

Carrera: Data Analytics



Módulo 2

‘Descubriendo la base de datos de Fast Food’

Nombre del autor: Facundo Ariel Sardo

Email: sardomkt@gmail.com

Cohorte: DA-FT08

Fecha de entrega: 09/10/2024 - 10/10/2024

Institución



Introducción

El proyecto "Descubriendo la base de datos de Fast Food" tiene como propósito central la creación y consolidación de una base de datos integral para la empresa Fast Food. A través de este proyecto, se busca optimizar la gestión y análisis de la información generada en los procesos de venta, permitiendo un acceso estructurado y eficiente a los datos. Esto permitirá a la organización mejorar su capacidad para tomar decisiones estratégicas basadas en datos, identificar patrones de comportamiento en ventas, evaluar la eficiencia de las sucursales y maximizar la satisfacción del cliente.

Los objetivos organizacionales alcanzados con este proyecto incluyen:

- Centralización y consolidación de los datos de ventas, productos y sucursales en una única fuente confiable.
- Mejora de la capacidad para analizar y comprender las tendencias de ventas y operaciones logísticas.
- Optimización de la toma de decisiones basadas en el análisis de indicadores clave de rendimiento (KPI).
- Escalabilidad de la base de datos para adaptarse al crecimiento futuro del negocio, asegurando su eficiencia y sostenibilidad.

Este proyecto también se enfoca en cumplir con los estándares de buenas prácticas en el manejo de bases de datos, garantizando la seguridad, integridad y coherencia de la información almacenada.

Desarrollo del proyecto

Avance 1

En esta etapa del proyecto, se ejecutó el contenedor de SQL Server utilizando Docker y se logró la conexión exitosa a SQL Server tanto desde Azure como desde DBeaver. A partir de allí, se elaboró el diagrama entidad-relación junto con el modelo relacional, lo que permitió crear la base de datos fast_food. Posteriormente, se crearon nueve tablas, estableciendo tanto las claves primarias como las foráneas, y se definieron relaciones clave entre las tablas de sucursales, empleados, productos y orígenes. Además, se seleccionaron los campos más relevantes para cada tabla y se elaboró el esquema relacional correspondiente.

Avance 2

En este avance del proyecto, se reemplazaron los nombres de varias tablas por los correspondientes, además de elaborar las tablas faltantes. También se modificaron los campos, eliminando o agregando nuevos según las necesidades del modelo de datos. Se procedió a la inserción de datos en las tablas y se realizaron actualizaciones en las mismas, ajustando o eliminando datos cuando fue necesario.

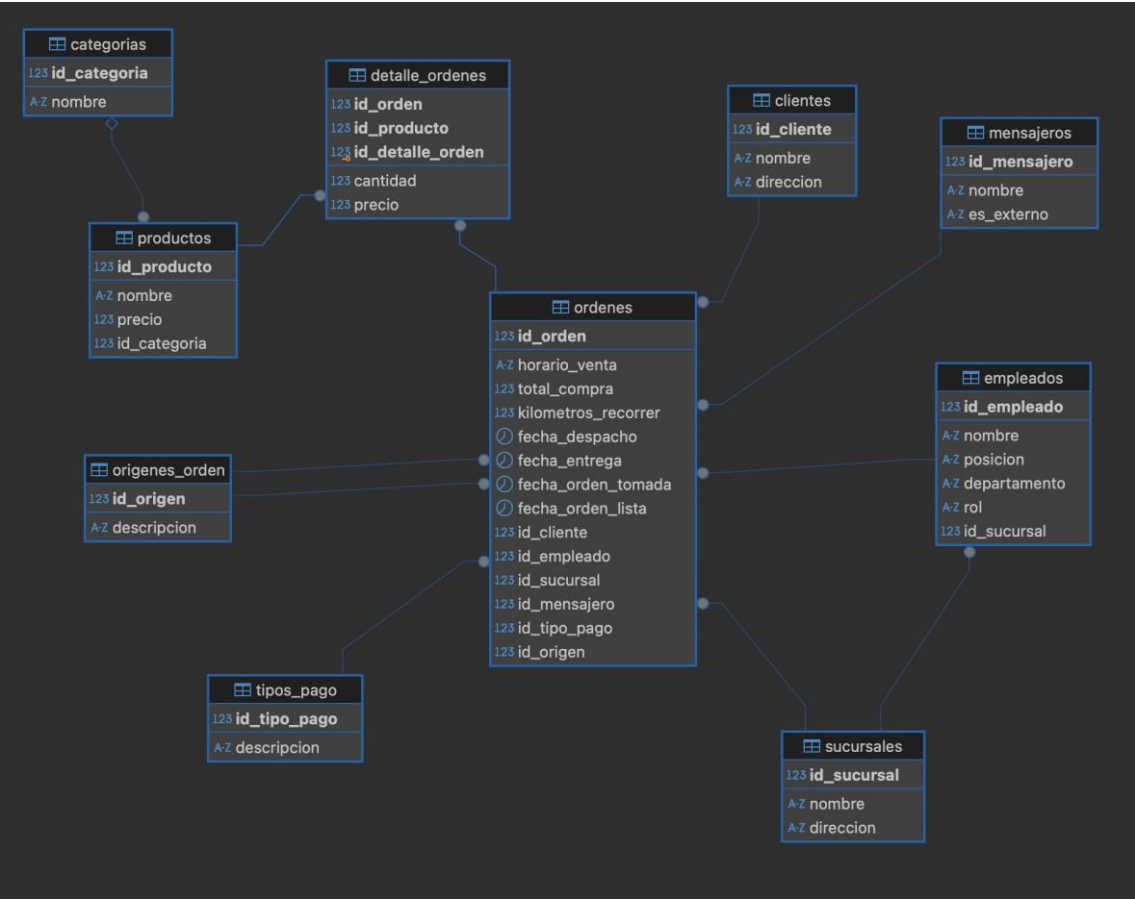
Módulo 2

Avance 3

En este avance del proyecto, se utilizaron funciones de agregación para extraer información clave de la base de datos. Se aplicaron cláusulas para organizar los resultados según diversas categorías, tales como sucursales y métodos de pago. Se respondieron diez preguntas de negocio específicas relacionadas con ventas, productos y sucursales. Además, se llevaron a cabo consultas avanzadas para obtener promedios, totales, máximos y mínimos en distintas dimensiones de los datos. Finalmente, se documentaron todas las consultas utilizando comentarios explicativos, con el objetivo de mejorar la comprensión del código.

Avance 4

En el avance del proyecto, se realizaron cinco consultas utilizando JOINS que respondieron a preguntas específicas sobre productos, categorías, empleados y ventas, lo que permitió obtener resultados más completos. Además, se implementaron alias para mejorar la claridad y legibilidad del código SQL, facilitando así la interpretación de las consultas.



Módulo 2

```
58
59 -- 1) Eficiencia de los mensajeros --
60 -- ¿Cuál es el tiempo promedio desde el despacho hasta la entrega de los pedidos gestionados por todo el equipo de mensajería? --
61
62 SELECT AVG(DATEDIFF(MINUTE, fecha_despacho, fecha_entrega)) tiempo_promedio_minutos
63 FROM ordenes
64 WHERE fecha_entrega IS NOT NULL;
```

Results Messages

	tiempo_promedio_minutos ▾
1	30

```
66 -- 2) Análisis de ventas por origen de orden: ¿Qué canal de ventas genera más ingresos? --
67
68 SELECT T1.id_origen, SUM( T1.total_compra ) total_ingresos
69 FROM ordenes T1
70 GROUP BY id_origen
71 ORDER BY total_ingresos DESC;
```

Results Messages

	id_origen ▾	total_ingresos ▾
1	2	2140.00
2	4	2025.00
3	3	2005.00
4	1	1998.51
5	5	955.00

```
73 -- 3) Productividad de los empleados: ¿Cuál es el nivel de ingreso generado por empleado? --
74
75 SELECT T1.nombre, T1.posicion, SUM(T2.total_compra) AS total_ingresos
76 FROM Empleados T1
77 INNER JOIN ordenes T2 ON T1.id_empleado = T2.id_empleado
78 GROUP BY T1.nombre, T1.posicion
79 ORDER BY total_ingresos DESC;
```

Results Messages

	nombre ▾	posicion ▾	total_ingresos ▾
1	Ethan Martinez	Mesero	1095.00
2	Olivia García	Encargado de Turno	1085.00
3	Jane Smith	Subgerente	1075.00
4	Lucas Miller	Atención al Cliente	1065.00
5	John Doe	Gerente	1053.51
6	Tom Brown	Barista	955.00
7	Emma Davis	Repartidor	945.00
8	Alice Johnson	Chef	930.00
9	Bill Jones	Cajero	920.00

Módulo 2

```

80
81 -- 4) Análisis de demanda por horario y día: ¿Cómo varía la demanda de productos a lo largo del día? --
82
83 SELECT DATEPART(HOUR, fecha_orden_tomada) AS hora, COUNT(*) cantidad_ordenes, SUM(total_compra) total_compra
84 FROM ordenes
85 GROUP BY DATEPART( HOUR, fecha_orden_tomada)
86 ORDER BY Hora

```

Results Messages

	hora	cantidad_ordenes	total_compra
1	8	2	2118.51
2	9	1	930.00
3	13	1	1075.00
4	14	1	1085.00
5	15	1	955.00
6	19	2	2015.00
7	20	1	945.00

```

88 -- 5) Comparación de ventas mensuales: ¿Cuál es la tendencia de los ingresos generados en cada periodo mensual? --
89
90 SELECT YEAR(fecha_orden_tomada) año, MONTH(fecha_orden_tomada) mes, SUM(total_compra) total_ingresos
91 FROM ordenes
92 GROUP BY YEAR(fecha_orden_tomada), MONTH(fecha_orden_tomada )
93 ORDER BY YEAR(fecha_orden_tomada) DESC;
94

```

Results Messages

	año	mes	total_ingresos
1	2023	1	1053.51
2	2023	2	1075.00
3	2023	3	920.00
4	2023	4	930.00
5	2023	5	955.00
6	2023	6	945.00
7	2023	7	1065.00
8	2023	8	1085.00
9	2023	9	1095.00

```

94
95 -- 6) Análisis de fidelidad del cliente: ¿Qué porcentaje de clientes son recurrentes versus nuevos clientes cada mes? --
96
97 WITH ordenespormes AS (
98 SELECT id_cliente, YEAR(fecha_orden_lista) AS año, MONTH(fecha_orden_lista) AS mes, COUNT(*) AS num_ordenes
99 FROM ordenes
100 GROUP BY id_cliente, YEAR(fecha_orden_lista), MONTH(fecha_orden_lista))
101 SELECT año, mes,
102 COUNT(CASE WHEN num_ordenes > 1 THEN 1 END) AS clientes_recurrentes,
103 COUNT(CASE WHEN num_ordenes = 1 THEN 1 END) AS clientes_nuevos,
104 FORMAT(ROUND(COUNT(CASE WHEN num_ordenes > 1 THEN 1 END) * 100.0 / COUNT(*), 2), 'P2') AS porcentaje_recurrentes
105 FROM OrdenesPorMes
106 GROUP BY año, mes;

```

Results Messages

	año	mes	clientes_recurrentes	clientes_nuevos	porcentaje_recurrentes
1	2023	1	0	1	0.00%
2	2023	2	0	1	0.00%
3	2023	3	0	1	0.00%
4	2023	4	0	1	0.00%
5	2023	5	0	1	0.00%
6	2023	6	0	1	0.00%
7	2023	7	0	1	0.00%
8	2023	8	0	1	0.00%
9	2023	9	0	1	0.00%

Optimización y sostenibilidad

La base de datos fast_food fue optimizada para el análisis externo a través de diversas estrategias. Se implementaron índices en las columnas más consultadas, lo que mejoró significativamente la velocidad de las consultas y el rendimiento general del sistema. Además, se estructuraron las tablas de manera que se minimizaran las redundancias y se maximizaran las relaciones entre los datos, facilitando así el acceso y la interpretación de la información.

Para garantizar la sostenibilidad a largo plazo, se establecieron procedimientos de mantenimiento regular, incluyendo copias de seguridad programadas y actualizaciones de software. Se documentaron las estructuras y relaciones de las tablas, lo que facilita la incorporación de nuevas funcionalidades en el futuro y asegura que el equipo de trabajo pueda adaptarse a posibles cambios en los requisitos del negocio. También se consideró el uso de herramientas de monitoreo para identificar y solucionar problemas de rendimiento antes de que afecten a los usuarios finales.

Desafíos y soluciones

Desafío: Conexión a SQL Server a través de Docker

Solución: Se configuró correctamente el contenedor de SQL Server utilizando Docker y se verificó la conexión desde Azure y DBeaver, asegurando una comunicación efectiva entre las herramientas y la base de datos.

Desafío: Creación de un modelo relacional adecuado

Solución: Se elaboró un Diagrama Entidad-Relación que facilitó la visualización de las relaciones entre las tablas, asegurando la correcta definición de claves primarias y foráneas, así como la estructura lógica de la base de datos.

Desafío: Eliminación de redundancias en las tablas

Solución: Se revisaron las tablas y se eliminaron las redundancias, como en el caso de la tabla orígenes, que fue eliminada para simplificar la estructura y mejorar la integridad de los datos.

Desafío: Mantenimiento de la calidad de datos al insertar y actualizar registros

Solución: Se implementaron validaciones al momento de insertar y actualizar datos, utilizando sentencias SQL adecuadas para asegurar que los datos cumplan con los criterios de calidad y consistencia.

Desafío: Análisis de datos de ventas en diferentes periodos

Solución: Se diseñaron consultas SQL específicas que utilizaron funciones de agregación y cláusulas para responder a preguntas sobre el comportamiento de las ventas, facilitando la comparación de datos antes y después de una fecha específica.

Desafío: Obtención de resultados completos mediante JOINS

Solución: Se realizaron consultas que involucraron múltiples tablas usando JOINS, permitiendo la integración de información relevante sobre productos, categorías, empleados y ventas, mejorando así la calidad del análisis.

Desafío: Optimización de consultas para el análisis externo

Solución: Se añadieron índices en columnas clave lo que mejoró la velocidad de las consultas y optimizó el rendimiento general de la base de datos.

Desafío: Documentación y mantenimiento de la base de datos

Solución: Se creó documentación exhaustiva sobre la estructura de la base de datos, las relaciones entre las tablas y los procedimientos de mantenimiento, lo que asegura que futuros desarrollos sean más fáciles y eficientes.

Reflexión personal

A lo largo de este proyecto, he aprendido significativamente sobre el manejo y la organización de bases de datos, así como sobre la importancia de estructurar la información de manera lógica y eficiente. Esta experiencia marcó mi primer contacto con la programación y el mundo de las bases de datos, lo que hizo que cada aspecto del proyecto fuera muy especial. La experiencia de crear la base de datos desde cero, definir las tablas, establecer relaciones y luego importar datos ha sido muy interesante y enriquecedora. Cada etapa del proceso me permitió comprender mejor la estructura subyacente de los datos y cómo se relacionan entre sí.

Elaborar consultas SQL complejas y realizar análisis de datos me ha permitido adquirir habilidades valiosas que son fundamentales en el campo del análisis de datos. Además, el proceso de trabajar con herramientas como SQL Server, Docker y DBeaver ha mejorado mi capacidad para conectar diferentes tecnologías y optimizar la gestión de datos. También he comprendido la relevancia de documentar cada paso del proceso, lo que facilita el trabajo en equipo y la sostenibilidad del proyecto a largo plazo.

Si tuviera que volver a empezar este proyecto, definitivamente aplicaría los conocimientos adquiridos. Sin embargo, me gustaría prestar más atención al diseño inicial de la base de datos, asegurándome de que cada tabla y relación estén bien definidas desde el comienzo. También consideraría involucrar a otros miembros del equipo desde etapas tempranas para obtener diferentes perspectivas y enriquecer el proceso de toma de decisiones.

En resumen, esta experiencia ha sido transformadora y me brindó un valioso primer contacto con la programación y el análisis de datos. Estoy agradecido por la oportunidad de aprender y crecer como Analista de Datos en Henry.

Extra Credit

Expansión de datos

Durante la fase de desarrollo de la base de datos para la cadena de comida rápida, se identificó la necesidad de ampliar la información disponible para obtener una visión más completa y detallada de las operaciones. La base de datos inicial era relativamente pequeña, lo que limitaba el análisis y la toma de decisiones basadas en datos. La expansión de datos es un enfoque estratégico que permite enriquecer la base de datos con conjuntos de datos adicionales, como información sobre proveedores, análisis de mercado, y datos demográficos de los clientes. Esta inclusión puede proporcionar una base más sólida para entender las tendencias del mercado y mejorar la eficiencia operativa. Al agregar nuevos conjuntos de datos, se pueden generar análisis más profundos sobre el comportamiento del consumidor y la dinámica del negocio.

Nuevas consultas estratégicas

Como parte del desarrollo del proyecto, se diseñaron consultas adicionales que abordan preguntas estratégicas específicas para el negocio de "Fast Food". Estas consultas son esenciales para extraer información relevante que puede influir en la toma de decisiones. En particular, se realizó una consulta para investigar patrones de consumo estacional. Este tipo de análisis permite identificar tendencias en las preferencias de los clientes durante diferentes épocas del año, lo que es crucial para planificar campañas de marketing, promociones y la gestión del inventario. Además, se desarrolló una consulta para realizar análisis de tendencias a lo largo del tiempo. Este enfoque permite observar cómo las preferencias de los consumidores cambian y evolucionan con los años, ayudando a la empresa a anticipar y adaptarse a las demandas del mercado. Con el tiempo, estas tendencias pueden ofrecer información valiosa sobre el ciclo de vida de los productos, facilitando decisiones más informadas sobre qué productos introducir, mantener o eliminar. En resumen, estas consultas estratégicas son un paso importante hacia la optimización del negocio y el aprovechamiento de los datos para impulsar el crecimiento y la satisfacción del cliente.