UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA” DE ICA

FACULTAD DE INGENIERIA DE SISTEMAS

TEMA

“APLICACIÓN DE UN SISTEMA DE TI UTILIZANDO MINERÍA DE DATOS PARA MEJORAR LOS PROCESOS EN LA EMPRESA FERRETERA LA ANGOSTURA REPRESENTACIONES DE LA CIUDAD DE ICA"

PARA OPTAR EL TITULO DE INGENIERO DE SISTEMAS

BACHILLER:

YATACO TASAYCO, Albert

ASESOR: Ing. Percy Hernández Rojas

ICA-PERU

2017

AGRADECIMIENTO

Con todo mi amor, respeto y consideración, a mis padres por su incondicional apoyo y comprensión; a la vida, por darle sentido y razón de ser a mi existencia y a cada uno de mis actos.

Albert

Í N D I C E.

Dedicatorias II

Introducción 1

Resumen 2

CAPÍTULO I: PLANTEAMIENTO METODOLÓGICO 3

1.1. Descripción de la Realidad Problemática 3

1.2. Delimitaciones y Definición del Problema 3

1.2.1 Delimitaciones 3

1.2.2 Definición del Problema 5

1.3 Formulación del Problema. 6

1.4 Objetivo de la Investigación 6

1.5 Hipótesis de la investigación 6

1.6 Variables e Indicadores 6

1.7 Tipo y Nivel de Investigación 7

1.8 Método y Diseño de la Investigación 8

1.9 Justificación e Importancia de la Investigación 10

1.10 Limitaciones 11

1.11 Viabilidad de la Investigación 11

1.12 Técnicas e Instrumentos de Recolección de Información 12

1.13 Cobertura de Estudio 12

CAPÍTULO II: MARCO TEÓRICO 13

2.1 Antecedentes de la Investigación 13

2.2 Marco Histórico 18

2.3 Marco Conceptual 27

Capitulo III: DESCRIPCION DE LA HERRAMIENTA 33

3.1. Análisis de requerimientos 33

3.2. Análisis de procesos 34

3.3. Análisis de estructura de datos 35

3.4. Diseño 38

3.5. Diagramas casos de uso 48

3.6. Diagrama de actividades 49

3.8. Desarrollo del prototipo 50

CAPITULO IV: ANALISIS E INTERPRETACION DE LOS DATOS 55

4.1 Análisis de Resultados 55

4.2. Grupo Experimental (Pre Prueba y Post Prueba) 57

CAPITULO V: CONCLUSIONES Y RECOMENDACIONES 78

5.1. Conclusiones 78

5.2. Recomendaciones 79

REFERENCIAS BIBLIOGRAFICAS 80

ANEXOS 82

INTRODUCCION

El presente trabajo de investigación, titulado " APLICACIÓN DE UN SISTEMA DE TI UTILIZANDO MINERIA DE DATOS PARA MEJORAR EL PROCESO EN LA EMPRESA FERRETERA LA ANGOSTURA REPRESENTACIONES DE LA CIUDAD DE ICA"**,** tiene como propósito principal proponer una solución donde se gestione el conocimiento de tal manera, que se desarrolle la creatividad y la innovación; permitiendo encontrar respuestas viables y sostenibles al problema del bajo nivel de competitividad de este segmento frente a las nuevas exigencias del mercado.

El presente proyecto de tesis, se desarrollaron cinco capítulos:

El primero capitulo: presenta el Planteamiento Metodológico, la formulación del problema, los objetivos, la hipótesis, las variables, la justificación y la importancia de la presente tesis.

El segundo, identifica el Marco Teórico donde se plantea los antecedentes materia de esta investigación. Así como también las bases teóricas y el marco conceptual.

En el tercer capítulo, se describen el desarrollo de la Herramienta a utilizar para el presente proyecto. De tesis

En el cuarto capítulo Los Resultados y la interpretación de los datos.

Finalmente el último capítulo presenta las conclusiones y recomendaciones, y se Adjunta la matriz de consistencia usada en el presente proyecto de tesis

RESUMEN

El presente proyecto de tesis demuestra que la Aplicación de un Sistema de TI Utilizando Minería de Datos mejoro los procesos en la empresa Ferretera La Angostura Representaciones, por cuanto en la tabla estadística descriptiva, se determinó que de las observaciones tomadas se obtuvo que el tiempo en tomar el control de los procesos se dio en una media de 4.54 dado que el tiempo más alto fue de 5,0 mientras que el más bajo fue de 4,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,24.

Asimismo, se obtuvo una media de 1,84. En general, el tiempo más alto obtenido fue de 2,0 mientras que el más bajo fue de 1,5. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,14.

Finalmente, de los resultados de la tabla 08, se dedujo que de las observaciones tomadas se obtiene que el tiempo de registro es una media de 3,60. En general, el tiempo más alto obtenido fue de 4,0 mientras que el más bajo fue de 3,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,25.

CAPÍTULO I: PLANTEAMIENTO METODOLÓGICO

1.1. Descripción de la Realidad Problemática

La investigación llevada a cabo para la realización del proyecto se basó en un estudio descriptivo, se revisó exhaustivamente la literatura existente relacionada, con el fin de obtener la información necesaria para dar inicio al proyecto. Durante la investigación se recabo cuantiosa información acerca de la empresa ferretera la Angostura representaciones de la ciudad de Ica, como así también acerca de la Inteligencia de Negocios, y sobre todo las herramientas a utilizar para el diseño y desarrollo del proyecto, las cuales sirvieron para llevar adelante la tesis.

Cabe destacar que para la realización del presente proyecto se analizaron detalladamente todos los requerimientos y en base a los conocimientos básicos, más las investigaciones realizadas, junto con los relevamientos del sistema actual que se posee, además del *hardware* con el que cuentan las empresas, se definió la solución que mejor se adecua al contexto actual, para luego definir las etapas necesarias para llevar adelante la realización de la tesis.

1.2. Delimitaciones y Definición del Problema

1.2.1 Delimitaciones

a) Delimitación Espacial.

El presente proyecto de tesis se llevara a cabo en las instalaciones de la Empresa Ferretera La Angostura Representaciones de la ciudad de Ica, sito en Carretera Panamericana Sur No 300 Ica.

b) Delimitación Temporal.

El presente proyecto de tesis se desarrolló en 2 etapas:

1ra Etapa: Consistió en el desarrollo del plan de tesis, para ello se recopilo información referente al tema, todo esto fue desarrollado durante los meses de Abril a Julio del año 2015.

2da Etapa: En esta etapa del proyecto se desarrolló la borrador de tesis realizando un estudio de lo que se va a realizar, todo esto fue desarrollado durante los meses de Agosto a Diciembre del 2015.

c) Delimitación Social.

Para el desarrollo del presente proyecto de tesis intervinieron los siguientes actores:

El Investigador

El Asesor

El Gerente de la Empresa

El personal Administrativo

d) Delimitación Conceptual

***Data Warehouse*:** Colección de datos orientado a un ámbito determinado (empresa, organización, etc.) sus características son no volátil, integrado y variable en el tiempo.

Data Marts: Es una versión especial de *data warehouse*, son subconjuntos de datos con la finalidad de ayudar a la toma de decisiones dentro de un área específica en la organización.

Minería de Datos: La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten una guía para la toma de decisiones.

Inteligencia de Negocios: Es un conjunto de tecnologías que tienen un fin común con el principal objetivo que es la de servir como soporte para la toma de decisiones.

1.2.2 Definición del Problema

La Empresa Ferretera Comercial la Angostura representaciones de la Ciudad de Ica, dedicada al rubro de venta de artículos de ferretería en general no cuenta con una base de datos en las cuales puedan registrar toda la información, es por ello que se hace necesario que se tenga que utilizar las tecnologías de información para poder competir con otras empresas dedicadas a este rubro, otro de los problemas que se presente es que el personal de la empresa no se encuentra capacitado para poder manejar estos sistemas.

El reto de este proyecto consiste en brindar un prototipo de solución de Inteligencia de Negocios capaz de transformar los datos en información útil, de manera que los gerentes puedan utilizar dicha información para incrementar la rentabilidad de las empresas. Brindándoles un soporte en el cual respaldar la toma de decisiones estratégicas.

1.3 Formulación del Problema.

¿De qué manera la aplicación de un sistema TI mejora los procesos en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica.?

1.4 Objetivo de la Investigación

Objetivo General.

Implementar un de Sistema de TI utilizando Minería de Datos para mejorar los proceso en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica..

1.5 Hipótesis de la investigación.

La Implementación un de Sistema de TI utilizando Minería de Datos mejorara los proceso en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica.

1.6 Variables e Indicadores.

1.6.1 Variable Independiente

Sistema de TI

Indicadores

Indicadores:	Índices:
X_1=Aplicación de un Sistema de control de asistencia	[No…Si]

1.6.2 Variable Dependiente

Mejora de los procesos

Indicadores

Y_1=Tiempo de entrada del personal [4...5seg] Y_2=Tiempo de salida del personal [4…5seg] Y_3=Tiempo de registro de del Personal [3…4seg]

1.7 Tipo y Nivel de Investigación.

1.7.1 Tipo de Investigación.

Investigación Aplicada. Porque depende de los descubrimientos y avances de la investigación básica, usa conocimientos previamente existentes y se enriquece con ellos pero se caracteriza por su interés en la aplicación, utilización y consecuencias prácticas de los conocimientos. La investigación aplicada busca el conocer para hacer, para actuar, para construir, para modificar[1].

1.7.2 Nivel de Investigación.

De acuerdo a la naturaleza del estudio de la investigación, reúne por su nivel las características de un estudio descriptivo y correlacional.

- **Descriptivo**. Porque tiene como objetivo determinar las características del problema, miden y evalúan diversos

[1] Tevni Grajales, En: http://tgrajales.net/investipos.pdf, 27 de marzo 2000

aspectos, dimensiones o componentes del fenómeno o fenómenos a investigar, en un estudio descriptivo se selecciona una serie de cuestiones y se mide cada una de ellas independientemente, para así y valga la redundancia describir lo que se investiga.

- **Correlacional.** Porque este tipo de estudios tienen como propósito medir el grado de relación que exista entre dos o más conceptos o variables, la utilidad y el propósito principal de los estudios correlaciónales es saber cómo se puede comportar un concepto o variable conociendo el comportamiento de otra u otras variables relacionadas.[2]

1.8 Método y Diseño de la Investigación.

1.8.1 Método de Investigación.

Método Científico. Porque se realiza con el fin de exponer y confirmar teorías en otras palabras, el método científico tiene su base y postura sobre la teoría mecanicista (todo es considerado como una máquina, y para entender el todo debemos descomponerlo en partes pequeñas que permitan estudiar, analizar y comprender sus nexos, interdependencia y conexiones entre el todo y sus partes)[3].

1.8.2 Diseño de investigación.

[2] Bedoya Sanchez, Enrique Osvaldo, Metodologia de la Investigacion Peru Ca. 3 En http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/empre/bedoya_se/cap3.pdf

[3] Aula Fácil, En: http://www.aulafacil.com/cursosenviados/metodo-cientifico.pdf

Este diseño muestra a un grupo se le aplica una prueba previa al estímulo o tratamiento experimental, después se le administre el tratamiento y finalmente se le aplica una prueba posterior al tratamiento. Hay un punto de referencia inicial para ver qué nivel tenía el grupo en las variables dependientes antes del estímulo. Es decir, hay un seguimiento del grupo[4].

G O_1 X O_2

G=Grupo Experimental

O_1= Datos de los indicadores de la variable dependiente sin el sistema de control de asistencia.

X= Sistema de control de asistencia

O_2= Datos de los indicadores de las misma variable dependiente con el sistema de control de asistencia.

[4] HERNÁNDEZ SAMPIERI y otros (1994). Metodología de la investigación, México, Mc Graw Hill, Cap. 4

1.9 Justificación e Importancia de la Investigación

1.9.1 Justificación.

El proyecto de tesis se justifica porque al implementar este sistema de TI beneficiara a la empresa dado que los trabajadores tendrán que ser más puntuales tanto en en ingreso y la salida y los horarios de trabajo se cumplirán tanto para beneficio de la empresa como para los trabajadores y así no se vean afectados en sus remuneraciones como consecuencia de esta desinformación.

Asimismo, permitirá mejorar la toma de decisiones que se realizan en el área e incrementar la capacidad de procesamiento de información; mejorando el rendimiento laboral, sin sobrecargas en funciones y sin repetir procesos innecesarios.

Además se justifica el costo porque la pérdida de información durante su traslado del campo a la oficina puede ocasionar gastos de recuperación de información que podrían ser prevenidos.

1.9.2 Importancia

La importancia de este proyecto radica en el impacto que generara a la empresa porque abrirá nuevos caminos para empresas dentro de la región de Ica que presenten situaciones similares, gracias a esta investigación se resolverá problemas que presentan estas empresas, también habla de una nueva manera de ver los procesos en una etapa con el uso de TI utilizando

minería de datos, y los beneficios que dejara en el cual este sistema mejoraría el nivel de la empresa.

1.10 Limitaciones

Las principales limitaciones que se tuvo para poder desarrollar el presente proyecto de tesis es la falta de información del personal por cuanto se piensa que utilizando este sistema se verán sometidos a un control estricto en el cumplimiento de sus obligaciones. Otra de las limitaciones que se tuvo es la falta de conocimiento del personal que labora en el área informática en el uso de TI y el poco conocimiento de lo que es una minería de datos.

1.11 Viabilidad de la Investigación.

El proyecto de tesis se considera viable por cuanto el investigador cuenta con los medios necesarios para poder realizar la investigación debido a que no se requiere de mucha inversión económica, sino de conocimiento sobre la herramienta que se va a utilizar. Asimismo, el proyecto investigación es viable económicamente, porque los gastos básicos en pasajes, movilidad local, comunicaciones por teléfono, y otros gastos administrativos son asumidos por el investigador en todo el proceso de desarrollo de la investigación. Finalmente la presente investigación cuenta con la disponibilidad de las herramientas y del software para poder implementar.

1.12 Técnicas e Instrumentos de Recolección de Información.

1.12.1 Técnicas:

Las técnicas de recolección de datos usadas en la presente investigación son:

a) Entrevistas

b) Análisis documental

c) Observación de campo

1.12.2 Instrumentos

Guía de entrevistas

Fichas documentales

Guías de Observación

1.13 Cobertura de Estudio.

1.13.1 Universo y Población:

El universo será todos los registros de envió de procesos de durante el lapso de 2 meses. Consistente en 80 procesos.

1.12.2 Muestra

Se considerara como muestra a 57 procesos de la empresa.

CAPÍTULO II: MARCO TEÓRICO

2.1 Antecedentes de la Investigación

TITULO: Confrontación de 2 Técnicas de minería de datos aplicadas a un dominio especifico.

AUTOR: Mario Gálvez, Fabricio Martínez

AÑO: 2004 - Colombia

RESUMEN:

Por ser la minería de datos una herramienta que dia a dia cobra mayor importancia en diversos entornos de nuestra sociedad, desde el académico, pasando por el comercial, hasta el de investigacion y desarrollo, este proyecto de investigacion es importante pues presenta una oportunidad a quienes deseen hacer uso de esta herramienta que ofrece tecnología.

Este proyecto de investigacion pretende mostrar las principales diferencias entre dos técnicas de Minería de datos, Arboles de decisión y reglas de asociación.

Para esto se tiene a disposición los siguientes recursos. Primero un conjunto de datos con información histórica de estudiantes de la Pontificia Universidad Javeriana y de su desempeño académico detallado. Segundo, una herramienta de Minería de Datos llamada WEKA desarrollada en un entorno académico que permitirá la ejecución de los algoritmos seleccionados y la posterior comparación de los resultados obtenidos para cada uno de ellos.

TITULO: APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA MEJORAR EL PROCESO DE CONTROL DE GESTIÓN EN ENTEL.

AUTOR: CLEMENTE ANTONIO MARTÍNEZ ÁLVAREZ

AÑO: 2012

Resumen

El presente estudio se enfoca en el análisis de ingresos no percibidos en la empresa de telecomunicaciones ENTEL, dentro del proceso de provisión de servicios privados de telefonía, internet y comunicaciones a los clientes de mercados no residenciales. Dicho proceso es controlado mediante indicadores de gestión, obtenidos a partir de la transformación de datos de clientes y servicios. La generación de estos indicadores demanda tiempo y esfuerzo por parte de los analistas de la empresa, debido a que es un trabajo realizado en forma manual. El objetivo principal de esta tesis consiste en reducir el tiempo de cálculo de los indicadores de servicios privados de ENTEL, para lo cual se aplicó modelamiento multidimensional, técnicas de minería de datos y automatización de procesos, y de este modo poder entregar información más oportunamente. La metodología de este trabajo se basa principalmente en las etapas del proceso conocido como Knowledge Discovery in Databases (KDD), implementadas de acuerdo a la metodología CRISP-DM, la cual es usada para el desarrollo de proyectos de minería de datos. Para comenzar, se hizo un levantamiento de las métricas existentes para la gestión de la provisión de servicios. Luego, se seleccionaron y procesaron las fuentes de datos para el estudio de forma automática, almacenando las variables más relevantes en un repositorio multidimensional (Data Mart), reduciendo drásticamente el tiempo de

cálculo de indicadores y liberando recursos humanos altamente calificados. A partir de lo anterior, se aplicaron técnicas de clustering para obtener grupos de elementos con datos de clientes y servicios cuyas características fueran similares, asociándoles un valor de precio según información histórica de consumo. Por último, se generó un modelo de clasificación que asignara, de acuerdo a una medida de similitud, elementos que no habían sido facturados a los grupos previamente definidos, y de esta manera estimar los ingresos no percibidos. Con ayuda de minería de datos se logró diseñar nuevas métricas para el proceso e identificar a los clientes y servicios más críticos, lo que permite llegar a valores más exactos de los ingresos perdidos en cada segmento, y aplicar estrategias diferenciadas para hacer el cobro a sus clientes. El trabajo realizado permitió una reducción del tiempo de obtención de los indicadores en un 78%, pasando de un total de 14 horas inicialmente a tan sólo 3 horas, logrando además estimar los ingresos perdidos mensualmente por servicios no facturados en un monto de MM $ 210, con un error de la estimación menor al 5%. Se espera que, con ayuda de este estudio, la empresa pueda tomar decisiones informadas y mejorar su capacidad de control del proceso de provisión de servicios privados, con el fin de regularizar su flujo de ingreso mensual.

TITULO: ESTUDIO DE MINERÍA DE DATOS PARA LA INFORMACIÓN DE MORTALIDAD EN MÉXICO

AUTOR: Paulina Galván Castro Alejandra Meza Mendoza

AÑO: 2012

Resumen

La Minería de Datos es usada con el fin de extraer patrones, de describir tendencias, predecir conceptos y en general de obtener información útil, novedosa y principalmente desconocida, que generalmente se encuentra en la información que tenemos a nuestro alrededor y que esta almacenada de forma heterogénea. Es de suma importancia que se entienda a la perfección que es la minería de datos, que nos ofrece, cuales son las herramientas que necesitamos, como se trabaja con ella y muchísimas cosas más por lo cual este trabajo consta de 3 capítulos teóricos para después en un cuarto capítulo apliquemos toda esta teoría en casos reales y prácticos. Se tiene la situación de que día a día hay una gran cantidad de personas que fallecen por diferentes razones. Para atender esta situación se han realizado diferentes estudios estadísticos en donde tratamos de ver casos muy precisos, para ser concretos se manejan 4 grandes temas de la mortalidad en México. La Minería de Datos no ayudara a darle un nuevo enfoque a esos estudios estadísticos con el fin de encontrar patrones de comportamiento de la ocupación más peligrosa, de la institución pública o privada de salud menos eficiente, de la violencia familiar ó del embarazo de alto riego. Sera bueno comentar un poco más de cada una de los temas que se analizarán posteriormente. Por mencionar un tema para la ocupación peligrosa, en principio se busca en número de personas que

hayan fallecido en su trabajo y esto para las diferentes ocupaciones, una vez obteniendo este resultado escogemos un algoritmo de minería de datos que nos arroje las posibles ocupaciones de dicho problema. Para la institución menos eficiente se busca obtener el número de personas que murieron por institución en el año 2000 y el 2009, en este caso se escogerá otro Introducción 10 algoritmo de minería de datos que nos muestre que institución es la menos eficiente en cada año.

TITULO: MINER´IA DE DATOS APLICADA A TEOR´IA DE JUEGOS

AUTOR: CRISTIAN DANILO BRAVO ROM ´ AN

AÑO: 2008

Resumen

En este trabajo se busca estudiar la factibilidad de desarrollar un modelo integrado que utilice de manera conjunta minería de datos y teoría de juegos, destacando los aspectos teóricos relevantes dentro de cada técnica que faciliten tal modelo sinérgico, para luego desarrollar una aplicación práctica de esta nueva aproximación en la industria financiera. El problema que inspira la aplicación práctica nace de la clasificación de clientes acorde a su propensión de compra, donde compiten cuatro empresas para capturarlo utilizando el precio que cobran y campañas de marketing. Es relevante ˜ utilizar una técnica como la minería de datos para abordar el problema, pues encuentra patrones relevantes para estos fenómenos en base a la explotación de grandes bases de datos, considerando las variables asociadas a los clientes y sus características. Sin perjuicio de lo anterior, este modelo fue complementado de manera novedosa con

variables del mercado, como los precios que observa el cliente y las campanas comerciales dirigidas a él, las que resultaron de importante capacidad ´ discriminante, siendo su estudio un aporte de esta tesis. Se utilizó la metodología de descubrimiento de conocimiento en bases de datos (KDD) en la construcción de los modelos de minería de datos, incorporando modificaciones para potenciar el contacto con el modelador. En cuanto a las técnicas particulares, se utilizaron Support Vector ´ Machines (SVMs), potente técnica de basada en el aprendizaje estadístico, y un modelo hibrido desarrollado en el marco de esta tesis que complementa estos resultados utilizando redes neuronales artificiales (ANNs), potente técnica en cuanto a su capacidad de aproximar funciones. Los resultados indican que el método que utiliza solo SVMs entrega mayores capacidades en la predicción, mientras el modelo de SVMs y ANNs posee ventajas importantes en tiempos de prueba y replicabilidad. No menos importante es el modelo de teoría de juegos creado, que considera el comportamiento de clientes a partir de los modelos de minera de datos y que incorpora también la estructura de ´ costos de las empresas competidoras y la eficiencia al competir utilizando el precio como variable estratégica. La intención fundamental es determinar los factores que determinan los costos de los ´ competidores y su capacidad de respuesta frente a las condiciones de mercado y demanda que enfrenten. Este modelo integrado entrega resultados estratégicos que utilizan, nuevamente como un aporte novedoso, funciones agregadas que nacen a partir de los modelos atómicos de minería de datos. Los resultados obtenidos reflejan fuentes de oportunidad para aumentar participación de mercado en base a

debilidades de los competidores, permite utilizar una medida cuantitativa de la eficiencia de la empresa y de los competidores y permite revisar el efecto de acciones comerciales en los clientes. En conclusión, las sinergias causadas por el uso de ambas técnicas en conjunto es ´ considerable.

2.2 Marco Histórico

2.2.1. El valor de la información

En la actualidad, cuya característica más importante, se basa en un crecimiento a gran escala de las tecnologías de la información y las telecomunicaciones, los activos más valiosos de una empresa pasan a ser los conocimientos, habilidades, valores y actitudes de sus empleados.

Partiendo de la premisa de que el conocimiento sobre un tema determinado y su optima utilización, se convierte en un factor determinante para el éxito en el mundo empresarial. El capital intelectual ha desplazado a los factores tradicionales, como ser la producción, el capital, la tierra ,el trabajo, como principal elemento a la hora de generar valor económico para la empresa, tal como se desprende del planteamiento de Cohen y Asín (2.000).

Todas las compañías de alguna manera han adoptado un modelo del mundo de negocios sustentado en la información, como por ejemplo: ¿qué factores influyen en la compra y la demanda?, ¿cómo hallar las oportunidades de negocio?, ¿existe directa relación entre la calidad del producto y la demanda de los

clientes? A medida que la exactitud de esta información crece, la capacidad de la empresa por competir se incrementa en forma proporcional.

La información puede transformarse en conocimiento tácito o explícito. Se entiende por conocimiento tácito aquel que poseen las personas producto de la experiencia adquirida, los estudios y la educación; mientras que los conocimientos explícitos son aquellos almacenados en bases de datos. Un ejemplo claro es aquel que se produce cuando una persona cambia de empleo, esta se está llevando consigo información, conocimientos y está ofertando su capacidad intelectual por un mayor precio; en el ámbito laboral, la experiencia y la educación son factores claves para aumentar el costo de la fuerza del trabajo intelectual. Así mismo, el hecho de que un sistema que maneja información, eventualmente falle, indefectiblemente generará pérdidas a la empresa.

Es una clara tendencia que las empresas están apostando fuertemente por la tecnología y las personas, para que en conjunto tengan un conocimiento suficiente que logre acercar la visión interna de ambos a la realidad exterior, en la misma medida que la brecha entre la visión interna y la realidad exterior disminuye, se consigue tomar decisiones más acertadas y en menos tiempo. Lo que busca esta tendencia es acercar lo máximo posible el mundo real a la visión interna para generar mayores ganancias, convirtiendo la información en utilidad y darle un valor a la

información.

2.2.2. Necesidad de información y conocimiento en la empresa

Desde el mismo instante en que las empresas iniciaron el proceso de acumular los datos de sus operaciones en medios de almacenamiento físico, y de esa forma conseguir una mejor administración y control de dicha información, ha surgido la necesidad de utilizarla para entender las necesidades particulares del negocio. En un mercado altamente competitivo, donde muchas marcas ofrecen productos similares, los clientes tienen una amplia gama de posibilidades para la elección del producto. Para obtener una porción del mercado, es indispensable para las empresas, tener la suficiente capacidad de transformar la información acumulada en conocimiento.

La necesidad de información en las empresas no surge de un día para el otro, el propósito de almacenar los datos radica en su utilización en algún momento, cuando así se requiera, caso contrario, cualquier dato de control sería desechado instantáneamente. Lo que si surge súbitamente, es la imperiosa necesidad de dar respuesta rápida a los requerimientos de información para la toma de decisiones y ayudar a mejorar de alguna manera los procesos internos de negocio.

2.2.3. La información que las empresas necesitan

En la actualidad las organizaciones demandan información en los niveles donde anteriormente la administración se basaba meramente en la intuición y el sentido común para la toma de

decisiones. Los mercados dinámicos obligan a las empresas a que la información estratégica esté disponible en las computadoras de los directivos y/o gerentes, esta práctica se ha generalizado principalmente motivada por la mayor utilidad que se obtiene de la información compartida.

Hoy en día la información está presente en todos los niveles de la organización con propósitos diferentes (comunicación, control, administración, evaluación, etc.) ayudando a la correcta y oportuna toma de decisiones desde el nivel gerencial, hasta el nivel operativo de la organización. Las empresas han entendido que a pesar de que los niveles directivos tienen una gran responsabilidad al tomar decisiones, por el peso que conllevan las mismas, existen también personas que toman decisiones y a pesar de que éstas no tienen un impacto global, deben ser de igual manera correctas y oportunas.

Directores, gerentes, jefes y todos aquellos que toman decisiones deben contar con la suficiente información para respaldar su trabajo cotidiano, la posición que ocupen en la pirámide organizacional se torna secundaria cuando la mirada es hacia la gestión de los procesos y como así también los puestos que tienen cierta relación y dependencia entre sí.

De modo general en una pirámide organizacional, los requerimientos informativos se dividen en tres niveles:

- Información Estratégica
- Información Táctica

- Información Técnico Operacional.

Información Estratégica: está pensada principalmente para ayudar a la toma de decisiones de las áreas gerenciales para alcanzar la misión empresarial. Se caracteriza porque son sistemas con poca carga diaria de trabajo y sin una gran cantidad de datos, sin embargo, la información que guarda está relacionada a un contexto cualitativo más que cuantitativo, que puede indicar cómo evolucionará en el futuro, el criterio es distinto, pero sobre todo es distinta su delimitación. Se asocia esta información a los niveles ejecutivos de las empresas. Es importante señalar que la información estratégica se nutre de grandes cantidades de datos de áreas relacionadas y no se enfoca puntualmente en una sola dirección, de ahí que las decisiones que puedan ser tomadas tienen un impacto directo en toda la organización.

Información Táctica: este tipo de información es la que soporta la coordinación de actividades y el nivel operativo de la estrategia, es decir, se analizan opciones y se trazan rutas posibles para alcanzar la estrategia definida por la dirección de la empresa. Se facilita la gestión independiente de la información por parte de los niveles intermedios de la organización. Este tipo de información es extraída puntualmente de un área o sección de la organización, por lo que su alcance es local y se asocia habitualmente a las gerencias.

Información Técnico Operacional: hace referencia a las operaciones diarias que son efectuadas de modo rutinario en las

corporaciones mediante la transacción masiva de datos y sistemas transaccionales. Las cargas son cotidianas y soportan la actividad de la empresa día tras día (contabilidad, facturación, almacén, presupuesto, etc.). Generalmente se asocia esta información con los jefes de área o las coordinaciones operativas, también llamadas de tercer nivel.

Podemos considerar entonces factores internos y externos de una empresa y así concluir que los requerimientos en la actualidad se orientan a descubrir y mejorar los beneficios de toda la cadena corporativa. Dichos requerimientos se ven reflejados en el interés por tener a mano los indicadores que arrojen información concreta y clave para determinada área de la empresa, y en el menor tiempo posible. La clara tendencia es que las áreas gerenciales necesitan en su mesa de trabajo, la información clave de su empresa; en todos los niveles el requerimiento es parecido, aunque es evidente que tendrá objetivos diferentes.

2.2.4. Las organizaciones y los distintos sistemas de información

Para que una empresa esté completamente automatizada es necesario contar con una gran infraestructura en tecnologías para que soporte todos los sistemas de información. El crecimiento en tecnología puede tener distintos orígenes partiendo desde la implementación, crecimiento, ampliación, mantenimiento, etc. Las necesidades actuales de las empresas han provocado contar con tecnología de la informaciones más sofisticados para responder a cada una de sus peticiones de

información, esto impulsó a que las compañías adquieran distintos tipos de sistemas de informaciones, entre estos sistemas podemos mencionar a los *OnLine Transaction Processing*(en adelante OLTP), Planificación de Recursos Empresariales(en adelante ERP), Sistemas de Soporte para la toma de decisiones (en adelante DSS), Administración de la Relación con Clientes (en adelante CRM), etc.

Estos sistemas siempre están utilizando bases de datos para almacenar la información generada, las mismas se utilizan como soporte para la toma de decisiones en las empresas.

Existen empresas que precisan información de una actividad específica, un ejemplo seria los Sistemas ERP (Planificación de Recursos Empresariales) son sistemas muy complejos y grandes donde un alto porcentaje de su contenido dedica a los procesos de producción, si una empresa se dedica a las bienes raíces, sería ilógico adquirir un sistema de alta complejidad y costoso como las ERP que no va a dar una solución a las necesidades de la empresa. Para empresas como estas, existen desarrollos de soluciones en el mercado comercializado como productos que pueden ser configurados en una organización en particular de acuerdo a sus necesidades, que dan soluciones a requerimientos específicos para cada industria. Hay en el mercado, *software* para distintos tipos de empresas como automotriz, hotelería, comercios, educativos entre otros.

Son distintas las herramientas utilizadas debido a que las actividades de misión crítica que soportan cada una de las organizaciones son diferentes, y por ende también son distintos los tipos de información que puede solicitar un directivo en cada una de las organizaciones, lo cual justifica que existan muchos productos de *software* dedicados a explotar la información de las bases de datos que no tienen características estándares, sino más bien son adaptables según las exigencias. Considerando las necesidades que se presentan en cada actividad.

La información que se genera en una empresa u organización está destinada a responder a diversos tipos de preguntas de los usuarios, de ahí nace la necesidad que existan sistemas de información para requerimientos muy específicos que permitan la recolección y el manejo de los datos. La estructura organizacional de una empresa es un factor importante para determinar la información que comúnmente es requerida por los funcionarios.

Los sistemas de procesamiento de datos (OLTP) utilizan medios de almacenamiento y técnicas para el cargado. Un alto porcentaje de las empresas recurren a los OLTP para guardar grandes cantidades de datos con un tiempo de respuesta corto en los miles de transacciones realizadas cotidianamente, sin embargo, su eficiencia no son las consultas masivas de grandes cantidades de información y mucho menos el análisis de la misma.

La tecnología tuvo que adaptar los medios necesarios para que sean eficientes en el ámbito específico de aplicación, así como para el diseño de estructuras de datos que ordenen la información como se desea, como en las herramientas o *software* que dan soluciones oportunas a los usuarios. Todos los sistemas de información tienen un fin muy particular, y se complementan para mantener de la manera más eficiente una organización; sin embargo, no todos pueden dar solución a las distintas demandas de los usuarios, ya que son diseñados para alguna área específica.

El motivo por el cual existen diferentes sistemas de información es porque las preguntas de los usuarios son muy específicas que no cualquier sistema puede resolver. De hecho la base de datos operacional, que es imprescindible en cualquier organización pero no está organizada para dar respuestas a preguntas globales sino más bien a pequeños grupos de datos. Preguntas que impliquen consultas complejas podrían resolverse en un lapso muy extenso, donde la posibilidad de que la vigencia desaparezca aumenta considerablemente. Es importante recalcar que una base de datos o sistema de información no está diseñada para resolver las necesidades informativas de la organización a nivel macro.

2.3 Marco Conceptual

2.3.1. Definición de Inteligencia de Negocios

La Inteligencia de Negocios cuenta con una diversidad de

interpretaciones como muchos otros términos o conceptos. Su uso es justificado a todo lo que sea considerado como tecnología de información, pero no hay un consenso en lo respecta a su definición[5].Examinando las distintas definiciones queda claro que la Inteligencia de Negocios no consiste en una metodología, sistema, *software* o herramienta en particular, si no es un conjunto de tecnologías que tienen un fin común con el principal objetivo que es la de servir como soporte para la toma de decisiones.

2.3.2. Los usuarios de las soluciones de Inteligencia de Negocios

A continuación se describen los diferentes tipos de usuarios que intervienen en una solución de inteligencia de negocios.

Productores de información:*"[..]Habitualmente son el 20% de los usuarios, que crean informes o modelos utilizando herramientas de escritorio. Donde predominan estadísticos que se valen de herramientas para minería de datos o son creadores de informes que utilizan herramientas para el diseño y/o programación de informes específicos. Regularmente son profesionales del área de sistema de información o usuarios muy avanzados con capacidades de comprender la información y la informática"(CANO, J 2.007).*

Los consumidores de información*: "[…] Son usuarios no habituales que regularmente consultan informes para la toma de decisiones,*

[5] (HOWARD DRESNER, H 1.989)

pero no acceden a los números o hacen análisis detallados diariamente. Los usuarios no habituales son directivos, gestores, responsables, colaboradores y usuarios externos. Este numeroso grupo está bien servido con cuadros de mando con análisis guiados, informes interactivos (por ejemplo: OLAP, informes parametrizados, vinculados,…) e informes de gestión estandarizados. La mayoría de estas herramientas proveen ahora acceso vía web para promover el acceso desde cualquier lugar y facilitar el uso y minimizar los costes de administración y mantenimiento"(CANO, J 2.007).

2.3.3. Historia de la Inteligencia de Negocios

En octubre de 1.958 H.P. Luhn de IBM, escribió un artículo llamado *Business Intelligence System* donde describe las características que debe tener un sistema de este tipo, en muchos aspectos de lo que escribió Luhn tiempo atrás, si se realiza un paralelismo de lo que se entiende hoy en día por inteligencia de negocios no varió mucho.

Edgar Frank Codd presenta el concepto de las bases de datos en el año 1.969, un año más tarde se desarrollaron las primeras base de datos con sus interfaz empresariales, estas aplicaciones, facilitan la entrada de datos en los sistemas, haciendo que aumente la información disponible, pero como el acceso a la información era de alta complejidad y difícil de acceder a las mismas no fue una solución completa. Otro de los avances

llegaron en el año 1.980 con la creación del concepto del *data warehouse* por Ralph Kimball y Bill Inmon, con ellas aparecieron los primeros sistemas de reportes, con todo esto la solución seguía siendo compleja y funcionalmente pobre, se contaba con potentes sistemas de bases de datos pero no existían aplicaciones que facilitaran su explotación.

En 1.989 Howard Dresner difundió el término de *Business Intelligence* escrito por H.P. Luhn en el año 1.958, en la década de los 90 llegaba la *Business Intelligence*1.0 y con ello la multiplicación de aplicaciones de Inteligencia de Negocios, logrando facilitar el acceso a la información considerablemente pero empeoraron el problema que se quería resolver ya que seguían apareciendo múltiples soluciones sin poder consolidarse. Con la llegada de *Business Intelligence* 2.0 en el año 2.000 se logró consolidar las aplicaciones en una pocas plataformas a partir de ahí las herramientas empezaron a dar soluciones reales a las empresas u organizaciones.

2.3.4. Características de la Inteligencia de Negocios

Información: el propósito de la Inteligencia de Negocios es proveer de información al usuario final para dar soporte a la toma decisiones, por ende la información es la esencia de la Inteligencia de Negocios, estas pueden originarse desde las bases de datos operacionales, como también de arquitecturas *data mart* y *data warehouse* diseñadas específicamente para el análisis.

El usuario requiere de información para apoyarse en el momento de tomar decisiones, pasando en segundo plano de dónde provenga esta, pudiendo ser de una fuente primitiva o derivada, para lo cual la inteligencia de negocios utiliza o crea fuentes de datos interna o externa con el fin de utilizarla como materia prima para lograr su objetivo.

Apoyo a la toma de decisiones: básicamente consiste en organizar y presentar los datos relevantes para que sirvan como soporte a la hora de tomar decisiones. Esto implica la utilización de tecnologías, técnicas de análisis y todo lo que sea necesario con el fin de obtener solamente aquella información relevante y útil.

Orientación al usuario final:un factor fundamentalque tuvo su incidencia en la tecnología de la Inteligencia de Negocios para explotar información, fue que el usuario final no contaba con conocimientos y técnicas que le permita acceder de una manera

sencilla y directa a los datos almacenados en los sistemas operacionales, ya que casi siempre necesitaba de ayuda de informáticos para acceder a la información, con la Inteligencia de Negocios, se elimina la dependencia de terceras personas para el acceso a los datos e información, siendo esta una herramienta sencilla y preparada para que interactúe directamente con el usuario final sin intermediarios.

2.3.5. Componentes de una solución de Inteligencias de Negocios

Teniendo en cuenta el esquema mencionado por Cano (2.007) la solución de Inteligencia de Negocios está compuesta por las fuentes de información, proceso de ETL (extracción, transformación y limpieza de datos) *datawarehouse* y motor OLAP como se puede apreciar en la figura 2.

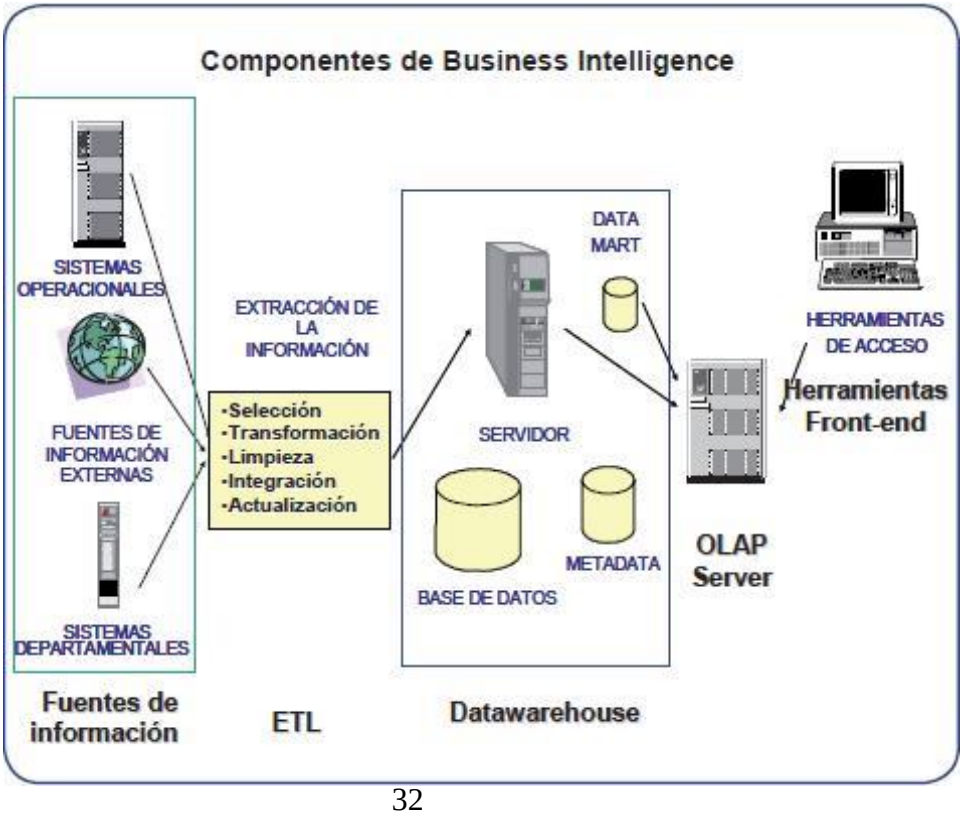


Figura 2: Componentes de Inteligencia de Negocios

Los componentes son:

- Fuentes de información: generalmente son los datos generados por sistemas operacionales, los cuales se utilizan para alimentar de información el *data warehouse*.

- Proceso ETL: de extracción, transformación y carga de los datos en el *datawarehouse*. Los datos antes de almacenarse en un *data warehouse*,pasan por procesos de filtrado, limpieza, trasformación y redefinición.

- El *datawarehouse*: en él se almacenan los datos de una manera que optimice su flexibilidad, facilidad de acceso y administración, en donde los datos están estructurados para generar informes que ayuden a la toma decisiones.

- El motor OLAP: es el que proporciona la capacidad de realizar cálculos, análisis, pronósticos, consultas en grandes volúmenes de datos.

Capitulo III: DESCRIPCION DE LA HERRAMIENTA

3.1. Análisis de requerimientos

A partir de las entrevistas realizadas a los miembros del directorio y a los gerentes de cada área de ambas empresas, se detectaron las dificultades con las que se encuentran en el momento de obtener información de vital importancia para la toma de decisiones, lo cual lleva a concluir en la necesidad de la implementación de una solución de Inteligencia de Negocios, que sea capaz de brindar reportes, por empresas y también un consolidado de ambas empresas, como así también un análisis de los datos históricos para predecir el probable comportamiento a futuro de las ventas.

De igual manera, se tomó la decisión de proporcionar a través de la implementación del prototipo, los reportes a ser utilizados por los departamentos de Venta y Marketing, esto se decidió en primera instancia, por pedido de los miembros del directorio, ya que consideran que son los departamentos a los que más beneficiará la solución de Inteligencia de Negocios, por otro lado, ambos departamentos se nutren del mismo tipo de información, por lo cual se optimiza la implementación del prototipo, debiéndose generar un solo grupo de reportes para ambos departamentos. A continuación se detallan los tipos de reportes analizados en detalle.

- Venta por Campaña: visualiza gráficamente los resultados de una o más campañas en función a tres variables (facturación, unidades y pedidos), de una o ambas empresas.

- Venta por División: visualiza gráficamente los resultados obtenidos por cada una de las divisiones, esto se puede generar para una o más campañas, en función a tres variables (facturación, unidades y pedidos), de una o ambas empresas.

- Venta por Departamento: visualiza gráficamente los resultados obtenidos en cada uno de los departamentos, esto se puede generar para una o más campañas, en función a tres variables (facturación, unidades y pedidos), de una o ambas empresas.

- Venta por *Quarter*: visualiza gráficamente los resultados de cada *quarter* (cada *quarter* se compone de 5 campañas, 1-5 1erQ, 6-10 2doQ, 11-15 3erQ y 16-20 4toQ) en función a tres variables (facturación, unidades y pedidos), de una o ambas empresas.

- *Top Salers*: Visualiza gráficamente los productos más vendedores de una o rango de campañas, en función a las dos variables (facturación y unidades), se podrá generar: consolidado general, por categoría, por división, por departamento, por promotora.

- Ventas por Categoría: visualiza gráficamente el resultado obtenido con cada una de las categorías.

3.2. Análisis de procesos

Contando con todos los procesos relevados, se procedió a la clasificación de los mismos en función a los requerimientos de la

solución, estos se clasificaron de acuerdo a la organización departamental del Grupo, la misma se detalla a continuación:

Luego de llevar a cabo estos análisis, se llegó a la conclusión que entorno a estos dos procesos seleccionados, se llevara adelante el diseño del proyecto.

3.3. Análisis de estructura de datos

Se analizaron minuciosamente las estructuras de datos provenientes de cada sistema de información, lo cual permitió definir qué tablas y datos serán necesarios para la realización del prototipo.

3.3.1. Análisis de herramientas de Inteligencia de Negocios

Con todos los datos relevados sobre las herramientas de Inteligencia de Negocios se realizó el análisis para la selección de la herramienta en la cual se desarrollará este trabajo de tesis.

En la tabla 5 se presenta un cuadro comparativo de las distintas herramientas mencionadas en el relevamiento.

Herramienta	Permite ETL	Reportes	Minería de Datos	Open Source
Oracle	SI	SI	SI	NO
Jasper	SI	SI	NO	SI
Pentaho	SI	SI	SI	SI
MicroStrategy	SI	SI	SI	NO

Tabla 5: Comparativa herramientas de Inteligencia de Negocios

Después de analizar minuciosamente cada componente las opciones de suite de Inteligencia de Negocios, se llegó a la

conclusión de que Pentaho BI Suite es la que mejor se adecua al proyecto teniendo en cuenta que con este trabajo se busca proporcionar una solución de Inteligencia de Negocios basada en *software* libre, además de la robustez de la herramienta y su larga trayectoria, sumado al amplio soporte en su versión de la comunidad y la facilidad de uso de todos sus componentes.

3.3.2. Análisis de sistemas gestores de bases de datos

Con todos los datos relevados sobre los sistemas gestores de bases de datos, se realizó el análisis para la selección del SGBD que se utilizará para el desarrollo del *data warehouse* necesario para la concreción del proyecto.

A continuación se presenta la tabla 6 donde se tiene un cuadro comparativo de los SGBD seleccionados y mencionados en el relevamiento

Característica	MySQL	SQL Server
Costo	Libre de Pago	De Pago
Open Source	Si	No
Plataformas	Linux, Windows, Mac y otras	Windows
Límite de tamaño de BD	Limitado por el Sistema Operativo	Limitado por el Sistema Operativo
Transacciones	Si	Si
Posibilidad de elegir diferentes formas de almacenamiento	Si	No
Claves Foráneas	Depende del motor	Si
Vistas	Si	Si
Procedimientos Almacenados	Si	Si
Triggers	Si	Si

Cursores	Si	Si
Subconsultas	Si	Si
Funciones definidas por el Usuario	Si	Si
Multiusuario	Si	Si

Tabla 6: Comparativa de SGBD

Finalmente, después de realizar un análisis detallado de las dos opciones seleccionadas, se optó por utilizar MySQL Server como SGBD para el desarrollo del *data warehouse*, basados en que no existen diferencias sustanciales entre las dos opciones, en función a lo que se necesita para la realización de esta tesis, como así también y principalmente teniendo en cuenta que se busca desarrollar una solución de Inteligencia de Negocios basada en *software* libre.

3.3.3. Análisis de factibilidad económica

A continuación se presenta una tabla donde pueden apreciarse los costos necesarios para llevar adelante este proyecto, en la misma se detallan los conceptos y los respectivos montos.

Concepto	Cant.	P.U u$s	Total u$s
Analista de Sistemas	1	0.0$s	0.0$s
Desarrollador	2	0.0$s	0.0$s
Especialista en BI	1	1.500$s	4.500$s
Servidor Dell PE R520 Rackeable	1	5.397$s	5.397$s
Licencia S.O. Servidor	1	0.0$s	0.0$s
Licencia Herramienta Desarrollo BI	1	0.0$s	0.0$s
Licencia Servidor BI	1	0.0$s	0.0$s
Total		6.647$s	6.647$s

Tabla 9: Cuadro de costos

Tal como puede observarse en la tabla anterior, el costo total que implica la implementación de este proyecto es por demás accesible, teniendo en cuenta la relación entre el costo mencionado y el volumen de ventas que manejan la empresa. Esto es favorecido también, gracias a la posibilidad de contar con los servicios de los profesionales de T.I, como ser Analistas de Sistemas y Desarrolladores, necesitándose solo la tutoría de un especialista en soluciones de Inteligencia de Negocios, para brindar un acompañamiento que asegure el éxito del proyecto. Así también recalcar una vez más, el ahorro que representa la utilización de herramientas con licencia de *software* libre.

3.4. Diseño

En este punto se describirá el diseño del *data warehouse*, ETL, reportes, cubos OLAP, cuadros de mandos, y los distintos diagramas que componen el prototipo.

Se utilizarán cuatro ejemplos de informes, un reporte dinámico que permite la interacción del usuario con el mismo, un reporte basado en el proceso de minería de datos, un análisis dinámico basado en cubos OLAP.

3.4.1. Diseño del *data warehouse*

Se diseñó e l*data warehouse* utilizando un modelo en estrella, y en cuanto a los estándares para el modelado, se define lo siguiente:

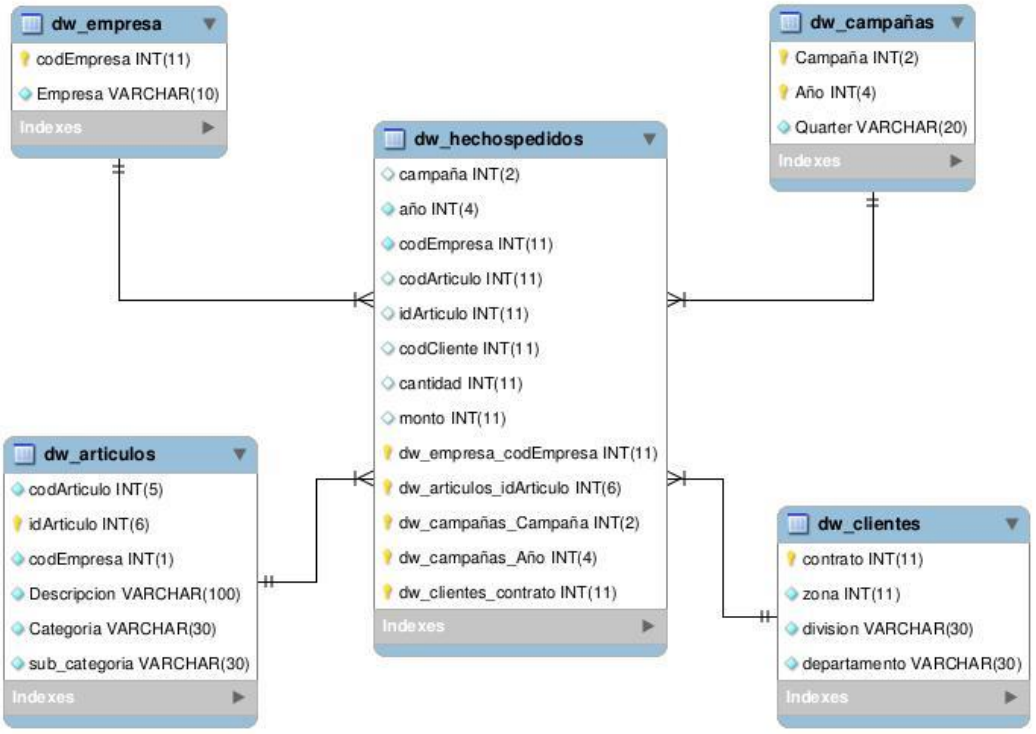
- Todas las tablas del *data warehouse* (hechos y dimensiones) poseerán el prefijo dw_.

- Todas las tablas intermedias utilizadas para el proceso de actualización poseerán el prefijo inter.

A continuación se presentan las distintas tablas dimensionales del *data warehouse* y sus relacionamientos con las tablas de hechos, para el diseño del diagrama entidad relación se utilizó la herramienta MySQL Workbench.

En las figuras 7,8 y 9 se pude observar los diagramas de entidad relación del *data warehouse*.

Figura 7: DER Hechos Pedidos



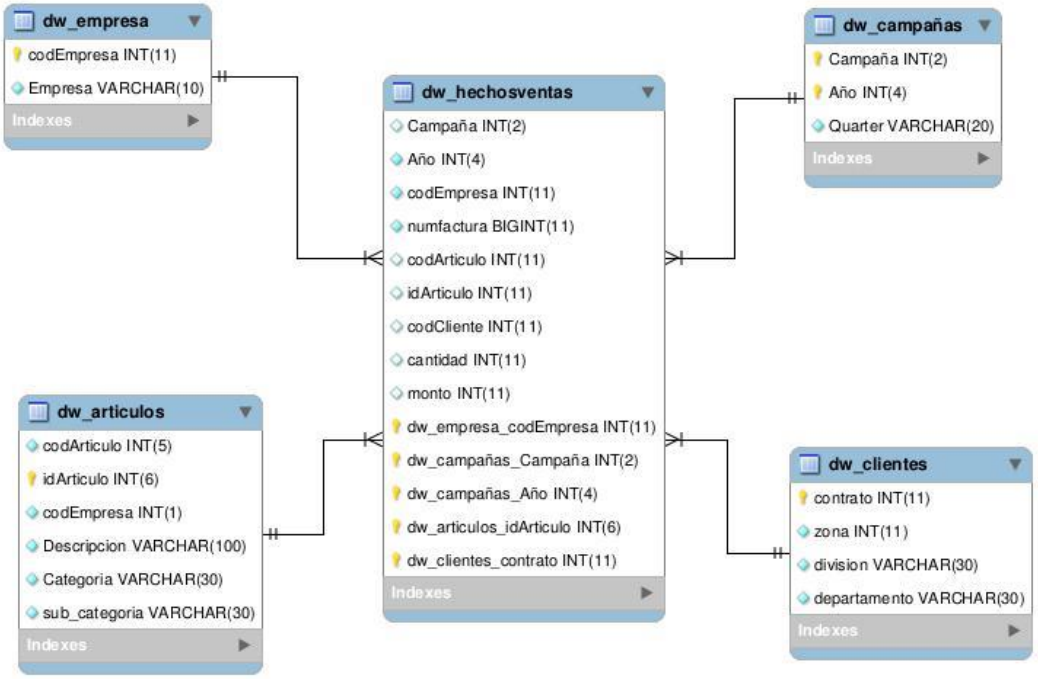


Figura 8: DER Hechos Ventas.

Figura 9: DER Proyección

3.4.2. Diseño del proceso ETL

A continuación se presenta en tablas, el mapeo del *data warehouse* indicando sus respectivas fuentes de datos.

Para la concreción de esta etapa del proyecto se utilizó la herramienta Pentaho Data Integration.

Columna	Tipo	Llave	Fuente Viru
CodArticulo	Int	Si	Artiuclos Viru.dbf.FSCODE
Descripción	Varchar	No	Artiuclos Viru.dbf.DESCRIPCION
categoría	Varchar	No	Artiuclos Viru.dbf.CATEGORIA
subcategoria	Varchar	No	Artiuclos Viru.dbf.SUBCATEGORIA

Tabla 10: Mapeo de tabla inter_articulos_viru

Columna	Tipo	Llave	Fuente Viru	Fuente Flayp
CodArticulo	int	No	inter_articulos_vi ru.codArticulo	Fwa.Articulos.fsCod e
IdArticulo	int	Si	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
CodEmpresa	int	No	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
Descripción	Varchar	No	inter_articulos_vi ru.Descripcion	Fwa.Articulos.descri pción
Categoria	Varchar	No	inter_articulos_vi ru.Categoria	Fwa.Categorias.Cate goria
SubCategoria	Varchar	No	inter_articulos_vi ru.sub_categori a	Fwa.subcategorias.s ubcategoria

Tabla 11: Mapeo de tabla dw_articulos

Columna	Tipo	Llave	Fuente Viru
Contrato	Int	Si	Fwa.clientes.contrato
Zona	Int	No	Fwa.clientes.zona
División	Varchar	No	Fwa.divisiones.division
departamento	varchar	No	Fwa.deparatmentos.departamento

Tabla 12: Mapeo de tabla dw_clientes

Columna	Tipo	Llave	Fuente Flayp
Orden	int	no	Fwa.campañas_actuales.orden

Tabla 13: Mapeo de tabla dw_campañas_actuales

Columna	Tipo	Llave	Fuente Flayp
CAMPA	varchar	No	Archivo DBF
CODIGO	int	No	Archivo DBF
PRODU	varchar	No	Archivo DBF
PRECIO	bigint	No	Archivo DBF
ZONA	varchar	No	Archivo DBF
CONTRATO	bigint	No	Archivo DBF
ESTADO	char	No	Archivo DBF
CANTIDAD	bigint	No	Archivo DBF
SERIE	char	No	Archivo DBF
SUCURSAL	varchar	No	Archivo DBF
TIMBRADO	varchar	No	Archivo DBF
DOC	bigint	No	Archivo DBF
TIPO	varchar	No	Archivo DBF
SUBTIPO	varchar	No	Archivo DBF
FECHA	datetime	No	Archivo DBF
OTRODOC	varchar	No	Archivo DBF
OTROFEC	datetime	No	Archivo DBF
OTROHOR	varchar	No	Archivo DBF
VENDEDOR	varchar	No	Archivo DBF
FECHADOC	datetime	No	Archivo DBF
IMPRESO	char	No	Archivo DBF
COSTO	bigint	No	Archivo DBF
IMPUESTO	char	No	Archivo DBF
PREMIO	char	No	Archivo DBF
EXPREMIO	char	No	Archivo DBF
FOLLETO	varchar	No	Archivo DBF
C_UNIDADES	bigint	No	Archivo DBF
C_VENTAS	bigint	No	Archivo DBF
C_CAJA	bigint	No	Archivo DBF
VERIFI	char	No	Archivo DBF
RESULT	char	No	Archivo DBF
BRASIL_1	char	No	Archivo DBF
BRASIL_2	char	No	Archivo DBF
COLUMNA	int	No	Archivo DBF
GRUPO	varchar	No	Archivo DBF
OPERADOR	varchar	No	Archivo DBF
PAGINA	int	No	Archivo DBF
FSCODE	bigint	No	Archivo DBF
C_CLIENTES	bigint	No	Archivo DBF
C_SUSTITU	char	No	Archivo DBF
C_PACKIN	char	No	Archivo DBF
XXUSR	varchar	No	Archivo DBF
XXFECHA	datetime	No	Archivo DBF
XXHORA	varchar	No	Archivo DBF

Tabla 14: Mapeo de tabla clonado_dbf_viru

Columna	Tipo	Llave	Fuente Flayp
Campaña	Int	No	Fwa.pedidos.campaña
Año	Int	No	Fwa.pedidos.año
codCliente	Int	No	Fwa.pedidos.numCliente
codArticulo	Int	No	Fwa.pedidodetalles.fsCode
idArticulo	Int	No	Agregado en el Proceso de ETL
Cantidad	Int	No	Fwa.pedidodetalles.cantidad
Monto	Int	No	Fwa.pedidodetalles.precio
codEmpresa	Int	No	Agregado en el Proceso de ETL

Tabla 15: Mapeo de tabla inter_pedido_avon

Columna	Tipo	Llave	Fuente Viru
Campaña	Int	No	clonado_dbf_viru.CAMPA
Año	Int	No	clonado_dbf_viru.ANO
codCliente	Int	No	clonado_dbf_viru.CONTRATO
codArticulo	Int	No	clonado_dbf_viru.PRODU
idArticulo	Int	No	Agregado en el Proceso de ETL
Cantidad	Int	No	clonado_dbf_viru.CANTIDAD
Monto	Int	No	clonado_dbf_viru.PRECIO
codEmpresa	Int	No	Agregado en el Proceso de ETL

Tabla 16: Mapeo de tabla inter_pedido_viru

Columna	Tipo	Llave	Fuente Flayp
numDocumento	Int	Si	Fwa.ventas.numDocumento
Campaña	Int	No	Fwa.ventas.campaña
Año	Int	No	Fwa.ventas.año
codCliente	Int	No	Fwa.ventas.contrato
codArticulo	Int	No	Fwa.ventasdetalle.codArticulo
idArticulo	Int	No	Agregado en el Proceso de ETL
Cantidad	Int	No	Fwa.ventasdetalle.cantidad
Monto	Int	No	Fwa.ventasdetalle.precio
codEmpresa	Int	No	Agregado en el Proceso de ETL

Tabla 17: Mapeo de tabla inter_ventas_avon

Columna	Tipo	Llave	Fuente Viru
CAMPA	bigint	No	clonado_dbf_viru.CAMPA
ANO	smallint	No	clonado_dbf_viru.ANO
PRODU	varchar	No	clonado_dbf_viru.PRODU
idArticulo	int	No	Agregado en el Proceso de ETL
PRECIO	bigint	No	clonado_dbf_viru.PRECIO
CONTRATO	bigint	No	clonado_dbf_viru.CONTRATO
CANTIDAD	bigint	No	clonado_dbf_viru.CANTIDAD
DOC	bigint	No	clonado_dbf_viru.DOC
codEmpresa	int	No	Agregado en el Proceso de ETL

Tabla 18: Mapeo de tabla inter_ventas_viru

Columna	Tipo	Llave	Fuente Viru	Fuente Flayp
Campaña	int	Si	inter_ventas_viru.CAMPA	inter_ventas_avon.Campaña
Año	int	Si	inter_ventas_viru.AÑOAgregado en el Proceso de ETL	inter_ventas_avon.Año
CodEmpresa	int	Si	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
NumFactura	bigin	No	inter_ventas_viru.DOC	inter_ventas_avon.numDocumento
CodArticulo	Int	No	inter_ventas_viru.PRODU	inter_ventas_avon.codArticulo
IdArticulo	int	Si	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
CodCliente	Int	Si	inter_ventas_viru.CONTRATO	inter_ventas_avon.codCliente
Cantidad	Int	No	inter_ventas_viru.CANTIDAD	inter_ventas_avon.cantidad
Monto	Int	No	inter_ventas_viru.PRECIO	inter_ventas_avon.monto

Tabla 19: Mapeo de tabla dw_hechosventas

Columna	Tipo	Llave	Fuente Viru	Fuente Flayp
Campaña	Int	Si	inter_pedido_viru.campaña	inter_pedido_avon.campaña
Año	Int	Si	inter_pedido_viru.año	inter_pedido_avon.año
codEmpresa	Int	Si	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
codArticulo	Int	Si	inter_pedido_viru.codArticulo	inter_pedido_avon.codArticulo
idArticulo	Int	No	Agregado en el Proceso de ETL	Agregado en el Proceso de ETL
codCliente	Int	Si	inter_pedido_viru.codCliente	inter_pedido_avon.codCliente
Cantidad	Int	Si	inter_pedido_viru.cantidad	inter_pedido_avon.cantidad
Monto	Int	No	inter_pedido_viru.monto	inter_pedido_avon.monto

Tabla 20: Mapeo de tabla dw_hechospedidos

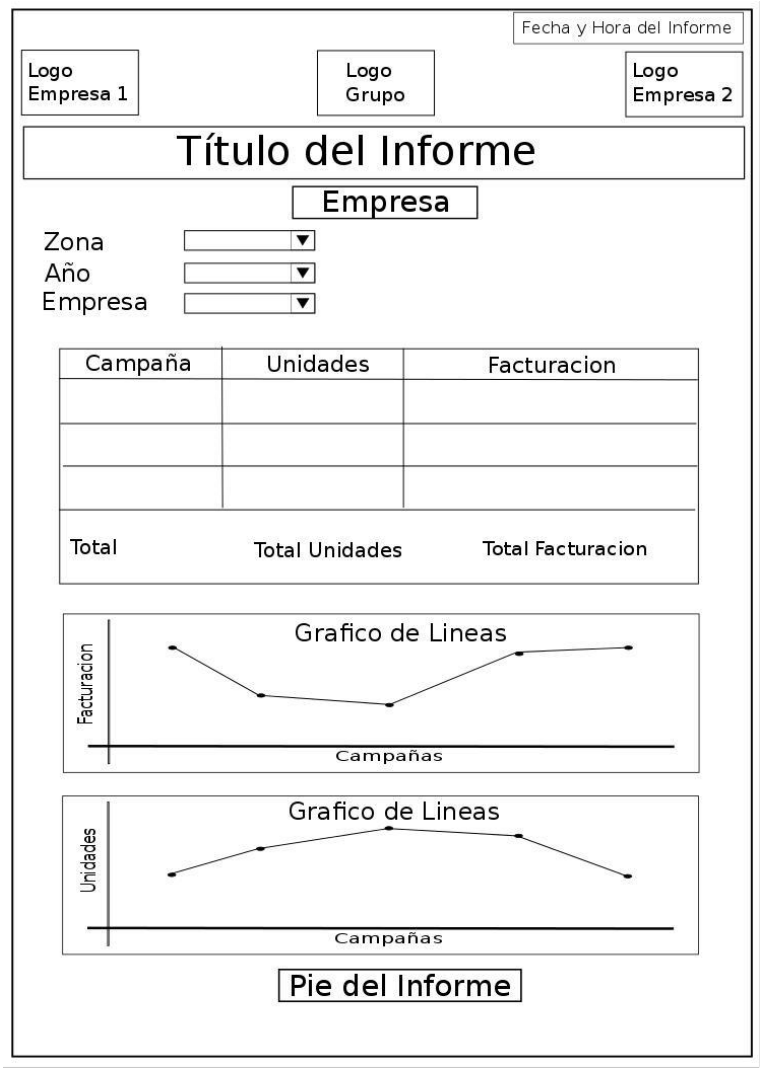
3.4.3. Diseño de reportes

En esta sección se detalla el diseño de los distintos tipos de informes previstos en el prototipo de solución de Inteligencia de Negocios.

La figura 10 representa el diseño del reporte ventas por campañas por zonas, el mismo está basado en la tabla dw_hechosventas y en las tablas de dimensiones dw_empresas, dw_campañas, dw_clientes del *data warehouse*, este reporte aporta información sobre la venta efectiva por zonas, en las variables de unidades y facturación de ambas empresas en conjunto, como así también por separado.

El reporte visualiza en un gráfico de líneas, el rendimiento de cada una de las campañas, tanto las campañas con datos cerrados, como así también las campañas en curso. Pudiéndose filtrar esa información por zonas y por años, como así también visualizar los datos de una empresa en particular, o de ambas en conjunto.

Figura 10 Ventas por Campaña



La figura 11 representa el diseño de uno de los reportes que se utiliza para representar el resultado de la minería de datos que predice las ventas, el mismo está basado en las tablas dw_proyección y dw_empresas del *data warehouse*, este reporte aporta información sobre la proyección de ventas en cuanto a las variables de unidades y facturación.

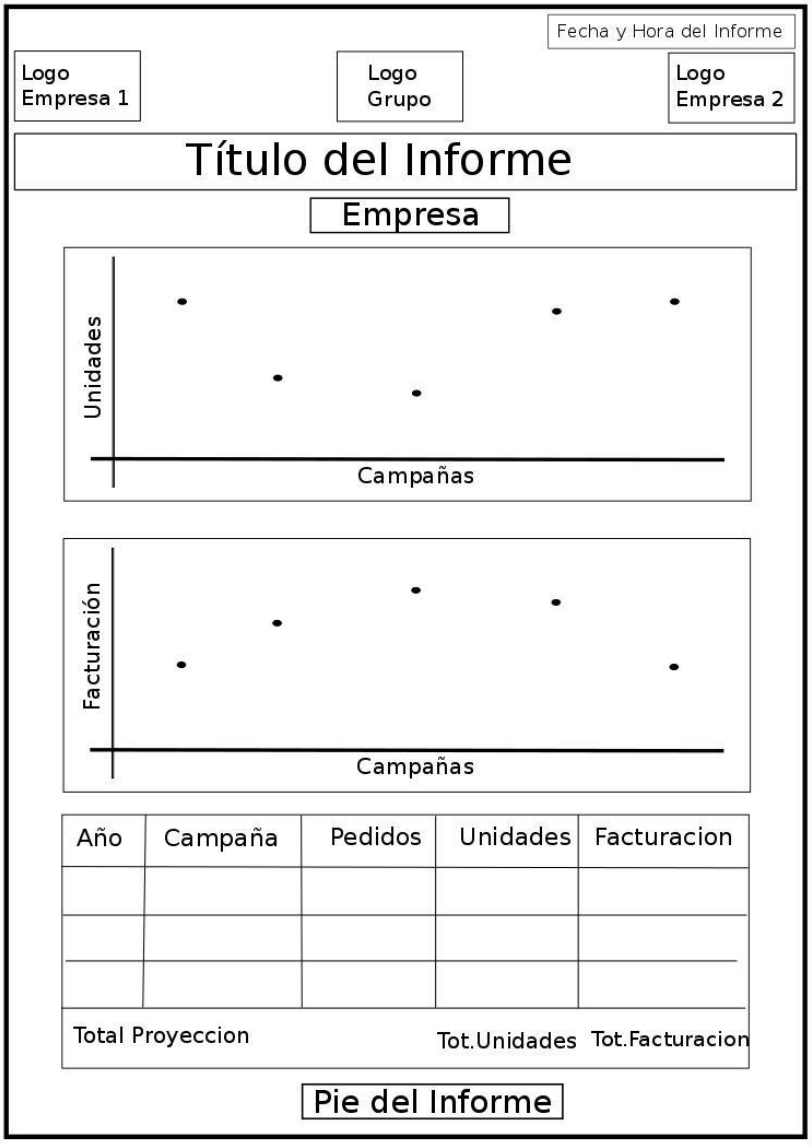


Figura 11: Reporte proyección de ventas

3.5. Diagramas casos de uso

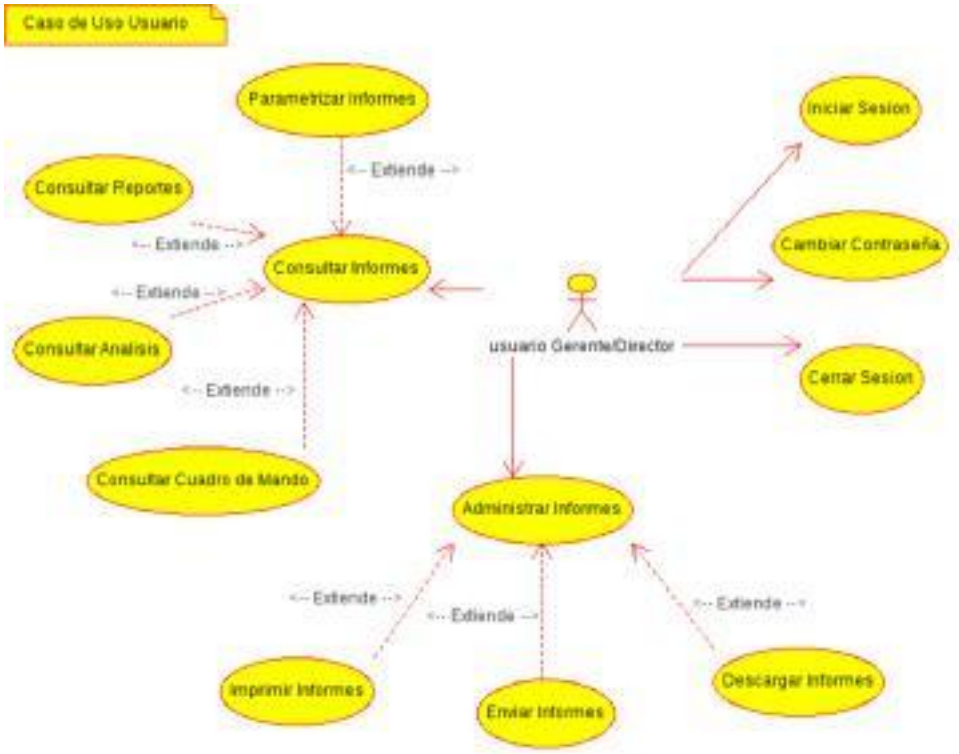


Figura 20: Caso de uso usuario

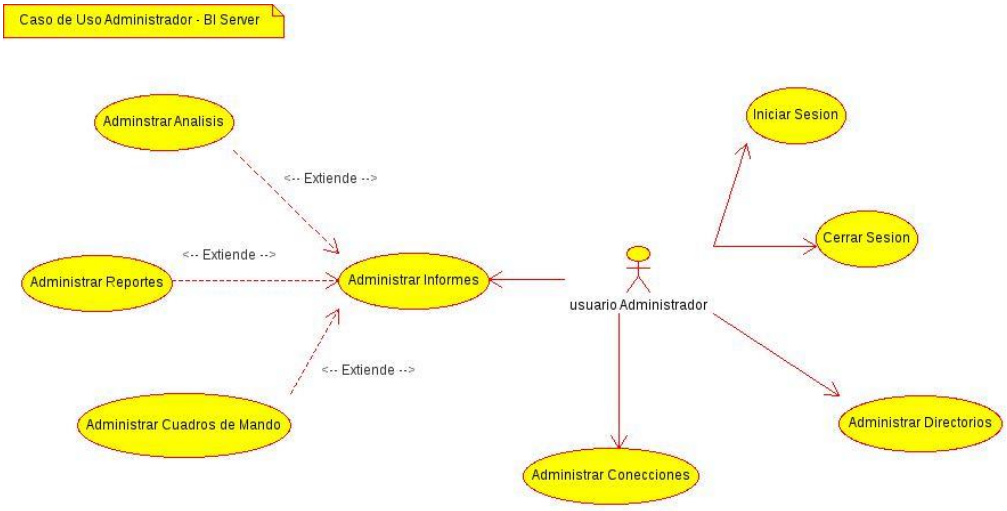


Figura 22: Caso de uso administrador – servidor

3.6. Diagrama de actividades

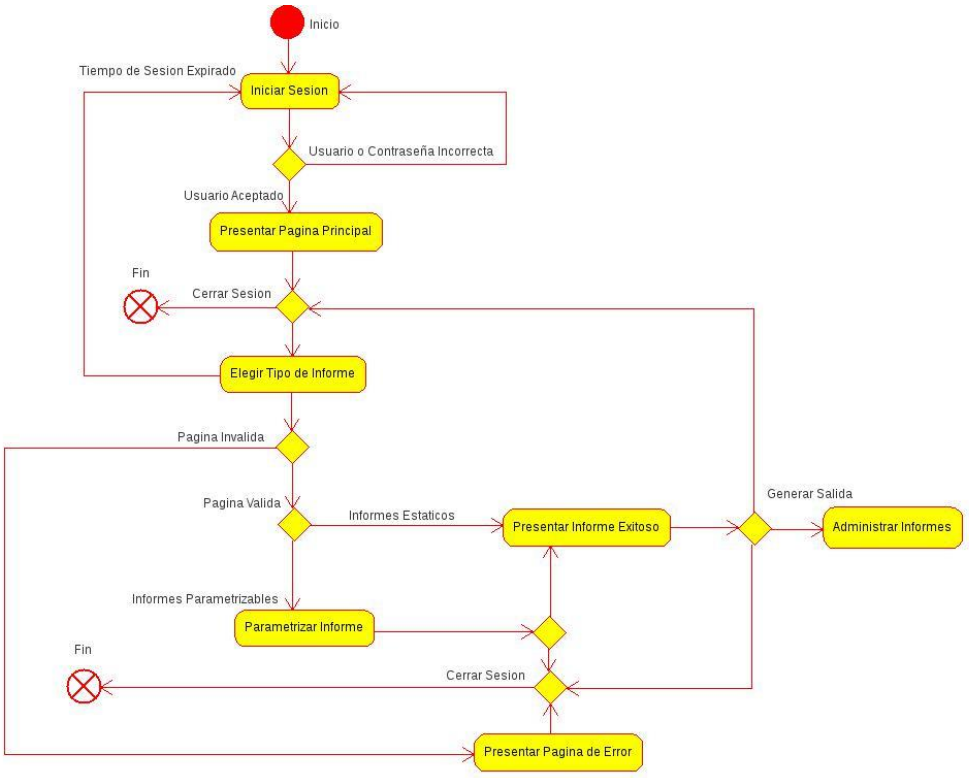


Figura 24: Diagrama de Actividades usuario

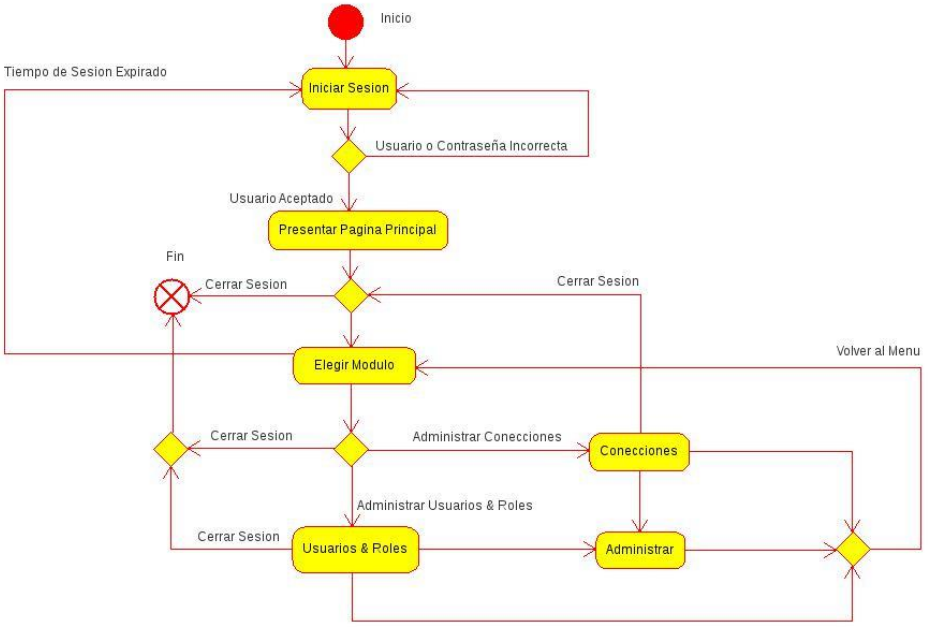


Figura 25: Diagrama de actividades administrador - usuarios

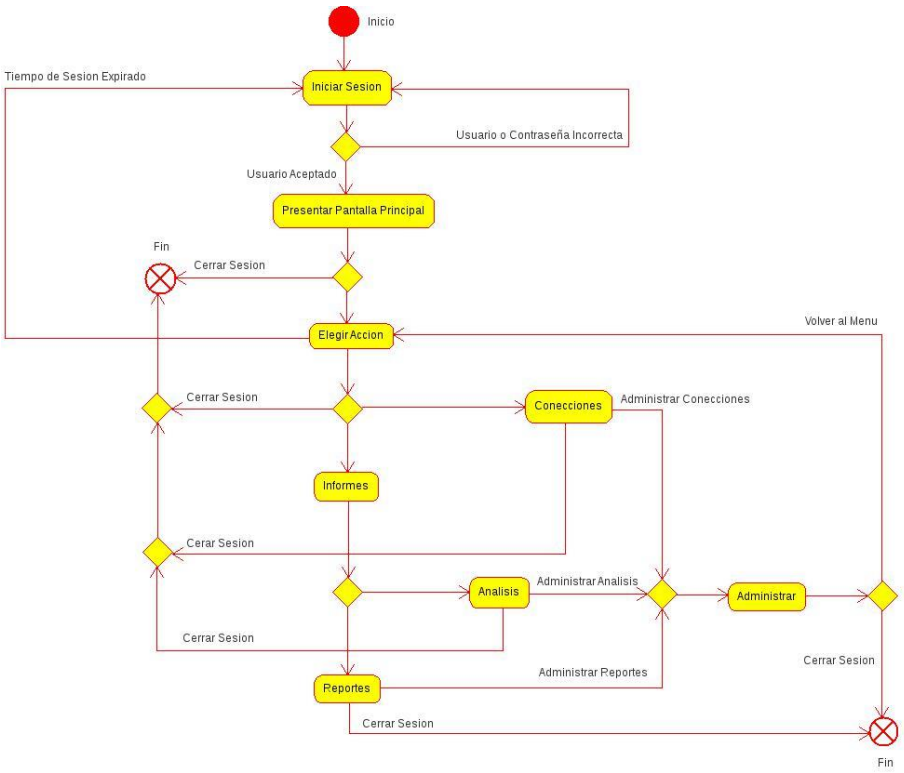


Figura 26: Diagrama de actividades administrador - servidor

3.8. Desarrollo del prototipo

3.8.1. Desarrollo del *data warehouse*

Para el desarrollo del *data warehouse* se utilizó el motor de base de datos MySQL utilizando interfaz de desarrollo la herramienta phpMyAdmin. Con la cual se desarrollaron las tablas intermedias, de dimensiones y de hechos además de las funciones y procedimientos almacenados requeridos para el correcto funcionamiento del *data warehouse*. En la Figura 33 se puede apreciar el *data warehouse* visto a través del phpMyAdmin.

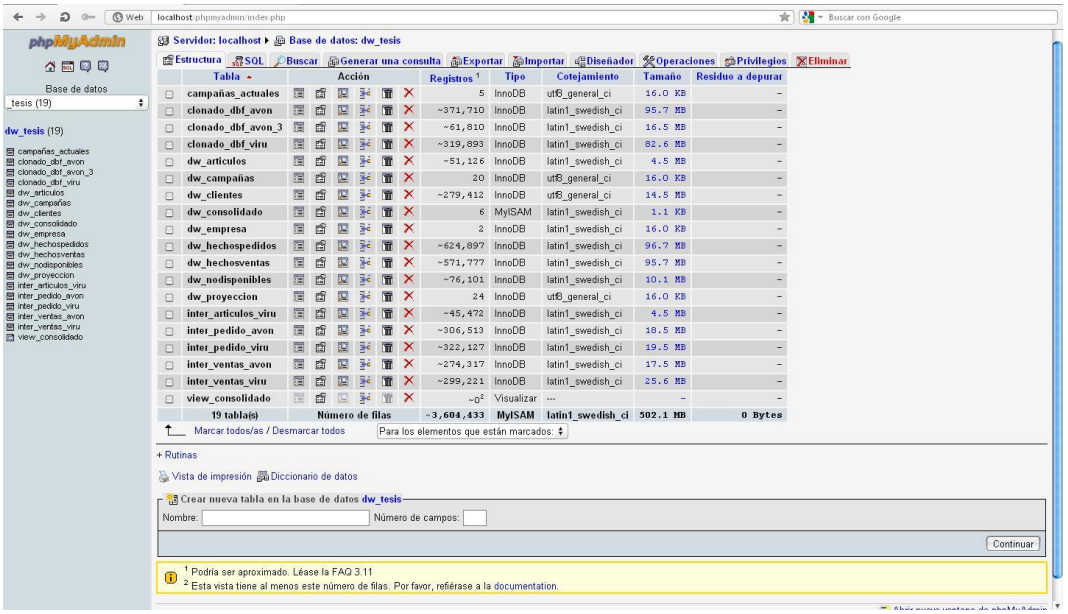


Figura 33: *Data warehouse* en PhpMyAdmin

3.8.2. Desarrollo de proceso ETL

En este apartado se detalla a modo de ejemplo, el conjunto de transformaciones que comprenden el trabajo para la carga de la tabla dw_hechosventas, en el *data warehouse*, en ella se describen las distintas etapas por las que pasaron los datos fuentes de ambas empresas hasta llegar a ser cargados en el *data warehouse*.

A continuación se presenta el proceso de ETL realizado con la herramienta Spoon incluida en el paquete de Pentaho Data Integration. Para una mejor compresión se detalla primeramente las transformaciones llevadas a cabo sobre los datos de Virú S.R.L. y luego las de la Empresa S.R.L.

En la figura 34 se observa la primera transformación, la misma utiliza información almacenada en la tabla dw_tesis.campañas actuales, donde se indican cuáles son las tres campañas vigentes, el sistema operacional de la empresa Virú S.R.L.

actualmente genera un archivo dbf por campaña, los mismos se actualizan periódicamente, conociendo las campañas vigentes, esta transformación selecciona el archivo dbf correspondiente de manera automática, e inserta esa información en la tabla clonado_dbf_viru tal cual reside en la fuente.

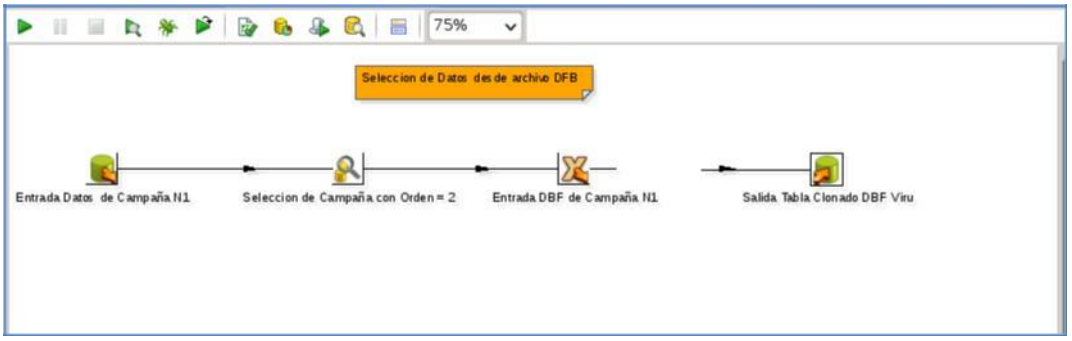


Figura 34: Transformación selección de campaña a cargar

3.8.3. Desarrollo de minería de datos

En esta etapa se presenta el desarrollo del algoritmo de regresión lineal con mínimos cuadrados utilizado para realizar el proceso de proyección de ventas con minería de datos.

A continuación se detalla el algoritmo escrito en lenguaje SQL utilizado en un procedimiento almacenado, para la proyección de ventas sobre las ventas ya concretadas, almacenadas en la tabla dw_consolidado, el resultado lo escribe en la tabla dw_proyeccion.

```
CREATE PROCEDURE `regresion_lineal`()
BEGIN
DECLARE _promedioX DECIMAL(20,10);
DECLARE _promedioY DECIMAL(20,10);
DECLARE _stdY DECIMAL(20,10);
DECLARE _varY DECIMAL(20,0);
DECLARE _maxY DECIMAL(20,10);
DECLARE _pendiente DECIMAL(20,10);

DECLARE _intercepto DECIMAL(20,10);
DECLARE _empresa INT;
DECLARE _campa INT;

SET _empresa = 1;

regresion_lineal: LOOP

SELECT AVG(Unidades) INTO _promedioY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT AVG(Pedidos) INTO _promedioX FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT STD(Unidades) INTO _stdY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT POW(STD(Unidades),2) INTO _varY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT MAX(Unidades) INTO _maxY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT SUM((Unidades-_promedioY)*(Pedidos-
_promedioX))/SUM(POW((Pedidos-_promedioX),2)) INTO _pendiente
FROM dw_consolidado WHERE codEmpresa = _empresa;
SET _intercepto:= _promedioY-_pendiente*_promedioX;
UPDATE dw_proyeccion SET Unidades = _pendiente * Pedidos +
_intercepto WHERE codEmpresa = _empresa;
```

```
SELECT AVG(Facturacion) INTO _promedioY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT AVG(Pedidos) INTO _promedioX FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT STD(Facturacion) INTO _stdY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT POW(STD(Facturacion),2) INTO _varY FROM
dw_consolidado WHERE codEmpresa = _empresa;
SELECT MAX(Facturacion) INTO _maxY FROM dw_consolidado
WHERE codEmpresa = _empresa;
SELECT SUM((Facturacion-_promedioY)*(Pedidos-
_promedioX))/SUM(POW((Pedidos-_promedioX),2)) INTO _pendiente
FROM dw_consolidado WHERE codEmpresa = _empresa;

SET _intercepto:= _promedioY-_pendiente*_promedioX;

UPDATE dw_proyeccion SET Facturacion = _pendiente * Pedidos +
_intercepto WHERE codEmpresa = _empresa;

SELECT MAX(año*100+campaña) INTO _campa FROM
dw_consolidado WHERE codEmpresa = _empresa;

DELETE FROM dw_proyeccion WHERE (año*100+campaña) <=
_campa AND codEmpresa = _empresa;
SET _empresa = _empresa + 1;
IF _empresa > 2 THEN
LEAVE regresion_lineal;
END IF;
END LOOP regresion_lineal;
END
```

CAPITULO IV: ANALISIS E INTERPRETACION DE LOS DATOS

4.1 Análisis de Resultados

A continuación se muestran los resultados obtenidos de la prueba de campo realizada, tanto para las variables independiente y dependientes, aplicando las métricas correspondientes a los indicadores seleccionados. Dichos resultados son sometidos a un minucioso análisis para extraer los principales rasgos de su comportamiento y, de este modo tener elementos de juicio para interpretar de manera global el comportamiento de las dos variables involucradas.

4.1.1 Para la Variable Independiente

X_1=Sistema de TI

INDICADORES	INDICE
Aplicación de un Sistema	[No…Si]

- **Para el Indicador: Tiempo de entrada de procesos.**

  En el caso de este indicador el análisis se basa en que si se implementa el sistema como responderán a dicho sistema en tiempo de prueba.

4.1.2 Para la Variable Dependiente

Mejora de los procesos

Tabla 01: Indicadores de Variables Dependientes

INDICADORES	UNIDAD DE MEDIDA	INDICES
Y_1:Tiempo de entrada	Segundos	[4...5]
Y_2:Tiempo de salida	Segundos	[4...5]
Y_3:Tiempo de registro	Segundos	[3...4]

4.2. Grupo Experimental (Pre Prueba y Post Prueba)

- **PARA LOS DATOS DE LA PRE PRUEBA**

PARA EL INDICADOR 1:

Para esta investigación el tiempo requerido se mide de acuerdo al índice así mismo la magnitud del error sea tolerable y el riesgo admisible, se requiere un tamaño de muestra en la que se asegure un 95% de probabilidad de que el error no sea superior al 5%.

Considerando que el tamaño de muestra en pre-prueba es de 57 observaciones realizadas en el mes de octubre referidas al tiempo que le toma realizar cada proceso, en la tabla se muestran las estadísticas descriptivas derivadas de la información recolectada durante la etapa de pre-prueba, las mismas que ayudará a interpretar los datos mostrados. Estos datos de pre-prueba ayudaran para contrastar los tiempos de control ante la implantación del sistema con los datos obtenidos después de la implantación del sistema.

Tabla 02: Tiempo de entrada de datos

FICHA DE OBSERVACION		
INDICADOR 1: Tiempo de entrada INSTRUMENTO DE APOYO: Reloj Analógico		
DATOS PRE PRUEBA		DATOS POST PRUEBA
Unidad de Análisis	Tiempo (Segundos)	Tiempo (Segundos)
1	4.5	1.5
2	4.8	2.0
3	5.0	1.5
4	4.6	1.7
5	4.3	2.0
6	4.5	2.0
7	4.3	1.8
8	4.5	1.9
9	4.3	1.6
10	4.7	2.0
11	4.3	1.8
12	4.4	1.9
13	4.9	2.0
14	5.0	1.8
15	4.5	1.9
16	4.3	2.0
17	4.7	1.8

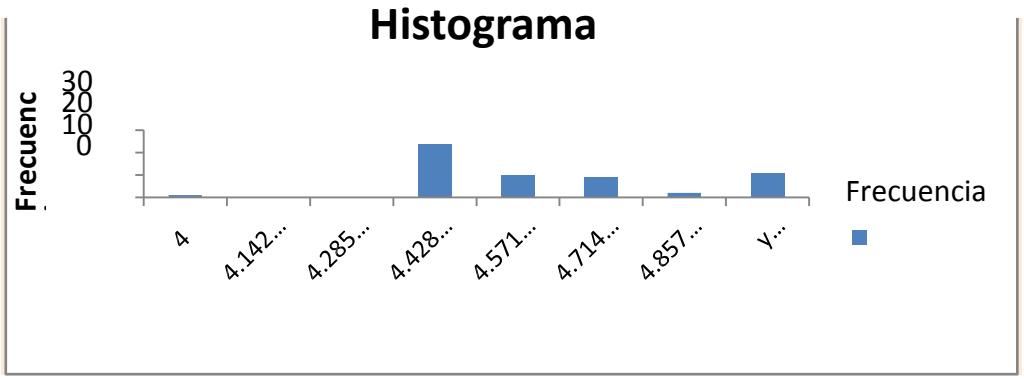
18	4.3	2.0
19	4.4	1.5
20	4.9	1.8
21	4.5	1.7
22	4.0	2.0
23	4.7	1.7
24	4.3	1.6
25	4.4	2.0
26	4.9	1.8
27	4.8	1.7
28	4.5	1.9
29	4.3	1.6
30	4.7	2.0
31	4.3	2.0
32	4.4	1.8
33	4.9	2.0
34	4.5	1.9
35	4.3	2.0
36	4.7	1.8
37	4.3	1.7
38	4.4	1.5
39	4.9	1.8
40	4.5	1.7

41	4.3	2.0
42	4.7	1.7
43	4.3	1.5
44	4.4	2.0
45	4.9	1.8
46	4.5	1.5
47	4.7	1.9
48	4.5	1.6
49	4.3	2.0
50	4.7	1.7
51	4.3	1.8
52	4.4	1.9
53	4.9	2.0
54	5.0	1.8
55	4.3	1.5
56	4.4	1.8
57	4.9	2.0

Tabla 03: Estadística Descriptiva de los Datos de Pre- Prueba

Media	4.543859649
Error típico	0.032638518
Mediana	4.5
Moda	4.3
Desviación estándar	0.246415404
Varianza de la muestra	0.060720551
Curtosis	-0.906179481
Coeficiente de asimetría	0.362123248
Rango	1
Mínimo	4
Máximo	5
Suma	259
Cuenta	57
Mayor (1)	5
Menor(1)	4
Nivel de confianza(95.0%)	0.065382807

Grafico 58: Histograma del Tiempo de entrada



Interpretación

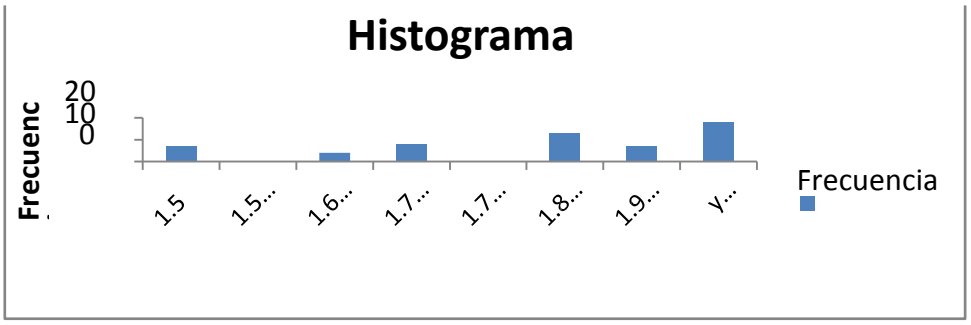
Analizando los resultados de la tabla anterior en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo que toma en controlar el proceso es una media de 4.54. En general, el tiempo más alto obtenido fue de 5,0 mientras que el más bajo fue de 4,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,24.

PARA LOS DATOS DE LA POST PRUEBA

Tabla 04: Estadística Descriptiva de los Datos de Post-Prueba

Media	1.810526316
Error típico	0.022761184
Mediana	1.8
Moda	2
Desviación estándar	0.17184317
Varianza de la muestra	0.029530075
Curtosis	-0.938704472
Coeficiente de asimetría	-0.475151411
Rango	0.5
Mínimo	1.5
Máximo	2
Suma	103.2
Cuenta	57
Mayor (1)	2
Menor(1)	1.5
Nivel de confianza(95.0%)	0.04559613

Grafico 59: Histograma del Tiempo de entrada de datos



Interpretación

En función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo empleado, tiene una media de 1,81. En general, el tiempo más alto obtenido fue de 2,0 , mientras que el más bajo fue de 1,5. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,17.

- **PARA LOS DATOS DE LA PRE PRUEBA**

PARA EL INDICADOR 2: Tiempo de salida

Para esta investigación el tiempo de salida se mide de acuerdo al índice así mismo la magnitud del error sea tolerable y el riesgo admisible, se requiere un tamaño de muestra en la que se asegure un 95% de probabilidad de que el error no sea superior al 5%.

Considerando que el tamaño de muestra en pre-prueba es de 57 observaciones realizadas en el mes de octubre referidas al tiempo de salida, en la tabla se muestran las estadísticas descriptivas derivadas de la información recolectada durante la etapa de pre-prueba, las mismas que ayudará a interpretar los datos mostrados. Estos datos de pre-prueba ayudaran para contrastar los tiempos de control antes de la implantación del sistema con los datos obtenidos después de la implantación del sistema.

Tabla 05: Tiempo de salida del proceso

FICHA DE OBSERVACION		
INDICADOR 2: Tiempo de salida INSTRUMENTO DE APOYO: Analógico		
DATOS PRE PRUEBA		DATOS POST PRUEBA
Unidad de Análisis	Tiempo (Segundos)	Tiempo (Segundos)
1	4.3	1.7
2	4.5	2.0
3	4.3	2.0
4	4.7	1.8
5	4.3	1.9
6	4.4	1.6

7	4.9	2.0
8	4.6	1.8
9	4.0	1.9
10	4.5	2.0
11	4.3	1.8
12	4.7	1.8
13	4.3	1.9
14	4.4	2.0
15	4.7	1.8
16	4.3	1.9
17	4.4	2.0
18	4.2	1.8
19	4.6	2.0
20	4.0	1.5
21	4.5	1.8
22	4.3	2.0
23	4.3	1.8
24	4.4	1.9
25	4.9	1.7
26	4.6	2.0
27	4.0	2.0
28	4.5	1.8
29	4.3	1.9

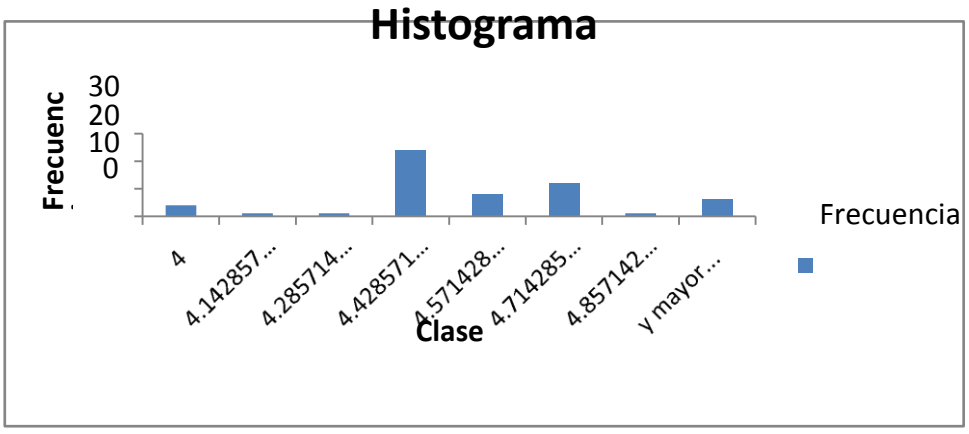
30	4.3	1.6
31	4.4	2.0
32	4.9	1.8
33	4.6	1.9
34	4.0	2.0
35	4.5	1.8
36	4.3	1.9
37	4.3	1.7
38	4.4	2.0
39	4.7	2.0
40	4.6	1.8
41	4.1	1.9
42	4.5	1.6
43	4.3	2.0
44	4.5	1.8
45	4.3	1.9
46	4.3	2.0
47	4.4	1.8
48	4.9	1.9
49	4.6	1.8
50	4.5	1.5
51	5.0	1.9
52	4.3	1.6

53	4.4	2.0
54	4.9	1.7
55	4.6	1.6
56	4.7	1.8
57	4.8	1.9

Tabla 06: Estadística Descriptiva de los Datos de Pre- Prueba

Media	4.461403509
Error típico	0.032129358
Mediana	4.4
Moda	4.3
Desviación estándar	0.242571334
Varianza de la muestra	0.058840852
Curtosis	-0.210289834
Coeficiente de asimetría	0.229256596
Rango	1
Mínimo	4
Máximo	5
Suma	254.3
Cuenta	57
Mayor (1)	5
Menor(1)	4
Nivel de confianza(95.0%)	0.064362838

Grafico 60: Histograma del Tiempo de salida del proceso



Interpretación

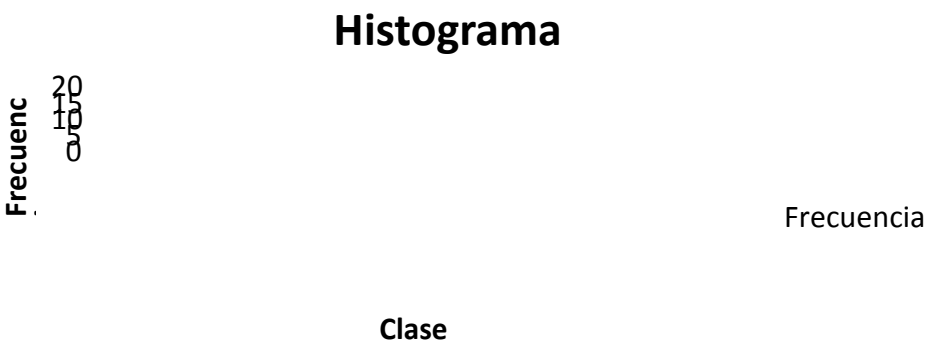
Analizando los resultados de la tabla anterior en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo que toma es una media de 4,46. En general, el tiempo más alto obtenido fue de 5,0 mientras que el más bajo fue de 4,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,24.

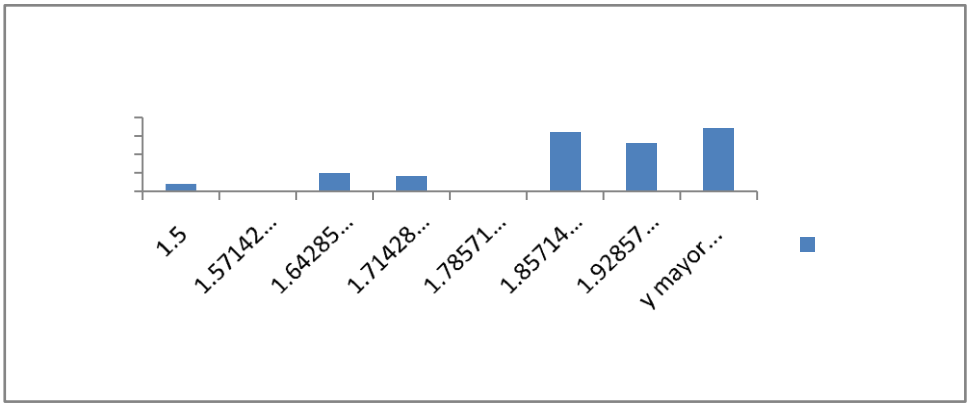
- **PARA LOS DATOS DE LA POST PRUEBA**

Tabla 07: Estadística Descriptiva de los Datos de Post- Prueba

Media	1.847368421
Error típico	0.018581471
Mediana	1.9
Moda	2
Desviación estándar	0.140287031
Varianza de la muestra	0.019680451
Curtosis	-0.132859933
Coeficiente de asimetría	-0.748549035
Rango	0.5
Mínimo	1.5
Máximo	2
Suma	105.3
Cuenta	57
Mayor (1)	2
Menor(1)	1.5
Nivel de confianza(95.0%)	0.03722316

Grafico 61: Histograma del Tiempo de salida del proceso


Histograma
Frecuenc
20
15
10
5
0
Frecuencia
Clase


1.5
1.57142...
1.64285...
1.71428...
1.78571...
1.85714...
1.92857...
y mayor...

Interpretación

Analizando los resultados de la tabla anterior en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo que toma es una media de 1,84. En general, el tiempo más alto obtenido fue de 2,0 mientras que el más bajo fue de 1,5. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,14.

- **PARA LOS DATOS DE LA PRE PRUEBA**

PARA EL INDICADOR 3: Tiempo de Registro del proceso

Para esta investigación el tiempo de registro del proceso, se mide de acuerdo al índice así mismo la magnitud del error sea tolerable y el riesgo admisible, se requiere un tamaño de muestra en la que se asegure un

95% de probabilidad de que el error no sea superior al 5%.

Considerando que el tamaño de muestra en pre-prueba es de 57 observaciones realizadas en el mes de octubre, en la tabla se muestran las estadísticas descriptivas derivadas de la información recolectada durante la etapa de pre-prueba, las mismas que ayudará a interpretar los datos mostrados. Estos datos de pre-prueba ayudaran para contrastar los tiempos antes de la implantación del sistema con los datos obtenidos después de la implantación del sistema.

Tabla 08: Tiempo de Registro

FICHA DE OBSERVACION		
INDICADOR 3: Tiempo de Registro INSTRUMENTO DE APOYO		
DATOS PRE PRUEBA		DATOS POST PRUEBA
Unidad de Análisis	Tiempo	Tiempo
1	3.8	1.6
2	3.6	2.0
3	3.9	1.8
4	4.0	1.9
5	3.5	2.0
6	3.8	1.8
7	3.7	1.8
8	3.6	1.9
9	3.9	2.0
10	3.5	1.8
11	3.6	1.5
12	3.4	2.0
13	3.8	1.8
14	3.5	1.9
15	3.8	2.0
16	3.7	1.8
17	3.3	1.8

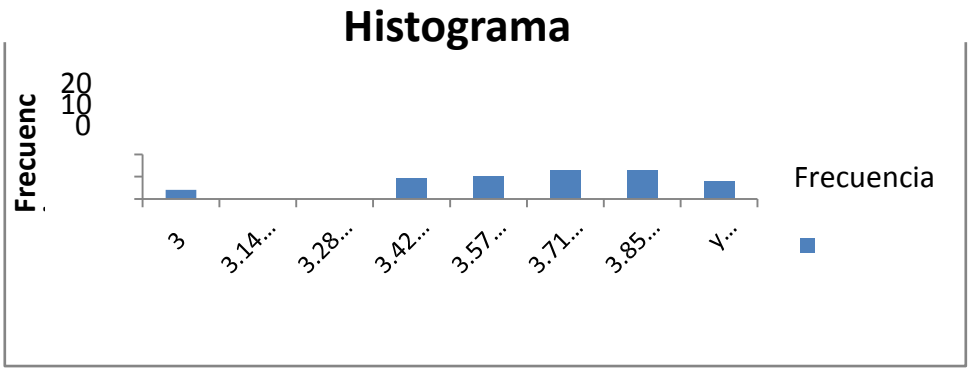
18	3.7	1.9
19	3.5	2.0
20	3.6	1.8
21	3.4	1.5
22	3.8	2.0
23	3.5	1.8
24	3.8	1.9
25	3.7	2.0
26	3.3	1.8
27	3.0	1.8
28	3.5	1.9
29	3.6	2.0
30	3.4	1.8
31	3.8	1.5
32	3.5	2.0
33	3.8	1.8
34	3.7	1.9
35	3.3	2.0
36	3.0	1.8
37	3.5	1.8
38	3.6	1.9
39	3.4	2.0
40	3.8	1.8

41	4.0	1.6
42	3.9	2.0
43	3.9	1.8
44	4.0	1.9
45	3.8	2.0
46	3.0	1.8
47	3.5	1.8
48	3.6	1.9
49	3.4	2.0
50	3.8	1.8
51	3.8	1.9
52	3.0	2.0
53	3.5	1.8
54	3.6	1.8
55	3.4	1.9
56	3.8	2.0
57	4.0	1.8

Tabla 09: Estadística Descriptiva de los Datos de Pre- Prueba

Media	3.607017544
Error típico	0.033663123
Mediana	3.6
Moda	3.8
Desviación estándar	0.254151002
Varianza de la muestra	0.064592732
Curtosis	0.288027167
Coeficiente de asimetría	-0.683351117
Rango	1
Mínimo	3
Máximo	4
Suma	205.6
Cuenta	57
Mayor (1)	4
Menor(1)	3
Nivel de confianza (95.0%)	0.067435338

Grafico 62: Tiempo de Registro de Tareo del Personal Obrero



Interpretación

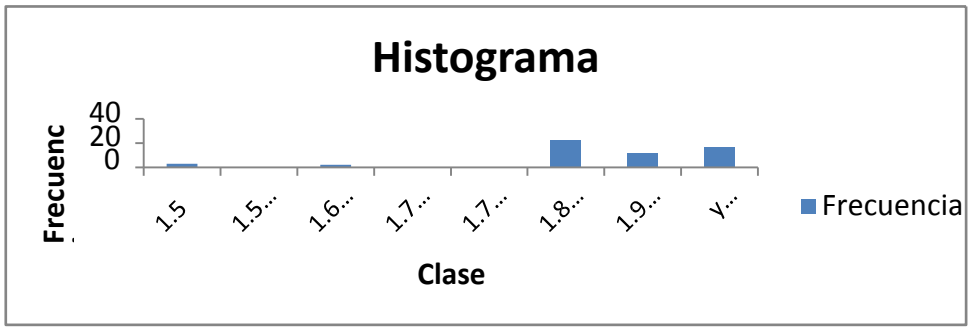
Analizando los resultados de la tabla anterior en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo de registro es una media de 3,60. En general, el tiempo más alto obtenido fue de 4,0 mientras que el más bajo fue de 3,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,25.

- **PARA LOS DATOS DE LA POST PRUEBA**

Tabla 10: Estadística Descriptiva de los Datos de Post- Prueba

Media	1.857894737
Error típico	0.017332689
Mediana	1.9
Moda	1.8
Desviación estándar	0.130858932
Varianza de la muestra	0.01712406
Curtosis	1.258513117
Coeficiente de asimetría	-1.045680571
Rango	0.5
Mínimo	1.5
Máximo	2
Suma	105.9
Cuenta	57
Mayor (1)	2
Menor(1)	1.5
Nivel de confianza (95.0%)	0.034721548

Grafico 62: Tiempo de Registro



Interpretación

Analizando los resultados de la tabla anterior en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo es una media de 1,85. En general, el tiempo más alto obtenido fue de 2,0 mientras que el más bajo fue de 1,5. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,13

CAPITULO V: **CONCLUSIONES Y RECOMENDACIONES**

5.1. Conclusiones

Finalizando el presente proyecto de tesis se puede llegar a las siguientes conclusiones:

1. Se concluye que de la tabla 03 en función a los datos mostrados en la tabla de estadística descriptiva, se puede decir que, de las observaciones tomadas se obtiene que el tiempo que toma en controlar el proceso es una media de 4.54. En general, el tiempo más alto obtenido fue de 5,0 mientras que el más bajo fue de 4,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,24.
2. Asimismo, de los resultados de la tabla 05, se puede decir que, de las observaciones tomadas se obtiene que el tiempo que toma es una media de 1,84. En general, el tiempo más alto obtenido fue de 2,0 mientras que el más bajo fue de 1,5. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,14.
3. Finalmente, de los resultados de la tabla 08, se puede decir que, de las observaciones tomadas se obtiene que el tiempo de registro es una media de 3,60. En general, el tiempo más alto obtenido fue de 4,0 mientras que el más bajo fue de 3,0. Además se puede observar que con respecto al promedio, los datos muestran una desviación estándar 0,25.

5.2. Recomendaciones:

Luego de finalizar el proyecto de tesis y cumpliendo con el objetivo de la tesis se pueden plantear las siguientes recomendaciones:

1. Se recomienda que se ponga en funcionamiento esta aplicación por cuanto va a redundar en beneficio para la empresa y que su información les va a permitir poder tomar las mejores decisiones.
2. Asimismo, se recomienda que se capacite al personal de la empresa involucrado en este tema para que pueda ayudar en la implementación y puesta en funcionamiento del esta aplicación.
3. Finalmente se recomienda que tanto el personal de la empresa como la parte de alta dirección de la empresa llámese los gerentes o los jefes de áreas se pongan de acuerdo en que el uso de esta tecnología beneficiara a la empresa y muy por el contrario pensar que se estará limitando al uso del personal, por cuanto se sabe que el personal administrativo es reacio al cambio.

REFERENCIAS BIBLIOGRAFICAS

1. Tevni Grajales, En: http://tgrajales.net/investipos.pdf, 27 de marzo 2000
2. Bedoya Sánchez, Enrique Osvaldo, Metodología de la Investigacion Perú Cap. 3 En. http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/empre/bedoya_se/cap3.pdf
3. Aula Fácil, En: http://www.aulafacil.com/cursosenviados/metodo-cientifico.pdf
4. HERNÁNDEZ SAMPIERI y otros (1994). Metodología de la investigación, México, Mc Graw Hill, Cap.4
5. (HOWARD DRESNER, H 1.989).
6. GÓMEZ GARCÍA, José Luis. Introducción al big data. Barcelona: UOC, 2015 12/1044313.
7. SOLANA, Albert. Big Data para directivos: guía rápida y ejemplos prácticos. Barcelona: Empresa Activa, 2015 12/1043310
8. LARA TORRALBO, Juan Alfonso. Minería de datos. Madrid: CEF, 2014 SDB (075)004.6 LAR
9. PÉREZ MARQUÉS, María. Minería de datos a través de ejemplos. San Fernando de Henares, Madrid: RC Libros, 2014. SDB 004.6 PER
10. SCHMARZO, Bill. Big data: el poder de los datos. Madrid: Anaya Multimedia, D.L. 2014 12/993264
11. JOYANES AGUILAR, Luis. Big data: análisis de grandes volúmenes de datos en organizaciones. Barcelona: Marcombo, 2013 SDB 004.6 JOY
12. MAYER-SCHÖNBERGER, Viktor. Big data: la revolución de los datos masivos. Madrid: Turner, 2013 SDB 004.6 MAY

13. SIEGEL, Eric. Analítica predictiva: predecir el futuro utilizando Big Data. Madrid: Anaya Multimedia, 2013 12/938162
14. TALLER DE MINERÍA DE DATOS Y APRENDIZAJE (5º. 2010. Valencia). Actas del V Simposio de Teoría y Aplicaciones de Minería de Datos. [Madrid]: Garceta, D.L. 2010 12/716926
15. NETTLETON, David F. Data mining: fundamentos y metodologías. Barcelona: UOC, 2007 SDB (075)004.6 NET.

ANEXO

ANEXO Nº 01: MATRIZ DE CONSISTENCIA

TÍTULO: "APLICACIÓN DE UN SISTEMA DE TI UTILIZANDO MINERÍA DE DATOS PARA MEJORAR LOS PROCESOS EN LA EMPRESA FERRETERA LA ANGOSTURA REPRESENTACIONES DE LA CIUDAD DE ICA"

PROBLEMA	OBJETIVO	HIPÓTESIS	VARIABLES	INDICADORES	ÍNDICES	MÉTODOS	TÉCNICAS	INSTRUMENTOS
Problema Principal	***Objetivo General***	***Hipótesis General***						
¿De qué manera la aplicación de un sistema TI mejora los procesos en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica.?	Implementar un de Sistema de TI utilizando Minería de Datos para mejorar los proceso en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica.	La Implementación un de Sistema de TI utilizando Minería de Datos mejorara los proceso en la Empresa Ferretera La Angostura Representaciones de la Ciudad de Ica	**DEFINICIÓN CONCEPTUAL** ***Variable Independiente:*** X = Sistema de TI ***Variable Dependiente:*** Mejora de los procesos.	Y_1 = Tiempo de entrada de datos. Y_2 = Tiempo de salida de datos. Y_3 = Tiempo de registro	No – Si 4-5 seg. 3-5 min. 3-4	**Tipo de Investigación:** Aplicada **Nivel de investigación:** Descriptiva-correlacional **Diseño de la investigación:** Cuasi Experimental Ge: O_1 X O_2 **Muestra:** Se considerara como muestra a 57 procesos de la empresa	Entrevista Observación Análisis Documental .	Guía de Entrevista Guías de observación de campo Fichas de Observación.