

SEP

SES

TNM

INSTITUTO TECNOLÓGICO DE CHIHUAHUA II



**DATA WAREHOUSE DE APOYO PARA LA GESTIÓN
DE LA EMPRESA Y MODELOS DE PREDICCIÓN**

TESIS
PARA OBTENER EL GRADO DE

MAESTRO EN SISTEMAS COMPUTACIONALES

PRESENTA

KARLA PATRICIA RAMOS MARTÍNEZ

DIRECTOR DE TESIS
M.C. LEONARDO NEVÁREZ CHÁVEZ

CODIRECTOR DE TESIS
DR. GREGORIO RONQUILLO MÁYNEZ

CHIHUAHUA, CHIH. 08 DE NOVIEMBRE DE 2023

Dictamen

Chihuahua, Chihuahua, 08 de noviembre 2023

M.C. MARIA ELENA MARTINEZ CASTELLANOS
COORDINADORA DE POSGRADO E INVESTIGACION.
PRESENTE

Por medio de este conducto el comité tutorial revisor de la tesis para obtención de grado de Maestro en Sistemas Computacionales, que lleva el nombre de: "**DATA WAREHOUSE DE APOYO PARA LA GESTIÓN DE LA EMPRESA Y MODELOS DE PREDICCIÓN**", que presenta el C. KARLA PATRICIA RAMOS MARTINEZ, hace de su conocimiento que después de ser revisado ha dictaminado la APROBACIÓN de la misma.

Sin otro particular de momento queda de usted.

Atentamente

La Comisión de Revisión de Tesis.



M.C. LEONARDO NEVÁREZ CHÁVEZ

Director de tesis



DR. GREGORIO RONQUILLO MAYNÉZ

Co-Director



DR. ALBERTO CAMACHO RIOS

Revisor



M.C. ARTURO LEGARDA SAENZ

Revisor

DEDICATORIAS

A Dios, por presentarme las oportunidades justo cuando las necesito, aunque a veces no entienda sus tiempos.

A mis padres, por darme la vida y brindarme lo que necesitaba en la medida que ellos podían durante mi rol de estudiante.

A mis hijos, que son mi motivo principal para seguir superándome y querer salir adelante, dándoles un buen ejemplo.

A mi esposo, por todos los años vividos hasta hoy a su lado.

AGRADECIMIENTOS

A Dios,

por permitirme haber alcanzado una meta más, mostrándome el camino para lograrlo.

A mis docentes de maestría,

que me brindaron herramientas y conocimientos que me permitieron acrecentar lo que antes
había aprendido.

A mi asesor de tesis,

por su paciencia y tolerancia que me brindó, por sus conocimientos que me facilitaron en gran
medida la realización de este proyecto.

A CONACYT,

por aceptarme como becaria, que sin ese beneficio me hubiera sido imposible la realización de
mi maestría.

RESUMEN

Las pequeñas y medianas empresas se enfrentan a diferentes tipos de cambios que pueden poner en riesgo su estabilidad financiera, lo que podría ocasionar el cierre temporal o permanente. Este proyecto sirve de apoyo a aquellas empresas que cuentan actualmente con cualquier tipo de sistema de ventas ya instalado y con suficiente información de ventas que permita realizar la extracción y transformación de los datos, para su posterior análisis, con el fin de tomar decisiones y/o predicciones. De esta forma las empresas no tendrán que cerrar ni de manera temporal o permanente, sino ver sus fortalezas y hacer a un lado sus debilidades.

SUMMARY

Small and medium-sized companies face different types of changes that may jeopardize their financial stability, which could lead to temporary or permanent closure. This project supports those companies that currently have any type of sales system already installed and with enough sales information that allows the extraction and transformation of data for further analysis, in order to make decisions and / or predictions. In this way, companies will not have to close either temporarily or permanently, but see their strengths and put aside their weaknesses.

CONTENIDO

Índice de figuras.....	ix
Índice de tablas.....	xii
CAPÍTULO I. INTRODUCCIÓN.....	1
1.1 Antecedentes.....	1
1.2 Definición del problema.....	3
1.3 Alcances y limitaciones.....	3
1.4 Justificación.....	4
1.5 Objetivos.....	4
1.6 Supuestos o hipótesis.....	5
CAPÍTULO II. ESTADO DEL ARTE.....	6
CAPÍTULO III. MARCO TEÓRICO.....	9
3.1 Metodología de desarrollo.....	9
3.2 Data warehouse.....	11
3.3 Machine learning.....	12
3.4 Inteligencia de negocios (business intelligence)	16
3.5 Metadatos.....	19
3.6 Pyme.....	19
3.7 Herramientas.....	20
CAPITULO IV. DESARROLLO.....	31
4.1 Sistema base y extracción de datos.....	31
4.2 Estadísticas.....	36
4.3 Predicciones.....	46
CAPÍTULO V. RESULTADOS	51
5.1 Importación a programas manejadores de bases de datos.....	51
5.2 Tablas agregadas al sistema para manejo en el data warehouse.....	51
5.3 Diagramas del data warehouse.....	52
5.4 Cubos de datos del data warehouse.....	62
5.5 Capacitación a la empresa.....	64

CAPÍTULO VI. CONCLUSIONES.....	66
6.1 Sistema de ventas.....	66
6.2 Empresa base.....	66
6.3 Dificultades encontradas durante la realización del proyecto.....	71
6.4 Avance.....	75
6.5 Capacitación.....	75
CAPITULO VII. BIBLIOGRAFÍA.....	77
APENDICE.....	79

INDICE DE FIGURAS

Figura 1.1 Estadísticas de censo económico 2019 y Estudio sobre la demografía de los negocios 2020 y 2021	2
Figura 3.1. Modelo incremental.....	11
Figura 3.2 Árbol de predicción.....	15
Figura 3.3 Librerías de SciPi.....	30
Figura 4.1 Ventana acceso.....	33
Figura 4.2 Ventana usuarios.....	33
Figura 4.3 Ventana principal.....	33
Figura 4.4 Respaldo.....	34
Figura 4.5 Guardar el archivo como .csv.....	35
Figura 4.6 Archivos convertidos a .csv.....	35
Figura 4.7 Ventana de estadísticas.....	36
Figura 4.8 Tabla tiempo.....	37
Figura 4.9 Exportación campo fecha_hora.....	37
Figura 4.10 Tabla sucursal	38
Figura 4.11 Extracción campo fecha.....	39
Figura 4.12 Extracción campo hora.....	39
Figura 4.13 Tabla ventatickets_articulos.....	39
Figura 4.14 Reporte general de ventas por meses del año	40
Figura 4.15 Reporte general de ventas por días del mes.....	41
Figura 4.16 Reporte general de días de la semana.....	41
Figura 4.17 Reporte general de ventas por minutos.....	42
Figura 4.18 Reporta general de ventas por intervalo de horario.....	42
Figura 4.19 Reporte general de ventas por departamento.....	43
Figura 4.20 Reporte general de monto vendido por usuario.....	43
Figura 4.21 Reporte general de clientes atendidos por usuario.....	44
Figura 4.22 Reporte general de movimientos de inventario por usuario.....	44

Figura 4.23 Reporte general de venta mensual de usuarios.....	45
Figura 4.24 Total venta sucursal determinada.....	45
Figura 4.25 Ventana predicciones.....	46
Figura 4.26 Consulta con campos especiales para realización de predicciones	47
Figura 4.27 Información del archivo.....	48
Figura 4.28 Estadísticas básicas.....	48
Figura 4.29 Agrupamientos.....	49
Figura 4.30 Valores más bajos.....	50
Figura 4.31 Valores más altos.....	50
Figura 5.1 Importación de tablas.....	51
Figura 5.2 Tabla tiempo.....	52
Figura 5.3 Tabla sucursal.....	52
Figura 5.4 Diagrama conceptual.....	53
Figura 5.5 Diagrama conceptual final del Data warehouse.....	54
Figura 5.6 Diagrama estrella	62
Figura 5.7 Cubo de dato 1.....	63
Figura 5.8 Cubo de dato 2.....	63
Figura 5.9 Cubo de datos 3.....	64
Figura 6.1 Reporte semestral.....	67
Figura 6.2 Estadísticas actuales de la empresa.....	67
Figura 6.3 Ventas del personal.....	68
Figura 6.4 Concentrado general.....	69
Figura 6.5 Ventas diarias.....	69
Figura 6.6 Cortes mensuales	70
Figura 6.7 Artículos diarios vendidos.....	70
Figura 6.8 Concentrado sueldos personal.....	71
Figura 6.9 Duplicado de encabezado.....	72
Figura 6.10 Tabla usuarios incompleta.....	72
Figura 6.11 Eliminación tabla usuarios.....	73
Figura 6.12 Importar tabla usuarios.....	73

Figura 6.13 Reimportación tabla usuarios.....	74
Figura 6.14 Encabezados incorrectos en tabla usuarios.....	74
Figura 6.15 Cambio de encabezado tabla usuarios.....	75

ÍNDICE DE TABLAS

Tabla 2.1 Proyectos que implementan Data warehouse.....	7
Tabla 5.1 Diccionario tabla de dimensión usuarios.....	55
Tabla 5.2 Diccionario tabla de dimensión cajas.....	55
Tabla 5.3 Diccionario tabla de hechos ventas_diarias.....	55
Tabla 5.4 Diccionario tabla de dimensión tiempo.....	57
Tabla 5.5 Diccionario tabla de dimensión productos.....	57
Tabla 5.6 Diccionario tabla de hechos ventas_articulos.....	58
Tabla 5.7 Diccionario tabla de dimensión departamentos.....	60
Tabla 5.8 Diccionario tabla de dimensión sucursal.....	60
Tabla 5.9 Diccionario tabla de dimensión clientes.....	60

I. INTRODUCCIÓN

1.1 ANTECEDENTES

Derivado de la reciente pandemia que atacó a nivel mundial por el virus COVID-19, muchos negocios tuvieron que cerrar sus puertas de manera obligatoria por parte de orden de gobierno a causa del riesgo que originaba seguir atendiendo y no ser negocios esenciales para la vida, algunos cerraron solo de manera temporal y aún así sufrieron pérdida en ventas y merma en mercancía, mientras que otros tuvieron que cerrar de manera permanente al no poder afrontar la situación que se presentaba. Con base a lo anterior se encontraron los siguientes hallazgos de acuerdo al Instituto Nacional de Estadística y Geografía (INEGI):

A 27 meses de concluidos los Censos Económicos 2019, el Instituto Nacional de Estadística y Geografía (INEGI) da a conocer los resultados del Estudio de Demografía de las Empresas (EDN) 2021 con el fin de hacerlos públicos y dar continuidad al marco analítico establecido por el Estudio de este tipo realizado en 2020. (INEGI, 2021).

Según los Censos Económicos de 2019, el 99.8% de las empresas del país son micro, pequeñas o medianas. El 99.8% de las empresas del país son pequeñas, medianas o microempresas, según los Censos Económicos de 2019. Estas unidades económicas suelen exhibir más cambios que las grandes corporaciones en términos de ingresos, personal empleado, ubicación, cierres y aperturas, entre otros factores. (INEGI, 2021).

Según la EDN 2020, se sumaron 619.443 nuevos establecimientos a los 4,9 millones totales, mientras que 1.010.857 cerraron sus puertas de forma definitiva. Según la EDN 2020, se sumaron 619.443 nuevos establecimientos a los 4,9 millones totales, mientras que 1.010.857 cerraron sus puertas de forma definitiva. De acuerdo a la EDN 2021 se estima que han nacido 1 millón de personas y han muerto 1 millón. Esto significa que entre octubre de 2020 y julio de 2021 nacieron más personas de las que murieron, en comparación con el periodo comprendido entre mayo de 2019 y septiembre de 2020. Los servicios privados no financieros tuvieron el mayor porcentaje de muertes de establecimientos en la EDN 2020, con un 24,92 por ciento, seguido del empresarial con un 18,98 por ciento y el manufacturero con un 15% de cierres concluyentes. Se registró las siguientes proporciones de mortalidad de establecimientos:

servicios privados no financieros con 38.16%, seguido del comercio con 29.88% y manufacturas con 25.69%. Los servicios privados no financieros tuvieron una tasa de mortalidad de 38,16 por ciento, seguidos por los empresariales con una tasa de 29,88 por ciento y las manufacturas con una tasa de 25,69 por ciento. (INEGI, 2021).

Sumado a este comportamiento normal, la emergencia sanitaria relacionada con la pandemia de COVID-19 ha provocado que los comercios reduzcan sus operaciones o incluso cierren permanentemente, como se vio con la EDN 2020, que fue publicada 17 meses después de finalizados los Censos. (INEGI, 2021).

Debido a lo anterior se puede apreciar que únicamente el 2% de los establecimientos son grandes, la mayor parte son negocios chicos o medianos, negocios que dan sustento a una familia, nacidos en las mismas ciudades o comunidades. De los lugares que cerraron permanentemente durante este periodo de pandemia, la tercera parte son comercios chicos o medianos, mismos que no tuvieron la posibilidad de esperar a que la situación económica se regularizara, muchos de ellos no tienen empleados, son atendidos solo por los dueños y sus familiares; los negocios que tenían empleados, tuvieron que hacer recorte de personal ya que por la baja en las ventas, muchas veces no era suficiente para pagar dichos sueldos.

A continuación en la Figura 1.1 se muestra la situación del 2019 al 2021.

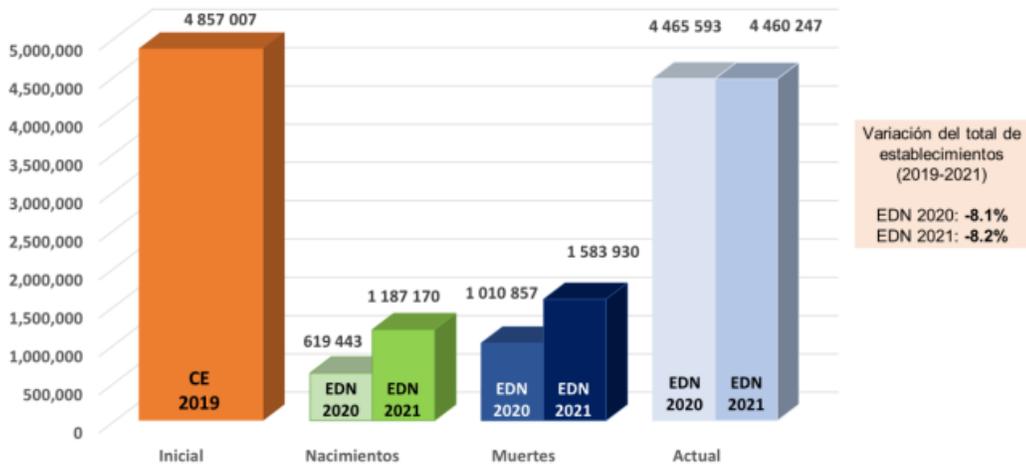


Figura 1.1 Estadísticas de censo económico 2019 y Estudio sobre la demografía de los negocios 2020 y 2021(INEGI, 2021)

Hoy en día existe software que permite conocer estadísticas o predicciones en los

negocios para conocer el panorama actual y a futuro sobre el comportamiento de las ventas, empleados, artículos, etc., dichos sistemas ayudarían en gran medida a la problemática descrita anteriormente, pero son costosos y prefieren la compra de software más económico, ese tipo de software no incluyen variedad de estadísticas ni predicciones. Los negocios no tendrían que cerrar permanentemente ni temporal por motivo de la pandemia o de la temporada baja, teniendo un sistema que proporcione los días y horarios de generalmente existe en el lugar, los dueños podrían tomar decisiones en cuanto a cambios de horarios abierto al público, así como surtir únicamente artículos de mayor demanda y otras situaciones y estadísticas.

Algunos negocios no pueden abrir los siete días de la semana debido a que genera gastos. El día de menor venta es ideal para cerrar al público. El propietario necesita atender personalmente al cliente los días de mayor afluencia de clientela, mientras que, los días a la semana de menor venta y más periodos muertos, puede dejar a una persona encargada del negocio.

En temporadas de baja venta, pandemia, cuestiones climáticas, enfermedad u otras diversas problemáticas, muchas empresas cierran sus puertas de manera temporal o definitiva; si las empresas tuvieran un sistema que realizara predicciones y gran cantidad de estadísticas en todos sus procesos, las empresas no tendrían necesidad de cerrar, solo realizar algunos cambios en sus horarios, personal, artículos, etc.

1.2 DEFINICIÓN DEL PROBLEMA

Existen negocios que cierran por periodos largos debido a la baja en sus ventas o cierran de manera permanente, habiendo en ellos, horas o días en donde si existen ventas a pesar no ser temporada, pero es difícil conocer estos datos para realizar la toma de decisiones y evitar cerrar la empresa. Con la ayuda de un Data Warehouse es posible conocer estadísticas y predicciones que ayudan de manera positiva, evitando así el cierre de la empresa.

1.3 ALCANCES Y LIMITACIONES

Alcances

Los alcances que tuvo el presente proyecto se mencionan a continuación:

- Para construir el Data warehouse se usará como base el sistema de ventas actual.
- Utilizar datos históricos de ventas, artículos, departamentos, inventario desde el año 2016 hasta 2022.
- Diseño, implementación y configuración del Data warehouse.
- Realizar predicciones sobre ventas, tipos de clientes, horarios y artículos.
- Obtener estadísticas como las siguientes: horas y días de mayor y menor afluencia de clientes, total de clientes atendidos por día, semana o mes; estadísticas referentes a ventas por categorías, tipos de clientes, días de la semana, etc.
- Identificación de artículos de mayor demanda de los clientes, para saber las áreas o departamentos que deben surtirse y cuales hay que dar de baja del inventario.
- Al conocer los tipos de artículos de mayor demanda, es posible identificar los tipos de clientes que son potenciales en la empresa.

Limitaciones

Las limitantes que ocurrieron por los cuales no fue posible realizar el proyecto como se esperaba, son las siguientes:

- En el Data warehouse se considera solamente información de ventas.
- No es posible hacer mención del nombre del sistema de ventas actual, debido a que se solicitó permiso de utilizarlo para esta investigación y no se obtuvo respuesta en su debido momento por parte de los miembros de la empresa.
- Que algunas tablas que genera el sistema de la base de datos actual, no tengan suficiente información como varias sucursales, sistema de facturación, crédito a clientes, etc.
- No hay sucursales en la empresa, por lo cual no será posible hacer datos estadísticos referentes a sucursal.
- No se maneja el sistema de facturación, por lo cual varias tablas no se contemplan.

1.4 JUSTIFICACIÓN

Con ayuda de este proyecto, será posible conocer días y horarios de mayor ventas y afluencia de clientes, los dueños de las empresas no tendrán que cerrar sus negocios por baja de temporada, podrán mantener abierto todos los meses, pero cambiando sus horarios y vendiendo únicamente los artículos o tipos de artículos que estén en la categoría con mayores ventas, gracias a las estadísticas y predicciones del sistema.

Predecir días de mayor y menor ventas y/o entrada de clientes, así como también conocer horas diarias de mayor y menor afluencia, conocer venta promedio por cliente y que el software genere variedad de estadísticas.

Las pymes de hoy en día requieren de capacitación sobre herramientas ETL que permitan el manejo de Data warehouse para la mejora en toma de decisiones. Existen empresas que ni siquiera cuenta con un sistema de ventas digital, en esos casos se requiere de inicio una capacitación sobre mejora en la administración.

1.5 OBJETIVOS

Objetivo General

Diseñar y construir un Data warehouse para análisis de datos históricos y predecir tendencias para apoyar en la toma de decisiones basadas en datos y que permitan una mejor gestión de la empresa.

Objetivos específicos

Los objetivos específicos de éste proyecto de investigación, son los siguientes:

- Utilizar herramientas ETL
- Diseña un sistema con Data warehouse
- Coordinar las consultas que permitan utilizar el Data warehouse.
- Validar el diseño del Data warehouse.

- Implementar modelos de predicción, gráficos, estadísticas que ayuden a obtener tendencias de la empresa en presente y futuro.
- Realizar consultas de interés para la empresa, sobre los diferentes criterios o necesidades.
- Determinar la o las tabla(s) de hechos.
- Realizar el diagrama conceptual de la base de datos

1.6 SUPUESTOS O HIPÓTESIS

El Data warehouse permite el análisis de datos históricos y facilita la predicción de comportamiento en las ventas, permitiendo una mejora en la toma de decisiones.

II. ESTADO DEL ARTE

Para el estudio de este contenido, fue necesaria la investigación de proyectos que involucren la realización de Data warehouse en cualquier tipo de empresa, a continuación se describen en la Tabla 2.1 algunos proyectos que implementan Data warehouse.

Tabla 2.1 Proyectos que implementan Data warehouse (Fuente: Elaboración propia)

PROYECTO	PROPUESTA DE SOLUCIÓN	RESULTADOS
<i>Implantación de un sistema de apoyo a las decisiones de venta mediante un mercado de datos basado en enfoques de OLAP, KPI y Minería de datos (Mohammed, 2022)</i>	<p><i>Etapas:</i></p> <ul style="list-style-type: none"> • <i>Pre procesamiento de los datos</i> • <i>Extracción, transformación y carga</i> • <i>Creación de cubos de ventas</i> • <i>Informes de construcción</i> • <i>Minería de datos</i> • <i>Interfaz web (Mohammed, 2022)</i> 	<i>Los datos de ventas consisten en filas obtenidas durante 2 años. El DSS de ventas propuesto proporciona una interfaz web para acceder a los informes de OLAP y KPI. (Mohammed, 2022)</i>
<i>Inteligencia empresarial aplicada a pequeñas empresas con ánimo de lucro.(D'Arconte,2018)</i>	<i>Se centra en la importancia de dos aspectos críticos, son el nivel de satisfacción de los clientes y la rentabilidad.(D'Arconte,2018)</i>	<i>Construcción de matriz satisfacción-beneficio. (D'Arconte,2018)</i>
<i>Mejore la toma de decisiones de RRHH basándose en Data Mart y OLAP. (Hamoud,2020)</i>	<i>El conjunto de datos tiene información detallada, se importa en formato CSV al SGBD SQL Server. Se diseña el Data warehouse.(Hamoud,2020)</i>	<i>Implementación OLAP basado en un conjunto de datos de empleados, puede ayudar a los responsables de la organización a mejorar los estándares, la gestión de costes, la innovación y la mejora del servicio y la calidad.(Hamoud,2020)</i>
<i>“Diseño de un prototipo de Data warehouse para una empresa de medicina prepagada” (Yaguachi y Jumbo, 2020).</i>	<i>“Técnicas de recopilación de información seleccionadas se podrá obtener información para entender las fuentes de datos e indicadores que se manejan en la actualidad en la compañía; para con los datos obtenidos proceder con el análisis y diseño de los procesos de Extracción, Transformación y Carga, así como también definir el modelo del nuevo Data warehouse” (Yaguachi y Jumbo, 2020).</i>	<i>Tablas de hechos, tablas de dimensiones, diseño de procesos ETL, implementación del sistema de Data warehouse. (Yaguachi y Jumbo, 2020).</i>

<p><i>“Sistema de soporte de decisiones con tecnología Data warehouse para la gestión de la información de la empresa Mallku Import” (Guillen, 2017).</i></p>	<p><i>“Tiene por finalidad el desarrollo de un Sistema de Soporte de Decisiones con Tecnología Data Warehouse que optimice la Gestión de la Información de la Empresa Mallku Import SAC” (Guillen, 2017).</i></p>	<p><i>“El sistema desarrollado ha optimizado la gestión de la información en un más del 83%, logrando validar y cumplir con el objetivo principal de la investigación” (Guillen, 2017).</i></p>
<p><i>“Análisis, diseño, construcción e implementación de un Data warehouse para toma de decisiones y construcción de los kpi, para la empresa Kronosconsulting Cia ltda” (Peñañiel et. al, 2013).</i></p>	<p><i>“Muestra el desarrollo de una aplicación de Inteligencia de Negocios, que incluye el análisis, diseño y construcción de un Data warehouse y Cubo multidimensional que permite el análisis de indicadores y Key Performance Indicators (KPI), aplicados a la recuperación de cartera y la eficiencia en el contacto de clientes deudores. La metodología seleccionada es HEFESTO versión 2.0, en la cual detalla todas las etapas necesarias para la construcción de un Data warehouse corporativo y los conceptos de Business Intelligence. En la etapa de construcción se trabajó con las herramientas de la suite de Microsoft SQL Management SQL 2012, Visual Studio 2010 y el componente DevExpress en la versión 13.1” (Peñañiel et. al, 2013).</i></p>	<p><i>“La teoría del Data Warehousing, Business Intelligence y la metodología Hefesto permite la implementación correcta de cada uno de los pasos para la construcción del Data warehouse, de forma estructurada y organizada. Por otra parte, un punto importante a destacar, es el costo en tiempo tomado en la fase de análisis de requerimientos y OLTP de la metodología Hefesto, por la carencia de documentación y procesos claros sobre el negocio. SATB logró muy buenos resultados al realizar la carga de la información y el posterior análisis de la información multidimensional” (Peñañiel et. al, 2013).</i></p>
<p><i>“Modelado e implementación de un Data warehouse Para un MDMS utilizando algoritmos ETL y VEE” (Mendez, 2018).</i></p>	<p><i>“Describe la estructura interna del MDMS así como las tecnologías implementadas y su justificación, también se describen algunos algoritmos que permiten realizar, optimizar y mejorar el rendimiento del MDMS, así como el funcionamiento a grandes rasgos de los sistemas que interactúan con el MDMS y como se realiza la comunicación entre los sistemas” (Mendez, 2018).</i></p>	<p><i>“Ubicación del MDMS en la Infraestructura Avanzada de Medición, arquitectura del MDMS y realización de un plan de pruebas” (Mendez, 2018).</i></p>

Para definir el estado del arte, se hace mención de la definición de un Data warehouse, “un Data warehouse es un sistema diseñado para almacenar datos antiguos y actuales que se recopilan para fuentes únicas o múltiples. Los datos recopilados se utilizan en las organizaciones principalmente para tomar la decisión o predecir el futuro” (Gravitar, 2022).

La toma de decisiones es el propósito fundamental del almacenamiento de datos en una

organización y las herramientas de consulta permiten a los usuarios interactuar con los sistemas de almacenamiento de datos. Estas herramientas se pueden dividir en herramientas OLAP, herramientas para extraer datos, herramientas para desarrollar aplicaciones y herramientas para generar informes. (Gravitar, 2022).

III. MARCO TEÓRICO

Dentro del marco teórico se describe a continuación la metodología incremental, que es el tipo de metodología que se recomienda y se utilizó para la realización de este proyecto, además se incluye el software y/o herramientas utilizadas que facilitaron su diseño e implementación.

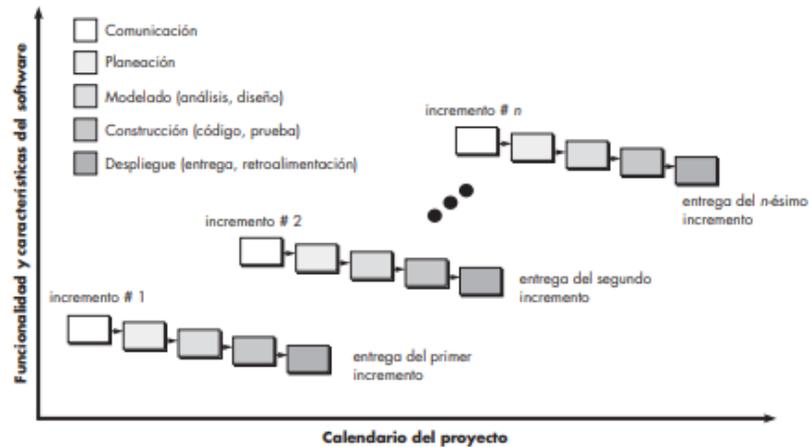
3.1 METODOLOGÍA DE DESARROLLO INCREMENTAL

La metodología de software seleccionada es el modelo de proceso incremental, debido a que se considera la más idónea porque es posible entregar incrementos en determinado sección de tiempo y de esta forma no tener que esperar hasta el final para ver un producto, sino ver varios productos, ir teniendo resultados donde sea posible ver mediante datos, gráficas, reportes, datos de interés para el cliente sin necesidad de tener que esperar hasta el término del proyecto. A continuación se describe el método antes mencionado:

Modelo de proceso incremental

Aunque los requisitos iniciales del software suelen estar bastante bien definidos, un proceso lineal suele ser imposible debido al alcance general del esfuerzo de desarrollo. Además, podría haber una necesidad urgente de proporcionar a los usuarios una cantidad limitada de funcionalidad de software de inmediato, con planes de ampliarla en próximas versiones de software. En estas situaciones, se elige un modelo de proceso creado para producir el software por etapas. (Pressman, 2010).

El modelo incremental incorpora aspectos de flujos de procesos lineales y paralelos. A medida que se desarrolla el cronograma de actividades, el modelo incremental aplica gradualmente secuencias lineales, como se muestra en la Figura 3.1. Cada secuencia lineal genera "incrementos" de software que se pueden distribuir de manera similar a los incrementos generados por el flujo de un proceso evolutivo. (Pressman, 2010).



3.1 Método incremental (Pressman, 2010)

Es típico que el primer incremento sea el producto crítico cuando se utiliza un modelo incremental. En otras palabras, se satisfacen las necesidades fundamentales, pero no se ofrecen numerosas funciones complementarias (algunas conocidas, otras no). El cliente utiliza (o realiza un análisis exhaustivo) el producto principal. Con cada incremento que se entrega, el modelo incremental implementa una serie de mejoras conocidas como incrementos que aumentan gradualmente la funcionalidad del cliente. La entrega dentro del plazo especificado por su cliente no es factible. En la fecha que solicite, proponer la entrega de uno o más incrementos; el software restante (incrementos adicionales) debe entregarse más tarde. El desarrollo de un plan para el siguiente incremento incluye ofrecer más funcionalidad y mejorar el producto principal para satisfacer mejor las necesidades de los clientes. Hasta que se complete el producto final, este proceso se repite después de entregar cada incremento. (Pressman, 2010).

El modelo de proceso incremental se centra en que en cada incremento se entrega un producto que ya opera. Los primeros incrementos son versiones desnudas del producto final, pero proporcionan capacidad que sirve al usuario y también le dan una plataforma de evaluación. (Pressman, 2010)

Cuando el personal no está disponible para implementar completamente el proyecto dentro del plazo establecido por la empresa, el desarrollo incremental es especialmente útil. Un pequeño número de trabajadores desarrolla los incrementos iniciales. Se puede agregar más personal (si es necesario) para trabajar en el siguiente incremento si el producto inicial es bien recibido. También se prevén aumentos para abordar los riesgos técnicos. Un sistema grande, por

ejemplo, podría necesitar nuevo hardware que esté en desarrollo y cuya fecha de entrega sea indeterminada. En esta situación, podría ser posible planificar los incrementos iniciales para evitar el uso de dicho hardware, brindando a los usuarios finales alguna funcionalidad sin sufrir un retraso significativo. (Pressman, 2010).

3.2 DATA WAREHOUSE

El concepto fundamental de esta investigación se centra en un Data warehouse, mismo que se describe a continuación:

“Los Data warehouse se pueden considerar como bases de datos multidimensionales que contienen grandes cantidades de datos donde los datos se organizan como un conjunto de dimensiones y tablas de hechos. Los hechos son eventos que tienen lugar en una empresa y para los cuales se requiere un proceso de toma de decisiones. Las dimensiones son propiedades que describen un hecho específico. Son los datos relativamente estáticos del DW” (Garani, 2009).

Una enorme cantidad de datos se está apoderando rápidamente del sector de las comunicaciones. Es cierto que para representar todos los datos pertinentes de llamadas, facturación y clientes, las empresas de telecomunicaciones necesitan muchas docenas de tablas relacionales. Se crea un registro de detalles de llamadas (CDR) para cada evento de usuario en la red de telecomunicaciones. El CDR contiene una amplia gama de información pertinente sobre cada llamada, como los números de teléfono de las partes que llaman y reciben, la ubicación de la torre, la hora de inicio de una llamada y su duración, así como información más específica, como el número de teléfono. número cobrado por la llamada, el monto facturado por cada llamada, el tiempo total de uso en el período de facturación y otros detalles. Es posible extraer la amplia gama de datos guardados en los CDR para buscar patrones en el comportamiento de las llamadas y el uso de las funciones del servicio. (Garani, 2009).

3.3 MACHINE LEARNING

Los sistemas que aprenden automáticamente son producidos por el aprendizaje automático, una disciplina científica en el campo de la inteligencia artificial. En este contexto, el aprendizaje implica encontrar patrones intrincados entre enormes cantidades de datos. El algoritmo que analiza datos y puede pronosticar el comportamiento futuro es la verdadera máquina de aprendizaje. Automáticamente implica que estos sistemas se desarrollan con el tiempo, independientemente de la intervención humana, también en este contexto. (<http://cleverdata>, 2021).

Big Data y Machine Learning aplicado a la empresa

Para emprender acciones comerciales que impidan que los clientes se pasen a la competencia, una compañía telefónica quiere saber qué clientes corren "peligro" de cancelar sus suscripciones a sus servicios. ¿Cómo se logra hacerlo? Antigüedad, planes contratados, uso diario, llamadas mensuales al servicio de atención al cliente, cambios recientes en los planes contratados: la empresa tiene un montón de datos de clientes, pero sin duda solo los utiliza para facturación y estadísticas. ¿Qué más se podría hacer con esa información? Se pueden utilizar para prever cuándo un cliente cancelará su suscripción y planificar el mejor curso de acción para evitarlo. En otras palabras, el aprendizaje automático le permite pasar de ser reactivo a proactivo. Cuando todos los datos históricos de los clientes se organizan y procesan adecuadamente en masa, se crea una base de datos que puede usarse para predecir comportamientos futuros, favorecer los comportamientos positivos y evitar los negativos. (<http://cleverdata>, 2021).

Es imposible que una sola persona analice esta gran cantidad de datos para sacar conclusiones y mucho menos hacer predicciones. Por otro lado, utilizando las variables que proporcionamos, los algoritmos pueden identificar patrones de comportamiento y determinar cuáles, en este caso, provocaron que un cliente se diera de baja. En la Figura 3 se muestra un ejemplo de una predicción condensada utilizando una herramienta de aprendizaje automático real e información de una compañía telefónica hipotética. (<http://cleverdata>, 2021).

Puede observar los patrones utilizados por clientes específicos que se han dado de baja utilizando la vista de árbol. En este caso, se resalta una de las ramas centrales, resaltando un patrón específico del cliente. (<http://cleverdata>, 2021).

- Se ha llamado al servicio de atención al cliente más de tres veces.
- Establece un máximo diario de 171,95 minutos para llamadas.
- Por la noche, las llamadas duran menos de 189,02 minutos. (<http://cleverdata>, 2021).

¿Dónde está la predicción cuando sólo hay un análisis de datos históricos? Si los clientes con estos rasgos ya han abandonado el negocio, es lógico que aquellos que todavía son clientes y exhiben el mismo comportamiento pronto sigan su ejemplo. Esta predicción tiene un nivel de confianza, en este caso 91 punto 97 por ciento, según el modelo predictivo, lo que indica que es bastante probable. Si el departamento de marketing tuviera acceso a estos datos, podría sugerir de forma proactiva un cambio en la estructura de precios o examinar las razones por las que los clientes llamaron al servicio de atención al cliente en un esfuerzo por retenerlos. (<http://cleverdata>, 2021).

El árbol de predicción completo se vería así. En este caso, se ha identificado una predicción errónea, puede ver el comportamiento típico de este grupo de clientes a la derecha de la Figura 3.2.: (<http://cleverdata>, 2021).

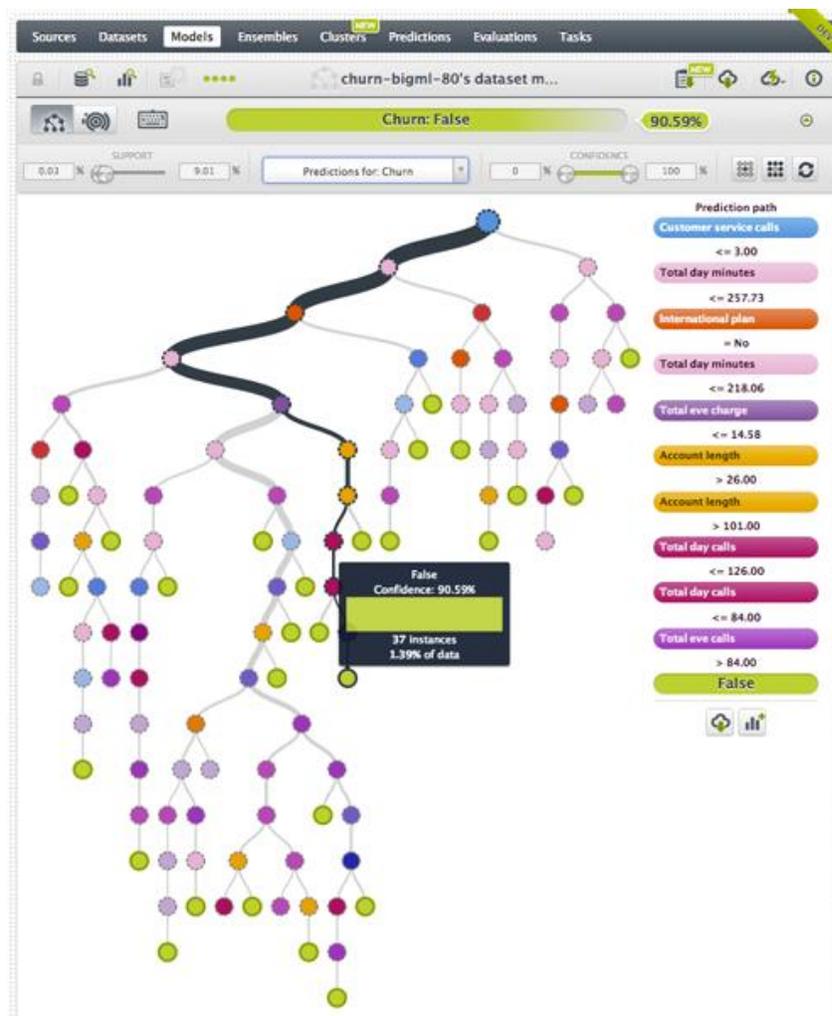


Figura 3.2 Árbol de predicción (<http://cleverdata>, 2021).

Importancia

Actualmente, las empresas están produciendo una cantidad exponencialmente mayor de datos. Sin duda una de tus ventajas competitivas es poder extraer de ellos información útil. El gran beneficio de esta tecnología es que actualmente puedes utilizarla sin necesidad de ser un experto en datos. Existen en el mercado herramientas de uso muy sencillo (incluso para profanos en análisis de datos) y asequibles económicamente para cualquier tamaño de empresa que permiten hacer predicciones como las descritas en el apartado anterior. (<http://cleverdata>, 2021).

La dificultad de utilizar los datos de forma eficaz se ha simplificado enormemente. El aprendizaje automático actual difiere de las versiones anteriores. Esto significa que actualmente es posible desarrollar modelos de comportamiento para analizar cantidades grandes y complejas

de datos si se utilizan datos de calidad, tecnologías apropiadas y análisis apropiados. Además, incluso a gran escala, los sistemas ofrecen resultados rápidos y precisos sin intervención humana. Como resultado, se produjeron pronósticos de alto valor que permitieron una mejor toma de decisiones y desarrollo de estrategias comerciales. (<http://cleverdata>, 2021).

Sin embargo, la cantidad de datos no debería desviar la atención. Para crear modelos que beneficien al negocio, no necesariamente es necesario tener acceso a tantos datos como Facebook o un banco importante. Es preferible tener datos de alta calidad (confiables y prácticos) que tener una gran cantidad de datos inútiles. (<http://cleverdata>, 2021).

Ámbitos de aplicación del Machine Learning

El aprendizaje automático se utiliza actualmente en numerosas actividades. Estas tecnologías se utilizan desde hace mucho tiempo en campos como las compras online, la publicidad online (dónde colocar un anuncio para que sea más visible según el usuario que visita la web) y los filtros antispam. (<http://cleverdata>, 2021).

La gama de aplicaciones potenciales depende de la creatividad y la información disponible dentro de la organización. Aquí hay algunos casos más. (<http://cleverdata>, 2021).

- Determinar engaños en las transacciones.
- Prever fallas en equipos en tecnología.
- Predecir qué trabajadores serán los más rentables el próximo año (la industria de recursos humanos está haciendo una apuesta importante por el aprendizaje automático).
- Elegir clientes potenciales en función de sus interacciones en línea y comportamientos en las redes sociales.
- Prever el tráfico urbano.
- Reconocer cuándo publicar actualizaciones, tweets o boletines de Facebook.
- Los diagnósticos premédicos deben realizarse a la luz de los síntomas del paciente.
- Modificar el comportamiento de una aplicación móvil para adaptarla a las preferencias y requisitos de cada usuario.

- Detectar brechas en una red para comunicaciones de datos.
- Elige el momento ideal para llamar a un cliente. (<http://cleverdata> 2021).

3.4 INTELIGENCIA DE NEGOCIOS (Business Intelligence BI)

Existe un término denominado Inteligencia de Negocios, que es necesario dentro de las empresas para la buena utilización de los recursos, mejoras en las ventas y la óptima toma de decisiones. A continuación se describen sus orígenes, definición y el proceso a seguir para implementar Business Intelligence.

Orígenes.

La frase "Business Intelligent" se le atribuye a Howard Dresner, quien la acuñó en 1989 mientras trabajaba como analista en el Instituto Gartner. Como él, los estadounidenses saltaron a la fama por la creación de herramientas de inteligencia empresarial de última generación. El Viejo Continente es la cuna y pionero en la aplicación del concepto BI, según Yves-Michel Marti, científico, profesor y fundador de Egideria, una de las mayores consultoras europeas de inteligencia empresarial. En sus investigaciones sobre economía inteligente, Martí afirma que un ejemplo muestra cómo, a finales del siglo XVI, la reina Isabel I decidió que la base del poder inglés sería "la información y el comercio" y pidió al filósofo Francis Bacon que creara un Sistema de información dinámico, que fue muy utilizado por los ingleses para poder ocupar los territorios conquistados. (Lozada, 2014).

El período anterior a la BI se puede encontrar en el pasado reciente, hace aproximadamente cuarenta o cincuenta años, en las décadas de 1960 y 1970 del siglo XX. En ese momento, las computadoras dejaron de ocupar espacios enormes y las empresas comenzaron a ver los datos como una fuente potencial e importante de información para tomar decisiones, incluso sin la financiación para un análisis de datos confiable. El entorno comenzó a cambiar en la década de 1970 con la introducción de las tecnologías DASD (Dispositivo de almacenamiento de acceso directo) y DBMS (Sistema de gestión de bases de datos) para el acceso y almacenamiento de datos. La base de datos fue considerada como un recurso fundamental para

la ventaja competitiva y a partir de ese momento la computadora pasó a ser vista como el coordinador central de las actividades corporativas (Lozada, 2014).

A principios de los años 90, la mayoría de las grandes empresas sólo contaban con Centros de Información (CI), que, a pesar de mantener un stock de datos, sólo tenían un acceso limitado a la información. Sin embargo, al ofrecer informes y análisis de gestión, los CI satisfacieron parcialmente las necesidades de los directivos. A medida que el mercado se volvió más complejo y la tecnología de la información avanzó, los programas informáticos produjeron información que no sólo era más precisa sino también oportuna para definir acciones que mejorarían el desempeño de las empresas. Entre 1992 y 1993 apareció el data warehouse, una importante base de datos de información o un depósito de datos consolidados, limpios y estandarizados. (Lozada, 2014).

Los expertos lo consideran crucial para cualquier proyecto de Business Intelligence. Este repositorio no necesariamente tiene que ser un almacén de datos; podría ser algo un poco menos complicado, como un data mart (una base de datos creada específicamente para un tema o área), o incluso una base de datos relacional estándar, dedicada a almacenar los datos utilizados como base para diversos análisis y proyecciones. En particular, alrededor de 1990, cuando la idea de inteligencia de negocios comenzó a difundirse como una evolución del EIS, Executive Information System, creado a finales de los años 1970 por investigadores del MIT (Massachusetts Institute of Technology-EE.UU.), llevó a las corporaciones a interesarse más en soluciones de BI debido al desarrollo tecnológico que permitió la creación de herramientas para facilitar la captura, extracción, almacenamiento, filtrado, disponibilidad y personalización de datos. La inteligencia empresarial ha ampliado su alcance con el tiempo para abarcar una serie de herramientas, incluido el propio EIS, soluciones DSS (Sistema de soporte a la decisión), hojas de cálculo electrónicas y generadores de consultas e informes. Todos tienen como objetivo promover la agilidad empresarial, mejorar los procesos de toma de decisiones y mejorar las estrategias de relación con los clientes. (Lozada, 2014).

Actualmente, las pequeñas, medianas y grandes empresas requieren BI para una amplia gama de propósitos, incluida la toma de decisiones, la optimización del trabajo, la reducción de costos, la previsión de crecimiento, el desarrollo de estrategias, la detección rápida de variaciones presupuestarias, la identificación de tendencias de ventas y el seguimiento eficaz de

los objetivos planificados. , accesibilidad a tablas comparativas de desempeño de empleados, asociados y colaboradores, y alineación de las operaciones en curso con objetivos de largo plazo. Se deben considerar dos factores: el costo y los objetivos a alcanzar, es decir, el proyecto debe estar en línea con los objetivos y estrategias corporativas, para que un proyecto de BI ayude a una empresa a rendir al máximo. (Lozada, 2014).

Definición.

Según Lozada (2014), la definición de Inteligencia de negocios es la siguiente:

“Business Intelligence (BI) o inteligencia de negocios, se lo define como un concepto que integra el procesamiento y el almacenamiento de grandes cantidades de datos, con el objetivo de poder transformar toda esta información en conocimiento de alto valor que permita tomar eficientemente decisiones estratégicas y en tiempo real, a través de un sencillo análisis y exploración” (Lozada, 2014).

Proceso para realizar Business Intelligence

El proceso para realizar Inteligencia de negocios según Lozada consta de 5 etapas, las cuales se describen a continuación:

ETAPA 1: Dirigir y Planear. En esta fase inicial se deben recoger las necesidades específicas de información de cada usuario, así como comprender sus distintas necesidades, para posteriormente generar con él preguntas que le ayuden a alcanzar sus objetivos.

ETAPA 2: Recolección de Información. Con el fin de recopilar los datos necesarios para encontrar las respuestas a las preguntas planteadas en la Fase 1, es aquí donde se lleva a cabo el proceso de extracción de información tanto interna como externa.

ETAPA 3: Procesamiento de Datos. Aquí es donde tiene lugar el proceso de extracción de información tanto interna como externa con el fin de recopilar los datos necesarios para encontrar las respuestas a las preguntas planteadas en la Fase 1.

ETAPA 4: Análisis y Producción. Para crear inteligencia empresarial, ahora trabajaremos en los datos extraídos e integrados utilizando herramientas y procedimientos comunes a la tecnología

BI. Al final, de esta fase se producirán informes, indicadores de desempeño, cuadros de mando, gráficos estadísticos, etc. que darán las respuestas a las preguntas.

ETAPA 5: Difusión. Finalmente, los usuarios que los necesiten recibirán las herramientas que necesitan para explorar los datos de una manera fácil e intuitiva.

3.5 METADATOS

Los metadatos, según Reddy, sirven de apoyo en el diseño del Data warehouse, se definen como: “Los metadatos son datos que ayudan en la definición de los datos disponibles en el almacén. Su propósito es generar, mantener y administrar los datos del almacén. Asimismo, ayuda en la especificación de datos según su origen, funcionalidad, valor y características. También describe el proceso de transformación y procesamiento de datos” (Reddy, 2019).

3.6 PYME

Las PYMES son consideradas como Pequeñas y Medianas Empresas, generalmente son negocios con un dueño directo y poco o ningún personal. A continuación se describe la definición y clasificación que utiliza el Banco BBVA:

Las PYMES han desempeñado un papel importante en el mundo del emprendimiento al ofrecer oportunidades de crecimiento tanto profesional como empresarial. (BBVA, 2023).

Este término se refiere a pequeñas y medianas empresas que cuentan con un número reducido de empleados y una facturación moderada. Las PYME se pueden dividir en tres categorías según un conjunto de rasgos distintivos. (BBVA, 2023).

En este sentido, crear una PYME puede ser un mejor aliado si tienes planes de poner en marcha un negocio para conseguir los objetivos empresariales que tienes en mente. Dependiendo de los objetivos, es posible seleccionar uno de sus diversos tipos, por ejemplo el que se describe a continuación. (BBVA, 2023)

Tipos de PYMES

El tamaño de las PYME, que se dividen en microempresas, pequeñas y medianas empresas, depende del número de empleados y de los ingresos anuales.

Microempresas: Describe negocios con no más de 10 empleados y con ingresos anuales de 4 punto 6 millones de pesos. (BBVA, 2023).

Pequeñas empresas: Las pequeñas empresas se definen como aquellas que ofrecen bienes o servicios y tienen entre 11 y 30 empleados o ventas anuales de 4 millones a 100 millones de pesos.

En México existen 4.1 millones de estas empresas, según datos de la más reciente Encuesta Nacional sobre Productividad y Competitividad de las Micro, Pequeñas y Medianas Empresas, realizada por el INEGI y la Secretaría de Economía. (BBVA, 2023).

Medianas empresas: Corporaciones de entre 31 y 100 empleados. También existen servicios para emplear entre 61 y 100 personas. Estas Pymes generan facturas anuales entre 100 y 250 millones de pesos. (BBVA, 2023)

3.7 HERRAMIENTAS

Los programas utilizados para la realización de este proyecto, se describen a continuación:

MySQL

MySQL fue utilizado para la importación de la base de datos, manipulación de los datos y principales consultas, a continuación se describe su definición y detalles principales de esta herramienta:

En Internet, MySQL es el sistema de gestión de bases de datos más utilizado. La mayoría de servidores y proyectos tienen acceso a este proyecto de código abierto. La mayoría de los proyectos web pueden satisfacerse con este sistema de gestión de bases de datos relacionales robusto y adaptable. (Desarrolloweb, 2023). Fue creado inicialmente por Sun Microsystems y luego adquirido por Oracle. Actualmente, MySQL se mantiene bajo dos licencias: una licencia comercial y una licencia de código abierto basada en GPL, ambas de uso gratuito para el público. Gracias a esta licencia dual, los desarrolladores originales de MySQL

podieron continuar trabajando independientemente de Oracle sin tener que depender de su guía al crear MariaDB, un sistema de gestión de bases de datos derivado de MySQL. (Desarrolloweb, 2023).

Debido a que es multiplataforma, este motor de base de datos se puede configurar en Windows, Linux o Mac. Normalmente se encuentra en servidores web que utilizan plataformas Linux y es un componente de la arquitectura LAMP (que significa Linux Apache MySQL y PHP). (Desarrolloweb, 2023).

Debido a su amplio conjunto de funciones y su capacidad para manejar cualquier tipo de proyecto, MySQL puede considerarse un software de nivel empresarial. No tiene restricciones de ningún tipo bajo su licencia comunitaria abierta. (Desarrolloweb, 2023).

Características destacadas de MySql

Las características más importantes del sistema de gestión de bases de datos MySQL son las enumeradas a continuación. (Desarrolloweb, 2023).

- Con él se crean y gestionan bases de datos relacionales.
- Utiliza numerosos motores de almacenamiento que varían en características y velocidades.
- Software de código abierto y de uso gratuito.
- Tiene una arquitectura cliente/servidor, y la instalación incluye tanto el programa servidor (el propio sistema gestor de bases de datos) como el programa cliente (vía línea de comandos).
- Sistema fácil de usar, fácil de mantener y liviano.
- Es estable y seguro. (Desarrolloweb, 2023). (Ver apéndice).

Xampp

La herramienta XAMPP es el conjunto de Apache, MariaDB/MySQL, PHP, Perl, para cualquiera de los diferentes sistemas operativos. Se define como: “XAMPP es el entorno más

popular de desarrollo con PHP. XAMPP es una distribución de Apache completamente gratuita y fácil. El paquete de instalación de XAMPP ha sido diseñado para ser increíblemente fácil de instalar y usar.

XAMPP tiene más de 10 años – hay una gran comunidad detrás del proyecto” (Apache, 2023).

Este programa fue utilizado en este proyecto para la manipulación de la base de datos y sus tablas respectivas, así como la realización de consultas de consultas.

Gliffy

Gliffy es una herramienta de diseño que se utilizó para la realización de los diagramas utilizados para el estudio de esta investigación. A continuación se describe más a detalle su uso y definición.

Gliffy, una de las primeras aplicaciones disponibles en Atlassian Marketplace, continúa brindando las integraciones más completas con las herramientas de Atlassian en este momento. Los usuarios pueden alternar entre capas e información en el visor de diagramas porque es el único creador de diagramas interactivo de Confluence. (Gliffy, 2023).

El software de diagramación de nivel empresarial de Gliffy, que cuenta con más de 18 millones de usuarios, permite a equipos de todos los tamaños crear diagramas y comunicarse visualmente. (Gliffy, 2023).

Existe una versión de prueba en la página oficial por 15 días, para este proyecto se tuvo que comprar la licencia por mes, ya que los diagramas se fueron modificando depende de las especificaciones que fueron surgiendo.

Phyton

Definición

Python es un lenguaje de programación muy utilizado hoy en día, contiene librerías que son utilizadas para la predicción de los datos, por lo cual se recomienda para este proyecto, cabe mencionar que por la premura del tiempo y seguridad de los datos, solo se realizaron algunas consultas base en este lenguaje para su estudio, no fue posible realizar todo el sistema de

predicción que se tenía en mente. A continuación se describe a detalle su definición y demás detalles.

Las aplicaciones web, el desarrollo de software, la ciencia de datos y el aprendizaje automático (ML) utilizan el lenguaje de programación Python. Debido a que Python es eficaz, fácil de aprender y compatible con una amplia gama de plataformas, los desarrolladores lo utilizan con frecuencia. La velocidad del desarrollo se acelera gracias a la disponibilidad de software Python gratuito que se integra bien con todo tipo de sistemas. (AWS, 2023).

Beneficios de Python

Las siguientes son algunas de las ventajas de Python:

Gracias a la sintaxis simple, similar al inglés, de Python, los desarrolladores pueden leer y comprender sus programas con facilidad. (AWS, 2023).

En comparación con muchos otros lenguajes, Python requiere menos líneas de código para crear un programa, lo que aumenta la productividad del desarrollador. (AWS, 2023).

Hay disponibles grandes bibliotecas estándar para Python, y estas bibliotecas contienen código reutilizable para casi cualquier tarea. Los desarrolladores pueden evitar escribir código nuevo desde cero de esta manera. (AWS, 2023).

Python se integra bien con otros lenguajes de programación conocidos, incluidos Java, C y C++, para desarrolladores. (AWS, 2023).

Hay millones de desarrolladores útiles que forman parte de la vibrante comunidad Python en todo el mundo. Si surge un problema, podrá recibir rápidamente ayuda del vecindario. (AWS, 2023).

Si desea aprender Python, Internet está repleto de muchos recursos útiles. Por ejemplo, es sencillo localizar videos, tutoriales, documentación y manuales para desarrolladores. (AWS, 2023).

Se pueden migrar diferentes sistemas operativos de computadora, incluidos Windows, macOS, Linux y Unix, para usar Python. (AWS, 2023).

Utilización de Python

Los siguientes son algunos ejemplos de casos de uso de desarrollo de aplicaciones en los que se utiliza el lenguaje Python.

Desarrollo de sitios web desde el servidor.

Las complejas operaciones de backend que llevan a cabo los sitios web para mostrar información a los usuarios se incluyen en el desarrollo web del lado del servidor. Los sitios web deben, por ejemplo, comunicarse con otros sitios web, interactuar con bases de datos y proteger las transmisiones de datos. (AWS, 2023).

Python ofrece muchas bibliotecas que contienen código preescrito para desarrollar funciones backend complejas, lo que lo hace útil para escribir código del lado del servidor. Además, los desarrolladores utilizan una amplia variedad de marcos de Python porque ofrecen todas las herramientas necesarias para crear aplicaciones web de forma rápida y sencilla. Por ejemplo, dado que no tienen que escribir la aplicación web básica desde cero, los desarrolladores pueden hacerlo en cuestión de segundos. Sin depender de herramientas de prueba externas, pueden probarlo utilizando las herramientas de prueba del marco. (AWS, 2023).

Automatización con scripts de Python

Un lenguaje de programación conocido como lenguaje de secuencias de comandos automatiza tareas que las personas completan con frecuencia. Los programadores utilizan con frecuencia scripts de Python para automatizar una variedad de tareas habituales, que incluyen:

- Se pueden cambiar el nombre de varios archivos a la vez.
- Transformar un archivo en un tipo diferente de archivo.
- Eliminar palabras redundantes de un archivo de texto.
- Ejecutar operaciones matemáticas elementales.
- Comunicaciones por correo electrónico.
- Descargar información.
- Realizar un análisis de registro simple.
- Encuentre problemas en múltiples archivos.
- Trabajar en proyectos de ciencia de datos y aprendizaje automático. (AWS, 2023).

Desarrollo de software

Los desarrolladores de software llevan a cabo muchas aplicaciones de software y tareas de desarrollo diferentes, incluidas las siguientes, que utilizan Python.:

- Monitorear la actividad de errores de software.
- Crear automáticamente el software.
- Controlar el avance del proyecto de software.
- Crear prototipos de programas informáticos.
- Crear aplicaciones de escritorio utilizando bibliotecas para interfaces gráficas de usuario (GUI).
- Crear videojuegos más complejos a partir de juegos simples basados en texto.
- Pruebas de software automatizadas. (AWS, 2023).

La prueba de software es el proceso de asegurarse de que el software esté libre de errores determinando si los resultados reales del software coinciden con los resultados esperados. (AWS, 2023).

Para probar las funciones que escriben, los desarrolladores utilizan marcos de pruebas unitarias de Python como Unittest, Robot y PyUnit. (AWS, 2023).

Los probadores de software utilizan Python para crear casos de prueba para varios escenarios de prueba. Lo utilizan, por ejemplo, para probar la interfaz de usuario de una aplicación web, diferentes partes del software y nuevas funciones. (AWS, 2023).

Hay diferentes herramientas disponibles para que los desarrolladores ejecuten scripts de prueba automáticamente. Las herramientas de integración e implementación continua (CI/CD) son el nombre que se les da a estos dispositivos. Los evaluadores de software y sus desarrolladores automatizan las pruebas utilizando herramientas CI/CD como Travis CI y Jenkins. Cada vez que los desarrolladores envían nuevos cambios de código, la herramienta CI/CD ejecuta automáticamente scripts de prueba de Python e informa los resultados de las pruebas. (AWS, 2023).

IDE de Python

El software llamado entorno de desarrollo integrado (IDE) proporciona a los programadores los instrumentos que necesitan para escribir, editar, probar y corregir código, todo en una sola ubicación. (AWS, 2023).

Los IDE de Python más conocidos son:

IDLE

El entorno de desarrollo integrado (IDE) de Python que se instala de forma predeterminada se denomina entorno integrado de desarrollo y aprendizaje (IDLE). El kit de herramientas GUI de Tkinter se utilizó exclusivamente en su desarrollo y ofrece las características que se enumeran a continuación. (AWS, 2023).

- Se ejecuta de manera eficiente en una variedad de sistemas operativos, incluidos Windows, Unix y macOS.
- La ventana del shell, que ejecuta comandos y muestra el resultado.
- Un editor de texto con múltiples ventanas que ofrece finalización automática de código y resaltado de sintaxis.
- Tener tu propio depurador. (AWS, 2023).

Spyder

Muchos científicos y analistas de datos utilizan el entorno de desarrollo integrado (IDE) de código abierto llamado Spyder. Ofrece un entorno de desarrollo completo con herramientas para análisis, visualización y depuración de datos sofisticados. Incluye las próximas características también: (AWS, 2023).

- Editor de código completo que admite múltiples idiomas.
- Una consola Python interactiva.
- Depurador básico.
- Bibliotecas especializadas en ciencia, como Matplotlib, SciPy y NumPy.
- La capacidad de investigar variables en código.
- La visualización de documentos en tiempo real es una posibilidad. (AWS, 2023).

Scikit-learn

Scikit-learn es una librería muy utilizada para la realización de predicciones en Python, se describe en este apartado ya que se pretendía incluirse en el proyecto y desarrollar una variedad de predicciones, mismas que no fueron posible por lo que se menciona anteriormente, la premura del tiempo y seguridad de los datos de la empresa. A continuación se describen más detalles sobre esta librería.

La biblioteca de Python más práctica para el aprendizaje automático es probablemente scikit-learn. Es gratuito, reutilizable en diferentes contextos y admite uso tanto académico como industrial. Se proporciona una colección de algoritmos de aprendizaje supervisados y no supervisados de Python. (Aprendeia, 2023).

Las siguientes bibliotecas o paquetes están incluidos en esta biblioteca, que está basada en SciPy (Scientific Python), como se muestra en la Figura 3.3:.

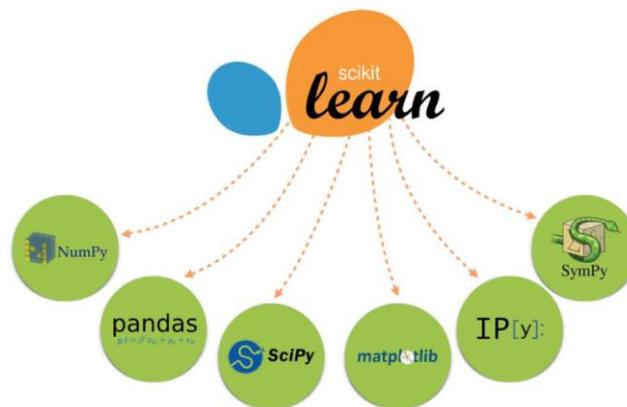


Figura 3.3 Librerías de SciPi (Aprendeia,2023)

- NumPy: librería de matriz n-dimensional base
- Pandas: estructura de datos y análisis
- SciPy: librería fundamental para la informática científica

- Matplotlib: trazado completo 2D
- IPython: consola interactiva mejorada
- SymPy: matemática simbólica” (Aprendeia, 2023).

Primero se deben importar los paquetes anteriores para poder implementar scikit-learn en la programación. Estos paquetes ya están instalados en este IDE si está utilizando Jupyter, o puede descargarlos usando la línea de comando. (Aprendeia, 2023).

Componentes de scikit-learn

Se incluyen numerosas funciones en la biblioteca scikit-learn; algunos de ellos se describen a continuación. (Aprendeia, 2023).

- **Algoritmos de aprendizaje supervisados**

Scikit-learn incluye la gran mayoría de algoritmos de aprendizaje automático que se incluyen en la categoría de clasificación de aprendizaje supervisado, incluidos modelos lineales generalizados como regresión lineal, máquinas de vectores de soporte (SVM), árboles de decisión e incluso métodos bayesianos. (Aprendeia, 2023).

- **Validación cruzada**

Hay varias formas de evaluar la precisión de los modelos supervisados y esta biblioteca contiene las instrucciones necesarias para poner en práctica rápidamente estos enfoques. (Aprendeia, 2023).

- **Algoritmos de aprendizaje no supervisados**

Esta biblioteca ofrece una gran selección de algoritmos de aprendizaje no supervisados, comenzando con agrupamiento, análisis factorial, análisis de componentes principales y redes neuronales no supervisadas. (Aprendeia, 2023).

- **Varios conjuntos de datos o dataset**

Esta biblioteca pone a disposición una serie de conjuntos de datos que son muy útiles para practicar el aprendizaje automático con Python y utilizar esta biblioteca. (Aprendeia, 2023).

- **Extracción y selección de características**

Esta biblioteca es muy útil para identificar atributos importantes a partir de los cuales construir modelos supervisados y para extraer características de texto e imágenes. (Aprendeia, 2023).

- **Comunidad**

Al igual que con scikit-learn, uno de los principales beneficios de utilizar herramientas de código abierto son sus comunidades activas. Actualmente, esta biblioteca está siendo mejorada por unos 35 colaboradores (Aprendeia, 2023).

Como se puede ver, scikit-learn es una biblioteca crucial para el aprendizaje automático; a medida que se aplique y utilice, se conocerán más detalles sobre sus características. (Aprendeia, 2023).

IV. DESARROLLO

Como se hizo mención anteriormente, la metodología seleccionada fue Incremental, donde se basa en 3 incrementos, cada uno de ellos con sus pasos respectivos: análisis, diseño, programación y pruebas. A continuación se detallan todas las actividades realizadas en el desarrollo de este proyecto de investigación, cada actividad realizada en el incremento que se indica.

4.1 SISTEMA BASE Y EXTRACCIÓN DE LOS DATOS

Análisis

Sistema base

Para la realización de este proyecto se toma como base el sistema que maneja actualmente la empresa para sus ventas diarias, para desarrollar el diseño del Data warehouse, cabe mencionar que se omite información detallada por cuestiones de privacidad. Las tablas de la base de datos del sistema son extraídas y convertidas a un formato más manejable y entendible por MySQL. Este procedimiento se describe más adelante.

Muchos negocios hoy en día tienen instalado en sus terminales este sistema, pero desafortunadamente no es posible realizar estadísticas específicas y predicciones.

Beneficios del sistema actual

- Poder registrar cada venta realizada.
- Poder dar de alta artículos, asignarlos a un departamento, realizar modificaciones o eliminarlos.
- Manejar variedad de usuarios, cambiar sus permisos de acuerdo a sus funciones.
- Ver las ventas diarias, ya sea del día actual o de un día en específico.
- Poder consultar precios específicos, sin necesidad de ir directamente a donde se encuentra el artículo físicamente.
- Poder consultar existencias de artículos.
- Poder utilizar un lector de código de barras, impresora de tickets o cajón de dinero para mejor administración del negocio.

Procesos básicos del sistema

Los procesos principales que permite realizar el sistema son los siguientes:

- Registrar venta.
- Hacer cierre, corte del día.
- Movimientos de usuarios.
- Reportes.
- Manejo de clientes.
- Productos e inventario.
- Ventas diarias y devoluciones.

Desventajas del sistema

- No muestra predicciones.
- No muestra totales diarios a detalle ni por cajero ni por departamentos o algún periodo en particular.
- No muestra estadísticas.
- No muestra graficas ni reportes.
- Está más enfocado a los abarroses.

Diseño

El diseño que se propone para la interfaz a realizar es la siguiente:

Diseño de interfaz

La interfaz básica se compone de una interfaz de acceso, como lo muestra la Figura 4.1, un menú principal donde se encuentran las opciones del sistema, (Ver figura 4.2) y la ventana usuarios que permite realizar movimientos en los cajeros o administrador, (Ver figura 4.3).



ACCESO

Nombre usuario:

Contraseña:

Aceptar

Figura 4.1 Ventana acceso

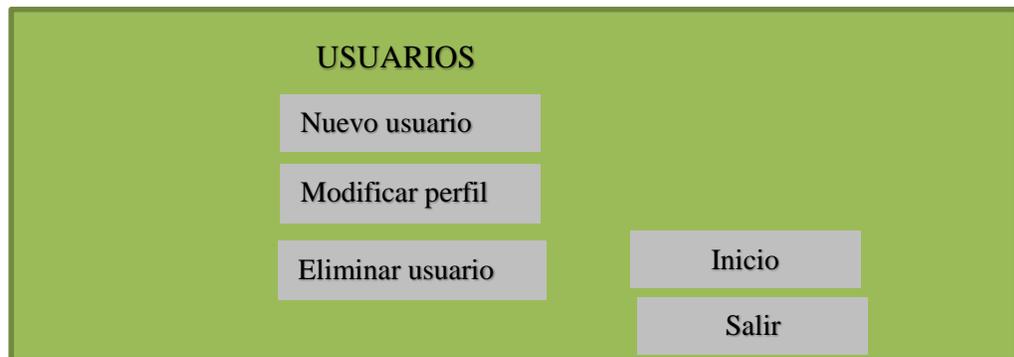


Figura 4.2 Ventana Usuarios

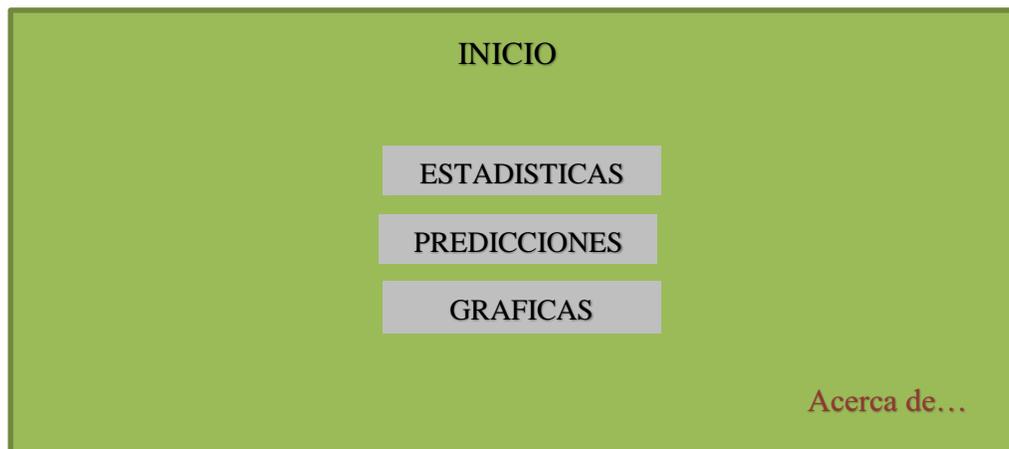


Figura 4.3 Ventana principal

Programación

Extracción de los datos

Para extraer la base de datos, hay que dirigirse al botón configuración, en la mayoría de los sistemas de punto de venta traen esta opción. Dentro del botón de configuración, se debe establecer la ruta donde se guardará periódicamente el respaldo de la base de datos del sistema, además de palomear la opción “Realizar respaldo automático al salir del programa”. La imagen del respaldo se encuentra en la Figura 4.4.

Después de conocer la ruta donde se encuentra el respaldo, se conoce la extensión del archivo, pero no es posible manipular la base de datos con esa extensión ya que no es compatible, debido a que no se puede importar directamente la base de datos en XAMPP, se dará a la tarea de convertir la base de datos a archivos de Excel mediante la siguiente página: <https://www.rebasedata.com/convert-fdb-to-excel-online>. Los datos ya convertidos se muestran en la figura 4.5.

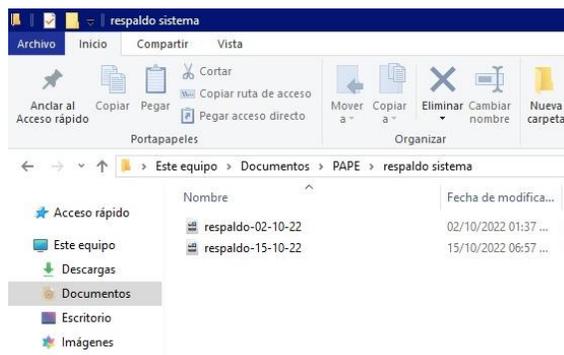


Figura 4.4 Respaldo

Es importante mencionar que para realizar el procedimiento anterior debe hacerse accediendo al sistema con la cuenta de tipo administrador.

Transformación de los datos

Mediante la transformación a Excel, es posible apreciar las tablas que contiene la base de datos del sistema base.

Debido a que la extensión .xls no es compatible con los manejadores de bases de datos que se trabajan (MySQL), es necesario convertir cada uno de los archivos de Excel a archivo .CSV,

como se muestra en la Figura 4.5. Para lo anterior, únicamente se tiene que abrir cada uno de los archivos de Excel y darle en la opción “Archivo/Guardar como” para seleccionar la opción Tipo “CSV” (delimitado por comas).

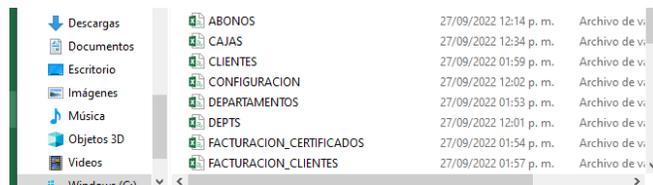


Figura 4.5 Guardar el archivo como .csv

A continuación en la Figura 4.6 se muestran los archivos ya convertidos a dicha extensión:

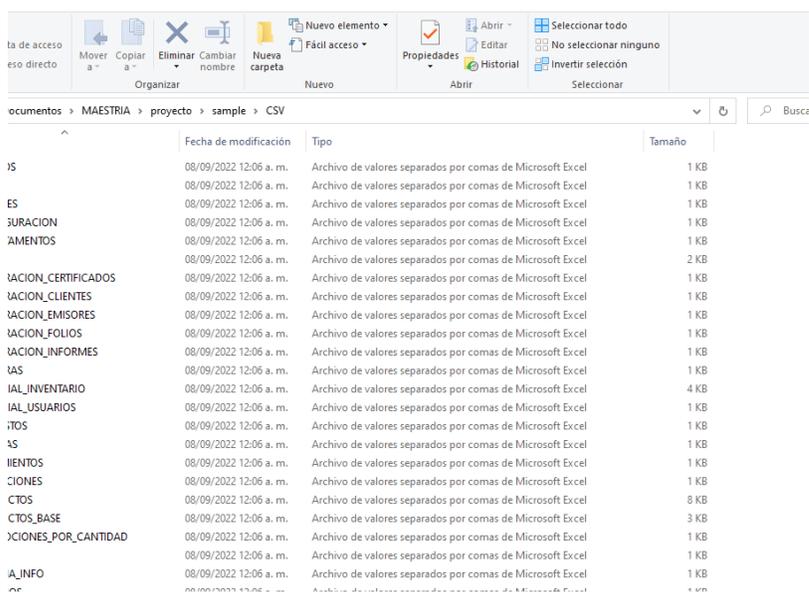


Figura 4.6 Archivos convertidos a .CSV

Pruebas

Las pruebas realizadas en este incremento se realizaron satisfactoriamente, se extrajo la totalidad de las tablas del sistema y se desarrolló su respectiva transformación de los datos.

4.2 ESTADISTICAS

Análisis

El programa donde se realizaron las consultas es XAMPP en su entorno phpmyadmin, dichas consultas se realizaron en varias dimensiones: tiempo, usuarios, artículos, ventas, departamentos y sucursales. A continuación se describen algunas de las consultas que se realizaron para conocer variedad de estadísticas de la empresa.

Diseño

En este incremento se desarrolla lo concerniente a las estadísticas de la empresa, el diseño sugerido y las principales se describen en la siguiente Figura 4.7.

ESTADISTICAS		
Descripción	Periodo	Fecha
Hora primer venta Hora última venta Mejor empleado clientes Mejor empleado venta Clientes atendidos Venta total Categoría menos vendida Categoría más vendida Artículo más vendido Artículo menos vendido Copias totales Impresiones totales Recargas Actas de nacimiento Dulcería Abarrotes Artículos escolares Regalos Didácticos Consultas Buscar artículo específico...	Diario Semanal Mensual Anual	
<input type="button" value="Visualizar"/>		

Figura 4.7 Ventana de estadísticas

Programación

Para poder realizar las consultas, fue necesario agregar 2 tablas a la base de datos: tiempo y sucursal, ya que se encuentran dentro de las tablas de hechos y por lo mismo son de suma importancia para la realización de las consultas ya que sin ellas no sería posible realizar las estadísticas y predicciones que se pretenden.

Agregación tablas para la implementación de Data warehouse.

Tabla tiempo

Se crea la tabla nueva y agregan los campos contenidos en el diagrama conceptual (Ver figura 5.5) en la tabla tiempo, como se aprecia en la Figura 4.8.

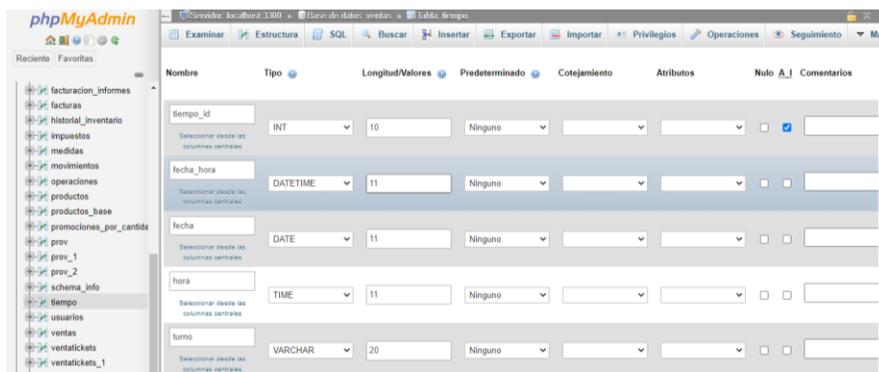


Figura 4.8 Tabla tiempo

Posteriormente se extraen los datos contenidos en la tabla ventatickets_articulos para llenar la tabla tiempo, se muestra un ejemplo a continuación en la tabla 4.9, de la primer consulta para exportar los datos de una tabla a otra.

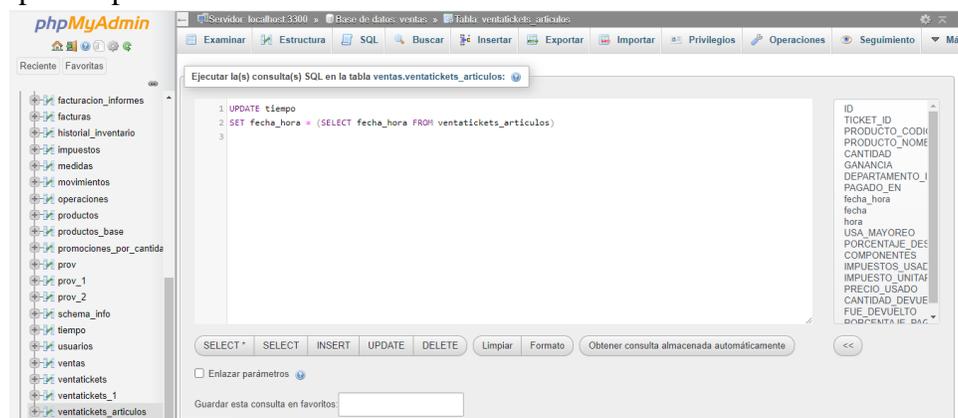


Tabla 4.9 Exportación campo fecha_hora

Además, se realizan también las siguientes consultas para la agregación de los demás

campos de la tabla “tiempo”:

Campo hora

```
UPDATE tiempo  
SET turno=IF(hora<'14:00:00','M','V')
```

Campos dia

```
UPDATE tiempo  
SET dia_mes= EXTRACT(DAY FROM fecha)
```

Campo mes

```
UPDATE tiempo  
SET mes= EXTRACT(MONTH FROM fecha)
```

Campo año

```
UPDATE tiempo  
SET anio= EXTRACT(YEAR FROM fecha)
```

Campo dia de la semana

```
UPDATE tiempo  
SET dia_semana=DAYOFWEEK(fecha)
```

Tabla sucursal

También se agregó la tabla “sucursal”, con los campos sucursal_id, nombre, dirección, usuario_id. Esta tabla es necesario agregarla ya que el sistema “Abarrotes” está diseñado solo para una sucursal y actualmente la empresa cuenta con 2 sucursales. Como lo muestra la Figura 4.10.



The screenshot shows a database management tool interface. At the top, there are menu options: Examinar, Estructura, SQL, Buscar, Insertar, Exportar, Importar, Privilegios, and Operac. Below the menu, a green status bar indicates 'Mostrando filas 0 - 1 (total de 2, La consulta tardó 0,0009 segundos)'. The main area displays a SQL query: 'SELECT * FROM `sucursal`'. Below the query, there are options for 'Perfilando' and 'Actualizar'. A table view shows two rows of data with columns: sucursal_id, nombre, direccion, and usuario_id. The first row has sucursal_id 1, nombre 'LA PAPE', direccion 'C. FRANCISCO A. NAVARRO #455, COL. PRI. CD. JIMENE', and usuario_id 0. The second row has sucursal_id 2, nombre 'BARATISIMO', direccion 'C. SOR JUANA INES DE LA CRUZ #25B, COL. BENITO JUA', and usuario_id 0. At the bottom, there are options for 'Seleccionar todo', 'Editar', 'Copiar', 'Borrar', and 'Exportar'.

sucursal_id	nombre	direccion	usuario_id
1	LA PAPE	C. FRANCISCO A. NAVARRO #455, COL. PRI. CD. JIMENE	0
2	BARATISIMO	C. SOR JUANA INES DE LA CRUZ #25B, COL. BENITO JUA	0

Figura 4.10 Tabla sucursal

Agregar campos faltantes a la tabla ventatickets_articulos

Fue necesario agregar los campos fecha, hora, que se extrajeron del campo fecha_hora de la misma tabla, para facilitar las extracciones que se realizaron para llenar la tabla “tiempo”. La Figura 4.11 muestra la extracción de la hora, la figura 4.12 muestra la extracción de la fecha y la Figura 4.13 muestra la tabla después de las extracciones.

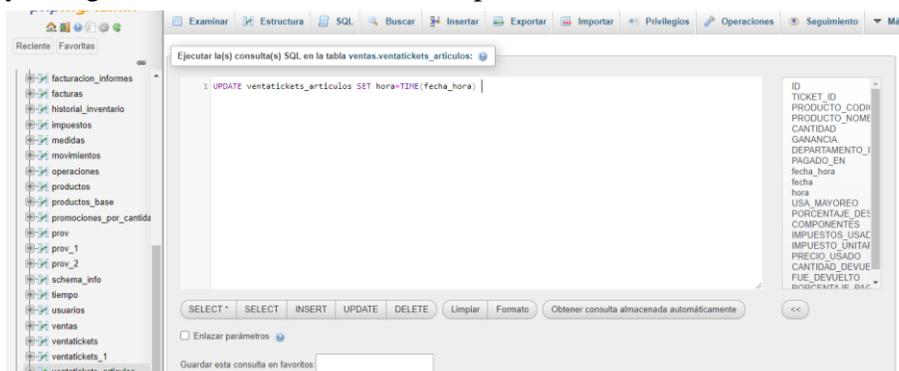


Figura 4.11 Extracción campo fecha

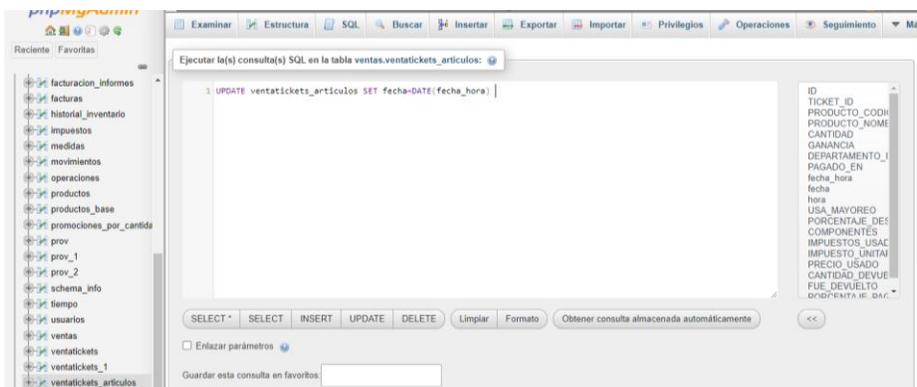


Figura 4.12 Extracción campo hora

The screenshot shows the table view of 'ventatickets_articulos' with the following data:

Q_NOMBRE	CANTIDAD	GANANCIA	DEPARTAMENTO_ID	PAGADO_EN	fecha_hora	fecha	hora	USA_MAYOREO	PORCENTAJE_DESCUENTO
	1	0	0	18.08.2019 13.33.33	2019-08-18 13:33:33	2019-08-18	13:33:33	f	0
NES COL	1	0	6	23.08.2019 07:37:09	2019-08-23 07:37:09	2019-08-23	07:37:09	f	0
NES BIN	11	1	6	23.08.2019 07:54:51	2019-08-23 07:54:51	2019-08-23	07:54:51	f	0
RTA IMP	2	1	6	23.08.2019 09:39:37	2019-08-23 09:39:37	2019-08-23	09:39:37	f	0
NES BIN	1	1	6	23.08.2019 11:05:51	2019-08-23 11:05:51	2019-08-23	11:05:51	f	0
NES COL	2	0	6	23.08.2019 11:05:51	2019-08-23 11:05:51	2019-08-23	11:05:51	f	0
RTA IMP	2	1	6	23.08.2019 11:11:54	2019-08-23 11:11:54	2019-08-23	11:11:54	f	0
	2	2.5	3	23.08.2019 12:44:11	2019-08-23 12:44:11	2019-08-23	12:44:11	f	0

Figura 4.13 Tabla ventatickets_articulos

Consultas en PHPMYADMIN

Dimensión tiempo

VENTAS TOTALES DE LOS MESES DEL AÑO

La figura 4.14 muestra el total de ventas por mes, utilizando el siguiente código para la consulta:

```
SELECT SUM(precio_usado) AS Total,  
       MONTHNAME(fecha) AS Mes  
FROM ventatickets_articulos  
GROUP BY Mes;
```

En la siguiente figura 4.14 se muestra el resultado de la consulta:

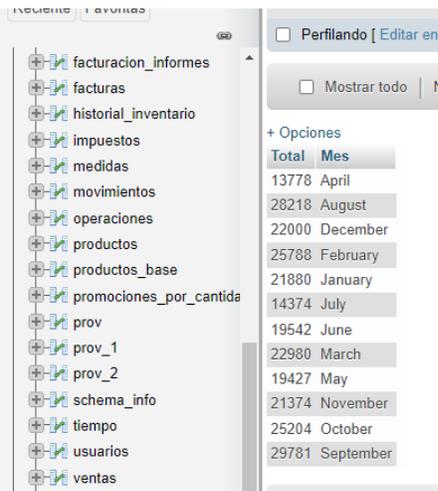


Figura 4.14 Reporte general de ventas por meses del año

VENTAS TOTALES POR DIAS DEL MES

La figura 4.15 muestra el total de ventas por días del mes, utilizando el siguiente código para la consulta:

```
SELECT SUM(precio_usado) AS Total,  
       DAYOFMONTH(fecha) AS Dia  
FROM ventatickets_articulos GROUP BY Dia;
```

En la siguiente figura 4.15 se muestra el resultado de la consulta:

	Total	Dia
facturacion_informes	7506	1
facturas	9426	2
historial_inventario	7800	3
impuestos	8801	4
medidas	9549	5
movimientos	6786	6
operaciones	7377	7
productos	7130	8
productos_base	7939	9
promociones_por_cantida	7832	10
prov	7852	11
prov_1	8645	12
prov_2	10424	13
schema_info	7836	14
tiempo	5514	15
usuarios	7148	16
ventas	10419	17
ventatickets	8279	18
ventatickets_1	8938	20
ventatickets_articulos	8430	21
prov_2	8899	22
schema_info	11202	23
tiempo	10268	24
usuarios	8660	25
ventas	8660	25
productos		+ Opciones
productos_base		Total Dia
promociones_por_cantida	8158	26
prov	8750	27
prov_1	10345	28
prov_2	10945	29
schema_info	9061	30
tiempo	6504	31
usuarios		

Figura 4.15 Reporte general de ventas por días del mes

VENTAS TOTALES DIAS DE LA SEMANA

La figura 4.16 muestra el total de ventas por días de la semana, utilizando el siguiente código para la consulta:

```
SELECT SUM(precio_usado) AS Total,
DAYNAME(fecha) AS Dia
FROM ventatickets_articulos GROUP BY Dia;
```

En la siguiente figura 4.16 se muestra el resultado de la consulta:

Total	Dia
28226	Friday
46554	Monday
29468	Saturday
36068	Sunday
38640	Thursday
44229	Tuesday
41161	Wednesday

Figura 4.16 Reporta general de días de la semana

VENTAS TOTALES POR MINUTOS

La figura muestra el total de ventas por determinado minuto de la hora, utilizando el siguiente código para la consulta:

```
SELECT SUM(precio_usado) AS Total,
MINUTE(hora) AS Minuto FROM ventatickets_articulos
GROUP BY Minuto;
```

En la siguiente figura 4.17 se muestra el resultado de la consulta:

Tabla	Total	Minuto
facturacion_informes	4970	0
facturas	5615	1
historial_inventario	5233	2
impuestos	5760	3
medidas	3558	4
movimientos	4771	5
operaciones	4291	6
productos	3852	7
productos_base	4859	8
promociones_por_cantida	4727	9
prov	3741	10
prov_1	3934	11
prov_2	4150	12
schema_info	4552	13
tiempo	4091	14
usuarios	4308	15
ventas	5484	16
ventatickets	3067	17
ventatickets_1	4230	18
ventatickets_articulos	4467	19
	3936	20

Figura 4.17 Reporte general de ventas por minutos.

MONTO DE VENTA EN INTERVALO DE HORARIO

La figura 4.18 muestra el total de ventas por intervalo de hora establecido, utilizando el siguiente código para la consulta:

```
SELECT SUM(precio_usado)
FROM ventatickets_articulos
WHERE hora >'07:00:00' AND hora <'08:00:00'
```

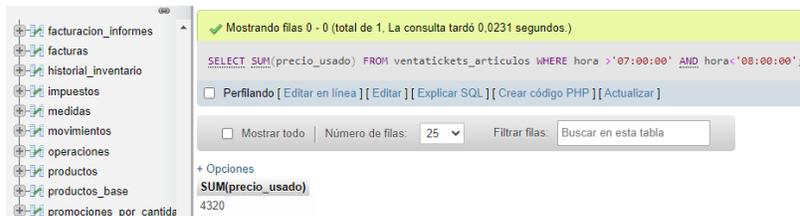


Figura 4.18 Reporte general de ventas por intervalo de horario

Dimensión departamentos

RELACIÓN DE MONTO DE VENTA POR DEPARTAMENTOS

```
SELECT SUM(ventatickets_articulos.PRECIO_USADO) AS Monto, departamentos.nombre
FROM ventatickets_articulos, departamentos
WHERE ventatickets_articulos.DEPARTAMENTO_ID=departamentos.ID
GROUP BY departamentos.nombre;
```

En la siguiente figura 4.19 se muestra el resultado de la consulta:

	Monto	nombre
<input type="checkbox"/>	114	CABALLERO
<input type="checkbox"/>	269	cosmeticos
<input type="checkbox"/>	226	CUADERNILLOS DIDACTI
<input type="checkbox"/>	397	DAMA
<input type="checkbox"/>	2481	DULCES
<input type="checkbox"/>	17930	ESCOLARES
<input type="checkbox"/>	1403	FIESTA
<input type="checkbox"/>	689	JUEGOS DIDACTICOS
<input type="checkbox"/>	2037	JUGUETES
<input type="checkbox"/>	85	MASCOTA
<input type="checkbox"/>	760	MERCERIA
<input type="checkbox"/>	795	MOCHILAS
<input type="checkbox"/>	703	PAPAS
<input type="checkbox"/>	21342	PAPELERIA
<input type="checkbox"/>	693	Salud
<input type="checkbox"/>	8531	SERVICIOS
<input type="checkbox"/>	683	TECNOLOGIA

Figura 4.19 Reporte general de ventas por departamento

Dimensión usuarios

TOTAL DE MONTO VENDIDO POR USUARIO

```
SELECT SUM(total) AS Monto, usuarios.nombre_completo
FROM ventatickets, usuarios
WHERE ventatickets.cajero_id=usuarios.id
GROUP BY usuarios.nombre_completo;
```

En la siguiente figura 4.20 se muestra el resultado de la consulta:

Monto	nombre_completo
135548	Administrador de la Tienda
2194	ana karen lopez diaz
9090	Angel Omar Moreno Lopez
912	ANGELA MUNOZ LOZANO
21084	Aresly Niño
4362	Azalia Rosas
62	Brisia Cordero
1434	Dania Gisella
2038	Dariana Gutierrez
49872	Denise Ogaz
740	Denisse Guadalupe Cruz Trejo
4490	estrella sarahi hernandez cordero
142	EVELYN BERENICE VALLES ZAPIEN
34696	fernando
2076	Guillermo Jimenez Martinez
12408	ISAI
37232	karla fernanda
21406	Maria del Carmen Ramos Martinez
810	marisol
40344	NOHELY

Figura 4.20 Reporte general de monto vendido por usuario

TOTAL DE CLIENTES ATENDIDOS POR USUARIO

```
SELECT COUNT(ventatickets.cajero_id) AS Clave, usuarios.nombre_completo
FROM ventatickets, usuarios
WHERE ventatickets.cajero_id=usuarios.id
GROUP BY ventatickets.CAJERO_ID;
```

En la siguiente Figura 4.21 se muestra el resultado de la consulta:



Clave	nombre_completo
104	Dania Gisella
22	EVELYN BERENICE VALLES ZAPIEN
3120	silvia
540	ISAI
2538	fernando
2382	NOHELY
106	ana karen lopez diaz
160	Dariana Gutierrez
292	Azalla Rosas
632	Angel Omar Moreno Lopez
52	Denisse Guadalupe Cruz Trejo
254	Yolanda Alvidrez Varela
1150	Aresly Niño
218	estrella sarahi hernandez cordero
3550	Denise Ogaz
3528	Rodolfo Laredo
1196	ULISES LOPEZ VILLALOBOS
4180	ramiro
5956	Administrador de la Tienda

Figura 4.21 Reporte general de clientes atendidos por usuario

MOVIMIENTOS DE INVENTARIO POR USUARIO

```
SELECT COUNT(usuario_id) AS Clave, usuarios.nombre_completo
FROM historial_inventario, usuarios
WHERE historial_inventario.usuario_id=usuarios.id
GROUP BY historial_inventario.USUARIO_ID;
```

En la siguiente figura 4.22 se muestra el resultado de la consulta:



Clave	nombre_completo
18	Rodolfo Laredo
82	Administrador de la Tienda

Figura 4.22 Reporte general de movimientos de inventario por usuario

Dimensión artículos

MOVIMIENTOS DE ARTÍCULOS EN INVENTARIO

```
SELECT COUNT(CODIGO_PRODUCTO) AS "Cantidad movimientos",
productos.DESCRIPCION
FROM historial_inventario, productos
WHERE historial_inventario.CODIGO_PRODUCTO=productos.CODIGO
GROUP BY productos.DESCRIPCION;
```

En la siguiente figura 4.23 se muestra el resultado de la consulta:

Cantidad movimientos	DESCRIPCION
1	BOLAS DE CHILE
1	BOLIGRAFO MAINSTAYS 4 PZ
1	CACHETADA
1	CARTONCILLO BLANCO
1	CARTONCILLO COLOR
1	CARTULINA BLANCA
2	CARTULINA COLOR PAST
1	CARTULINA FOSFORESCENTE
3	CHICLE AIRMINTS
1	CHOCOLATE BONOBON
1	COLORES BLANCA NIEVES 12 PZ
1	COLORES CRAYOLA 12 PZ
1	COLORES CRAYOLA 24 PZ
1	COPIA CARTA IMPRESORA
1	COPIA POR LOS DOS LADOS
1	IMPRESIONES B/N
1	IMPRESIONES COLOR
1	OBLEAS
1	PALETA DE CORAZON CUPIDO
1	PALETA DE CORAZON GRANDE
1	PAPEL BOND LISO

Figura 4.23 Reporte general de movimientos de artículos en inventario

Actualmente en la empresa existen muy pocos movimientos en inventario debido a que no existe una persona encargada específicamente para esa tarea, es por esto que aparecen muy pocos movimientos al momento de la consulta.

TOTAL DE ARTICULOS EN INVENTARIO POR DEPARTAMENTO

```
SELECT SUM(cantidad) AS Cantidad,
FROM ventatickets_articulos
GROUP BY DEPARTAMENTO;
```

En la siguiente Figura 4.24 se muestra el resultado de la consulta:

	Cantidad	Departamento
<input type="checkbox"/>	54	ABARROTOS (Eliminado 08/11/2021)
<input type="checkbox"/>	25	JUEGOS DIDACTICOS
<input type="checkbox"/>	104	FIESTA
<input type="checkbox"/>	354	MERCERIA
<input type="checkbox"/>	5	MOCHILAS
<input type="checkbox"/>	8	CUADERNILLOS DIDACTI
<input type="checkbox"/>	79	PAPAS
<input type="checkbox"/>	508	ABARROTOS
<input type="checkbox"/>	146	JUGUETES
<input type="checkbox"/>	2	MASCOTA
<input type="checkbox"/>	99	Salud
<input type="checkbox"/>	9540	PAPELERIA
<input type="checkbox"/>	1994	ESCOLARES
<input type="checkbox"/>	85	ACCESORIOS
<input type="checkbox"/>	17661	SERVICIOS
<input type="checkbox"/>	88	BEBE
<input type="checkbox"/>	11	DAMA

Figura 4.24 Reporte general de total de artículos en inventario por departamento

Pruebas

En esta etapa hubo algunos errores en la realización de las consultas, ya que había campos que el usuario introdujo de manera incorrecta, es decir, como “producto común”, sin seleccionarlo de la lista directa de artículos del inventario, por lo cual se tuvo que modificar manualmente los campos que tuvieran letra Ñ, números intermedios o caracteres especiales.

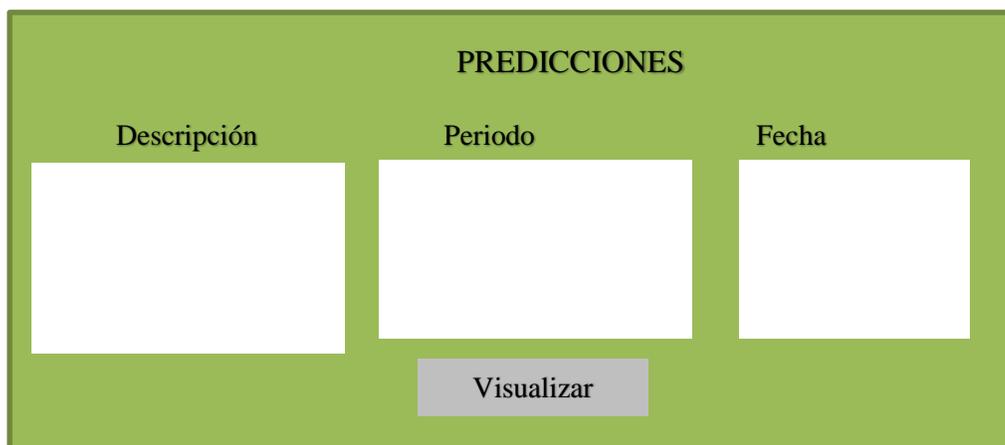
4.3 PREDICCIONES

Análisis

Para las predicciones de la base de datos se hace uso de la librería scikit-learn, también llamada sklearn. Dicha librería es actualmente usada para predicciones en programas como python, se queda acentado esta sugerencia, pero por la premura del tiempo no fue posible desarrollarlo, en su lugar se inicia en Python con extracción de datos que serán usados para desarrollar estas predicciones y con las cuales se realizan las consultas básicas que se implementan para dicha cuestión.

Diseño

En esta forma se realizan proyecciones en base a las estadísticas actuales, como lo muestra la Figura 4.25.



The image shows a web interface for making predictions. It has a green header with the title "PREDICCIONES". Below the header, there are three input fields labeled "Descripción", "Periodo", and "Fecha". At the bottom center, there is a grey button labeled "Visualizar".

Figura 4.25 Ventana predicciones

Programación

Las predicciones se realizan con ayuda de programas externos al manejador de bases de datos, en este caso Spyder (Python), se realiza una consulta con campos claves, se extraen los datos y se guarda el archivo como 'ventas_anio_mes_diasem.csv'. Como se muestra en la Figura 4.26.



Figura 4.26 Consulta con campos especiales para realización de predicciones.

Posteriormente dentro de Spyder se realiza la importación de las librerías pandas y numpy y se lee el archivo de datos:

```
#import pandas
import pandas as pd
import numpy as np
archivo=pd.read_csv('ventas_anio_mes_diasem.csv')
print(archivo)
```

Tipos de datos

```
print('Mostrando tipos de datos')
archivo.info()
print(archivo.info())
```

Lo anterior se muestra en la siguiente Figura 4.27:

```

[18456 rows x 5 columns]
Filas y columnas de los datos
(18456, 5)
Mostrar primeros registros de los datos
  PRODUCTO_NOMBRE  ANIO  MES  DIA_SEMANA  ARTICULOS_VENDIDOS
0  bolas de unice  2022    6         1           2
1  cubrebocas     2020   10         3           1
2  HOJA DE VACUNA  2021    8         5           1
3  internet busqu  2021    3         5           1
4  reparacion com  2021   10         7           1
Mostrando tipos de datos
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18456 entries, 0 to 18455
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   PRODUCTO_NOMBRE      18456 non-null  object
1   ANIO                  18456 non-null  int64
2   MES                   18456 non-null  int64
3   DIA_SEMANA           18456 non-null  int64
4   ARTICULOS_VENDIDOS   18456 non-null  int64
dtypes: int64(4), object(1)
memory usage: 721.1+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18456 entries, 0 to 18455
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype

```

Figura 4.27 Información del archivo

Estadísticas básicas

Se realizan estadísticas básicas con los datos extraídos.

```
#Obtener estadísticas básicas
print(archivo.describe())
```

Lo anterior se muestra en la siguiente Figura 4.28:

```

4  ARTICULOS_VENDIDOS  18456 non-null  int64
dtypes: int64(4), object(1)
memory usage: 721.1+ KB
None
      ANIO      MES      DIA_SEMANA  ARTICULOS_VENDIDOS
count  18456.000000  18456.000000  18456.000000  18456.000000
mean   2020.766905    6.568704    3.772269    3.440778
std     0.996315     3.525092    1.898792   10.992397
min     2019.000000    1.000000    1.000000    0.000000
25%    2020.000000    3.000000    2.000000    1.000000
50%    2021.000000    7.000000    4.000000    1.000000
75%    2022.000000   10.000000    5.000000    2.000000
max     2022.000000   12.000000    7.000000   338.000000

In [10]:

```

Figura 4.28 Estadísticas básicas

Agrupamientos

Se realizan agrupamientos de tipo determinado.

```
#Hacer agrupamientos
archivo.groupby('ANIO').mean()
print(' ')
print(archivo.groupby('ANIO').mean())
```

Lo anterior se muestra a continuación en la Figura 4.29

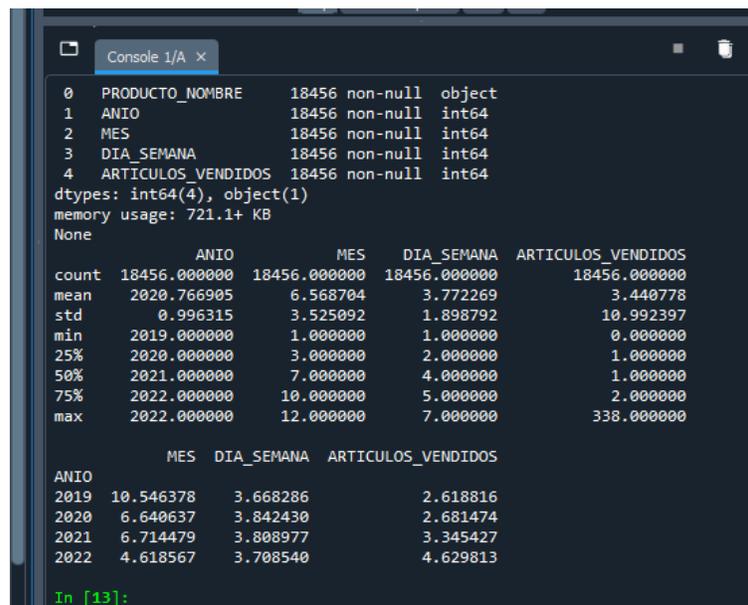


Figura 4.29 Agrupamientos

Valores más altos

Se muestran los 3 valores más altos, es decir los artículos más vendidos.

```
#Mostrar 3 valores más altos
archivo.sort_values(by='ARTICULOS_VENDIDOS',ascending=False)[['ANIO','MES','DIA_SEMANA','ARTICULOS_VENDIDOS']].head(3)
print(' ')
print(archivo.sort_values(by='ARTICULOS_VENDIDOS',ascending=False)[['ANIO','MES','DIA_SEMANA','ARTICULOS_VENDIDOS']].head(3))
```

Lo anterior se muestra en la Figura 4.30:

	ANIO	MES	DIA_SEMANA	ARTICULOS_VENDIDOS
9586	2022	6	1	338
9567	2022	3	3	326
8970	2022	1	2	277

Figura 4.30 Valores más altos

Valores más bajos

Se muestran los 3 valores más bajos, es decir los artículos menos vendidos.

```
#Mostrar 3 valores más bajos
archivo.sort_values(by='ARTICULOS_VENDIDOS',ascending=True)[['ANIO','MES','DIA_SEMANA','ARTICULOS_VENDIDOS']].head(3)
print(' ')
print(archivo.sort_values(by='ARTICULOS_VENDIDOS',ascending=True)[['ANIO','MES','DIA_SEMANA','ARTICULOS_VENDIDOS']].head(3))
```

Lo anterior se muestra en la siguiente Figura 4.31

	ANIO	MES	DIA_SEMANA	ARTICULOS_VENDIDOS
30	2019	9	4	0
12677	2021	1	5	0
9227	2021	4	4	1

In [17]:

Figura 4.31 Valores más bajos

Pruebas

En las pruebas realizadas se obtuvo satisfactoriamente los datos establecidos en las consultas. Es importante mencionar que por la privacidad de los datos y la premura del tiempo sólo se realizaron pruebas en la extracción de datos 'ventas_anio_mes_diasem.csv', pero es posible realizar otras extracciones de campos de datos de diversas tablas y realizar estas mismas consultas y similares, que ayude a la empresa a la toma de decisiones. Por las mismas razones anteriores, la etapa de predicciones tuvo que quedar pendiente para su realización.

V. RESULTADOS Y DISCUSIÓN.

Se obtuvieron hasta ahora varios resultados, a continuación se describen cada uno de ellos:

5.1 IMPORTACIÓN A PROGRAMAS MANEJADORES DE BASES DE DATOS

Una vez que se tienen los archivos con extensión .CSV, se procede a la creación de la base de datos e importación de las tablas, esto en los programas XAMPP y MySQL WORKBENCH, lo anterior para su posterior manejo de los datos y análisis. En la siguiente Figura 5.1 se muestra dicha importación.

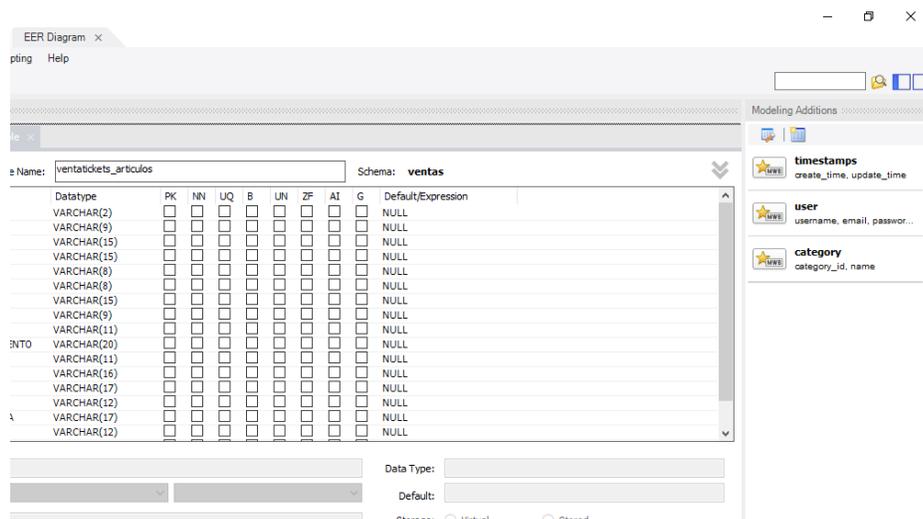


Figura 5.1 Importación de tablas

5.2 TABLAS AGREGADAS AL SISTEMA PARA MANEJO EN EL DATA WAREHOUSE

Después del análisis que se hizo a las tablas contenidas en la base de datos, se detectó que para el fácil manejo del Data warehouse es necesario agregar las tablas “tiempo” y “sucursal”.

La tabla “tiempo” con los siguientes campos: tiempo_id, fecha_hora, fecha, hora, turno, dia, dia_semana, mes, semestre, año. Esta tabla es necesaria para poder manejar la mayoría de las estadísticas y/o predicciones, mismas que ya maneja la empresa, pero todo lo realiza de

manera manual, como lo muestra la Figura 5.2:

Figura 5.2 Tabla tiempo

El llenado de esta tabla se hizo de manera automática, mediante una serie de consultas, extrayendo los datos del campo fecha_hora existentes de la tabla ventatickets_articulos, posteriormente realizar extracciones de este mismo campo, para cada uno de los siguientes campos en la tabla, como fue fecha, hora, día del mes, día de la semana, mes, año y semestre.

También se agregó la tabla “sucursal”, con los campos sucursal_id, nombre, dirección, usuario_id. Esta tabla es necesaria ya que el sistema Abarrotes está diseñado solo para una sucursal y actualmente la empresa cuenta con 2 sucursales. Como lo muestra la Figura 5.3.

Figura 5.3 Tabla sucursal

El llenado de esta tabla se hizo de manera manual, agregando cada sucursal existente en la empresa.

5.3 DIAGRAMAS DEL DATA WAREHOUSE

A continuación se describen los diagramas principales realizados para el diseño del

sistema.

Diagrama conceptual

El diagrama conceptual es el primero que se diseñó para el Data warehouse. Para el diagrama conceptual se determinaron 2 tablas de hechos como lo muestra la Figura 5.4, la primera es ventas_diarias, la cual se relaciona con las tablas clientes, cajas, tiempo y también se relaciona con la segunda tabla de hechos, ventas_articulos.

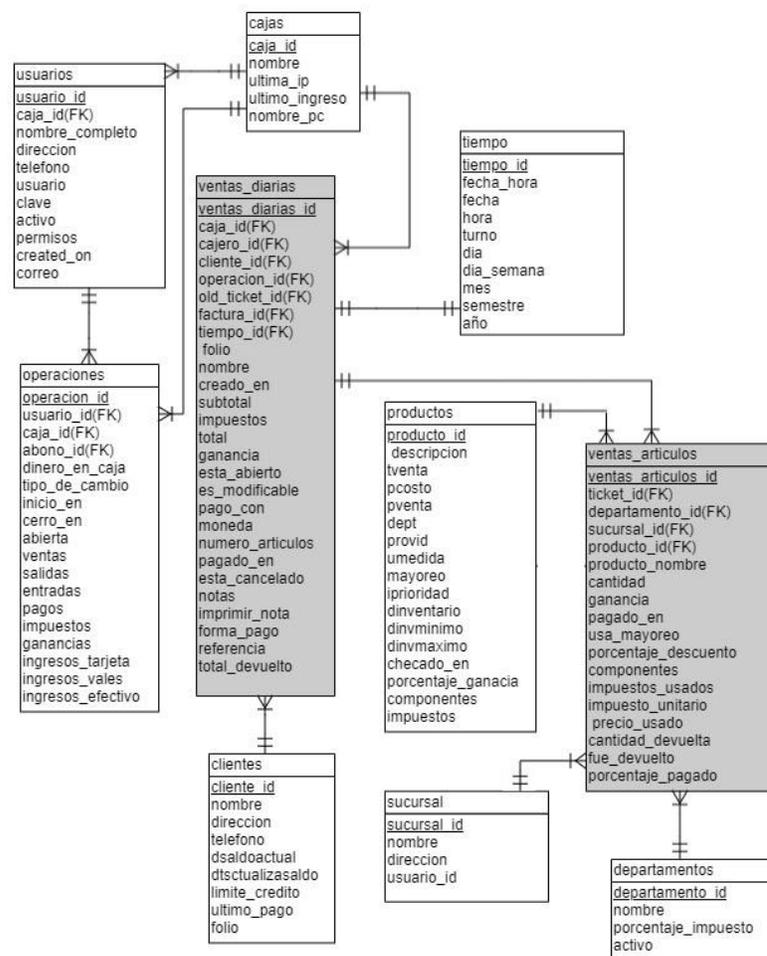


Figura 5.4 Diagrama conceptual

La segunda tabla de hechos como ya se hizo mención, es ventas_articulos que se relaciona

de manera directa con las tablas sucursal, departamentos, productos y con la otra tabla de hechos, ventas_diarias.

Las tablas de dimensión son: tiempo, usuarios, productos, departamentos, cajas, sucursal, clientes.

Para el diseño del Data warehouse hubo que hacer unos ajustes al diagrama anterior, se agregaron unas relaciones entre algunas tablas, se eliminó la tabla operaciones, ya que no existen consultas o predicciones que puedan salir de dicha tabla. Además, se eliminación una serie de campos que se considera no son útiles para la realización de las operaciones del Data warehouse. A continuación, en la Figura 5.5 se muestra el diagrama conceptual optimizado para la realización del Data warehouse:

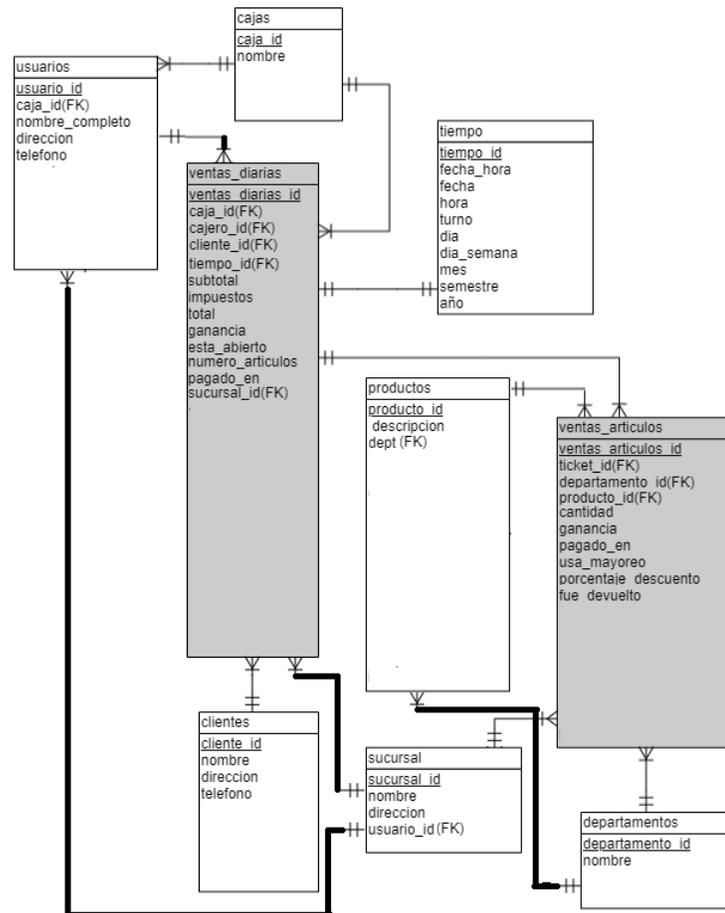


Figura 5.5 Diagrama conceptual final para el Data warehouse.

DICCIONARIOS DE DATOS

Tabla 5.1 Diccionario tabla de dimensión usuarios

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
usuario_id	VARCHAR	2	PK	Clave de identificación del cajero.	cajas ventas_diarias sucursal
caja_id	VARCHAR	3	FK	Clave de identificación de la caja atendida.	
nombre_completo	VARCHAR	34		Nombre completo del cajero(a).	
Dirección	VARCHAR	9		Dirección completa del cajero.	
Teléfono	VARCHAR	8		Teléfono del cajero.	

Tabla 5.2 Diccionario tabla de dimensión cajas

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
caja_id	VARCHAR	2	PK	Clave de identificación de la caja de venta.	usuarios ventas_diarias
Nombre	VARCHAR	14		Nombre completo de la caja de venta.	

Tabla 5.3 Diccionario tabla de hechos ventas_diarias

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
ventas_diarias_id	VARCHAR	2	PK	Clave de identificación de la venta diaria por cliente.	ventas_articulos clientes sucursal tiempo cajas usuarios
caja_id	VARCHAR	7	FK	Clave de identificación de la caja donde	

				se efectúa la compra.	
cajero_id	VARCHAR	9	FK	Clave de identificación del usuario que brinda la venta.	
cliente_id	VARCHAR	10	FK	Clave de identificación del cliente.	
tiempo_id	VARCHAR		FK	Clave de identificación del tiempo.	
Subtotal	VARCHAR	8		Subtotal de venta sin impuestos ni descuento.	
Impuestos	VARCHAR	9		Cantidad monetaria de impuestos cargados a la venta.	
Total	VARCHAR	5		Total monetaria de la venta.	
Ganancia	VARCHAR	5		Ganancia del artículo	
esta abierto	VARCHAR	6		Muestra si el ticket esta abierto o cerrado	
numero_articulos	VARCHAR	16		Numero de artículos vendidos en por ticket de compra.	
pagado_en	VARCHAR	9		Fecha de la	

				compra.	
sucursal_id	VARCHAR	10	FK	Clave de identificación de la sucursal.	

Tabla 5.4 Diccionario tabla de dimensión tiempo

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
tiempo_id	Int	10	PK	Clave de identificación del tiempo.	ventas_diarias
fecha_hora	Datetime			Fecha y hora de la venta.	
fecha	Date			Fecha de la venta.	
Hora	Time			Hora de la venta.	
turno	text			Turno de la venta.	
dia_mes	tinyint	4		Día del mes de la venta.	
dia_semana	tinyint	4		Número del día de la semana de la venta.	
mes	tinyint	4		Número de mes de la venta.	
semestre	tinyint	4		Número de semestre de la venta.	
año	year	4			

Tabla 5.5 Diccionario tabla de dimensión productos

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
producto_id	VARCHAR	13	PK	Clave de identificación del artículo.	venta_articulos departamentos

Descripción	VARCHAR	48		Nombre completo del artículo.	
Dept	VARCHAR	4	FK	Clave de identificación del departamento.	

Tabla 5.6 Diccionario tabla de hechos ventas_articulos

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
venta_articulos_id	VARCHAR	9	PK	Clave de identificación de venta por artículo.	ventas_diarias productos sucursal departamentos
ticket_id	VARCHAR	10	FK	Clave de identificación del ticket de compra.	
departamento_id	VARCHAR	15	FK	Clave de identificación del departamento.	
tiempo_id	INT	11	FK	Clave de identificación del tiempo.	
producto_id	VARCHAR	15	FK	Clave de identificación del producto.	
Cantidad	VARCHAR	8		Número de artículos vendidos de la misma especie.	
ganancia	VARCHAR	8		Ganancia monetaria obtenida en la	

				venta.	
pagado_en	Text			Fecha en que fue pagada la compra.	
fecha_hora	Datetime			Fecha y hora en que fue pagada la compra.	
Fecha	Date			Fecha en que fue pagada la compra.	
Hora	Time			Hora en que fue pagada la compra.	
usa_mayoreo	VARCHAR	11		Valor falso 'f' cuando no usa descuento mayoreo, valor verdadero 'v' cuando si usa descuento por mayoreo.	
porcentaje_descuento	VARCHAR	20		Porcentaje otorgado de descuento.	
precio_usado	VARCHAR	12		Precio del artículo cobrado en caja.	
fue_devuelto	VARCHAR	12		Valor falso 'f' cuando el artículo no fue devuelto a la tienda, valor verdadero 'v' cuando el	

				artículo si fue devuelto a la tienda.	
--	--	--	--	---------------------------------------	--

Tabla 5.7 Diccionario tabla de dimensión departamentos

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
departamento_id	VARCHAR	2	PK	Clave de identificación del departamento.	ventas_articulos sucursal
Nombre	VARCHAR	32		Nombre completo del departamento.	

Tabla 5.8 Diccionario tabla de dimensión sucursal

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
sucursal_id	INT	11	PK	Clave de identificación de la sucursal.	ventas_articulos clientes ventas_Diarias
nombre	VARCHAR	50		Nombre completo de la sucursal.	
dirección	VARCHAR	50		Dirección completa de la sucursal.	
usuario_id	INT	11	FK	Usuario responsable de la sucursal.	

Tabla 5.9 Diccionario tabla de dimensión clientes

Nombre	Tipo	Tamaño	Campo llave	Descripción	Relaciones
cliente_id	VARCHAR	6	PK	Clave de identificación del cliente.	ventas_diarias

Nombre	VARCHAR	6		Nombre completo del cliente.	
Dirección	VARCHAR	9		Dirección completa del cliente.	
Teléfono	VARCHAR	8		Número telefónico del cliente.	

Diagrama estrella

El diagrama estrella es el segundo diagrama diseñado para el Data warehouse, a diferencia del diagrama conceptual, el diagrama estrella contiene la tabla central “ventas_diarias” donde se relacionan las 5 tablas que se despliegan, es decir todas las tablas del diagrama; en el diagrama conceptual, existen 2 tablas de hechos y en cada tabla de hechos existen varias relaciones, es el diagrama de estrella la única relación se dirige hacia la tabla de hechos de todas las tablas del diagrama. En el diagrama de estrella visto en la Figura 5.5, la tabla de hechos es ventas_diarias de la cual se relacionan las tablas ventas, operaciones, ventas_articulos y la tabla clientes. Mediante sus respectivas clave principal y clave foráneas.

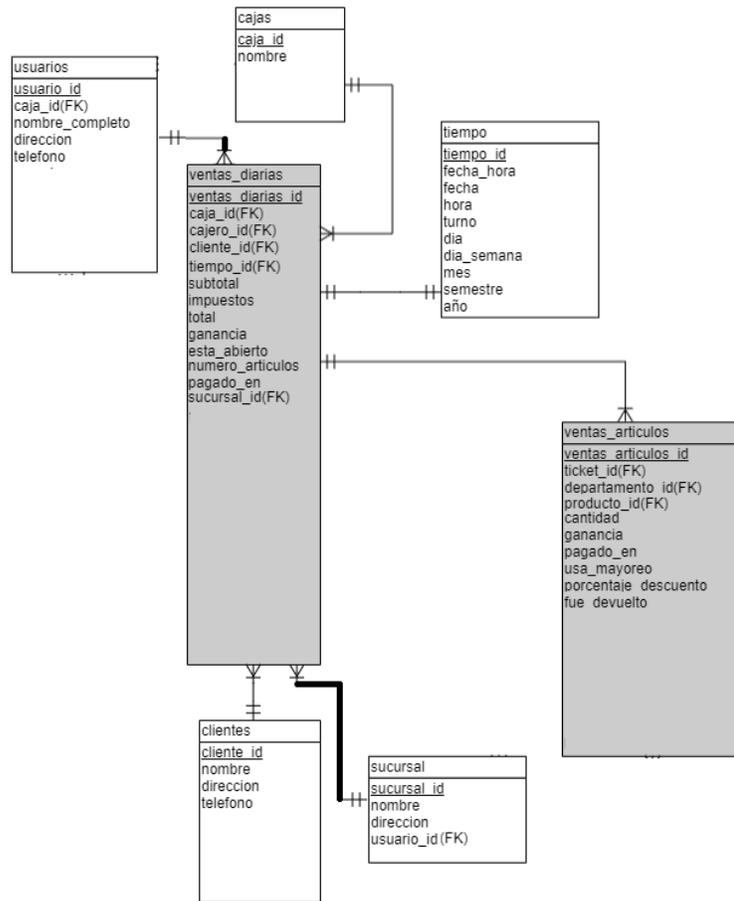


Figura 5. 6 Diagrama estrella

5.4 CUBOS DE DATOS DEL DATA WAREHOUSE

Los cubos de datos contienen dentro de las celdas datos de ventas que son importantes para conocer diferentes estadísticas de la empresa y con base a ello realizar toma de decisiones.

A continuación se muestran los cubos de datos OLAP referentes al Data warehouse que se diseña en este documento:

La Figura 5.8 muestra como coordenada en X la tabla departamentos, como coordenada en Y la tabla artículos y como coordenada en Z la tabla tiempo, es decir, es posible conocer en diversos lapsos de tiempo que departamentos se venden más, que artículos se venden

más, que artículo de cierto departamento se ha vendido tantas veces, qué artículo de cada diferente departamento es el más vendido o el que nunca ha sido vendido, etc. esto para diversas toma de decisiones sobre ampliar inventario o dar de baja.

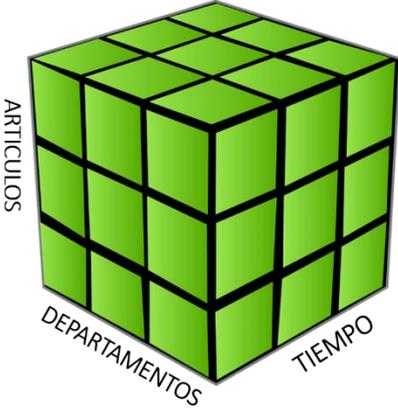


Figura 5.7 Cubo de dato 1

La Figura 5.9 muestra como coordenada en X la tabla artículos, como coordenada en Y la tabla usuarios y como coordenada en Z la tabla departamentos, es decir, es posible conocer en diversos departamentos que artículos venden más cada usuario, que artículo de cierto departamento se ha vendido tantas veces, que usuario tiene poca o mucha venta en determinado artículo, qué artículo de cada diferente departamento es el más vendido y que usuario lo vende o el usuario que menos ventas tiene de artículos en algún departamento, etc. esto para diversas toma de decisiones sobre ampliar inventario o dar de baja.

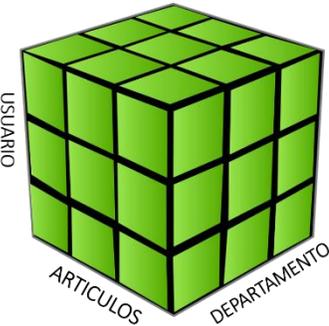


Figura 5.8 Cubo de dato 2

La Figura 5.10 muestra como coordenada en X la tabla artículos, como coordenada en Y la tabla sucursal y como coordenada en Z la tabla tiempo, es decir, es posible conocer en diversos lapsos de tiempo que sucursales venden más, que sucursales venden menos, que

artículos se venden más, que artículo de cierto departamento se ha vendido tantas veces, qué artículo de cada diferente departamento es el más vendido o el que nunca ha sido vendido, esto para diversas toma de decisiones sobre ampliar inventario o dar de baja.

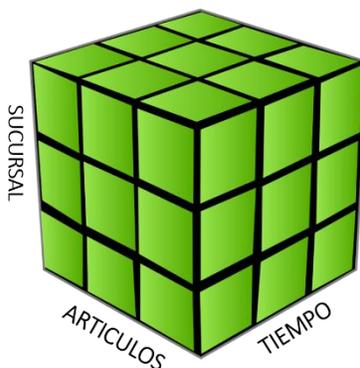


Figura 5.9 Cubo de datos 3

5.5 CAPACITACIÓN A LA EMPRESA

El personal de la empresa fue capacitado de manera satisfactoria en el diseño y utilización de herramientas ETL que permiten el análisis de los datos, extracción y manejo de los mismos. El proceso de capacitación consistió en lo siguiente:

1. Detección de necesidades. Es importante conocer que procesos requieren de apoyo en la empresa, ya que determinará la mayoría de las consultas y estadísticas del Data warehouse.
2. Aspectos informáticos. Es necesario conocer si la empresa cuenta ya con un sistema de punto de ventas o algún otro programa donde maneje su información de ventas, como puede ser Excel, Access u otro programa previamente instalado. En el caso de La empresa en cuestión, ya tenía un sistema de ventas instalado, así que fue necesario capacitar en la utilización de todo el sistema, ya que había apartados del mismo que los usuarios desconocían su uso.
3. Herramientas ETL. Consiste en la utilización de programas que permitan la extracción, transformación y uso de los datos, se dio a conocer la explicación de lo anterior.
4. Diagramas. Después de conocer las tablas contenidas en la base de datos se determinaron cuáles serían las tablas de hechos y las demás tablas que se usarían en los diferentes

diagramas, se hizo el diseño en conjunto con el personal y se dio la instrucción correspondiente.

5. Estadísticas. Se determinaron las consultas principales y se capacitó sobre la realización de las mismas así como su posterior modificación, ya que hay consultas que se pueden ejecutar con diferentes argumentos.
6. Predicciones. Se explicó sobre lo realizado en esta investigación en el apartado de predicciones, así mismo se hizo la mención que pudiera realizarse posteriormente una investigación enfocada únicamente a predicciones dada la grandeza del tema y la falta de tiempo para realizarla en este proyecto.

6 CONCLUSIONES.

6.1 SISTEMA DE VENTAS

Algunos sistemas de punto de venta existentes ya incluyen la opción de “Reportes” donde permite ver estadísticas básicas en cuanto a clientes, monto, departamentos y cajeros, esas estadísticas son de mucha ayuda a la empresa, pero no son tan específicas como la empresa base lo requiere. La empresa necesita datos más detallados, no solo de monto vendido, sino de lapso de tiempo, artículos específicos, y otras estadísticas que se describen en el apartado diseño de este documento y los hallazgos del siguiente apartado.

6.2 EMPRESA BASE

La empresa base tiene giro de papelería, actualmente maneja estadísticas y contabilidad pero lo realiza de manera manual en documentos de Excel, día tras día y semana a semana, por ejemplo, se hacen las cuentas de diversos aspectos: mejor vendedor, peor vendedor, día de la semana de menor venta, día de la semana de mayor venta, día de la semana con más clientes, día de la semana con menos clientes, monto diario o semanal vendido por cajero, clientes semanales atendidos, clientes mensuales atendidos, monto vendido por hora, clientes atendidos por hora, concentrado de ventas por hora, diarias, semanales, mensuales y anuales por cajeros y por sucursales y muchas otras más estadísticas que se llevan a cabo en estos archivos. En estos documentos de Excel, es donde se almacena la mayoría de las estadísticas de la empresa y se revisa periódicamente para determinar las estadísticas, pero realizando conteo mediante este programa. En este apartado se mencionan algunas de las estadísticas principales que la empresa maneja actualmente en Excel.

Para este proyecto se realizaron algunas de las consultas anteriores pero su realización en XAMPP, como se describe en el apartado 4.2.3. No fue posible realizar todas las consultas que se realizan actualmente en la empresa, ya que son muy numerosas y algunas de mayor dificultad que requieren más tiempo de estudio, por la premura de tiempo no fue posible.

A continuación en la Figura 6.1 se muestra el concentrado de los cálculos por semestre

por empleado en las 2 sucursales, todos estos datos se obtuvieron al contar de manera manual los datos existentes en el sistema de ventas actual.

EMPLEADO	VENTA GENERAL	CLIENTES ATENDIDOS	DIAS TRABAJADOS	VENTA DIARIA	CLIENTES POR DIA	VENTAS POR CLIENTE
DENISSE	\$21,606.00	1334	79	\$273.49	16.88607595	\$16.20
RUBEN	\$10,395.00	554	40	\$259.88	13.85	\$18.76
KARLA	\$20,706.00	959	88	\$235.30	10.89772727	\$21.59
DANIA	\$4,816.00	285	23	\$209.39	12.39130435	\$16.90
RAMIRO	\$4,799.00	162	23	\$208.65	7.043478261	\$29.62
ARELY	\$1,894.00	129	10	\$189.40	12.9	\$14.68
ESTRELLA	\$742.00	68	5	\$148.40	13.6	\$10.91
FERNANDA	\$1,805.00	52	15	\$120.33	3.466666667	\$34.71

Figura 6.1 Reporte semestral

Enseguida se muestran las ventas diarias en monto y clientes atendidos diarios de una sucursal, además se puede apreciar la hora de la primer venta y la hora de la última venta, el total en monto vendido y el total de clientes atendidos, se aprecia con letra rojo los días de menor venta en monto y menor clientes atendidos, para ir conociendo los días que sería recomendable cerrar al público por la baja en ventas, esto para saber el panorama y cambiar los horarios de atención en base a las proyecciones que se observan, además se aprecia con letra amarilla el segundo día de menor venta y menos clientes atendidos, esto para en un futuro determinar cuál día se apertura únicamente con medio turno, todo esto se calculó también de manera manual del sistema actual, ya que no hay ningún apartado en el sistema que arroje esta información. La Figura 6.2 muestra esta información.

DIA	fecha	ramir	fernan	karl	rube	dania	deniss	estrella	aresly	p. venta	u. venta	total venta	total cliente	total sem	t. clienté	retiro co
119	VIERNES 29/04/2022									11:21 a.m	07:20 p.m	177	10			
120	SABADO 30/04/2022					390	10			10:36 a.m	04:30 p.m	390	10			
121	DOMINGO 01/05/2022				50	7				10:32 a.m	04:57 p.m	50	7	2689	147	
122	LUNES 02/05/2022			619	29			358.5	27	07:38 a.m	08:34 p.m	977.5	56			
123	MARTES 03/05/2022			499.5	20			337	23	07:36 a.m	08:27 p.m	836.5	43			
124	MIERCOLES 04/05/2022							141	12	08:06 a.m	02:35 p.m	141	12			
125	JUEVES 05/05/2022				346.5	15	39	4	54	8	08:52 a.m	08:46 p.m	439.5	27		
126	VIERNES 06/05/2022			263	14					08:30 a.m	06:39 p.m	263	14			
127	SABADO 07/05/2022			50	1		160.5	13		09:11 a.m	04:53 p.m	210.5	2	14		
128	DOMINGO 08/05/2022							531	19	09:56 a.m	04:18 p.m	651	19	4072	185	
129	LUNES 09/05/2022				187.5	15		594.5	30	07:19 a.m	09:02 p.m	782	45			
130	MARTES 10/05/2022		27	2	85	6				10:55 a.m	01:29 p.m	112	2	8		
131	MIERCOLES 11/05/2022				216	15		251	13	07:48 a.m	08:00 p.m	467	28			
132	JUEVES 12/05/2022				244	13		234	23	07:31 a.m	08:07 p.m	478	36			
133	VIERNES 13/05/2022				138	4				08:02 a.m	10:22 a.m	138	4			
134	SABADO 14-may				60	4				07:38 p.m	08:06 p.m	60	1	4		
135	DOMINGO 15/05/2022				430	12				10:03 a.m	04:12 p.m	430	12	2477	137	
136	LUNES 16/05/2022				220	15		299.5	10	07:41 a.m	09:06 p.m	519.5	25			
137	MARTES 17/05/2022				516.5	17		212	15	07:48 a.m	08:46 p.m	728.5	32			
138	MIERCOLES 18/05/2022				130	13		347	31	08:06 a.m	08:12 p.m	477	44			
139	JUEVES 19/05/2022			266	5	627	17			09:56 a.m	08:24 p.m	893	22			
140	VIERNES 20/05/2022							166	9	07:55 a.m	10:10 a.m	166	2	9		
141	SABADO 21/05/2022						118	8		09:44 a.m	04:51 p.m	118	1	8		
142	DOMINGO 22/05/2022				30	3	401	18		10:57 a.m	08:07 p.m	431	21	3330	161	
143	LUNES 23/05/2022				424	28				09:28 a.m	08:17 p.m	424	28			
144	MARTES 24/05/2022				494	25		157.5	9	07:45 a.m	08:37 p.m	651.5	34			

Figura 6.2 Estadísticas actuales de la empresa

Las estadísticas anteriores de la Figura 6.2, no es posible visualizarlas en la versión instalada del punto de venta; es importante hacer mención que la última versión del sistema de

ventas en su actualización incluye ya un apartado de reportes, pero no con las características que lo requiere la empresa.

	A	B	C	D	E	F	G	H	I	J
1	mes	año	empleado	pape venta	pape cliente	bara venta	bara cliente	total venta	total cli	
55				13653	766	5537	386	19190	1152	
56	JULIO	2022	dania	2019	143	811	52	2830	195	
57	JULIO	2022	ruben	725	63	47	10	772	73	
58	JULIO	2022	karla	889	34	153	13	1042	47	
59	JULIO	2022	ramiro	60	1	85	4	145	5	
60	JULIO	2022	fernanda	30	1	0	0	30	1	
61	JULIO	2022	Aresly	2299	135	774	55	3073	190	
62	JULIO	2022	Estrella	498	29	1655	68	2153	97	
63	JULIO	2022	YOLANDA	0	0	251	9	251	9	
64				4705	406	3776	211	8481	617	
65	AGOSTO	2022	RUBEN	379	26	30	1	409	27	
66	AGOSTO	2022	DANIA	2362	74	4423	337	6785	411	
67	AGOSTO	2022	RAMIRO	2837	133	83	5	2920	138	
68	AGOSTO	2022	FERNANDA	1917	40	0	0	1917	40	
69	AGOSTO	2022	ARESLY	654	41	5698	259	6352	300	
70	AGOSTO	2022	KARLA	6926	138	3122	44	10048	182	
71	AGOSTO	2022	ESTRELLA	164	9	0	0	164	9	
72	AGOSTO	2022	YOLANDA	0	0	150	5	150	5	
73				15239	461	13506	651	28745	1112	
74	SEPTIEMBRE	2022	karla	3986	150	321	20	4307	170	
75	SEPTIEMBRE	2022	dania	1870	53	6665	393	8535	446	
76	SEPTIEMBRE	2022	ARESLY	4752	334	4982	291	9734	625	
77	SEPTIEMBRE	2022	RAMIRO	1060	37	12.5	2	1072.5	39	
78	SEPTIEMBRE	2022	FERNANDA	625	3	0	0	625	3	

Figura 6.3 Ventas del personal

En la Figura 6.3 anterior se muestra un concentrado de las ventas mensuales de cada cajero en cada sucursal, en monto y en total de clientes atendidos, además el total del mes en cada uno de estos aspectos. Con esto es fácil identificar el empleado de mayor venta; en la empresa se le otorga un porcentaje de comisión al empleado que obtuvo la mayor venta en el mes.

En la Figura 6.4 se muestra un concentrado general por mes y por sucursal tanto en monto y en clientes, mediante los filtros de Excel la empresa conoce una variedad de estadísticas, que se pretende con este trabajo de investigación, sea posible lograrlo mediante la realización del diseño del Data warehouse, no solo de todas las estadísticas que maneja la empresa de manera manual en Excel, sino también de otras estadísticas y predicciones que no maneja actualmente y se requiere conocerse para conocer la situación real de la empresa en diversos aspectos.

	A	B	C	D	E	F	G
1	AÑO	MES	SUCURSA	MONTO VENTA	CLIENTES		
2	2021	JUNIO	PAPE	6948	443		
3	2021	JUNIO	BARA	3211	163		
4	2021	JULIO	PAPE	7787	525		
5	2021	JULIO	BARA	4126	290		
6	2021	AGOSTO	PAPE	7847	412		
7	2021	AGOSTO	BARA	3888	273		
8	2021	SEPTIEMBRE	PAPE	12180	593		
9	2021	SEPTIEMBRE	BARA	5366	310		
10	2021	OCTUBRE	PAPE	11012	623		
11	2021	OCTUBRE	BARA	6148	363		
12	2021	NOVIEMBRE	PAPE	4657	270		
13	2021	NOVIEMBRE	BARA	5165	351		
14	2021	DICIEMBRE	PAPE	8693	465		
15	2021	DICIEMBRE	BARA	5293	269		
16	2022	ENERO	PAPE	10477	571		
17	2022	ENERO	BARA	6965	470		
18	2022	FEBRERO	PAPE	15638	831		
19	2022	FEBRERO	BARA	10776	625		
20	2022	MARZO	PAPE	17072	926		
21	2022	MARZO	BARA	5061	256		
22	2022	ABRIL	PAPE	10100	513		
23	2022	ABRIL	BARA	10100	513		

Figura 6.4 Concentrado general

En la siguiente Figura 6.5 se conoce la situación de venta por día de la semana en las 2 sucursales.

	A	B	C	D	E	F	G	H	I	J	K
1	EV	EC	DIA	LUGAR	MES	DIAS	CLIENTES	VENTA MENSU	VENTA PRO	CLIENTES PRO	
2	1	1	LUNES	PAPE	ENERO	5	158	2994	598.8	31.6	
3	4	3	MARTES	PAPE	ENERO	4	89	1097	274.25	22.25	
4	2	2	MIERCOLES	PAPE	ENERO	4	99	2050	512.5	24.75	
5	5	4	JUEVES	PAPE	ENERO	4	67	1088	272	16.75	
6	3	5	VIERNES	PAPE	ENERO	4	64	1260	315	16	
7	7	6	SABADO	PAPE	ENERO	4	54	953	238.25	13.5	
8	6	7	DOMINGO	PAPE	ENERO	5	50	1033	206.6	10	
9	6	6	LUNES	PAPE	FEBRERO	3	98	1791	597	32.66666667	
10	3	2	MARTES	PAPE	FEBRERO	4	137	2142	535.5	34.25	
11	2	4	MIERCOLES	PAPE	FEBRERO	4	132	2488	622	33	
12	4	1	JUEVES	PAPE	FEBRERO	4	155	2138	534.5	38.75	
13	5	5	VIERNES	PAPE	FEBRERO	4	111	2119	529.75	27.75	
14	7	7	SABADO	PAPE	FEBRERO	4	65	1259	314.75	16.25	
15	1	3	DOMINGO	PAPE	FEBRERO	4	133	3702	925.5	33.25	
16	1	1	LUNES	BARATISIMO	ENERO	5	160	2461	492.2	32	
17	2	2	MARTES	BARATISIMO	ENERO	4	79	1087.5	271.875	19.75	
18	3	3	MIERCOLES	BARATISIMO	ENERO	3	67	990	330	22.33333333	
19	4	5	JUEVES	BARATISIMO	ENERO	4	43	664	166	10.75	
20	6	4	VIERNES	BARATISIMO	ENERO	4	56	572	143	14	
21	7	7	SABADO	BARATISIMO	ENERO	4	27	548	137	6.75	
22	5	6	DOMINGO	BARATISIMO	ENERO	5	38	593	118.6	7.6	
23	3	3	LUNES	BARATISIMO	FEBRERO	3	93	1999	666.3333333	31	
24	1	2	MARTES	BARATISIMO	FEBRERO	4	152	2414.5	603.625	38	
25	2	1	MIERCOLES	BARATISIMO	FEBRERO	4	175	2373	593.25	43.75	
26	5	4	JUEVES	BARATISIMO	FEBRERO	4	78	1221	305.25	19.5	
27	6	5	VIERNES	BARATISIMO	FEBRERO	4	70	1081	270.25	17.5	

Figura 6.5 Ventas diarias

La tabla 6.6 muestra los cortes mensuales para conocer si los usuarios tuvieron diferencia de dinero en la caja y lo reportado de ventas en el sistema, se muestra con signo negativo las pérdidas de la empresa en el campo diferencia, debido a que dan mal el cambio de dinero o no registran la venta antes del corte diario y en negro las cantidades que sobraron de dinero en caja y que al usuario se le olvido registrar en el sistema.

	A	B	C	D	E	F	G	H	I	J	K	L
1	FECHA	MONTO INICIO	MONTO FIN	GANANCIA elid	ganancia rez	VENTA SISTEN	DIFERENCI	EMPLEADO	ESTATUS	retiros		
142	24/05/2022	\$ 106.00	\$ 181.00	\$ 75.00	\$ 75.00	\$ 157.50	-\$ 50.00	Denisse	Debe	32	250	
143	25/05/2022	\$ 99.00	\$ 327.00	\$ 228.00	\$ 228.00	\$ 265.00	-\$ 12.00	Denisse	Debe	25	49	
144	26/05/2022	\$ 98.00	\$ 235.00		\$ 137.00	\$ 183.50	-\$ 13.50	Denisse	Debe	46	60	
145	27/05/2022	\$ 105.00	\$ 127.00	\$ 22.00	\$ 22.00	\$ 22.00	\$ -	Denisse				
146	28/05/2022	\$ 102.00	\$ 325.00	\$ 223.00	\$ 223.00	\$ 221.00	-\$ 2.00	Dania	Olvido registrar			
147	29/05/2022	\$ 79.00	\$ 591.00	\$ 512.00	\$ 512.00	\$ 538.00	-\$ 26.00	Ruben	Debe			
148	30/05/2022											
149	31/05/2022	\$ 84.00	\$ 143.00	\$ 59.00	\$ 59.00	\$ 119.00	\$ -	Denisse	Debe	260		
150	01/06/2022	\$ 137.00	\$ 321.00	\$ 184.00	\$ 184.00	\$ 194.00	-\$ 10.00	denisse				
151	02/06/2022	\$ 90.00	\$ 272.00	\$ 182.00	\$ 182.00	\$ 203.00	-\$ 21.00	denisse				
152	03/06/2022	\$ 113.00	\$ 297.50	\$ 184.50	\$ 184.50	\$ 229.00	\$ 1.50	denisse				
153	04/06/2022	\$ 199.50	\$ 314.50	\$ 115.00	\$ 115.00	\$ 102.00	\$ 13.00	dania				
154	05/06/2022	\$ 167.50	\$ 829.50	\$ 662.00	\$ 662.00	\$ 757.00	\$ 5.00	ruben				
155	06/06/2022	\$ 100.50	\$ 127.50	\$ 27.00	\$ 27.00	\$ 281.00	\$ 6.00	denisse	FALTANTE JUNIO			
156	07/06/2022	\$ 82.00	\$ 167.50	\$ 85.50	\$ 85.50	\$ 134.00	-\$ 8.50	denisse	RUBEN	-\$ 38.50		
157	08/06/2022	\$ 95.50	\$ 263.50	\$ 168.00	\$ 168.00	\$ 135.00	\$ 33.00	denisse	ESTRELLA	-\$ 39.00		
158	09/06/2022	\$ 70.00	\$ 162.00	\$ 242.00	\$ 242.00	\$ 262.00	\$ -	denisse	ARESLY	\$ 1.00		
159	09/06/2022	\$ 327.00	\$ 560.00	\$ 233.00	\$ 233.00	\$ 313.50	-\$ 10.50	ruben				
160	10/06/2022	\$ 95.00	\$ 240.00	\$ 145.00	\$ 145.00	\$ 304.00	-\$ 53.00	denisse				
161	11/06/2022											
162	12/06/2022	\$ 131.00	\$ 374.00	\$ 243.00	\$ 243.00	\$ 231.00	\$ 12.00	estrella				
163	12/06/2022	\$ 374.00	\$ 831.00	\$ 457.00	\$ 457.00	\$ 485.00	-\$ 28.00	ruben				
164	13/06/2022	\$ 278.00	\$ 437.00	\$ 159.00	\$ 159.00	\$ 153.00	\$ 6.00	ruben				
165	14/06/2022	\$ 109.00	\$ 148.00	\$ 39.00	\$ 39.00	\$ 61.00	-\$ 22.00	estrella	DENISSE	-\$ 92.50		
166	14/06/2022	\$ 190.00	\$ 500.50	\$ 310.50	\$ 310.50	\$ 307.50	\$ 3.00	ruben				
167	15/06/2022	\$ 179.50	\$ 549.50	\$ 370.00	\$ 370.00	\$ 83.00	\$ 85.00	aresly				

Figura 6.6 Cortes mensuales

Debido a que no existe forma de visualizar un reporte de los artículos que vende diario cada usuario, en Excel se lleva un control al respecto, es importante hacer mención que el nuevo sistema en su apartado de Reportes tampoco realiza este tipo de estadísticas, ni por artículo, ni categoría del artículo, ni rango de horario, etc, aún le falta mucho al sistema de ventas en su apartado de Reporte, para que se muestren más detalladas las estadísticas de la empresa. En la siguiente Figura 6.7 se muestra cada artículo diario vendido por cajero:

	F	G	H	I	J	K	L	M	N	O	P	Q
1	FECHA	DEDESCRIPCION	CANTIDA	EMPLEA				FECHA	DEDESCRIPCION	CANTIDA	EMPLEA	
2	09/03/2022	Carpeta tamaño oficina	1	Angel				12/03/2022	Folleto	1	Ramiro	
3	09/03/2022	Tijeras delgadas chicas	1	Angel				12/03/2022	Foami normal	2	Ramiro	
4	09/03/2022	Curp b/n	2	Angel				12/03/2022	Ojos	1	Ramiro	
5	09/03/2022	Corrector de lapiz	1	Angel				13/03/2022	Copias	4	Angel	
6	09/03/2022	Corrector	1	Angel				13/03/2022	Copias credencial	1	Angel	
7	10/03/2022	Cinta de medir	1	Denisse				13/03/2022	Mameluco de bebe	1	Angel	
8	10/03/2022	Pulpitas	9	Denisse				13/03/2022	Bolsa de regalo	7	Angel	
9	10/03/2022	Zipper	2	Denisse				13/03/2022	Papel china	8	Angel	
10	10/03/2022	Dulce combo	1	Denisse				13/03/2022	Carrulina blanca	1	Angel	
11	10/03/2022	Bocadines	1	Denisse				13/03/2022	Chicle	2	Angel	
12	10/03/2022	Caja de dulces san vale.	1	Denisse				13/03/2022	100 hojas de maquina	1	Angel	
13	10/03/2022	Pintura vegetal	1	Denisse				13/03/2022	Hojas de maquina	12	Angel	
14	10/03/2022	Impresión color	8	Denisse				13/03/2022	Hojas de color	2	Angel	
15	10/03/2022	Estrellitas arcoiris	1	Denisse				13/03/2022	Carpeta de color	2	Angel	
16	10/03/2022	Sabritas	6	Denisse				13/03/2022	Cuaderno cuadricula gde	1	Angel	
17	10/03/2022	Envio Whatsapp	3	Denisse				13/03/2022	Impresión b/n	41	Angel	
18	10/03/2022	Copias	20	Denisse				13/03/2022	Envio Whatsapp	6	Angel	
19	10/03/2022	Copia credencial	4	Denisse				13/03/2022	Bolas de tamarindo	2	Angel	
20	10/03/2022	No afiliacion ISSTE	1	Denisse				13/03/2022	Lapiz baco teacher	1	Angel	
21	10/03/2022	Impresión b/n	3	Denisse				13/03/2022	Pegamento en barra Bully	2	Angel	

Figura 6.7 Artículos diarios vendidos

Actualmente en la empresa se separa del total de la venta un 30% del dinero, se destina a sueldo de empleados, hay semanas en que si se solventa ese gasto y hay otras semanas en que el monto de sueldos supera por mucho el 30% de las ventas, habiendo pérdida de dinero. Se

lleva también un archivo con ese concentrado como lo muestra la siguiente Figura 6.8:

	A	B	C	D	E	F
1	FECHA	EMPLEADO	SUELDO SEMAN	VENTA TOTAL	COMISION	TOTAL
2	30/09/2022	ARESLY				
3	30/09/2022	DANIA				
4	07/10/2022	ARESLY	768.00	2,641.00	79.23	847.23
5	07/10/2022	DANIA	768.00	2,979.00	89.37	857.37
6	14/10/2022	ARESLY		358.00	10.74	600
7	14/10/2022	DANIA	832.00	1,411.00	42.33	874.33
8	22/10/2022	DANIA	640.00	1,521.00	45.63	685.63
9	21/10/2022	ARESLY	768.00	2,039.00	61.17	829.17
10	28/10/2022	ARESLY	640.00		0.00	699.00
11	29/10/2022	DANIA	768.00	2,113.00	63.39	831.39
12	03/11/2022	ARESLY				738.00
13	04/11/2022	DANIA				699.00
14	10/11/2022	ARESLY	960.00	2,390.00	71.70	1,030.00
15	01/12/2022	ARESLY	940.00	2,338.00	70.14	1,010.14
16	03/12/2022	DANIA	900.00	1,095.00	32.85	932.85
17	09/12/2022	ARESLY	920.00	2,055.00	61.65	981.65
18	10/12/2022	DANIA	920.00	2,398.00	71.94	991.94
19						
20						
21						

Figura 6.8 Concentrado sueldos personal

Hace falta actualmente en Excel, realizar un archivo donde haga la comparación del monto mensual vendido para determinar el 30% y comparar con el monto total de sueldos, actualmente solo se hace de manera visual, pero se pretende que también se realice con el Data warehouse, para conocer bien la situación real, si es de pérdida o ganancia y además poder conocer predicciones al respecto para ver la viabilidad de seguir con la empresa o las 2 sucursales. También es necesario hacer una comparación de las ventas de cada sucursal, para conocer cual sucursal se pronostica a la alza y cuál va a la baja o cual se ha mantenido en ventas, esa tampoco existe ni en archivos de Excel.

6.4 DIFICULTADES ENCONTRADAS DURANTE LA REALIZACIÓN DEL PROYECTO

Eliminar primer registro de todas las tablas

Después de la importación de la base de datos, se observó que en el primer registro de todas las tablas aparecían datos como encabezados, que fue necesario eliminar. La siguiente Figura 6.9 muestra el ejemplo que se describe anteriormente.

CODIGO	DESCRIPCION	TVENTA	PCOSTO	PVENTA	DEPT	PROVID	UMEDIDA	MAYOREO	IPRIORIDAD	DINVENTARIO	DINVMINIMO	DINVMAXIMO
DDUB	DUBALIN	U	0	2.5	11	0	1375264	0		-1	0	
SCCC	COPIA POR LOS DOS LADOS	U	0	2	6	0	1375264	0		-1	0	
SCCI	COPIA CARTA IMPRESORA	U	0	1	6	0	1375264	0.5		-1	0	

Figura 6.9 Duplicado de encabezado

Para eliminar el primer registro de cada tabla, se tiene que poner el nombre de la tabla en cuestión y poner dentro de la cláusula where la condición donde el nombre de la primer columna sea igual al texto ubicado en el primer registro, de la forma siguiente:

```
DELETE FROM productos
WHERE CODIGO='CODIGO'
```

Tabla usuarios tenía datos erróneos

Al momento de querer realizar consultas donde se hiciera uso de la tabla usuarios, se encontró que solo hacía referencia a un solo usuario debido a que la información de la tabla no se había importado completa, de 26 usuarios que tenía la base de datos, sólo aparecía un usuario, como aparece en la Figura 6.10, fue necesario eliminarla y volver a importarla como lo muestra la Figura 6.11.

ID	NOMBRE_COMPLETO	DIRECCION	TELEFONO	USUARIO	CLAVE	ACTIVO	PERMISOS	CREATED_ON	CORREO	ESTA_EN_CA.
7	karla ramos	39617		admin	cmrk37	1	618475290660	16/06/2008 00:00:00	soporte@abarotespuntodeventa.com	1

Figura 6.10 Tabla usuarios incompleta

La eliminación de la tabla se realiza mediante la sentencia siguiente:

```
DROP TABLE usuarios;
```

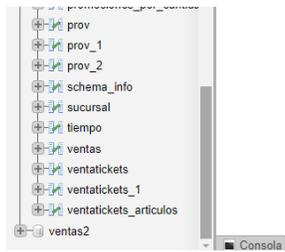


Figura 6.11 Eliminación tabla usuarios

Se puede apreciar en la Figura 6.11 que la tabla usuarios ya no aparece dentro de la base de datos.

Para importar nuevamente la tabla en formato .CSV se debe dirigir a la pestaña importar de phpmyadmin, dentro del botón seleccionar archivo dirigirse a la ruta en donde se encuentra el archivo usuarios.csv, como lo muestra la figura 6.12.

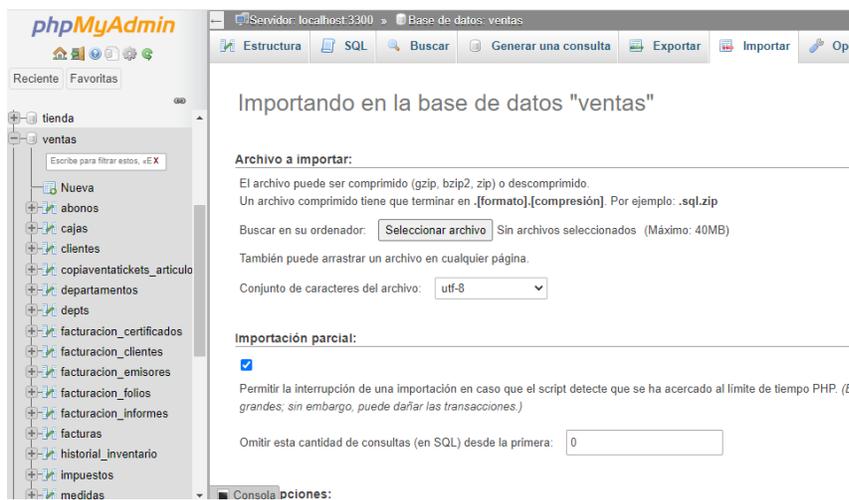


Figura 6.12 Importar tabla usuarios

A continuación debe aparecer la ventana de éxito de la importación, como se aprecia en la figura 6.13:

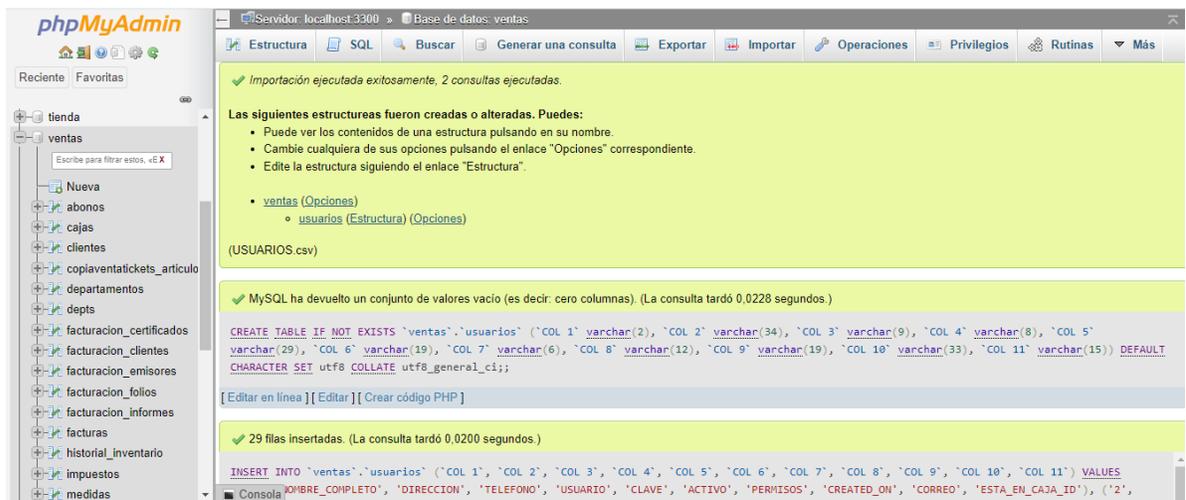


Figura 6.13 Reimportación tabla usuarios

Después de importar y dirigirse a la tabla usuarios, se encuentra que los nombres de los campos no son los que corresponden, como lo muestra la figura 6.14, por lo cual se procede a nombrar cada campo con el nombre que están en la primer fila.



Figura 6.14 Encabezados incorrectos en tabla usuarios

Además se aprecia que la primer fila tiene también los encabezados como datos, esta fila hay que eliminarla con el procedimiento del punto 6.3. Después de realizar lo anterior, debe quedar como en la siguiente Figura 6.15:

ID	NOMBRE_COMPLETO	DIRECCION	TELEFONO	USUARIO	CLAVE	ACTIVO	PERMISOS	CREATED_ON	COF
2	fernando			fernando	9540	1	618475290624	06.10.2020 08:40:36	
7	Administrador de la Tienda	39617		admin	cmrki	1	618475290625	16.06.2008 00:00:00	sop
3	Denise Ogaz			denise (Eliminado 23/01/2021)	313396	0	618475290626	06.10.2020 18:55:27	
4	Rodolfo Laredo			rodolfo	968524xy	1	618475290627	07.10.2020 08:44:52	
5	ULISES LOPEZ VILALOBOS			ulises (Eliminado 23/01/2021)	36524	0	618475290628	07.10.2020 13:43:27	
28	Aresly Niño			aresly	464046	1	618475290629	13.06.2022 16:32:33	
6	ramiro			ramiro	krxyzw536	1	618475290630	26.10.2020 16:23:05	

Figura 6.15 Cambio de encabezado tabla usuarios

6.4 AVANCE

De acuerdo a todo lo realizado en este proyecto se llega a un avance de los siguientes rubros:

- Comprensión del problema 100%
- Identificar las fuentes de datos 100%
- Implementación herramientas ETL 100%
- Creación del modelo del Data warehouse 100%
- Diseño y Programación 80%
- Toma de decisiones 80%
- Predicciones 30%
- Documentación 90%

En base a lo anterior se considera como verdadera la hipótesis dada a inicio de este documento: “El Data warehouse permite el análisis de datos históricos y facilita la predicción de comportamiento en las ventas”.

6.5 CAPACITACIÓN

El proceso de capacitación se puede aplicar a cualquier empresa, independientemente si tiene actualmente ya instalado un sistema de ventas, el propósito es capacitar desde cero. El propósito es concientizar al dueño o encargado del lugar que sólo las nuevas tecnologías, programas computacionales como un sistema de venta y herramientas informáticas como las ETL y Data

warehouse pueden apoyar en gran modo un buena de decisiones en la empresa que conlleva a aumentar ventas, aprovechar bien los recursos, buen manejo y selección de personal.

7 BIBLIOGRAFÍA.

- Alicia CONCYTEC. (2016). Sistema de Soporte de Decisiones con tecnología data warehouse para la gestión de la información de la empresa Mallku Import SAC - Juliaca 2016
https://alicia.concytec.gob.pe/vufind/Record/RNAP_e3ceebe6d62818b5a105fe3a440fd741/Description#tabnav
- Apache Friends (2023). Página de descarga de XAMPP
<https://www.apachefriends.org/es/index.html>
- Aprendeia (2023). Librería-scikit-learn-de-python
<https://aprendeia.com/libreria-scikit-learn-de-python/>
- AWS Amazon Web Services (2023). Qué es Python.
<https://aws.amazon.com/es/what-is/python/>
- BBVA (2023). Definición de una PYME.
<https://www.bbva.mx/educacion-financiera/creditos/que-es-una-pyme.html>
- Cleverdata (2021). Definición de Machine Learning.
<https://cleverdata.io/que-es-machine-learning-big-data/>
- D'Arconte, C. (2018). Business intelligence applied in small size for profit companies. *Procedia computer science*, 131, 45-Desarrollo Web (2023).
<https://desarrolloweb.com/home/mysql>
- Gliffy (2023). Página de compra o descarga del programa de diseño Gliffy.
<https://www.gliffy.com/>
- Georgia. Garani, A. Chernov, I. Savvas and M. Butakova, "A Data Warehouse Approach for Business Intelligence," *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Napoli, Italy, 2019, pp. 70-75, doi: 10.1109/WETICE.2019.00022.
- Gravitar (2022). Servicio de consultoría en inteligencia de negocios.
<https://gravitar.biz/datawarehouse/metodologias-data-warehouse/#:~:text=Un%20Data%20Warehouse%20es%20un,consultas%20y%20toma%20de%20decisiones.>
- Guillén Quisca, R. N. (2017). Sistema de soporte de decisiones con tecnología Data warehouse para la gestión de la información de la empresa mallku import sac-juliaca 2016.
- Hamoud, A. K., Abd Ulkareem, M., Hussain, H. N., Mohammed, Z. A., & Salih, G. M. (2020,

May). Improve HR decision-making based on data mart and OLAP. In Journal of Physics: Conference Series (Vol. 1530, No. 1, p. 012058). IOP Publishing.

INEGI (2021). El inegi presenta los resultados del estudio sobre la demografía de los negocios 2021. https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/edn/edn_2021pdf

Mohammed Dahr Jasim, Khalaf Hamoud Alaa. (2022). Implementing Sales Decision Support System Using Data Mart Based On Olap, Kpi, And Data Mining Approaches

Lozada Peñafiel, Ximena Nathalie (2014). Análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los kpi, para la empresa kronosconsulting cia ltda.

Méndez Carmona Mario Alberto (2018). Modelado e implementación de un Data warehouse para un mdms utilizando algoritmos etl y vee. <https://rinacional.tecnm.mx/jspui/handle/TecNM/728>

Peñafiel, X. L., Tamayo, h. C., Argudo, w. P., & de la torre dÍaz, a. (2013). Análisis, diseño, construcción e implementación de un Data warehouse para toma de decisiones y construcción de los kpi, para la empresa kronosconsulting cia ltda. Internet] <http://repositorio.Espe.Edu.Ec/bitstream/21000/7942/1/ac-si-espe-047750.Pdf>

[Pressman, Roger S. \(2010\).](#)

Ingeniería del software un enfoque práctico. Editorial Mc. Graw Hill Séptima edición.

Reddy Nadikattu, Rahul (2019). Arquitectura de almacenamiento de datos: liderando la próxima generación de ciencia de datos

Repositorio Nacional Tecnológico Nacional de México (2018). <https://rinacional.tecnm.mx/handle/TecNM/728>

Universidad Israel (2019). <http://repositorio.uisrael.edu.ec/handle/47000/2180>

Vaisman Alejandro, Zimányi Esteban. (2014). Data warehouse Systems Desing and implementation, Springer.

Yaguachi, D. E. C., & jumbo, j. L. C. (2017). Diseño de un prototipo de Data warehouse para una empresa de medicina prepagada. Apropiación, generación y uso edificador del conocimiento de estudiantes sentipensantes.

8 APENDICE

Componentes de la instalación de MySql

Nuestros dos componentes principales de instalación de MySQL son:

- El servidor de la base de datos, que es el sistema de gestión de bases de datos MySQL.
- A través del terminal, el cliente se conecta al sistema gestor de bases de datos.

Aunque las dos piezas de software normalmente se utilizan en conjunto, es importante distinguirlas. A modo de ilustración, cuándo utilizar la línea de comando:

(Desarrolloweb, 2023).

```
mysql -h localhost -u root -p
```

El comando "mysql" se utiliza para conectarse al servidor de base de datos instalado localmente mediante el cliente MySQL. (Desarrolloweb, 2023).

Aunque tenemos acceso a otros clientes de bases de datos como MySQL Workbench, HeidiSQL, etc., no siempre utilizamos el cliente, que nos facilitan el acceso a una interfaz gráfica de usuario muy útil y eficaz. (Desarrolloweb, 2023).

Instalación de MySQL sobre Windows

Para la realización de este proyecto, se instaló MySQL sobre la plataforma Windows, a continuación se describen los pasos para su instalación:

Sólo requiere que un instalador, que incluye el asistente estándar, esté instalado en Windows. Instale la versión comunitaria simplemente descargándola. (Desarrolloweb, 2023).

Se pueden considerar varios factores, entre ellos:

1. Si se instala MySQL para trabajar con PHP puede que algunos sistemas como Xampp ya te lo instalen todo de una vez.
2. Usando el administrador de paquetes Chocolatey, se puede instalar MySQL en Windows. Para usar Chocolatey, simplemente emita el comando: después de la instalación. (Desarrolloweb, 2023)

```
choco install mysql
```

Instalación de MySQL en Linux

En caso que la instalación se quiera desarrollar sobre el sistema operativo Linux, se debe seguir el siguiente procedimiento:

Como los administradores de paquetes varían de una distribución de Linux a otra, esto depende de la distribución que esté utilizando. (Desarrolloweb, 2023).

Ejecute el siguiente comando para instalar MySQL en Ubuntu u otra distribución basada en Debian: (Desarrolloweb, 2023).

```
apt install mysql-server
```

Este sería el caso si necesitara ejecutar el comando como superusuario. (Desarrolloweb, 2023).

```
sudo apt install mysql-server
```

Rocky Linux necesita MySQL instalado. (Desarrolloweb, 2023).

```
sudo dnf install mysql-server
```

De acuerdo con las instrucciones de la página de instalación de MySQL para Linux en la documentación oficial, es posible que deba activar los repositorios de versiones de la comunidad MySQL para otras distribuciones. (Desarrolloweb, 2023).