

# DATAWAREHOUSE

DATA BASE:

ClassicModel



Proyecto Bases de Datos II sección 1100

Integrantes:

- Kency Pamela Osegura Valdez 20201004556
- Kenia Paola Romero Nuñez 20171003359
- Cristian Rodil Zuniga 20181002386



# Contenido

PROYECTO DE DATA WAREHOUSE .....	1
OBJETIVOS .....	2
GENERAL .....	2
ESPECIFICOS.....	2
MARCO TEÓRICO .....	3
INTELIGENCIA DE NEGOCIOS .....	3
Beneficios de BI .....	4
DATA WAREHOUSE.....	6
Funcionamiento.....	7
Data Mart.....	7
Diferencia entre almacén de datos y data mart .....	7
Pasos para construir un Data Warehouse.....	7
OLTP .....	8
OLAP .....	8
ETL.....	8
Análisis de datos y toma de decisiones .....	9
Ventajas e inconvenientes de los data warehouse .....	9
TABLEAU .....	10
BASE DE DATOS OLTP: ClassicModels .....	12
PREGUNTAS DEL NEGOCIO .....	13
MÉTRICA UTILIZADA .....	14
CONCLUSIONES.....	15
RECOMENDACIONES .....	16
Bibliografía .....	17

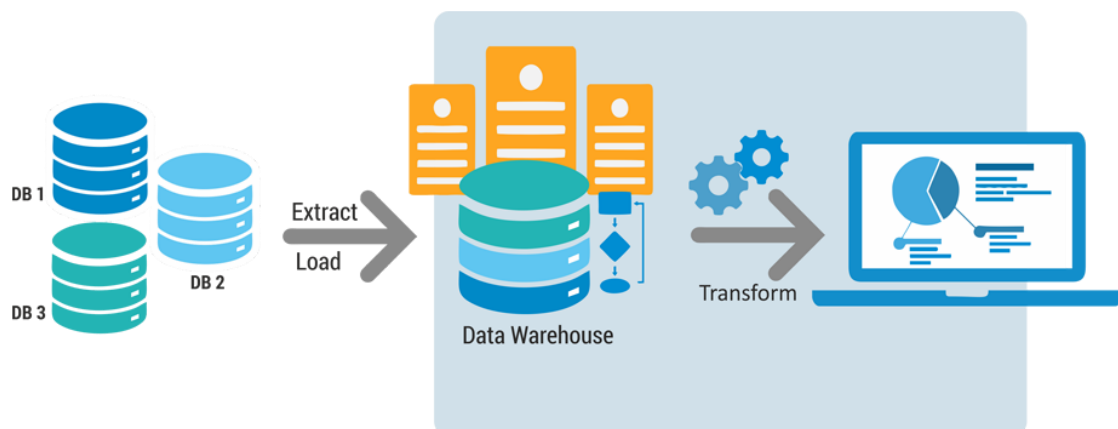
# Proyecto Data Warehouse

Se ha desarrollado un proyecto de Data Warehouse utilizando la base de datos OLTP conocida como ClassicModels, la cual se centra en ser un minorista de modelos a escala de autos clásicos.

El objetivo principal de este almacén de datos es proporcionar una solución centralizada que permita a una alta gama de usuarios acceder a la información necesaria a través de informes personalizados que aborden sus necesidades específicas.

Este enfoque busca no solo mejorar la accesibilidad de los datos, sino también brindar un marco integral para el análisis y la toma de decisiones. Además, se busca mejorar la eficiencia operativa y la capacidad de respuesta a través de la consolidación y transformación de datos en un formato más útil y comprensible.

Al proporcionar a los usuarios la capacidad de realizar análisis detallados y generar informes, se fomenta la toma de decisiones fundamentadas en evidencia y se impulsa la innovación en todos los niveles.



# OBJETIVOS

## GENERAL

Implementar un Data Warehouse basado en la base de datos OLTP ClassicModels para proporcionar una solución integral que permita el acceso eficiente a la información y facilite la toma de decisiones de la organización.

## ESPECIFICOS

- Establecer una infraestructura que integre y transforme los datos de la base de datos OLTP en un formato adecuado para el análisis y la generación de informes.
- Desarrollar un conjunto de informes que cumplan con las preguntas solicitadas para el desarrollo de la data mart.

El equipo conformado por Oseguera, Romero y Zuniga, (2024) llevo a cabo el proyecto "Data Warehouse en la base de datos ClassicModels", haciendo uso de la inteligencia de negocios; dicho proyecto se centró en la implementación de un data mart destinado a explorar las relaciones entre empleados, clientes y productos. Este enfoque se orientó hacia la satisfacción de la necesidad de obtener información detallada sobre estas dimensiones, permitiendo así un análisis exhaustivo de los datos. El propósito fundamental fue mejorar la toma de decisiones empresariales, proporcionando una orientación precisa para el negocio hacia las mejores oportunidades y estrategias.

El éxito en la implementación de la Data Warehouse está estrechamente ligado a la presencia de facilitadores dentro de la organización, encargados de suministrar información y requisitos pertinentes. La utilización de software libre se presenta como una herramienta fundamental en este proceso, ya que ofrece la flexibilidad necesaria para adaptarse de manera más personalizada a las necesidades del negocio. Este enfoque, por ende, conlleva a la generación de informes más precisos y ajustados a las demandas específicas de la empresa, fortaleciendo así su capacidad de análisis y toma de decisiones.

## INTELIGENCIA DE NEGOCIOS

El Business Intelligence (BI) o inteligencia de negocios, minería de datos, visualización de datos, herramientas e infraestructuras de datos, y las practicas recomendadas para ayudar a las organizaciones a tomar decisiones más basadas en los datos. En la práctica sabes que tienes una inteligencia de negocios moderna cuando tienes una visión integral de los datos de tu organización y los utilizas para impulsar el cambio, eliminar las ineficiencias y adaptarte rápidamente a los cambios del mercado o suministro.

La inteligencia de negocios muestra datos actuales e históricos dentro de su contexto empresarial para que las empresas tomen mejores decisiones. Los analistas pueden aprovechar BI para proporcionar puntos de referencia de rendimiento y de la competencia para que la organización funcione de manera más fluida y eficiente. Los analistas también pueden detectar más fácilmente las tendencias del mercado para

aumentar las ventas o los ingresos. Cuando los datos correctos se utilizan de forma eficaz, pueden ayudar con cualquier cosa, desde el cumplimiento hasta los esfuerzos de contratación.

El BI funciona dando respuesta a las preguntas y objetivos que presentan las empresas y organizaciones. Para responder a estas preguntas y hacer un seguimiento en relación con los objetivos, recopilan los datos necesarios, los analizan y determinan las acciones que deben tomarse para que alcancen los objetivos.

En el aspecto técnico, los datos sin procesar se recopilan de la actividad del negocio. Los datos se procesan y luego se guardan en los almacenes de datos. Una vez que los datos están almacenados, los usuarios pueden acceder a ellos, a fin de iniciar el proceso de análisis para responder preguntas de negocios.

Una interesante definición para inteligencia de negocios, según el Data Warehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial.

### Beneficios de BI

Dentro del marco de beneficios que representa una solución de inteligencia de negocios podemos mencionar que esta nos permite:

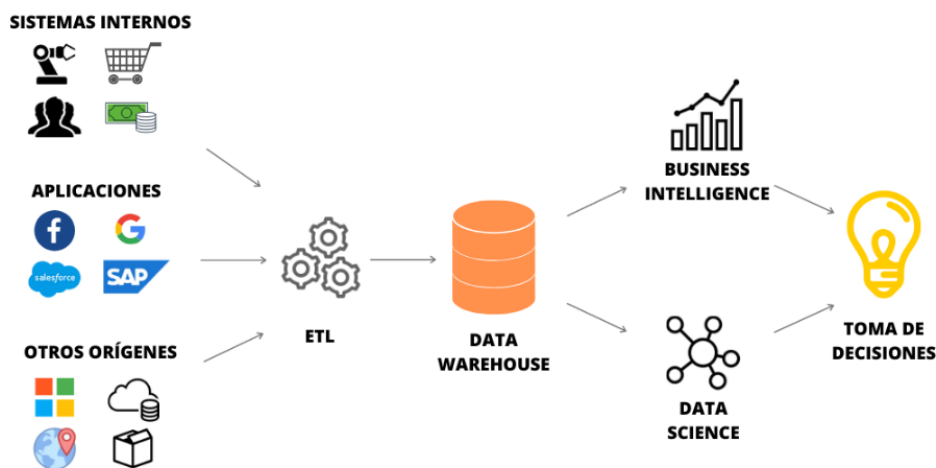
- **Manejar el crecimiento:** el reto para las empresas es evolucionar, es crecer y esto significa “cambio”. Que tan ágiles son mis procesos para enfrentar los cambios y las necesidades puntuales de la empresa.
- **Control de costos:** el manejo de costos es el detonador que fuerza muchas empresas a considerar una solución de inteligencia de negocios, para tener la capacidad de medir gastos y ver esto a un nivel de detalle que identifique la línea de negocios, producto, centro de costo, entre otras.
- **Entender mejor a los clientes:** las empresas almacenan toneladas de información valiosa relacionada a sus clientes. El reto es transformar esta información en conocimiento y este conocimiento dirigido a una gestión comercial que

representa algún tipo de ganancia para la empresa. Ventas crudas, ventas incrementales, retención de clientes, anticipar nuevas oportunidades, medición de efectividad de campañas e identificar los patrones de compras y/o comportamiento de los clientes hacen de la inteligencia de negocios un elemento vital para lograr los objetivos de la empresa.

- **Indicadores de gestión:** los indicadores de desempeño permiten representar medidas enfocadas al desempeño organizacional con la capacidad de representar la estrategia organizacional en objetivos, métricas, iniciativas y tareas dirigidas a un grupo y/o individuales en la organización. Dentro de las capacidades funcionales de los indicadores de gestión podemos mencionar: el monitoreo, análisis y la administración. “monitorea” los procesos críticos de negocios y las actividades utilizando métricas que me den una alerta sobre un problema potencial o alguna gestión que debo realizar. “Analiza” la raíz de los problemas explorando la información de múltiples perspectivas en varios niveles de detalle. “Administra” los recursos y procesos para dirigir la toma de decisiones, optimizar el desempeño. Esto nos permite tener una visión global de la empresa con la capacidad de dirigir la organización en la dirección correcta.

### Solución de arquitectura de BI

Es importante visualizar de alguna forma que comprende una arquitectura de inteligencia de negocios.



En la imagen se representa gráficamente esta arquitectura.

Analicemos este diagrama de izquierda a derecha. Los primeros dibujos representan las distintas fuentes de datos que pudieran utilizarse para extraer los datos de múltiples fuentes simultáneamente. El segundo dibujo representa el proceso de extracción, transformación y carga (ETL). Este proceso es en el que se definen de las fuentes heterogéneas que campos se van a utilizar, se necesitan algún tipo de modificación y/o transformación y donde quiere ubicar estos datos, este proceso se le conoce como “mapping”. El tercer dibujo representa el repositorio de datos. En este repositorio se encuentran los datos transformados representados visualmente en modelos multidimensionales, dimensiones y tablas de datos. Existe un proceso entre el repositorio de datos y la interfase de acceso al usuario, este es el motor de BI que me permite habilitar componentes, administrar consultas, monitorear procesos, cálculos, métricas. La interfase de acceso a usuarios permite interaccionar con los datos, representar de forma gráfica con aquellos resultados de las consultas y los indicadores de gestión que fueron contruidos.

## DATA WAREHOUSE

### Repositorio Central de Datos

Un Data Warehouse es una base de datos integrada y orientada a temas que almacena datos de múltiples fuentes para su análisis y reporte. Recopila datos históricos y actuales de la organización, lo que permite una visión integral del negocio a lo largo del tiempo.

El término “Data Warehousing” se refiere al proceso que consiste en recolectar y manipular datos provenientes de diversas fuentes, con el fin de recuperar informaciones valiosas para una empresa.

Esta plataforma reúne diversas tecnologías y componentes que permiten explotar los datos. Permite almacenar una gran cantidad de datos, así como también su tratamiento y análisis. El objetivo es transformar los datos brutos en información útil, y volverlos disponibles y accesibles para los usuarios.



Un Data Warehouse se encuentra generalmente separado de la base de datos operacional de la empresa. Permite a sus usuarios apoyarse en datos históricos y actuales para tomar mejores decisiones.

#### Funcionamiento

Un Data Warehouse funciona como un repositorio central. La información proviene de una o varias fuentes de datos, como sistemas transaccionales u otras bases de datos relacionales.

Los datos pueden ser estructurados, semiestructurados o no estructurados. Una vez que están integrados en el warehouse, son tratados y transformados. Los usuarios pueden luego acceder a ellos con la ayuda de herramientas de Business Intelligence, de clientes SQL o de tablas.

Además, los Data Warehouse hacen posible el Data Mining (exploración de datos). Este procedimiento consiste en buscar tendencias y patrones en los datos y apoyarse sobre estos para argumentar las ventas y las ganancias de la empresa.

#### **Data Mart**

Es un sistema de almacenamiento de datos que contiene información específica de la unidad de negocio de una organización. Contiene una parte pequeña y específica de los datos de la empresa que la almacena en un sistema de almacenamiento más grande. Las empresas utilizan una data mart para analizar la información específica de cada departamento de manera más eficiente. Brinda datos resumidos que las partes interesadas clave pueden utilizar para tomar decisiones informadas rápidamente.

#### **Diferencia entre almacén de datos y data mart**

Un data mart comparte muchas de las cualidades de un almacén de datos. La diferencia que hay entre ellos es que un almacenamiento de datos contiene datos de toda la empresa sobre diversos temas. Mientras tanto, un data mart almacena información estrechamente relacionada con un tema específico.

#### **Pasos para construir un Data Warehouse**

1. Definir los objetivos: establecer los objetivos de negocio que se desean alcanzar con el Data Warehouse. Esto guiara el diseño y la construcción de la solución.

2. Analizar las fuentes de datos: identificar y evaluar las diferentes fuentes de datos, tanto internas como externas, que alimentarán el Data Warehouse.
3. Diseñar el modelo dimensional: crear un modelo de datos optimizado para análisis, con dimensiones y hechos, que permita responder a las necesidades de información.
4. Implementar el ETL: desarrollar los procesos de extracción, transformación y carga de datos desde las fuentes al Data Warehouse.
5. Desplegar la solución: poner en funcionamiento el Data Warehouse e integrar las herramientas de análisis y reporteria para los usuarios finales.
6. Mantener y escalar: asegurar la actualización continua del Data Warehouse y la adaptación a las nuevas necesidades de la organización.

### **OLTP**

Una base de datos OLTP (Procesamiento de Transacciones en Línea) es un sistema diseñado para manejar transacciones en tiempo real. Se utiliza principalmente para aplicaciones operativas en las que la velocidad y la precisión son críticas. Estas bases de datos están optimizadas para insertar, actualizar y eliminar datos de manera eficiente.

### **OLAP**

Por otro lado, una base de datos OLAP (Procesamiento Analítico en Línea) se utiliza para el análisis y la generación de informes. Estas bases de datos están diseñadas para consultas complejas y análisis de grandes volúmenes de datos. Son ideales para extraer información agregada y realizar análisis en profundidad.

### **ETL**

La extracción, transformación y carga (ETL) es un proceso que integra y transfiere información de varios orígenes de datos a una sola base de datos física. Los data marts utilizan la ETL para recuperar información de fuentes externas cuando esta no proviene de un almacenamiento de datos. El proceso implica los siguientes pasos:

- Extracción: se recopila información sin procesar de varias fuentes.
- Transformación: se estructura la información en un formato común.
- Carga: se transfieren los datos procesados a la base de datos.

Las herramientas ETL copian información de fuentes externas, como hojas de cálculo, aplicaciones y documentos de texto. Luego, el data mart procesa, organiza y almacena la información de forma estructurada.

### **Análisis de datos y toma de decisiones**

Una vez que los datos han sido recopilados y almacenados en el Data Warehouse, las empresas pueden realizar análisis exhaustivos para extraer información valiosa.

Mediante herramientas de visualización y minería de datos, los gerentes y analistas pueden identificar tendencias, patrones y oportunidades clave que les permitan tomar decisiones informadas y estratégicas.

El data warehouse se convierte en la fuente única de verdad, permitiendo a la organización alinear sus objetivos y acciones con base en datos confiables.

### **Ventajas e inconvenientes de los data warehouse**

Resultan útiles para permitir a las empresas acceder de forma rápida y fácil a los datos provenientes de diversas fuentes de forma centralizada.

Gracias a estas herramientas, se puede acceder a informaciones coherentes y actualizadas sobre todas las actividades de la empresa. Permite también generar informes y efectuar búsquedas con los datos.

Gracias a los grandes volúmenes de datos históricos, los usuarios pueden analizar las tendencias sobre diferentes periodos temporales para realizar predicciones futuras.

Sin embargo, los data warehouse presentan también inconvenientes, no es la mejor solución para los datos no estructurados.

Además, la creación e implementación de depósitos de datos llevan tiempo y requieren en general mucho trabajo. Paradójicamente, un data warehouse puede rápidamente volverse obsoleto.

Por otro lado, es difícil realizar cambios en el tipo de datos, los esquemas fuentes de datos, los index y las búsquedas. La utilización de una plataforma como esta puede resultar demasiado compleja para un usuario promedio.

Así, las empresas invierten muchos recursos para formar a sus empleados y para implementar el warehousing.

Es importante sopesar las ventajas y desventajas antes de decidir utilizar este tipo de solución.

## TABLEAU

Tableau es una plataforma de visualización de datos. Uno de los softwares más empleados en Business Intelligence, ya que permite realizar una analítica completa de datos.

Se trata de una herramienta muy útil, ya que permite administrar y realizar un análisis super completo de una segura, eficaz y adaptable. Por tanto, es uno de los softwares más utilizados por las empresas.

Una interfaz muy intuitiva que mejora aspectos tan importantes en Big Data como la estadística o la propia preparación de datos inteligente.

Es una herramienta clave para Big Data. Bases de datos, hojas de cálculo o servicios cloud son algunas de sus claves.

### **Tableau es de uso sencillo**

Destaca por su uso fácil. Analizar los diferentes datos, arrastrando y soltando (no es necesario contar con un lenguaje de programación).

### **Es de análisis rápido**

Otra de las características y ventajas de tableau es la conexión de datos. En cuestión de minutos vas a poder visualizarlos con esta herramienta.

### **Actualización automática**

Asimismo, las actualizaciones se llevan a cabo en tiempo real. También es posible programarlas con la frecuencia que el usuario quiera-

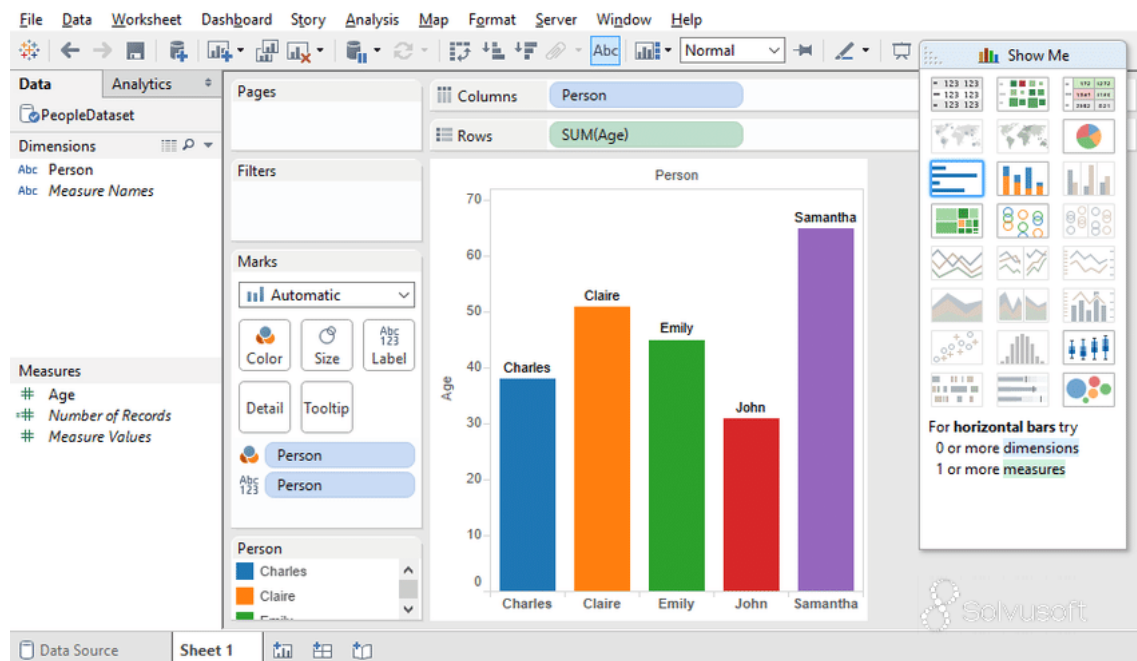
### **Información inmediata**

La información va a ser compartida y procesada en tiempo real, también.

## TABLEAU DESKTOP

Es una de las versiones más demandadas. Ayuda a crear reportes, dashboards e historias usando diferentes tablas y gráficos. Los libros de trabajo y los dashboards creados con tableau desktop pueden ser compartidos local o públicamente.

Tableau desktop es un software de visualización de datos que permite ver y comprender datos en cuestión de minutos.



## BASE DE DATOS OLTP: ClassicModels

La base de datos OLTP llamada ClassicModels que incluye varias tablas esenciales que representan entidades como oficinas, empleados, clientes, pedidos, líneas de productos, productos, detalles de pedidos y pagos es estructura típica de un entorno transaccional donde se gestionan operaciones diarias como ventas, compras, relaciones con clientes y gestión de empleados.

ClassicModels es una base de datos minorista de modelos a escala de autos clásicos.

La base de datos OLAP se construye tomando en cuenta la base de datos OLTP. Las tablas en la base de datos OLAP se organizan de manera que permiten consultas complejas y se enfoca en proporcionar un entorno optimizado para el análisis de datos empresariales.

Las características clave de la base de datos OLAP incluye:

**Dimensiones:** Las dimensiones representan las características por las cuales se desea analizar los datos. En la base de datos OLAP, las dimensiones realizadas son la dimensión de tiempo (TBL\_TIEMPO), clientes (CUSTOMERS), empleados (EMPLOYEES) y líneas de productos (PRODUCTLINES).

La tabla de **dimensión de tiempo** almacena información relacionada con el tiempo, incluyendo detalles como nombre del mes, semana, trimestre, cuatrimestre, semestre y día de la semana. Estos atributos temporales permitirán analizar los datos en función del tiempo y realizar comparaciones en base al tiempo.

La tabla de **dimensión Clientes** representa información sobre los clientes, incluyendo el número de cliente, nombre del cliente, nombre de contacto, ciudad y país. Esta dimensión se utiliza para analizar las ventas y relaciones con los clientes.

La tabla **dimensión Empleados** contiene datos sobre los empleados, como el número de empleado, nombre del empleado y código de oficina al que están asignados.

La dimensión **Líneas de Producto** (productlines) almacenara información sobre las líneas de productos, incluyendo el nombre de la línea y una descripción. Esta dimensión se utiliza para categorizar y analizar las ventas por línea de productos.

La tabla de **dimensión Productos** almacena detalles específicos sobre cada producto en tu inventario. Cada producto tiene un código único (productcode) y está asociado a una línea de producto a través de la columna productline, que es una clave foránea que hace referencia a la tabla productlines.

En el caso de la **tabla de hechos** almacena datos clave relacionados con las transacciones de ventas. Incluye información como el número de empleado, número de cliente, código de producto, tiempo de la transacción y el total de venta del producto. Estos datos numéricos son esenciales para el análisis de ventas y para analizar el rendimiento de la empresa. Esta tabla está relacionada con las dimensiones antes mencionadas.

La base de datos OLAP utiliza un modelo copo de nieve que refleja una estructura de datos normalizada donde las líneas de productos se separan en una tabla independiente para reducir la redundancia y facilitar el análisis jerárquico de productos.

Enlace a la base de datos OLTP [aquí](#).

## Preguntas del negocio

1. Análisis del total de ventas por empleado que su nombre comience con la letra L, de los días lunes y miércoles en el semestre (1) del año 2003.
2. Se reportan las ventas de motocicletas según la marca dentro de los tres cuatrimestres; meses enero, febrero y marzo del año 2003.
3. Es importante obtener el Nombre, ciudad y país de los clientes que compraron motocicletas 2002 Yamaha YZR M1.
4. Conocer la cantidad en Stock y ventas de Classic Cars (1948 Porsche y Jaguar XK 120) en el cuatrimestre (1-2-3) del año 2003.

5. Se necesita saber el código de producto, nombre, cantidad en stock, código de la línea de producto y una descripción de los barcos vendidos en 2003.
6. Conocer la cantidad en Stock de Classic Cars (Ford y Dodge) en el año 2003.
7. Conocer el nombre y código de oficina de los empleados que vendieron trenes en el segundo cuatrimestre del año 2003.
8. Ventas de los empleados de apellido Patterson de martes a viernes en el semestre (0-1) del año 2003.
9. Ventas de los empleados con la letra G que vendieron Classic Cars y Vintage Cars y la cantidad en stock de estos durante el año 2003.
10. Se desea saber los empleados que atendieron a aquellos clientes que compraron aviones y trenes, y la cantidad que queda en stock de estos.

## Métrica Utilizada

Basándonos en las preguntas del negocio podemos identificar que lo que queremos conocer es el total de ventas por ende esta seria nuestra métrica que tendremos en la tabla de hechos.

La consulta va a representar el monto total generado por las ventas de cada producto en un período de tiempo específico. Esta métrica se calcula multiplicando la cantidad de unidades vendidas (`quantityordered`) por el precio unitario de cada producto (`priceeach`), y luego sumando estos valores para obtener el total de ventas por producto.

La interpretación de esta métrica implica entender cuánto dinero se ha generado por la venta de cada producto dentro de un período determinado. Esto es crucial para analizar el desempeño de ventas de productos individuales, identificar los productos más populares o rentables, y tomar decisiones estratégicas relacionadas con el inventario, precios, entre otros aspectos comerciales.



A grandes rasgos podemos destacar un sinnúmero de conocimientos provechosos entre los que podemos destacar los siguientes:

- ✓ La implementación de un Data Warehouse ofrece numerosas ventajas que impactan directamente en la eficiencia y la competitividad de una organización. Al consolidar datos dispersos en una ubicación centralizada y optimizada para el análisis, se facilita el acceso rápido y confiable a la información clave.
- ✓ Se obtiene una visión integral y coherente de los datos empresariales, y promueve la alineación de los objetivos y procesos internos, optimizando la eficiencia operativa.
- ✓ La implementación exitosa de un data warehouse fortalece la capacidad de la empresa para adaptarse a los cambios del mercado, anticipar las necesidades del cliente y mantener una ventaja competitiva sostenible en un entorno empresarial dinámico y desafiante.
- ✓ La implementación efectiva de un proceso ETL conlleva una serie de beneficios para una organización. En primer lugar, el extraer datos de múltiples fuentes y transformarlos en un formato coherente y utilizable, también proporciona una base sólida para la toma de decisiones informadas y estratégicas. Optimiza la integridad y la calidad de los datos, lo que a su vez mejora la precisión de los análisis y generación de reportes.
- ✓ La implementación de reportes proporciona una implementación clara y concisa de los datos almacenados, permitiendo a los usuarios comprender fácilmente las tendencias, patrones y resultados clave dentro de la organización.

Para implementar un Data Warehouse, haciendo uso de Data Mart, ETL y reportes podemos tomar en cuenta las siguientes recomendaciones:

- ✓ Debemos realizar un análisis exhaustivo de los requerimientos comerciales y de los usuarios finales. Debemos detectar las fuentes de los datos relevantes, entender las necesidades de información de los diferentes departamentos.
- ✓ Se debe planificar la arquitectura del Data Warehouse, debemos asegurarnos de que pueda adaptarse y crecer con las necesidades cambiantes de la organización. Esto implica el diseño de diversos data mart específicos para las diferentes áreas de la organización.
- ✓ Garantizar la calidad de los datos, implementado procesos de limpieza, transformación y validación de datos durante el proceso ELT para garantizar la integridad y consistencia de la información almacenada.
- ✓ Garantizar la accesibilidad y la comprensión intuitiva de las herramientas para usuarios sin experiencia en programación es esencial. Esto facilitara su adaptación y comprensión de las operaciones necesarias para satisfacer sus requerimientos de información.

## Bibliografía

Tableau. (2024) ¿Qué es el Business Intelligence? Tu guía para la inteligencia de negocios y por qué es importante. Recuperado de <https://www.tableau.com/es-mx/learn/articles/business-intelligence>

Oracle - Whitepaper. (2017) ¿Qué es inteligencia de negocios? Recuperado de <https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529esa.pdf>

Lopez, M. (2022) Tableau: ¿Qué es y cómo trabajar con él? IMMUNE TECHNOLOGY INSTITUTE. <https://immune.institute/blog/tableau-que-es/>

DataScientest. (2024) Data Warehouse: ¿qué es y cómo utilizarlo? Recuperado de [https://datascientest.com/es/data-warehouse-que-es-y-como-utilizarlo#:~:text=Un%20Data%20Warehouse%20\(depósito%20de,que%20permiten%20explotar%20los%20datos.](https://datascientest.com/es/data-warehouse-que-es-y-como-utilizarlo#:~:text=Un%20Data%20Warehouse%20(depósito%20de,que%20permiten%20explotar%20los%20datos.)

Aws re: Invent. (2023) ¿Qué es un data mart? Recuperado de <https://aws.amazon.com/es/what-is/data-mart/>