**TÍTULO DEL PROYECTO: “Descubriendo la base de datos de Fast Food”**

[M2 PROYECTO INTEGRADOR](https://drive.google.com/drive/folders/1i7qMBX8_Esi5lLrE4uLLOf8lof-egqrJ?usp=sharing)

**Nombre del autor:** Jaime Alberto Orejarena Carvajal

**Email:** [orejarenajaime1979@gmail.com](mailto:orejarenajaime1979@gmail.com)

**Cohorte:** DAPT-01

**Fecha de entrega:**

**Institución: FAST FOOD**

****

**INTRODUCCIÓN**

Como Analista de Datos, me complace presentar el informe final que abarca el proceso completo de análisis de datos para la empresa FAST FOOD (Best delivery), desde la creación de la base de datos hasta la generación de insights significativos para la toma de decisiones estratégicas.

El objetivo primordial de este informe es destacar las distintas etapas involucradas en el ciclo de vida de los datos, desde la concepción inicial de la estructura de la base de datos hasta la extracción de perspectivas valiosas a través de consultas avanzadas y análisis comparativos entre las tablas.

En este informe, abordaremos las siguientes etapas clave:

**1. Esquema Relacional y Creación de la Base de Datos:**

En esta fase, está incluida la elaboración del esquema relacional, se describe el proceso detallado de diseño y creación de la base de datos en Microsoft SQL Server, lo que representa la estructura lógica de la base de datos.

**2. Inserción de Datos y Gestión de la Información:**

Aquí, se detalla el proceso de inserción de datos en las tablas creadas previamente, así como las estrategias para mantener la integridad y coherencia de los datos mediante operaciones de eliminación y actualización, aquí podemos ver la fortaleza de nuestro esquema relacional y así garantizar la optimización y sostenibilidad de los datos y de la arquitectura del esquema relacional

**3. Consultas con Funciones de Agregación:**

Esta sección se centra en el uso de funciones de agregación en consultas SQL para calcular estadísticas y métricas clave dentro de cada tabla, lo que nos permite descubrir patrones y tendencias significativas en los datos.

**4. Consultas con "JOIN" y "UNION":**

En esta etapa, exploraremos el poder de las operaciones de "JOIN" y "UNION" para combinar datos de múltiples tablas y realizar análisis comparativos entre diferentes conjuntos de datos, lo que nos proporcionará una visión más completa, contextualizada y detallada de la información.

**5. Informe Final:**

Por último, consolidaremos nuestros hallazgos en un informe final que sintetiza los insights más relevantes y las recomendaciones basadas en el análisis realizado, con el objetivo de respaldar la toma de decisiones informadas por parte de la alta dirección.

Cada una de estas etapas desempeña un papel crucial en el proceso de análisis de datos, y su comprensión integral nos permite aprovechar al máximo el potencial de los datos para impulsar el éxito de **FAST FOOD.**

Vamos a explorar en detalle cada una de estas etapas y sus implicaciones en el contexto del análisis de datos utilizando Microsoft SQL Server.

**DESARROLLO DEL PROYECTO**

**Avance 1:** Diseñar el esquema o estructura de la base de datos. (Sublenguaje DDL)

* **Creación base datos:** Como buena práctica y especificado en la query del avance 1, creamos la base de datos con especificaciones dadas para guardar en el disco C, el nombre de la carpeta es “SQL\_DB”
* **Creación de tablas:** Iniciamos con la sentencia “USE FastFoodDB”
* Como una buena práctica en grupo, decidimos iniciar la creación de las tablas a través de un borrador en excel del esquema relacional y así determinar el modelo semántico de cada tabla, identificando entidades y atributos de las tablas, establecimos relaciones entre tablas definiendo específicamente las claves primarias y secundarias, definimos restricciones para especificar las reglas de datos que no deberían ir nulos, estas dos actividades para garantizar la integridad referencial de la base de datos, además para normalizar, para minimizar la redundancia y mejorar la eficiencia en el almacenamiento y manipulación de datos, al mismo tiempo definimos el tipo de datos.
* Posteriormente seguimos a la creación de las tablas de acuerdo al modelo semántico establecido en el punto anterior.
* Por último ejecutamos el esquema relacional de las tablas para identificar visualmente entidades y atributos, las relaciones entre tablas y restricciones.
* **Archivo Queries:** Guardamos las queries creadas con el nombre de “DA\_Orejarena\_Jaime\_avance1PI”. Adjuntas a este informe.

**Avance 2:** Población de tablas mediante queries. Consultas generales (Sub Lenguaje DML 1).

* **Poblar las tablas:** Utilizamos la sentencia **INSERT** para agregar datos a las tablas previamente creadas. (Queries proporcionadas en el avance)
* **Actualizar datos de tablas:** Usamos las sentencias UPDATE para actualizar y datos de las tablas, ejemplo:
* Tabla productos, aumentando el precio en 1 de la CategoríaID 1
* En la tabla empleados actualizamos la posición “cocina” por “chef”.
* En la tabla sucursales a la SucursalID 1 le actualizamos la dirección.
* **Eliminar datos de tablas:** Usamos las sentencias DELETE, para eliminar datos de las tablas:
* Se solicitó eliminar la **Orden ID = 10** de la base de datos, por tal razón para ejecutar la sentencia **DELETE** la tuvimos que hacer para la tabla DetalleOrden y la tabla Ordenes, debido a la relación entre las tablas a través de las primary keys y una primary key compuesta.
* La organización también decidió eliminar la **Categoría ID = 2** correspondiente a “postres”, las tablas referidas, productos y categorías tienen una relación a través de Foreign Keys, por lo tanto tuvimos que desactivar y activar nuevamente la Foreign Keys para eliminar la categoría postres y garantizar la sostenibilidad y disposición de la tabla.
* Por último la organización decidió cerrar la **Sucursal ID = 10** necesitaba eliminar todos los empleados asociados o relacionados a esta sucursal. Para esta sentencia no fue necesario asociar tablas o desactivar claves, solo indicamos que de la tabla empleados borrara la Sucursal asociada número 10.
* **Consultas básicas y generales:** Aplicamos el sublenguaje DML para responder las preguntas que se plantean, usamos la sentencia **SELECT** para seleccionar las columnas de una tabla establecida y **FROM** para determinar la tabla a Consultar, iniciamos con un conocimiento debutante de algunas funciones de agregación.
* **Consulta 1** ¿Cuál es la cantidad total de registros únicos en la tabla de órdenes?
* **Consulta 2** ¿Cuántos empleados existen en cada departamento?
* **Consulta 3** ¿Cuántos productos hay por código de categoría?
* **Consulta 4** ¿Cuántos clientes se han importado a la tabla de clientes?
* **Consulta 5** ¿Cuáles son las sucursales con un promedio de ventas por orden superior a $40, ordenadas por el promedio de kilómetros recorridos para las entregas de mayor a menor?
* **Archivo Queries:** Guardamos las queries creadas con el nombre de “DA\_Orejarena\_Jaime\_avance2PI”. Adjuntas a este informe.

**Avance 3:** Consultas con funciones de agregación (Sub Lenguaje DML 2).

* **Aplicación del sublenguaje DML:** Para responder las preguntas que se plantean, usamos funciones de agregación como **AVG, SUM, MAX, MIN, COUNT.** Utilizamos **GROUP BY** para agrupar la(s) columna(s) consultada(s), **ORDER BY** para ordenar la función de agregación, **WHERE** y **HAVING** para agrupar de acuerdo a un valor específico, también usamos la función **CAST** para convertir valores de un tipo a otro y la sentencia **UNION** para unir dos consultas.
* **Consultas para responder a las preguntas establecidas por la organización:**
* **Ventas globales** ¿Cuál es el total de ventas a nivel global?
* **Promedio de precios de productos por categoría** ¿Cuál es el precio promedio de los productos dentro de cada categoría?
* **Orden mínima y máxima por sucursal** ¿Cuál es el valor de la orden mínima y máxima por cada sucursal?
* **Mayor número de kilómetros recorridos para entrega** ¿Cuál es el mayor número de kilómetros recorridos para una entrega?
* **Promedio de cantidad de productos por orden** ¿Cuál es la cantidad promedio de productos por orden?
* **Sucursal con la venta promedio más alta**  ¿Cuál sucursal tiene la venta promedio más alta?
* **Sucursal con la mayor cantidad de ventas por encima de un umbral** ¿Cuáles son las sucursales que han generado ventas por orden por encima de $50?
* **Comparación de ventas promedio antes y después de una fecha específica** ¿Cómo se comparan las ventas promedio antes y después del 1 de julio de 2023?
* **Análisis de actividad de ventas por horario** ¿Durante qué horario del día (mañana, tarde, noche) se registra la mayor cantidad de ventas, cuál es el valor promedio de estas ventas, y cuál ha sido la venta máxima alcanzada?
* **Archivo Queries:** Guardamos las queries creadas con el nombre de “DA\_Orejarena\_Jaime\_avance3PI”. Adjuntas a este informe.

**Avance 4:** Consultas con múltiples tablas y JOINs.

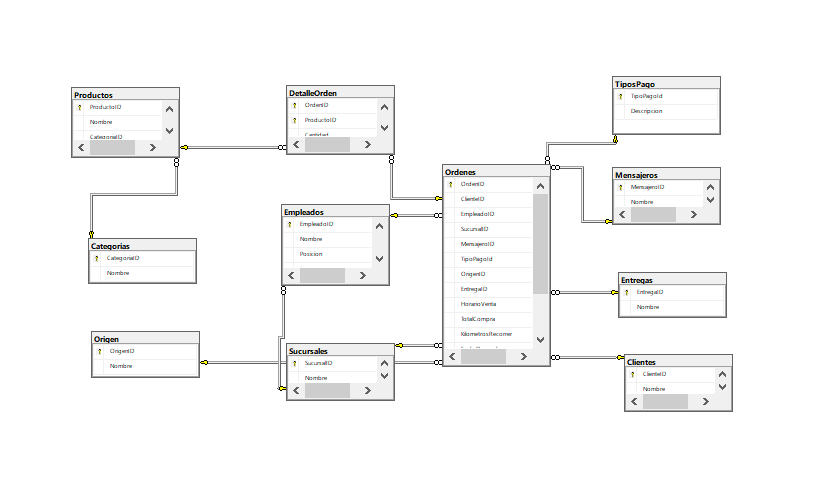
* **Aplicación de JOINs y Alias:** aprendimos mediante la práctica a usar **INNER JOIN**, **LEFT JOIN**, **RIGHT JOIN** y **FULL JOIN** múltiples consultas con relación de las diferentes tablas, aprendimos igualmente a usar los “alias” para mejorar la legibilidad y claridad del código en SQL. Respondimos a preguntas específicas sobre total de ventas por categoría y origen de venta además abordamos problemas que involucran empleados, tiempos de despacho y otros aspectos operativos.
* **Consultas para responder a las preguntas establecidas por la organización:**
* **Listar todos los productos y sus categorías** ¿Cómo puedo obtener una lista de todos los productos junto con sus categorías?
* **Obtener empleados y su sucursal asignada** ¿Cómo puedo saber a qué sucursal está asignado cada empleado?
* **Identificar productos sin categoría asignada** ¿Existen productos que no tienen una categoría asignada?
* **Detalle completo de órdenes** ¿Cómo puedo obtener un detalle completo de las órdenes, incluyendo cliente, empleado que tomó la orden, y el mensajero que la entregó?
* **Productos vendidos por sucursal** ¿Cuántos productos de cada tipo se han vendido en cada sucursal?
* **Archivo Queries:** Guardamos las queries creadas con el nombre de “DA\_Orejarena\_Jaime\_avance4PI”. Adjuntas a este informe.

# 

# 

# **RESULTADOS Y CONSULTAS**

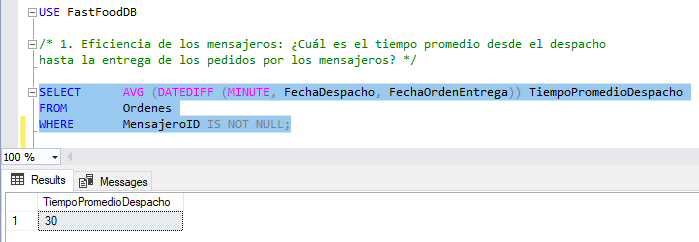
1. **Esquema Relacional base de datos FastFoodDB**

****

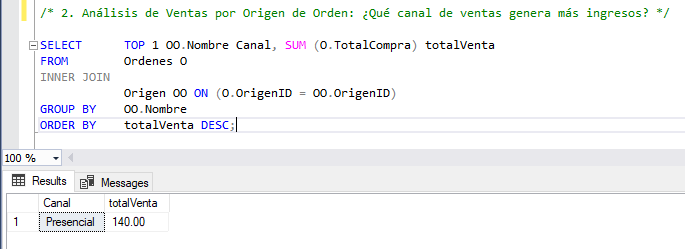
# 

1. **Consultas finales:**

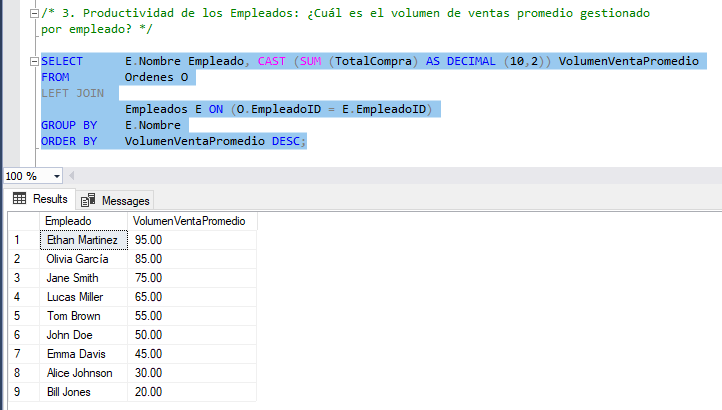
* **Consulta 1 Eficiencia de los mensajeros:** ¿Cuál es el tiempo promedio desde el despacho hasta la entrega de los pedidos por los mensajeros?



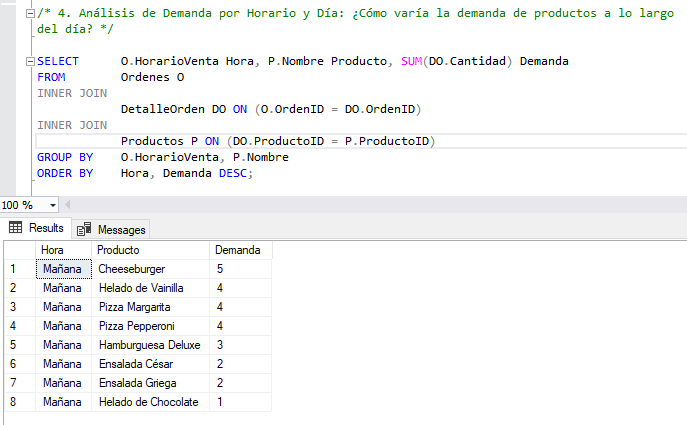
* **Consulta 2 Análisis de Ventas por Origen de Orden:** ¿Qué canal de ventas genera más ingresos?



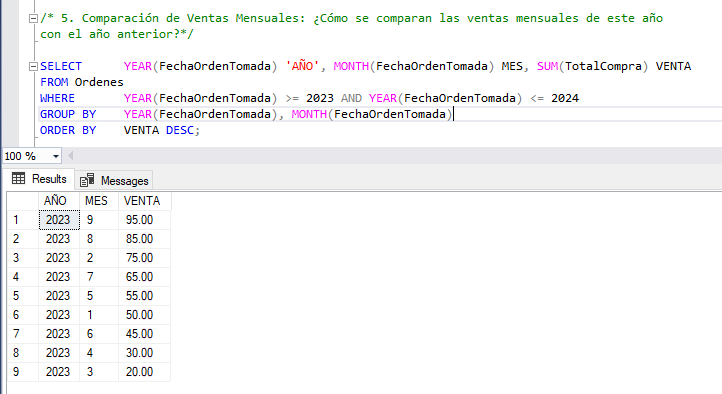
* **Consulta 3 Productividad de los empleados:** ¿Cuál es el volumen de ventas promedio gestionado por empleado?



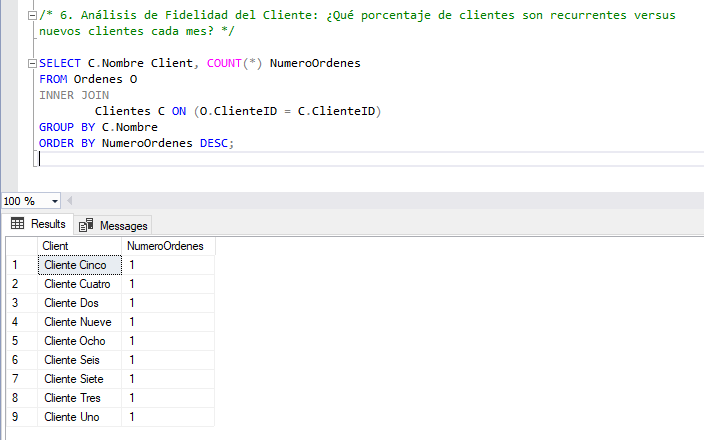
* **Consulta 4 Análisis de Demanda por Horario y Día**: ¿Cómo varía la demanda de productos a lo largo del día?



* **Consulta 5 Comparación de Ventas Mensuales**: ¿Cómo se comparan las ventas mensuales de este año con el año anterior?



* **Consulta 6 Análisis de Fidelidad del Cliente**: ¿Qué porcentaje de clientes son recurrentes versus nuevos clientes cada mes?



1. **Hallazgos claves:**

* **Consulta 1 Eficiencia de los mensajeros:** ¿Cuál es el tiempo promedio desde el despacho hasta la entrega de los pedidos por los mensajeros?

Se evidencio un tiempo promedio de 30 minutos para las entregas, indagando en otras consultas, nos dimos cuenta que todos los mensajeros tenían un tiempo de entrega del producto de 30 minutos, esto significa: que se tiene buen control de los tiempos de entrega o que faltan datos para realizar un buen análisis de tiempos y movimiento para medir la eficiencia de los mensajeros.

* **Consulta 2 Análisis de Ventas por Origen de Orden:** ¿Qué canal de ventas genera más ingresos?

Se evidencio que el canal de ventas que genera más ganancias es el presencial con un total de ventas de $140. Correspondiente a un 26,9% del total de ventas.

* **Consulta 3 Productividad de los empleados:** ¿Cuál es el volumen de ventas promedio gestionado por empleado?

Se evidencio que el empleado con mayor promedio o cantidad de ventas medida es dinero fue Ethan Martinez, que de acuerdo a la consulta tiene unas ventas de $95 que comparada con el total de las compras de la compañía que son $520, esto corresponde al 18,27%.

* **Consulta 4 Análisis de Demanda por Horario y Día**: ¿Cómo varía la demanda de productos a lo largo del día?

Se evidencio que el producto más vendido fue la “Cheeseburger” durante la jornada de la mañana, además no se evidencio coincidencia de Órdenes y productos para los otros horarios de la jornada de trabajo.

* **Consulta 5 Comparación de Ventas Mensuales:** ¿Cómo se comparan las ventas mensuales de este año con el año anterior?

Se evidencio que el mes que más ventas tuvo del año 2023 fue el mes 9 correspondiente al mes de septiembre y el que menos ventas tuvo del año 2023 fue el mes 3 correspondiente al mes de marzo, se evidencia inexistencia de datos para los meses de Octubre (10), Noviembre (11) y Diciembre (12), además no se puede hacer un comparativo con el año 2024 por que no existen registros de datos para los meses de ese año.

* **Consulta 6 Análisis de Fidelidad del Cliente:** ¿Qué porcentaje de clientes son recurrentes versus nuevos clientes cada mes?

Se evidencio que no hubo recurrencia de clientes ya que todos son nuevos. Todos los clientes tuvieron una compra, por lo tanto no se puede evidenciar fidelidad por parte de los clientes.

1. **Recomendaciones estratégicas:**

* **Consulta 1 Eficiencia de los mensajeros:** para esta consulta se recomienda estratégicamente no solo medir la eficiencia de la entrega de los mensajeros, sino también el tiempo de entrada de la orden y la salida de preparación, además como se evidencia con el hallazgo que todos los empleados tienen un tiempo de 30 minutos, recomendaría hacer este análisis cuando se cuente con más datos disponibles con respecto a las órdenes despachadas. También recomendaria la implementación de nuevas métricas a partir de un análisis de tiempos y movimientos para analizar la eficiencia de los mensajeros, además se podría considerar la implementación de herramientas de seguimiento como sistemas GPS o aplicaciones, para mejorar el control en tiempo real.
* **Consulta 2 Análisis de Ventas por Origen de Orden:** Dado que el canal presencial tiene una representación significativa de los ingresos de la compañía, se recomienda el fortalecimiento de este canal y su mejora continua, por ejemplo podría considerarse: optimización de la experiencia del cliente, mejorar la experiencia del cliente en los puntos de venta físicos para fomentar la lealtad y aumentar las ventas; diversificación de productos y ofertas especiales, aumentar la variedad de productos disponibles en los puntos de venta físicos y ofrecer ofertas especiales y promociones exclusivas para atraer a más clientes y aumentar las ventas.
* **Consulta 3 Productividad de los empleados:** una excelente recomendación sería la implementación de programas de reconocimiento y recompensa para motivar y mantener el alto rendimiento de Ethan, así como de otros empleados destacados, además es importante analizar las acciones y estrategias implementadas por Ethan para obtener los resultados y así compartirlas con otros empleados, fomentando las buenas prácticas entre compañeros, por último es crucial invertir en el desarrollo profesional de Ethan, esto aumentará el compromiso con la empresa y lo retara y prepara para roles de liderazgo dentro de la organización.
* **Consulta 4 Análisis de Demanda por Horario y Día:** la primera recomendación sería la optimización del menú, dado que la cheeseburger es el producto estrella de la mañana, es importante optimizar la oferta, aumentando la disponibilidad de esta en este horario, además, como no se evidencia una demanda significativa de otros productos durante los otros horarios de la jornada de trabajo, se podría revisar y optimizar el menú para reflejar mejor los patrones de consumo del cliente a lo largo del día. Esto podría implicar la introducción de nuevos productos o la eliminación de aquellos que tienen una demanda baja en ciertos horarios, por último se recomienda la optimización de inventarios para garantizar la disponibilidad de ingredientes, haciendo un seguimiento preciso de las ventas por hora.
* **Consulta 5 Comparación de Ventas Mensuales:** Dado que no se encontraron datos completos para el año 2023 y para lo corrido del año 2024, se recomienda la mejorar los procesos de recopilación y análisis de datos para obtener una visión más completa y precisa del desempeño de las ventas a lo largo del tiempo, entre los cuales estaría la Implementación de un sistema de Registro de datos y un proceso para la validación de los mismos, paso a seguir cuando se cuente con los registros requeridos realizar un análisis comparativo de ventas y por último la planificación de los recursos necesarios y estrategias para las ventas.
* **Consulta 6 Análisis de Fidelidad del Cliente:** Basándonos en el hallazgo de que todos los clientes son nuevos y no hay recurrencia de clientes, se podría sugerir la siguiente recomendación estratégica: Desarrollo de Estrategias de Retención de Clientes, Dado que todos los clientes son nuevos y no se observa recurrencia en las compras, es crucial desarrollar estrategias efectivas para fomentar la fidelidad y retención de clientes a largo plazo.

# **OPTIMIZACIÓN Y SOSTENIBILIDAD**

Durante la creación de nuestra base de datos 'FastFoodDB', hemos enfocado nuestros esfuerzos en garantizar su optimización y sostenibilidad a largo plazo, considerando el continuo crecimiento del negocio. Para lograrlo, hemos empleado el sublenguaje DDL de manera efectiva, estableciendo cuidadosamente el modelo semántico para cada tabla.

Esto incluyó la declaración concienzuda de las **PRIMARY KEYS**, **FOREIGN KEYS** y **PRIMARY KEYS COMPUESTAS**, así como la especificación de tipos de datos y el esquema de relaciones. Esta atención al detalle se refleja en la robustez de nuestra base de datos; incluso al ingresar o eliminar datos, y al realizar consultas, la estructura y la arquitectura de la misma permanecen intactas.

Esta sólida ingeniería o arquitectura de datos garantiza una estructura de datos flexible, acceso eficiente de datos en todo momento, capaz de desenvolverse de la mejor forma, junto con la organización en este mundo cambiante y de crecimiento de los negocios.

# **DESAFÍOS Y SOLUCIONES**

Durante todo este modulo y el desarrollo del proyecto integrador hubo muchos desafíos, pero en general todos los pudimos desarrollar durante los CR junto al profesor Eduard, ahora hubo dos puntualmente muy significativos, enumerados a continuación:

1. **Activar y desactivar FOREIGN KEYS:** Durante el avance 2 tuve un inconveniente pues se pedía eliminar una categoría de producto, el cual tenía una clave foránea relacionada, durante el desarrollo autónomo no sabia ni tenia idea de como se hacía, tuvimos que preguntar en el CR como se realizaba y solucionar el desafío que determinaba esta consigna, ahora en el mundo real no es considerado como una buena práctica, ya que esta acción podría dañar la arquitectura de la base de datos.
2. **Sintaxis en SQL:** para mi fue y es un desafío aún sin solucionar, el aprendizaje de la sintaxis de los diferentes consultas, procedimientos almacenados, vistas, etc, son un reto fantástico para mi vida de data analytics junior, pero que se que con la práctica y disciplina cada dia se ira mejorando.

# **REFLEXION PERSONAL**

Cuando inicie mi reto en hacerme DATA ANALYST me intrigaba cómo poder hacer consultas y análisis de datos en base de datos dentro grandes empresas, hoy después de este desafío empiezo a ver como puedo yo como analista de datos y además como ingeniero establecer mis recomendaciones estratégicas poder hacer crecer organizaciones a nivel global. Durante este módulo y este proyecto me encontré con temas que realmente me dejaron ver que aun soy principiante en este mundo de los datos pero, que con mi disciplina y mi habilidad de aprender podré sacarlo adelante.

Solo queda por decir que estoy listo para un próximo nivel. Mil gracias

# **EXTRA CREDITS:**

* **Recomiendo:** Implementar índices clustered en la tabla de órdenes, nuestra tabla principal de transacciones, estos índices mejorarán significativamente la eficiencia de las consultas relacionadas con las órdenes, se optimizará el rendimiento de las consultas, garantizara un acceso más rápido en las consultas, operaciones y filtrados, en conclusión la implementación de índices clustered mejorará la velocidad y la eficiencia en el procesamiento de transacciones en nuestra base de datos.