

Carrera: Data Analytics

Módulo 2

Análisis de Base de Datos de Fast Food

Nombre del autor: LAURA OCAMPO CAMPO

Email: LAURA80173@GMAIL.COM

Cohorte: DA-FT06

Fecha de entrega: 14 de Agosto del 2024

Institución: Fast Food Inc



Introducción

El proyecto tiene como propósito central consolidar la información dispersa de la empresa "Fast Food" en una base de datos unificada. La empresa enfrenta desafíos significativos relacionados con la gestión ineficiente de datos, lo que afecta su capacidad de tomar decisiones informadas y de escalar operaciones. Este proyecto ha buscado mejorar la gestión de datos, optimizar la toma de decisiones a través de informes detallados, y aumentar la productividad y satisfacción del cliente mediante una oferta y servicio mejorados.

Carrera: Data Analytics

Módulo 2

Desarrollo del proyecto

DDL avance 1

1. Creación de base de datos
2. Verificación de la creación de la DB
3. No se trabajara con schemas (code review m2l2)
4. Naming conventions: snake_case en nombres de tablas y columnas para mantener consistencia y claridad
5. Creación de tablas según modo semántico (ER)
 - Identificación de entidades [id único - IDENTITY(1,1)]
 - Definir las relaciones
 - Definición del tipo de datos por columna, PK y FK
 - Creación del diagrama con la herramienta: <https://dbdiagram.io/d>
 - Validación del diseño
 - Creación del dataset
6. Guardar datos y crear una copia
7. Creación de tablas en SQL server
8. En el avance 2 dan unas nuevas columnas

DML avance 2

9. **Modificaciones de Tablas:**
 - **Productos:**
 - Se agregó la columna price (DECIMAL).
 - **Cientes:**
 - Se agregó la columna address (VARCHAR).
 - **Origenes de Orden:**
 - Se agregó la columna description (VARCHAR).
 - **Empleados:**
 - Se agregó la columna branch_id (INT, FK).
 - Se agregó la columna department_name (VARCHAR).
 - **Órdenes:**
 - Se agregó la columna employee_id (INT, FK).
 - **Elementos de Orden:**
 - Se modificó la columna price de INT a DECIMAL.
10. **Operaciones:**
 - Se insertaron datos en las tablas categories, products, branches, customers, order_origins, departments, orders, order_items, payment_types y messengers.
 - Se resolvieron conflictos de FK en products y order_items.
 - Se abordaron errores de conversión de tipos de datos en employees.
11. **Consultas:**
 - Conteo de registros únicos en la tabla orders.
 - Número de empleados por departamento.
 - Conteo de productos por categoría.
 - Total de clientes en la tabla customers.
 - Sucursales con ingresos promedio > 1000.00 y costos optimizados en base al promedio de kilómetros recorridos.

Consultas DML avance 3

1. **Total de Ventas Globales**
 - **Descripción:** Calcular el total de ventas acumulado en la tabla de orders.
 - **Proceso:** Se suma el valor de la columna total_amount para obtener el total de ventas global.
2. **Precio Promedio de los Productos por Categoría**
 - **Descripción:** Determinar el precio promedio de los productos en cada categoría.
 - **Proceso:** Se unen las tablas products y categories usando la columna category_id. Luego, se agrupan los resultados por el nombre de la categoría y se calcula el promedio del precio de los productos.

Carrera: Data Analytics

Módulo 2

3. **Valor Mínimo y Máximo de Órdenes por Sucursal**
 - **Descripción:** Encontrar el valor mínimo y máximo de las órdenes en cada sucursal.
 - **Proceso:** Se unen las tablas orders y branches usando la columna branch_id. Se agrupan los resultados por el nombre de la sucursal y se calculan los valores mínimo y máximo de total_amount para cada sucursal.
4. **Mayor Número de Kilómetros Recorridos para una Entrega**
 - **Descripción:** Identificar el mayor número de kilómetros recorridos para una entrega.
 - **Proceso:** Se obtiene el valor máximo de la columna kilometers_to_cover en la tabla de orders.
5. **Promedio de Cantidad de Productos por Orden**
 - **Descripción:** Calcular la cantidad promedio de productos en cada orden.
 - **Proceso:** Se calcula el promedio de la columna quantity en la tabla order_items.
6. **Distribución de la Facturación Total por Métodos de Pago**
 - **Descripción:** Ver cómo se distribuyen las ventas totales según los métodos de pago.
 - **Proceso:** Se unen las tablas orders y payment_types usando payment_type_id. Luego, se agrupan los resultados por el nombre del método de pago y se suman las ventas totales para cada tipo de pago.
7. **Sucursal con el Ingreso Promedio Más Alto**
 - **Descripción:** Encontrar la sucursal con el mayor ingreso promedio.
 - **Proceso:** Se usa una CTE (Common Table Expression) para calcular el ingreso promedio de cada sucursal. Luego, se selecciona la sucursal con el mayor promedio de ventas.
8. **Sucursales con Ventas Totales por Encima de \$1000**
 - **Descripción:** Identificar las sucursales que han generado ventas totales superiores a \$1000.
 - **Proceso:** Se suman las ventas para cada sucursal y se filtran los resultados para mostrar solo aquellas sucursales cuyo total de ventas es mayor a \$1000.
9. **Comparación de Ventas Promedio Antes y Después de una Fecha Específica**
 - **Descripción:** Comparar el promedio de ventas antes y después del 1 de julio de 2023.
 - **Proceso:** Se usa una CTE para clasificar las órdenes en dos periodos (antes y después del 1 de julio de 2023) y se calcula el promedio de ventas para cada periodo.
10. **Distribución de Ventas por Horario del Día**
 - **Descripción:** Analizar las ventas según el horario del día (mañana, tarde, noche).
 - **Proceso:** Se agrupan las órdenes por el horario del día (sale_time). Se cuentan las ventas, se calcula el ingreso promedio y se determina el importe máximo alcanzado por una orden en cada horario del día.

Consultas DML avance 4

1. **Lista de Todos los Productos con Sus Categorías**
 - **Descripción:** Obtener una lista completa de productos junto con sus categorías.
 - **Proceso:** Se realiza una unión izquierda (LEFT JOIN) entre las tablas products y categories usando category_id para combinar productos con sus respectivas categorías. Si un producto no tiene categoría asignada, aparecerá en el resultado con un valor nulo en Category Name.
2. **Sucursal Asignada a Cada Empleado**
 - **Descripción:** Ver a qué sucursal está asignado cada empleado.
 - **Proceso:** Se realiza una unión izquierda (LEFT JOIN) entre las tablas employees y branches usando branch_id para combinar empleados con las sucursales a las que están asignados.

Carrera: Data Analytics

Módulo 2

3. Productos Sin Categoría Asignada

- **Descripción:** Identificar productos que no tienen una categoría asignada.
- **Proceso:** Se realiza una unión izquierda (LEFT JOIN) entre las tablas products y categories. Se filtra para mostrar productos cuyo category_id es nulo. Además, se utiliza una verificación condicional para mostrar un mensaje si no existen productos sin categoría, y en caso contrario, se muestra una lista de productos sin categoría.

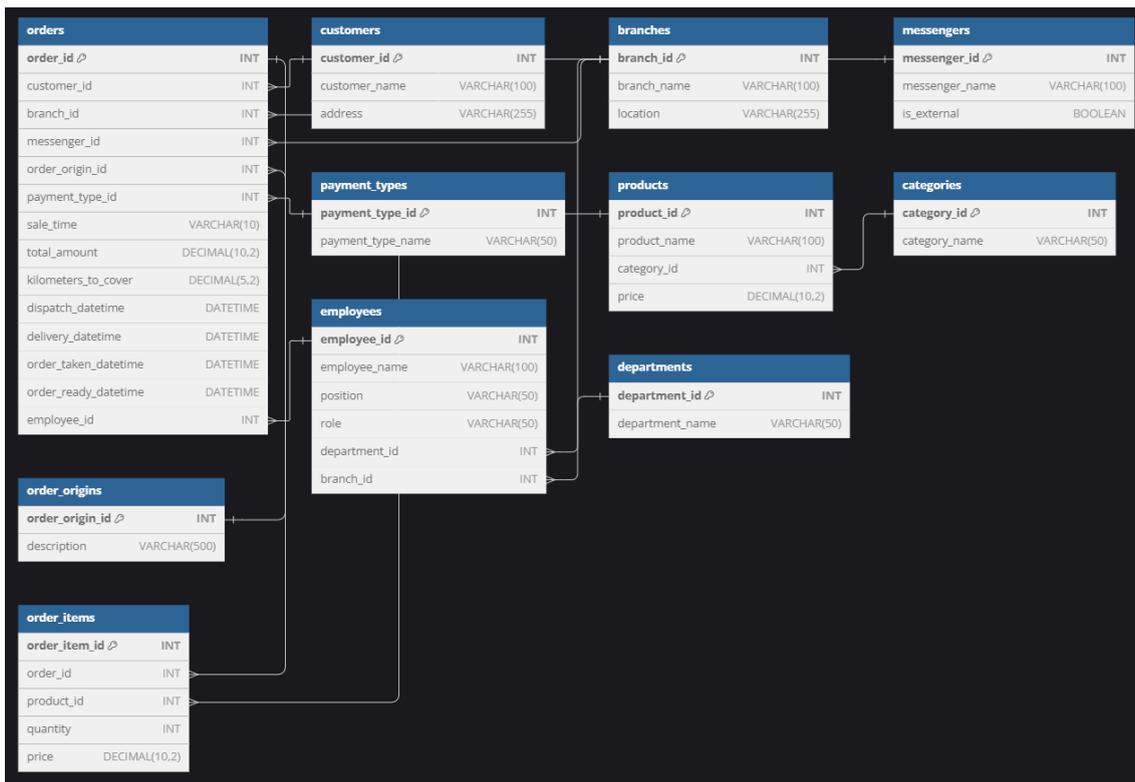
4. Detalle Completo de Ordenes

- **Descripción:** Obtener un detalle completo de las órdenes, incluyendo el nombre del cliente, el nombre del empleado que tomó la orden y el nombre del mensajero que la entregó.
- **Proceso:** Se realiza una unión interna (INNER JOIN) entre las tablas orders, customers, employees y messengers usando las claves correspondientes para combinar la información de cada orden con los detalles del cliente, empleado y mensajero.

5. Número de Artículos Vendidos por Categoría en Cada Sucursal

- **Descripción:** Contar cuántos artículos de cada categoría se han vendido en cada sucursal.
- **Proceso:** Se realiza una unión interna (INNER JOIN) entre las tablas order_items, products, categories, orders y branches para combinar la información de artículos vendidos con categorías y sucursales. Luego, se agrupan los resultados por nombre de sucursal y categoría, y se cuenta el número total de productos vendidos

Resultados y consultas



Esquema relacional de la base de datos final (Dataset)

Hallazgos Clave

1. Tiempo Promedio de Entrega

- **Hallazgo:** El tiempo promedio de entrega es de 30 minutos.
- **Interpretación:** Esto sugiere que en promedio, las entregas se completan en 30 minutos desde el despacho hasta la entrega.

2. Ventas Totales por Canal de Venta

Carrera: Data Analytics

Módulo 2

- **Hallazgo:**
 - **Presencial:** 8560.00
 - **En línea:** 4414.04
 - **Drive Thru:** 4380.00
 - **Teléfono:** 4340.00
- **Interpretación:** El canal de venta presencial genera las mayores ventas, seguido por en línea, drive thru y teléfono.
- 3. **Ingresos Totales Generados por Empleado**
 - **Hallazgo:**
 - **John Doe:** 4414.04
 - **Ethan Martinez:** 4380.00
 - **Olivia García:** 4340.00
 - **Jane Smith:** 4300.00
 - **Lucas Miller:** 4260.00
 - **Interpretación:** John Doe es el empleado que ha generado más ingresos, mientras que Lucas Miller es el que ha generado menos.
- 4. **Análisis de Demanda por Hora del Día**
 - **Hallazgo:**
 - **Mañana:** 8 pedidos
 - **Tarde:** 8 pedidos
 - **Noche:** 4 pedidos
 - **Interpretación:** La demanda de pedidos es igual durante la mañana y la tarde, pero menor durante la noche.
- 5. **Tendencia de Ventas Mensuales**
 - **Hallazgo:**
 - **Enero 2023:** 4414.04
 - **Febrero 2023:** 4300.00
 - **Julio 2023:** 4260.00
 - **Agosto 2023:** 4340.00
 - **Septiembre 2023:** 4380.00
 - **Interpretación:** Las ventas mensuales muestran variaciones a lo largo del año, con un incremento en septiembre y una caída en julio.
- 6. **Porcentaje de Clientes Recurrentes vs. Nuevos**
 - **Hallazgo:** Todos los clientes son recurrentes (100%).
 - **Interpretación:** No hay clientes nuevos, lo que podría indicar una falta de adquisición de nuevos clientes.

Recomendaciones Estratégicas

1. **Tiempo Promedio de Entrega**
 - **Recomendación:** Si el tiempo promedio de entrega es demasiado alto, considerar optimizar las rutas de entrega o mejorar la logística para reducir este tiempo.
2. **Ventas Totales por Canal de Venta**
 - **Recomendación:** Potenciar los canales que generan mayores ventas como el presencial y el en línea. Revisar las estrategias de marketing y ventas para los canales con menor rendimiento.
3. **Ingresos Totales Generados por Empleado**
 - **Recomendación:** Evaluar las estrategias de incentivos y bonificaciones para empleados de alto rendimiento y proporcionar formación adicional a los que generan menos ingresos.
4. **Análisis de Demanda por Hora del Día**
 - **Recomendación:** Ajustar las operaciones para mejorar la cobertura durante las horas con alta demanda y evaluar la posibilidad de ajustar horarios de personal en base a la demanda.
5. **Tendencia de Ventas Mensuales**
 - **Recomendación:** Implementar estrategias de promoción durante los meses con ventas bajas y aprovechar las tendencias de los meses con mayores ventas para planificar campañas y ofertas.
6. **Porcentaje de Clientes Recurrentes vs. Nuevos**

Carrera: Data Analytics

Módulo 2

- **Recomendación:** Desarrollar campañas de marketing y estrategias de adquisición de clientes para atraer nuevos clientes y diversificar la base de clientes recurrentes.

Comentarios y Sugerencias:

- **Optimización del Despacho:** Mejorar la logística y el uso de tecnologías para optimizar las rutas de entrega puede reducir el tiempo promedio de entrega y aumentar la eficiencia operativa.
- **Estrategias de Marketing:** Revisar y fortalecer las estrategias de marketing para canales con menor rendimiento puede aumentar las ventas y diversificar las fuentes de ingresos.
- **Desarrollo del Personal:** Implementar programas de formación y incentivos para empleados puede mejorar el rendimiento y aumentar los ingresos generados por cada empleado.
- **Ajuste de Operaciones:** Adaptar las operaciones para coincidir con la demanda horaria y ajustar el personal según las horas de mayor demanda puede mejorar la eficiencia y reducir costos.
- **Promociones Estacionales:** Aprovechar las tendencias de ventas mensuales para planificar promociones y campañas puede maximizar los ingresos durante los meses de alta demanda y mitigar las caídas en otros periodos.
- **Atraer Nuevos Clientes:** Desarrollar estrategias para atraer nuevos clientes y diversificar la base de clientes puede asegurar un crecimiento sostenido y equilibrado para la empresa.

Optimización y sostenibilidad

Para garantizar que la base de datos fuera eficiente en el análisis externo, se implementaron varias técnicas de optimización. Se crearon índices en columnas clave que se utilizan frecuentemente en consultas, como `dispatch_datetime` en la tabla `orders` y `product_name` en la tabla `products`. Esto permite una recuperación más rápida de datos y mejora el rendimiento general de las consultas. Además, se revisaron y ajustaron las consultas para minimizar el tiempo de ejecución.

Por otro lado, para asegurar la sostenibilidad a largo plazo, se adoptaron prácticas de diseño de base de datos que facilitan la escalabilidad y el mantenimiento. La estructura de la base de datos se diseñó para ser modular, permitiendo la adición de nuevas tablas y relaciones sin afectar el rendimiento del sistema. Se establecieron procedimientos de respaldo regulares y se documentaron las prácticas recomendadas para el mantenimiento de la base de datos, asegurando que el sistema pueda evolucionar con el crecimiento de la empresa.

Desafíos y soluciones

Durante el proyecto, me enfrenté a varios desafíos que requirieron soluciones específicas para garantizar el éxito de la implementación de la base de datos. El primer desafío fue la consolidación de datos dispersos, ya que la información estaba distribuida en múltiples documentos (avance uno y dos), lo que complicaba la creación de una vista unificada. Para superar esto, centralice la información en una base de datos relacional, integrando todos los datos relevantes en tablas estructuradas que mejoraron la accesibilidad y el análisis.

Reflexión personal

A lo largo del proyecto, he adquirido valiosas habilidades en el análisis de bases de datos, la creación de consultas SQL avanzadas y la interpretación de resultados. Estas competencias me han permitido entender mejor cómo estructurar y optimizar bases de datos para cumplir con las necesidades empresariales. Si tuviera que comenzar de nuevo, mantendría el enfoque en la optimización de consultas y en una planificación exhaustiva de la estructura de la base de datos. No obstante, consideraría la implementación de herramientas adicionales para un análisis más

Carrera: Data Analytics

Módulo 2

profundo y para la visualización de datos, mejorando la capacidad de tomar decisiones informadas y ofreciendo insights más detallados sobre el desempeño de la empresa.

EXTRA CREDIT

Expansión de Datos

- 1. Inserción de Datos en la Tabla customers**
 - **Descripción:** Agregar información sobre clientes en la tabla customers.
 - **Proceso:** Se insertan nuevos registros con nombres y detalles de contacto para varios clientes.
- 2. Inserción de Datos en la Tabla branches**
 - **Descripción:** Agregar información sobre sucursales en la tabla branches.
 - **Proceso:** Se insertan registros con nombres y ubicaciones de diferentes sucursales.
- 3. Inserción de Datos en la Tabla messengers**
 - **Descripción:** Agregar datos sobre mensajeros en la tabla messengers.
 - **Proceso:** Se insertan registros con nombres y estatus de los mensajeros (internos o externos).
- 4. Inserción de Datos en la Tabla order_origins**
 - **Descripción:** Añadir información sobre los orígenes de los pedidos en la tabla order_origins.
 - **Proceso:** Se insertan registros que describen los diferentes métodos a través de los cuales se realizan los pedidos.
- 5. Inserción de Datos en la Tabla payment_types**
 - **Descripción:** Insertar tipos de pago en la tabla payment_types.
 - **Proceso:** Se añaden registros para diferentes métodos de pago.
- 6. Inserción de Datos en la Tabla categories**
 - **Descripción:** Añadir categorías de productos en la tabla categories.
 - **Proceso:** Se insertan registros para distintas categorías de productos.
- 7. Inserción de Datos en la Tabla departments**
 - **Descripción:** Agregar información sobre departamentos en la tabla departments.
 - **Proceso:** Se insertan registros con nombres de departamentos.
- 8. Inserción de Datos en la Tabla products**
 - **Descripción:** Insertar productos en la tabla products.
 - **Proceso:** Se añaden registros de productos junto con su categoría correspondiente.
- 9. Inserción de Datos en la Tabla employees**
 - **Descripción:** Añadir información sobre empleados en la tabla employees.
 - **Proceso:** Se insertan registros con detalles sobre empleados, sus sucursales y departamentos.
- 10. Inserción de Datos en la Tabla orders**
 - **Descripción:** Insertar detalles de ordenes en la tabla orders.
 - **Proceso:** Se añaden registros de ordenes con información completa sobre clientes, sucursales, mensajeros, tipos de pago, y tiempos relevantes.

Nuevas Consultas Estratégicas

- 1. Patrones de Consumo Estacionales**
 - **Descripción:** Analizar el número de pedidos por mes.
 - **Proceso:** Se agrupan los pedidos por mes y se cuenta el número total de pedidos en cada mes.
- 2. Preferencias Regionales de Productos**
 - **Descripción:** Evaluar la cantidad de productos vendidos por sucursal.

Carrera: Data Analytics

Módulo 2

- **Proceso:** Se agrupan los productos vendidos por sucursal y se cuenta la cantidad vendida de cada producto.

Optimización de Rendimiento

1. Implementación de Índices

- **Descripción:** Mejorar el rendimiento de consultas mediante la creación de índices en columnas clave.
- **Proceso:** Se crean índices en las columnas `dispatch_datetime` de la tabla `orders` y `product_name` de la tabla `products` para acelerar las búsquedas y consultas.

Consultas de Verificación

1. Productos Sin Categoría

- **Descripción:** Verificar si existen productos sin categoría asignada.
- **Proceso:** Se revisa la tabla `products` para encontrar productos cuyo `category_id` es nulo. Si no existen productos sin categoría, se imprime un mensaje indicando esto.

2. Detalle Completo de Órdenes

- **Descripción:** Obtener detalles completos de las órdenes.
- **Proceso:** Se realiza una unión interna (INNER JOIN) entre las tablas `orders`, `customers`, `employees` y `messengers` para combinar y mostrar detalles completos de las órdenes.

3. Contar Artículos Vendidos por Categoría y Sucursal

- **Descripción:** Contar los artículos vendidos por categoría en cada sucursal.

Proceso: Se agrupan los artículos vendidos por categoría y sucursal, y se cuenta la cantidad total de productos vendidos en cada combinación.