de máquina) y preservando las inversiones realizadas; de este modo adaptando al sistema tecnológico nuevas arquitecturas al unificar los modelos de datos, modelos de referencia y estándares generando beneficios en la portabilidad de las aplicaciones y en la productividad de las empresas.

2.4 CLASIFICACIÓN DE BASES DE DATOS (SGBD)

Las bases de datos que usan y que implementan en la gestión del manejo de datos, en la actualidad, por las grandes empresas y medianas también por sus características y funcionalidades que estas proponen se pueden clasificar bajo un estándar o marco de referencia en el sector de sistemas. Cabe destacar que el almacenamiento de información y grandes volúmenes en la manipulación de datos juega un papel importante en la toma de decisiones. Se definirá los conceptos claves de los gestores y motores de datos en este documento como también la clasificación dependiendo el uso de estas.

Una base de datos es una colección de archivos interrelacionados, creados con un sistema manejador de bases de datos. El contenido que engloba a la información concerniente (almacenadas en archivos) de una organización, de tal manera que los datos estén disponibles para los usuarios. Una finalidad es eliminar la redundancia o al menos minimizarla.

Datos por acá, datos por allá, documentos con información valiosa en cajas, contratos encima del escritorio y la gente necesita tener análisis de datos de forma rápida y muchas veces no sabe dónde buscar o guardar información. Los datos son lo más importante a la hora de tomar decisiones, los datos que se obtienen en el día a día puede ser clasificados dentro de dos grupos: los *no estructurado* y *estructurados*. Estos mismos y con gran contenido, pueden agregar valor al negocio de la empresa ya que las personas interesadas consiguen respuestas efectivas por el análisis de métricas, informes y ventas, gracias a manipular la información que se guarda en las bases de datos o repositorios.

Nacen nuevas oportunidades de negocio con la utilización e implementación de tecnologías para manipular los datos *no estructurados*. Forman parte de este grupo documentos con imágenes y texto, archivos PDF, hojas de cálculo, imágenes, videos, archivos de audio, presentaciones con diapositivas, publicaciones en medios sociales. Estos últimos mediante técnicas *big data* y *datawarehouse* facilitan que las organizaciones puedan analizar con informes reales la situación exacta de la empresa, tarea, proyecto y no justamente gracias a las consultas a bases de datos.

Las bases de datos pueden clasificarse de acuerdo con el uso que se va a dar a los datos, a la información que se desee manipular (Noticias, artículos y guías sobre estrategia y operaciones de TI en las empresas, s.f.).

- Bases de datos estáticas

Son base de datos de sólo lectura se utilizan mucho para realizar ejecución en la toma de decisiones y realizar análisis de datos, por ejemplo, base de datos de empresas del rubro periódicos ya que almacenan información por si se requiere consultarla tiempo después.

- Bases de datos dinámicas

Estas son base de datos donde la información almacenada se modifica con el tiempo, permiten que se haga una actualización, borrado y adición de datos, además de las operaciones fundamentales de consulta. Son utilizadas en muchos ámbitos de negocios, en un sistema de información de un supermercado, una farmacia.

- Bases de datos centralizadas

Es muy similar al modelo de cliente/servidor en el sentido que la base de datos está centralizada en un lugar y los usuarios están distribuidos. Este modelo sólo brinda la ventaja de tener el procesamiento distribuido ya que en sentido de disponibilidad y fiabilidad de los datos no se gana nada.

- Bases de datos distribuidos

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos que se encuentran distribuidas en diferentes espacios lógicos (por ejemplo, un servidor corriendo 2 máquinas virtuales) e interconectados por una red de comunicaciones. Dichas BDD tienen la capacidad de realizar procesamiento autónomo, esto permite realizar operaciones locales o distribuidas.

- Bases de datos federada

Una base de datos federada es un sistema en el que varias bases de datos parecen funcionar como una sola entidad. Cada componente de la base de datos en el sistema es completamente auto sostenido y funcional. Cuando una aplicación consulta la base de datos el sistema analiza cuál de los componentes de la base contiene los datos que se solicitan y pasa la solicitud a la misma. SQL Server es un ejemplo de una base de datos federada.

- Base de datos NoSQL

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

También llamadas no solo SQL, son un enfoque hacia la gestión de datos y el diseño de base de datos para grandes volúmenes de información, abarca una amplia gama de tecnologías y arquitecturas, gracias a esto busca resolver los problemas de escalabilidad y rendimiento de big data que las bases de datos relacionales no fueron diseñadas para abordar.

Las grandes y pequeñas empresas gracias a esta característica acceden y analizan grandes cantidades de datos no estructurados o datos que se almacenan de forma remota en varios servidores virtuales en la nube, en escritorios, cajones de oficina, cajas o dispositivos de almacenamiento tipo USB. Las bases de datos más populares NoSQL son Apache Cassandra, simpleDB, Google BigTable, Mapreduce, Memcachedb y Voldemort. Las empresas que utilizan NoSQL incluyen NetFlix, LinkedIn y Twitter.

2.5 BIG DATA

2.5.1 OBJETIVO DEL BIG DATA

Siempre que una persona o empresa se conecta a la red está generando información que es almacenada en la “nube”. Toda esa información sobre consumos, pagos, actitudes, intereses, estados de ánimo o comportamientos puede ser utilizada con múltiples fines, pero antes debe ser organizada y analizada. “Los usuarios han dejado de ser meros receptores de información y han comenzado a ser generadores de esta” (CESEDEN - Centro Superior de Estudios de la Defensa Nacional, 2016).

Cada vez son más los equipos que se conectan a Internet. Entre los de uso cotidiano, primero fueron las computadoras, de escritorio o portátiles, luego los teléfonos celulares, consolas de videojuegos y los televisores comenzaron a aprovechar a hacer uso de la conectividad. De este modo, lo que era propio de la computación paso a ser parte de otros dispositivos: actualizaciones automáticas de software, descarga desde servidores remotos, uso y acceso a contenidos digitales, interacción con otros usuarios y generación de entornos colaborativos y de interacción. Así, cada tipo de producto empezó a incorporar posibilidades de conexión a Internet, heladeras, aires acondicionados, sistemas de iluminación, entre otros. Dispositivos que pueden manejarse desde un Smartphone, activar y desactivar remotamente es solo una posibilidad básica si pensamos el impacto potencial de estas tecnologías en generación de datos y análisis de conducta. La heladera que hace el pedido de productos faltantes al supermercado de manera autónoma puede ser un ejemplo (Bodenheimer, 2017).

La *big data* es un término impreciso, que se usa cuando queremos hablar de los datos que la sociedad crea y procesa en forma digital con cada vez más creciente velocidad, volumen y variedad (Brust, 2020).

Big data ya no pertenece únicamente al ámbito de la tecnología. En la actualidad, un imperativo de negocio, que brinda soluciones para desafíos de negocio de incremento de datos a los que se enfrentan las empresas de los mercados bancarios y financieros de todo el mundo. Las compañías de servicios financieros están aprovechando las soluciones *big data* para transformar procesos, sus organizaciones y, muy pronto, el sector en su totalidad (David Turner, 2013).

No está claro que el término *big data* tenga un significado preciso. El término refiere a al volumen y tipo de datos provenientes de la interacción con dispositivos interconectados, como teléfonos celulares, tarjetas de crédito, cajeros automáticos, relojes inteligentes, computadoras personales, dispositivos de GPS y cualquier objeto capaz de producir información y enviarla electrónicamente a otra parte. La *big data* se refiere a la copiosa cantidad de datos producidos espontáneamente por la interacción con dispositivos interconectados (Escudero, 2019).

Pensar en *big data* es observar el desarrollo tecnológico y las capacidades de almacenamiento que se van desarrollando y a su vez la cantidad de datos que se genera segundo a segundo. Muchos han podido entender que eso puede ser aprovechada, convirtiéndolo en información y mejor aún, en conocimiento (Bodenheimer, 2017).

2.5.2 ¿QUÉ DATOS UTILIZA *BIG DATA*?

En el mundo de *big data*, existe un consenso en que los *big data* poseen tres V.

Se describe *big data* en términos de tres V: volumen, velocidad, variedad. Los datos son demasiado grandes para procesar con las herramientas actuales; llega demasiado rápido para un almacenamiento e indexación óptimos; y es demasiado heterogéneo para encajar en un esquema rígido. Existe una gran presión sobre los investigadores de bases de datos para estudiar, explicar y resolver los desafíos técnicos en *big data*, pero no encontramos inspiración en las tres V (Gottlob, 2013).

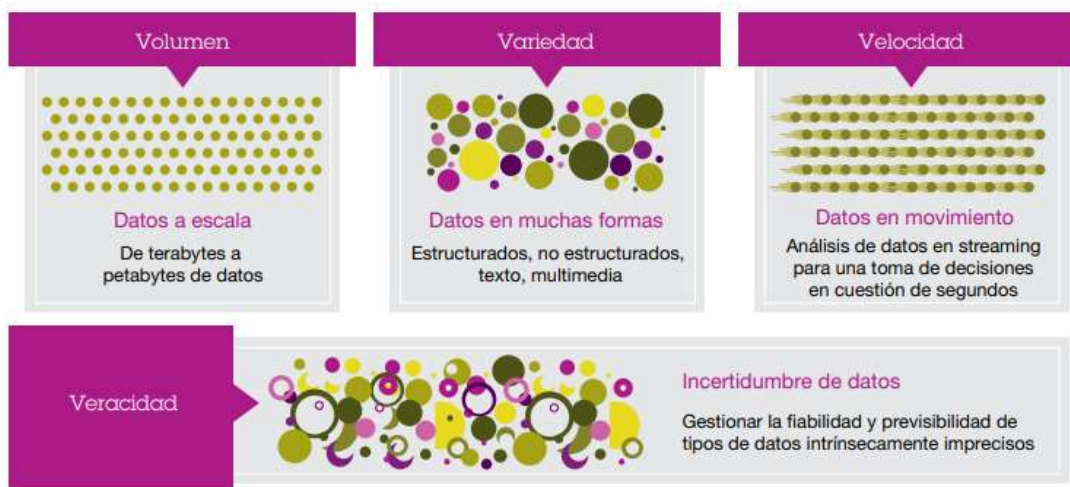
Un estudio realizado a 1.144 profesionales de diferentes empresas ha consultado a los encuestados sobre la definición de *big data*. Los resultados más altos corresponden a las siguientes frases: “un mayor ámbito de información”, “nuevos tipos de datos y análisis” e “información en tiempo real” (Michael Schroeck, 2012).

Estos resultados coinciden con la forma más común de caracterizar tres dimensiones de *big data* –‘las tres V:’ volumen, variedad y velocidad. Y si bien estas dimensiones engloban los principales atributos de *big data*, creemos que las empresas deben tener en cuenta una cuarta e importante dimensión: la

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

veracidad. Incluir la veracidad como el cuarto atributo de *big data* muestra la importancia de abordar y gestionar la incertidumbre inherente a algunos tipos de datos (Michael Schroeck, 2012).

Figura 8 Dimensiones del big data



Fuente: (Michael Schroeck, 2012)

El 71% de esas empresas de los mercados bancarios y financieros indican que uso de la información (con *big data*) y del uso de *analytics* está dando lugar a una ventaja competitiva para sus empresas. En comparación con el 63% de los encuestados de los demás sectores (David Turner, 2013).

Entonces, *big data* corresponde a una gran cantidad de datos que poseen las siguientes características: volumen, variedad, veracidad y velocidad. *Analytics* es una herramienta que permite abordar esos datos. El importante desafío para *big data*: si los datos están, pero no sistematizados convenientemente, es casi lo mismo que si no están. Creer que la información está por el mero hecho de que los datos existen es un serio error.

Desarrollando las tres V, la primera de ellas, volumen, quiere decir que la cantidad de datos debe ser grande ¿Cuán grande? Hoy quizás 10 terabytes es considerada una gran cantidad de datos (Wang, 2012). “Expertos como Ernesto Mislej, expresan que no hay un número exacto que determine qué es *big data* y que no, sino que está determinado por las herramientas que se utilizan para llevar adelante la gestión de los datos y su análisis, eso es lo que define si es o no es *big data*” (Bril, 2015).

La segunda V es velocidad, esta característica hace referencia a la frecuencia en la que los datos son generados, capturados, analizados y distribuidos para luego poder utilizarlos y accionar sobre los resultados. Hoy en día se requiere que los datos sean prácticamente en tiempo real, ya que, si se poseen los datos rápidamente, uno puede predecir con exactitud lo que va a suceder (Wang, 2012).

La tercera es la variedad, con una gran proliferación de tipos de datos, ya sean estructurados o desestructurados, uno debe captar datos de diferentes tipos y fuentes para poder mejorar y optimizarla calidad de la información generada (Wang, 2012).

2.5.3 CASOS DE EJEMPLIFICADORES, PORQUÉ *BIG DATA*

Junto con la baja del costo de la tecnología y los nuevos softwares de administración y análisis de información, genera un momento histórico para el análisis de datos. Estos elementos significan que tienen las competencias necesarias para analizar grandes cantidades de datos de forma económica y rápida. Por lo que esta capacidad representa una oportunidad de para ganar en términos de eficiencia, productividad y rentabilidad (Minelli, 2013).

En (PowerData, 2013) Se resume una lista de drivers para la utilización de *big data*, estos son consumidores más sofisticados, la automatización y por último la monetización:

En primer lugar, los consumidores más sofisticados provocan que estos sean más difíciles de entender y sus exigencias sean mayores. *Big data* es un método para poder comprender a los consumidores y responder a sus necesidades y expectativas mediante análisis que antes eran impensados y no se podían realizar. Este análisis puede permitir entre otras cosas, la obtención de correlaciones de comportamiento entre una serie de usuarios. Permitiendo optimizar y personalizar el trato hacia ese segmento de consumidores. Ofreciendo quizás productos más atractivos para ellos, mediante canales más efectivos.

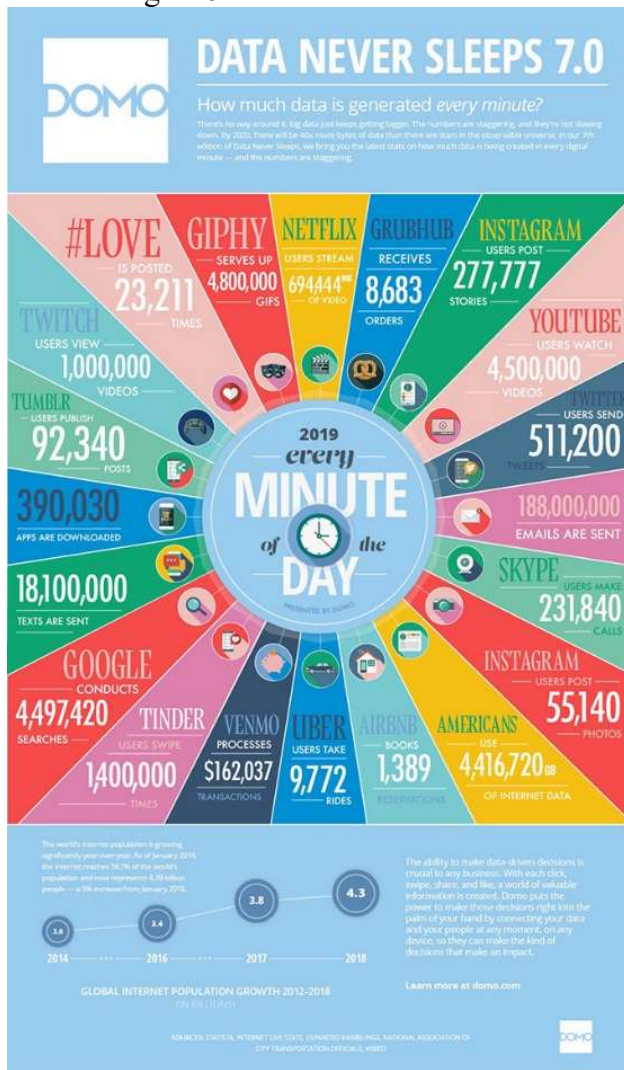
La automatización, ya que la tecnología permite gestionar de forma automática y con mayor precisión los datos y su utilización. La automatización busca ganar en tiempos, eficiencia y costos.

La monetización, es decir, mejorar los resultados económicos, ya sea utilizando la herramienta para asistir por ejemplo en campañas de marketing personalizadas en tiempo real. Asistiendo al proceso de toma de decisiones, uno puede obtener mejores resultados de rentabilidad.

La cantidad de datos con la que se puede contar es abrumadora, a continuación, se expresa los que ocurre en cada plataforma o red, minuto a minuto.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 9 “Cada minuto del día”



Fuente: (DOMO, s.f.)

Cada minuto 4.500.000 de usuarios observan videos en Youtube, se generan 277.777 publicaciones en Instagram, se generan 511.200 tweets de usuarios, se envían 188.000.000 de mails, se realizan 231.840 llamadas a través de Skype, se contratan 9.772 servicios de traslado de Uber, se generan 4.497.420 de consultas en Google, se producen 1.389 reservas en Airbnb, son algunos de los ejemplos que nos manifiesta la figura 9; esto nos lleva a determinar que definitivamente estamos rodeados de datos que, siendo procesados de la manera correcta, podrían hablarnos sobre gustos de clientes y posibles clientes, qué es lo que están pensando los colaboradores de una organización, la reputación de una compañía y la valoración de los productos que se comercializan en el mercado, cuáles son las tendencias y las costumbres de la sociedad, entre otras.

“Piensen en lo que hicieron en las últimas dos horas. Si caminaron con su celular, muy posiblemente hayan generado datos de su ubicación geográfica, y ni hablar si activaron el *GPS* para viajar en auto” (Escudero, 2019).

Tabla 1 Clasificación de los datos

Datos Administrativos	Datos de gestión del talento	Datos de tecnologías sociales y de conducta
Datos demográficos de colaboradores	Gestión del desempeño	Datos de conducta en tiempo real
Liquidación de sueldos	Encuestas de clima	Redes sociales y profesionales
Fechas de selección	Evaluaciones de talento	Uso de mail

Fuente: Elaboración Propia

Cabe destacar que, en algunas ocasiones, los datos no son generados por el propósito de crearlos, como en las respuestas a una encuesta tradicional, sino como resultado de otra acción.

2.5.4 TÉCNICA DE SISTEMATIZACIÓN DE INFORMACIÓN DE *BIG DATA*

La sistematización de la información es muchas veces la más importante de las tareas. Los datos son materia prima para la discusión. Éstos pueden ser anárquicos y en apariencia inconexos (de censos, encuestas *online*, etc.) a un subconjunto pequeño pero que puede ser estudiado como si hubiese provenido de un experimento, aun cuando dicho experimento jamás fue implementado.

La selección de los datos debe ser la apropiada, la misma debe incluir comprensión del dominio de negocios, del problema y del objetivo de análisis. A su vez se debe determinar las características de validez, consistencia y precisión de los datos seleccionados.

2.5.5 COMO SE IMPLEMENTA *BIG DATA*

Se definen tres modelos de implementación de *big data*, estos dependen del *business intelligence* (BI) de la empresa, la infraestructura de la empresa y el nivel de conocimiento del entorno. El primer modelo es el revolucionario, que lo define como la aplicación en el tiempo real de *big data*. Este tipo de implementación tiene como ventajas la agilidad, la optimización de datos, tiempos y la flexibilidad, pero como desventajas el mayor costo y tiempo de adaptación para poder comenzar a utilizar la herramienta (PowerData, 2013).

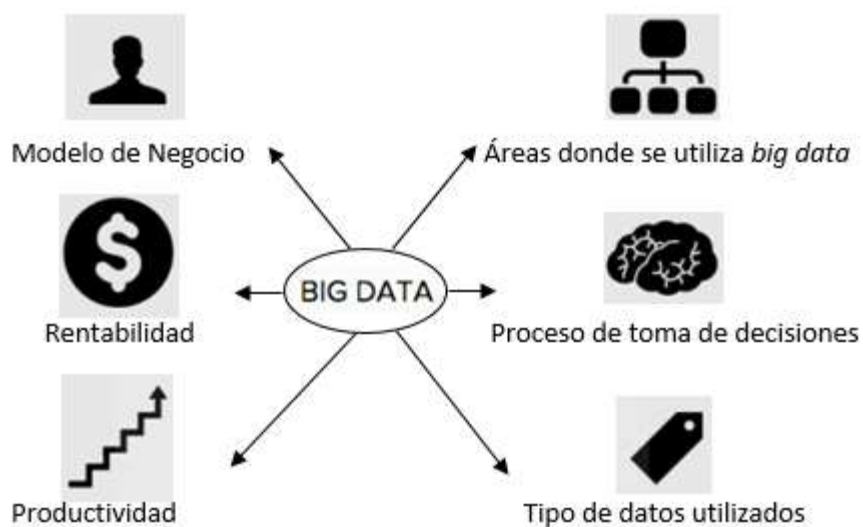
El segundo modelo de implementación es el evolutivo. Este modelo es usual en empresas ya que tienen dentro de ellas un nivel de *business intelligence* (BI) avanzado, por lo que su implementación sería una

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

evolución y no una revolución. Las ventajas de este modelo son un mayor rendimiento de la herramienta, un mayor volumen de capacidad en el momento de implementación y ahorro en costos, ya que ya poseen *business intelligence* (BI) que sirva de base para abastecer e implementar la herramienta. Los inconvenientes son que la velocidad y rendimiento de *big data* están limitados por la capacidad del *business intelligence* (BI) existente.

El tercer y último modelo de implementación es la híbrida, una mezcla entre los dos modelos anteriores. Las ventajas de este modelo son la precisión y rentabilidad. Las desventajas son su dificultad y posible resistencia para revolucionar y evolucionar al mismo tiempo. En este caso el objetivo es adaptar el *business intelligence* (BI) existente y mejorar su implementación según los requerimientos y capacidad necesaria, con la dificultad de que ambas tareas se deben realizar el mismo tiempo.

Figura 10 Variables de estudio



Fuente: Elaboración Propia

2.5.6 *BIG DATA* APLICADO EN LA GESTIÓN DE RECURSOS HUMANOS

Big data se refiere a cantidades masivas y exponencialmente crecientes de empleados, cliente y datos transaccionales disponibles en organizaciones. En el caso de los recursos humanos, Las organizaciones tienen grandes cantidades de talento o datos relacionados con las personas (por ejemplo, habilidades, calificaciones de desempeño, edad, tenencia, registro de seguridad, desempeño de ventas, antecedentes de educación, gerenciamiento, roles anteriores, etc.) que se pueden utilizar para comprender mejor la composición actual de la organización, el rendimiento y el riesgo de mejorar desarrollo de empleados, productos y servicios. *Big data* en recursos humanos se establece en evaluar y mejorar prácticas que

incluyen adquisición de talento, desarrollo, retención y desempeño organizacional general. Esto implica integrar y analizar métricas internas, puntos de referencia externos, datos de redes sociales y datos gubernamentales para ofrecer una solución más informada al problema comercial frente a su organización (Stone, <https://www.cornerstoneondemand.com/>, s.f.).

Se necesitan nuevas herramientas y tecnología porque los grandes datos son muy grandes, cambian rápidamente y están potencialmente desestructurados. Con estas herramientas, las organizaciones de recursos humanos pueden realizar análisis y pronósticos para tomar decisiones más inteligentes y precisas, medir mejor la eficiencia e identificar los "puntos ciegos" de la administración para responder preguntas importantes sobre la productividad de la fuerza laboral, el impacto de los programas de capacitación en el desempeño empresarial, predictores de desgaste de la fuerza laboral y cómo identificar líderes potenciales. La capacidad de capturar y analizar grandes datos ha permitido que muchas compañías aumenten los ingresos al comprender mejor y dirigirse con mayor precisión a los clientes y reducir los costos a través de procesos comerciales mejorados. El mayor problema para las personas de recursos humanos que gestionan el talento ha sido la falta de números y la falta de datos para poner sobre la mesa en las discusiones comerciales (Stone, <https://www.cornerstoneondemand.com/>, s.f.).

Para que el análisis de *big data* tenga éxito en el campo de recursos humanos también se deben abordar desafíos regulatorios y éticos:

- Si los datos pronosticaran que el desempeño de la empresa mejoraría si todas las minorías de ciertas etnias fueran degradadas, la gran mayoría de los gerentes de recursos humanos rechazarían la conclusión analítica por ser falsa e ilegal. Sin embargo, es probable que los problemas de discriminación con el análisis *big data* sean más sutiles y en particular pueden servir para exacerbar patrones discriminatorios previos. Incluso si los antecedentes raciales o éticos no se incluyeron en los sets de datos podría concluir que un grupo particular de colaboradores son estrellas y que la capacitación deba basarse en los antecedentes y experiencias de ese grupo. Pero si la discriminación sutil en el pasado hubiera sido tolerada por la cultura organizacional, las conclusiones del *big data analytics* podrían ser técnicamente correctas, pero solo porque ciertas minorías fueron previamente excluidas del entrenamiento efectivo (Sodeman, 2020).
- El uso intrusivo de análisis podría generar problemas de privacidad. Algunas empresas han desarrollado sistemas de seguimiento que recopilan audio, geolocalización y otros datos de los empleados durante su jornada laboral. Estos sistemas pueden identificar las acciones más productivas y las actividades de intercambio de conocimientos por parte de individuos y

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

compañeros de equipo, incluida la identificación de grupos de trabajo informales que tienen miembros en múltiples ubicaciones. Si bien estos investigadores y empleadores quizás tengan buenas intenciones en este esfuerzo, los colaboradores tienden a preocuparse por su privacidad en sus trabajos (Sodeman, 2020).

Sin embargo, todas estas preocupaciones legales y éticas, tienen un tema central: debe haber transparencia y cuidado por parte del empleador y el consentimiento de los empleados. Es probable que los empleadores cumplan con las leyes cuando recursos humanos toma las siguientes medidas: se comunica con los empleados sobre cómo y por qué se recopilan los datos; ejerce el debido cuidado de que los datos sean seguros, se utilicen para el propósito previsto y no sean discriminatorios; y obtiene el consentimiento explícito, escrito y voluntario de los empleados sobre la recopilación de datos (Sodeman, 2020).

2.5.7 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

Es el nombre que reciben las técnicas computacionales, matemáticas y estadísticas asociadas al fenómeno *big data*.

La práctica ha acordado en traducir el término *machine learning* como “aprendizaje automático”. Como se puede apreciar en la Figura 1, los métodos de aprendizaje automático caen en la frontera entre la computación y la estadística: ambas reclaman su paternidad. Su principal objetivo es explorar los datos pasados para construir el modelo de una fórmula matemática que prediga de la mejor manera el acontecimiento que estemos interesados. Una vez construido el modelo, se lo podría alimentar de información para tener como resultado una predicción sobre un posible suceso (Escudero, 2019).

“En la vieja visión de la estadística, la idea era estimar el modelo, propuesto por una teoría o tal vez por experiencia previa. El modelo venía de afuera del problema y los datos se usaban solo para estimarlo. La revolución de *machine learning* cambia por completo esta estrategia. La profusión de datos permite, construir, estimar y reevaluar el modelo a medida que se lo usa. Es la idea de aprender, en vez de estimar” (Escudero, 2019).

Cuanto más flexible sea el modelo y menos se conozca sobre él, más serán los datos que se necesiten para alimentarlo mediante datos masivos. Es aquí donde aparece *big data*.

2.6 BODEGA DE DATOS (DATA WAREHOUSE)

Una bodega de datos, según define Bill Inmon, considerado el padre del *data warehouse*, es un conjunto de datos integrados u orientados a una materia, que varían con el tiempo y que no son transitorios, los

cuales apoyan el proceso de toma de decisiones de la administración y está orientada al manejo de grandes volúmenes de datos provenientes de diversas fuentes o diversos tipos (Rainardi, 2007).

El concepto de *data warehouse* (DW) llegó de la mano de Bill Inmon. Él pensó en un único repositorio de información para poder integrar y explotar información de diversos sistemas fuentes.

Data warehouse o almacén de datos, es definido como un proceso de gestión centralizada de datos, en donde se aplican herramientas que permiten resumir, describir y analizar los datos históricos para dar soporte a la toma de decisiones (Sivanandam, 2006). Un *data warehouse* proporciona una visión global, común e integrada de los datos de la organización, independiente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios. Normalmente en el almacén de datos habrá que guardar información histórica que cubra un amplio período de tiempo. Pero hay ocasiones en las que no se necesita la historia de los datos, sino sólo sus últimos valores, siendo además admisible generalmente un pequeño desfase o retraso sobre los datos operacionales.

Podemos entender un *data mart* como un subconjunto de los datos del *data warehouse* con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. Al igual que en un *data warehouse*, los datos están estructurados en modelos de estrella o copo de nieve y un *data mart* puede ser dependiente o independiente de un *data warehouse*. El *data mart* está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. El *data warehouse* es la organización en su conjunto es el almacén natural para los datos corporativos comunes. El *data warehouse* es la organización en su conjunto, es el almacén natural para los datos corporativos comunes (Curto, 2007).

2.7 DEFINICIÓN DE ETL

Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra base de datos, *data mart*, o *data warehouse* para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

Dentro de este conjunto de herramientas encontramos las que nos ayudan a realizar una programación de la ejecución de las tareas de extracción; este factor es uno de los más relevantes debido a que se depende de la latencia en el proceso de la generación de la información, en otras palabras el tiempo que debe transcurrir entre el momento en que los datos son generados y el momento en que estos son almacenados para su procesamiento; sin embargo otro factor importante es la seguridad de los datos, lo cual se encuentra sujeto a las políticas de la empresa, dando que si se va a realizar una integración de los datos

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

en los múltiples sistemas ya sea en una bodega o un depósito, se debe prevenir la vulnerabilidad de este requiriendo una planeación muy cuidadosa.

Otro aspecto que se debe tener en cuenta en estas herramientas de extracción, transformación y carga es el punto de partida de la generación de información, la calidad de los datos que se van almacenar es muy relevante y demasiado crítica para la generación de información que sea útil y coherente para la organización.

En idioma inglés su acrónimo es *extract, transform and load* (extraer, transformar y cargar), frecuentemente abreviado en ETL también se pueden utilizar para la integración con sistemas heredados (Wikipedia, 2012).

2.8 MINERÍA DE DATOS (DATAMINING)

2.8.1 OBJETIVO DE LA MINERÍA DE DATOS

Seleccionar las variables de los colaboradores para poderlos agrupar y asociar, de esa forma detectar posibles tendencias o patrones de comportamiento relacionadas con el desempeño laboral de los empleados.

Una de las ventajas de contar con bases de datos es la posibilidad de descubrir información de interés y adquirir conocimiento mediante el análisis de los datos. Sin embargo, el volumen mismo de las bases de datos es con frecuencia una limitante para análisis manuales por lo que se han desarrollado tecnologías especializadas que faciliten su manejo.

La minería de datos es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. A continuación, algunas de las definiciones que han dado diferentes autores para la minería de datos son:

La minería de datos es el proceso de descubrir patrones en los datos. Los datos se presentan en grandes cantidades. Los patrones descubiertos deben ser significativos de manera que se permitan ventajas, por lo general, de tipo económicas (Frank, 2005).

La minería de datos es el proceso eficiente, no trivial, de extraer información valiosa (patrones y tendencias) de una gran colección de datos (Sivanandam, 2006).

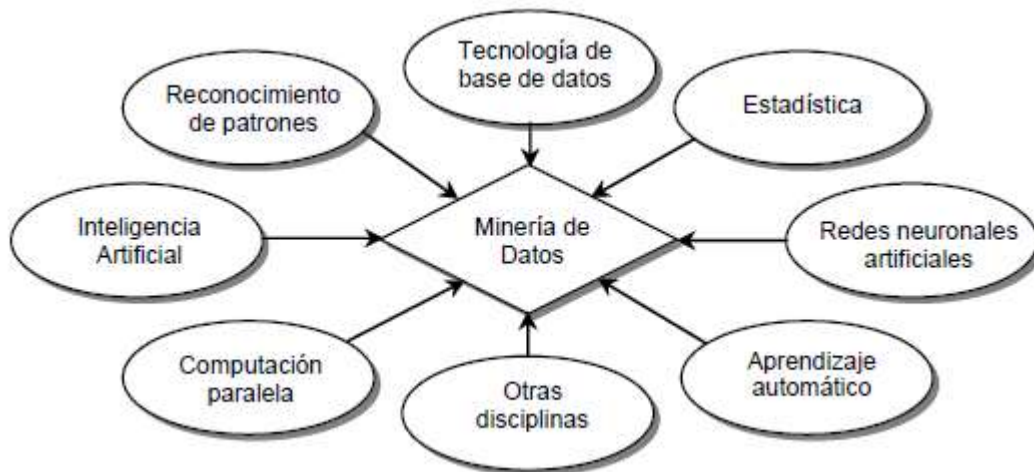
La minería de datos es el análisis de conjuntos de datos observados, a menudo extensos, para encontrar relaciones insospechadas y resumir los datos en forma comprensible y útil para el usuario de la información (Smyth, 2001).

La minería de datos es la exploración y análisis de grandes cantidades de datos para descubrir un resultado particular y patrones significativos (Linoff, 2004).

Los autores encuentran en estas definiciones datos en común en la extracción de información útil, estos mismos son patrones, asociaciones y relaciones entre los datos para los usuarios finales. Debido a las muchas formas de definir la minería de datos, otros autores resumen el objetivo de la minería de datos como el proceso de convertir datos extraídos de grandes volúmenes de datos en conocimiento útil. (Ramírez, 2004). El establecimiento de patrones es una práctica útil, puesto que permite explicar algo con relación a los datos que se tienen (Frank, 2005).

La minería de datos facilita el relevar patrones existentes en el conjunto de datos, los cuales generan información válida, novedosa, útil, comprensible y explotable. Estos factores han llevado a considerar a la minería de datos como un campo interdisciplinario, el cual es desarrollado de manera conjunta o como apoyo de otras disciplinas como: tecnología de bases de datos, sistemas para la toma de decisiones, estadística, inteligencia artificial, aprendizaje automático, redes neuronales, reconocimiento de patrones y visualización de datos, entre otras.

Figura 11 Asociación de diversas disciplinas de minería de datos



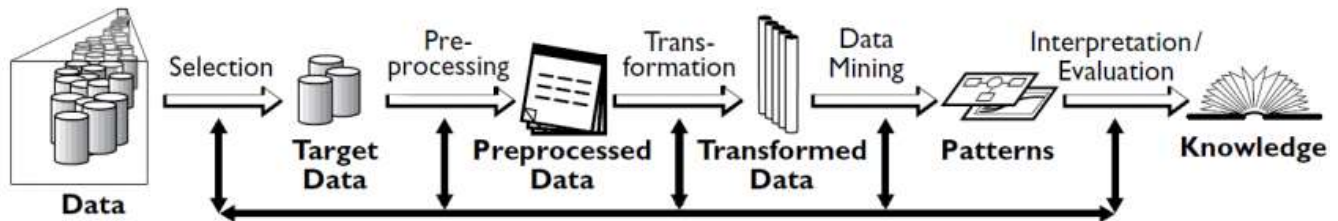
Fuente: Adaptado de (Ramírez, 2004)

Una aplicación de minería de datos podría identificar varios grupos en los datos que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Ni la recolección de datos, preparación de datos, ni la interpretación de los resultados y la información son parte de la etapa de minería de datos, pero pertenecen al proceso de KKD (Knowledge Discovery in Database) como pasos adicionales.

Figura 12 Knowledge Discovery in Database



Fuente: Adaptado de (Ramírez, 2004)

Los términos relacionados con el dragado de datos, la pesca de datos y el espionaje de los datos se refieren a la utilización de métodos de minería de datos a las partes de la muestra que son (o pueden ser) demasiado pequeños para las inferencias estadísticas fiables que se hicieron acerca de la validez de cualquiera de los patrones descubiertos. Estos métodos pueden, sin embargo, ser utilizados en la creación de nuevas hipótesis que se prueban contra las poblaciones de datos más grandes.

Según (Ramírez, 2004) define que un proceso típico de minería de datos consta de los siguientes pasos generales:

- Selección del conjunto de datos:

Tanto en lo que refiere a las variables objetivo (aquellas que se quieren predecir, calcular o inferir), como las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.

- Análisis de las propiedades de los datos:

En especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).

- Transformación del conjunto de datos de entrada:

Se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema. A este paso también se le conoce como pre-procesamiento de los datos.

- Seleccionar y aplicar la técnica de minería de datos:

Se construye el modelo predictivo, de clasificación o segmentación.

- Extracción de conocimiento:

Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre procesamiento diferente de los datos.

- Interpretación y evaluación de datos

Una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

Si el modelo final no supera la evaluación, el proceso se podría repetir desde el principio o, si el experto lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido. Una vez validado el modelo, éste ya está listo para su exploración.

Los modelos obtenidos por técnicas de minería de datos se aplican incorporándolos en los sistemas de análisis de información de las organizaciones e incluso, en los sistemas transaccionales.

2.8.2 TIPOS DE MINERÍA DE DATOS

La minería de datos ha sufrido transformaciones en los últimos años de acuerdo con los cambios tecnológicos, de estrategias de marketing, la extensión de los modelos de compra en línea, etc. Los más importantes de ellos son:

- La importancia que han cobrado los datos no estructurados (texto, páginas de Internet).
- La necesidad de integrar los algoritmos y resultados obtenidos en sistemas operacionales, portales de internet, etc.
- La exigencia de que los procesos funcionen prácticamente en línea (por ejemplo, en casos de fraude con una tarjeta de crédito).

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

- Los tiempos de respuesta. El gran volumen de datos que hay que procesar en muchos casos para obtener un modelo válido es un inconveniente pues esto implica grandes cantidades de tiempo de proceso y hay problemas que requieren una respuesta en tiempo real.

2.8.2.1 PREDICCIÓN

Muchas formas de minería de datos son predictivas. Por ejemplo, un modelo podría predecir el ingreso basado en la educación y otros factores demográficos. Las predicciones tienen una probabilidad asociada y las probabilidades de predicción son también conocidas como confianza. Algunas formas de minería de datos predictiva generan reglas, las cuales son condiciones que implican una salida dada. Por ejemplo, una regla podría especificar que una persona que tiene un grado universitario y vive en cierto barrio o ciudad, probablemente tiene un ingreso mayor al promedio en la región

2.8.2.2 AGRUPACIÓN

La agrupación es otra forma en la que la minería de datos identifica grupos naturales en los datos. Por ejemplo, un modelo podría identificar el segmento de la población que tiene un ingreso dentro de un rango específico, que tiene un buen registro de manejo (en un sistema de puntuación o *scoring* de conductores), y que compran un nuevo automóvil.

El agrupamiento ha sido reconocido como un método eficaz para tratar problemas de clasificación de variables de estudio para la búsqueda de patrones y tendencias (Rodrigo, 2002).

Existen dos tipos principales de agrupamiento (Larose, 2005); el agrupamiento jerárquico que se caracteriza por el desarrollo recursivo de una estructura en forma de árbol; y el agrupamiento particional que organiza los registros dentro de k grupos. Los métodos particionales tienen ventajas en aplicaciones que involucran gran cantidad de datos y los jerárquicos, por lo general, se utilizan para identificar el número deseado de grupos.

2.8.2.3 FUNCIONES Y TAREAS DE LA MINERÍA DE DATOS

La minería de datos se apoya en la aplicación de métodos matemáticos de análisis, utilizando diferentes algoritmos y técnicas de clasificación, tales como *clustering*, regresión inteligencia artificial, redes neuronales, reglas de asociación, árboles de decisión, algoritmos genéticos, entre otras, que son de gran utilidad para para llevar a cabo el análisis inteligente de grandes volúmenes de información digital.

Estos elementos permiten que los algoritmos aprendan conceptos por sí solos, sin tener que ser programados, es decir, se trata de un conjunto de reglas abstractas que por sí solas son construidas, lo que se ha conseguido es una “autoconfiguración” propiamente dicha.

- Problemas de clasificación

Cuando la variable a predecir es un conjunto de estados discretos o categóricos de los cuales a su vez pueden ser:

- Binaria: {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}, etc.
- Múltiple: Comprar {Producto1, Producto2...}, etc.
- Ordenada: Riesgo {Bajo, Medio, Alto}

Al igual que la clasificación, la regresión es una tarea de aprendizaje inductivo que ha sido ampliamente estudiada y utilizada. Los modelos de regresión predicen variables y valore numéricos, por ejemplo, las ventas de una empresa a partir de los precios a fijar.

Estos métodos proporcionan una forma directa de descubrir y representar la similitud en los datos, formando conjuntos de datos similares entre ellos. Una vez que se crean los grupos, los algoritmos pueden predecir a qué conjunto corresponde una nueva instancia, lo que les permite implementar análisis predictivos.

Tabla 2 Tipología de los algoritmos de minería de datos

METODOS	SUPERVISADOS		NO SUPERVISADOS
	Clasificación	Regresión	Agrupamiento
Agrupamiento Jerárquico			X
K-means y derivados			X
k-NN	X		
SVM	X	X	
Redes Neuronales	X	X	
Árboles de decisión	X	X	
Métodos probabilísticos	X	X	

Fuente: Adaptado de (Caihuelas, 2017)

2.8.2.3.1 MINERÍA DE DATOS SUPERVISADA

“Los métodos supervisados son algoritmos que basan su proceso de aprendizaje en un juego de datos de entrenamiento convenientemente etiquetado. Por etiquetado entendemos que para cada ocurrencia del juego de datos de entrenamiento conocemos el valor de su atributo objetivo o clase. Esto le permitirá al algoritmo poder deducir una función capaz de predecir el atributo objetivo para un juego de datos nuevos.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Las grandes familias de algoritmos de aprendizaje supervisado son: Algoritmos de clasificación y Algoritmos de regresión” (Caihuelas, 2017, pág. 39).

El aprendizaje supervisado es una técnica para deducir una función a partir de datos de entrenamiento que consisten en pares de objetos normalmente vectores; datos de entrada y los resultados deseados. La precisión de la función aprendida depende en gran medida de cómo el objeto de entrada está representado. La salida de la función puede ser un valor numérico como en los problemas de regresión o una etiqueta de clase como en los de clasificación. El objetivo de este aprendizaje es el de crear una función capaz de predecir el valor correspondiente a cualquier objeto de entrada válida después de haber ajustado los parámetros y haber visto una serie de ejemplos con los datos de entrenamiento.

El desempeño del algoritmo se puede medir utilizando un conjunto de pruebas independiente del entrenamiento que se usa con el mismo. Los algoritmos más utilizados son las redes neuronales, las máquinas de vectores soporte; el algoritmo de K-vecinos más cercanos (*KNN*), los modelos de mixturas, el clasificador bayesiano ingenuo, los árboles de decisión y las funciones de base radial.

2.8.2.3.1.1 KNN VECINOS MÁS CERCANOS

Es un algoritmo de aprendizaje supervisado del tipo de clasificación, con los parámetros y los datos de entrenamiento su objetivo será clasificar correctamente todas las instancias nuevas. En contraste con los algoritmos de aprendizaje supervisado, el algoritmo KNN no genera un modelo fruto del aprendizaje con datos de entrenamiento, sino que el aprendizaje sucede en el mismo momento en el que se pide clasificar una nueva instancia. A este tipo de algoritmos se les llama métodos de aprendizaje perezoso o *lazy learning method*, en inglés. El funcionamiento del algoritmo es muy simple. Para cada nueva instancia a clasificar, se calcula la distancia con todas las instancias de entrenamiento y se seleccionan las *k* instancias más cercanas.

2.8.2.3.1.2 MÁQUINAS DE SOPORTE VECTORIAL

Las máquinas de soporte vectorial en inglés (*support vector machines* o SVM) son algoritmos de aprendizaje supervisado capaz de resolver problemas de clasificación, tanto lineales como no lineales. Actualmente es considerado uno de los algoritmos más potentes en reconocimiento de patrones. Su eficiencia y los buenos resultados obtenidos en comparación con otros algoritmos han convertido esta técnica en la más utilizada en campos como el reconocimiento de textos y habla, predicción de series temporales y estudios sobre bases de datos de marketing, entre otros, pero también en otros campos como la secuenciación de proteínas y el diagnóstico de varios tipos de cáncer.

Las SVM utilizan las técnicas de optimización cuadrática propuestas para buscar maximizar el margen entre los puntos pertenecientes a los distintos grupos a clasificar, de tal forma que el máximo número de futuros puntos queden bien clasificados. El objetivo de las SVM es encontrar el hiperplano óptimo que maximiza el margen entre clases del conjunto del set de datos de entrenamiento.

2.8.2.3.2 MINERÍA DE DATOS NO SUPERVISADA

Dentro de la minería de datos, las técnicas y métodos que usan los algoritmos no supervisados son utilizados para la detección de patrones ocultos en la información y fuentes de datos de la empresa. Dichos patrones representan por si mismos información útil que puede ser utilizada directamente en la toma de decisiones. Este tipo de métodos no supervisados se pueden clasificar en tres grandes grupos, dependiendo del esquema de organización interna que empleen para actualizar las características de las neuronas de la red:

- Métodos de aprendizaje basados en la regla de Hebb. Una de las primeras aproximaciones al aprendizaje no supervisado fue la realizada por Hebb que ha supuesto una de las aportaciones fundamentales en el campo de las redes de neuronas. Basados en ella, se han propuesto diversos tipos de reglas hebbianas empleadas con diversos fines, aunque una de las aplicaciones más importantes es en el análisis de componentes principales.
- Métodos de aprendizaje competitivos. En este tipo de sistemas el objetivo es agrupar o categorizar los datos que sean similares. Generalmente, se trata de similitud geométrica en el espacio de entrada. Estos modelos utilizan redes neuronales que compiten para ser activadas. A diferencia de los métodos anteriores, donde varias neuronas pueden ser activadas al mismo tiempo, en este tipo de sistemas sólo una de ellas será la ganadora. Por tanto, la red tratará de agrupar los patrones similares que serán representados por una neurona de la red.
- Métodos de aprendizaje basados en modelos de teoría de la información. Un esquema de aprendizaje basado en los principios de la teoría de la información. Este tipo de sistemas trata de maximizar la cantidad de información que se conserva en el procesamiento de los datos minimizando la entropía. Este tipo de métodos tiene multitud de aplicaciones prácticas, entre ellas, el análisis de componentes principales y sistemas de compresión de información.

La mayoría de los métodos descritos antes tienen como objetivo seleccionar las mejores reglas de asociación teniendo en cuenta los factores de soporte y de confianza para obtener patrones consistentes e interesantes.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

2.8.2.3.2.1 CLUSTERING

Este algoritmo intenta encontrar patrones en los datos identificando principalmente los que se forman como grupos separados. Esos criterios son por lo general distancia o similitud. Cada grupo encontrado se puede usar para realizar una segmentación de clientes/empleados y de esa forma ofrecer así distintos productos a cada grupo. El objetivo principal es encontrar grupos de instancias que en un *cluster* sean similares entre sí y diferentes de las otras en otros *clusters*. Otra posible aplicación es agrupar documentos por temática, donde cada *cluster* o grupo pertenece a un tipo de documento, sin embargo, puede usarse como paso previo a otras técnicas de aprendizaje. Algunos ejemplos son exploración de datos y pre procesamiento de datos.

2.8.2.3.2.2 K-MEANS

El algoritmo K-medias es probablemente el algoritmo de agrupamiento más conocido. El algoritmo está basado en la minimización de la distancia interna (la suma de las distancias de los patrones asignados a un agrupamiento al centroide de dicho agrupamiento). De hecho, este algoritmo minimiza la suma de las distancias al cuadrado de cada patrón al centroide de su agrupamiento.

El algoritmo es sencillo y eficiente. Además, procesa los patrones secuencialmente (por lo que requiere un almacenamiento mínimo).

Sin embargo, está sesgado por el orden de presentación de los patrones (los primeros patrones determinan la configuración inicial de los agrupamientos) y su comportamiento depende enormemente del parámetro K.

El algoritmo de K-medias, también llamado algoritmo de Lloyd en la comunidad informática es un método de *clustering* o agrupamiento que consiste en clasificar un conjunto de elementos con base en propiedades de estos de forma iterativa.

Otra característica importante es que emplea la media estadística para el cálculo los nuevos *clusters*. Concretamente, se calcula del *cluster* su centroide; esto es el punto resultante de la media de todos los elementos asignados al *cluster* en cuestión. Los dos pasos principales del procedimiento son:

Paso de asignación:

Se asigna a cada elemento el *cluster* o grupo más cercano, aplicando una función para el cálculo de distancia.

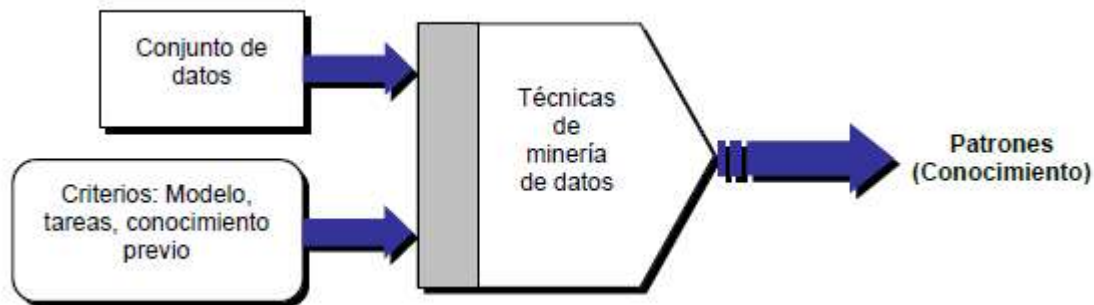
Paso de actualización:

Consistente en calcular los nuevos centroides para cada clúster. Aunque el algoritmo presenta una complejidad computacional difícil, existen heurísticos que hacen que converja rápidamente a un óptimo local.

2.8.2.3.3 TÉCNICAS DE MINERÍA DE DATOS

Las técnicas de *machine learning* nos abren una amplia ventana al mundo de los datos, ofreciéndonos grandes posibilidades para mejorar la eficiencia y el rendimiento de nuestro negocio y elevar, en consecuencia, la competitividad de nuestra empresa. En este sentido, apenas hemos comenzado a vislumbrar el gran potencial que nos ofrece el aprendizaje automático para abordar problemas reales y crear modelos que proporcionen predicciones de alto valor, permitiendo, de esta forma, tomar decisiones más adecuadas y desarrollar mejores acciones de negocio.

Figura 13 Proceso de extracción de patrones mediante el uso de técnicas de minería de datos



Fuente: Adaptado de (Ramírez, 2004)

Por tanto, es conveniente aprovechar la oportunidad de obtener información valiosa a partir de nuestros datos y no perder de vista las ventajas competitivas que nos puede brindar el asomarnos al extenso e incipiente universo de *machine learning*. La técnica más utilizada en minería de datos es la clasificación que emplea métodos como árboles de decisión o redes neuronales. Cada proceso de clasificación que se realiza implica un aprendizaje y una propia clasificación (Larose, 2005).

Ese aprendizaje es donde se entrenan los datos mediante los diferentes algoritmos, para posteriormente realizar las pruebas y comprobar resultados. En esta etapa del proyecto seleccionamos los algoritmos posibles que nos ayudarán a determinar los factores que afectan el aprovechamiento académico.

2.8.2.3.3.1 REDES NEURONALES

Son un paradigma del aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir estímulo de salida.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Una red neuronal es básicamente una interconexión de neuronas que trabajan entre sí para poder producir una salida, en la cual se generan los procesos necesarios asociados al aprendizaje como respuesta al estímulo generado en el ambiente (Larose, 2005).

“Una red neuronal es un procesador masivamente paralelo distribuido que es propenso, por naturaleza, a almacenar conocimiento experimental y hacerlo disponible para su uso” (Haykin, 2009). Con la ayuda de las redes neuronales se puede llegar a identificar factores en los alumnos con buenas o malas perspectivas de aprovechamiento académico. Calcular la probabilidad de que un alumno pueda desertar. Clasificar los diferentes atributos de los alumnos y explorar los factores relacionados.

2.8.2.3.3.2 ALGORITMO DE RED NEURONAL DE SQL (MICROSOFT)

En SQL Server Analysis Services el algoritmo de red neuronal combina cada posible estado del atributo de entrada con cada posible estado del atributo de predicción y usa los datos de entrenamiento para calcular las probabilidades. Posteriormente puede usar estas probabilidades para la clasificación o la regresión, así como para predecir un resultado del atributo de predicción basándose en los atributos de entrada.

Los modelos de minería de datos construidos con este algoritmo pueden contener varias redes, en función del número de columnas que se utilizan para la entrada y la predicción, o solo para la predicción. El número de redes que contiene un único modelo de minería de datos depende del número de estados que contienen las columnas de entrada y las columnas de predicción que utiliza el modelo.

Las redes artificiales de neuronas tratan, en cierto modo, de replicar el comportamiento del cerebro, donde tenemos millones de neuronas que se interconectan en red para enviarse mensajes unas a otras. Esta réplica del funcionamiento del cerebro humano es uno de los modelos de moda por las habilidades cognitivas de razonamiento que adquieren.

El reconocimiento de imágenes o videos, por ejemplo, es un mecanismo complejo para mostrar una red neuronal para hacer (Haykin, 2009).

2.8.2.3.3.3 AGRUPAMIENTO O *CLUSTERING*

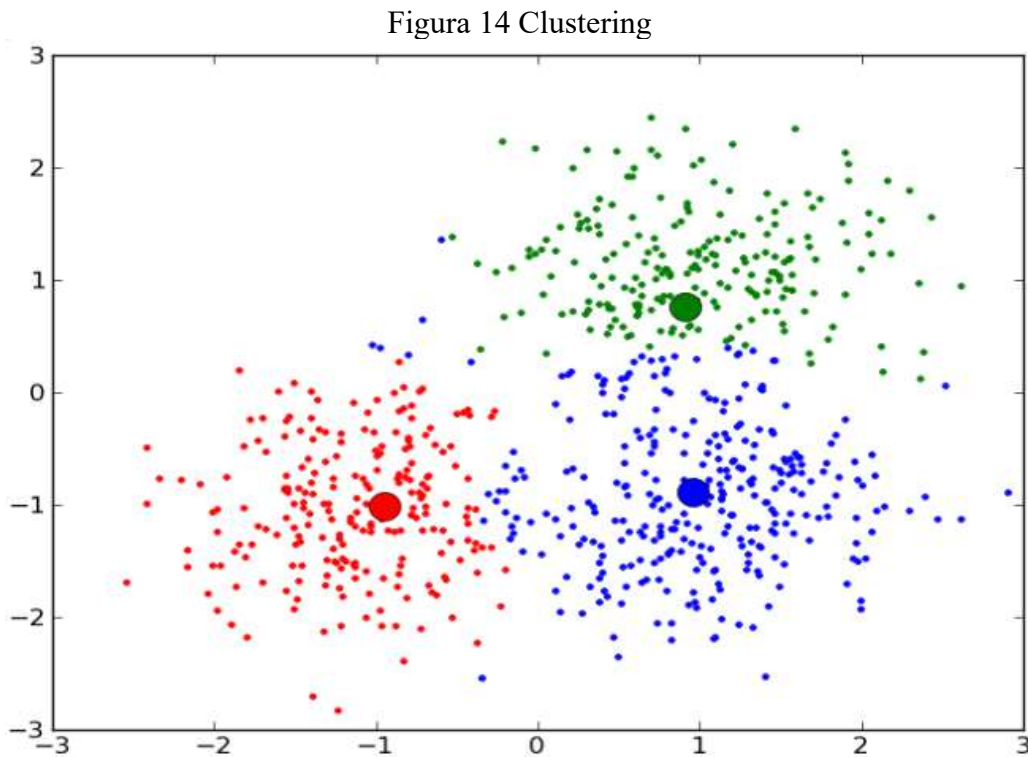
Un algoritmo de agrupamiento (en inglés, *clustering*) es un proceso de agrupación de una serie de vectores que utiliza técnicas interactivas para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

Se considera al agrupamiento como, la tarea de dividir una población heterogénea en un número de subgrupos homogéneos de acuerdo con las similitudes de sus registros (Linoff, 2004). Dentro de esta tarea existen dos tipos principales de agrupamiento (Larose, 2005): el jerárquico que se caracteriza por el desarrollo recursivo de una estructura en forma de árbol, y el particional que organiza los registros dentro de k grupos.

2.8.2.3.3.4 CLASIFICACIÓN

Es una de las principales tareas en el proceso de minería de datos que se emplea para asignar datos a un conjunto predefinido de variables. El objetivo, es encontrar algún tipo de relación entre los atributos de entrada y los registros de salida para comprender el comportamiento de los datos, así mediante el conocimiento extraído se puede predecir el valor de un registro desconocido (Sivanandam, 2006).

Es la técnica que permite descubrir grupos de elementos con similitudes sin conocer la estructura de los datos. Sin embargo, el mayor problema de la clasificación es que muchas veces no es representativo y no proporciona un conocimiento detallado, solo otorga predicciones.



Fuente: Elaboración propia

2.8.2.3.3.5 REGRESIÓN

La regresión es el aprendizaje de una función cuyo objetivo es predecir valores de una variable continua a partir de la evolución de otra variable también continua, la cual por lo general es el tiempo (Sivanandam,

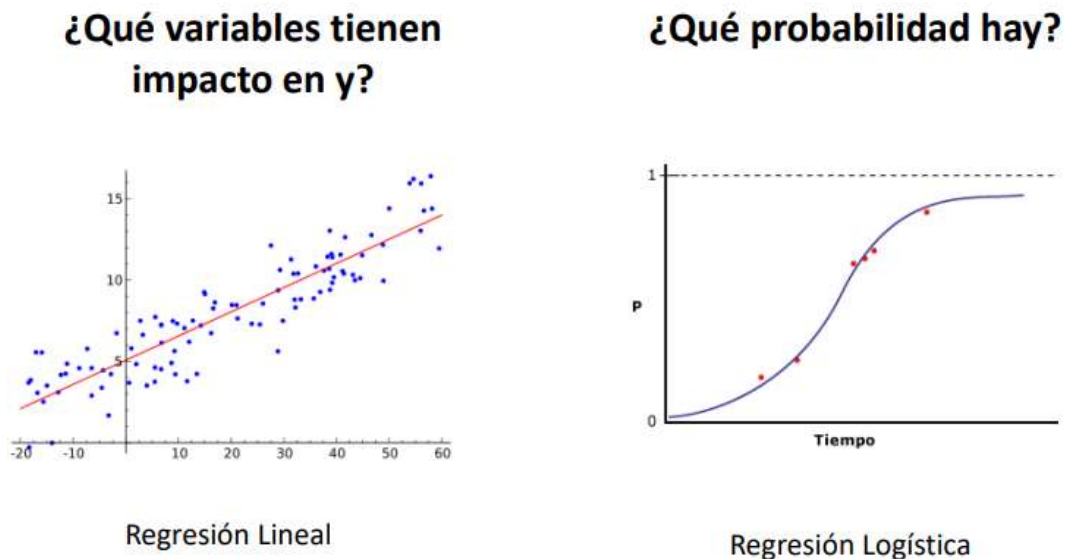
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

2006). En la regresión, la información de salida es un valor numérico continuo o un vector con valores no discretos. Esta es la principal diferencia respecto de la clasificación donde el valor a predecir es numérico.

Incluyen diversas técnicas estadísticas para modelar y analizar diversas variables, buscando explicar la relación entre una variable dependiente y una o más variables independientes, comprendiendo el impacto de éstas en la variable a explicar.

Si solo se dispone de una variable definida se trata de un problema de regresión simple, mientras que si se dispone de varias variables se trata de un problema de regresión múltiple. A esta tarea también se le conoce como: interpolación, cuando el valor o valores predichos están en medio de otros; o estimación, cuando se predice valores futuros (Ramírez, 2004).

Figura 15 Regresión, ejemplos regresión lineal y regresión logística



Fuente: Elaboración propia

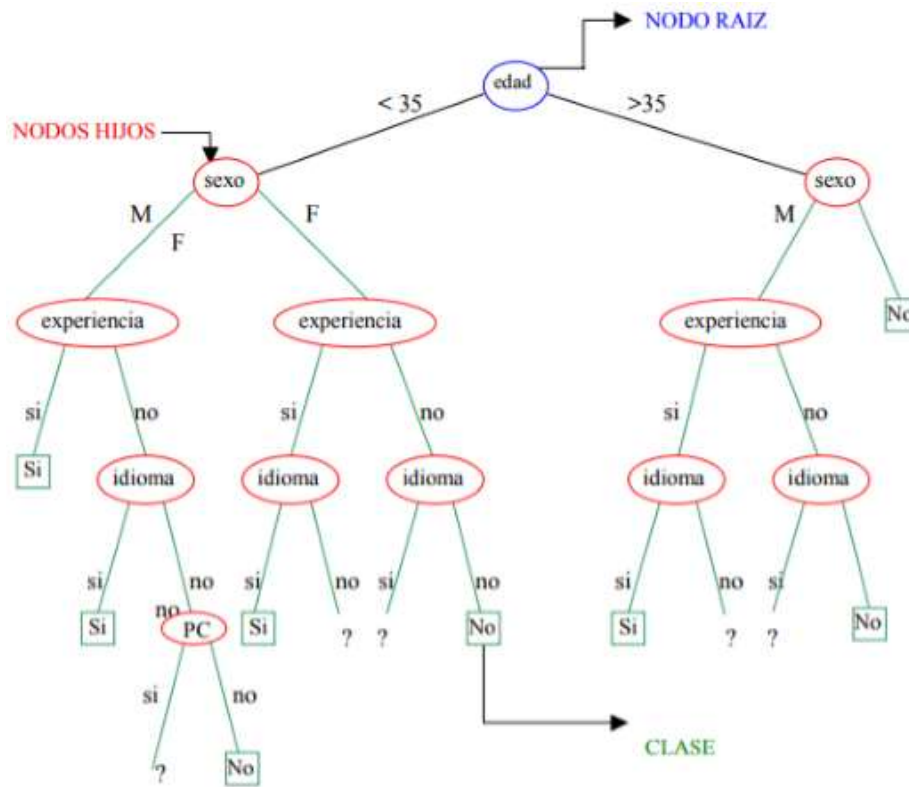
2.8.2.3.3.6 ÁRBOLES DE DECISIÓN

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial. Dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.

Los árboles de decisión son una técnica de minería de datos que establece un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas (Davidson, 2007).

Se puede decir que los árboles de decisión se adecuan más a la clasificación para poder determinar las clases que se puedan generar, y por tal motivo poder identificar a que clase pertenece un objeto.

Figura 16 Árbol de decisión, ejemplo entrevista laboral



Fuente: Elaboración propia

A diferencia de los métodos lineales, pueden representar relaciones no lineales para resolver problemas. Al ser más precisos y elaborados, se gana en capacidad predictiva, pero se pierde en rendimiento.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Tabla 3 Relación existente entre las tareas y técnicas de minería de datos

Nombre	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones
Redes neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
K-medias			✓		
A priori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis de comp. principales					✓
Two-step, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Maquinas de soporte vectorial	✓	✓	✓		
Reglas CN2 (cobertura)	✓			✓	
Análisis discriminante multivariante	✓				

Fuente: (Ramírez, 2004)

En la tabla 3 se observa que todas las áreas pueden ser resueltas por más de una técnica, y que algunas técnicas pueden resolver más de una tarea, como es el caso de las redes neuronales, los árboles de decisión y las regresiones (lineal, logarítmica y logística).

2.8.2.3.4 TIPOS DE MODELOS PARA ANÁLISIS PREDICTIVOS

El análisis predictivo agrupa una variedad de técnicas estadísticas de modelización, aprendizaje automático y minería de datos que analiza los datos actuales e históricos reales para hacer predicciones acerca del futuro o acontecimientos no conocidos.

En el ámbito de los negocios los modelos predictivos extraen patrones de los datos históricos y transaccionales para identificar riesgos y oportunidades. Los modelos predictivos identifican relaciones entre diferentes factores que permiten valorar riesgos o probabilidades asociadas sobre la base de un conjunto de condiciones, guiando así al decisor durante las operaciones de la organización.

El efecto funcional que pretenden estas iniciativas técnicas es que el análisis predictivo provea una puntuación (probabilidad) para cada sujeto (cliente, empleado, paciente, producto, vehículo, componente, máquina y otra unidad dentro de la organización) con el objeto de determinar, informar o influir procesos en la organización en el que participen un gran número de sujetos, tal y como ocurre en marketing, evaluación de riesgo crediticio, detección de fraudes, fabricación, salud y operaciones gubernamentales como el orden público.

Cuando se habla de análisis predictivo generalmente se quiere hablar de “modelos predictivos”, datos de puntuaciones sobre la base de modelos predictivos y previsiones. No obstante, se está generalizando el uso del término para relacionarlo con disciplinas analíticas y está muy extendido su uso para la segmentación entre usuarios de negocio y decisores.

2.8.2.3.4.1 MODELOS PREDICTIVOS

Los modelos predictivos son modelos de la relación entre el rendimiento específico de un sujeto en una muestra y uno o más atributos o características del mismo sujeto. El objetivo del modelo es evaluar la probabilidad de que un sujeto similar tenga el mismo rendimiento de una muestra diferente. Esta categoría engloba modelos en muchas áreas, como el marketing, donde se buscan patrones de datos ocultos que respondan preguntas sobre el comportamiento de los clientes o modelos de detección de fraude.

Los modelos predictivos a menudo ejecutan cálculos durante las transacciones en curso, por ejemplo, para evaluar el riesgo o la oportunidad de un cliente o transacción en particular, de forma que aporte conocimiento a la hora de tomar una decisión. Gracias a los avances en el análisis de grandes volúmenes de datos estos modelos son capaces de simular el comportamiento humano frente a estímulos o situaciones específicas (Nyce, 2007).

2.8.2.3.4.2 MODELOS DESCRIPTIVOS

En el modelo descriptivo se identifican patrones que describen los datos mediante tareas, por ejemplo, agrupamiento (*clustering*) y reglas de asociación. Mediante este modelo se identifican patrones que explican o resumen el conjunto de datos, siendo estos útiles para explorar las propiedades de los datos examinados.

Los modelos descriptivos siguen un tipo de aprendizaje no supervisado, que consiste en adquirir conocimiento desde los datos disponibles, sin requerir influencia externa que indique un comportamiento deseado al sistema (Sivanandam, 2006). Los modelos descriptivos cuantifican las relaciones entre los datos de manera que es utilizada a menudo para clasificar clientes o contactos en grupos. A diferencia de los modelos predictivos que se centran en predecir el comportamiento de un cliente en particular, los modelos descriptivos identifican muy diferentes relaciones entre los clientes y los productos.

Los modelos descriptivos no clasifican u ordenan a los clientes por su probabilidad de realizar una acción particular de la misma forma en la que lo hacen los modelos predictivos. Sin embargo, los modelos

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

descriptivos, pueden ser utilizados por ejemplo para asignar categorías a los clientes según su preferencia en productos o su franja de edad.

2.8.2.3.4.3 MODELOS DE DECISIÓN

Los modelos de decisión describen la relación entre todos los elementos de una decisión; es decir, los datos conocidos incluyendo los resultados de los modelos predictivos, la decisión y el pronóstico de los resultados de una decisión con el objeto de predecir los resultados de una decisión involucrando una gran cantidad de variables. Estos modelos pueden ser utilizados en la optimización o maximización de determinados resultados al mismo tiempo que otros son minimizados. Los modelos de decisión son generalmente usados para el desarrollo de la decisión lógica o conjunto de reglas de negocio que deberían producir el resultado deseado para cada cliente o circunstancia (Coker, 2014).

2.8.2.3.4.4 CARACTERÍSTICAS DE LAS HERRAMIENTAS DE ANÁLISIS PREDICTIVO

Utilizar análisis predictivo y entender sus resultados requería de los usuarios habilidades muy avanzadas en el pasado. Con la popularización del análisis avanzado las herramientas más modernas han mejorado sustancialmente el acceso a todo tipo de usuarios analistas.

“Organizaciones de todas las industrias se están animando a darle valor al análisis predictivo. Con el crecimiento de su demanda, los proveedores de análisis predictivo están proveyendo herramientas que reducen la barrera y aumentan el atractivo a aquellos con menos habilidades estadísticas” (Curran, 2015). Al mismo tiempo que el análisis predictivo es utilizado en los procesos de toma de decisiones y es integrado en los procesos de las organizaciones, también se está operando un cambio en el mercado con relación al tipo de usuario que consume la información siendo cada vez más relevante el papel del usuario de negocio.

Los usuarios de negocio demandan herramientas que ellos mismos puedan utilizar con autonomía y los proveedores están creando nuevas aplicaciones que eliminan la complejidad matemática, proviniendo interfaz gráfica fácil de usar e incorporan accesos rápidos para, por ejemplo, reconocer qué tipo de datos están disponibles o sugerir un análisis predictivo apropiado (Fern, 2011).

Las plataformas de análisis predictivo buscan con empeño la sofisticación que les permita analizar y visualizar la información oculta en sus datos de forma que cualquier cargo en la organización pueda extraer conocimiento útil para la toma de decisiones. La visualización de los datos se hace habitualmente

a través de tablas, cuadros, gráficos y clasificaciones que muestran visualmente las clasificaciones más probables resultadas de las predicciones.

2.8.3 METODOLOGÍAS DE MINERÍA DE DATOS

Las metodologías permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial. Estas metodologías ayudan a entender el proceso de descubrimiento de conocimiento para proveer una guía de planificación y ejecución en los proyectos. Algunos modelos conocidos como metodologías son en realidad un modelo de proceso definiendo un conjunto de actividades y tareas organizadas para llevar a cabo un trabajo.

Por esta razón, diversas empresas y consultorías en el mundo han desarrollado metodologías de trabajo para guiar al usuario para una aplicación exitosa de las técnicas de minería de datos a través de una sucesión de pasos (Sivanandam, 2006).

La diferencia fundamental entre metodología y modelo de proceso radica en que el modelo de proceso establece qué hacer y la metodología especifica cómo hacerlo. Dentro de las metodologías más adecuadas para la planificación de proyectos se encuentran las siguientes:

2.8.3.1 METODOLOGÍA SEMMA

SEMMA (Sampling, Exploration, Modification, Modeled, Assessment) desarrollado por la empresa Statistical Analysis Systems Inc. (SAS Institute), quien la define como: “el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos”.

Las fases de SEMMA (Santos, 2008) y las tareas relacionadas son las siguientes:

- Muestra:

El proceso comienza con el muestreo de datos, por ejemplo, seleccionando el conjunto de datos para modelar.

El conjunto de datos debe ser lo suficientemente grande como para contener suficiente información para recuperar, pero lo suficientemente pequeño como para ser utilizado de manera eficiente. Esta fase también se ocupa de la partición de datos.

- Explora:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Esta fase cubre la comprensión de los datos al descubrir relaciones anticipadas e imprevistas entre las variables, y también anomalías, con la ayuda de la visualización de datos.

- Modificar

La fase de modificación contiene métodos para seleccionar, crear y transformar variables en preparación para el modelado de datos.

- Modelo

En la fase de Modelo, la atención se centra en la aplicación de diversas técnicas de modelado (extracción de datos) en las variables preparadas con el fin de crear modelos que posiblemente proporcionen el resultado deseado.

- Evaluar:

La última fase es Evaluar. La evaluación de los resultados del modelado muestra la fiabilidad y la utilidad de los modelos creados.

2.8.3.2 METODOLOGÍA KDD

La extracción de conocimiento está principalmente relacionada con el proceso de descubrimiento conocido como Knowledge Discovery in Databases (KDD), que descubre conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información.

No es un proceso automático, es un proceso repetitivo que explora volúmenes muy grandes de datos para determinar relaciones. Es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos. “El proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos” (Zytkow, 2002).

Se definen las etapas del proceso del descubrimiento de información:

- Selección de datos: En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.
- Pre-procesamiento: Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes

o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.

- **Transformación:** Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.
- **Minería de datos:** Es la fase de modelamiento propiamente en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u —ocultos‖ en los datos.
- **Interpretación y evaluación:** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

KDD es un proceso iterativo e interactivo dividido en una secuencia de pasos, de los cuales la minería de datos es considerada como uno de los pasos más importantes en todo el proceso, definiéndolo como "el corazón del proceso KDD mismo" (Fayyad, 1997).

2.8.3.3 METODOLOGÍA CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining). Esta metodología se encuentra dentro de las más utilizadas para la elaboración de proyectos de minería de datos, está basado en actividades ordenadas en seis fases que recorren todo el proceso de minería de datos, desde la definición de los objetivos del negocio que se pretende obtener, hasta la vigilancia y el mantenimiento del modelo que se proponga e implemente (Chapman, 2000).

La metodología consta de cuatro niveles, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos. En el nivel general, el proceso está organizado en seis fases: análisis del problema, análisis de datos, preparación de datos, modelado, evaluación e implementación.

A su vez, cada fase consta de tareas generales de un segundo nivel. Estas tareas generales se proyectan a tareas específicas, donde se describen las acciones que se deben desarrollar para cada situación específica.

Se definen las fases a continuación:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

- Fase de comprensión del negocio: Se centra en la comprensión de los objetivos del proyecto de minería de datos desde un punto de vista de negocios. Esta fase es equivalente, por lo tanto, a una fase de análisis de requerimientos de un proyecto de desarrollo de software, y es importante porque el cliente puede no tener claro qué es lo que quiere. Las tareas a realizar en esta fase incluyen determinar los objetivos de negocios, evaluar la situación del proyecto en términos de recursos, restricciones y suposiciones, determinar objetivos de minería de datos que traduzcan a criterios técnicos los objetivos de negocios y, finalmente, producir el plan del proyecto.
- Fase de comprensión de los datos: Comprende la recolección inicial de datos, identificando la calidad de estos y estableciendo las relaciones más evidentes entre ellos. Incluye la tarea de recolección de datos iniciales, en la cual deben inscribirse los datos en términos de número de registros, número de campos por registro y significado de cada campo. Incluye también la tarea de descripción de los datos en términos de tipo, distribución, tablas de frecuencia y estadísticas. Tareas adicionales de esta fase son la exploración de los datos mediante gráficos y tablas, y la verificación de la calidad de los mismos.
- Fase de preparación de los datos: En esta fase debe construirse una base de datos, la cual debe contener todas las características consideradas candidatas para estimar el valor de una variable que se espera predecir. Esta fase incluye la tarea de selección de los datos a los que se va aplicar la técnica del modelo, la tarea de limpieza de los mismos para alcanzar el nivel de calidad requerido por las técnicas de minería de datos que sean seleccionadas, la tarea de construir datos adicionales, la tarea de integrar diferentes bases de datos, y la tarea de formatear los datos.
- La fase de preparación de los datos debe entregar datos que estén en un formato adecuado para la técnica del modelo que se empleará en la siguiente fase. Por esto, la fase de modelado puede requerir regresar una o más veces a la fase de preparación de los datos.
- Fase de modelado: Esta es la fase medular de un proyecto de minería de datos y consiste en descubrir una relación entre un conjunto de variables y una variable que se espera predecir. Contempla la selección de una técnica de modelado, entre las cuales pueden mencionarse las redes de KOHONEN o modelos K-MEAN para agrupamiento, árboles de decisión C5 o C&R para segmentación, redes neuronales o regresión logística para predicción, inducción de reglas generalizadas para descubrimiento de patrones y análisis de factores para reducir la complejidad de los datos, entre otras. Entre sus tareas se encuentran la selección de la técnica del modelo, la generación del diseño de las pruebas del modelo, la construcción del mismo y, finalmente, la evaluación técnica a la que debe someterse el modelo a través de criterios estadísticos.

- Fase de evaluación de negocios: Se evalúa el modelo de forma técnica en relación a factores tales como su precisión y generalidad. En esta fase, en cambio, debe evaluarse el nivel de satisfacción de los objetivos de negocios perseguidos por el proyecto de minería de datos. Incluye la tarea de evaluar los resultados, la tarea de revisar el proceso de minería de datos y, finalmente, la tarea de determinar los próximos pasos a seguir (momento en el que debe decidirse si debe darse por terminado el proyecto de minería de datos y entrar en la fase de despliegue, si deben iniciarse iteraciones adicionales, o si debe iniciarse un nuevo proyecto de minería de datos).
- Fase de despliegue del modelo: En esta fase deberá definirse una estrategia para implementar los resultados de la minería de datos. Incluye las tareas de planificar el despliegue del modelo, de planificar el monitoreo y el mantenimiento de los modelos, de generar el reporte final del proyecto, y de revisar el proyecto en relación a evaluar lo que ocurrió correctamente y lo que necesita ser mejorado.

2.8.3.4 COMPARACIÓN DE METODOLOGÍAS

En el campo en expansión de la minería de datos, se ha solicitado una metodología estándar o simplemente una lista de mejores prácticas para el proceso diversificado e iterativo de minería de datos que los usuarios pueden aplicar a sus proyectos de minería de datos, independientemente de la industria. Mientras que el proceso estándar de *cross industry* para minería de datos o CRISP-DM, fundado por la iniciativa del Programa Estratégico Europeo de Investigación en Tecnología de la Información, tenía como objetivo crear una metodología neutral, SAS también ofreció un patrón a seguir en sus herramientas de minería de datos.

Algunos modelos profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos (como CRISP-DM), mientras que otros proveen sólo una guía general del trabajo a realizar en cada fase (como el proceso KDD o SEMMA).

2.9 ANALYTICS

El concepto de *people analytics* puede asumir diferentes definiciones y enfoques, sin embargo, existe un amplio consenso para que sea comprendido como el proceso sistemático para obtener, procesar, analizar e interpretar información, con el propósito de mejorar las decisiones que se toman en el ámbito de la gestión del capital humano.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

En el mundo laboral, los *big data* posibilitaron el surgimiento de los *people analytics*. Técnica resultante del análisis y aplicación de macrodatos para el desarrollo de prácticas vinculadas con la administración del personal (Zangaro, 2020).

2.9.1 OBJETIVO DE *PEOPLE ANALYTICS*

People analytics, también llamado análisis de talento o personas es la aplicación de técnicas considerables de minería de datos y análisis de negocios a los datos de talento. Análisis que miden el rendimiento y la eficiencia que solo son importantes para RRHH. Los ejemplos incluyen: tiempo para llenar una solicitud de trabajo, número de personas capacitadas, número de personas con ciertas competencias, deserción del año pasado, deserción estimada para el próximo año, cantidad estimada de candidatos a tener en cuenta en función de la deserción estimada, cuya fuente proporciona el mejores candidatos, informes de cumplimiento, informes de diversidad (Stone, <https://www.cornerstoneondemand.com>, s.f.).

Analytics es el encuentro entre el arte y la ciencia. El arte nos enseña cómo mirar el mundo. La ciencia nos enseña cómo hacer las cosas. El propósito de *analytics* es encontrar el mejor recorrido a través de la masa de datos para descubrir el valor que está oculto allí (Fitz-Enz J. R., 2014).

Es el proceso de examinar grandes volúmenes de datos de una gran variedad de tipos (*big data*) para describir patrones ocultos, correlaciones desconocidas y otras informaciones útiles. El objetivo principal es ayudar a las compañías a tomar decisiones (Aguilar, 2013).

“El objetivo de *analytics* es encontrar la mejor ruta a través de una masa de datos para descubrir los valores que están ocultos allí” (Fitz-Enz J. R., 2014, pág. 4).

El objetivo de *people analytics* es proporcionar a una organización información para administrar de manera efectiva a los empleados para que los objetivos comerciales se puedan alcanzar de manera rápida y eficiente. El análisis de recursos humanos no solo se ocupa de recopilar datos sobre la eficiencia de los empleados. En cambio, su objetivo es proporcionar información sobre cada proceso mediante la recopilación de datos y luego usarlos para tomar decisiones relevantes sobre cómo mejorar los procesos. El desafío de la analítica de recursos humanos es identificar qué datos se deben capturar y cómo usar los datos para modelar y predecir capacidades para que la organización obtenga un retorno óptimo de la inversión en su capital humano (Stone, <https://www.cornerstoneondemand.com/>, s.f.).

Predecir conductas futuras, basándose solo en conductas pasadas, puede tratarse de un verdadero disparate, si no se hace con la suma rigurosidad que sólo puede darnos un profesional de las ciencias del

comportamiento. La historia personal no se preserva, se crea y recrea en un presente en el que hemos pasado de ser meros relatores a ser protagonistas a constructores de realidades. Cada interlocutor nos obliga, sin saberlo y buscarlo siquiera, a recrear un nuevo relato, una nueva realidad de nuestro pasado.

El uso de la inteligencia artificial es la tendencia y, parece, que la gestión de los recursos humanos es una de las áreas que está sumándose a esta corriente. Se utiliza para el análisis de los datos y la aplicación de la inteligencia artificial en la evaluación de perfiles, para medir el nivel de integridad de un candidato o para visualizar la experiencia cotidiana de los trabajadores y hasta “monitorear” el clima laboral en las empresas.

“Xpocketes” se trata de una herramienta, online, que ofrece medir las experiencias, visualizarlas en el momento en que se producen efectivamente y permite a la empresa accionar rápidamente. Se realiza a través de una medición constante de consultas ad hoc y el procesamiento inmediato de la información. Así, los jefes obtienen una foto de la experiencia de su propio equipo y ven el impacto de su gestión a nivel personas. Monitoreado, todo, por el área de recursos humanos (Samela, 2019).

Cuando nos referimos al término analítica, debemos comprender que posee dos tipos de uso; puede aprovecharse tanto desde la perspectiva del análisis de situaciones pasadas como para la predicción o prescripción de escenarios futuros.

Cabe destacar el concepto de indicador de gestión, también llamado *key performance indicator* (KPI), los cuales son utilizados desde hace años por muchas empresas. Éstos sirven para responder a lo que ya pasó, de esta manera las decisiones se toman de modo reactivo. En cambio, cuando se hace referencia a *analytics*, se podría responder a “porqué pasó lo que pasó” y a “qué podría pasar” con la posibilidad de predecir escenarios futuros; de esta manera las decisiones se toman proactivamente.

Una vez definidos los conceptos básicos de la temática, se presenta un modelo de cuatro niveles en los cuales se propone comprender cuáles son los objetivos propios de cada momento, qué tareas realizan, qué habilidades son necesarias para el nivel y cuáles serían los resultados esperados. Son así útiles formas de entender cómo puede ir evolucionando y madurando *analytics* en la organización (Bersin, www.deloite.com, 2013):

Reactivo – informes operacionales, se trata de objetivos simples, pero relevantes para la generación de los cimientos necesarios para las etapas posteriores. Es el adquirir de los que se necesita técnicamente para poder hacer cierto uso de datos (tener capacidad de entender qué datos se almacenan, cómo se registra y qué significa cada variable). Se relevó que el 56% de las empresas está en este nivel.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Proactivo – informes avanzados, este nivel incluye la realización de informes operacionalizados para hacer comparaciones (*benchmarking*) y tomar decisiones. Estamos aquí frente a un análisis multidimensional y tableros, en términos de análisis: haciendo cortes, viendo y comparando segmentos. En cierta forma podría ser comparado a lo que se denomina *business intelligence* (BI), herramientas de software, hardware, redes y bases de datos para organizar, analizar y proveer acceso a datos para ayudar a los managers y a los demás usuarios empresariales a estar más informados y preparados para la toma de decisiones (Laudon, 2012). Se relevó que el 30% de las empresas está en este nivel.

Análisis estratégico, en este nivel se distinguen diferencias cualitativas en término de enfoque de análisis y las herramientas utilizadas. Implica y requiere herramientas de análisis más sofisticadas que el simple uso de estadística descriptiva (por ejemplo, análisis de distribución). En este nivel se utiliza, análisis estadístico, diseño de modelos relativos a fenómenos propios de la conducta humana. Se relevó que el 10% de las empresas está en este nivel.

Análisis predictivo, en este nivel se alcanza la posibilidad de poder diseñar e implementar modelos predictivos del talento, por ejemplo, predecir quién va a renunciar y quién tendrá alto desempeño. Se relevó que el 4% de las empresas está en este nivel.

Figura 17 “El modelo de madurez en analítica”



Fuente: (Bersin, 2013)

2.9.2 CASOS DE EJEMPLIFICADORES DEL USO DE *PEOPLE ANALYTICS*

Algunas de las acciones que son llevadas adelante en las diferentes áreas de las organizaciones que implementan esta tecnología son:

Las compañías de seguros han analizado los mejores perfiles de desempeño en ventas y ahora saben que los candidatos con un alto perfil académico no necesariamente tienen un excelente desempeño comercial (Deloitte University Press, 2016, pág. 88).

Una compañía de tecnología desarrolló un modelo predictivo que le permite saber qué personas pueden convertirse en “colaboradores tóxicos” (aquellos que mienten, hacen trampa, o cometen algún crimen) y este se ha convertido en un buen filtro en su proceso de selección (Deloitte University Press, 2016, pág. 88).

Las compañías del sector automotriz estudian patrones de ausentismo para predecir cuándo la gente podría tomar días libres y poder así programar sus reemplazos para cubrir estas ausencias (Deloitte University Press, 2016, pág. 89).

El Ministerio de energía de México usa un modelo predictivo de planeamiento de su fuerza laboral, para identificar brechas de talento crítico actuales y futuras, en cargos jerárquicos del sector petrolero, de aquí a diez años. El modelo incorpora tendencias macroeconómicas tales como el precio del petróleo y la tasa de cambio, las cuales se correlacionan directamente con la necesidad de talento especializado. Basado en un entendimiento de la brecha de competencias críticas, el ministerio puede trabajar de manera proactiva con varios proveedores para suplir estas brechas. Con esta iniciativa se ha expandido la capacidad de planificar la fuerza laboral y anticiparse a las demandas futuras (Deloitte University Press, 2016, pág. 89).

Las compañías de software, el sector bancario y la manufactura identifican las características de sus mejores vendedores, para descubrir la manera en que trabajan con sus pares, así como el tiempo que pasan con sus clientes y la manera como construyen *networking*, lo cual significa operar con un mejor predictor que la cantidad de entrenamiento o experiencia directa en venta que puedan tener (Deloitte University Press, 2016, pág. 88).

Information technology y recursos humanos analizan datos extraídos de mails para entender por qué unas personas son más productivas que otras (Deloitte University Press, 2016, pág. 88).

Una empresa de cosmética creó un “centro de excelencia de productividad en ventas” para estudiar patrones que le sirvieran en sus programas de contratación y compensación y así poder optimizar su productividad (Deloitte University Press, 2016, pág. 88).

Una compañía farmacéutica y otra de software han usado información de LinkedIn y otras redes sociales de trabajo para predecir el “riesgo de perder el talento” (Deloitte University Press, 2016, pág. 89).

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Deloitte Canadá, ha encontrado que aquellas áreas que tienen mejor iluminación, salas más grandes de reuniones y una mejor colaboración entre sus equipos, son aquellas que tienen mayores índices de retención y productividad (Deloitte University Press, 2016, pág. 89)..

Mastercard ha desarrollado también modelos predictivos que mejoran la experiencia de sus empleados. Analiza patrones que permiten tomar mejores decisiones en temas de retención de altos potenciales y prepararse ante posibles renunciaciones (Deloitte University Press, 2016, pág. 89).

Los bancos estudian patrones de fraude e incumplimiento y pueden ahora predecir comportamientos que pueden atentar contra la ética del negocio (Deloitte University Press, 2016, pág. 89).

Una compañía financiera de Reino Unido usa analítica para evaluar a sus colaboradores e identificar quiénes pueden ser potencialmente deshonestos. Esta herramienta les permite anticiparse a sus riesgos y manejarlos de manera proactiva (Deloitte University Press, 2016, pág. 89).

Una compañía eléctrica investigó un accidente de trabajo y pudo constatar que, si hubiese analizado su encuesta reciente de compromiso, habría podido predecir los errores antes de que ocurrieran. (Deloitte University Press, 2016).

Un equipo de desarrollo organizacional de *eBay* mide la internalización de sus valores culturales a través de métricas internas y externas. Compara las percepciones de sus colaboradores con la percepción externa de su compañía, a través de análisis temáticos obtenidos de artículos que circulan en los medios y hacen referencia a la percepción de la empresa en el mercado. (Deloitte University Press, 2016, pág. 90).

2.9.3 FUNCIONES Y TAREAS DE *PEOPLE ANALYTICS*

Principalmente, *analytics* es una herramienta de administración que puede ser usada para identificar oportunidades, resolver problemas actuales y predecir futuros rendimientos de las inversiones. Como cualquier instrumento, puede ser usado incorrectamente, pero no significa que sea una mala herramienta. (Fitz-Enz, 2014).

La analítica predictiva es la parte de la analítica que trata de predecir el comportamiento de los usuarios. Es una rama de la minería de datos (*data mining*) centrada en la predicción de las probabilidades y tendencias futuras. La analítica predictiva pretende la mejora de decisiones y respuesta rápidas al cambio y comprende técnicas de minería de datos, estadística y modelado. Trata de analizar hechos actuales o históricos con el propósito de hacer predicciones sobre sucesos futuros. (Aguilar, 2013)

Por ello, resulta importante describir qué tipo de información se recopila y con qué fin. La analítica descriptiva, utiliza datos históricos para identificar comportamiento y gráfica respecto de cómo se están haciendo las cosas. La analítica predictiva, hace posible la creación de modelos que permiten vaticinar lo que va a ocurrir con antelación. La analítica prescriptiva, analiza los datos para encontrar la solución entre una gama de variantes, siendo su tarea optimizar recursos y aumentar la eficiencia operativa (Bodenheimer, 2017).

Los registros de los empleados contienen datos en bruto sobre la fecha de contratación, la revisión del rendimiento, cualquier cambio de estado y la fecha de salida. Existe una rica base de datos de investigación sobre rotación, que arroja teorías sobre los motivos para quedarse y partir. Sin embargo, hasta el momento, se han realizado muy pocos intentos para conectar los cambios de retención en los resultados comerciales. (Fitz-Enz, 2014).

John Boudreau, profesor de la *USC University*, formuló una tesis en la que establece que el siguiente paso en la evolución de la gestión de recursos humanos, será abandonar el método intuitivo (como base para tomar decisiones) y reemplazarlo por el del análisis fáctico. Ese paso es el que propone el movimiento de *people analytics* y ya se encuentra en curso. Esto no significa que la intuición sea esencialmente mala. Sin embargo, la gestión actual requiere de un rigor metodológico basado en el procesamiento inteligente de una maravillosa cantidad de información de la que hoy disponemos (Cravino, 2011, pág. 3).

Analytics puede aprovechar la información disponible en los sistemas para identificar patrones de aquellos que pudieron crecer en la organización sosteniendo su desempeño, y así tomar acciones vinculadas al potencial, basadas en modelos y no suposiciones de qué es potencial. Los datos por sí solos son cacofonía pura (Escudero, 2019).

Con todas estas posibilidades de análisis de datos la toma de decisiones se vuelve más fácil. Sin embargo, es importante no dejar de lado el “*gut feeling*” o la intuición humana. Con esto no nos referimos a una toma de decisiones impulsiva, sino a un proceso cuidadoso, basado en años de experiencia (Horacio Gómez, 2020).

CAPÍTULO 3: GESTIÓN DE RECURSOS HUMANOS Y SECTOR FINANCIERO DE LA REPÚBLICA ARGENTINA

Es mucho el tiempo en que se viene sosteniendo que quienes conforman una organización tienen sobradas razones para intentar ejercer influencia a los nuevos miembros, tales como el modo correcto de percibir, pensar, sentir y afrontar los problemas. Es que los miembros de la empresa han ido aprendiendo a enfrentarse a la complejidad que plantea el día a día y saben que lo han hecho con suficiente éxito, éxito que les ha permitido permanecer.

Las empresas deben lograr la máxima eficacia y eficiencia en sus colaboradores determinando necesidades de competencia y formación para el personal, evaluando constantemente las acciones y los objetivos pactados por las gerencias. Las organizaciones saben que, para permanecer en el futuro, necesitan desarrollar una cultura sustentada en la confianza. Asegurar que el personal sea y esté consciente de la importancia de sus actividades y como deben contribuir a alcanzar el objetivo de satisfacción al cliente.

En las compañías en las que el capital intelectual llega a ser una decisiva fuente productora de valor, los profesionales de recursos humanos deben ser activos y enérgicos en el desarrollo de este capital. Los profesionales de recursos humanos se convierten así en adalides de los colaboradores vinculando las contribuciones de los empleados al éxito de la organización. Con activos adalides de los empleados que entiendan las necesidades y aseguren que se responda a esas necesidades, la contribución de los empleados aumenta (Ulrich, 2016).

Se toma como base de la investigación los conceptos desarrollados por diferentes autores quienes elaboran bases que sirven de apoyo para el desarrollo del trabajo en cuestión.

3.1 LA MOTIVACIÓN.

Todos tenemos determinadas capacidades que podríamos distinguir como innatas o adquiridas. Cada uno de estos grupos está compuesto por tres elementos básicos, a saber:

Capacidades innatas:

- La capacidad física.
- La capacidad intelectual.
- La resistencia psíquica

Capacidades adquiridas:

- Las habilidades.
- El conocimiento.
- La experiencia

La motivación es un concepto crucial en el estudio de las organizaciones. Su relación con el desempeño individual, la satisfacción y la productividad organizacional la ha convertido en tópico esencial del Comportamiento Organizacional. Sin embargo, tanto la definición como la explicación del fenómeno cuentan con un amplio espectro de posibilidades.

La motivación es el catalizador que pone a las capacidades en juego, traduciendo en potencialidad interna en realidad externa (Vola-Luhrs R. &., 2010).

La motivación es el proceso mediante el cual cada trabajador cumple con su tarea laboral con eficiencia, para lograr una meta o resultado mediante el cual puede satisfacer sus necesidades particulares (Aponete, 2006).

Para fines de introducción de la temática, se considera la motivación como un proceso mediante el cual se inicia, se sostiene y se direcciona una conducta para alcanzar un incentivo que satisface una necesidad importante en ese momento para el individuo.

3.2 EL ENTORNO LABORAL Y LOS TRABAJADORES

En primera instancia consideremos, el entorno social, el cual hace referencia a un entorno donde la persona se desarrolla, el cual se nutre de diferentes circunstancias, tal como la cultura, la sociedad y la

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

comunidad donde vive el individuo. El entorno social, permite involucrar a las personas en diversas situaciones, dándole la posibilidad de crear diferentes roles en entornos de interacción con otros.

En segunda instancia, el entorno laboral, es aquel entorno que permite desencadenar en el individuo, una serie de herramientas, las cuales le posibilitan ir obteniendo competencias que podrán ser utilizadas en un entorno profesional y personal.

En tercera instancia, y no menos importante, se considerará el entorno profesional, el cual refiere el logro y alcance paulatino de diversos escalones, por los que debe subir el ser humano, y en el que diversas habilidades como adaptación y el compromiso entre otras le generan un estilo de vida.

El entorno personal se refiere a aquellas características demográficas de la persona, tales como la edad, género, situación socioeconómica y nivel educativo alcanzado, todas estas características permiten crear una percepción de la persona, y analizar su evolución individual.

Por último, consideraremos el entorno socio sanitario implica que el bienestar de cada persona es una continuidad no solo recursos si no de cuidados que deben estar seguidos de acciones que permitan la salud integral y así mismo la calidad de vida

“Una forma nueva y eficaz de humanizar el trabajo es humanizar la concepción de la organización, concibiéndola como un ser organizacional. Sistémicamente tiene estructura, dinámicas emergentes, morfología y fisiología propia (o estilo de gestión si se prefiere) y todas las características de un ser vivo. Concebirlo así facilita la gestión organizacional” (Reyes, 2009). Es así que las empresas son dinámicas y gestionan sus recursos de acuerdo a las divergencias del mercado y a las demandas de sus trabajadores como pilares de dicha dinamización.

La evolución del concepto de trabajador, ha pasado por definiciones basadas en la preocupación por aumentar la eficiencia de la industria a través de la racionalización del trabajo-operario, y en otras instancias, autores como Fayol que en 1841, consideraban su preocupación basada en el aumento de la eficiencia de la empresa a través de la organización y estructura de la misma. Aspectos como la producción en masa, el aumento de costos de la mano de obra, fueron impactando en las contrataciones del personal, en la división del trabajo y particularmente en la configuración de los vínculos de las personas con el mundo del trabajo.

Desde la revolución industrial hasta nuestros días, estas configuraciones se van transformando en función de un proceso de cambio económico, social y psicológico. La transición desde una economía basada en el trabajo manual, reemplazada por otra dominada por la industria y la manufactura hasta la revolución

de las comunicaciones y de la tecnología conlleva la formulación de nuevos principios de administración y gestión de las empresas. Las importantes modificaciones en el plano diacrónico tendrán consecuencias en el tipo de vínculo que se establece entre las personas y el mundo del trabajo:

- Demográficas: Traspaso de la población del campo a la ciudad. Migraciones internacionales. Crecimiento sostenido de la población. Descolonizaciones.
- Político - Económicas: Desarrollo del capitalismo. Aparición de las grandes empresas y grandes monopolios.
- Sociales: El surgimiento del proletariado urbano. Reivindicaciones y conquistas sociales.
- Ambientales: Deterioro del ambiente y degradación del paisaje — Explotación irracional de la tierra. Uso de nuevas fuentes de energía.
- Tecnológicas: La aplicación de la ciencia y tecnología en los procesos productivos. Las máquinas.
- Comunicaciones: La revolución en el transporte, la telefonía, Internet.

Es primordial hacer hincapié en los desafíos que ciertas tareas conllevan a los responsables de recursos humanos a partir de los elementos que emergen de la función de dicha área para alcanzar los resultados que se espera de esta unidad: Identificar y aumentar el capital humano que produce riqueza, contratar personal en tiempo y forma, retener personal idóneo y necesario, lograr competitividad y productividad en el trabajo humano, lograr que la empresa sobreviva a sus miembros (Vola-Luhrs, 2010).

El objetivo de estos planes es unificar aspiraciones de un empleado con las oportunidades y desafíos existentes en la organización. En otros términos, incrementar la probabilidad de un empleado para que alcance sus logros y asegurarse de que la organización ubique a la gente adecuada en el lugar y tiempo correctos. Tanto individuo como organización necesitan de estas actividades, dado que además de aumentar las posibilidades de éxito individual, reducen los costos que se asocian a carreras frustradas. Por otro lado, ayuda a la organización a identificar y desarrollar a aquellos individuos que desean y tienen aptitudes para puestos directivos. El planeamiento estratégico es un proceso mediante el cual la gerencia evalúa e integra en las decisiones de políticas, tanto el estado actual como la repercusión futura de los cambios en su medio de operación. Las decisiones estratégicas tienen que ver, ante todo, con los problemas de la organización en su afán por reestructurar los recursos de la organización de manera que generen un potencial de desempeño máximo (Vola-Luhrs, 2010).

El área de recursos humanos necesita aprender a hablar en términos cuantitativos y objetivos, utilizando números para expresar la actividad y valor agregado. Las empresas usan números para explicarse. Las ventas, los gastos operativos, los ciclos de tiempo y los volúmenes de producción son indicadores que

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

expresan la actividad comercial. En la década del '70, la productividad era la cuestión clave. En la década del '80, el movimiento de calidad enfatizó la eficacia del proceso como una ventaja competitiva. Ambos se basaron en números para expresar grados de cambio. Estamos en el umbral de la fase más emocionante y prometedora de la evolución de los recursos humanos y la gestión del capital humano. Hemos pasado del caballo y el cochecito al automóvil y al avión. Ahora es el momento de montar el cohete y dirigirse a la estratosfera. Al igual que la aritmética, la estadística no tiene sesgo y es aplicable a una amplia gama de oportunidades. Se puede usar en estudios de problemas únicos y localizados o para apoyar cambios de imagen en toda la organización. La salsa secreta de la estadística es como el código fuente de los programas de computadora, una lógica oculta que puede ir paso a paso o avanzar, utilizando macros para acelerar la solución (Fitz-Enz J. , 2010).

La gestión estratégica de la cadena comienza con los altos ejecutivos construyendo su plan trascendental de negocios haciendo la siguiente pregunta básica: ¿Cómo hacemos dinero?

La gerencia también mira hacia adentro las capacidades de la empresa. Estos incluyen la visión de las empresas, el liderazgo, la marca, la cultura, la fortaleza financiera y las capacidades de los empleados. A partir de esta doble evaluación, se hacen planes para producir, vender y dar servicio a la oferta de la empresa. Este escenario lleva a respuestas presuntas y/o probadas de los clientes. El siguiente paso en la cadena son las operaciones, donde los gerentes de línea planifican el diseño y administran el sistema de producción que presumiblemente servirá y respaldará las respuestas de los clientes. Las operaciones dependen del talento humano (Fitz-Enz, 2014).

En relación a la motivación de las personas en las organizaciones, se enuncia que el pago y las condiciones por sí solos no determinan la motivación. Por qué dar a una persona un ascenso o un aumento de sueldo puede tener un efecto desmotivador en los demás. La creencia en la teoría de la equidad laboral (Adam, 1963), es que las personas valoran el trato justo, lo que hace que se sientan motivadas a mantener la equidad dentro de las relaciones de sus compañeros de trabajo y la organización. Palabras como esfuerzos y recompensas, o trabajo y paga, son una simplificación excesiva, de ahí el uso de los términos entradas y salidas. Los *inputs* son lógicamente lo que damos o ponemos en nuestro trabajo. Los *outputs* son todo lo que sacamos a cambio. Se trata de que el balance entrada-salida compense al trabajador y/o perciba que lo compensa.

El mayor misterio del área de gestión humana en la actualidad es el concepto del potencial y su evaluación. A pesar de ser un tema de mención frecuente por los profesionales, poco hay escrito sobre este asunto y menor cantidad de trabajos desarrollados con estudio y análisis objetivo. Frecuentemente

los consultores ofrecen herramientas, pero “misteriosos” son los estudios que justifican la validez y confiabilidad de las mismas preguntas ¿Qué diferencia a aquellos que pueden ir creciendo en las organizaciones y sostener el nivel de desempeño a medida que pasan a nuevas posiciones? ¿Las herramientas que supuestamente evalúan potencial son válidas o confiables? ¿Estamos tomando decisiones equivocadas al confiar en instrumentos que miden en realidad otra cosa? (Bodenheimer, 2017).

3.3 RELACIÓN ENTRE ACTIVIDADES DE FORMACIÓN Y DESEMPEÑO

Los puestos de trabajo constituyen la intersección de las personas con las organizaciones, y conforman la base de la gestión de recursos humanos y del desarrollo de las funciones básicas de selección, capacitación, desempeño y carrera.

La identidad que caracteriza a la educación es la libertad académica, que está en contraste con el proceso de capacitación donde no existe libertad del instructor en la elección de los contenidos, ni tiene libertad el participante de aplicar o no los conocimientos adquiridos. Quien participa en un proceso de capacitación, efectivamente, debe aplicar lo aprendido en forma inmediata en su puesto de trabajo. Aquí el rol del superior inmediato es el de auditar la aplicación efectiva, en el día a día, de los conocimientos adquiridos, de las habilidades perfeccionadas o de las actitudes nuevas o modificadas del trabajador. Por otro lado, el proceso de Desarrollo se lleva adelante con la mirada puesta en el futuro, es decir, se diferencia de la capacitación en que lo aprendido por el participante no será de aplicación inmediata en su puesto de trabajo sino en otros puestos o responsabilidades en el futuro a veces aún incierto.

Entonces, en la educación, la libertad de cátedra es una característica que la identifica y diferencia de la capacitación y al desarrollo. Es uno de los derechos incluidos dentro de los derechos fundamentales como lo es la libertad académica.

La capacitación del personal es toda actividad que se realiza en una determinada organización respondiendo a sus propias necesidades, tendiendo a provocar un cambio positivo en la actitud mental, los conocimientos y habilidades que se traducirán en las conductas de su personal. Cuando decimos cambio positivo, entendemos lo que la propia organización, en forma arbitraria y en forma unilateral, ha definido como meta operativa de la acción a realizar (Vola-Luhrs R. &., 2010).

Es difícil encontrar una empresa en la que las necesidades de capacitación no sean mayores que lo que el presupuesto permitiría satisfacer. Más allá de la creatividad puesta en juego para optimizar los recursos escasos, hay esquemas que permiten jerarquizar, dar un orden de prioridad a las demandas relevadas de

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

capacitación. En la figura 17, “esquema preparado por la Consultora McKinsey y presentado por Le Boterf sugiere dar prioridad a las necesidades de capacitación que tienen mayor interés estratégico y mayor rentabilidad” (Vola-Luhrs R. &., 2010).

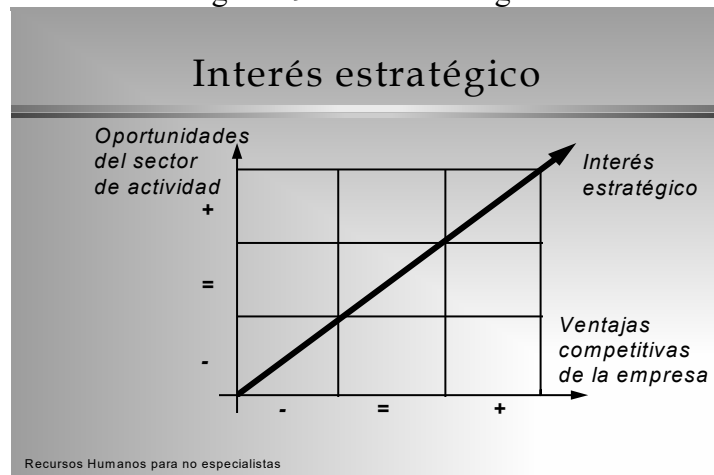
Figura 18 Ordenamiento de las necesidades



Fuente: (Vola-Luhrs R. &., 2010)

Cuando el mercado presenta oportunidades para la venta de determinado equipo, por ejemplo, y la empresa tiene un buen modelo y una marca prestigiosa, capacitar a los vendedores en este producto tiene un alto valor estratégico, según lo representamos en el siguiente cuadro:

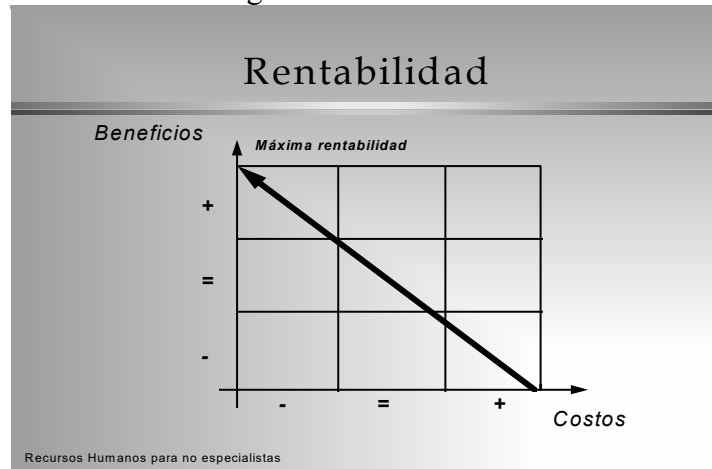
Figura 19 Interés estratégico



Fuente: (Vola-Luhrs R. &., 2010)

En cuanto a la rentabilidad, es tema conocido, una relación entre costos y beneficios.

Figura 20 Rentabilidad



Fuente: (Vola-Luhrs R. &., 2010)

La relación entre la gestión de las actividades de formación y el desempeño es sustentada por diferentes autores a partir de investigaciones realizadas.

Existe la opinión de que la formación es importante no solo porque es necesaria para construir y mantener una fuerza laboral, sino también porque fomenta el bienestar corporativo y mejora el desempeño organizacional (Guterresa, 2020).

“La Gestión del Talento Humano en la actualidad se convierte en una de las mejores herramientas que permite mejorar los diferentes procesos que lleva a cabo una institución. Como también coadyuva en el mejoramiento del desempeño laboral de las personas en sus funciones y obligaciones que cumplen en sus puestos de trabajo. Y, por lo tanto, la función de la Gestión del Talento Humano se vuelve esencial para el logro de ventajas competitivas. Las personas como Talento Humano. En esta nueva concepción, dejan de ser simples recursos (humanos) organizacionales para ser abordadas como seres dotados de inteligencia, personalidad, conocimientos, habilidades, destrezas, aspiraciones y percepciones singulares” (Peralta, 2014).

La formación del trabajador contribuye a que éste realice sus actividades de forma más eficiente, por lo que el conocimiento institucionalizado (educación formal) y el conocimiento certificado por instancias profesionalizantes o de capacitación contribuye a incrementar la productividad laboral. En este sentido, una mayor formación por parte del trabajador se encuentra incentivada debido a que en sí misma una mayor educación encierra la promesa de una mejora salarial en términos laborales (Becker, 1984).

La importancia de la formación del recurso humano se encuentra ligada a las mejoras en eficiencia que estos experimentan en sus áreas de trabajo, un mayor y mejor conocimiento de calidad les permite desempeñar mejor sus labores, abrir la posibilidad de movilidad dentro de la organización, ocupar áreas

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

de trabajo de mayor responsabilidad y de mejor retribución económica. Un nivel de ingresos más elevado le brinda al trabajador una mayor satisfacción y un mejor bienestar de vida a nivel personal y familiar. Es por ello por lo que el trabajador se empeña en adquirir conocimientos en la escuela y al interior de la organización donde labore, es decir, avanzar en los grados académicos y capacitarse. La educación y capacitación que el individuo posee se suma a los factores determinantes de los puestos de trabajo y de la remuneración que percibe como, por ejemplo, el género, la edad, etc. (Zambrano, 2018).

“La preparación permanente de todos los trabajadores para desempeñarse con eficacia en su lugar de trabajo, asumiendo con mente flexible los cambios y nuevas tecnologías que se suceden vertiginosamente. Abordar el desempeño profesional desde la capacitación en el contexto actual del progreso científico técnico y tecnológico en todas las ramas de la economía y la sociedad, implica tener en cuenta la necesidad de elevar la apropiación de conocimientos, el desarrollo de habilidades y valores que le permitan al individuo desenvolverse con éxito en la profesión y alcanzar los objetivos que le confiere la sociedad” (Machín, 2015).

3.4 RELACIÓN ENTRE AUSENTISMO Y CLIMA LABORAL

El ausentismo es un factor que desencadena dos tipos de costos, el directo, que implica el costo monetario que tiene la compañía, y el costo indirecto, que implica las estrategias de compensación ante ausencias no planeadas, como el pago de horas extras. Pero de fondo no hay que perder de foco los motivos por los cuales se genera el ausentismo. Basta como ejemplo verificar el abordaje típico del ausentismo caracterizado como desvío de pautas reglamentarias y generalmente gestionadas a partir de acciones disciplinarias correctivas y/o punitivas.

Se ha comprobado que en el ausentismo laboral influyen diversos factores, variables propias del individuo, como las motivaciones, necesidades, hábitos, valores, habilidades y conocimientos; variables de tipo ocupacional, como el tipo de empresa, los métodos de producción y el tamaño de la planta laboral, y variables relacionadas con la organización, como el clima laboral de la empresa y las políticas institucionales (Medina-Campos, 2020).

El clima laboral es un indicador fundamental en la vida de la empresa, condicionado por múltiples cuestiones; desde las normas internas de funcionamiento, las condiciones ergonómicas del lugar de trabajo y equipamientos, pasando por las actitudes de las personas que integran el equipo, los estilos de dirección de los líderes y jefes, los salarios y remuneraciones, hasta la identificación y satisfacción de cada persona con la labor que realiza (Trigine, 2012).

Las empresas tienen vida propia y tienen una influencia determinante sobre el comportamiento humano en las mismas. Las personas actúan dentro del marco de referencia formal o informal que dan los sistemas. Por esto, la idea de actuar sobre los individuos en los procesos de cambio llevará a la organización al fracaso. Hay que actuar sobre los sistemas para que los individuos modifiquen sus comportamientos, mitos, valores, tradiciones, costumbres y hasta cambien sus héroes. Las venas de la organización son los sistemas y que la estructura de valores debe estar alineada con la de los individuos para generar un clima organizacional que se manifieste en todo su esplendor.

La relación entre el ausentismo y el clima laboral es sustentada por diferentes autores a partir de investigaciones realizadas.

“Si un trabajador está cansado del agotamiento físico o no puede soportar la presión psicológica en el trabajo, es más probable que él o ella esté ausente del trabajo que una persona que trabaja en un Ambiente de mejor calidad. Los aspectos físicos y psicológicos del trabajo. se supone que la situación influye en la decisión de ir a trabajar en un día determinado, y causar ausencias voluntarias e involuntarias” (Ose, 2005).

“El absentismo puede ser considerado como un resultado final y una consecuencia de la satisfacción de los profesionales, y seguramente las intervenciones efectivas para disminuir el absentismo de enfermería hospitalaria lo son porque están aumentando la satisfacción laboral de este personal” (Joaquín Jesús Blanca-Gutiérrez, 2013).

“Para analizar cuáles son las cuestiones que generan la conducta de no asistir al trabajo en los individuos, se debe examinar el grado de satisfacción laboral que poseen los trabajadores. La satisfacción laboral es el estado emocional positivo o placentero de la percepción subjetiva de las experiencias laborales del sujeto” (Locke, 1976).

“El tercer enfoque para el estudio del absentismo, el que toman los sociólogos, incorpora conceptos de los modelos económicos y psicológicos, pero añade una dimensión estructural relacionada con el entorno laboral y el clima” (Ambrose, 2015).

La satisfacción laboral se ha convertido en un objetivo básico para la organización y la dirección de recursos humanos, ya que se ha constatado su influencia sobre aspectos tales como la productividad, el absentismo y la rotación de personal (Jesus, 2010).

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

3.5 SECTOR FINANCIERO DE LA REPÚBLICA ARGENTINA

Con la intención de contextualizar al sector financiero de la República Argentina, el cual se replica para la Ciudad Autónoma de Buenos Aires, se dispone en la Ley 24.144/92 del Marco Legal del Sistema Financiero (Banco Central de la República Argentina, 2019), el cual ofrece un panorama comprensivo de la normativa vigente relativa al sistema financiero argentino. Es elaborado por el Banco Central sobre la base de los datos publicados en el Boletín Oficial de la Nación y en Infoleg (Información Legislativa, fuente del Ministerio de Economía y Finanzas Públicas).

A modo de complemento, cabe destacar la Ley de Entidades Financieras N° 21.526, que en sus dos primeros artículos menciona a quiénes comprende:

Artículo 1 - Quedan comprendidas en esta ley y en sus normas reglamentarias las personas o entidades privadas o públicas (oficiales o mixtas) de la Nación, de las provincias o municipalidades que realicen intermediación habitual entre la oferta y la demanda de recursos financieros (Banco Central de la República Argentina, 2019).

Artículo 2 - Quedan expresamente comprendidas en las disposiciones de esta ley las siguientes clases de entidades: a) Bancos comerciales, b) Bancos de inversión, c) Bancos hipotecarios, d) Compañías financieras, e) Sociedades de ahorro y préstamo para la vivienda u otros inmuebles, f) Cajas de crédito. La enumeración que precede no es excluyente de otras clases de entidades que, por realizar las actividades previstas en el artículo 1, se encuentren comprendidas en esta ley (Banco Central de la República Argentina, 2019).

Se considera al sector financiero, el sector económico en el que será indispensable la aplicación y el buen uso de herramientas tecnológicas innovadoras en gestión de las personas, para asegurar un crecimiento sustentable y eficiente como fuente de atracción del desarrollo económico, generación de conocimiento, desarrollo de tecnología y revalorización de profesionales.

CAPÍTULO 4: DISEÑO DE LA SOLUCIÓN

En el mundo de los negocios, los tomadores de decisiones necesitan tener acceso a información precisa y oportuna a fin de alcanzar sus objetivos. Históricamente la inteligencia de negocios (BI) había sido utilizada por los analistas para procesar los datos haciendo uso de herramientas complejas y hojas de cálculo. En la actualidad la toma de decisiones implica una amplia gama de roles de negocio. Los mayores vendedores de inteligencia de negocios (BI) se centran en proporcionar suites completas, lo que permite a los tomadores de decisiones acceder a los datos de origen en casi cualquier ambiente.

Mediante el uso de la inteligencia de negocios (BI) se logra unir el mundo de los datos y el de los negocios. Esta permite a las empresas analizar grandes cantidades de datos de forma rápida y sencilla, que puede ser procesada utilizando diversas reglas o criterios del negocio en cuestión, de forma inteligente. La inteligencia de negocios (BI) integra los datos y puede cruzarlos entre varios procesos, evitando de esta forma los análisis aislados e incompletos, a la vez que facilita y apoya el proceso de toma de decisiones.

La toma de decisiones implementada con inteligencia de negocios, a través de las herramientas de *big data* y *people analytics*, contribuirá de gran manera a una mejor planeación en el área de recursos humanos para optimizar la calidad de la toma de decisiones con respecto a la asignación y el uso eficiente de recursos financieros y recursos no financieros destinados al personal. Con el objetivo de determinar los efectos que causan las actividades de formación en el desempeño y el impacto del clima organizacional en el ausentismo.

En esta fase se comprende a la recolección y comprensión de datos, validar la calidad de los mismos, depurar los datos y establecer una relación entre los campos y atributos provenientes de diferentes fuentes de datos que se utilizarán posteriormente.

La investigación consiste en implementar un proceso predictivo para la eficientización de la asignación de los recursos con los que dispone el área de recursos humanos; aplicando *big data*, minería de datos basado en técnicas de *clustering* y *people analytics*. Como principal fuente de datos para llevar a cabo esta investigación, en la empresa existe un historial de cuatro años y con 952 registros. Se considera el ordenamiento de la información las bases contemplando los registros de los colaboradores y las unidades de negocios, su continuidad con la relación laboral y la perdurabilidad en la entidad financiera.

Las variables a estudiar, por colaborador, son: cantidad de horas de formación atravesadas en cada año y los resultados de las evaluaciones de desempeño.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Las variables a estudiar, por unidad de negocio, son: registros de ausentismo y los resultados de las encuestas de clima laboral y compromiso.

Para este proyecto se utiliza *Oracle Analytics* para el desarrollo de las visualizaciones y como gestor de datos *Oracle SQL Developer*. Se propone las herramientas de Oracle ya que la entidad financiera en la que se lleva adelante el caso de estudio utiliza como sistema de gestión de recursos humanos llamado *Oracle Human Capital Management (HCM)*, el cual es compatible y adaptable a las dos mencionadas.

Oracle Human Capital Management Es una solución integral diseñada de forma nativa para la nube, que conecta todos los procesos de recursos humanos, desde la contratación hasta la jubilación, e incluye funcionalidad RR. HH. globales, administración del talento, administración del personal y nóminas. También es la solución más conectada de toda la empresa, con una nube que unifica HCM en las finanzas, la cadena de suministro y la experiencia del cliente. Esto proporciona una experiencia uniforme en todos los dispositivos, habilita una fuente de confianza para datos de RR. HH. que permite mejorar la toma de decisiones, y le ofrece innovación vanguardista para abordar sus necesidades actuales y futuras. A esto nos referimos cuando hablamos de trabajo humanizado (<https://www.oracle.com>, s.f.).

4.1 DATA VISUALIZATION

4.1.1 ¿QUÉ ES LA VISUALIZACIÓN DE DATOS?

La visualización de datos describe de forma gráfica la presentación de información abstracta. La visualización de datos nos permite identificar patrones, tendencias y correlaciones que, de otra forma, pasarían desapercibidas en los informes, las tablas o las hojas de cálculo tradicionales. Los negocios de la actualidad tienen acceso a una gran cantidad de datos que se generan tanto dentro como fuera de la organización. La visualización de datos ayuda a que usted los comprenda. Le permite mirar los datos de un modo diferente para descubrir nuevas respuestas y conocimientos (Oracle, s.f.).

El desafío de *data visualización* es contar una historia de datos con imágenes. Ir más allá de la simple presentación de números y hechos. Construir un relato en torno a sus datos para que la relevancia de estos se comunique de forma clara. Una imagen vale más que mil palabras. Un solo gráfico expresa problemas complejos de forma más clara y también les permite a los usuarios identificar valores atípicos y anomalías mucho más rápido.

4.1.1.1 ORACLE DATA VISUALIZATION

Es una herramienta de visualización de la suite de *Oracle* la cual permite explorar datos y descubrir estadísticas importantes mediante visualizaciones intuitivas e interactivas. Permite crear reportes que puedan ser comprendidas por cualquier usuario y que a su vez permitan proveer datos para complementar la tomar decisiones. Es una herramienta intuitiva para su uso.

Los beneficios de la herramienta se caracterizan por, facilidad de uso; no necesariamente se requieren recursos especializados para la explotación de la herramienta, independencia del área de técnica.

Versatilidad para compartir; variedad de perfiles de usuario con diferentes permisos. Velocidad en la carga de datos, procesamiento en memoria más rápido, facilidad para manejar millones de registros a la vez y realizar consultas complejas. *Self Service*; usuarios capaces de crear sus propios documentos de análisis para ser compartidos. Existen cuatro versiones de *Oracle Data Visualization*.

4.1.1.1.1 ORACLE ANALYTICS VERSIÓN DESKTOP

Oracle Analytics Desktop o también conocida como *Data Visualization*, proporciona una exploración y una visualización autónoma de los datos en una descarga por escritorio de usuario. *Oracle Analytics Desktop* es la herramienta para la exploración rápida de datos de ejemplo de varios orígenes o para un análisis y una investigación de sus propios juegos de datos locales.

Oracle Analytics Desktop permite visualizar los datos para que pueda centrarse en explorar patrones de datos. Basta con cargar archivos de datos o bien conectarse a *Oracle Applications* o a una base de datos, seleccionar los elementos de interés y permitir que *Oracle Analytics Desktop* busque la mejor manera de visualizarlos. Seleccione una variedad de visualizaciones para ver los datos de un modo determinado.

Oracle Analytics Desktop también le ofrece una vista previa de las capacidades de visualización de autoservicio incluidas en *Oracle Analytics Cloud*, la plataforma de análisis en la nube de Oracle.

4.1.1.1.2 ORACLE ANALYTICS VERSIÓN CLOUD

Oracle Analytics Cloud brinda a los analistas de negocios y consumidores capacidades de análisis modernas, de autoservicio y con tecnología de inteligencia artificial para la preparación de datos, visualización, elaboración de informes empresariales, análisis aumentado y procesamiento/generación de lenguaje natural.

Es fácil de usar, tiene un atractivo visual sorprendente, accede a datos de fuentes gobernadas y personales, y permitir análisis sofisticados que se puedan compartir ampliamente en toda su comunidad de

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

consumidores. Impulsa conocimientos más profundos al integrar el aprendizaje automático y la inteligencia artificial en todos los aspectos del proceso de análisis, lo que hace que su trabajo sea más fácil que nunca.

Oracle tiene como objetivo eliminar la necesidad de traducir sus preguntas para amoldarse al modo en que funcionan los sistemas de análisis. Con NLP y NLG, puede hacer sus preguntas y obtener respuestas rápidamente y con más claridad. Los paneles agrupan contenido de diversas fuentes y sistemas, le presentan una vista personalizada de los datos y le permiten interactuar completamente con esos datos en una experiencia única e integrada.

Oracle ha desarrollado la exploración móvil basada en inteligencia artificial, la inteligencia empresarial móvil aprende lo que a usted le interesa, cuándo y dónde le interesa, y con quién quiere compartir y colaborar.

La preparación integrada de datos aumenta, refina, corrige y crea conjuntos de datos más ricos que mejoran los conocimientos y agudizan su comprensión del mundo en el que trabaja. *Oracle* garantiza que tenga acceso a datos híbridos, ya sea en las instalaciones, en la nube o en su escritorio, proveniente de las principales aplicaciones y almacenes de datos. Más acceso a más datos produce análisis más ricos y diversos; brinda la posibilidad de compartir y publicarlos con quien se desee. Cabe destacar que la herramienta cuenta con analítica predictiva.

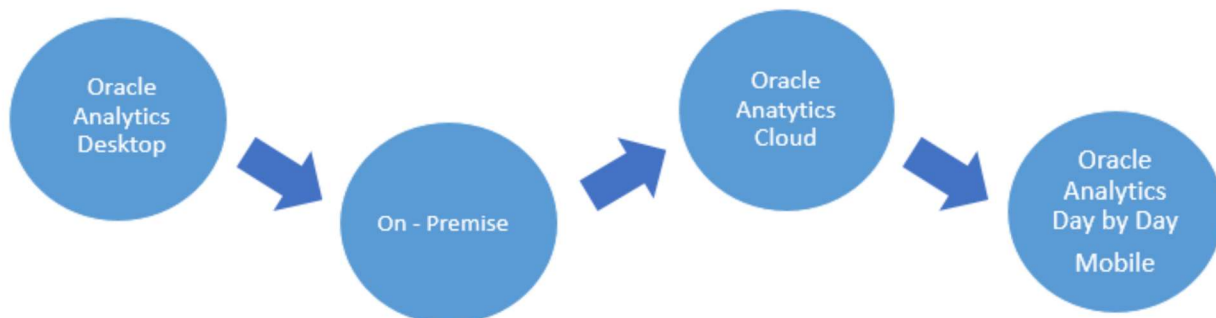
4.1.1.1.3 ORACLE ANALYTICS VERSIÓN ON PREMISE

Es un desprendimiento de las dos, consiste en instalarla en los servidores de la empresa.

4.1.1.1.4 ORACLE ANALYTICS DAY BY DAY

Aplicación *mobile* de *Oracle Analytics*.

Figura 21 Oracle Analytics

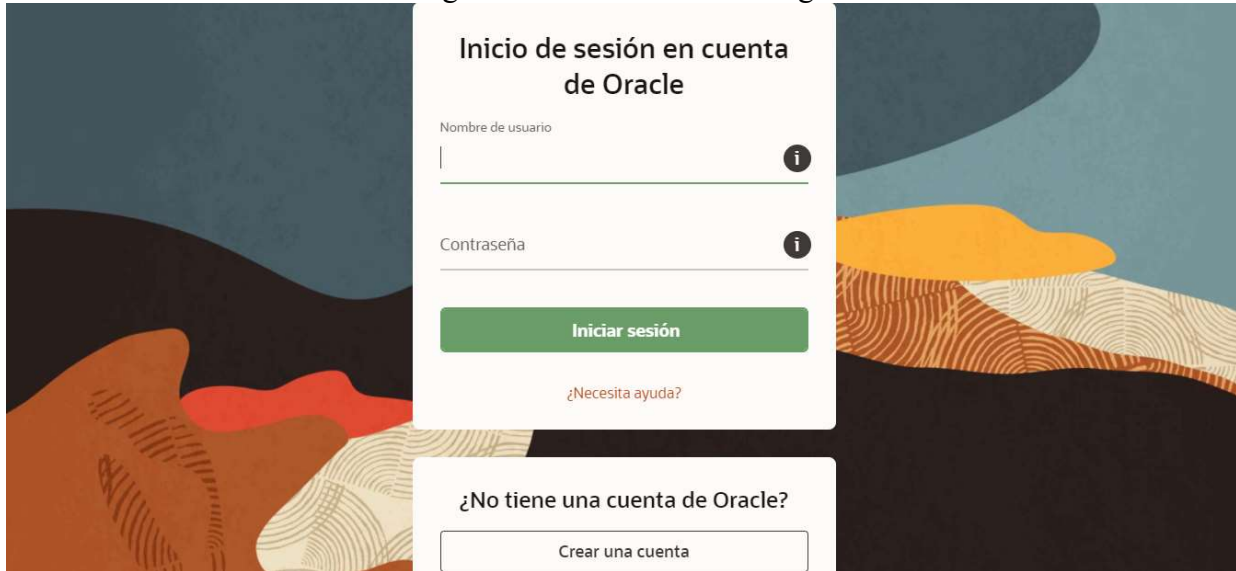


Fuente: Elaboración Propia

4.1.1.2 DETALLE DE LA DESCARGA DE LA APLICACIÓN

Oracle Data Visualization permite descargar una versión gratuita partir de la generación de una cuenta en su página web oficial en la que comercializa sus productos : <https://login.oracle.com/myssso/signon.jsp>
<https://www.oracle.com/ar/solutions/business-analytics/analytics-desktop/oracle-analytics-desktop.html>
Se accede a una cuenta versión *cloud* o *desktop* por 30 días.

Figura 22 Detalle de la descarga



Fuente: (Oracle, s.f.)

A partir de la carga de los datos requeridos, cualquier persona que desee generar un acceso, puede lograrlo. Una vez completados los datos se debe verificar la confirmación en el correo electrónico.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 23 Generación de cuenta
Crear una cuenta Oracle



The image shows a web form for creating an Oracle account. The form is titled "¿Ya tiene una cuenta de Oracle? Iniciar sesión" and "Crear una cuenta Oracle". It includes fields for "Dirección de correo electrónico", "Contraseña", "Volver a escribir contraseña", "país" (set to Argentina), "Nombre" (split into "Nombre de pila" and "Apellidos"), "Cargo", "Teléfono de trabajo", "Nombre de empresa", "Dirección", "Ciudad", "Estado/provincia" (set to "-Seleccionar-"), and "Código postal". There is a "Crear una cuenta" button at the bottom. The form is flanked by abstract, colorful background graphics.

Fuente: (Oracle, s.f.)

Una vez completados los datos se debe verificar la confirmación en el correo electrónico.

Figura 24 Comprobación de usuario



Fuente: (Oracle, s.f.)

Correo de verificación de creación de cuenta.

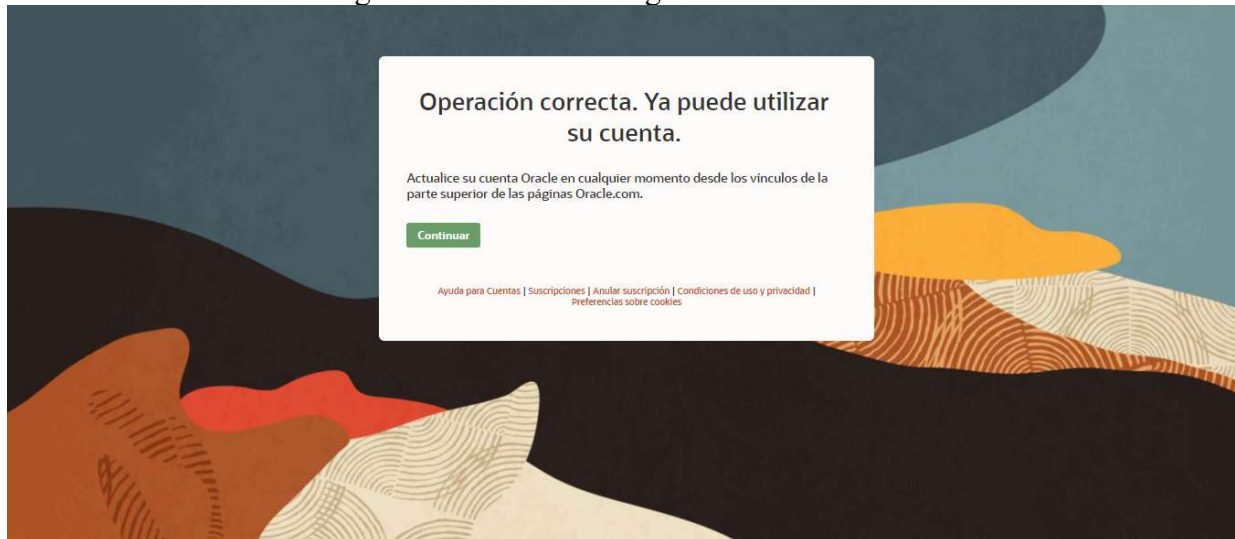
Figura 25 Verificación de correo electrónico



Fuente: (Oracle, s.f.)

Confirmación de instalación.

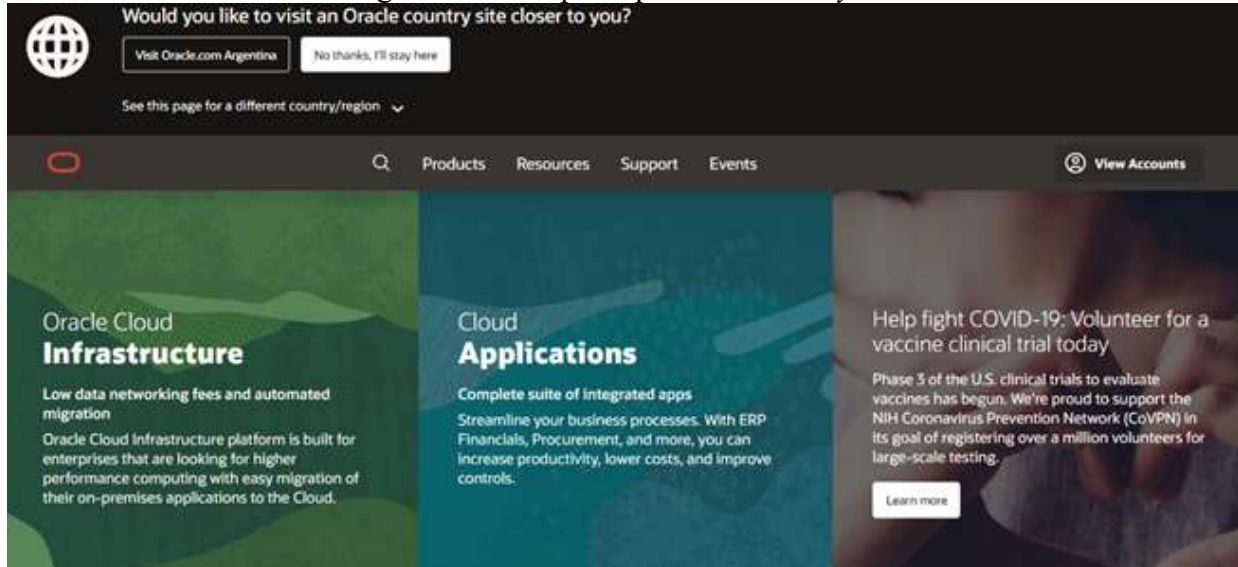
Figura 26 Confirmación generación de cuenta



Fuente: (Oracle, s.f.)

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

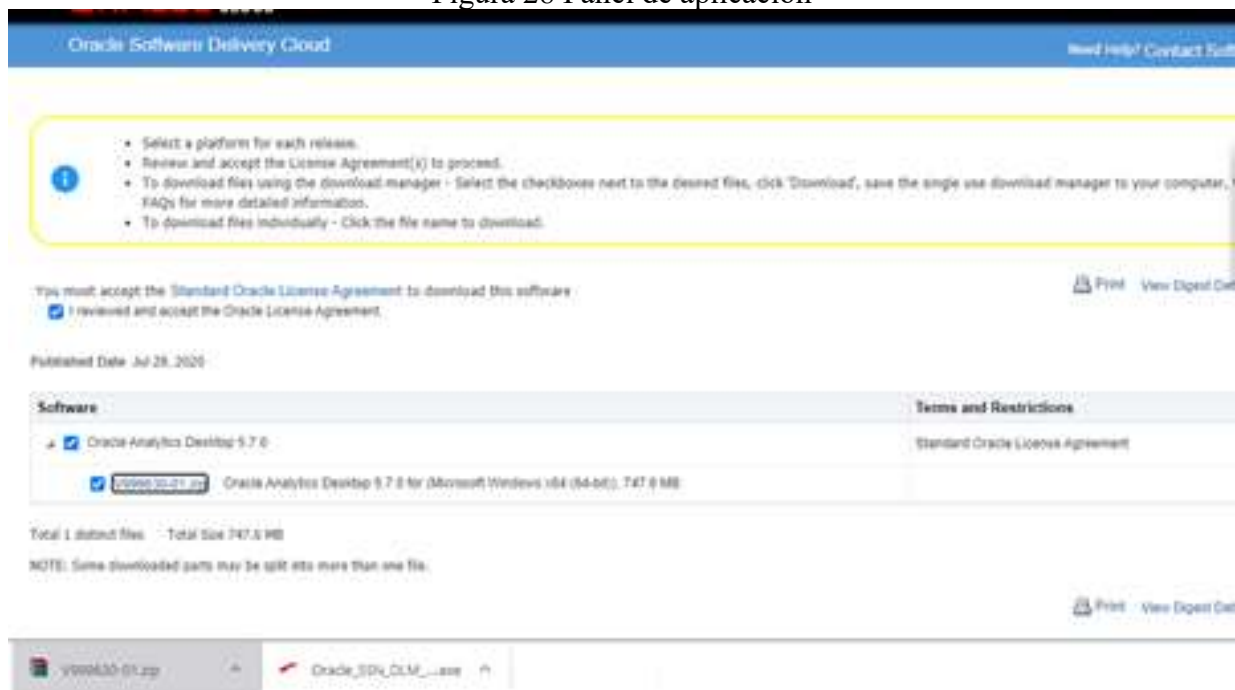
Figura 27 Panel principal *Oracle Analytics*



Fuente: (Oracle, s.f.)

Luego se procede a la descarga de la aplicación.

Figura 28 Panel de aplicación



Fuente: (Oracle, s.f.)

Se genera un archivo con extensión .rar.

Figura 29 Descarga archivo .rar



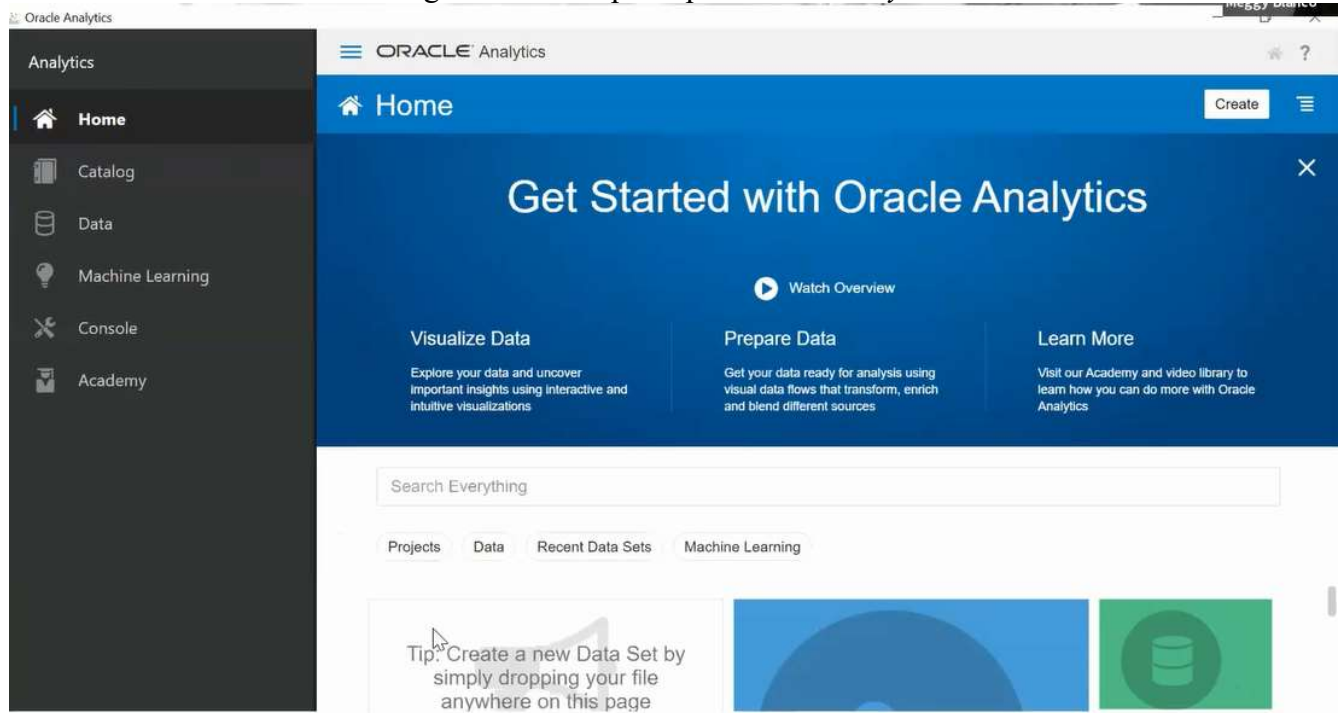
Fuente: (Oracle, s.f.)

Al descomprimir el archivo .rar nos genera un ejecutable .exe el cual debemos ejecutar como administrador. Una vez ejecutado podemos acceder a la aplicación desde el escritorio de la PC.

Se inicia la herramienta y se accede al menú principal, en el *home* y en el “*catalog*”, podemos visualizar los proyectos que tengamos agregados, conexiones y set de datos.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

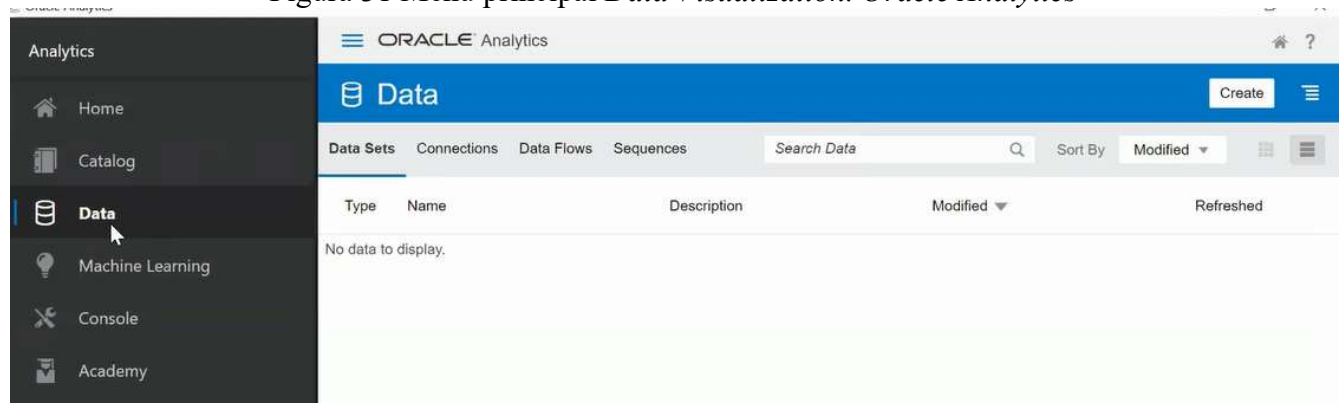
Figura 30 Menu principal *Oracle Analytics*



Fuente: (Oracle, s.f.)

Al hacer click sobre *data*, se despliega una sección referente a la creación de los datos, se parte de las conexiones, en caso de ser necesario, los data set pueden ser atados a una conexión de base de datos o pueden ser subidos como un archivo plano, los *data flow* y las secuencias.

Figura 31 Menu principal *Data Visualization. Oracle Analytics*



Fuente: (Oracle, s.f.)

Dentro de las conexiones de *Data Visualization*, existe una amplia gama de visualizaciones. Se parte de la suite de Oracle, puede ser aplicaciones, base de datos autónoma, contenedores de datos para extraer los archivos, *big data*, *teradata* y *Spark* y finalmente los data set que podamos generar como archivos propios.

Figura 32 Menú generación de conexión. *Oracle Analytics*



Fuente: (Oracle, s.f.)

El *data flow* permite crear un flujo de datos con capacidades para crear ETL. Este es el proceso que permite simular la creación de nuestro *data set*, cuando hacemos este proceso en el *data flow*, se puede tener dos conexiones, dos set de datos y unirlos por *join*, aplicar filtros y realizar cálculos matemáticos, entre otros.

Figura 33 *Data flow*. *Oracle Analytics*



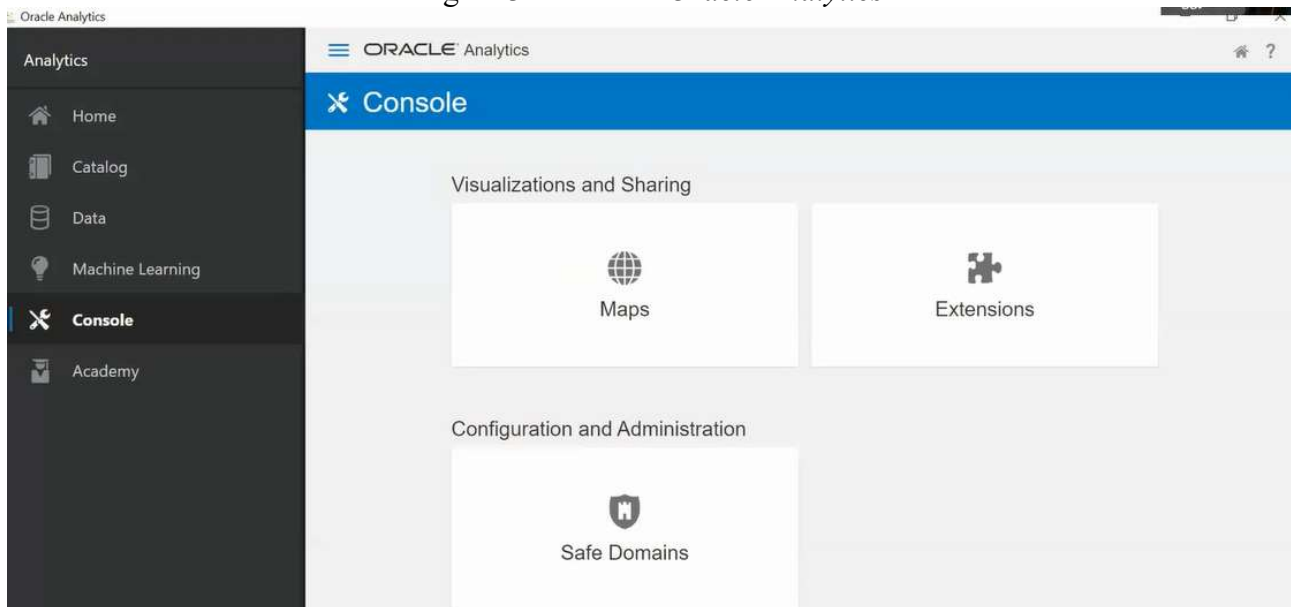
Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

En el panel se distinguen las capacidades, características y componentes que podemos asignar en el *data flow*. En el mismo se distingue un *preview data* del proceso en estado de elaboración. Finalmente, cuenta con un paso de *save data*, que permite crear el *data set* dentro de nuestra aplicación local, el cual se puede almacenar en una base de datos a partir de la generación de un archivo o se puede generar una vista de la base de datos para luego poder consumir.

Esta herramienta, ofrece la posibilidad de aplicar *machine learning*, tiene integrado un algoritmo que permite realizar múltiples posibilidades para el aprendizaje. A partir de un *data flow*, se genera un modelo para implementar *machine learning* dentro de las visualizaciones.

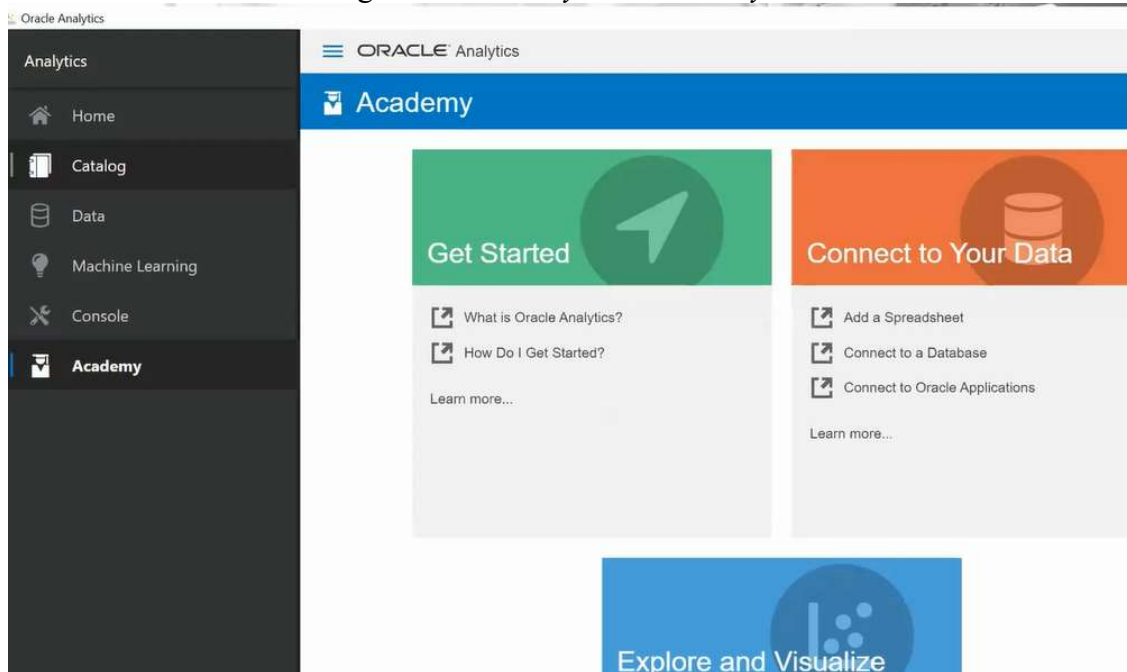
Figura 34 Console. *Oracle Analytics*



Fuente: Elaboración Propia con *Oracle*

En *Console*, permite cargar mapas del mundo para generar visualizaciones territoriales; además cuenta las *Extensions* que podemos añadir visualizaciones que no estén incorporadas. *Safe Domains* es para administrar la seguridad e incorporar dominios que deseemos utilizar que no estén incluidas en la herramienta. Podemos cargar URL para cargar imágenes, cargamos el dominio.

Figura 35 *Academy. Oracle Analytics*



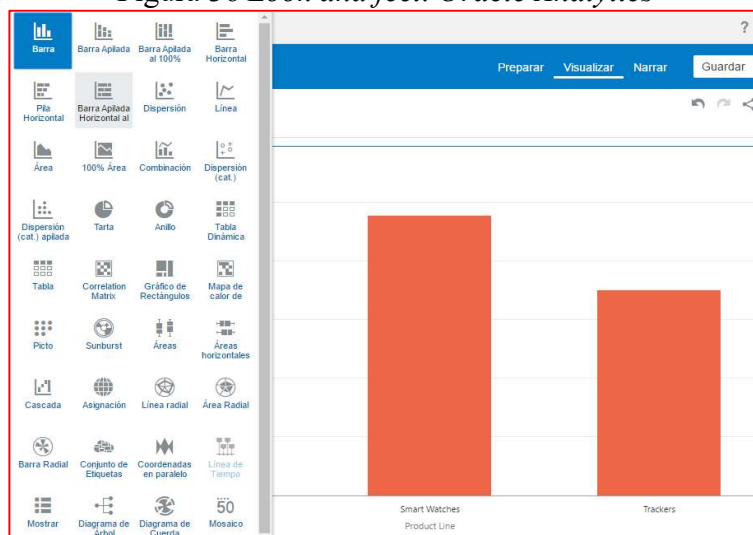
Fuente: Elaboración Propia con *Oracle*

Academy, por su parte, es instrucción sobre el uso de la herramienta. Nos ofrece información a través de la web de *Oracle* y nos permite el acceso a videos con el paso a paso para proceder a la instalación y uso.

4.1.1.3 LOOK AND FEEL: EXPLORACIÓN Y VISUALIZACIÓN

A partir de un *data set* se podrán crear tableros con distintos gráficos descriptivos según el contenido de los datos y la forma en que deseen ser visualizados por el usuario.

Figura 36 *Look and feel. Oracle Analytics*

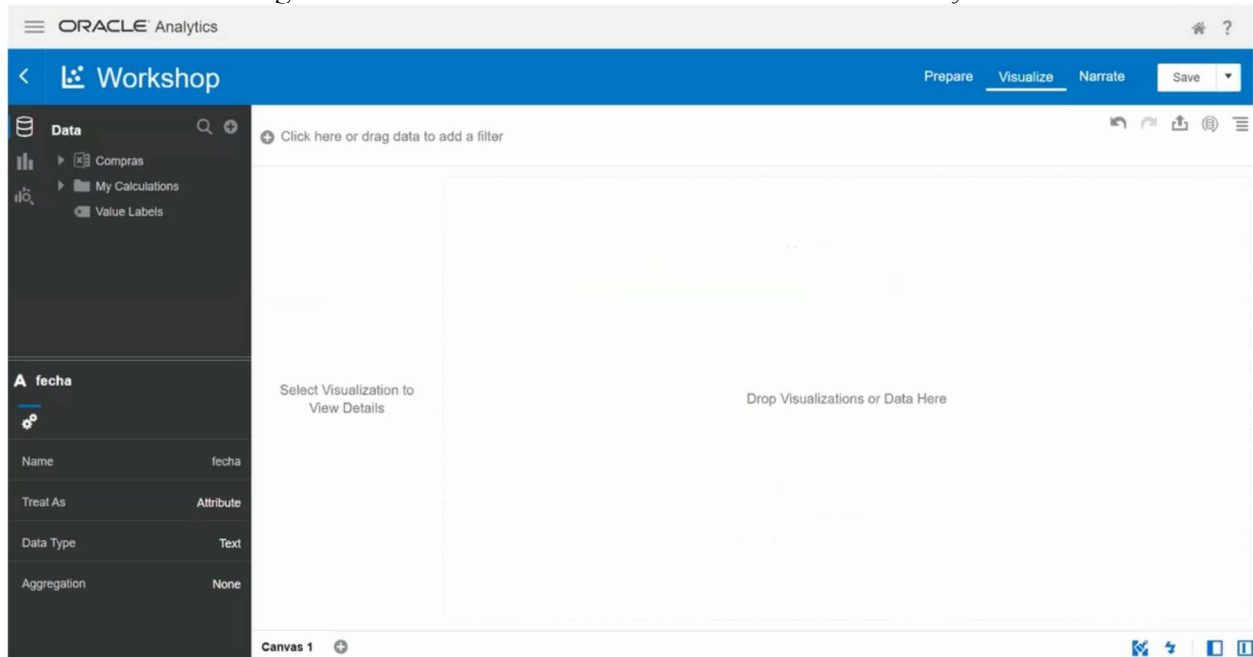


Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

En las fases del proceso analítico está la parte de conexión de set de datos y a partir de él podemos crear nuestras visualizaciones. Las visualizaciones pueden ser múltiples, tablas, torta, tablas comparativas y añadir otros gráficos que no estén incorporados.

Figura 37 Elaboración de la visualización. *Oracle Analytics*



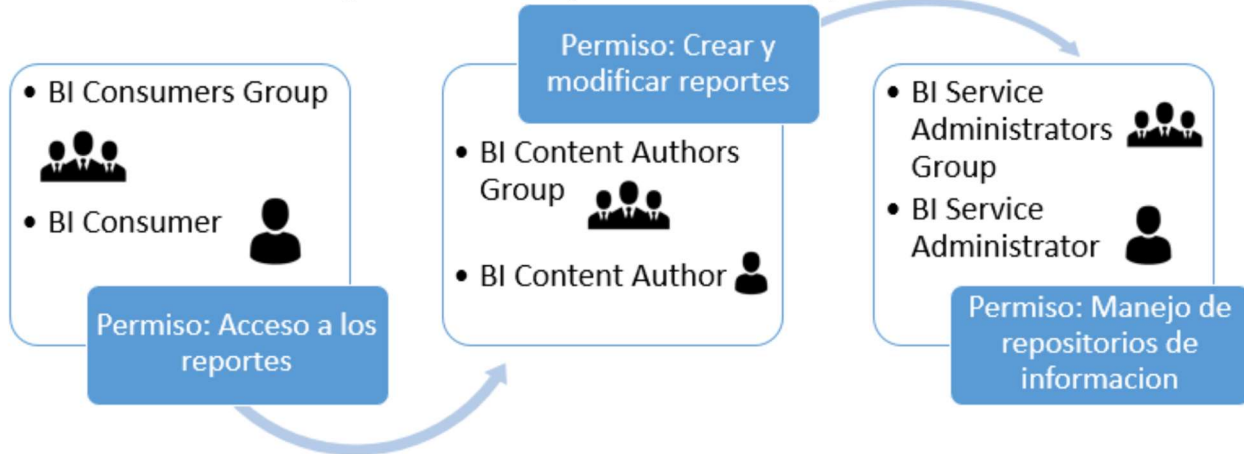
Fuente: Elaboración Propia con *Oracle*

Una vez definido el set de datos, debemos crear el proyecto para hacer nuestras visualizaciones. En el proyecto, se obtiene una vista como (última imagen) en el enlace *Prepare*. Para crearlo tenemos tareas importantes, preparar la data, en caso que tenemos un *data set* y no un *data flow*, podemos definir columnas, filas, asignar nombres a las columnas, generar fórmulas en base a esas columnas, modelos relacionales, cruzar información.

Visualize (visualizar), es donde podemos crear nuestros tableros y ver nuestro *data set* discriminado por diferentes niveles. Los campos tienen características asociadas y se discriminan por dos tipos, métricas, todo aquello que podemos contabilizar y atributos, es una característica propia de la métrica.

Una vez creado el Proyecto, se podrán compartir y exportar los informes construidos con usuarios o grupos de usuarios fácilmente. También podrán ser vistos en la aplicación móvil o ser compartidos en la nube para facilitar la colaboración.

Figura 38 *Look and feel*: Distribución y permisos



Fuente: Elaboración Propia

En lo que respecta a la distribución de informes va a estar supeditado a la versión disponible de la herramienta *Oracle Data Visualization*. En la versión *Desktop*, no tenemos roles ni permisos, como alternativa se puede encriptar el proyecto, asignarle una clave, para que solo puedan acceder ciertos usuarios. Mientras que en la versión *Cloud* y *On premise*, tenemos una serie de permisos de roles y usuarios que pueden asignarse.

4.1.2 ORACLE SQL DEVELOPER

Oracle SQL Developer es una interfaz gráfica de usuario gratuita que permite a los usuarios y administradores de bases de datos realizar sus tareas con menos clics y pulsaciones de teclas. *SQL Developer* es una herramienta de productividad cuyo objetivo principal es ayudar al usuario final a ahorrar tiempo y maximizar el retorno de la inversión en el paquete de tecnología de *Oracle Database*.

SQL Developer proporciona potentes editores para trabajar con SQL, PL/SQL, procedimientos almacenados de Java y XML. Con *SQL Developer*, puede ejecutar consultas, generar planes de ejecución, exportar datos al formato deseado (XML, Excel, HTML, PDF, etc.), ejecutar, depurar, probar y documentar sus programas de base de datos, y mucho más.

A continuación, se muestra un resumen de varias funcionalidades de *SQL Developer*:

- Crear conexiones
- Examinar objetos
- Crear objetos
- Modificar objetos
- Consultar y actualizar datos

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

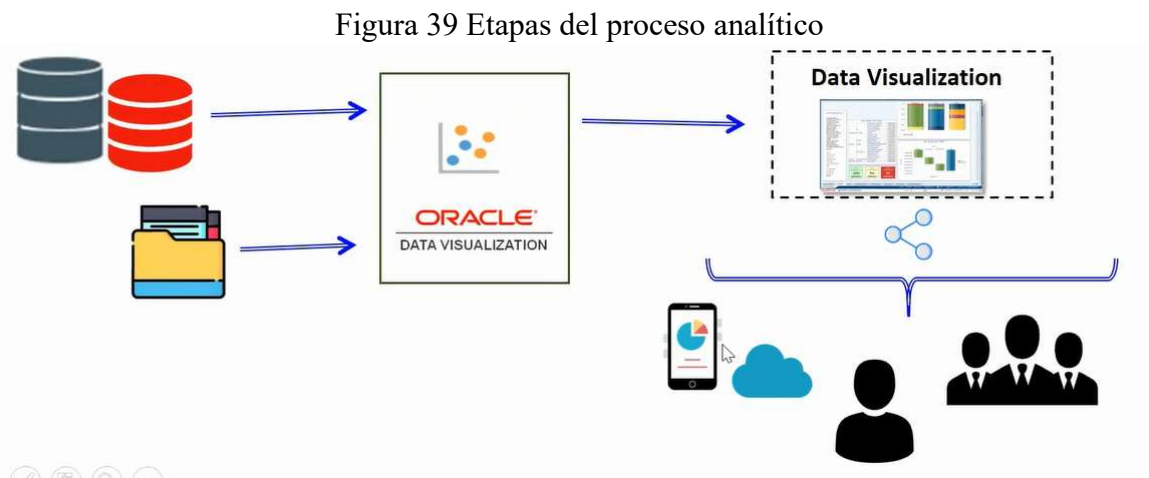
- Exportar datos y DDL, importar datos
- Ejecutar y crear informes

4.2 ANÁLISIS DE DATOS

En esta fase se integraron, transformaron y normalizaron los sets de datos de las 120 sucursales de la entidad financiera para la ejecución en el prototipo utilizando las técnicas descritas en los párrafos anteriores.

La fase de análisis de datos comprende la recolección de los datos de los colaboradores de la entidad financiera en la cual se realiza el caso de estudio respectivo y que fueron obtenidos desde diferentes sistemas de información de la empresa. Las variables a estudiar, por colaborador, son: cantidad de horas en actividades de formación técnica atravesadas, los resultados de las evaluaciones de desempeño, el resultado de la medición de compromiso o *engagement*, lo que se considera en la entidad financiera como una variable *proxy* de clima laboral de la sucursal en la que se desempeñan, y el ausentismo de cada una de las personas.

Dado que los indicadores buscados varían a través del tiempo, he tomado como muestra el total de colaboradores que se ha desempeñado ininterrumpidamente durante los años 2016, 2017, 2018 y 2019 asciende a 952 (novecientos cincuenta y dos).



Fuente: Elaboración Propia

4.3 PREPARACIÓN DE LOS DATOS

En esta etapa intervienen una serie de pasos y componentes fundamentales que acompañan el proceso previo para implementar correctamente las técnicas de *big data* y *people analytics*. El primer paso

comprende la comprensión del dominio del caso de estudio junto con un detalle correcto de los objetivos puesto que es fundamental tener muy claros los límites de la investigación.

Para el segundo paso, es clave para la creación del set de datos la correcta selección e integración de los datos provenientes de distintas fuentes, los datos que se necesitan están esparcidos en diferentes registros de documentos de Microsoft Excel y algunos de ellos son dispares respecto a la información que poseen, estos van desde colecciones de documentos, correos electrónicos, bases de datos de procesos, bases de datos transaccionales propias de la entidad financiera, registros *web*, etc. Es importante homogeneizar todos los formatos de los datos para que la información sea más fácil de procesar y analizar.

Dando continuidad al tercer paso, limpieza y procesamiento de datos, es importante eliminar todo el ruido y datos aislados como inconsistencias y duplicidad de contenido, los datos no depurados confunden el proceso y conllevan a resultados inválidos o poco confiables. Los pasos implementados hasta este momento tienen el objetivo de mejorar la calidad de los datos y los resultados de la utilización de técnicas de *big data* y *people analytics* en el proceso de esta investigación.

4.4 CREACIÓN DE LA BASE DE DATOS

El siguiente paso consiste en la preparación de los datos. En esta fase se hace necesario construir una base de datos, la cual va a contener todas las características consideradas candidatas para estimar el valor de una variable que se espera predecir.

Se incluyen tareas de selección de los datos a los que se va a aplicar la técnica modelo, la tarea de limpieza de los mismos para alcanzar el nivel de calidad, la tarea de construir datos adicionales y la tarea de dar formato a los datos.

Una vez que se obtiene la estructura final con la información, se procede con la transformación de los datos al *Oracle Data Visualization*. De esta manera se crea la base de datos para los modelos de minería de datos.

4.5 CONSTRUCCIÓN DE ETL

Para la construcción del proceso ETL se integraron los datos de la empresa del caso de investigación. En las siguientes secciones se definen y explican los procesos realizados en dicho proceso.

4.5.1 PREPARAR LOS DATOS ETL

ETL por sus siglas significa extraer, transformar y cargar. Es uno de los procesos importantes dentro de las empresas para mover datos desde múltiples fuentes, cambiar el formato y depurar datos innecesarios

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

para cargarlos en otra base de datos, *data mart*, o *data warehouse*, este proceso se usa para el análisis de metadatos o para apoyar un proceso de negocio.

4.5.2 SELECCIÓN DE LA FUENTE DE DATOS

Se trabajó con datos de diferentes fuentes y repositorios, se necesitó pre procesar y preparar los datos antes del desarrollo de modelos predictivos. Se detectó que los datos carecían de ciertos valores y que estos eran erróneos y tenían diferentes formatos en su estructura.

Al desarrollar la investigación se encontraron algunas limitaciones, no solo en el acceso a los datos, sino también en la interpretación de los mismos.

Una primera limitación proviene del acceso a los datos y documentos, los mismos no fueron proporcionados en su totalidad por la empresa. La confidencialidad de dicha información fue una característica principal para el acceso a los datos.

La primera parte del proceso ETL consistió en extraer los datos desde los sistemas de origen, para la investigación y caso de estudio de la empresa, los datos provienen de carpetas de los empleados del área de recursos humanos, archivos planos en .txt, archivos de Excel y sistema *Oracle Human Capital Management*, propio de la entidad financiera.

Una vez que se identificaron los orígenes y fuentes de datos, se procede a la construcción de las tablas físicas del modelo y al desarrollo de los procesos de carga. Se generan todos los registros necesarios con un periodo de cuatro años, que va desde el 1º de enero de 2016 hasta el 31 de diciembre de 2019.

La fase de carga va a interactuar directamente con la base de datos de destino. Al realizar esta operación se aplicarán todas las restricciones que se definieron anteriormente, por ejemplo, valores únicos, integridad referencial, campos obligatorios, rangos de valores. Éstas restricciones contribuyen a que se garantice la calidad de los datos en el proceso ETL y deben ser tenidos en cuenta.

4.5.3 CONSTRUCCIÓN DEL *DATA WAREHOUSE*

Con el resultado de los dos procesos anteriores se arma el archivo final con el resultado del consolidado de la base de datos. A continuación, se detalla el diagrama Copo de nieve.

Figura 40 Construcción del data warehouse



Fuente: Elaboración Propia

En esta fase, se debe comenzar con una matriz donde se determina la dimensionalidad de cada indicador, para luego especificar los diferentes grados de detalle dentro de cada concepto del negocio.

En el modelo dimensional se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones, con base en las consultas hechas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos de negocio.

El Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Cada Modelo Dimensional está compuesta por una tabla con una llave combinada, llamada tabla de hechos, y con un conjunto de tablas más pequeñas llamadas tablas de dimensiones. Los elementos de estas tablas se pueden definir de la siguiente manera (Wikipedia, s.f.):

- Hechos: es una colección de piezas de datos y datos de contexto. Cada hecho representa una parte del negocio, una transacción o un evento.
- Dimensiones: es una colección de miembros, unidades o individuos del mismo tipo.
- Medidas o métricas: son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.
- Cada punto de entrada a la tabla de hechos está conectado a una dimensión, lo que permite determinar el contexto de los hechos.

Entre las principales fortalezas para este modelo se pueden mencionar:

- La resistencia a cambios en la conducta del usuario. El diseño lógico de este modelo puede realizarse independientemente de los patrones esperados de consulta, pues todas las dimensiones son generadas como puntos de entrada simétricos a la tabla de hechos.
- Flexible para aceptar datos nuevos e inesperados. La tabla de hechos y las tablas de dimensiones pueden ser alteradas simplemente agregando los registros de los nuevos datos en la tabla correspondiente. La información no tiene que ser cargada nuevamente. Las herramientas para

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

generar reportes o consultas no necesitan ser reprogramadas para adaptarse a los cambios. Y, finalmente, las aplicaciones utilizadas continúan ejecutándose sin cambios en su rendimiento.

Se denomina modelo estrella o esquema de estrella-uniión a un modelo dimensional como una tabla de hechos rodeada por las tablas de dimensiones.

Una base de datos dimensional se puede concebir como un cubo de tres o cuatro dimensiones (OLAP), en el que los usuarios pueden acceder a una porción de la base de datos a lo largo de cualquiera de sus dimensiones. Tal como se mencionó en esta investigación en el “II.3 Gestores de base de datos (SGBD)”

Dado que es muy común representar a un modelo dimensional como una tabla de hechos rodeada por las tablas de dimensiones, frecuentemente se le denomina también modelo estrella o esquema de estrella-uniión.

Otra variante es la que se conoce como *snowflake* o copo de nieve, en donde se presentan ramificaciones a partir de las tablas de dimensiones y no solo a partir de la tabla de hechos.

4.6 IMPLEMENTACIÓN DE LA SOLUCIÓN

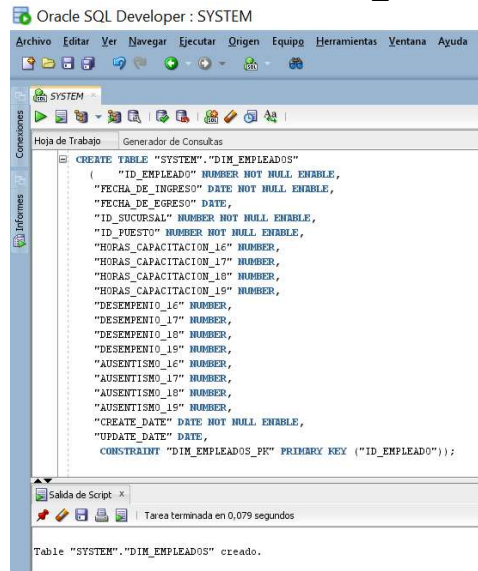
Para la ejecución del proyecto analítico se realizó un modelado estrella para el cual fueron creadas las dimensiones DIM_EMPLEADOS, DIM_PUESTOS y DIM_SUCURSALES mediante el gestor de base de datos de *Oracle SQL Developer*. Posteriormente se utilizó la herramienta *Oracle Analytics* para exportar dichas dimensiones como set de datos a fin de generar la tabla de hechos FACT_ANALYTICS_RRHH en la conexión de la base de datos de *Oracle* por medio de la creación de un *data flow*.

4.6.1 CREACIÓN DE TABLAS DIMENSIONALES ORACLE SQL DEVELOPER

4.6.1.1 DIM_EMPLEADOS

Posterior a la configuración de la conexión local se ejecutó el script que se visualiza a continuación para la creación de la tabla DIM_EMPLEADOS:

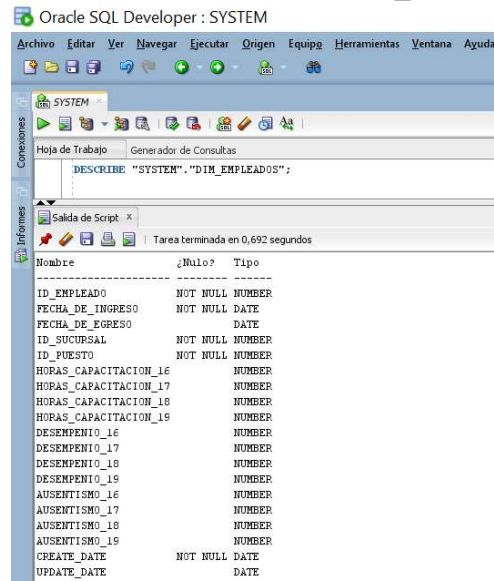
Figura 41 Creación de la tabla DIM_EMPLEADOS



Fuente: Elaboración Propia con *Oracle*

A fin de validar la creación de la DIM_EMPLEADOS bajo la estructura definida se utilizó el script describe tal como se muestra:

Figura 42 Validación de la tabla DIM_EMPLEADOS



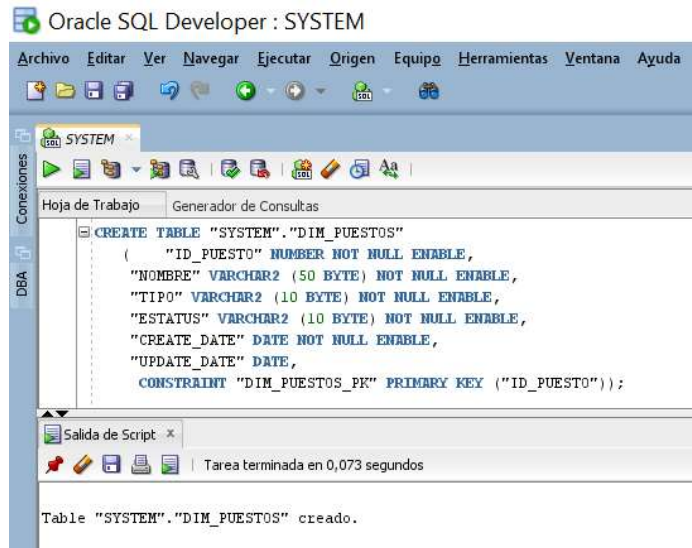
Fuente: Elaboración Propia con *Oracle*

4.6.1.2 DIM_PUESTOS

Posterior a la configuración de la conexión local se ejecutó el script que se visualiza a continuación para la creación de la tabla DIM_PUESTOS:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

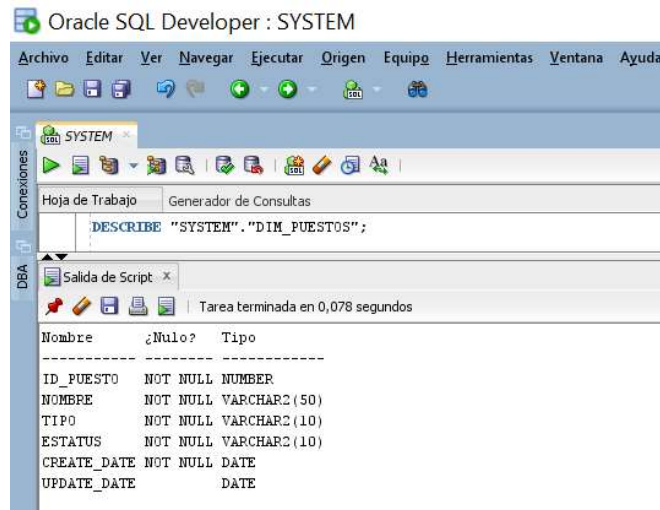
Figura 43 Creación de la tabla DIM_PUESTOS



Fuente: Elaboración Propia con *Oracle*

A fin de validar la creación de la DIM_PUESTOS bajo la estructura definida se utilizó el script describe tal como se muestra:

Figura 44 Validación de la tabla DIM_PUESTOS

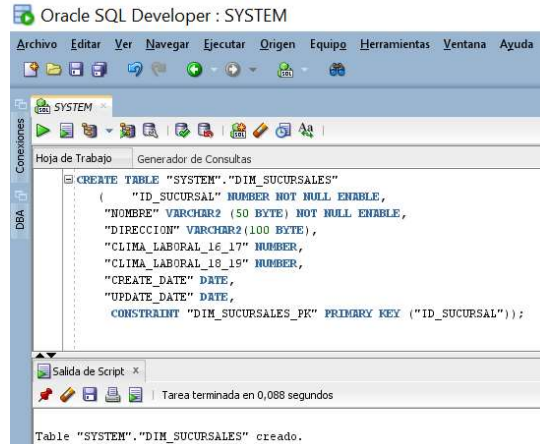


Fuente: elaboración propia con *Oracle*

4.6.1.3 DIM_SUCURSALES

Posterior a la configuración de la conexión local se ejecutó el script que se visualiza a continuación para la creación de la tabla DIM_SUCURSALES:

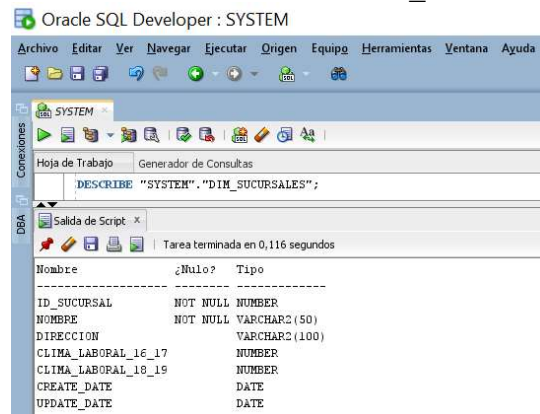
Figura 45 Creación de la tabla DIM_SUCURSALES



Fuente: elaboración propia con *Oracle*

A fin de validar la creación de la DIM_SUCURSALES bajo la estructura definida se utilizó el script describe tal como se muestra:

Figura 46 Validación de la tabla DIM_SUCURSALES



Fuente: elaboración propia con *Oracle*

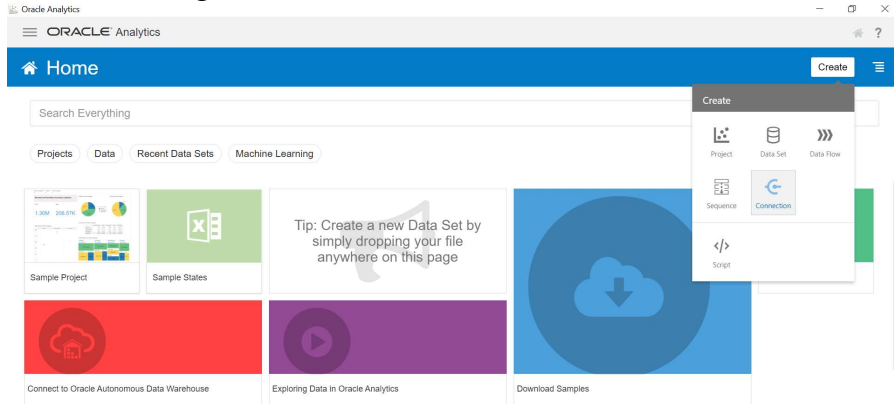
4.6.2 PROCESAMIENTO DE DATOS EN *ORACLE ANALYTICS*

4.6.2.1 CONFIGURACIÓN DE LA CONEXIÓN A LA BASE DE DATOS ORACLE

La conexión a la base de datos se realizó de la siguiente manera:

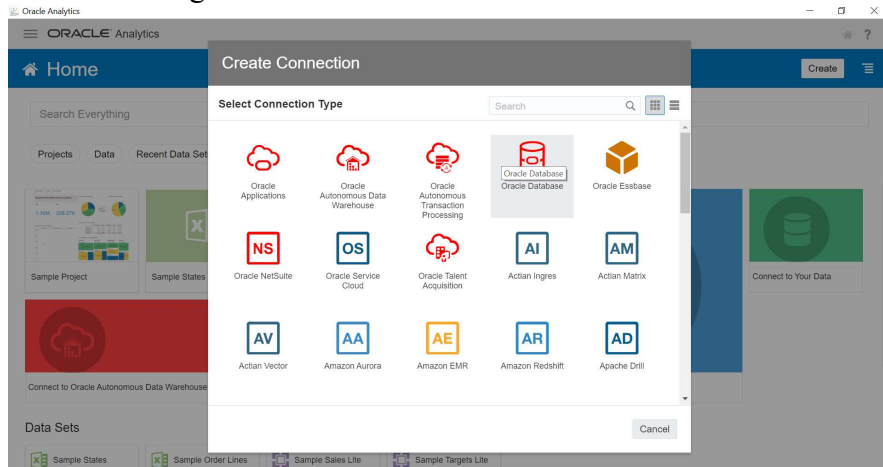
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 47 Conexión de base de datos. Paso 1



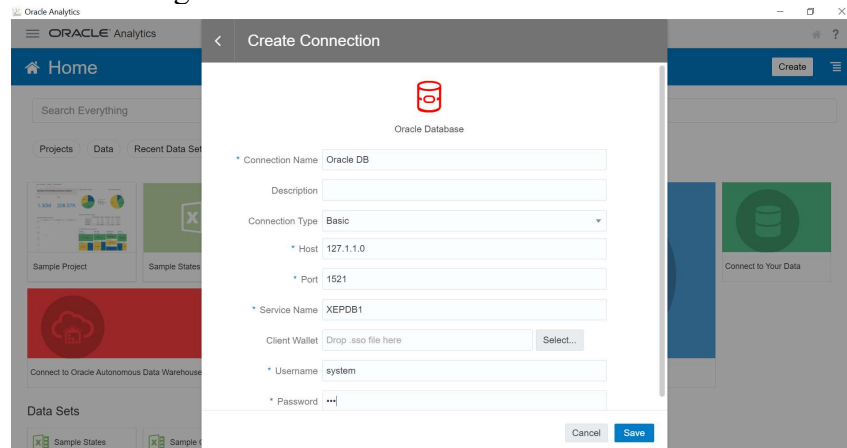
Fuente: elaboración propia con *Oracle*

Figura 48 Conexión de base de datos. Paso 2



Fuente: elaboración propia con *Oracle*

Figura 49 Conexión de base de datos. Paso 3

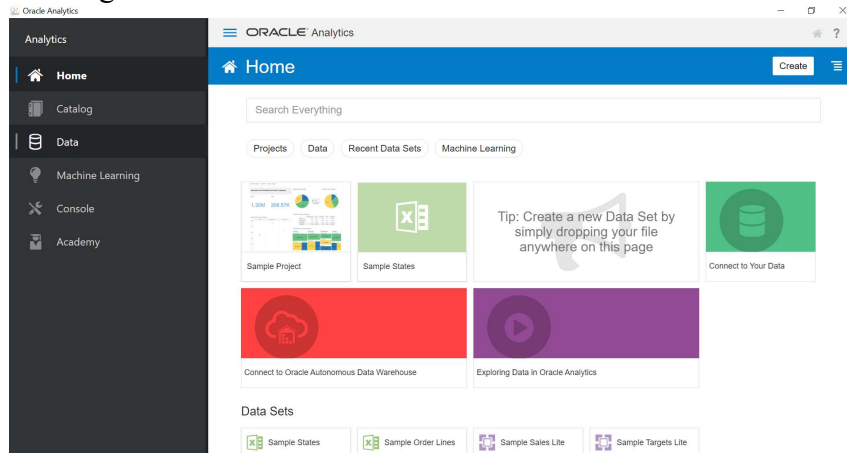


Fuente: elaboración propia con *Oracle*

La configuración que se visualiza es una conexión local por lo cual el *host* utilizado es 127.1.1.0.

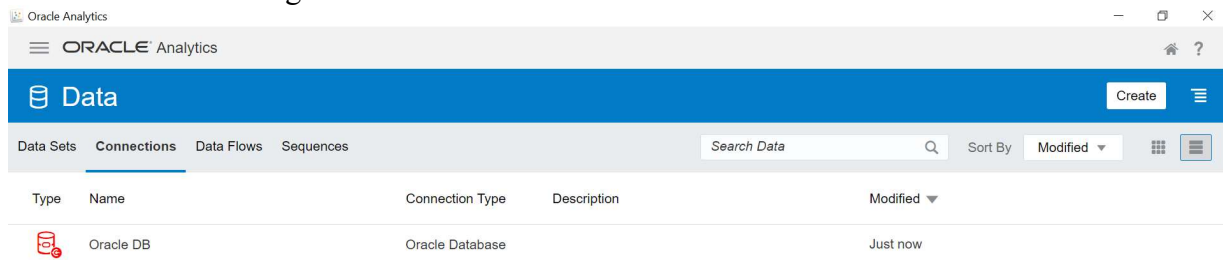
Para la validación de la creación de la conexión se siguieron los siguientes pasos:

Figura 50 Validación de creación de la conexión. Paso 1



Fuente: elaboración propia con Oracle

Figura 51 Validación de creación de la conexión. Paso 2

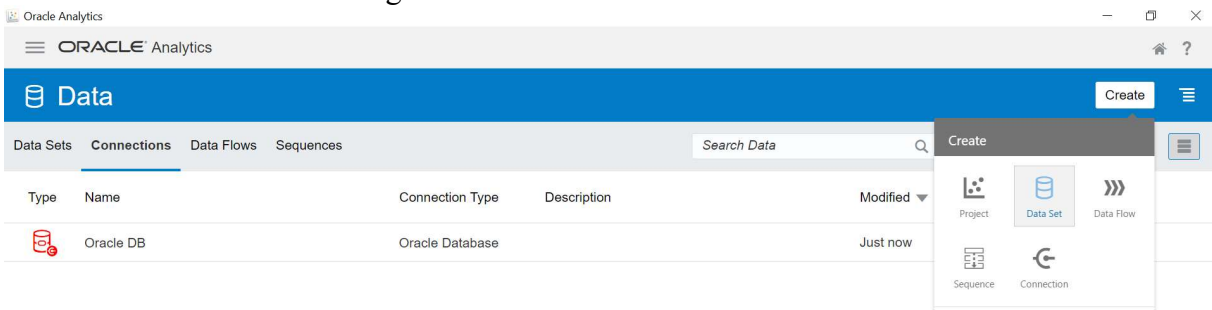


Fuente: elaboración propia con Oracle

4.6.3 CREACIÓN DE SET DE DATOS EN ORACLE ANALYTICS

4.6.3.1 DIM_EMPLEADOS

Figura 52 Creación set de datos. Paso 1

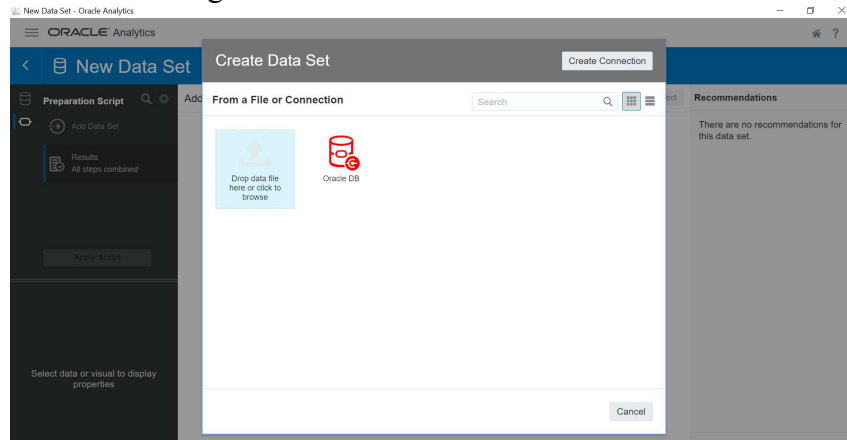


Fuente: elaboración propia con Oracle

Puesto que para esta tesis los sets de datos se alojan en una base de datos de tipo Oracle se debió crear previamente la conexión a la misma:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

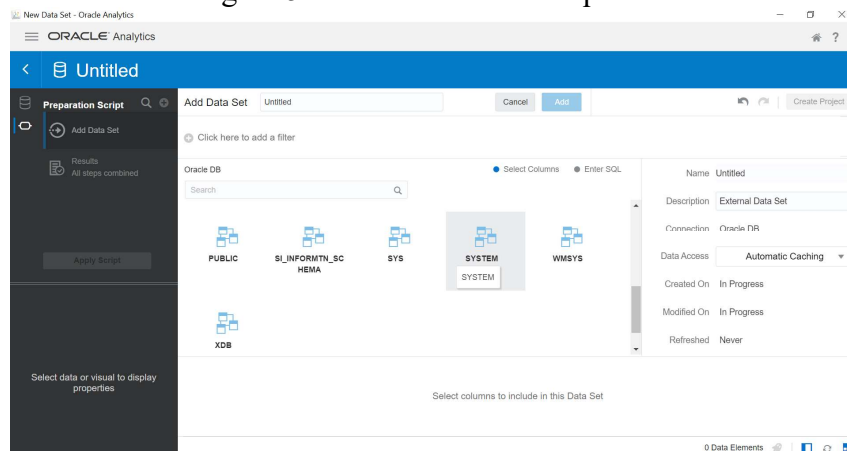
Figura 53 Creación set de datos. Paso 2



Fuente: elaboración propia con *Oracle*

Luego de definir la conexión se visualizan los esquemas que se encuentran en la base de datos, para el proyecto se utilizó el esquema “SYSTEM”:

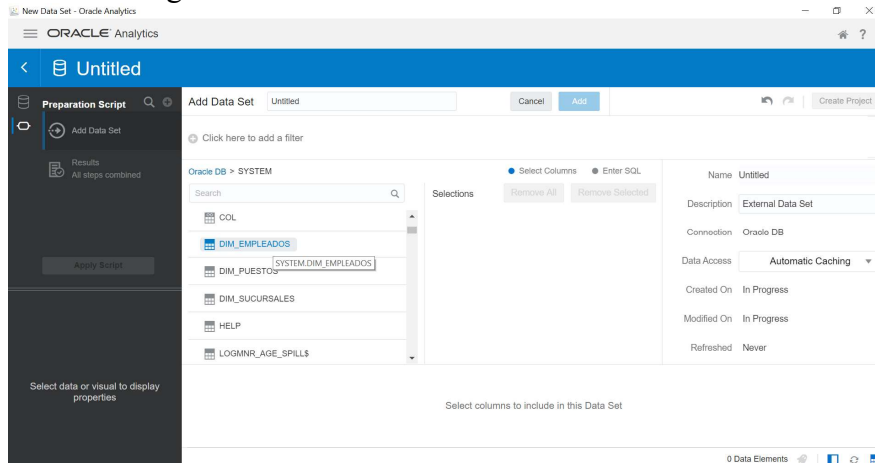
Figura 54 Visualización de esquemas



Fuente: elaboración propia con *Oracle*

Finalmente, al elegir el esquema se visualizan todas las tablas contenidas en él. Para la creación del *data set* DIM_EMPLEADOS se selecciona la tabla DIM_EMPLEADOS como fuente:

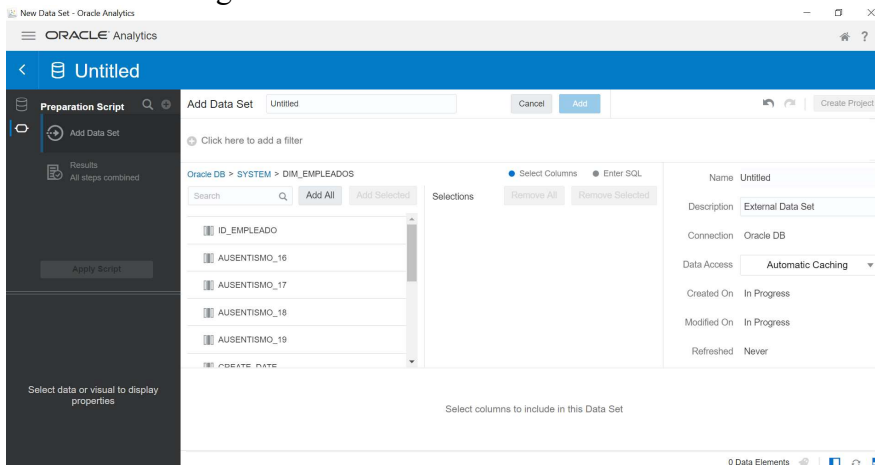
Figura 55 Visualización de tablas de contenidos



Fuente: elaboración propia con *Oracle*

Una vez seleccionada la tabla DIM_EMPLEADOS se deben seleccionar los atributos que se desean incorporar en el data set para el dimensional:

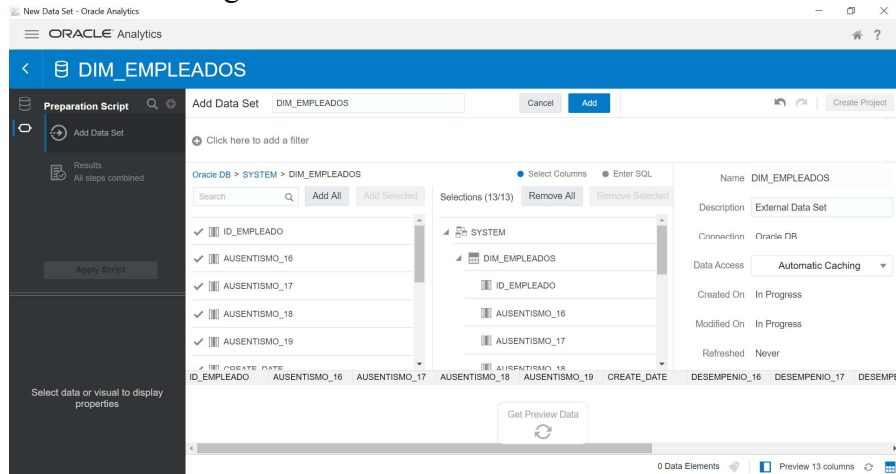
Figura 56 Selección de atributos. Paso 1



Fuente: elaboración propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

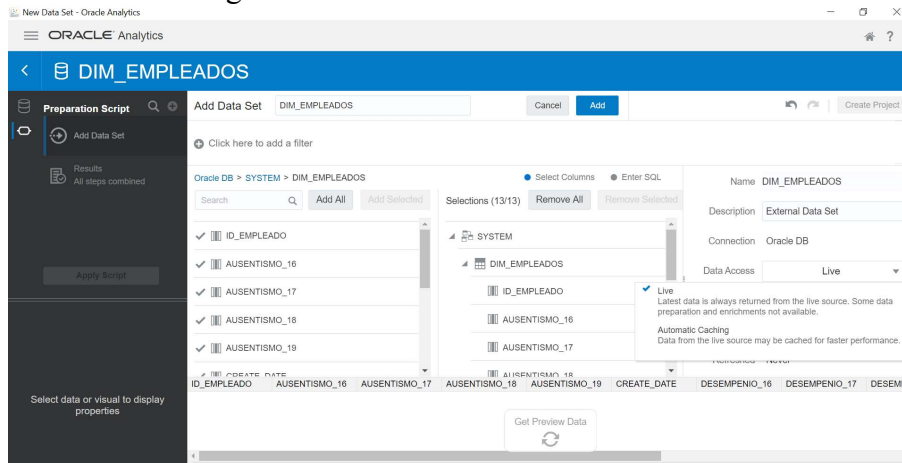
Figura 57 Selección de atributos. Paso 2



Fuente: elaboración propia con Oracle

Por *default* la conexión contra la base de datos es de tipo “Automatic Caching”, para el proyecto fue configurada como *Data Access* “Live”:

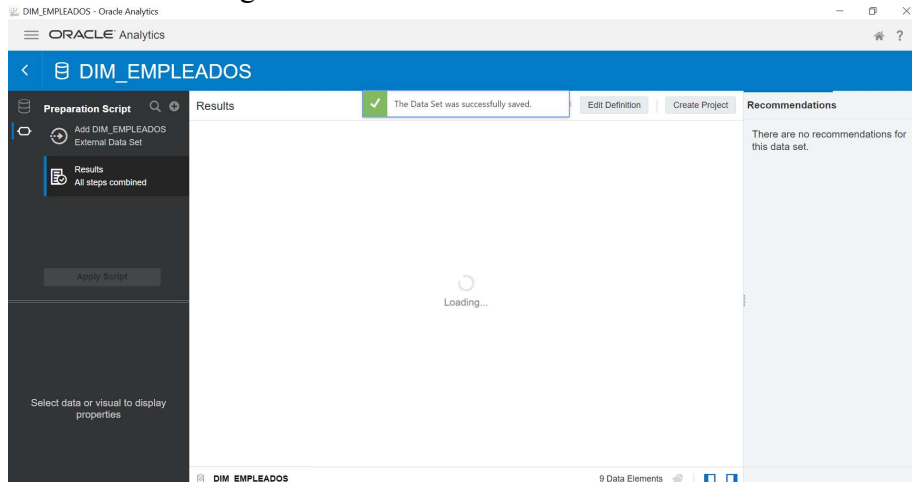
Figura 58 Conexión contra base de datos



Fuente: elaboración propia con Oracle

Luego de establecer todos los parámetros y haber definido el nombre del *data set* se confirma la creación haciendo click sobre el botón “Add”

Figura 59 Confirmación de la conexión



Fuente: elaboración propia con *Oracle*

Validación de la creación del set de datos “DIM_EMPLEADOS”

Figura 60 Validación creación set de datos

The screenshot shows the Oracle Analytics interface displaying a table of data sets. The table has columns for Type, Name, Description, Modified, and Refreshed. The data set 'DIM_EMPLEADOS' is listed as an External Data Set, created 6 minutes ago and refreshed just now.

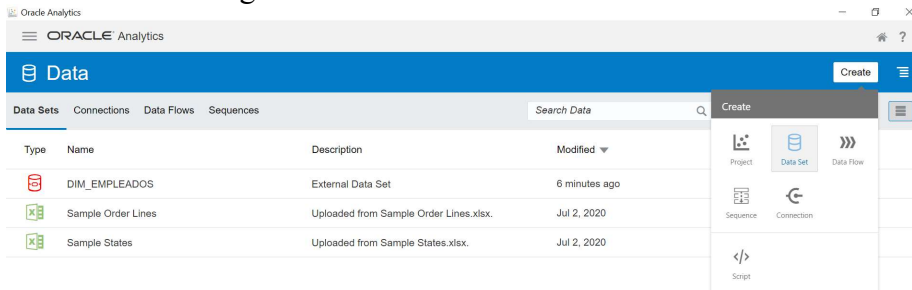
Type	Name	Description	Modified	Refreshed
External Data Set	DIM_EMPLEADOS	External Data Set	6 minutes ago	Just now
Sample Order Lines	Sample Order Lines	Uploaded from Sample Order Lines.xlsx	Jul 2, 2020	Jul 2, 2020
Sample States	Sample States	Uploaded from Sample States.xlsx	Jul 2, 2020	Jul 2, 2020

Fuente: elaboración propia con *Oracle*

Los *data sets* creados posteriormente se realizaron siguiendo los pasos ejecutados anteriormente tal como se muestra a continuación.

4.6.3.2 DIM_PUESTOS

Figura 61 Creación set de datos. Paso 1

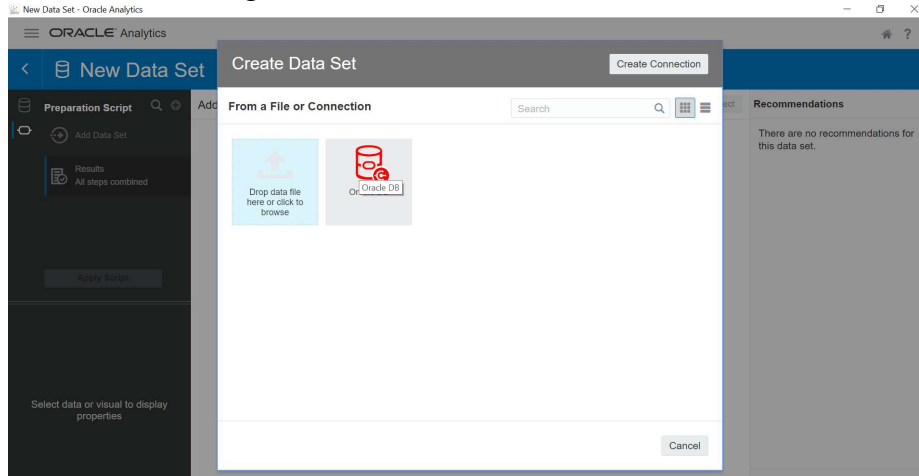


Fuente: elaboración propia con *Oracle*

Puesto que para esta tesis los sets de datos se alojan en una base de datos de tipo *Oracle* se debió crear previamente la conexión a la misma:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

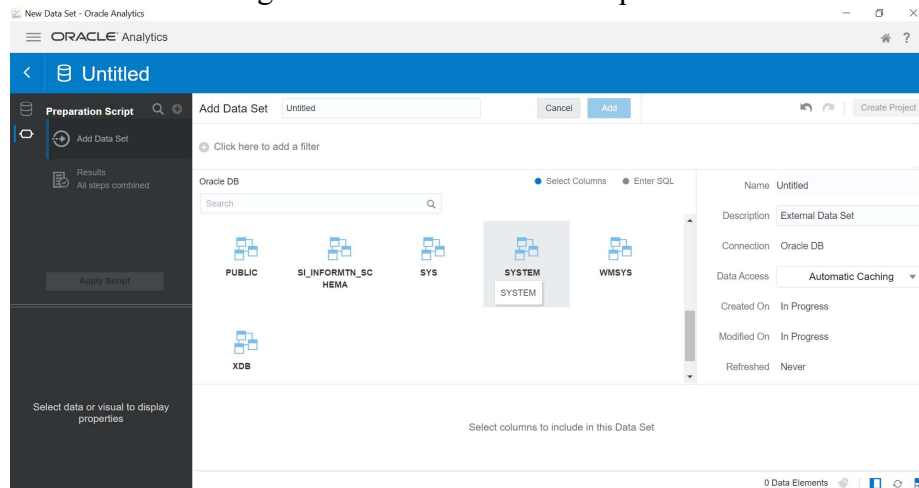
Figura 62 Creación set de datos. Paso 2



Fuente: elaboración propia con *Oracle*

Luego de definir la conexión se visualizan los esquemas que se encuentran en la base de datos, para el proyecto se utilizó el esquema “SYSTEM”:

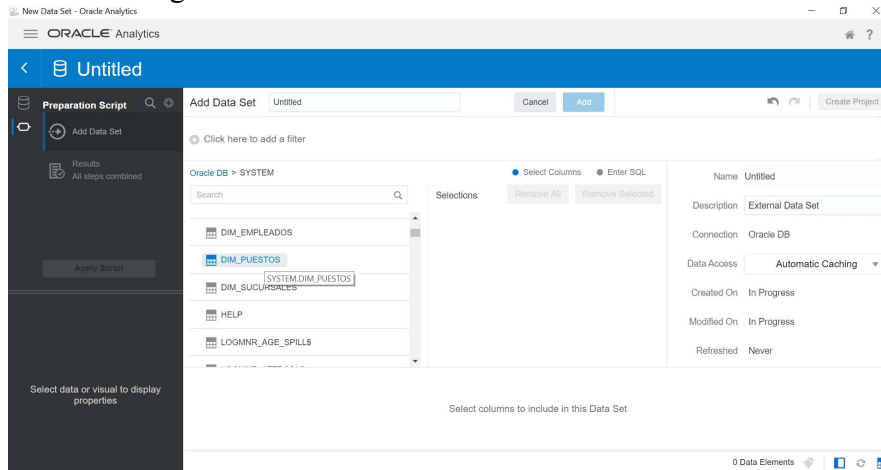
Figura 63 Visualización de esquemas



Fuente: elaboración propia con *Oracle*

Finalmente, al elegir el esquema se visualizan todas las tablas contenidas en él. Para la creación del *data set* DIM_PUESTOS se selecciona la tabla DIM_PUESTOS como fuente:

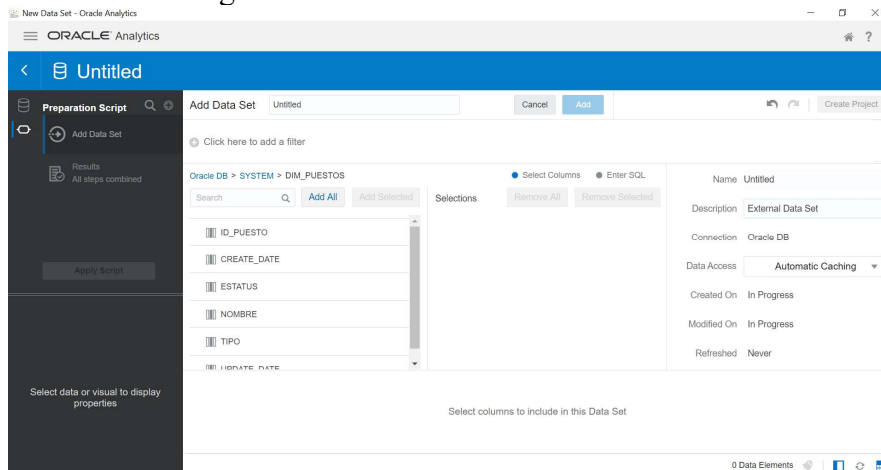
Figura 64 Visualización de tablas de contenidos



Fuente: elaboración propia con *Oracle*

Una vez seleccionada la tabla DIM_PUESTOS se deben seleccionar los atributos que se desean incorporar en el *data set* para el dimensional:

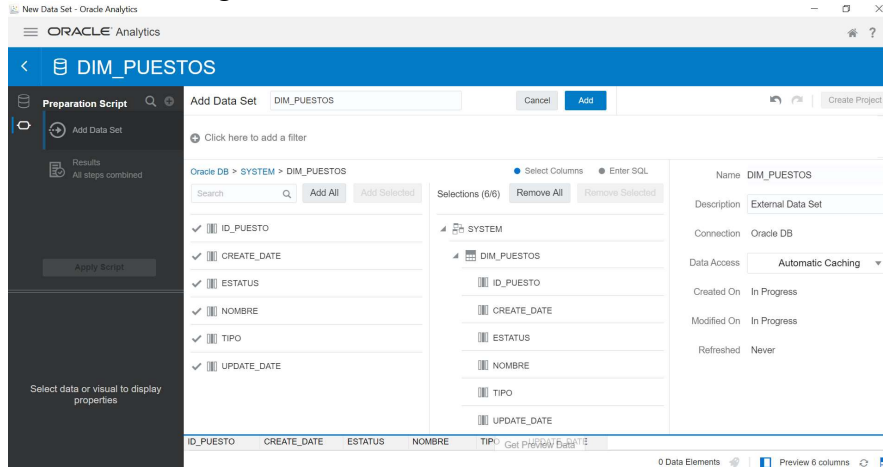
Figura 65 Selección de atributos. Paso 1



Fuente: elaboración propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

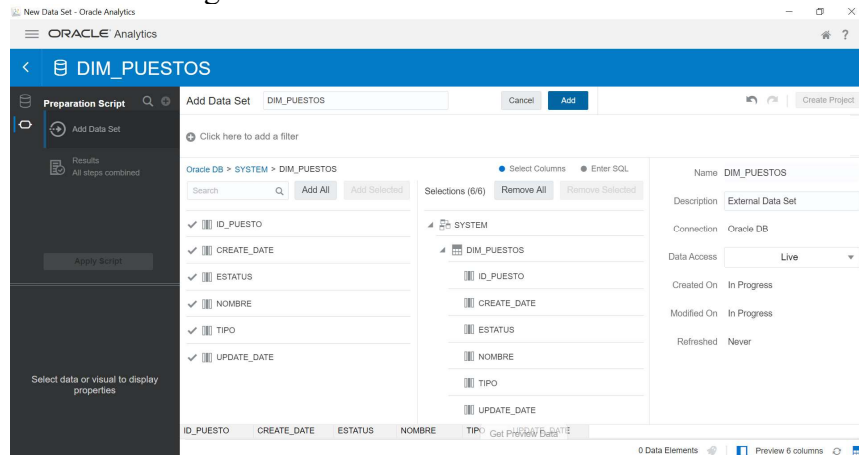
Figura 66. Selección de atributos. Paso 2



Fuente: elaboración propia con *Oracle*

Por *default* la conexión contra la base de datos es de tipo “*Automatic Caching*”, para el proyecto fue configurada como *Data Access* “*Live*”:

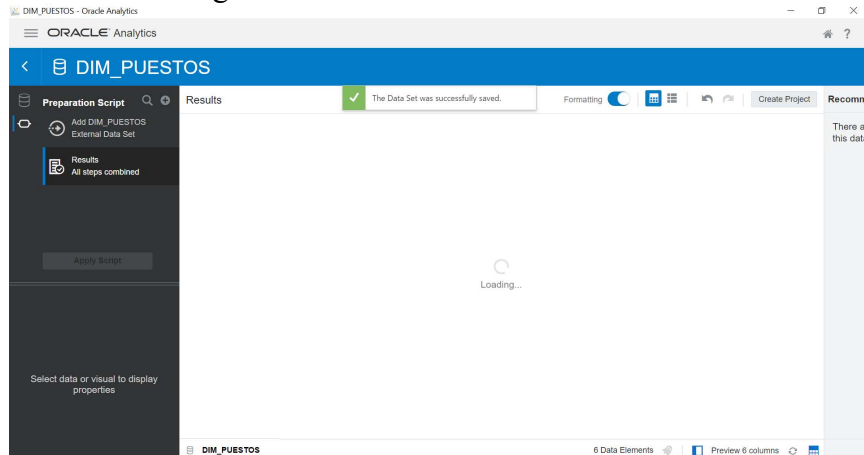
Figura 67 Conexión contra base de datos



Fuente: elaboración propia con *Oracle*

Luego de establecer todos los parámetros y haber definido el nombre del *data set* se confirma la creación haciendo click sobre el botón “*Add*”

Figura 68 Confirmación de la conexión



Fuente: elaboración propia con *Oracle*

Validación de la creación del set de datos “DIM_PUESTOS”

Figura 69 Validación creación set de datos

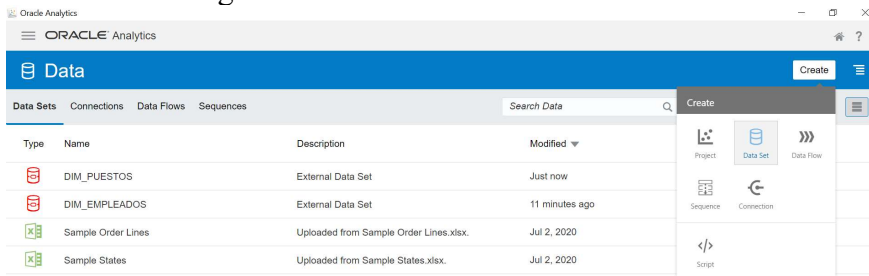
The screenshot shows the Oracle Analytics 'Data' page. It features a table with columns: Type, Name, Description, Modified, and Refreshed. The table lists several data sets, including DIM_PUESTOS, DIM_EMPLEADOS, Sample Order Lines, and Sample States.

Type	Name	Description	Modified	Refreshed
External Data Set	DIM_PUESTOS	External Data Set	Just now	Just now
External Data Set	DIM_EMPLEADOS	External Data Set	10 minutes ago	Just now
Sample Order Lines	Sample Order Lines	Uploaded from Sample Order Lines.xlsx	Jul 2, 2020	Jul 2, 2020
Sample States	Sample States	Uploaded from Sample States.xlsx	Jul 2, 2020	Jul 2, 2020

Fuente: elaboración propia con *Oracle*

4. 6.3.3 DIM_SUCURSALES

Figura 70 Creación set de datos. Paso 1

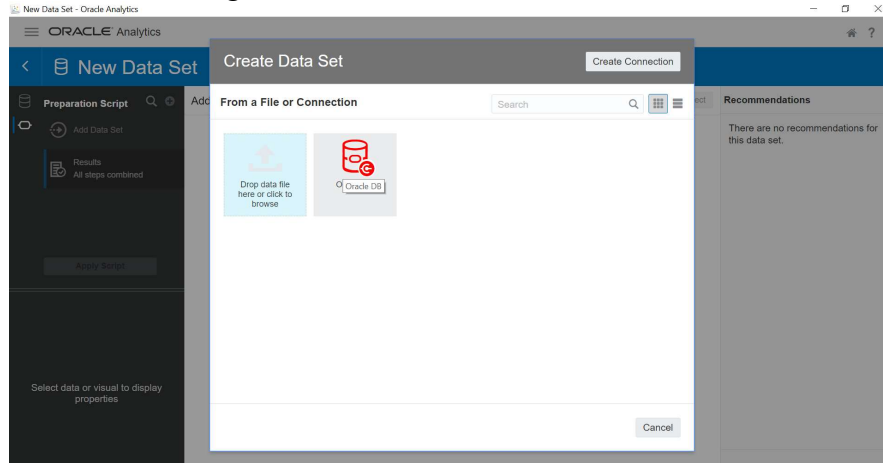


Fuente: elaboración propia con *Oracle*

Puesto que para esta tesis los sets de datos se alojan en una base de datos de tipo *Oracle* se debió crear previamente la conexión a la misma:

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

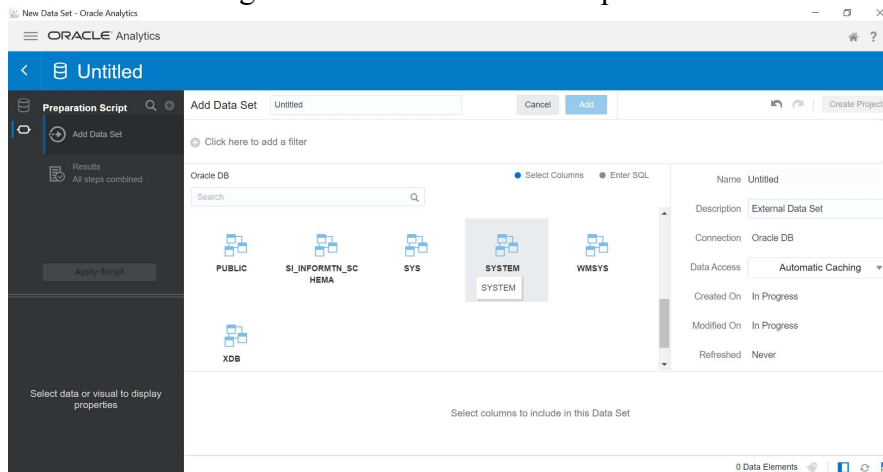
Figura 71 Creación set de datos. Paso 2



Fuente: elaboración propia con *Oracle*

Luego de definir la conexión se visualizan los esquemas que se encuentran en la base de datos, para el proyecto se utilizó el esquema “*SYSTEM*”:

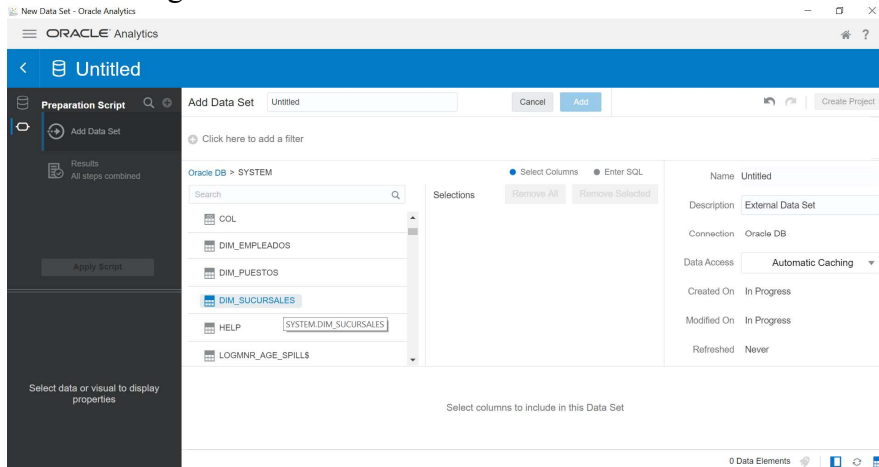
Figura 72 Visualización de esquemas



Fuente: elaboración propia con *Oracle*

Finalmente, al elegir el esquema se visualizan todas las tablas contenidas en él. Para la creación del *data set* DIM_SUCURSALES se selecciona la tabla DIM_SUCURSALES como fuente:

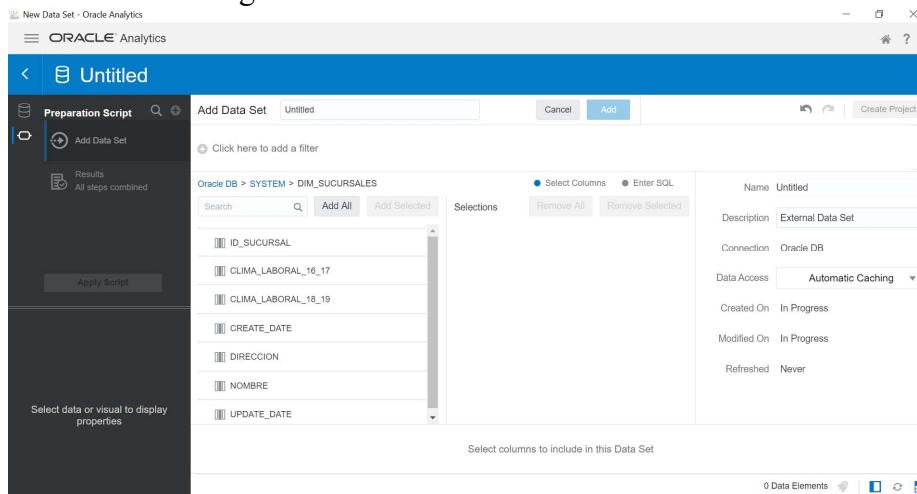
Figura 73 Visualización de tablas de contenidos



Fuente: elaboración propia con Oracle

Una vez seleccionada la tabla DIM_SUCURSALES se deben seleccionar los atributos que se desean incorporar en el *data set* para el dimensional:

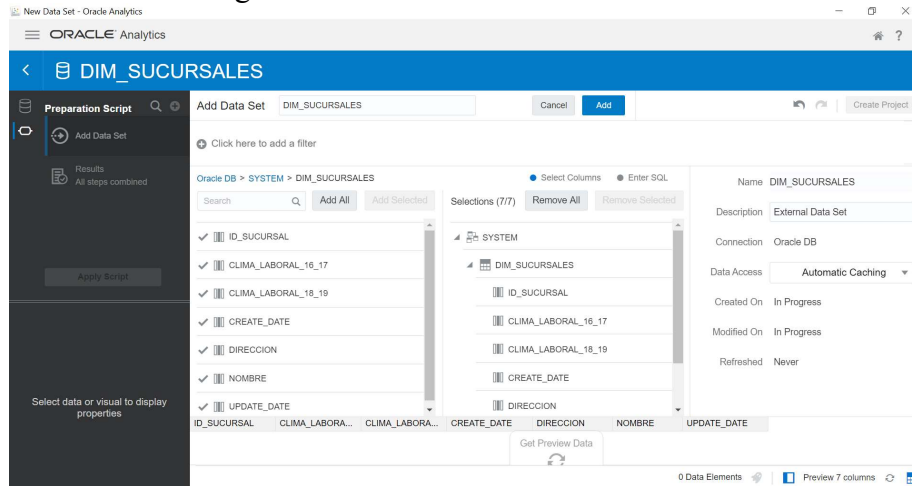
Figura 74 Selección de atributos. Paso 1



Fuente: elaboración propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

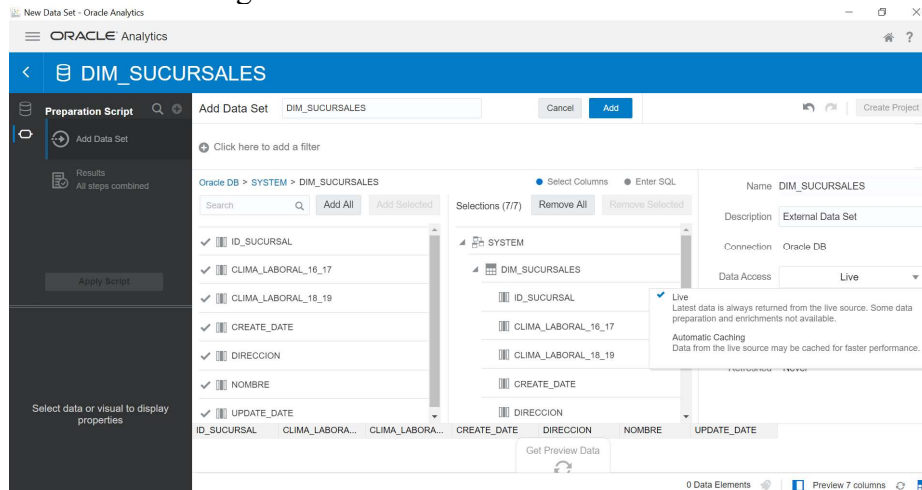
Figura 75 Selección de atributos. Paso 2



Fuente: elaboración propia con *Oracle*

Por default la conexión contra la base de datos es de tipo “*Automatic Caching*”, para el proyecto fue configurada como *Data Access “Live”*:

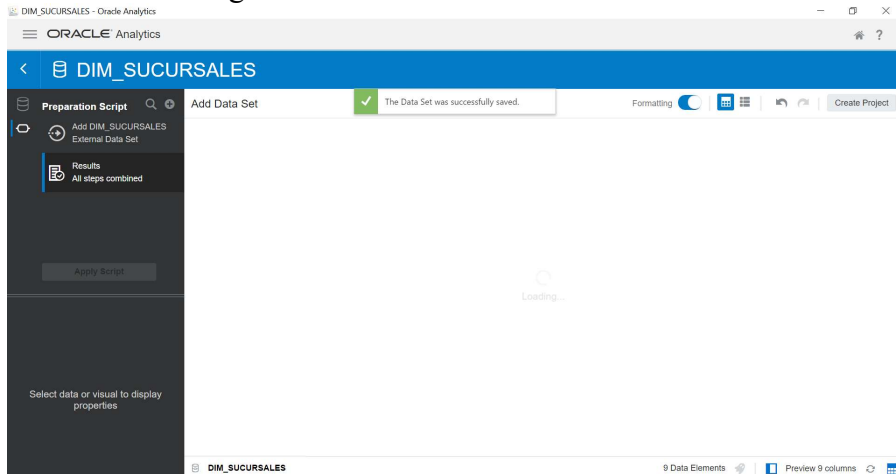
Figura 76 Conexión contra base de datos



Fuente: elaboración propia con *Oracle*

Luego de establecer todos los parámetros y haber definido el nombre del *data set* se confirma la creación haciendo click sobre el botón “*Add*”

Figura 77 Confirmación de la conexión



Fuente: Elaboración Propia con Oracle

Validación de creación de set de datos “DIM_SUCURSALES”

Figura 78 Validación creación set de datos

The screenshot shows the 'Data' section of Oracle Analytics. A table lists the following data sets:

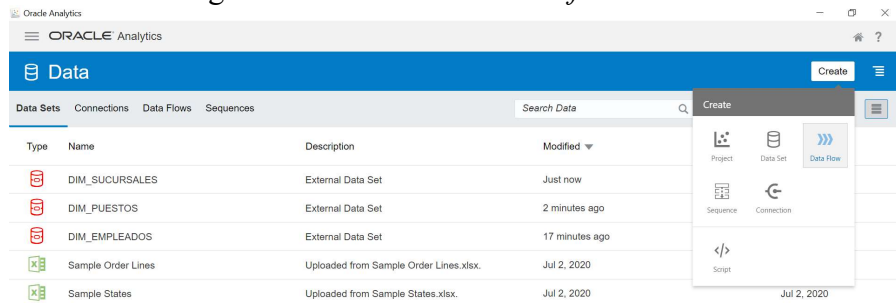
Type	Name	Description	Modified	Refreshed
External Data Set	DIM_SUCURSALES	External Data Set	Just now	Just now
External Data Set	DIM_PUESTOS	External Data Set	Just now	4 minutes ago
External Data Set	DIM_EMPLEADOS	External Data Set	15 minutes ago	3 minutes ago
Sample Order Lines	Sample Order Lines	Uploaded from Sample Order Lines.xlsx.	Jul 2, 2020	Jul 2, 2020
Sample States	Sample States	Uploaded from Sample States.xlsx.	Jul 2, 2020	Jul 2, 2020

Fuente: Elaboración Propia con Oracle

4.6.4 Generación de FACT_ANALYTICS_RRHH

Para la generación de la FACT_ANALYTICS_RRHH se creó un *data flow* siguiendo los pasos que se muestran a continuación. En este *data flow* se relacionaron las dimensiones previamente elaboradas a fin de centralizar y consolidar la información que se explotó mediante la reportería.

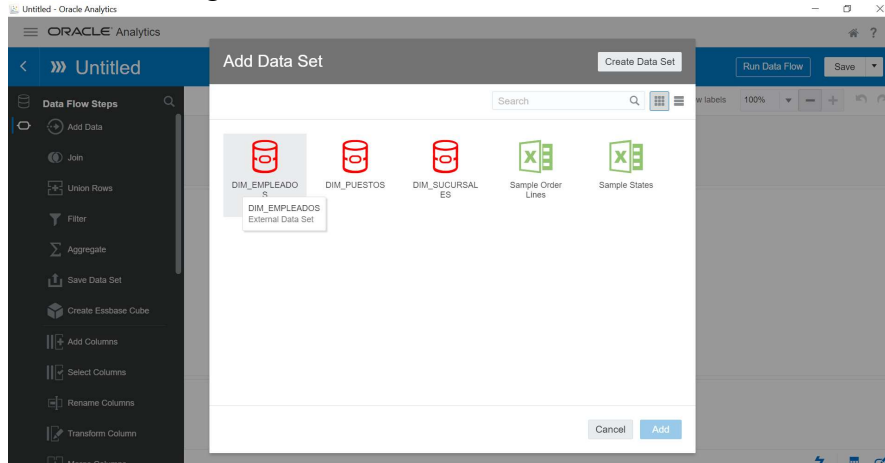
Figura 79 Generación de *data flow*. Paso 1



Fuente: Elaboración Propia con Oracle

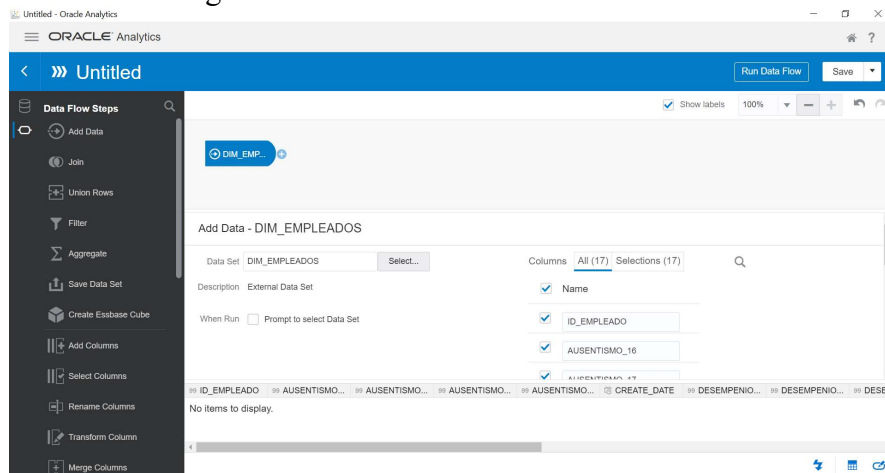
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 80 Generación de data flow. Paso 2



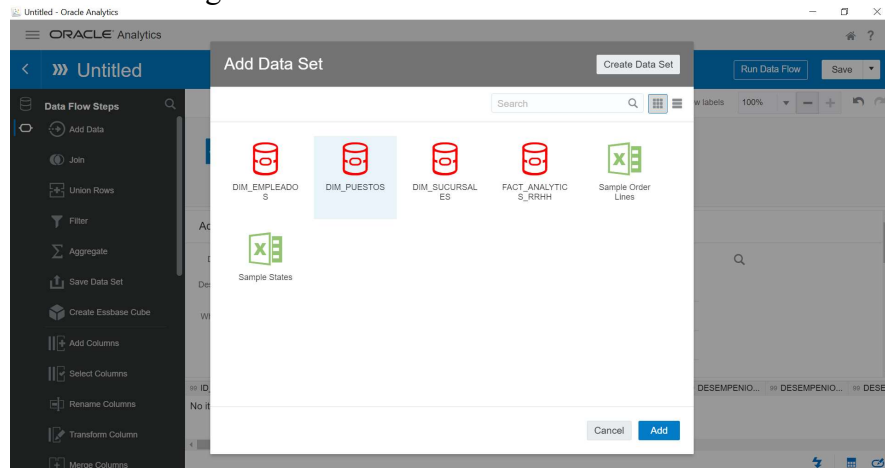
Fuente: Elaboración Propia con Oracle

Figura 81 Generación de data flow. Paso 3



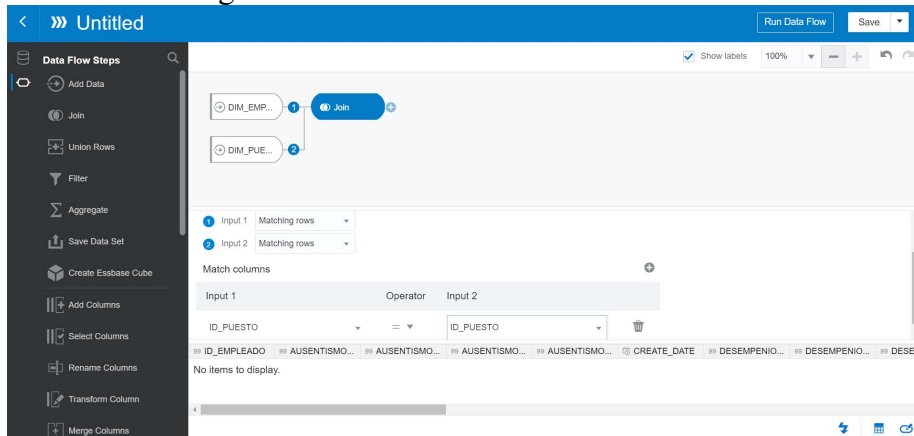
Fuente: Elaboración Propia con Oracle

Figura 82 Generación de data flow. Paso 4



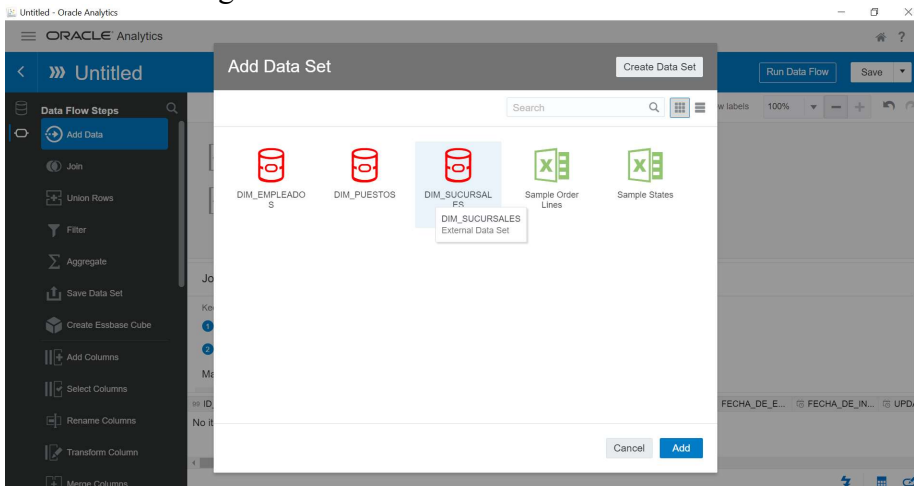
Fuente: Elaboración Propia con Oracle

Figura 83 Generación de data flow. Paso 5



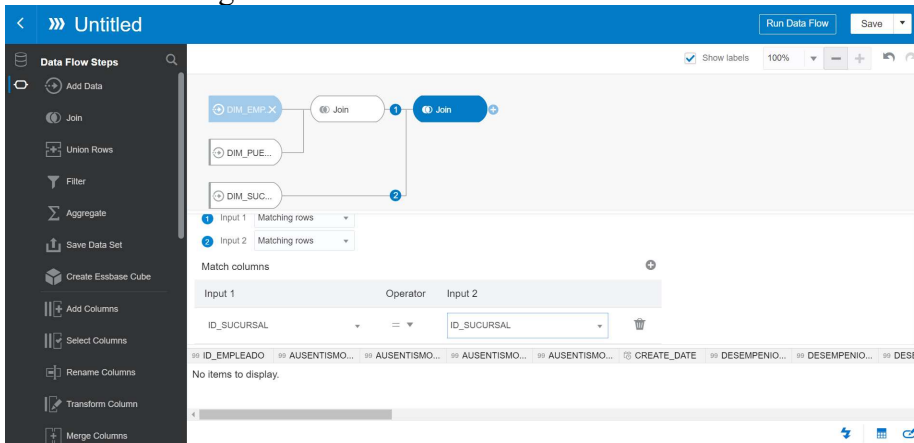
Fuente: Elaboración Propia con Oracle

Figura 84 Generación de data flow. Paso 6



Fuente: Elaboración Propia con Oracle

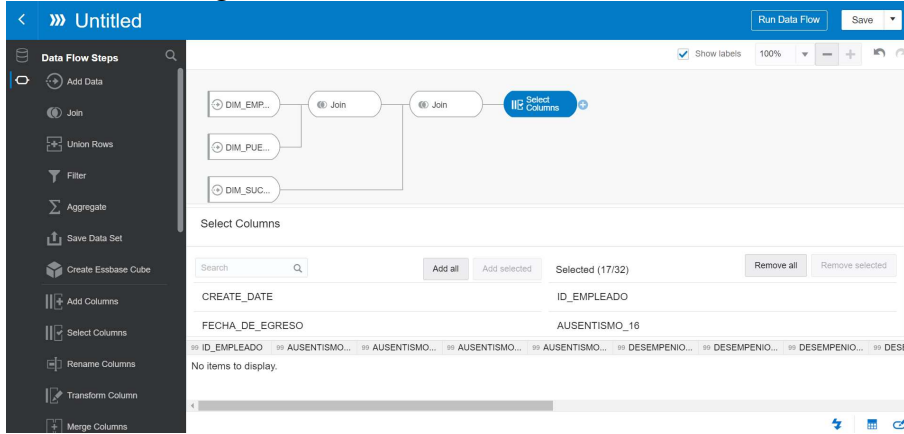
Figura 85 Generación de data flow. Paso 7



Fuente: Elaboración Propia con Oracle

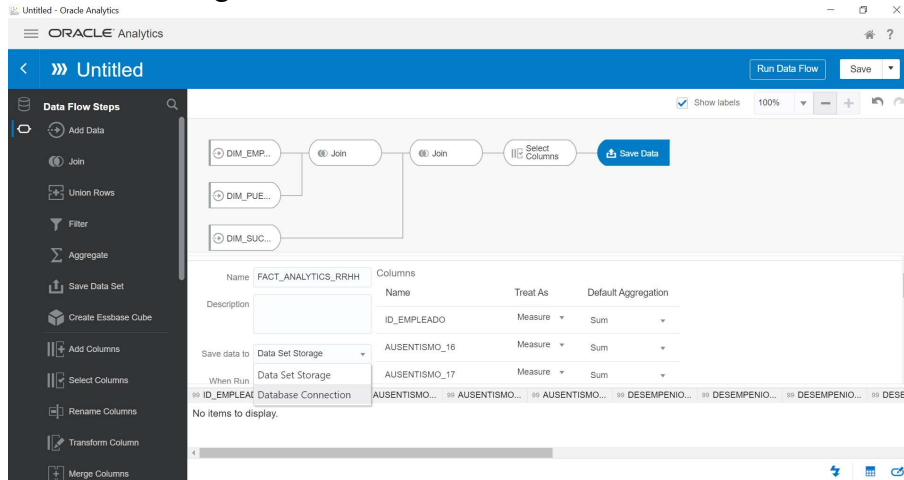
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 86 Generación de data flow. Paso 8



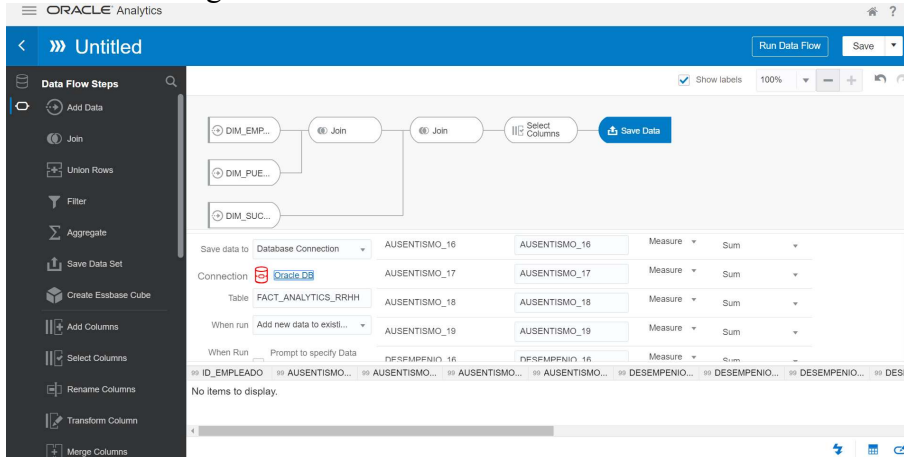
Fuente: Elaboración Propia con *Oracle*

Figura 87 Generación de data flow. Paso 9



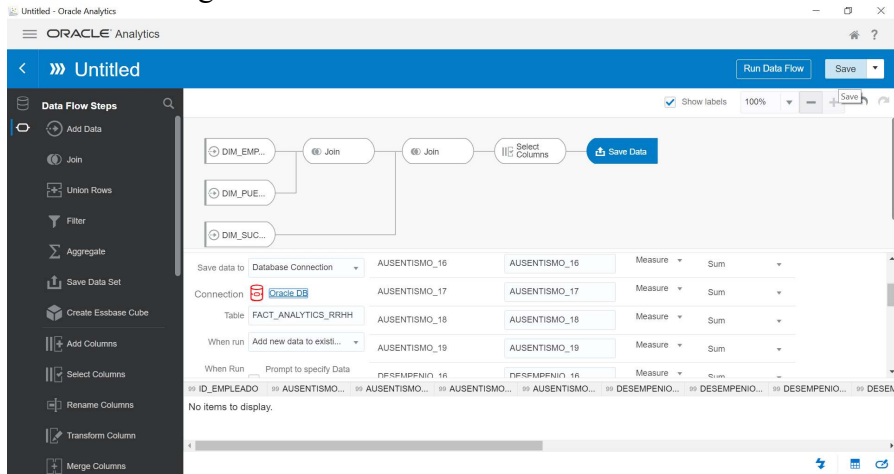
Fuente: Elaboración Propia con *Oracle*

Figura 88 Generación de data flow. Paso 10



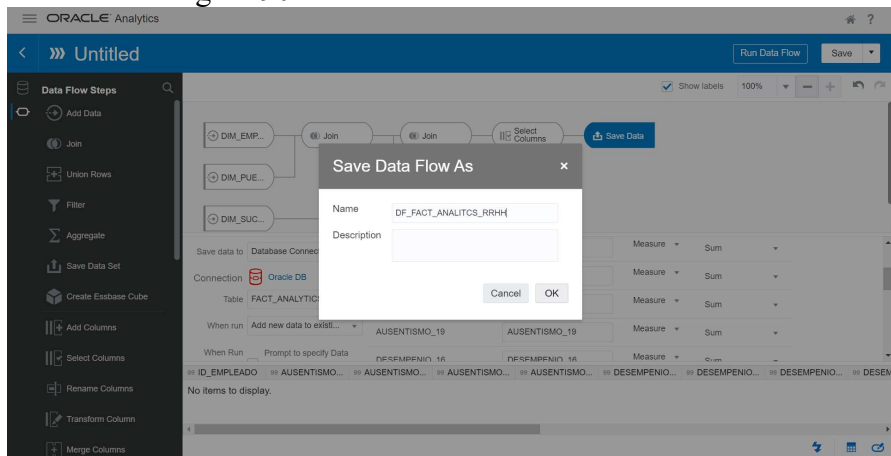
Fuente: Elaboración Propia con *Oracle*

Figura 89 Generación de data flow. Paso 11



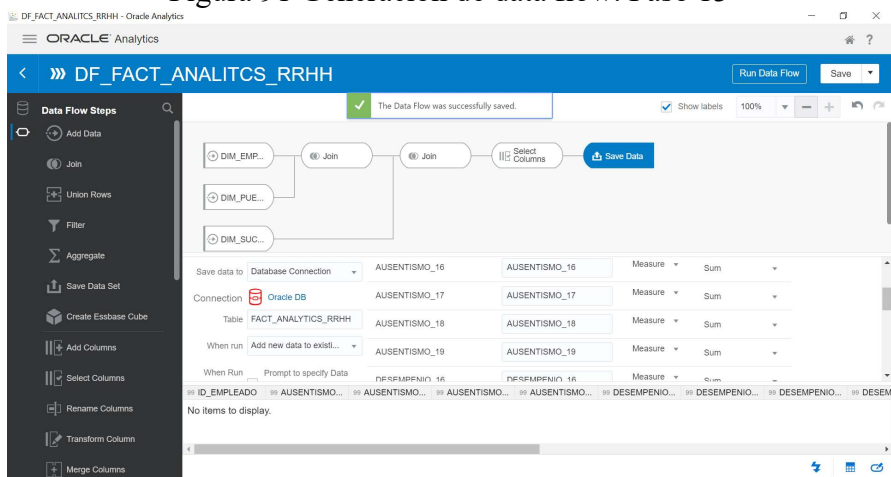
Fuente: Elaboración Propia con Oracle

Figura 90 Generación de data flow. Paso 12



Fuente: Elaboración Propia con Oracle

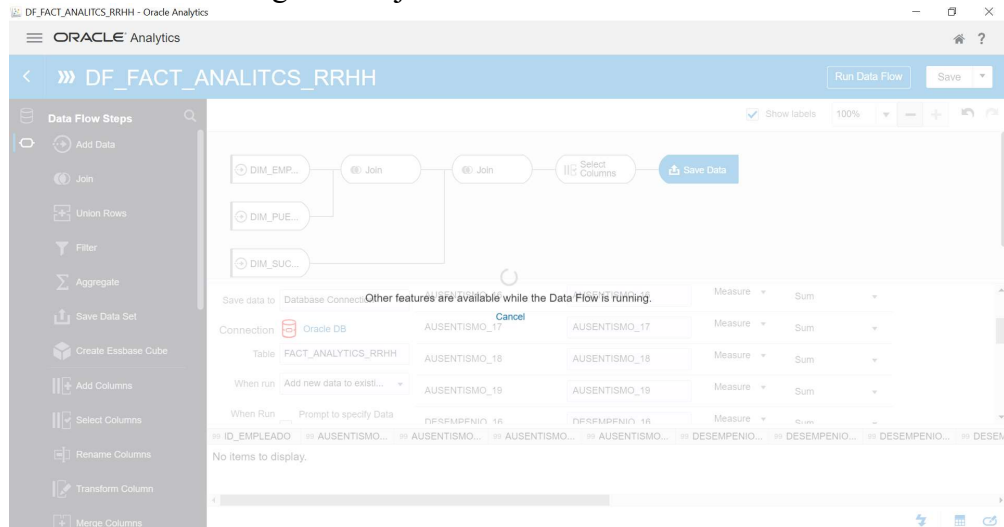
Figura 91 Generación de data flow. Paso 13



Fuente: Elaboración Propia con Oracle

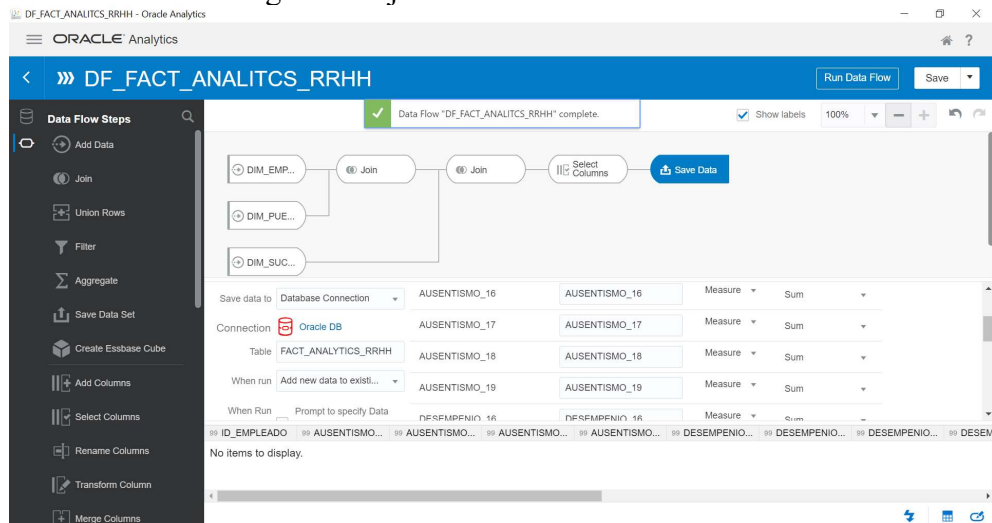
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
 Un caso de estudio”
 Ejecución del *data flow* para la creación de la tabla “FACT_ANALYTICS_RRHH” en la base de datos
 Oracle.

Figura 92 Ejecución de data flow. Paso 1



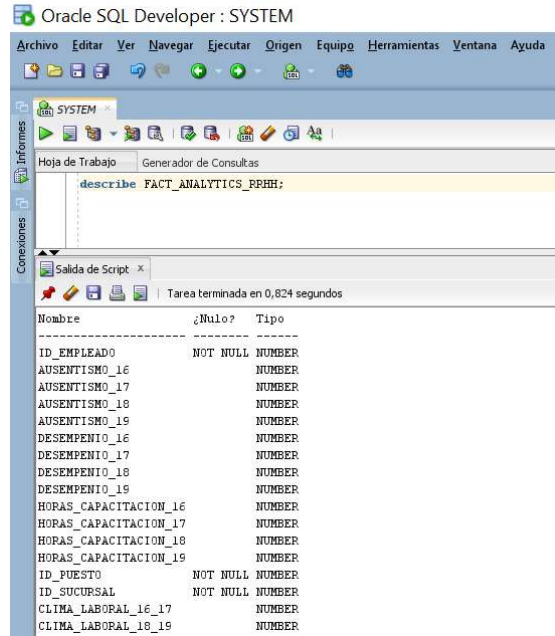
Fuente: Elaboración Propia con Oracle

Figura 93 Ejecución de data flow. Paso 2



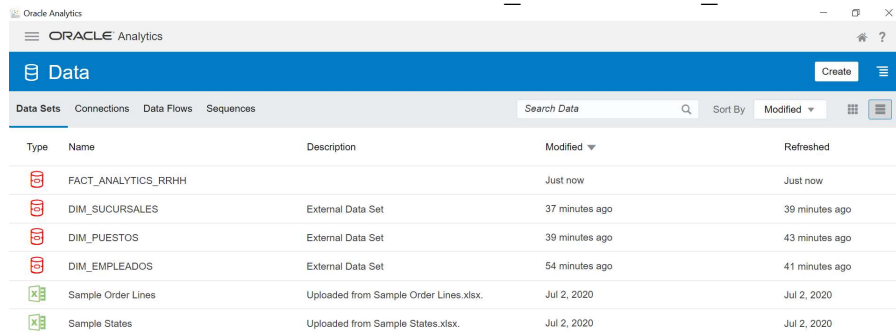
Fuente: Elaboración Propia con Oracle

Figura 94 Validación de la creación de la tabla en Oracle database



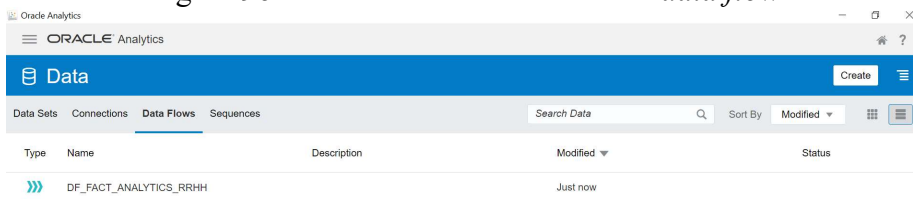
Fuente: Elaboración Propia con Oracle

Figura 95 Validación de la creación del data set “FACT_ANALYTICS_RRHH” en Oracle Analytics



Fuente: Elaboración Propia con Oracle

Figura 96 Validación de la creación del data flow



Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

4.6.5 IMPORTACIÓN DE DATOS EN LOS DIMENSIONALES

4.6.5.1 DIM_EMPLEADOS

Para importar los datos mediante el *tools* de *Oracle SQL Developer* se debió ingresar a la tabla creada previamente en el esquema y seleccionar la opción “Importar Datos”:

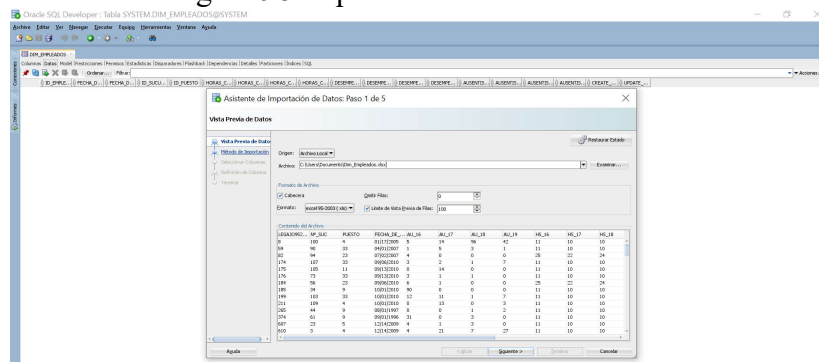
Figura 97 Importación de datos. Paso 1



Fuente: Elaboración Propia con *Oracle*

A continuación, se selecciona la ruta en donde se encuentra ubicado el archivo. *xlxs* utilizado para cargar la dimensión DIM_EMPLEADOS

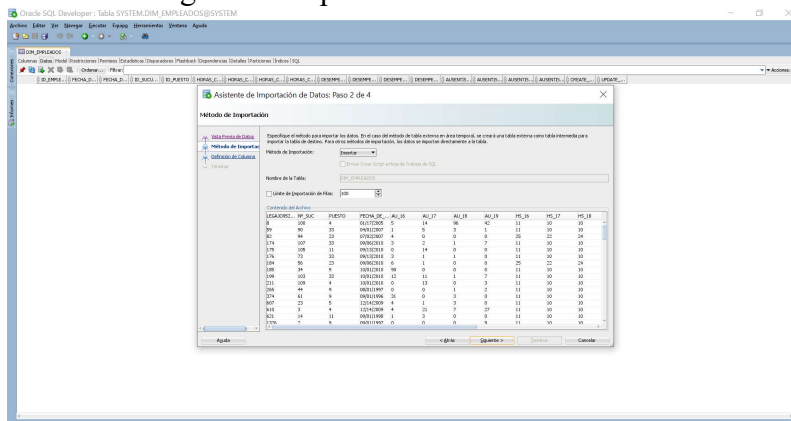
Figura 98 Importación de datos. Paso 2



Fuente: Elaboración Propia con *Oracle*

Se identifica que el método de importación utilizado para la carga de datos era “Insertar”

Figura 99 Importación de datos. Paso 3



Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Se valida que la cantidad de datos del origen (archivo .xlsx) fueran la misma cantidad que se inserta en el destino (DIM_EMPLEADOS) y que el mapeo se haya realizado de manera exitosa

Figura 103 Validación importación datos de origen

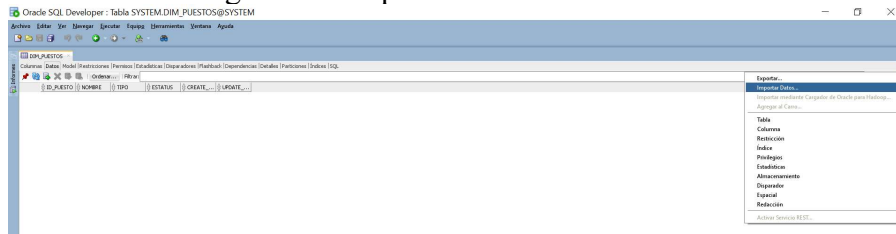
ID_EMPLEADO	NOMBRE	FECHA_HIRO	SALARIO	COMISION
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

Fuente: Elaboración Propia con Oracle

4.6.5.2 DIM_PUESTOS

Para importar los datos mediante el *tools* de Oracle SQL Developer se debe ingresar a la tabla creada previamente en el esquema y seleccionar la opción “Importar Datos”:

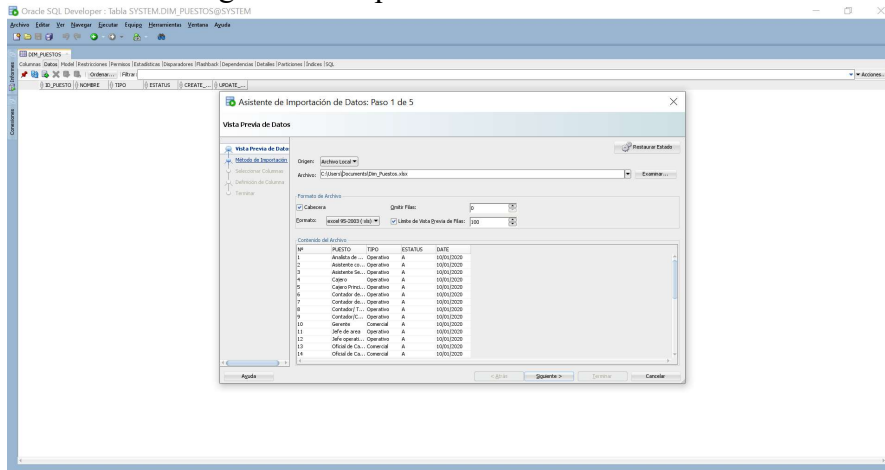
Figura 104 Importación de datos. Paso 1



Fuente: Elaboración Propia con Oracle

A continuación se selecciona la ruta en donde se encuentra ubicado el archivo .xlsx utilizado para cargar la dimensión DIM_PUESTOS

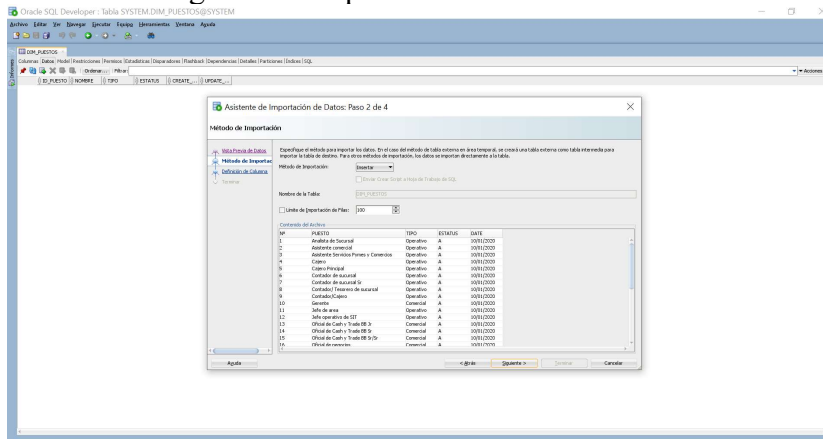
Figura 105 Importación de datos. Paso 2



Fuente: Elaboración Propia con Oracle

Se identifica que el método de importación utilizado para la carga de datos era “Insertar”

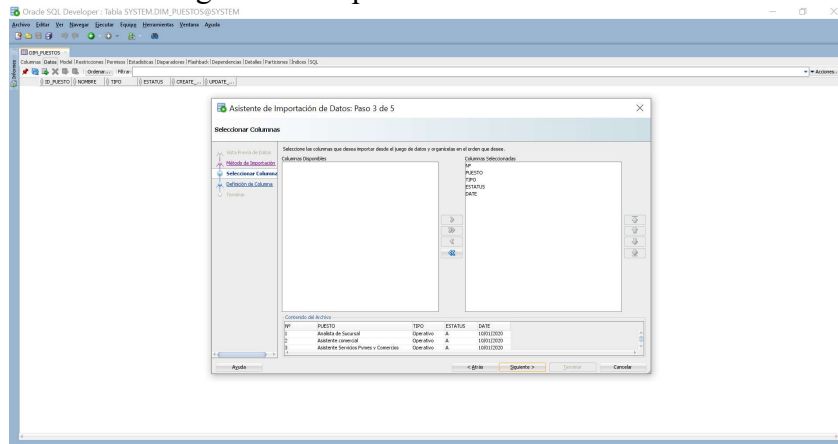
Figura 106 Importación de datos. Paso 3



Fuente: Elaboración Propia con Oracle

Se seleccionaron los atributos contenidos en el archivo utilizado para la carga de la tabla

Figura 107 Importación de datos. Paso 4

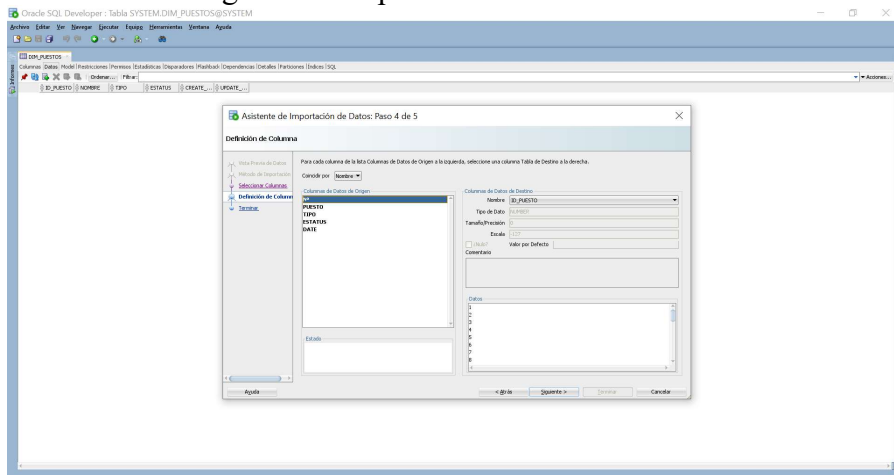


Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Se mapearon los atributos del archivo .xlsx contra los atributos de la DIM_PUESTOS

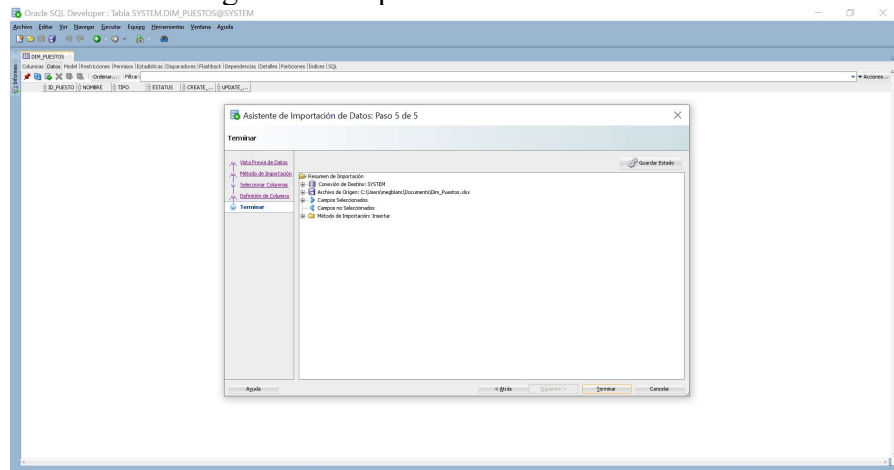
Figura 108 Importación de datos. Paso 5



Fuente: Elaboración Propia con *Oracle*

Se confirma la importación de los datos

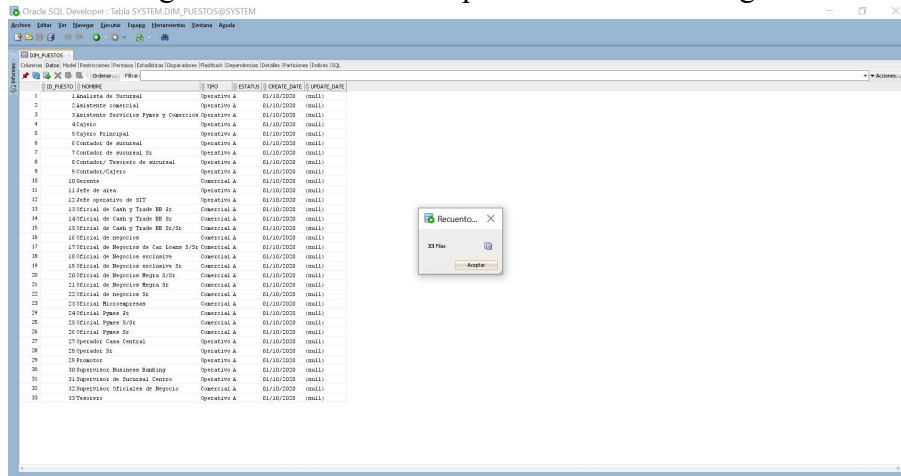
Figura 109 Importación de datos. Paso 6



Fuente: Elaboración Propia con *Oracle*

Se valida que la cantidad de datos del origen (archivo .xlsx) fueran la misma cantidad que se inserta en el destino (DIM_PUESTOS) y que el mapeo se haya realizado de manera exitosa

Figura 110 Validación importación datos de origen

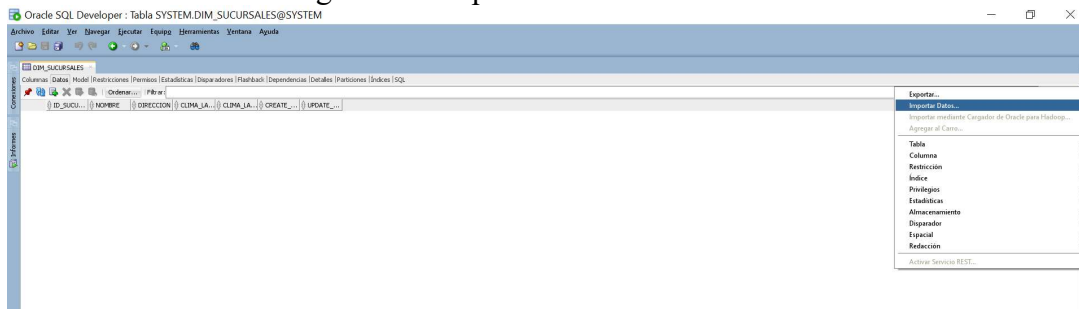


Fuente: Elaboración Propia con Oracle

4.6.5.3 DIM_SUCURSALES

Para importar los datos mediante el *tools* de Oracle SQL Developer se debió ingresar a la tabla creada previamente en el esquema y seleccionar la opción “Importar Datos”:

Figura 111 Importación de datos. Paso 1

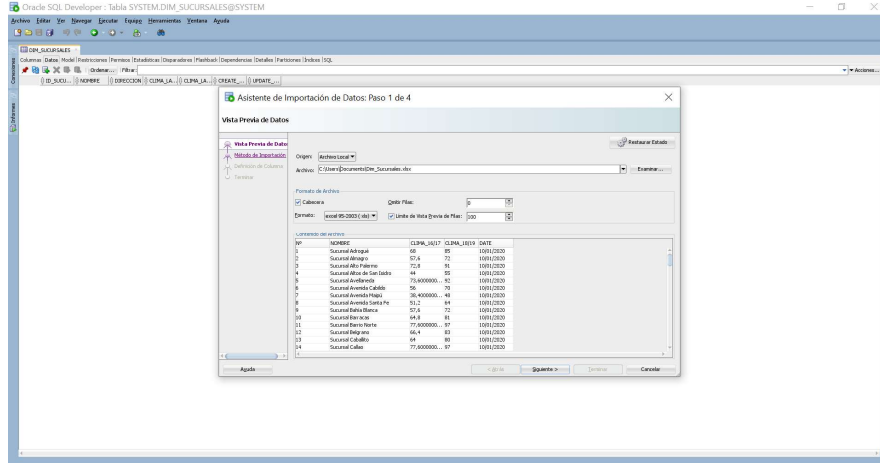


Fuente: Elaboración Propia con Oracle

A continuación se selecciona la ruta en donde se encuentra ubicado el archivo .xlsx utilizado para cargar la dimensión DIM_SUCURSALES

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

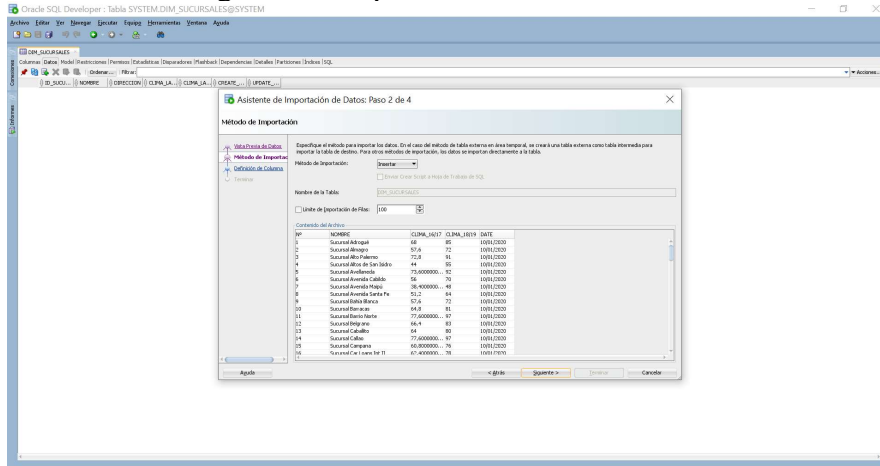
Figura 112 Importación de datos. Paso 2



Fuente: Elaboración Propia con *Oracle*

Se identifica que el método de importación utilizado para la carga de datos era “Insertar”

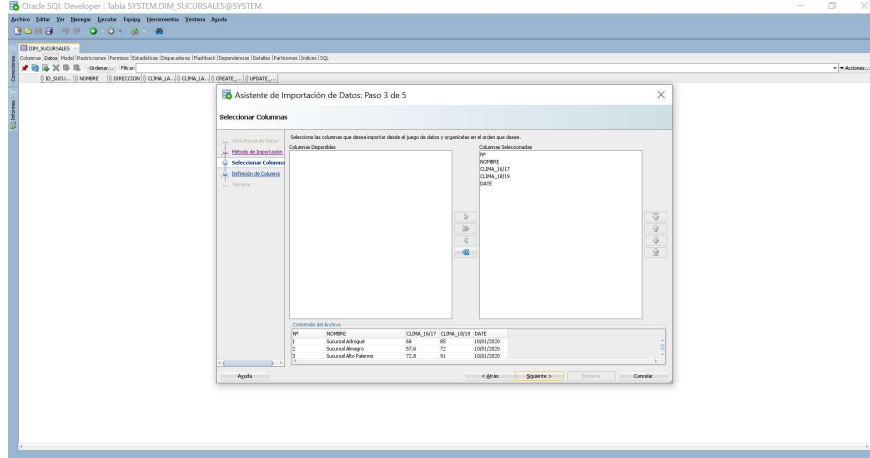
Figura 113 Importación de datos. Paso 3



Fuente: Elaboración Propia con *Oracle*

Se seleccionaron los atributos contenidos en el archivo utilizado para la carga de la tabla

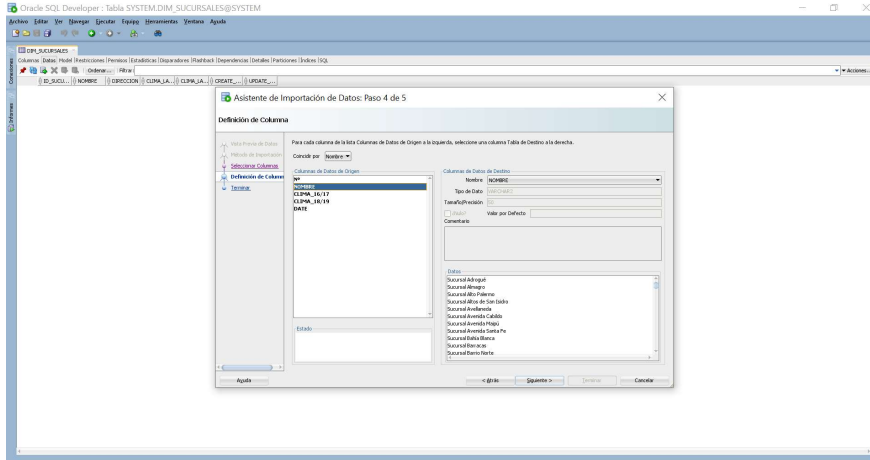
Figura 114 Importación de datos. Paso 4



Fuente: Elaboración Propia con *Oracle*

Se mapearon los atributos del archivo .xlsx contra los atributos de la DIM_SUCURSALES

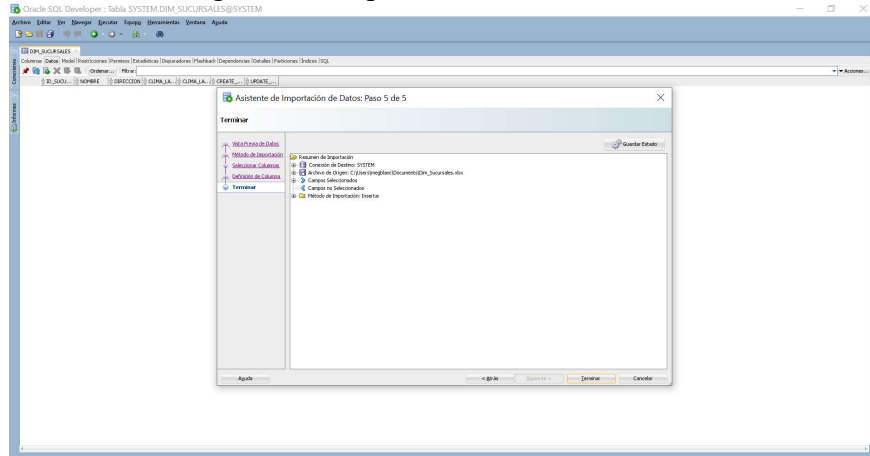
Figura 115 Importación de datos. Paso 5



Fuente: Elaboración Propia con *Oracle*

Se confirma la importación de los datos.

Figura 116 Importación de datos. Paso 6

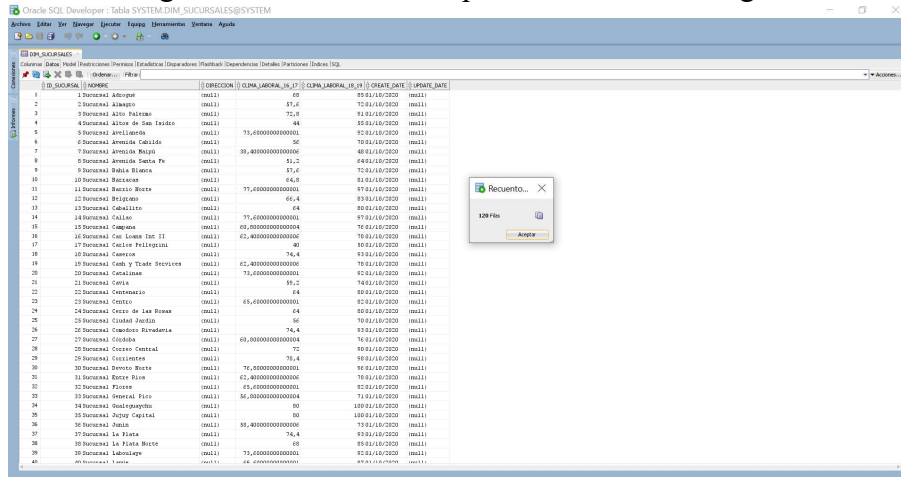


Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Se valida que la cantidad de datos del origen (archivo xlxs) fueran la misma cantidad que se inserta en el destino (DIM_SUCURSALES) y que el mapeo se haya realizado de manera exitosa.

Figura 117 Validación importación datos de origen



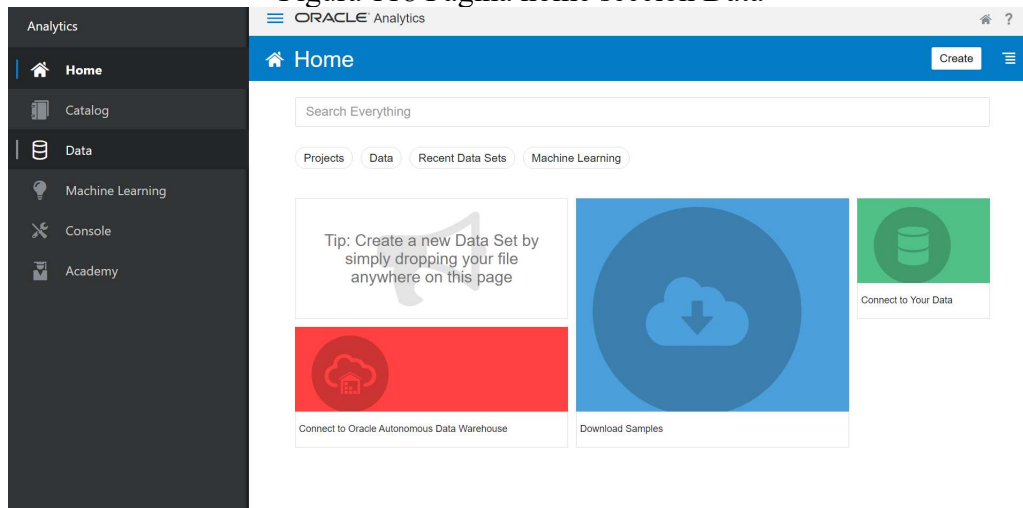
Fuente: Elaboración Propia con Oracle

4.6.5.4 FACT_ANALYTICS_RRHH

Para la carga de la *fact* se ejecutó el *data flow* de la herramienta *Oracle Data Visualization* siguiendo los pasos que se muestran a continuación:

Se ingresó desde la página de home hasta la sección *Data*.

Figura 118 Página home sección Data



Fuente: Elaboración Propia con Oracle

Posteriormente se ingresó a la sección de “*Data Flows*”.

Figura 119 Elaboración *data flow*. Paso 1

Type	Name	Description	Modified	Refreshed
	FACT_ANALYTICS_RRHH		Oct 6, 2020	Oct 6, 2020
	DIM_EMPLEADOS	External Data Set	Oct 6, 2020	Oct 6, 2020
	DIM_SUCURSALES	External Data Set	Sep 24, 2020	Sep 24, 2020
	DIM_PUESTOS	External Data Set	Sep 13, 2020	Sep 13, 2020
	Sample Order Lines	Uploaded from Sample Order Lines.xlsx.	Jul 2, 2020	Jul 2, 2020
	Sample States	Uploaded from Sample States.xlsx.	Jul 2, 2020	Jul 2, 2020

Fuente: Elaboración Propia con Oracle

Figura 120 Elaboración *data flow*. Paso 2

Type	Name	Description	Modified	Status
	DF_FACT_ANALYTICS_RRHH		Oct 6, 2020	

- Run
- Open
- Open in a New Window
- Inspect
- Export
- Delete

Fuente: Elaboración Propia con Oracle

Se seleccionó desde el *data flow* creado para la carga de la FACT_ANALYTICS_RRHH la opción “Run” y se visualizó el estatus de ejecución del mismo.

Figura 121 Ejecución de *data flow*

Type	Name	Description	Modified	Status
	DF_FACT_ANALYTICS_RRHH		Oct 6, 2020	Running

Type	Name	Description	Modified	Status
	DF_FACT_ANALYTICS_RRHH		Oct 6, 2020	Completed

Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”
 Posterior a la confirmación de la ejecución del flujo de datos se validó mediante el gestor de base de datos de Oracle la carga en la Fact.

Figura 122 Creación del proyecto

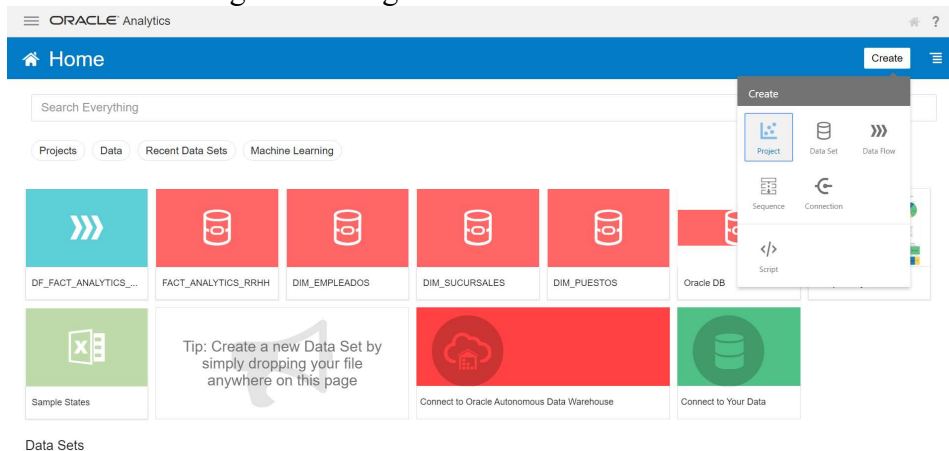
ID_EMPLEADO	AUSENTO_16	AUSENTO_17	AUSENTO_18	AUSENTO_19	DESEMPEÑO_16	DESEMPEÑO_17	DESEMPEÑO_18	DESEMPEÑO_19	HORAS_CAPACITACION_16	HORAS_CAPACITACION_17	HORAS_CAPACITACION_18	HORAS_CAPACITACION_19	ID_PUESTO	ID_SUCURSAL	CLIMA_LABORAL_16_17
1	0	5	14	56	42	3	4	(null)	3	11	10	10	11	4	100,73,60000000000001
2	59	1	5	3	1	3	4	(null)	3	11	10	10	11	33	90,69,9
3	82	4	0	0	0	3	3	3	3	25	22	24	27	23	94,75
4	174	3	2	1	7	3	3	3	3	11	10	10	11	33	107,64,8
5	175	0	14	0	0	3	3	3	3	11	10	10	11	11	105,77,60000000000001
6	176	3	1	1	0	4	3	3	3	11	10	10	11	33	73,80
7	184	6	1	0	0	3	3	3	3	25	22	24	27	23	86,69,60000000000001
8	185	90	0	0	0	3	4	4	3	11	10	10	11	5	34,80
9	195	12	11	1	7	3	3	3	3	11	10	10	11	33	103,60
10	211	0	13	0	3	3	3	3	3	11	10	10	11	4	109,75,2
11	265	0	0	1	2	(null)	(null)	4	4	11	10	10	11	9	44,40,00000000000004
12	374	11	0	3	0	(null)	4	4	3	11	10	10	11	9	61,65,60000000000001
13	697	4	1	3	0	(null)	3	3	3	11	10	10	11	5	23,65,60000000000001
14	610	4	21	7	27	3	3	3	3	11	10	10	11	4	3,72,9
15	421	1	3	0	0	3	3	3	3	11	10	10	11	11	19,77,60000000000001
16	1126	0	0	0	6	3	3	3	3	11	10	10	11	9	2,97,6
17	1505	9	0	0	0	3	3	3	3	25	22	24	27	24	13,64
18	1524	2	33	0	1	(null)	4	4	(null)	11	10	10	11	9	55,80
19	1934	216	1	0	3	3	3	3	3	11	10	10	11	11	57,72,8
20	1536	85	07	17	6	3	3	3	3	11	10	10	11	4	40,69,60000000000001
21	1937	15	0	0	1	3	4	3	3	25	22	24	27	16	24,64
22	1535	7	2	1	5	(null)	(null)	2	2	11	10	10	11	4	12,66,4
23	1542	6	4	1	0	3	3	3	3	11	10	10	11	11	21,59,2
24	1544	5	2	0	0	3	3	2	2	25	22	24	27	16	69,67,2
25	1551	0	0	0	7	3	3	3	3	25	22	24	27	16	36,54,4000000000000006
26	1558	90	0	128	0	3	3	3	3	25	22	24	27	16	78,67,2
27	1561	1	2	1	1	(null)	3	2	2	25	22	24	27	18	86,68,8
28	1562	26	72	109	113	(null)	3	2	2	25	22	24	27	16	67,67,2
29	1563	184	5	4	2	(null)	3	2	(null)	25	22	24	27	16	80,80
30	1566	1	0	2	1	(null)	4	3	3	25	22	24	27	10	110,73,60000000000001
31	1577	1	139	16	5	(null)	4	3	3	25	22	24	27	16	110,70,4
32	1581	5	3	4	0	3	3	3	2	11	10	10	11	9	26,74,4
33	1586	0	4	3	12	(null)	3	3	3	11	10	10	11	4	57,72,8
34	1598	17	3	6	1	3	3	3	3	11	10	10	11	33	10,64,8
35	1592	1	7	1	36	3	3	3	3	11	10	10	11	27	23,65,60000000000001
36	1598	101	1	6	3	2	3	3	3	25	22	24	27	18	108,74,4
37	1603	1	4	1	5	3	3	3	2	11	10	10	11	4	109,75,2
38	1606	0	21	0	4	3	3	3	2	25	22	24	27	10	15,68,00000000000004

Fuente: Elaboración Propia con Oracle

Una vez culminada la fase de modelado de datos se procedió a la creación del proyecto para la explotación y análisis de los mismo.

4.6.5.5 DESARROLLO DEL PROYECTO

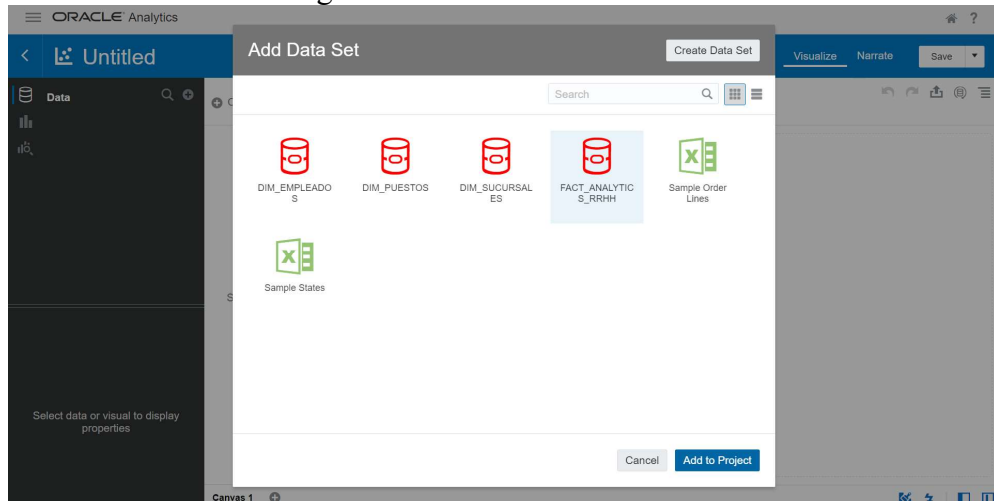
Figura 123 Página home Data Visualization



Fuente: Elaboración Propia con Oracle

Desde la página de home se seleccionó la opción “*Create - Project*”.

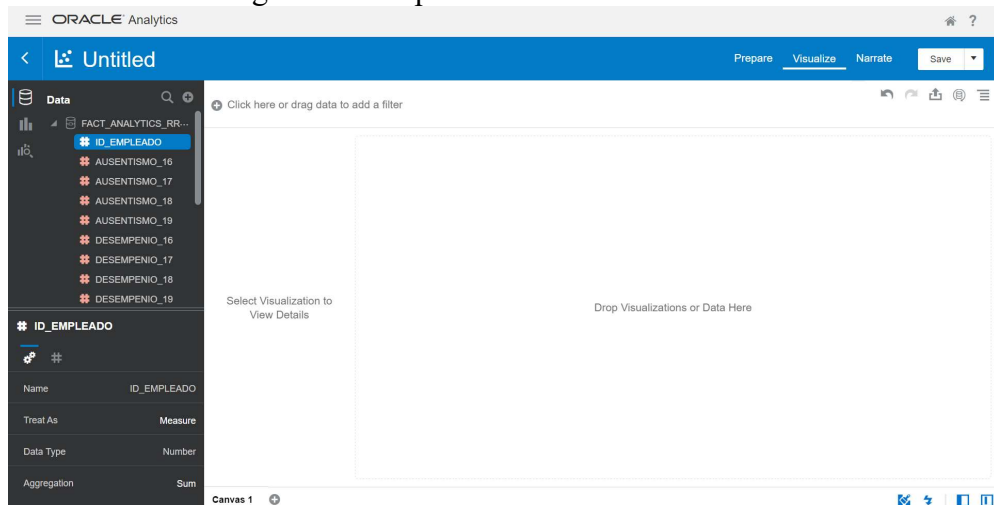
Figura 124 Selección del *data set*



Fuente: Elaboración Propia con *Oracle*

Para la creación se debió seleccionar previamente uno de los sets de datos a utilizar para la creación de las visualizaciones.

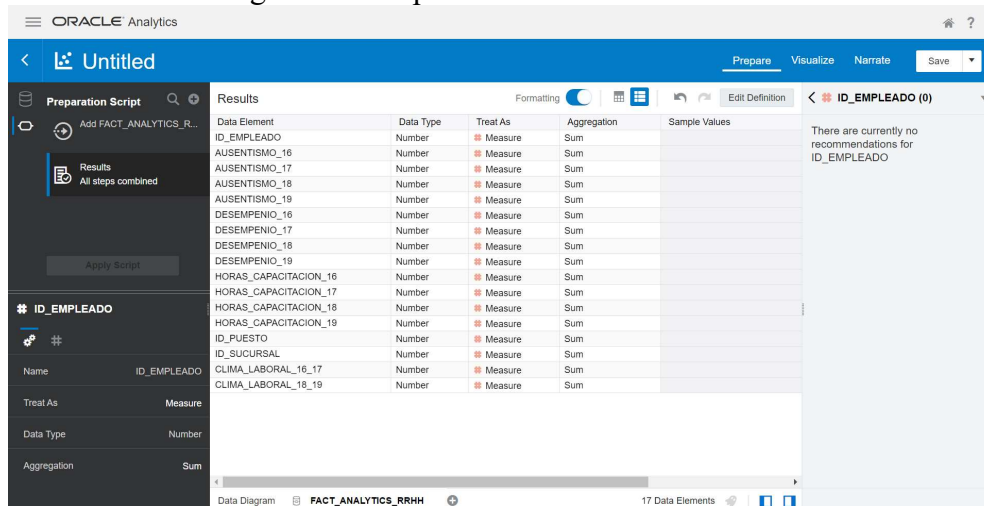
Figura 125 Preparación del *data set*. Paso 1



Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

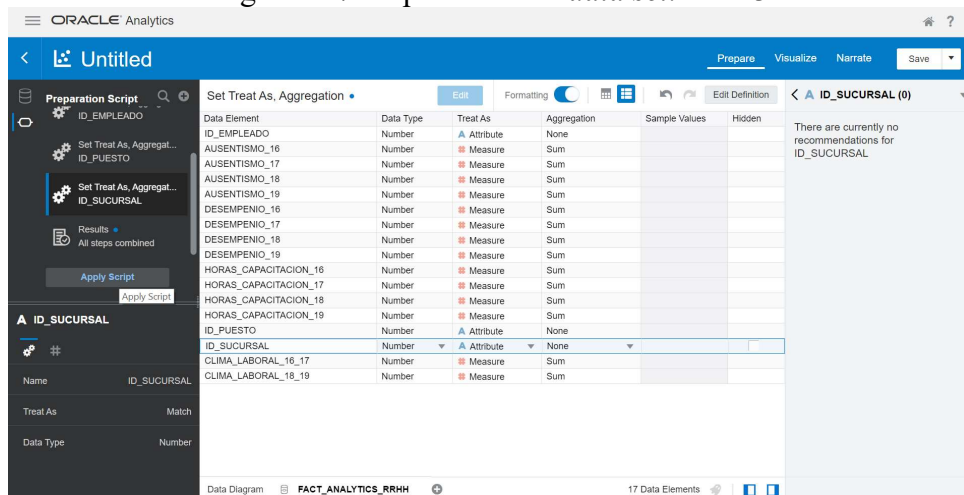
Figura 126 Preparación del *data set*. Paso 2



Fuente: Elaboración Propia con Oracle

Luego de seleccionar el set de datos principal “FACT_ANALYTICS_RRHH” se ingresó a la solapa del *Analytics* “Prepare” para realizar la adecuación de los datos previas a la generación de las visualizaciones.

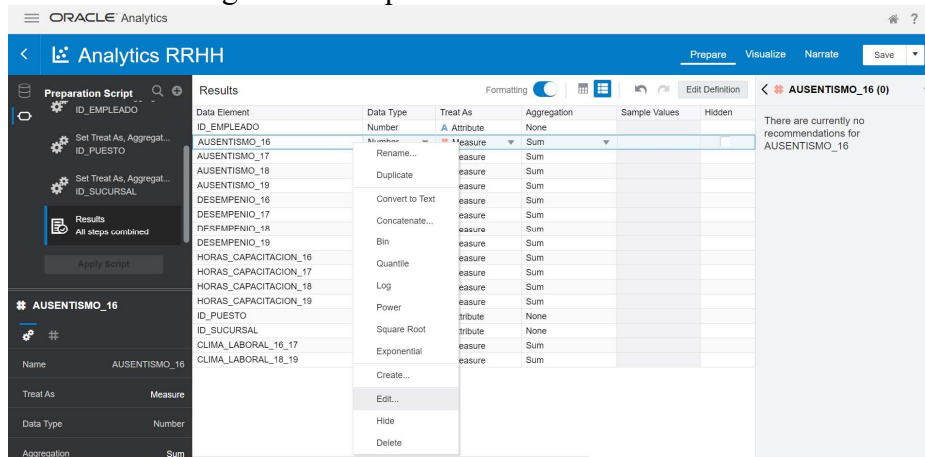
Figura 127 Preparación del *data set*. Paso 3



Fuente: Elaboración Propia con Oracle

Se definió el tratamiento para los datos acorde a lo definido según el modelo indicando “Attribute o Measure”.

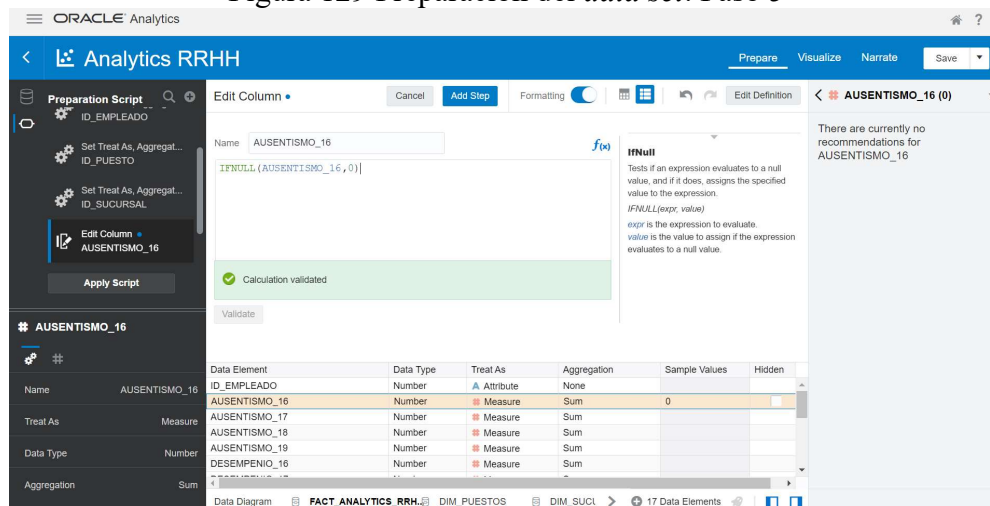
Figura 128 Preparación del *data set*. Paso 4



Fuente: Elaboración Propia con Oracle

Así mismo fueron adecuadas todas las medidas de la *fact* con la función *ifnull* para estandarizar los datos visualizados en caso de que tuvieran un valor *null*.

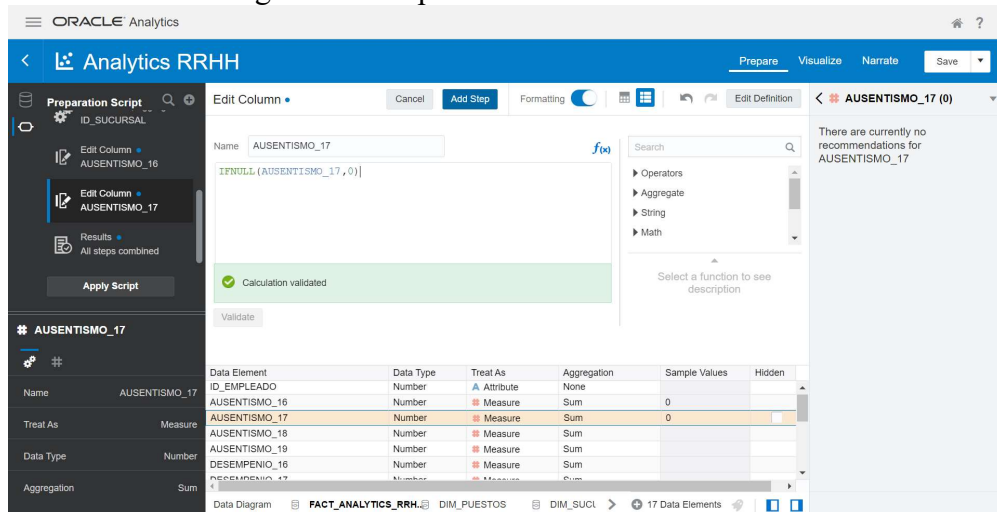
Figura 129 Preparación del *data set*. Paso 5



Fuente: Elaboración Propia con Oracle

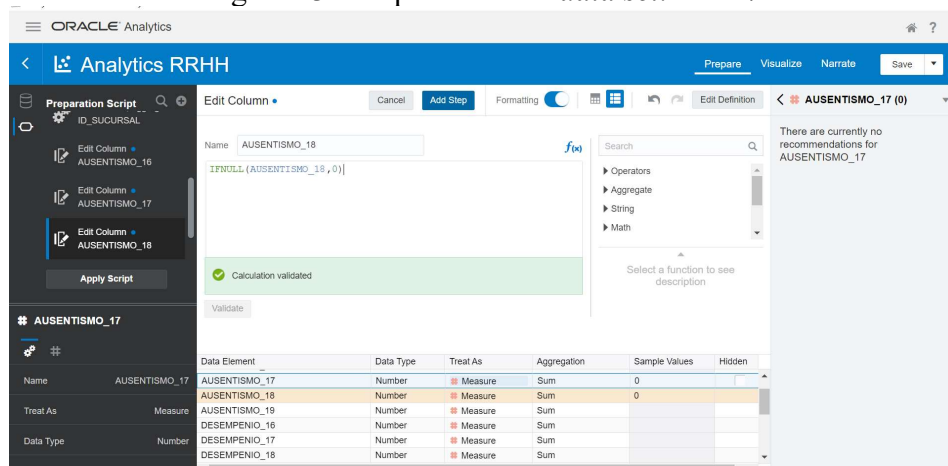
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 130 Preparación del *data set*. Paso 6



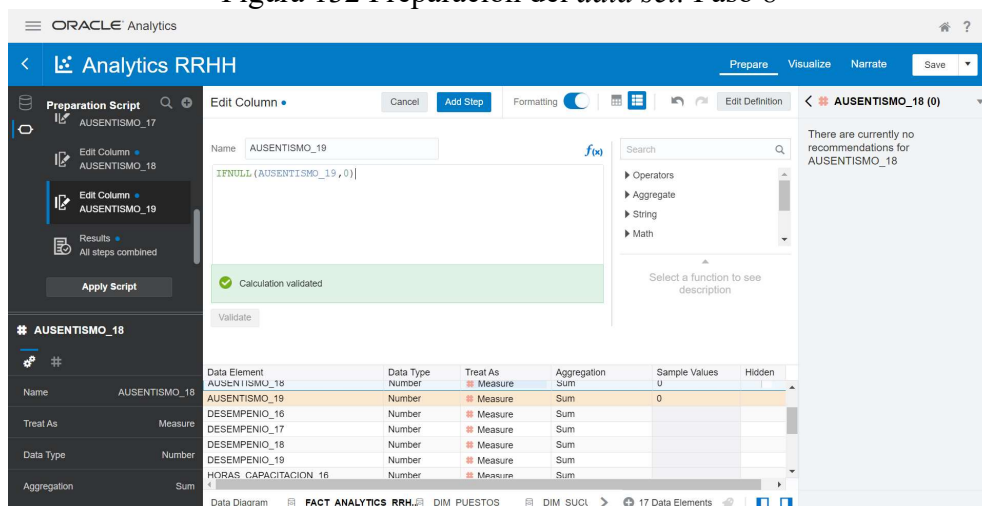
Fuente: Elaboración Propia con Oracle

Figura 131 Preparación del *data set*. Paso 7



Fuente: Elaboración Propia con Oracle

Figura 132 Preparación del *data set*. Paso 8



Fuente: Elaboración Propia con Oracle

Figura 133 Preparación del *data set*. Paso 9

The screenshot shows the Oracle Analytics RRHH interface during the preparation of a data set. The main window is titled 'Analytics RRHH' and has tabs for 'Prepare', 'Visualize', and 'Narrate'. A 'Preparation Script' sidebar on the left shows a list of columns: AUSENTISMO_17, AUSENTISMO_18, AUSENTISMO_19, and DESEMPENIO_16. The 'Edit Column' dialog is open for 'DESEMPENIO_16', showing the formula `IFNULL(DESEMPENIO_16, 0)`. A green checkmark indicates 'Calculation validated'. Below the dialog is a table of data elements:

Data Element	Data Type	Treat As	Aggregation	Sample Values	Hidden
DESEMPENIO_16	Number	Measure	Sum	0	
DESEMPENIO_17	Number	Measure	Sum		
DESEMPENIO_18	Number	Measure	Sum		
DESEMPENIO_19	Number	Measure	Sum		
HORAS_CAPACITACION_16	Number	Measure	Sum		
HORAS_CAPACITACION_17	Number	Measure	Sum		

On the right, there is a 'Function Library' with categories like Operators, Aggregate, String, and Math. A message on the far right states: 'There are currently no recommendations for AUSENTISMO_19'.

Fuente: Elaboración Propia con Oracle

Figura 134 Preparación del *data set*. Paso 10

The screenshot shows the Oracle Analytics RRHH interface during the preparation of a data set. The main window is titled 'Analytics RRHH' and has tabs for 'Prepare', 'Visualize', and 'Narrate'. A 'Preparation Script' sidebar on the left shows a list of columns: AUSENTISMO_19, DESEMPENIO_16, and DESEMPENIO_17. The 'Edit Column' dialog is open for 'DESEMPENIO_17', showing the formula `IFNULL(DESEMPENIO_17, 0)`. A green checkmark indicates 'Calculation validated'. Below the dialog is a table of data elements:

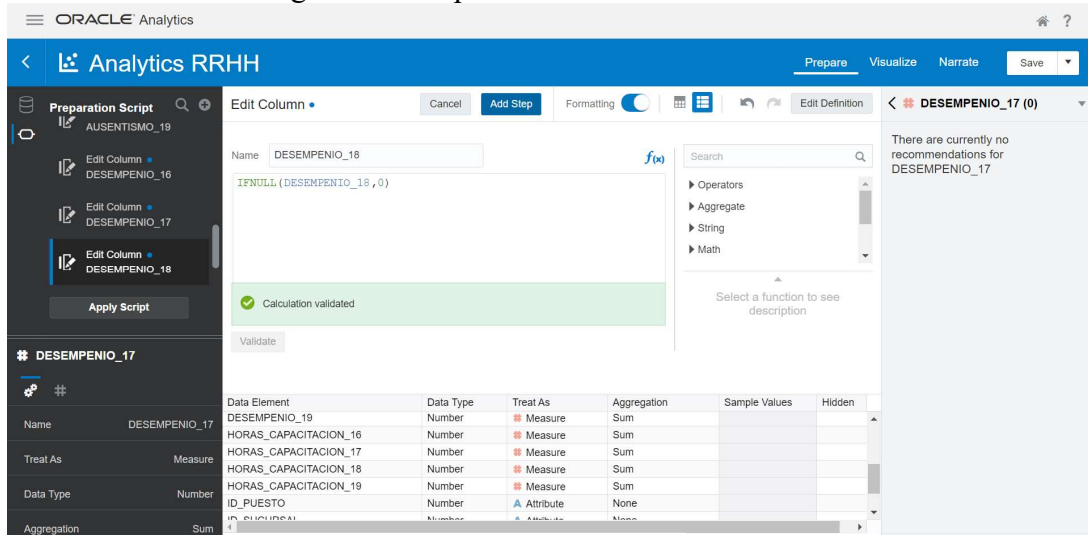
Data Element	Data Type	Treat As	Aggregation	Sample Values	Hidden
DESEMPENIO_16	Number	Measure	Sum		
DESEMPENIO_18	Number	Measure	Sum		
DESEMPENIO_19	Number	Measure	Sum		
HORAS_CAPACITACION_16	Number	Measure	Sum		
HORAS_CAPACITACION_17	Number	Measure	Sum		
HORAS_CAPACITACION_18	Number	Measure	Sum		
HORAS_CAPACITACION_19	Number	Measure	Sum		

On the right, there is a 'Function Library' with categories like Operators, Aggregate, String, and Math. A message on the far right states: 'There are currently no recommendations for DESEMPENIO_16'.

Fuente: Elaboración Propia con Oracle

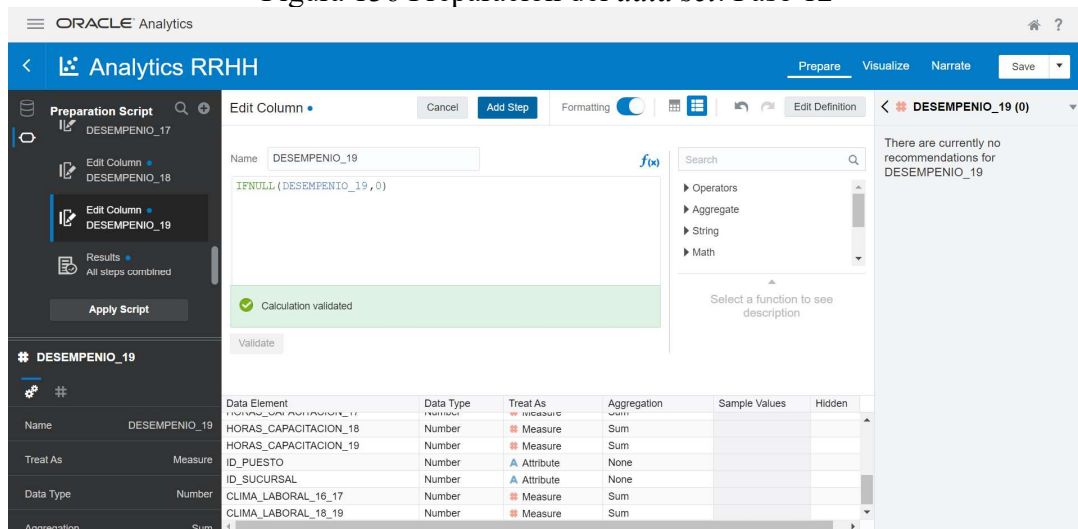
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 135 Preparación del *data set*. Paso 11



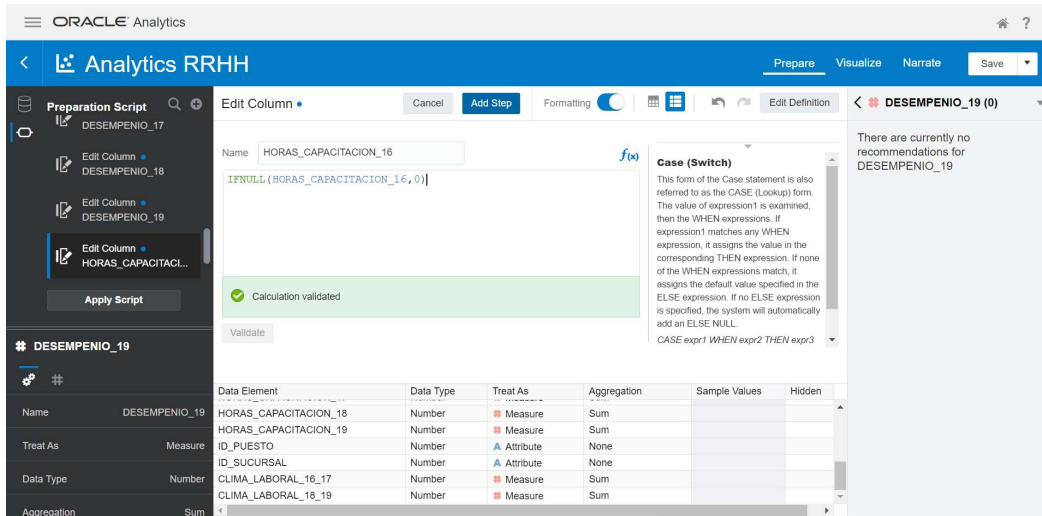
Fuente: Elaboración Propia con Oracle

Figura 136 Preparación del *data set*. Paso 12



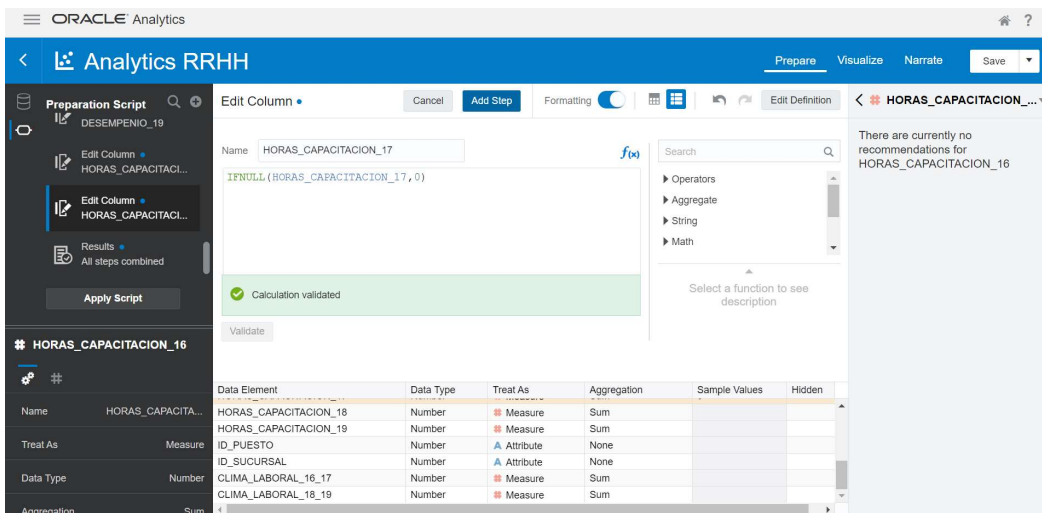
Fuente: Elaboración Propia con Oracle

Figura 137 Preparación del *data set*. Paso 13



Fuente: Elaboración Propia con Oracle

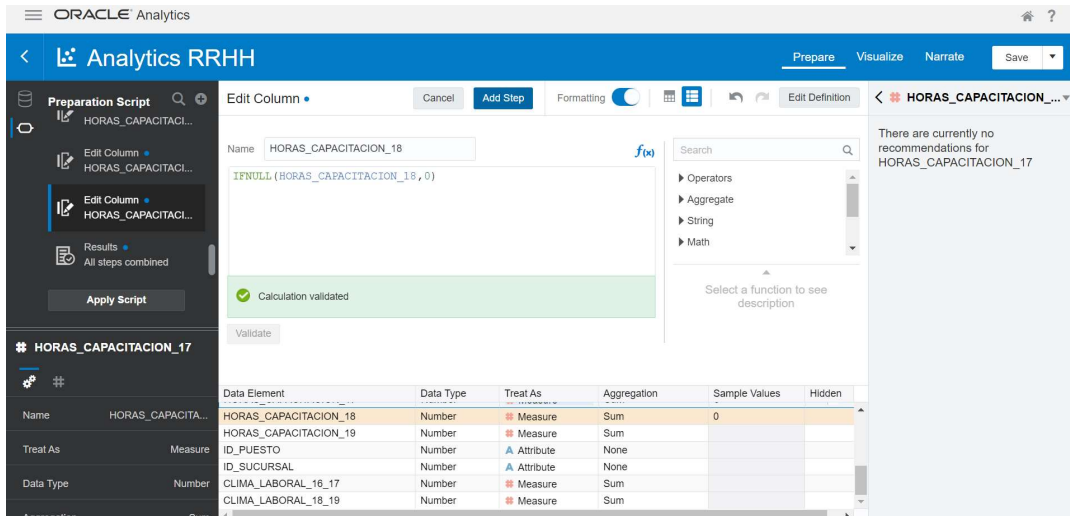
Figura 138 Preparación del *data set*. Paso 14



Fuente: Elaboración Propia con Oracle

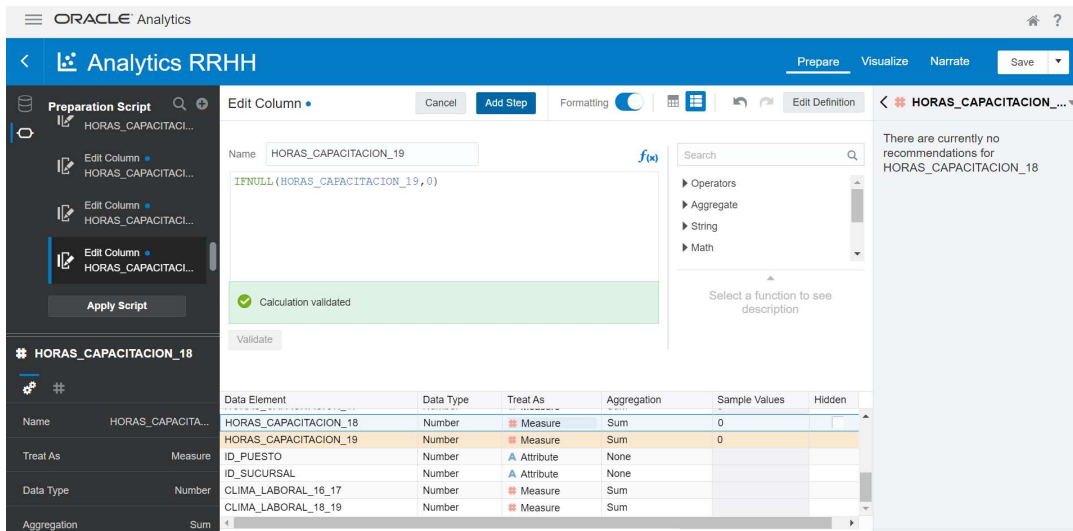
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 139 Preparación del *data set*. Paso 15



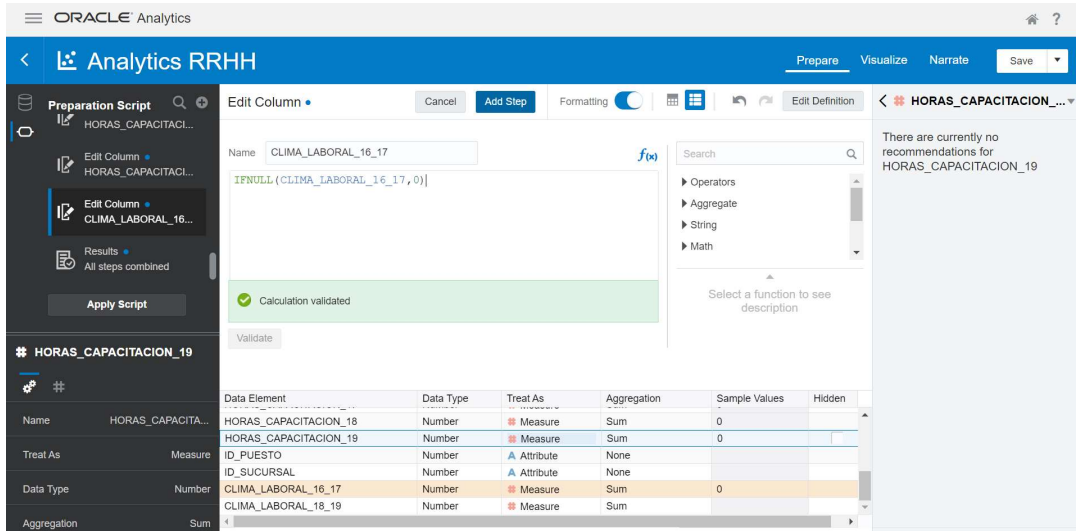
Fuente: Elaboración Propia con Oracle

Figura 140 Preparación del *data set*. Paso 16



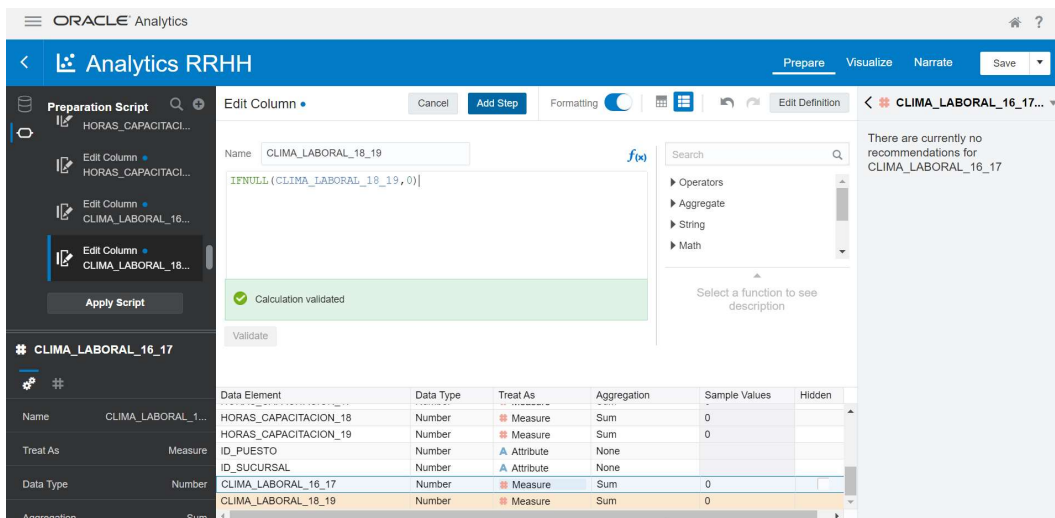
Fuente: Elaboración Propia con Oracle

Figura 141 Preparación del *data set*. Paso 17



Fuente: Elaboración Propia con Oracle

Figura 142 Preparación del *data set*. Paso 18

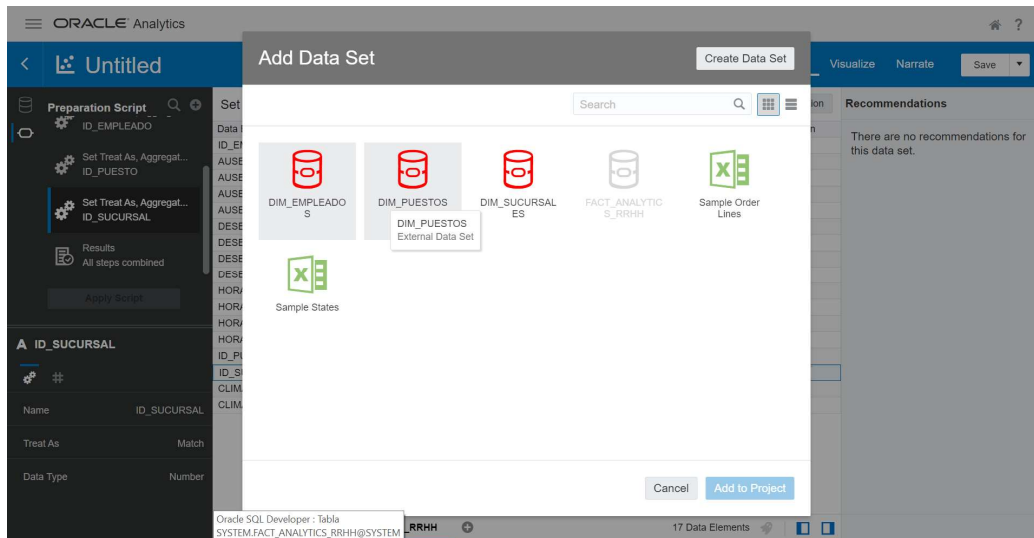


Fuente: Elaboración Propia con Oracle

Una vez realizado todo el tratamiento de los datos para la *fact* se hizo click sobre la opción “*Apply Script*” a fin de salvar los cambios ejecutados.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

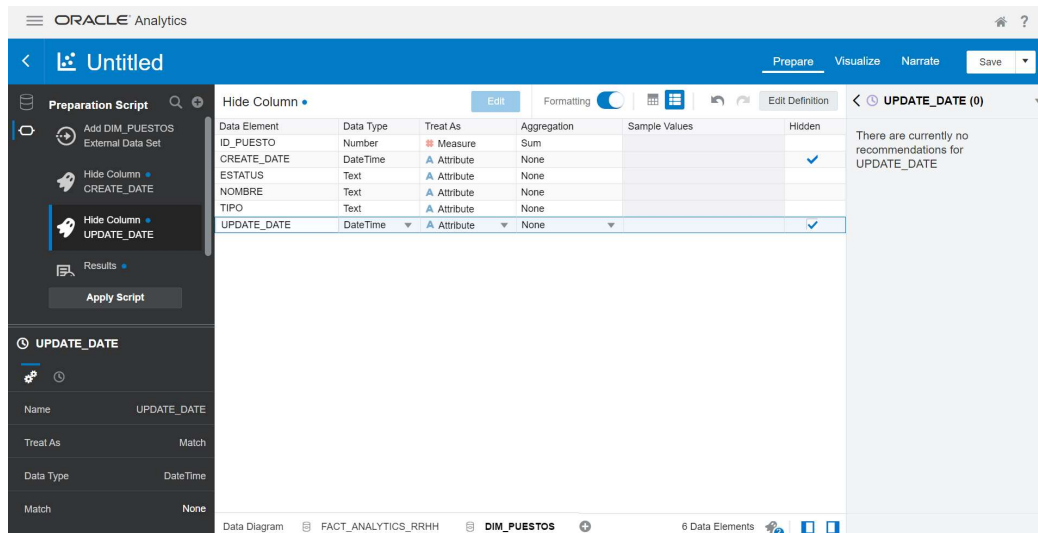
Figura 143 Complementación del *data set*. Paso 1



Fuente: Elaboración Propia con Oracle

Haciendo click sobre el icono + ubicado en la parte inferior de la pantalla se procedió a adicionar la DIM_PUESTOS para complementar el set de datos de los reportes.

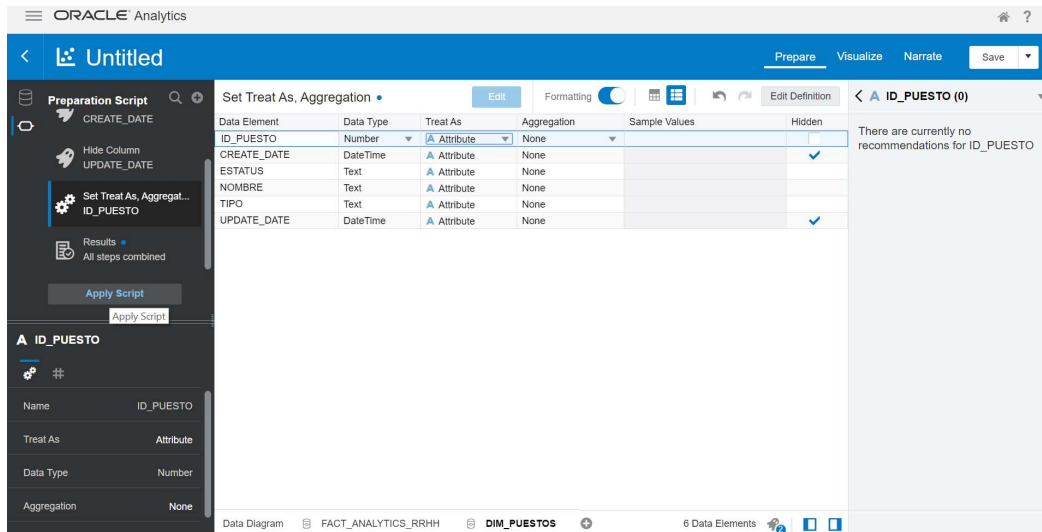
Figura 144 Depuración de datos. Paso 1



Fuente: Elaboración Propia con Oracle

Se ocultaron los campos que no serían requeridos para el desarrollo de las visualizaciones a fin de depurar el modelo de datos haciendo click en la columna “*Hidden*”

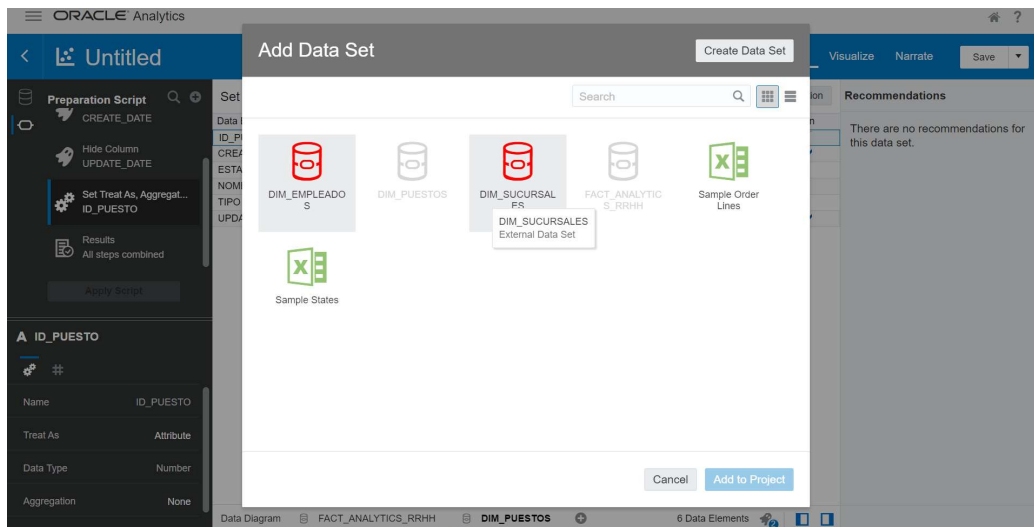
Figura 145 Definición de tratamiento de datos



Fuente: Elaboración Propia con Oracle

Se definió el tratamiento para los datos acorde a lo definido según el modelo indicando “Attribute o Measure” y se confirmaron los cambios sobre la DIM_PUESTOS mediante la opción “Apply Script”.

Figura 146 Complementación del data set. Paso 2

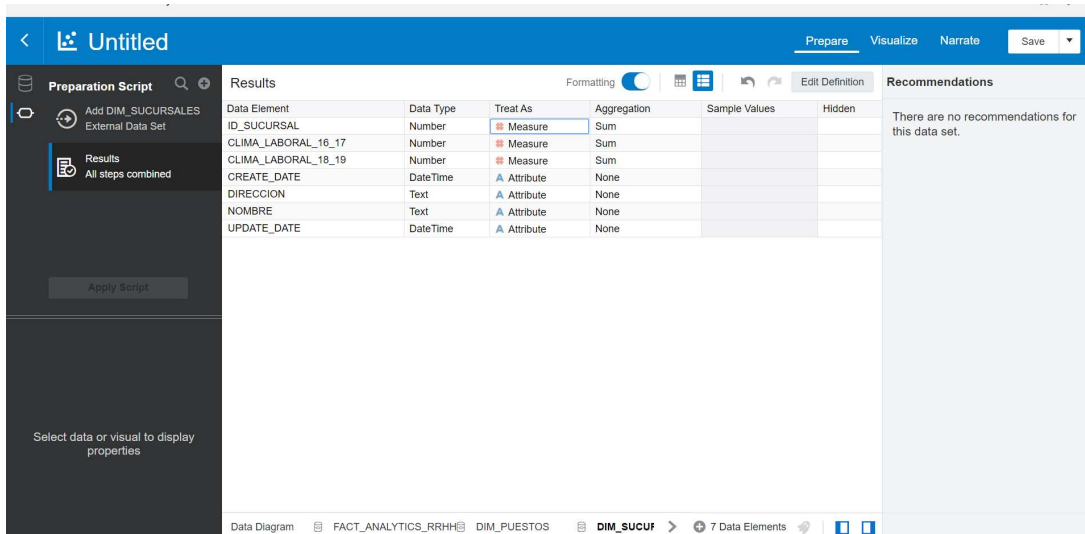


Fuente: Elaboración Propia con Oracle

Finalmente se incorporó la DIM_SUCURSALES para completa el modelo en analytics y proceder a la creación de los diferentes tableros.

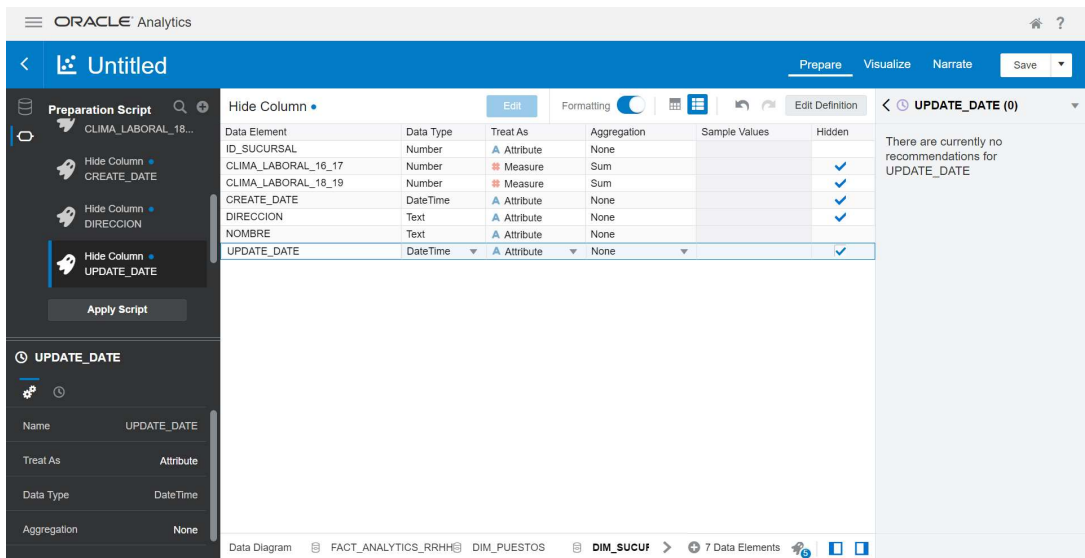
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 147 Complementación del *data set*. Paso 3



Fuente: Elaboración Propia con *Oracle*

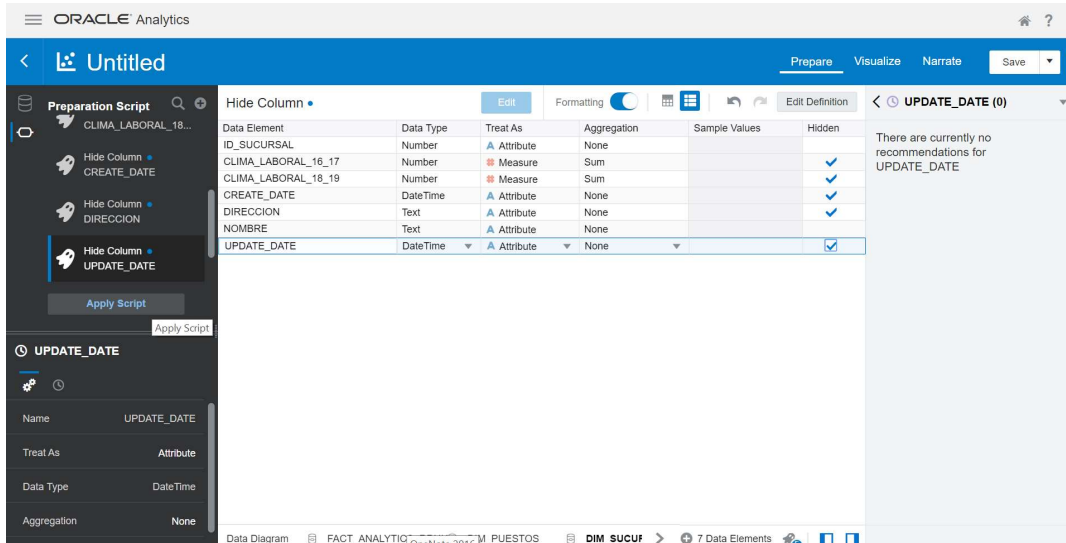
Figura 148 Depuración de datos. Paso 2



Fuente: Elaboración Propia con *Oracle*

Se ocultaron los campos que no serían requeridos para el desarrollo de las visualizaciones a fin de depurar el modelo de datos haciendo click en la columna “*Hidden*”.

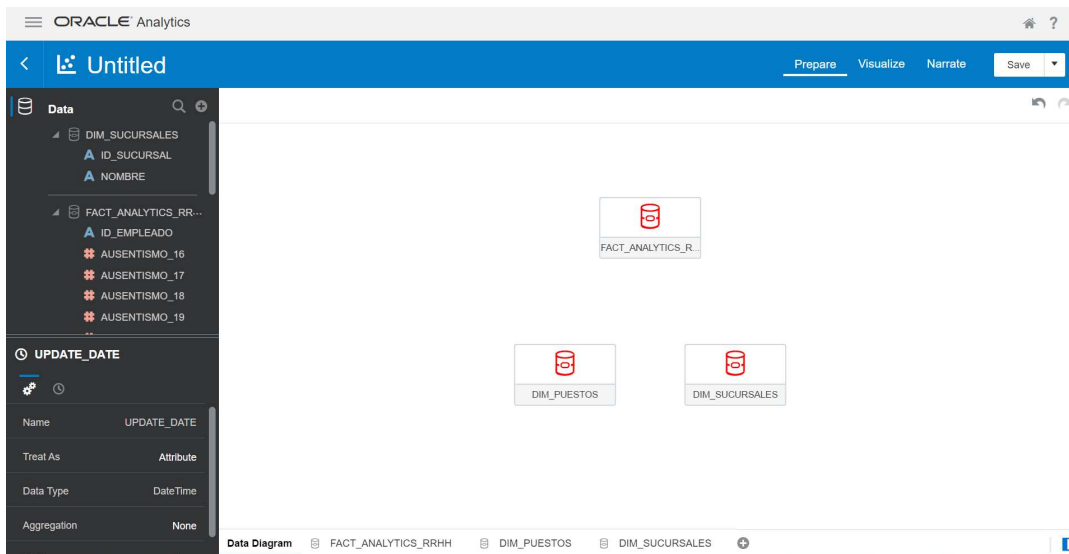
Figura 149 Definición de tratamiento de datos



Fuente: Elaboración Propia con Oracle

Se aplicaron los cambios haciendo click sobre la opción “Apply Script”.

Figura 150 Diagrama de datos las relaciones existentes. Paso 1

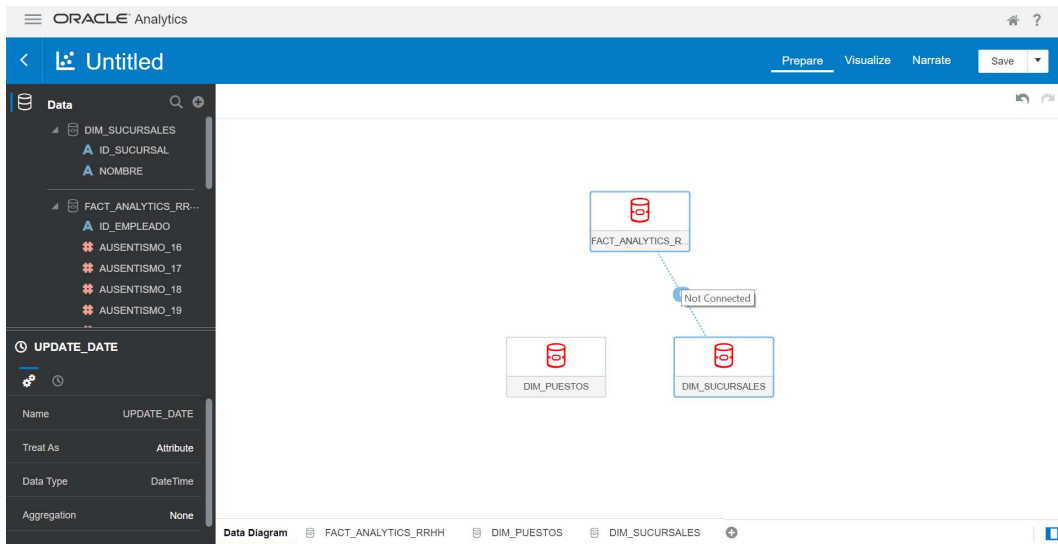


Fuente: Elaboración Propia con Oracle

Para culminar la preparación de los datos previos a la creación de las visualizaciones se estableció en el Diagrama de datos las relaciones existentes sobre cada uno de los sets de datos mediante su *primary key* definido en el modelo.

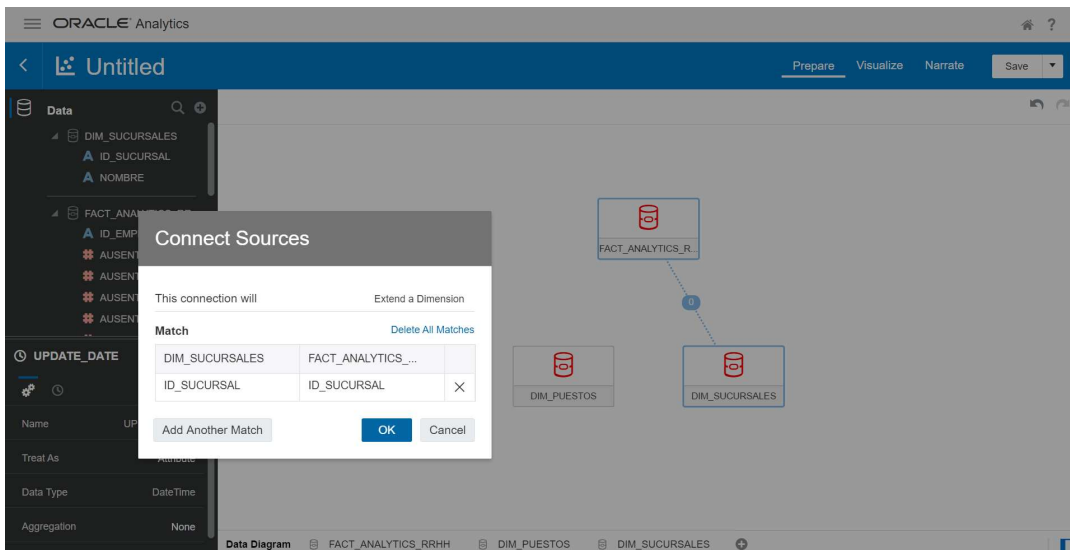
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 151 Diagrama de datos las relaciones existentes. Paso 2



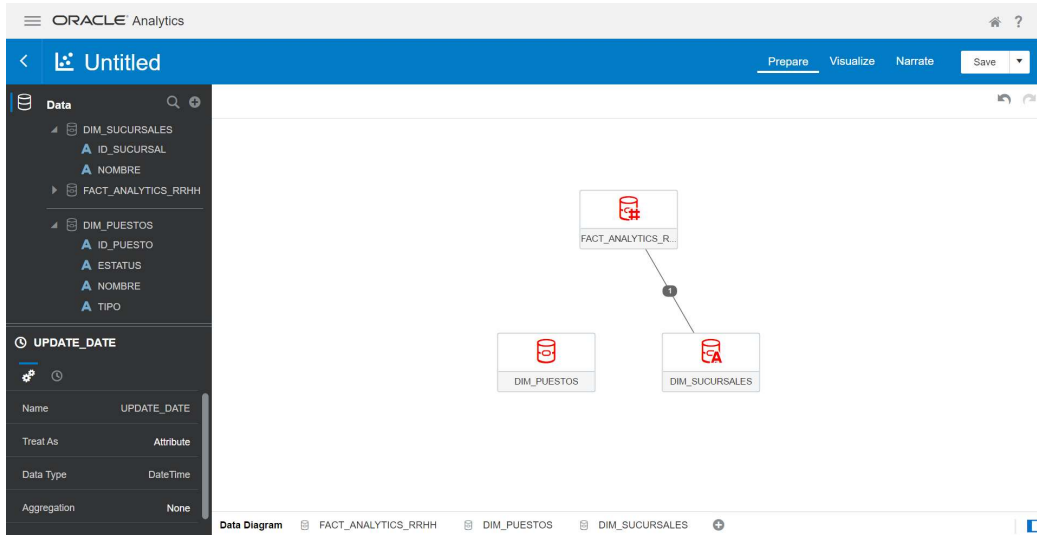
Fuente: Elaboración Propia con *Oracle*

Figura 152 Diagrama de datos las relaciones existentes. Paso 3



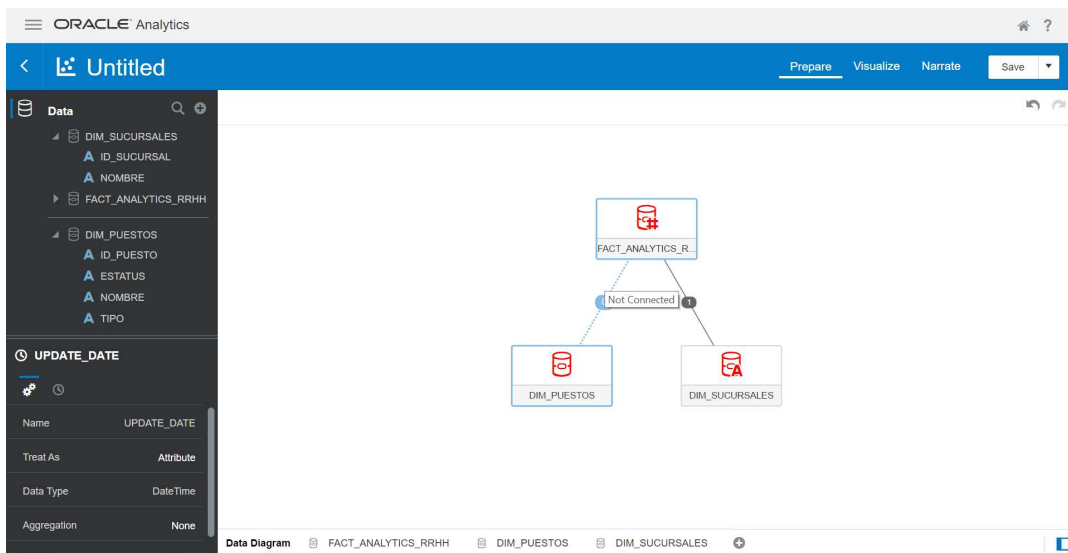
Fuente: Elaboración Propia con *Oracle*

Figura 153 Diagrama de datos las relaciones existentes. Paso 4



Fuente: Elaboración Propia con *Oracle*

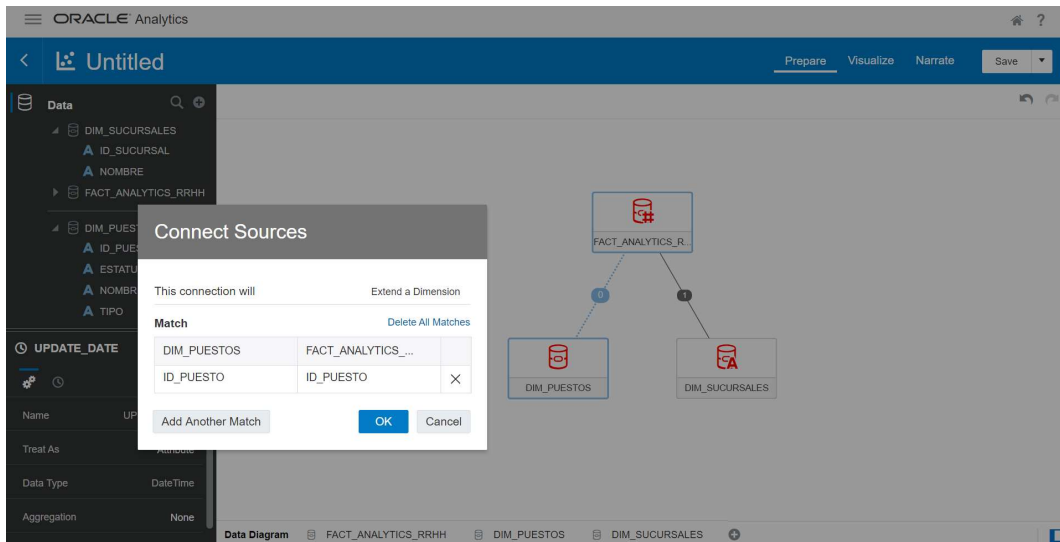
Figura 154 Diagrama de datos las relaciones existentes. Paso 5



Fuente: Elaboración Propia con *Oracle*

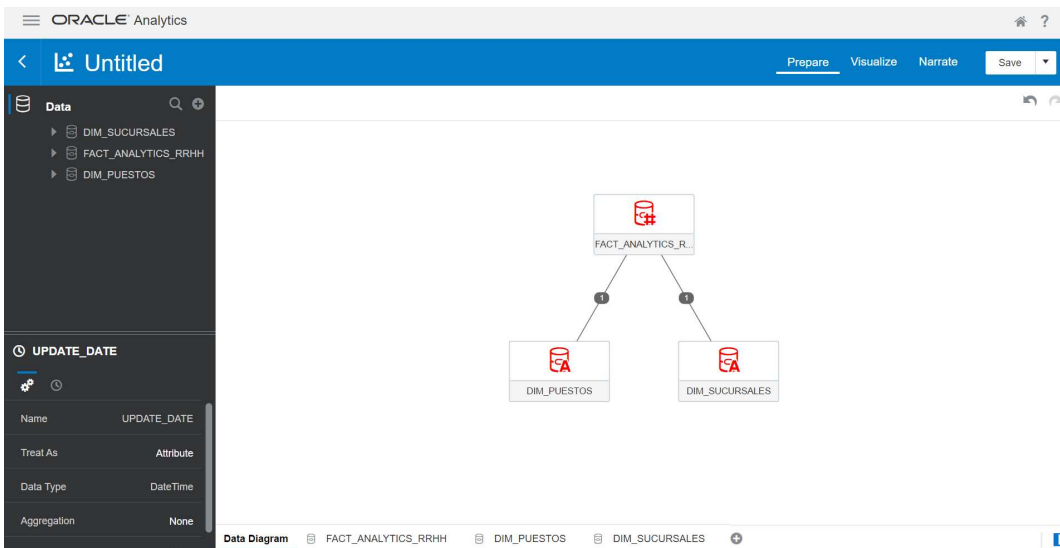
“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 155 Diagrama de datos las relaciones existentes. Paso 6



Fuente: Elaboración Propia con *Oracle*

Figura 156 Diagrama de datos las relaciones existentes. Paso 7



Fuente: Elaboración Propia con *Oracle*

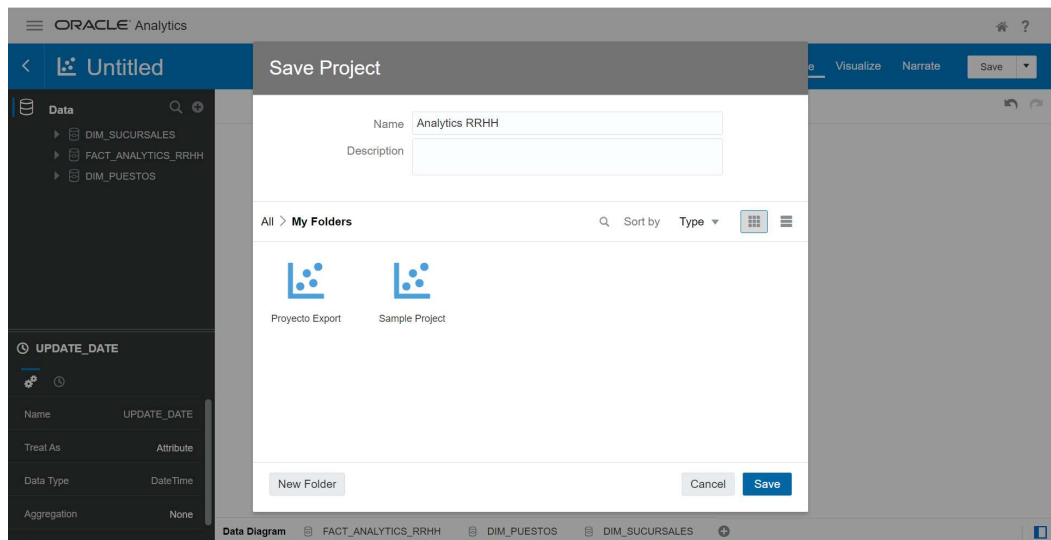
Figura 157 Diagrama de datos las relaciones existentes. Paso 8



Fuente: Elaboración Propia con *Oracle*

Luego de culminar la preparación de los datos se procedió a guardar el proyecto bajo el nombre “*Analytics RRHH*” y desarrollar las visualizaciones que serían analizadas en apartados posteriores.

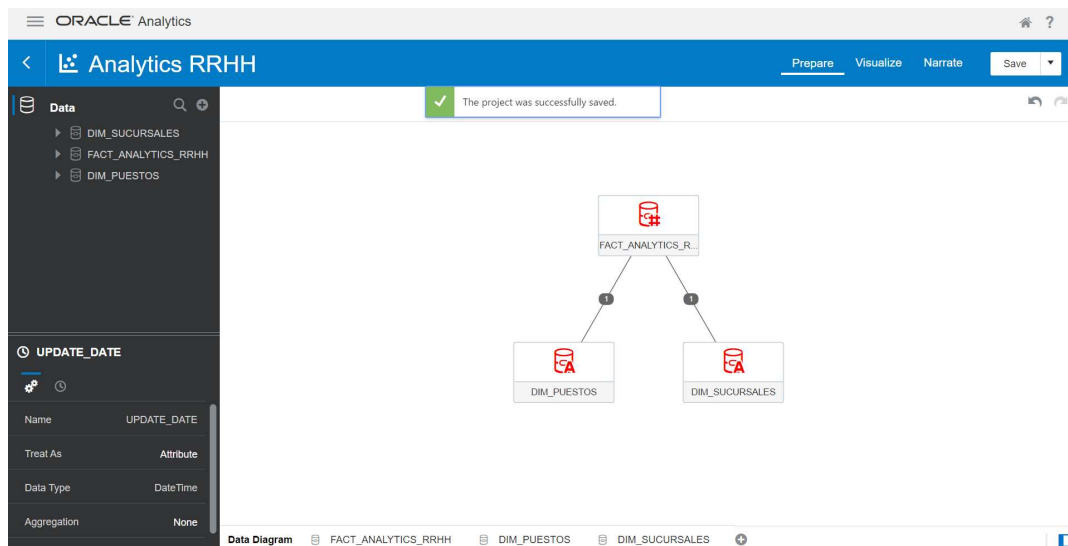
Figura 158 Guardado de proyecto. Paso 1



Fuente: Elaboración Propia con *Oracle*

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 159 Guardado de proyecto. Paso 2



Fuente: Elaboración Propia con *Oracle*

4.6.6 DISEÑO DE LAS PRUEBAS

En este apartado se profundiza con ayuda del prototipo las variables analizadas y que se utilizaron para relevar los datos de la investigación, así como también, cual es la importancia de estos mismos para el área de recursos humanos. Las variables fueron seleccionadas para poder relacionar el impacto de las actividades de formación en el desempeño de las personas y el impacto del clima organizacional en el ausentismo.

Estas mismas le dan la posibilidad y la capacidad a la gerencia de medir con objetividad los datos que proporcionan las acciones, no solamente para analizar los desvíos presupuestarios de recursos financieros y recursos no financieros, sino también los efectos que causan y que a su vez podrían ser consecuencia de la falta de atención en estos factores. De tal manera que se puede visualizar los resultados en cada una de las sucursales a partir de la información relevada.

La empresa debe ser un sujeto que facilite el estado de bienestar, salud y potenciar estrategias para garantizar el bienestar de los trabajadores y la calidad de vida. Debe además garantizar la administración equitativa de sus recursos y velar por asegurar una productividad mayor o igual a la de la competencia.

En el primer caso se analiza la relación entre ausentismo y el clima laboral; esto en dos períodos el bienio 2016-2017 y 2018-2019. En la figura 1 se ve de qué manera correlacionan estas variables, aunque a primera vista se dificulta ver la incidencia de una sobre otra. Para subsanar esta situación en la figura 2

se ordena las sucursales de mayor a menor clima laboral. Ésta imagen nos permite llegar a la conclusión que en el bienio 2016-2017 no hay una relación clara entre clima laboral y ausentismo.

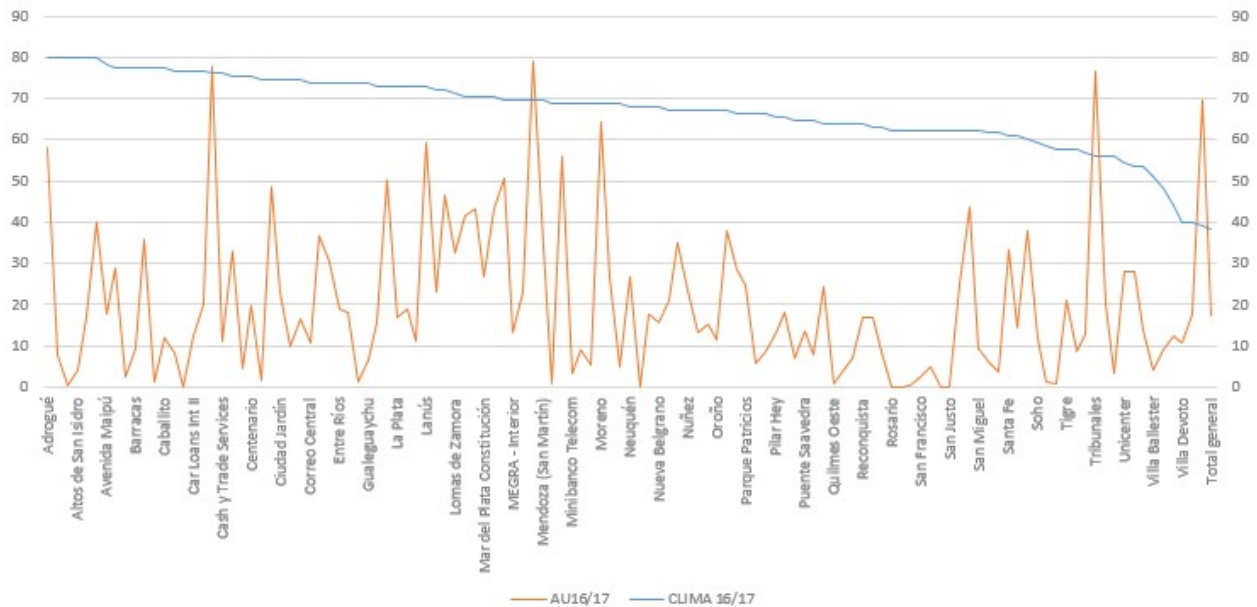
Figura 160 Ausentismo 2016-2017 vs Clima laboral 2016-2017



Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 161 Ausentismo 2016-2017 vs Clima laboral 2016-2017. Ordenado por clima laboral de mayor a menor



Fuente: Elaboración Propia con *Oracle*

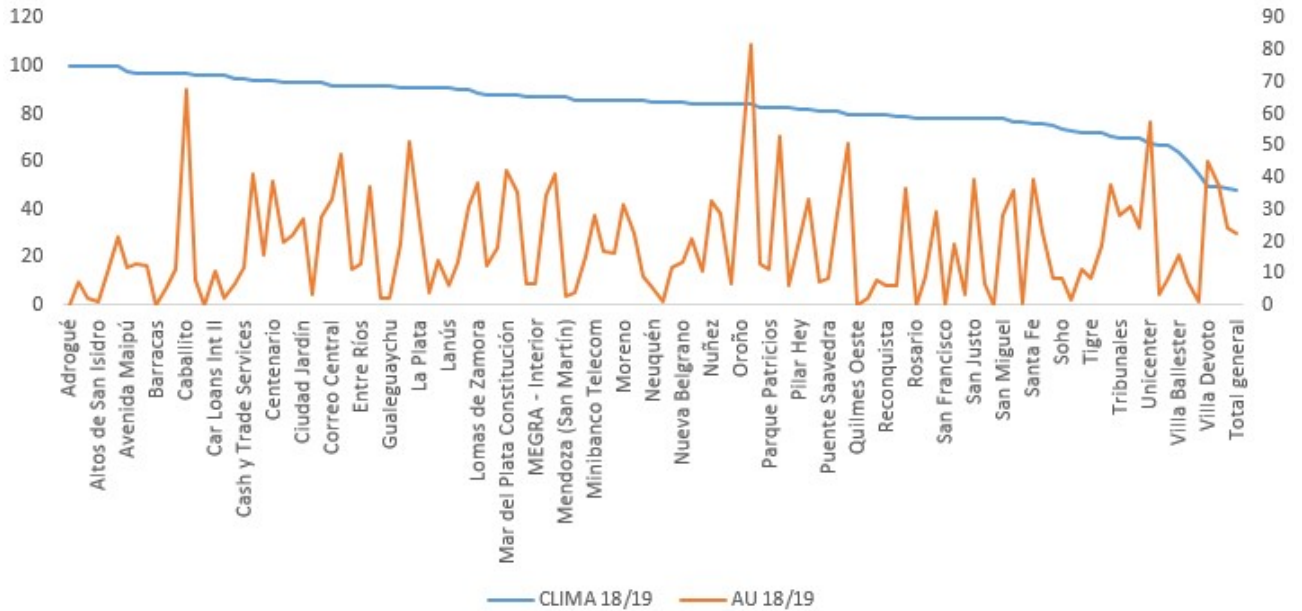
Se repite la metodología para el período 2018-2019. Se ordena las sucursales de mayor a menor clima laboral. Nuevamente se observa que no hay relación entre clima laboral y ausentismo. Al realizar este análisis intraperíodo se puede concluir que el clima laboral no determina el ausentismo.

Figura 162 Ausentismo 2018-2019 vs Clima laboral 2018-2019



Fuente: Elaboración Propia con Oracle

Figura 163 Ausentismo 2018-2019 vs Clima laboral 2018-2019. Ordenado por clima laboral de mayor a menor

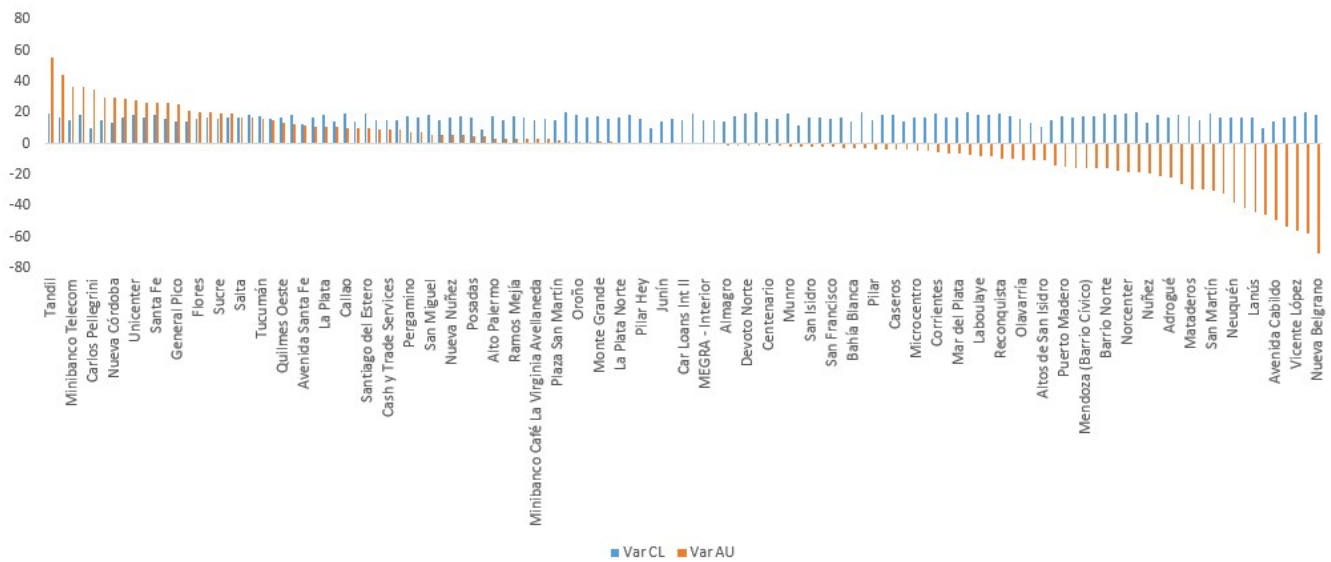


Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Para avanzar en el análisis se realiza una comparación entre períodos para observar si el ausentismo está afectado por los cambios, mejor o peor clima laboral. A tal fin es que se comparó, primero, la evolución entre el clima laboral entre las sucursales entre períodos como así el ausentismo y se los plasmó en la figura 164. En esta ocasión tampoco se observa una relación, ni positiva, ni negativa entre ambas variables dado que el clima laboral ha mejorado en todas las sucursales, mientras que el ausentismo se ha incrementado o disminuido independientemente de esto.

Figura 164 Diferencia clima laboral 2016/2017 – clima laboral 2018/2019 vs ausentismo 2016/2017 – ausentismo 2018/2019



Fuente: Elaboración Propia con *Oracle*

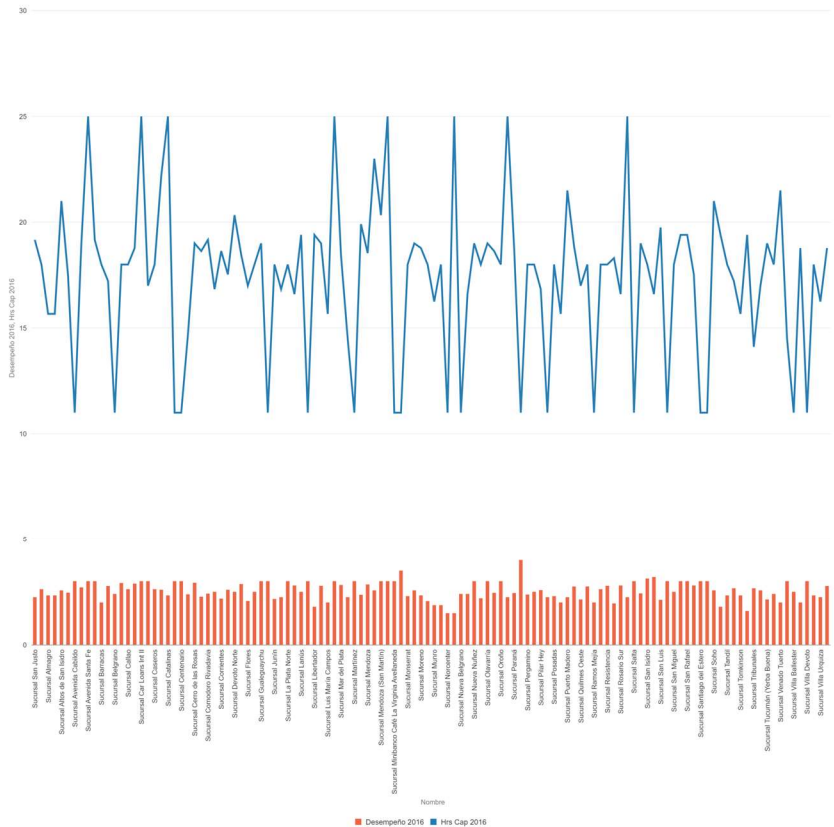
Al realizar este análisis, se concluye que el clima laboral no es determinante en el ausentismo, ya sea al verlo intra períodos o entre períodos. No realizar este tipo de estudios previamente al asignar recursos financieros o no financieros con el objetivo de reducir el ausentismo, puede llevar a hacer una asignación ineficiente e inadecuada de los mismos.

El segundo eje que se analiza es la correspondencia entre las horas de capacitación otorgadas por la empresa a los colaboradores y el desempeño de los mismos. En este caso se tomaron datos para las mismas 120 sucursales de la entidad financiera para cuatro años consecutivos 2016 a 2019.

Figura 165 Horas de capacitación 2016 vs desempeño 2016

Desempeño vs Hrs Cap 2016

Nombre	Desempeño 2016	Hrs Cap 2016
Sucursal San Justo	2,25	19,17
Sucursal Adrogue	2,63	18,00
Sucursal Almagro	2,33	15,67
Sucursal Alto Palermo	2,33	15,67
Sucursal Altos de San Isidro	2,57	21,00
Sucursal Avellaneda	2,46	17,46
Sucursal Avenida Cabildo	3,00	11,00
Sucursal Avenida Mapá	2,71	19,00
Sucursal Avenida Santa Fe	3,00	25,00
Sucursal Bahía Blanca	3,00	19,17
Sucursal Barracas	2,00	18,00
Sucursal Barrio Norte	2,78	17,22
Sucursal Belgrano	2,40	11,00
Sucursal Caballito	2,92	18,00
Sucursal Calles	2,63	18,00
Sucursal Campana	2,89	18,78
Sucursal Car Loans Int II	3,00	25,00
Sucursal Carlos Pellegrini	3,00	17,00
Sucursal Caseros	2,63	18,00
Sucursal Cash y Trade Services	2,60	22,20
Sucursal Catalinas	2,33	25,00
Sucursal Cavia	3,00	11,00
Sucursal Centenario	3,00	11,00
Sucursal Centro	2,38	14,66
Sucursal Cerro de las Rosas	2,63	19,00
Sucursal Ciudad Jardín	2,27	18,64
Sucursal Comodoro Rivadavia	2,42	19,17
Sucursal Conco Cental	2,50	16,83
Sucursal Comentes	2,18	18,64
Sucursal Córdoba	2,60	17,53
Sucursal Devoto Norte	2,50	20,33
Sucursal Entre Ríos	2,87	18,47
Sucursal Flores	2,07	17,00
Sucursal General Pico	2,50	18,00
Sucursal Gualeguaychú	3,00	19,00
Sucursal Iny Capas	3,00	11,00
Sucursal Jardín	2,17	18,00
Sucursal La Plata	2,25	16,83
Sucursal La Plata Norte	3,00	18,00
Sucursal Laboye	2,80	16,60
Sucursal Lanús	2,50	19,40
Sucursal Las Heras	3,00	11,00
Sucursal Libertador	1,80	19,40
Sucursal Lomas de Zamora	2,78	19,00
Sucursal Luis María Campos	2,00	15,67
Sucursal MEGRA - Interior	3,00	25,00
Sucursal Mar del Plata	2,82	18,41
Sucursal Mar del Plata Constitución	2,25	14,50
Sucursal Martínez	3,00	11,00
Sucursal Mataderos	2,36	19,91
Sucursal Mendoza	2,85	18,54
Sucursal Mendoza (Barrio Chico)	2,57	23,00
Sucursal Mendoza (San Martín)	3,00	20,33
Sucursal Microcentro	3,00	25,00
Sucursal Minibanco Café La Virginia Avellaneda	3,00	11,00

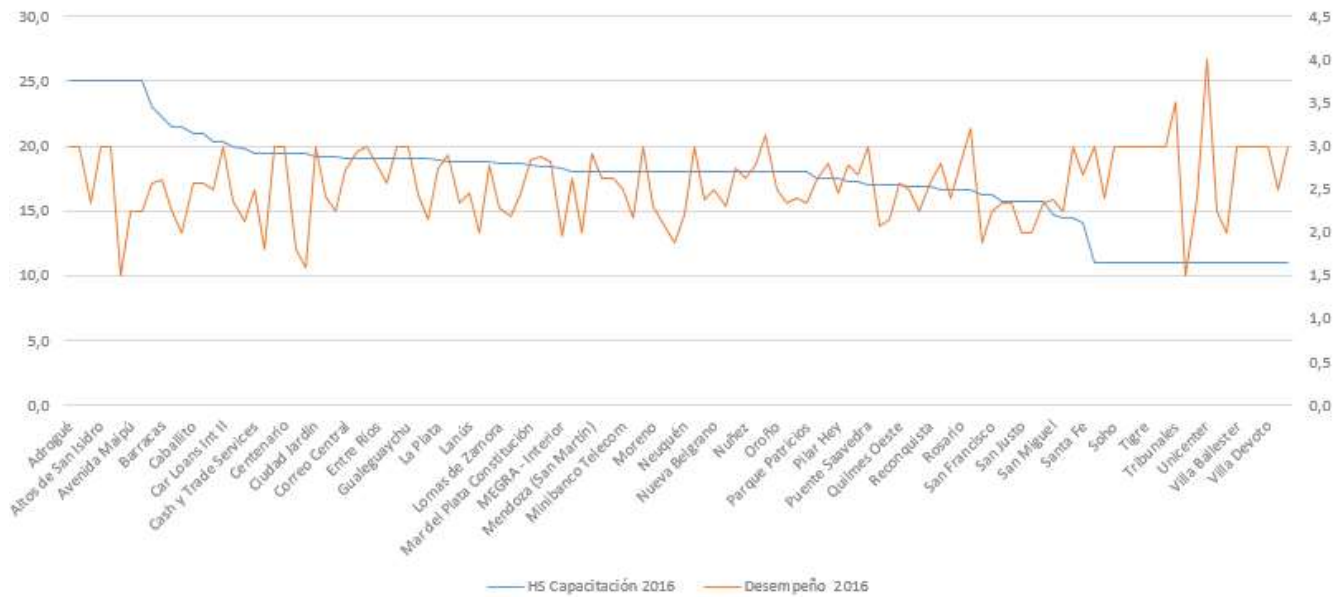


Fuente: Elaboración Propia con Oracle

En las figuras 165, 166, 167, 168, 169, 170, 171 Y 172 se divisa que para los cuatro años en análisis no hay relación dominante entre horas de capacitación y desempeño. A priori esto es un resultado contraintuitivo, ya que inicialmente las personas esperarían que a mayor capacitación del personal redunde mejor desempeño. Quedará para un futuro trabajo confirmar si esta relación se verifica o si existen problemas de diseño para los indicadores de desempeño u capacitación que no permitan captar las particularidades de cada caso.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 166 Horas de capacitación 2016 vs desempeño 2016. Ordenado por horas de capacitación de mayor a menor



Fuente: Elaboración Propia con *Oracle*

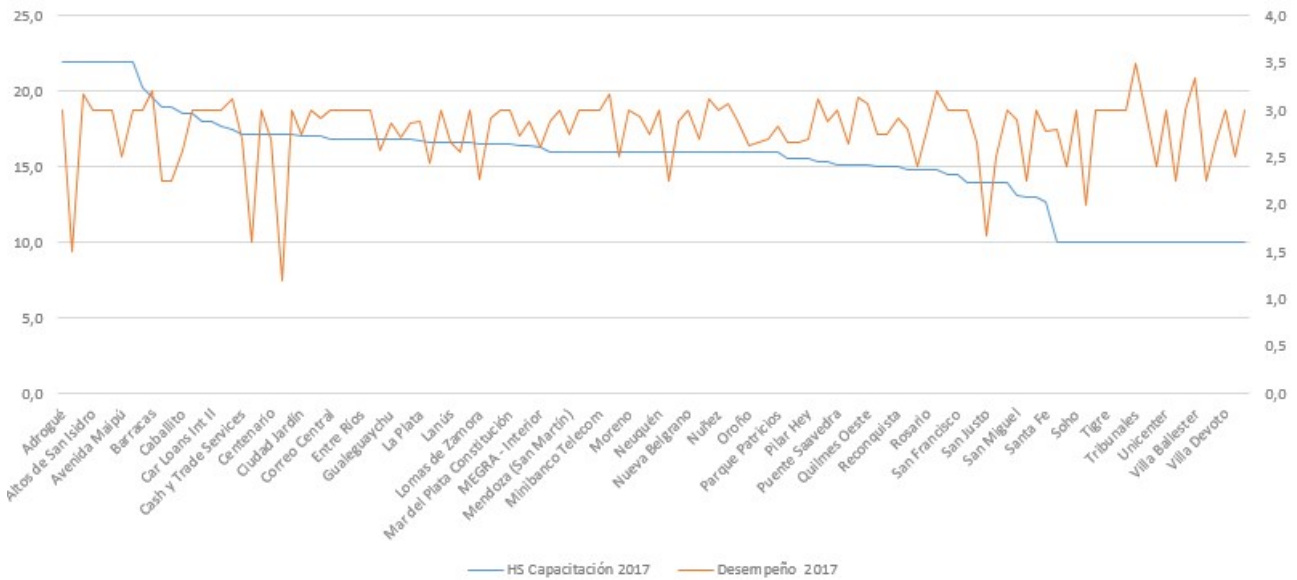
Figura 167 Horas de capacitación 2017 vs desempeño 2017

Desempeño vs Hrs Cap 2017



Fuente: Elaboración Propia con *Oracle*

Figura 168 Horas de capacitación 2017 vs desempeño 2017. Ordenado por horas de capacitación de mayor a menor



Fuente: Elaboración Propia con Oracle

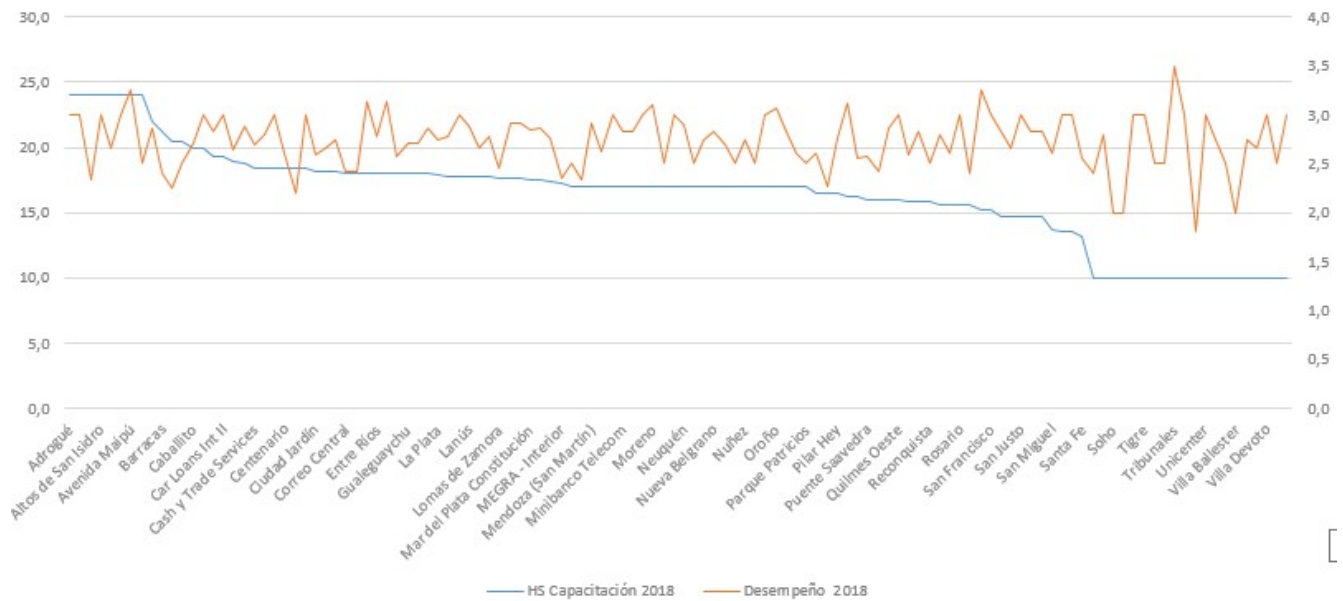
Figura 169 Horas de capacitación 2018 vs desempeño 2018



Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Figura 170 Horas de capacitación 2018 vs desempeño 2018. Ordenado por horas de capacitación de mayor a menor



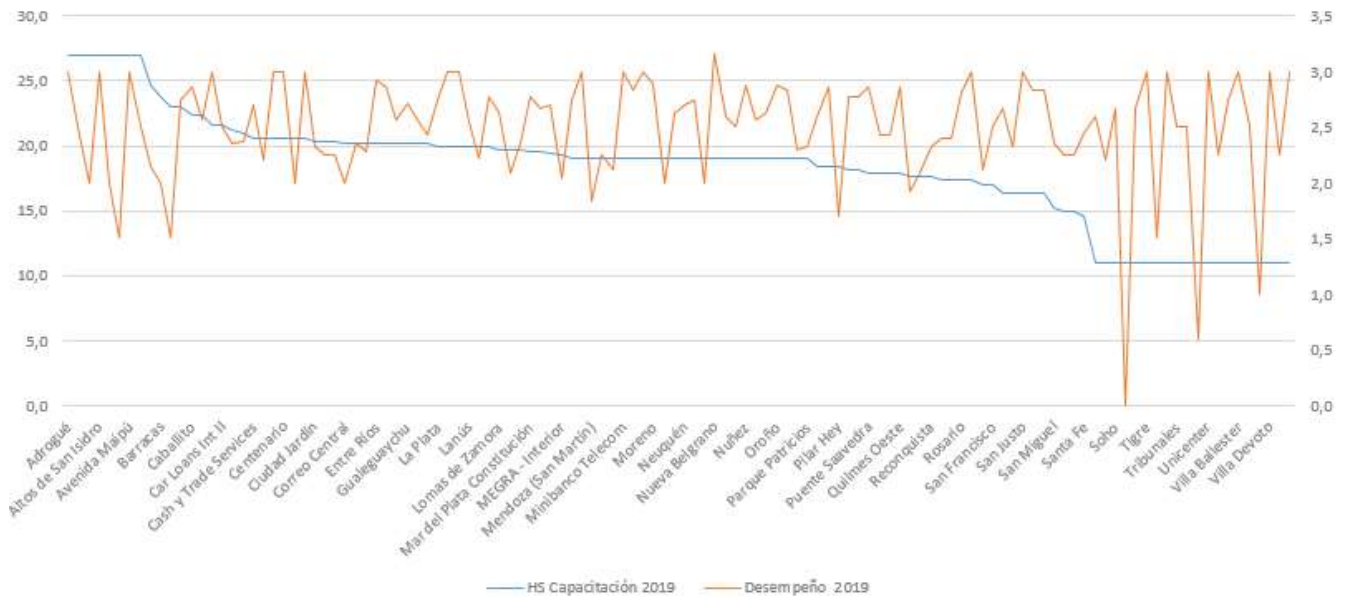
Fuente: Elaboración Propia con Oracle

Figura 171 Horas de capacitación 2019 vs desempeño 2019



Fuente: Elaboración Propia con Oracle

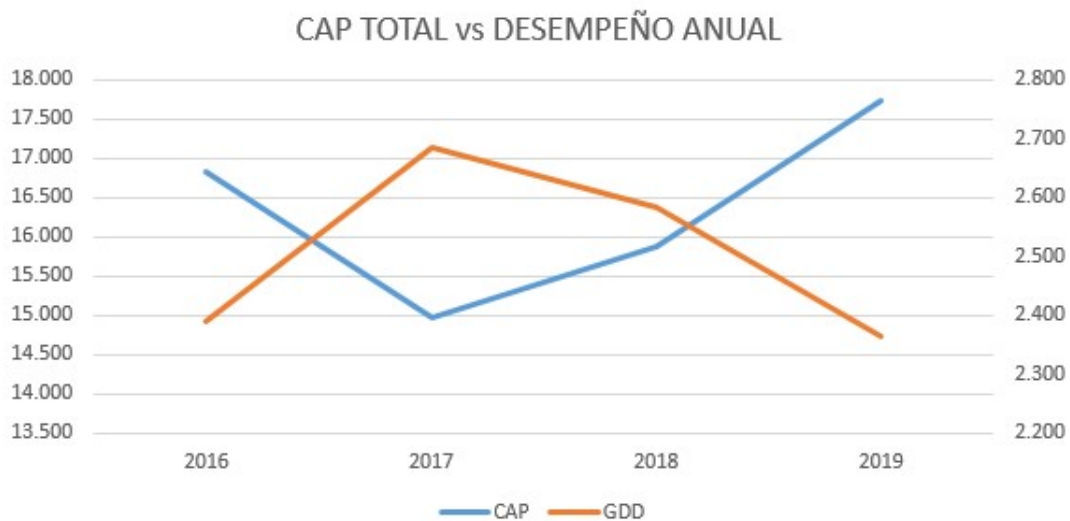
Figura 172 Horas de capacitación 2019 vs desempeño 2019. Ordenado por horas de capacitación de mayor a menor



Fuente: Elaboración Propia con Oracle

Al realizar un análisis agregado anualmente sin distinguir por sucursal se observa una relación inversa entre las horas de capacitación y el desempeño, es decir, a mayor horas de capacitación, menor desempeño. Nuevamente es un resultado contraintuitivo que puede deberse entre otros por dos factores no captados en los indicadores; mal diseño del indicador de desempeño o la existencia de un desfase entre ambas variables, es posible que los efectos de capacitación se evidencien seis o doce meses después en el puesto de trabajo.

Figura 173 Horas de capacitación total anual 2016 a 2019 y desempeño total anual 2016 a 2019



Fuente: Elaboración Propia con Oracle

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera.
Un caso de estudio”

Tal como se mencionó en los párrafos anteriores, es de vital importancia, realizar este tipo de análisis previa a la asignación de recursos donde se permita encontrar las variables que mejor correlacionen a fin de lograr el objetivo planteado inicialmente por el área de recursos humanos.

CAPÍTULO 5: CONCLUSIONES

5.1 CONCLUSIONES

La gestión de recursos humanos, las condiciones laborales y de vida son los principales determinantes del desempeño laboral en los trabajadores. Los estudios de caso ayudan a obtener información precisa para realizar una investigación completa de un contexto particular, identificar nuevos problemas, focalizar conceptos o variables promisorias, establecer prioridades para investigaciones futuras o sugerir afirmaciones y postulados.

Las organizaciones comprenden que deben formar profesionales eficientes y al hacerlo lo deben llevar a cabo de manera más asertiva, porque de otro modo no les resulta posible ganar las batallas de competitividad y expansión del mercado y en muchos casos se pone en juego la supervivencia de la misma.

Muchas empresas buscan trascender la concepción básica del vínculo trabajador - empleador mediante políticas de compensaciones, beneficios o recompensas, estrategias de motivación o al implementar programas de liderazgo. Todo lo mencionado tiene como contraparte, la asignación de recursos escasos con los que cuentan las organizaciones que deben ser administrados de manera eficiente para cumplir con los objetivos trazados por los diferentes estamentos y gerencias de las compañías.

Así es que, la dirección de las empresas en general y el área de recursos humanos en particular, debe valorar al recurso humano en una dimensión más completa –intelectual, social y afectiva- con el objetivo de aplicar las prácticas y políticas que puedan potenciar cada una de estas dimensiones del capital humano de sus empleados, las cuales se configuran en características de difícil observación, mensura y relación. Todas ellas son necesarias para maximizar el valor que puedan aportar los empleados a la organización a través de un capital humano que en el siglo XXI debe pensarse como multidimensional.

Los procesos de *big data*, minería de datos y *analytics* están orientados al desarrollo de métodos que exploran las bases de datos de las empresas generadas *ad-hoc*, así como el uso de las técnicas para transformar dichos datos en información y entender mejor el proceso de aprendizaje y conocimiento de los mismos. El motivo de la investigación de tesis es identificar los beneficios en la calidad de la toma de decisiones a partir del uso de *big data* y analítica aplicada a la gestión de recursos humanos, *people analytics* con respecto a la gestión tradicional, más relacionada a la subjetividad de los observadores.

En el caso que da origen a esta tesis, los analistas de recursos humanos de la entidad financiera podrán usar la metodología propuesta por este trabajo para identificar y establecer procedimientos que permitan

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

detectar con modernas herramientas las variables relevantes con el objeto de mejorar la administración de recursos financieros y no financieros al mejorar la calidad de la toma de decisiones.

Las empresas, se enfrentan a un gran desafío: comprender los diversos entornos por los que su personal atraviesa y así mismo generar políticas de fortalecimiento para su superación diaria. La implementación del prototipo, *Oracle Data Visualization*, apunta a lograr los objetivos de las organizaciones ya que con la misma se analiza el comportamiento de las personas en el ámbito laboral de una manera sistemática, ágil y a través del tiempo con parámetros objetivos.

La estructura de una organización debe ser resultado de la selección de elementos, teniendo en consideración la búsqueda de consistencia interna (procesos internos) y externa (medio ambiente y contexto). Administrar el clima laboral, el ausentismo, las actividades de formación y el desempeño de los colaboradores es necesario para mejorar la productividad de la organización al buscar las debilidades, promover los cambios y administrar recursos.

El análisis previo a la investigación que se desarrolló permitió apreciar la importancia que tiene el proceso de recopilación de datos, abarcando las fases de análisis y preparación de los datos asociado a la metodología CRISP-DM.

El análisis de parámetros objetivos y su evolución en el tiempo, permite llegar a conclusiones que a priori se configuran contraintuitivas pero que de otra forma llevarían a una asignación ineficiente de recursos. En el estudio se pudo observar que si la entidad financiera quería mejorar el ausentismo de los trabajadores no debía destinar recursos a mejorar el clima laboral, ya que no se observa una relación entre ambas variables, ni intra períodos, ni entre períodos (esto se traduce que el clima laboral en un momento dado o el mejoramiento del clima laboral entre dos períodos no tienen un impacto significativo sobre el ausentismo).

Así mismo el análisis a través del *Oracle Data Visualization* sobre la relación de horas de capacitación y desempeño, arroja un resultado que inicialmente hubiese parecido contradictorio dado que no se hayan evidencias significativas de que una mayor cantidad de horas de capacitación, redunde en mejoras en el desempeño. Vale aclarar que quedará para un futuro trabajo confirmar si esta relación se verifica o si existen problemas de diseño para los indicadores de desempeño o capacitación que no permitan captar las particularidades de cada caso.

Una vez probados y evaluados los modelos de *big data*, minería de datos y *analytics*, se comprueba que los factores principales estudiados influyen en la administración de los recursos financieros y no

financieros. Los profesionales de recursos humanos están en condiciones de tomar decisiones basadas en datos.

5.2 TRABAJO A FUTURO

Mejorar indicadores sobre la gestión del desempeño. Analizar posibles desfases entre las variables de capacitación y performance.

Redeterminar, temporalmente, que los efectos de la capacitación se evidencien en mayor plazo en el puesto de trabajo que el que se utiliza en la actualidad.

Relacionar las dinámicas y frágiles estructuras de las organizaciones de hoy con las competencias de los integrantes de las mismas, incluyendo conocimientos y habilidades transportables. Transferencia a la gestión.

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

CAPÍTULO 6: BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Adam, J. S. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 422-436. Obtenido de <https://es.wikipedia.org>:
https://es.wikipedia.org/wiki/Teor%C3%ADa_de_la_equidad_laboral
- Aguilar, L. J. (2013). Big Data análisis de grandes volúmenes de datos en organizaciones. México: Alfaomega. Obtenido de
https://books.google.com.ar/books?id=1GywDAAAQBAJ&pg=PT417&lpg=PT417&dq=grandes+datos+simulacion+y+visualizacion&source=bl&ots=_WVdR42fUT&sig=ACfU3U3ae3VRQfAKXYchk8rjYpqKUBEHVQ&hl=es&sa=X&ved=2ahUKewjBjp2zqbnpAhWsIrkGHWEtD8YQ6AEwCXoECAkQAQ#v=onepage&q=gra
- Alles, M. (2015). *Dirección Estratégica de RRHH*. Ciudad Autónoma de Buenos Aires: Granica. doi:978-950-641-849-6
- Ambrose, S. J. (2015). The determinants of absenteeism: evidence from Australian blue-collar employees. *University of Melbourne , Australia Published online*.
- Anaconda*. (s.f.). Obtenido de <https://www.anaconda.com/products/enterprise>
- Anderson, C. (2016). *Charlas TED*. Ciudad Autónoma de Buenos Aires: Paidós Empresa. doi:978-950-12-9449-1
- Aponte, J. C. (2006). *Administración de Personal*. Bogotá: ECOE. doi:978-958-648-446-7
- Banco Central de la República Argentina. (2019). *Banco Central de la República Argentina*. Obtenido de BCRA: <https://www.bcra.gob.ar>
- Becker, G. (1984). *El Capital Humano: Un análisis teórico y empírico referido fundamentalmente a la educación*. Madrid: Alianza Universidad Textos. doi:84-206-8063-X
- Bersin, J. (07 de octubre de 2013). *Big Data in Human Resources: A World of Haves And Have-Nots*. Obtenido de FORBES: <https://www.forbes.com/sites/joshbersin/2013/10/07/big-data-in-human-resources-a-world-of-haves-and-have-nots/#5ffb33eb200f>

- Bersin, J. (2013). *www.deloitte.com*. Obtenido de Predictions for 2014:
<https://www2.deloitte.com/content/dam/Deloitte/au/Documents/human-capital/deloitte-au-con-predictions-bersin-2014-1213.pdf>
- Bersin, J. (2015). The geeks arrive in HR: People analytics is here. *FORBES*, <https://www.forbes.com>.
Obtenido de <https://www.forbes.com>
- Bodenheimer, J. M. (2017). *HR Analytics*. Ciudad Autónoma de Buenos Aires: Alfaomega. doi:978-987-3832-35-2
- Bolo, M. (5 de octubre de 2017). Menos intuición, más análisis. *El Cronista*, págs.
<https://www.cronista.com/columnistas/Menos-intuicion-mas-analisis-20171005-0010.html>.
- Bottlepy*. (s.f.). Obtenido de <https://bottlepy.org/docs/dev/>
- Bril, A. (2015). <http://repositorio.udesa.edu.ar/>. Obtenido de
<http://repositorio.udesa.edu.ar/jspui/bitstream/10908/11970/1/%5bP%5d%5bW%5d%20T.L.%20Adm.%20Bril%2c%20Alex.pdf>
- Brust, A. V. (31 de Mayo de 2020). *Ciencia de Datos para Gente Sociable*. Obtenido de
https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/
- Caihuelas, J. G. (2017). *Minería de Datos modelos y algoritmos*. Barcelona: UOC. doi:978-84-9116-904-8
- Cardozo, A. (2012). *Desarrollo Humano en las Organizaciones*. Buenos Aires: Temas UADE.
- Cardozo, A. P. (2012). *Desarrollo Humano en las Organizaciones*. Ciudad Autónoma de Buenos Aires: Temas UADE. doi:9789871826476
- Casa Rosada*. (s.f.). Obtenido de Casa Rosada: <https://www.casarosada.gob.ar/>
- Castañeda, B. &. (2010). *Procesamiento de Datos y Análisis Estadísticos utilizando SPSS*. Porto Alegre: EDIPUCRS. doi:978-85-7430-973-6
- CESEDEN - Centro Superior de Estudios de la Defensa Nacional. (2016). Big Data vs Seguridad Mundial. *Expresión Forense*, 61.
- Chapman, P. (2000). *CRISP-DM 1.0*. SPSS. Obtenido de <https://the-modeling-agency.com/crisp-dm.pdf>

- “Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”
- Cleveland, W. S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review* , 21-26.
- Coker, F. (2014). *Pulse: Understanding the Vital Signs of Your Business*. Ambient Light Publishing. doi:9780989308601
- Cood, E. F. (1970). *A Relational Model of Data for Large Shared Data Banks* . ACM.
- Cravino, L. M. (2003). *Un trabajo feliz*. Buenos Aires: Temas. doi:9789879164853
- Cravino, L. M. (2011). *Medir lo importante*. Buenos Aires: Temas. doi:978-950-9445-37-6
- Cravino, L. M. (2011). *Medir lo Importante*. Buenos Aires: TEMAS. doi:978-950-9445-37-6
- Cravino, L. M. (2011). *Medir lo Importante*. Buenos Aires: Temas. doi:978-950-9445-37-6
- Curran, R. &. (2015). *The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015*. Cambridge, USA: Forrester.
- Curto, J. (07 de Octubre de 2007). Obtenido de <http://josepcurto.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>
- David Turner, M. S. (2013). *Analítica: el uso de Big Data en el mundo real para el sector de servicios financieros*. Obtenido de IBM: ftp://public.dhe.ibm.com/la/documents/swg/es/analytics/IBM_Analitica_uso_de_Big_Data_en_mundo_para_sector_servicios_financieros.pdf
- Davidson, I. &. (2007). *Knowledge Discovery and Data Mining Challenges and Realities*. IGI Global. doi:9781280825286
- Deloitte University Press. (2016). *Deloitte University*. Obtenido de Human Capital: <https://www.deloitte.com>
- DOMO. (s.f.). Obtenido de <https://www.domo.com/learn/data-never-sleeps-7>
- Edwards, M. E. (2016). *Predictive HR Analytics: Mastering the HR Metric*. London: Kogan Page. doi:9780749473914
- Efrain, T. (2005). *Decision Support Systems and Intelligent Systems*. Upper Saddle River, NJ: Prentice Hall.
- Escudero, W. S. (2019). *Big Data*. Argentina: Siglo veintiuno. doi:978-987-629-899-5

- Fayyad, U. (1997). *Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases* .
- Fern, H. (2011). *THE TOP 5 TRENDS IN PREDICTIVE ANALYTICS: Maturing user adoption brings vision, viability, validity and value*. New York: Information Management. Obtenido de <https://search.proquest.com/openview/1f7afd15aead352df8e821b15b2fbb32/1?pq-origsite=gscholar&cbl=51938>
- Fitz-Enz, J. (2010). *The New HR Analytics*. New York: AMACOM. doi: 978-0-8144-1643-3
- Fitz-Enz, J. R. (2014). *Predictive Analytics for Human Resources*. New Jersey: Wiley. doi:9781118893678
- Frank, I. H. (2005). Data Mining: Practical machine learning tools and . En I. H. Frank, *Data Mining: Practical machine learning tools and* (pág. 525). San Francisco: Morgan Kaufmann Publishers .
- Gottlob, G. G. (2013). *Big Data*. Oxford: Springer. doi:978-3-642-39467-6
- Granja Camila Belén, K. M. (2016). <https://repositorio.uade.edu.ar>. Obtenido de <https://repositorio.uade.edu.ar/xmlui/handle/123456789/5257?show=full>
- Guterres, L. F. (2020). The role of work motivation as a mediator on the influence of education-training and leadership style on employee performance. *Management Science Letters*, 1497–1504. doi:10.5267/j.msl.2019.12.017
- Haykin, S. (2009). *Neural Networks*. Prentice Hall. doi:9780130956484
- Horacio Gómez, L. B. (2020). *PANDEMIA SIN AMNESIA: NO DEJEMOS EL CAPITAL HUMANO AL OLVIDO*. Monterrey: Sintec Consulting. Obtenido de <http://sintec-uploads.s3.amazonaws.com/wp-content/uploads/2020/07/08135515/Pandemia-sin-amnesia.pdf>
- <https://www.oracle.com>. (s.f.). Obtenido de <https://www.oracle.com/ar/human-capital-management/>
- Jaramillo, J. M. (2004). *Indicadores de gestión*. Bogotá: 3R. doi:978-9588017006
- Jesus, T. J.-G. (2010). La satisfacción laboral como valor intangible de los recursos humanos . *Teoría y Praxis*, 35-53.
- Joaquín Jesús Blanca-Gutiérrez, M. d.-D.-F. (2013). Intervenciones eficaces para reducir el absentismo. *Gaceta Sanitaria*, 545-551.
- Jupyter*. (s.f.). Obtenido de <https://jupyter.org/>

- “Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”
Kozyrkov, C. (22 de Diciembre de 2018). *www.medium.com*. Obtenido de <https://medium.com/datos-y-ciencia/qu%C3%A9-diablos-es-ciencia-de-datos-flc8c7add107>
- Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: Wiley-Interscience. doi:9781280275296
- Laudon, K. C. (2012). *Management Information Systems*. New Jersey: Pearson. doi:978-0-13-214285-4
- Linoff, M. J. (2004). *Data Mining Techniques*. Indianapolis, Indiana: Wiley Publishing. doi:9780471470649
- Llamas, E. H. (2000). *Productividad y mercado de trabajo en México*. Iztapalapa, México: Universidad Autónoma Metropolitana. doi:9706545514
- Locke, E. (1976). La naturaleza y las causas de satisfacción en el trabajo. *Rand McNally*, 155-156.
- Machín, O. H. (29 de septiembre de 2015). La capacitación para el desempeño profesional de especialistas en Gestión de los Recursos Humanos de la Dirección de Trabajo en Pinar del Río. *Mendive*, 508-515. Obtenido de <http://mendive.upr.edu.cu/index.php/MendiveUPR/article/view/787>
- Manual Dataprix TI. . (2009). <https://www.dataprix.com>. Obtenido de <https://www.dataprix.com/es/data-warehousing-y-metodologia-hefesto/34-datawarehouse-manager>
- Marin, R. (16 de Abril de 2019). Los gestores de bases de datos más usados en la actualidad. *Revista Digital INESEM*. Obtenido de <https://revistadigital.inesem.es/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/>
- Medina-Campos, X. L. (2020). Satisfacción laboral y ausentismo en el personal de enfermería de un hospital público de Monterrey. *Revista de Enfermería del Instituto Mexicano del Seguro Social*, 37-48. Obtenido de <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=94017>
- Michael Schroeck, R. S.-M. (2012). *Analytics: el uso de big data en el mundo real*. Obtenido de IBM: <https://www.fundacionseres.org/Lists/Informes/Attachments/951/IBM%20Analytics%20el%20uso%20de%20big%20data%20en%20el%20mundo%20real%20-%20Como%20las%20empresas%20mas%20innovadoras%20extraen%20valor%20de%20datos%20inciertos.pdf>

- Minelli, A. D. (2013). *Big data, big analytics*. New Jersey: Wiley. doi:9781118147603
- Minerva Data Mining*. (s.f.). Obtenido de <https://mnrva.io/kdd-platform.html>
- Morton, M. S. (1971). *Management decision systems: Computer based support for decision Making*. Massachusetts: Universidad de Harvard. doi:978-0875840901
- Noticias, artículos y guías sobre estrategia y operaciones de TI en las empresas. (s.f.). *SearchDataCenter.com/es*. Obtenido de <https://searchdatacenter.techtarget.com/es>
- Nyce, C. (2007). *Predictive Analytics White Paper*. American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America. Obtenido de <https://www.the-digital-insurer.com/wp-content/uploads/2013/12/78-Predictive-Modeling-White-Paper.pdf>
- Oracle*. (s.f.). Obtenido de <https://www.oracle.com/ar/business-analytics/data-visualization.html>
- Oracle*. (s.f.). Obtenido de <https://www.oracle.com>
- Ose, S. O. (2005). Working conditions, compensation and absenteeism. *Journal of Health Economics*, 161–188.
- Pandas*. (s.f.). Obtenido de https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html
- Peralta, H. O. (2014). GESTIÓN DEL TALENTO HUMANO Y SU RELACIÓN CON EL DESEMPEÑO LABORAL DEL PERSONAL DE LA MUNICIPALIDAD DISTRITAL DE PACUCHA. Andahuaylas, Perú: UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS. Obtenido de <http://repositorio.unajma.edu.pe/bitstream/handle/123456789/192/03-2015-EPAE-Oscco%20Peralta-gestion%20del%20talento%20humano.pdf?sequence=1&isAllowed=y>
- PowerData. (2013). *Big World, Big Data*. PowerData. Obtenido de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/337030/big-world-big-data>
- Rainardi, V. (2007). Building a Data Warehouse With Examples in SQL Server. En V. Rainardi, *Building a Data Warehouse With Examples in SQL Server*. New York: APRESS. doi:978-1-59059-931-0
- Ramírez, J. H. (2004). Introducción a la Minería de Datos. Madrid: Pearson. doi:84-205-4091-9
- Redeszone*. (s.f.). Obtenido de <https://www.redeszone.net/2016/02/13/postgresql-lanza-una-actualizacion-de-seguridad-critica-para-todas-sus-versiones-con-soporte/>

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Reyes, I. F. (12 de junio de 2009). Obtenido de <http://ignaciofernandez.blogspot.com/2009/06/la-organizacion-como-ser-vivo-cuerpo.html>

Rodrigo, D. M. (2002). Aplicación del análisis cluster para el estudio de la relación Nao-Precipitaciones de invierno en el sur de la Península Ibérica. Almería: Universidad de Almería. Obtenido de https://repositorio.aemet.es/bitstream/20.500.11765/9148/1/0030_PU-SA-III-2002-D_MUNOZ.pdf

Russom, P. (31 de Diciembre de 2011). *TDWI*. Obtenido de <https://tdwi.org/>

Samela, G. (09 de junio de 2019). IA y datos para medir y gestionar los recursos humanos. *Clarín*. Obtenido de https://www.clarin.com/economia/ia-datos-medir-gestionar-recursos-humanos_0_jtomBdThr.html

Santos, A. A. (2008). *KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW*. IADIS European Conference Data Mining. Obtenido de https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview

Scikit-learn. (s.f.). Obtenido de <https://scikit-learn.org/stable/>

SCIPY. (s.f.). Obtenido de https://web.archive.org/web/20070927015018/http://www.scipy.org/History_of_SciPy/

Sivanandam, S. S. (2006). Introduction to Data Mining and its Applications. En S. Sumathi, *Introduction to Data Mining and its Applications* (pág. 827). New York: Springer. doi:9783540343509

Smyth, D. H. (2001). Principles of Data Mining. Massachusetts, Estados Unidos: The Massachusetts Institute of Technology Press. doi:0-262-08290- x

Sodeman, H. R. (2020). The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources. *Business Horizons*, 85-95. doi:<https://doi.org/10.1016/j.bushor.2019.10.001>

Stone, C. (s.f.). <https://www.cornerstoneondemand.com>. Obtenido de <https://www.cornerstoneondemand.com/glossary/hr-analytics>

Stone, C. (s.f.). <https://www.cornerstoneondemand.com/>. Obtenido de <https://www.cornerstoneondemand.com/glossary/big-data-hr>

- Talend. (s.f.). Obtenido de <https://es.talend.com/>
- Tan, K. S. (2016). *Introduction to Data Mining*. Edinburgh: Pearson. doi:9788131714720
- Trigine, F. G. (2012). *Manual de instrumentos de gestión y desarrollo de las personas en las organizaciones*. Madrid: Ediciones Díaz de Santos. doi:978-84-9969-483-2
- Ulrich, D. (2016). *Recursos Humanos Champions*. Ciudad Autónoma de Buenos Aires: Granica. doi:978-950-641-244-9
- Vola-Luhurs, R. &. (2010). *Recursos Humanos*. Buenos Aires: Pearson. doi:978-987-615-097-2
- Vola-Luhurs, R. (2010). *Lo que sobra es el Talento*. Ciudad Autónoma de Buenos Aires: Cefiro. doi:978-24132-2-4
- Wang, R. (2012). Monday's Musings: Beyond The Three V's of Big Data - Viscosity and Virality. *Forbes*. Obtenido de <https://www.forbes.com/sites/raywang/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/#6db708fe5fff>
- Wasan, S. G. (2001). *A Proposal for Data Mining Management System*. Obtenido de https://www.researchgate.net/publication/2403357_A_Proposal_for_Data_Mining_Management_System
- Wegener, R. (2013). *The value of Big Data: How Analytics differentiates winners*. Obtenido de Brain and Company: <https://www.bain.com/>
- Wikipedia*. (s.f.). Obtenido de https://en.wikipedia.org/w/index.php?title=Dimensional_modeling&oldid=839460441
- Wikipedia*. (s.f.). Obtenido de <https://es.wikipedia.org/wiki/Python>
- Wikipedia*. (23 de noviembre de 2012). Obtenido de https://es.wikipedia.org/w/index.php?title=Extract,_transform_and_load&oldid=107626881
- Wu, J. (1997). *Statistics = Data Science?* Michigan, EEUU: University of Michigan, Ann Arbor. Obtenido de <https://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf>
- Zambrano, D. M. (08 de junio de 2018). La importancia de la capacitación en el rendimiento del personal administrativo de la Universidad Técnica de Ambato. *La importancia de la capacitación en el rendimiento del personal administrativo de la Universidad Técnica de Ambato*. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442018000300268#B1

“Tecnología *big data* y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio”

Zangaro, D. F. (2020). Big data y people analytics: intimidad y emociones en la gestión de los recursos humanos. *INNOVAR*, 75-87. doi:<https://doi.org/10.15446/innovar.v30n78.90306>

Zytkow, W. K. (2002). *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford University Press.