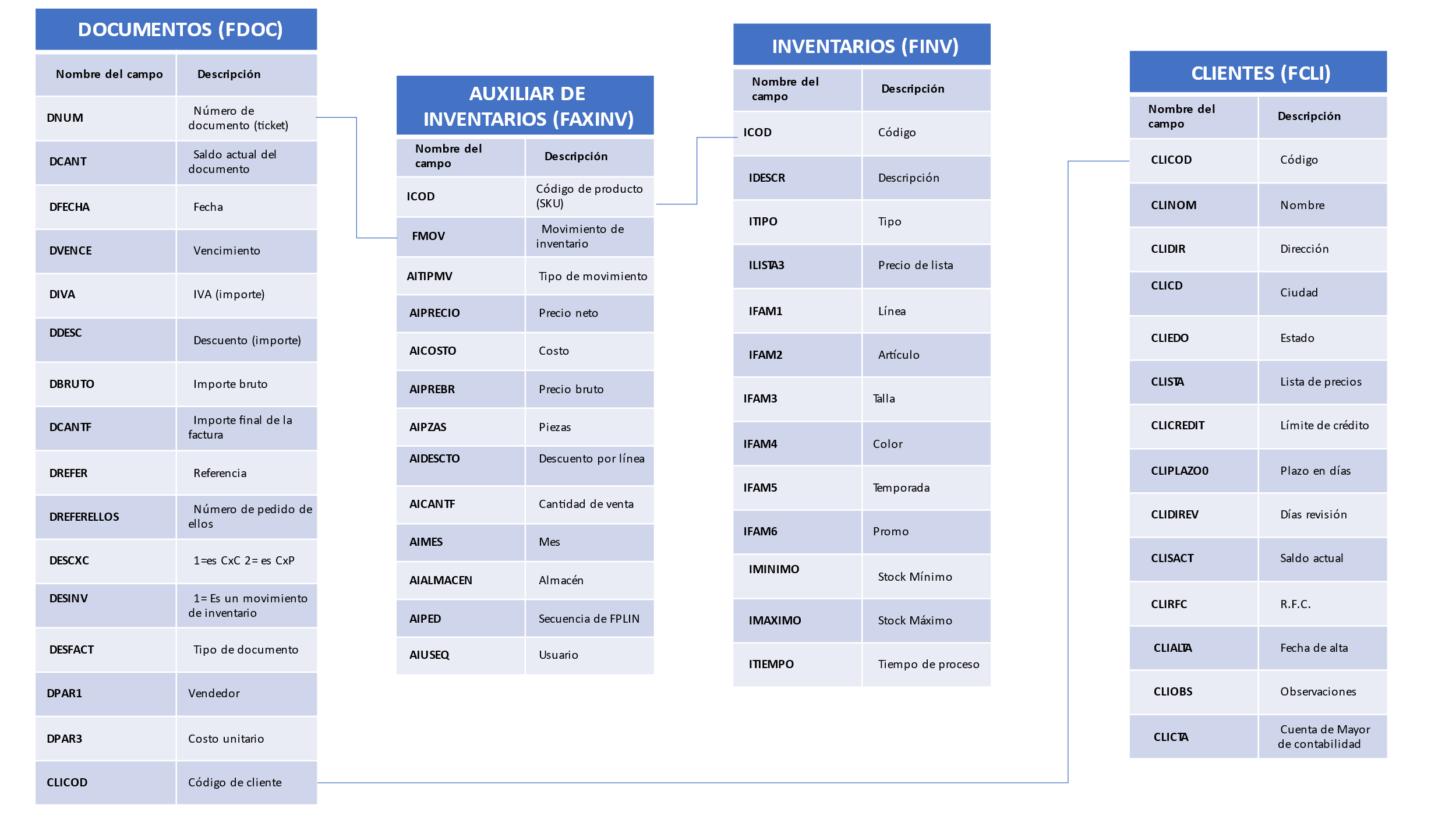
1. **Construcción de query en SQL**

**Instrucciones:** Con el diagrama de base de datos compartido, crea un query para obtener la información de ventas del cliente, incluyendo los siguientes campos:

* # de ticket
* Fecha
* Monto (DCANT) con y sin IVA
* Vendedor
* Cliente
* Productos comprados (SKU)
* Descripción
* Cantidad (unidades)
* Almacén que vendió
* Precio de lista
* Talla
* Color
* Temporada



**Entregable:** Código escrito con la consulta solicitada. Puedes asumir que las tablas ya están creadas en la base de datos.

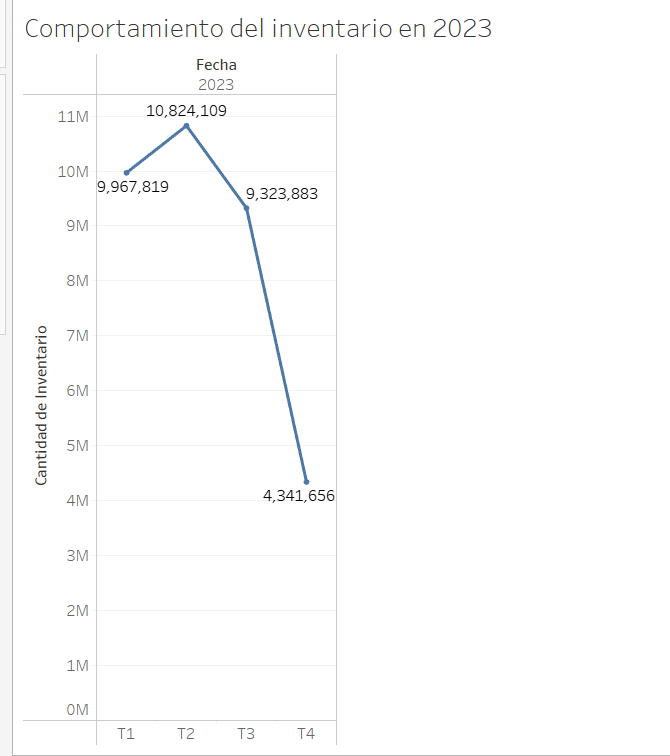
1. **Desarrollo de visualizaciones en Business Intelligence**

**Instrucciones:** Con el archivo “Fuente para dashboard”, desarrolla un dashboard que permita responder la mayor cantidad de preguntas de la siguiente lista, para dar una recomendación al cliente acerca de su operación.

**Preguntas:**

1. ¿Qué materiales me voy a quedar sin inventario y en qué fecha me quedo sin inventario?

La muestra total nos indica que durante todo el año hay abasto de la mayoría de los materiales hay que destacar que se observa una tendencia para el cuarto trimestre de exhaustación del inventario lo que es lógico ya que lo esperado es acabar con el inventario inicial presupuestado, lo cual sería lo más optimo en cuestión de ventas



El grafico 1. muestra la tendencia para el año 2023

Tabla 1.Materiales que han escaseado a lo largo del histórico y sus respectivas fechas son los siguientes

|  |  |
| --- | --- |
| Materiales | Trimestre |
| 063311ZM | T4 |
| 464256 | T4 |
| 534754 | T4 |
| 464255 | T4 |
| 534753 | T4 |

1. ¿Cuáles productos son los más importantes en venta?

Tabla 2. Los 8 productos con mayor importancia son los siguientes:

|  |  |
| --- | --- |
| Producto | Cantidad promedia por año |
| 063082 | 2,218,725.00 |
| 534753 | 1,225,763 |
| 534754 | 1,153,450 |
| 63080 | 922,150 |
| 146810ZM | 905,267 |
| 464255 | 823,586 |
| 063111ZM | 700,920 |
| 464256 | 691,481 |

1. ¿Cuáles productos tengo riesgo de quedarme con mucho inventario al final del año?

534753,534754, a pesar de ser de los productos mas vendidos se observa una cantidad considerable al termino de cada año en el 4T (cuarto trimetre) en realidad la mayoría de los productos con mas volumen de venta tienen probabilidad de que un considerable stock quede después de cerrar el ciclo de año lo cual podría significar ahorro ya que por la tendencia es una certeza que se continuaran vendiendo por su parte 146810ZM se vende poco y tiene un gran cantidad de inventario habría que priorizar si en realidad ese material es benéfico o perdida, Ya que a lo largo del año se adquiere más de este producto pero no se vende de manera prolifera.

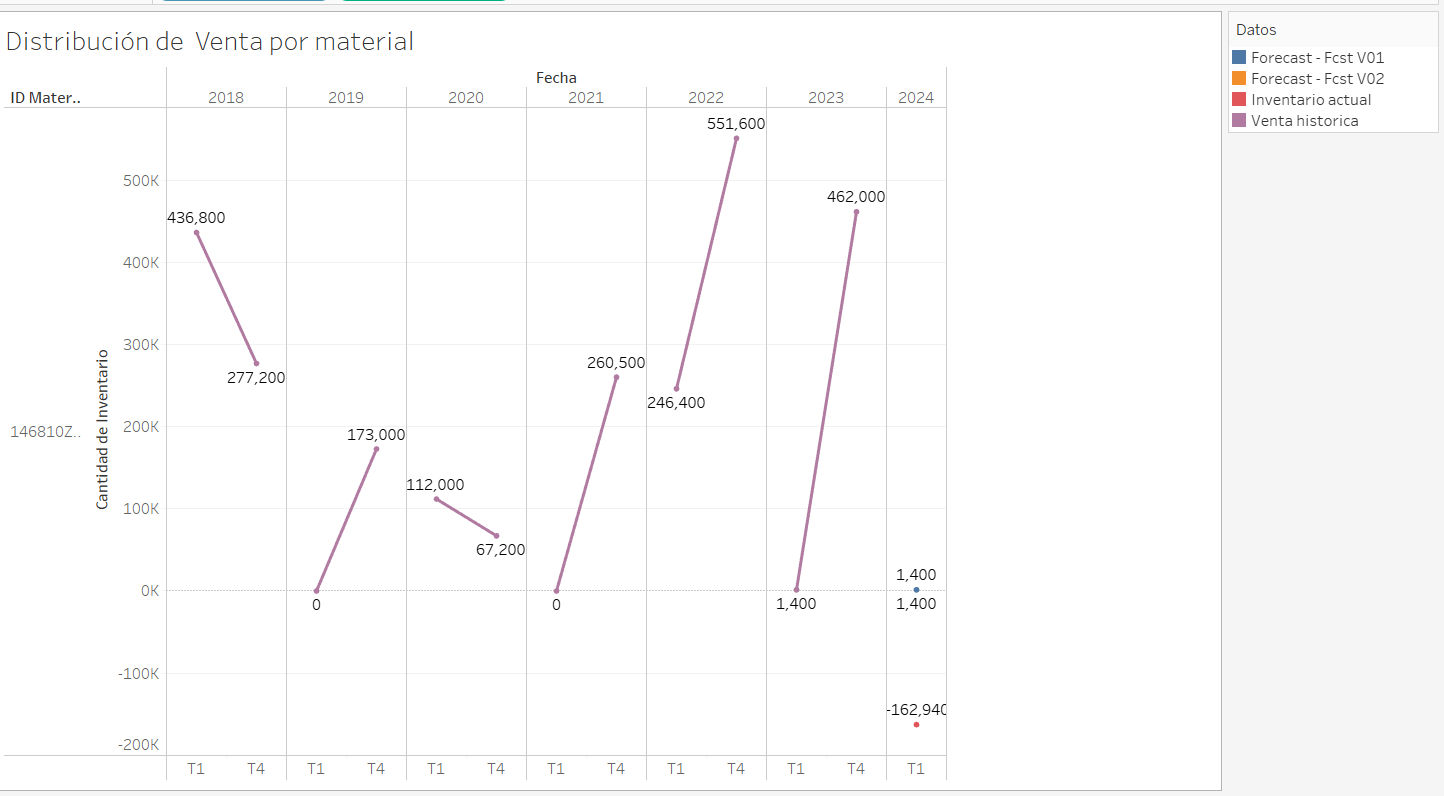


Grafico 2. Comportamiento del material 146810ZM

1. ¿Hay mucha variación entre la venta de años pasados y el forecast? En general la diferencia del modelo de forecast es un tanto inexacta respecto al primer trimestre(T1) y el Tercer trimestre (T3) esto se repite desde 2018 al 2023 por motivos de comodidad agregue una visualización donde se puede observar solamente el año 2023, se debería ajustar esa parte del modelo pero como una aproximación o proyección inicial debería funcionar para establecer parámetros y objetivos de venta

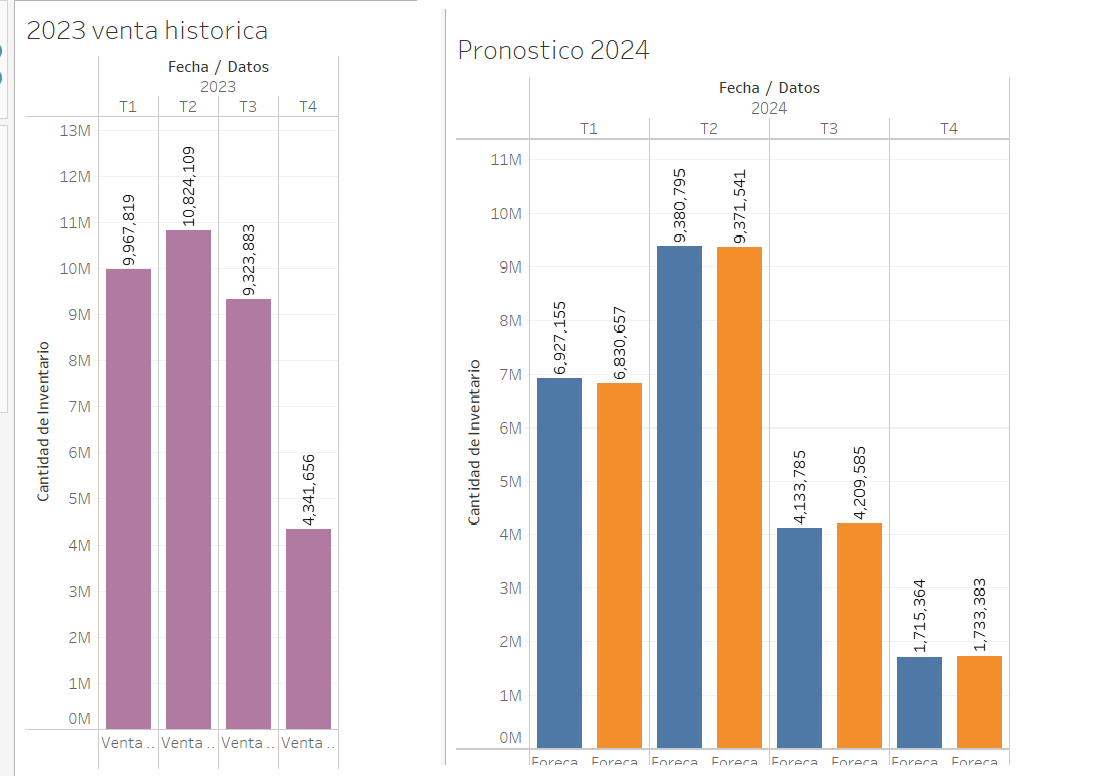


Gráfico 3. Comparación de modelo V1 y v2 de Forecast contra año 2023

1. ¿Cuál sería la cobertura del inventario contra la versión 1 del forecast y con la versión 2 del paraforecast? En el grafico 2 podemos observar que ambos modelos se comportan muy similar respecto al desglose del inventario en función del tiempo

Tabla 3. Comparación de Forecast V1 y v2

|  |  |  |  |
| --- | --- | --- | --- |
| Trimestre | V1 forecast | V2 forecast | Diferencia numérica entre los modelos |
| T1 | 6927155 | 6830657 | |  | | --- | | 96498 | |
| T2 | 9380795 | 9371541 | 9254 |
| T3 | 4133785 | 4209585 | 75800 |
| T4 | 1715364 | 1733383 | |  | | --- | | 18019 | |

1. Después de haber presentado esta primera versión del dashboard, ¿qué pasos recomendarías seguir para liberarlo para el uso en el día a día por parte del cliente?

Darle seguimiento con un plan estratégico para cubrir las necesidades e inquietudes del cliente cubriendo aspectos como la validación y prueba, el feedback del usuario es importante, se debe continuar ajustando el dashboard con las demandas del cliente y también proponiendo nuevos enfoques de negocio para ayudar y optimizar los recursos del mismo , crear una documentación de navegación y fácil uso para la capacitación del personal que tenga acceso al dashboard, una constante automatización y actualización para ayudar a la satisfacción del cliente

1. ¿Cómo esperarías que fuera la interacción entre IBSO y el cliente por los siguientes 12 meses?

Durante los próximos 12 meses, la interacción entre IBSO y el cliente debería ser continua, colaborativa y enfocada en alcanzar los objetivos de negocio mediante la implementación efectiva de soluciones de analítica y business intelligence. Desde el inicio, se establecerán reuniones de inicio para definir claramente los objetivos del proyecto, roles, responsabilidades y un cronograma detallado. La comunicación será regular, con reuniones semanales o quincenales para revisar el progreso y discutir cualquier problema, complementadas con informes mensuales detallados y revisiones trimestrales para evaluar el desempeño y el impacto en los indicadores clave de rendimiento (KPIs). IBSO proporcionará talleres y capacitaciones mensuales para asegurar que el equipo del cliente esté capacitado en el uso de las herramientas y tecnologías implementadas. Utilizando metodologías ágiles, se desarrollarán y entregarán soluciones en sprints, permitiendo iteraciones rápidas y mejoras continuas. Además, IBSO ofrecerá soporte técnico continuo y mantenimiento, asegurando que las herramientas y sistemas se mantengan actualizados y funcionales. El feedback regular del cliente será recopilado y analizado para adaptar las estrategias y soluciones según sea necesario, garantizando que se cumplan las expectativas y necesidades cambiantes del cliente. Hacia el final del año, se llevará a cabo una planificación estratégica para definir los objetivos del próximo año, informando al cliente sobre nuevas tendencias y oportunidades en el ámbito de la analítica y BI. Este enfoque integral asegurará una relación productiva y de alto valor, alineada con los objetivos del cliente y adaptable a sus necesidades dinámicas**.**

**Entregable:** Dashboard (En Tableau / Power BI / Qlik) con las tablas, gráficas y/o visualizaciones que justifiquen los resultados y recomendaciones propuestas. El dashboard debería seguir una estructura lógica y deberá permitir al usuario entenderlo sin mucha capacitación.

**Referencias para estructura del dashboard:**

<https://medium.com/@mokkup/8-essential-dashboard-design-principles-for-effective-data-visualization-40653c5fd135>

Tableau: <https://data.ucop.edu/support-training/tableau-files/building_effective_dashboards.pdf>

1. **Ejercicios – Python**

**Instrucciones:** Con las preguntas mostradas, trata de desarrollar los códigos para responder el mayor número de ejercicios usando Python.

**Entregable:** Códigos en Python mostrando los resultados de los ejercicios.

**Ejercicio 1:**

1. Crea un diccionario llamado datos\_climaticos donde cada clave es el nombre de una ciudad y el valor es una lista de temperaturas (en grados Celsius) registradas a lo largo de una semana.
2. Para cada ciudad, calcula la temperatura promedio, la temperatura máxima y la mínima de la semana.
3. Determina cuál fue la ciudad con la temperatura promedio más alta y la más baja durante la semana.

**Ejercicio 2:**

Asigna a cada letra minúscula un valor, desde 1 para la 'a' hasta 26 para la 'z'.

Crea una función que pida al usuario una cadena de letras minúsculas y responde la suma de los valores de las letras en la cadena. (Ejemplo: hola = 8 + 15 + 12 + 1 = 36).

Si el usuario te da un número o una letra mayúscula, pídele que lo cambie (Input: Hola. Output: Cambia a minúscula la letra “H” en la posición 1. Input: int2. Output: Cambia el número en la posición 4 por una letra minúscula).

**Ejercicio 3:**

1. Extraer la información del csv **Prueba\_Promociones**

2. ⁠Generar un código donde el usuario pueda ingresar las siguientes variables

- fecha inicio (convertir a datetime)

- ⁠fecha fin (convertir a datetime)

- ⁠categoría (validar que sea string)

- ⁠uso (validar que sea string)

- ⁠sku (no permitir al usuario avanzar si no ingreso un valor string en el campo de SKU)

- ⁠% (validar que sea número decimal)

- ⁠inventario inicial (validar que sea entero)

3. Generar una nueva columna donde se coloque el # de semana correspondiente de la fecha

4. ⁠Con él % ingresado por el usuario, impactar ese porcentaje como crecimiento (columna de piezas) para todos los datos que cumplan con las siguientes condiciones:

- estén dentro del rango de fecha seleccionado. Si el usuario no coloca fecha fin entonces desde fecha inicio hasta el final, si el usuario no coloca fecha inicio entonces desde la fecha fin hasta el principio y si no coloca ninguna, a todas las fechas disponibles

- ⁠la columna **modelo** sea diferente a “real”

- ⁠la columna **uso** sea igual al valor ingresado por el usuario, en caso de que esté vacío ese campo, la columna categoría sea igual al valor ingresado por el usuario. En caso de que ambas vengan vacías aplicar a todo

5. Una vez aplicado el incremento %, generar un nuevo dataframe en donde solo venga información del SKU seleccionado y se pueda visualizar el consumo de inventario, es decir, el inventario inicial colocado por el usuario será la cantidad inicial el 01-01-2024 y de ahí se tendrá que ir consumiendo (restando) la venta proyectada (**modelo** diferente de real). Se deberá imprimir la primera fecha en la que se vuelve negativo el inventario

**Ejercicio 4: Recomendador de libros**

Fuente: <https://developer.nytimes.com/docs/books-product/1/overview>

**Elementos a utilizar:**

1. Crea una cuenta de Developer en la página de New York Times para tener un API key (<https://developer.nytimes.com/>)
2. Usa el “Books API”.

**¿Qué debe poder hacer tu código?**

1. Pedirle al usuario decidir qué lista de “Best Sellers” quiere consultar.
2. Poder escoger si quiere ver los “Best Sellers” actuales o de alguna fecha en específica.
3. Poder escoger un precio específico del libro que quiere adquirir.
4. Poder escoger un rango de edades dirigido para el libro.

**Resultado:**

1. Poder verlo de manera estructurada, poniendo la información clave del libro que estás recomendado para la información que te dio el usuario.
2. Poder acceder rápidamente a la reseña generada por el NYT acerca del libro recomendado.
3. Para los Best Sellers actuales, decirle al usuario dónde lo puede comprar.
4. Poder mostrar toda esta información en una aplicación / pantalla / interfaz gráfica amigable para que el usuario pueda interactuar con los resultados.
5. **Examen técnico**

Explica con el mayor detalle las respuestas a estas preguntas (45 minutos).

1. La información del ERP del cliente vive en un data center controlado por un proveedor, con reglas que no permiten adicionar campos, columnas, construir tablas o extraer información. ¿Qué pasos harías para pasar de la información en el ERP, cruzar y transformarla, y mandarla a una herramienta de Business Intelligence con actualizaciones automáticas?

Lo que haría para abordar la extracción y uso de datos de un ERP ubicado en un data center con restricciones estrictas, primero se deben evaluar los métodos permitidos de acceso, como APIs, informes programados, o exportaciones periódicas de datos. Luego, estos datos deben ser almacenados temporalmente en un data warehouse intermedio, utilizando herramientas de ETL (como Talend, Apache NiFi, o servicios en la nube como AWS Glue) para automatizar la carga y transformación de datos. Una vez transformados y normalizados, los datos se integran con una herramienta de BI (como Tableau, Power BI, Looker, o Qlik Sense), configurando dashboards y reportes que se actualicen automáticamente según las actualizaciones de datos en el data warehouse. Es crucial implementar monitoreo continuo y revisión periódica para asegurar la precisión y relevancia de los datos y reportes.

1. El cliente de la pregunta 1 está evaluando adquirir un servidor físico o en la nube para poder alojar ahí un espejo de la información de su ERP. ¿Qué recomendarías en cuanto a costo o inversión y el rendimiento esperado dadas las características de cada una de las opciones?

Yo pienso que para ayudar a tomar la decisión entre un servidor físico y un servidor en la nube para alojar un espejo de la información del ERP, es crucial considerar diversos factores como costo de inversión, mantenimiento, rendimiento, escalabilidad y flexibilidad. Un servidor físico implica una alta inversión inicial en hardware y software, junto con costos continuos de mantenimiento, personal técnico, electricidad, refrigeración y actualizaciones de hardware y software. Sin embargo, ofrece control total sobre la infraestructura, lo cual puede optimizar el rendimiento y minimizar la latencia si el servidor se encuentra cerca del cliente. No obstante, la escalabilidad es limitada y costosa, ya que requiere la compra de más hardware para aumentar la capacidad. Por otro lado, un servidor en la nube presenta un costo inicial bajo, con un modelo de pago basado en demanda o suscripción, y la mayoría de los costos de mantenimiento y actualizaciones son gestionados por el proveedor de servicios de nube como AWS, Azure o Google Cloud. La nube ofrece alta flexibilidad y escalabilidad, permitiendo ajustar recursos en tiempo real según las necesidades, aunque la latencia podría ser mayor dependiendo de la ubicación del servidor y la distancia de los datos. Además, los proveedores de nube suelen proporcionar robustas medidas de seguridad, aunque también depende de la correcta configuración y gestión por parte del cliente. Dado que un servidor en la nube permite una implementación rápida y eficiente, es recomendable debido a sus menores costos iniciales y de mantenimiento, junto con su capacidad para escalar y ajustarse a las necesidades cambiantes del negocio. En resumen, aunque un servidor físico ofrece un mayor control y potencialmente menor latencia, las ventajas significativas en términos de costo y flexibilidad de una solución en la nube generalmente superan estas consideraciones, haciendo que la nube sea la opción más adecuada para la mayoría de los escenarios.

1. Un cliente solamente cuenta con un servidor de producción donde viven los datos, ¿qué sugerirías para no afectar el desempeño de esta al conectar BI?, ¿Cuáles serían las implicaciones (tiempo, conocimiento, administración, costo) de esta sugerencia?

Nuestro cliente podria evitar afectar el rendimiento del servidor de producción al conectar una herramienta de Business Intelligence (BI), recomendaría implementar una solución de réplica o sincronización de datos que permita realizar consultas y análisis en un entorno aparte. Esto que implica una configuraración en una base de datos de réplica, utilizando herramientas de replicación en tiempo real o casi en tiempo real como MySQL Replication, Oracle Data Guard, SQL Server Replication, o servicios en la nube como AWS DMS o Google Cloud Dataflow. Adicionalmente, se debe implementar un proceso de ETL (Extract, Transform, Load) para extraer datos de la base de datos de producción, transformarlos y cargarlos en el data warehouse o base de datos de réplica. Las implicaciones de esta solución incluyen varios aspectos. En términos de tiempo, la configuración inicial puede tardar desde unos días hasta varias semanas, y el establecimiento de una sincronización precisa requerirá tiempo adicional para pruebas y ajustes. En cuanto al conocimiento, se necesitará expertise técnico en bases de datos y herramientas de replicación, así como experiencia en gestión de bases de datos y administración de servidores, y posiblemente conocimientos en servicios de nube. La administración implicará mantenimiento continuo para asegurar la correcta sincronización y precisión de los datos, además de la monitorización y optimización para detectar problemas de rendimiento y errores en la replicación. Los costos incluirán la infraestructura adicional, ya sea en servidores físicos o en la nube, licencias de software o suscripciones para herramientas de replicación y ETL, y gastos asociados a la contratación o formación de personal técnico especializado. A pesar de estas implicaciones, los beneficios de mantener el servidor de producción libre de cargas adicionales, asegurando su disponibilidad y rendimiento, hacen que la inversión en una base de datos de réplica o un data warehouse sea altamente justificable.

1. Un query que escribiste regresa 10,000 filas únicamente, pero su ejecución toma entre 3 y 4 horas. Por lo mismo, muchas veces no llegan los resultados completos a BI o se pierde la conexión mientras esto ocurre. ¿Qué harías?

Pienso que la mejor manera de resolver se debe seguir una estrategia de optimización. Primero, es crucial revisar y analizar el query en detalle, buscando ineficiencias como joins innecesarios, subconsultas redundantes o filtros ineficaces, y asegurarse de que las columnas utilizadas en las cláusulas WHERE, JOIN y ORDER BY estén debidamente indexadas. Utilizar la función EXPLAIN para obtener el plan de ejecución del query puede ayudar a identificar cuellos de botella específicos. La simplificación del query, eliminando subconsultas innecesarias y optimizando las cláusulas JOIN, puede reducir significativamente el tiempo de ejecución. Además, es fundamental revisar la configuración del servidor de la base de datos para asegurarse de que esté optimizada para el tipo de consultas que se ejecutan, incluyendo ajustes en parámetros de memoria y tamaño de caché. El particionamiento de tablas grandes puede mejorar el rendimiento al segmentar los datos de manera lógica, y la creación de vistas materializadas puede ser beneficiosa si los datos subyacentes no cambian frecuentemente, permitiendo el almacenamiento de resultados precomputados. Para manejar grandes volúmenes de datos, se puede considerar dividir la consulta en varias más pequeñas que se ejecuten en paralelo o secuencialmente, acumulando los resultados, y utilizar cargas incrementales de datos para procesar solo los nuevos o modificados desde la última carga en lugar de ejecutar consultas sobre todo el conjunto de datos. Mejorar la infraestructura del servidor, aumentando la capacidad de procesamiento, memoria o almacenamiento, también puede contribuir a reducir el tiempo de ejecución. Finalmente, es importante optimizar la conectividad entre la base de datos y la herramienta de BI, asegurando una red rápida y fiable para minimizar interrupciones. Implementar estas medidas de optimización debería mejorar significativamente el tiempo de ejecución del query y la fiabilidad de la entrega de datos a la herramienta de BI.

1. Te busca un cliente para decirte que los resultados que ve en los tableros no le hacen sentido (ayer si mostraba resultados correctos). ¿Qué pasos harías para darle una respuesta al cliente?

La mejor manera de atender al cliente acerca de la falta de coherencia en los resultados de los tableros de BI, primero me sumergiría en una investigación exhaustiva para identificar la causa subyacente del problema. Esto implicaría entrevistar al cliente para comprender completamente la naturaleza de las discrepancias y cualquier cambio reciente en los datos o configuraciones. Luego, verificaría los datos en la fuente para asegurarme de que estén intactos y no hayan sido afectados por errores de procesamiento. Paralelamente, revisaría los registros de los procesos ETL para detectar posibles fallos o inconsistencias durante la carga y transformación de los datos. Posteriormente, inspeccionaría detenidamente la configuración de los tableros de BI para identificar cualquier ajuste inadvertido que pudiera haber alterado los resultados. Una vez identificada la causa raíz, procedería a implementar las correcciones necesarias, ya sea rectificando los datos en la fuente, ajustando los procesos ETL o corrigiendo la configuración del tablero. Finalmente, mantendría una comunicación clara y continua con el cliente, informándole sobre las acciones tomadas y proporcionando cualquier seguimiento necesario para garantizar que el problema se resuelva de manera satisfactoria y se eviten recurrencias futuras.

1. ¿Qué mejores prácticas sugerirías para garantizar la seguridad de la información del cliente, ya sea en un servidor físico o en la nube?

Existen hoy en día ciertos requerimientos para garantizar la seguridad de la información del cliente, tanto en servidores físicos como en la nube, es esencial adoptar una serie de mejores prácticas que aborden múltiples dimensiones de la seguridad. Primero, se debe implementar el cifrado de datos en reposo y en tránsito, utilizando certificados SSL/TLS para proteger las comunicaciones web. La gestión de acceso debe reforzarse mediante la autenticación multifactor (MFA) asegurando además el uso de contraseñas fuertes. En el entorno de la nube, es crucial seleccionar proveedores confiables y configurar adecuadamente los servicios, realizando auditorías regulares y asegurando la seguridad de los contenedores. El monitoreo continuo y la auditoría de actividades permiten detectar y responder rápidamente a incidentes de seguridad. La protección contra malware y amenazas debe incluir el uso de software anti-malware actualizado y la implementación de firewalls y sistemas de detección y prevención de intrusiones (IDS/IPS). Asimismo, se deben realizar copias de seguridad periódicas y desarrollar planes de recuperación ante desastres, probándolos regularmente. La seguridad física es igualmente importante, restringiendo el acceso a las instalaciones y utilizando vigilancia. Mantener todos los sistemas y software actualizados con los últimos parches de seguridad es fundamental para prevenir vulnerabilidades. Finalmente, la concienciación y capacitación continua del personal sobre prácticas seguras y la realización de simulacros de phishing son cruciales, así como asegurar el cumplimiento con normativas y estándares de seguridad relevantes, realizando evaluaciones periódicas de cumplimiento para garantizar la alineación con las regulaciones aplicables. Al integrar estas prácticas, se puede crear un entorno seguro y resiliente para la información del cliente.

1. El director comercial de una empresa nos pide un proyecto para poder visualizar información en un tablero de BI con actualizaciones cada 15 minutos con información proveniente de distintas fuentes:
   1. ERP (SQL Server)
   2. CRM (API)
   3. Catálogos de información (Excel)
   4. Google Analytics
   5. Google Ads
   6. Shopify
   7. Facebook Ads

Prepara una propuesta técnica que incluya tiempos, costos y herramientas a utilizar para entregarle al director comercial este tablero. En caso de que haya información que necesites para hacer esta propuesta y no la tengas, genera supuestos para los que sería valida la propuesta.

Objetivo:

Desarrollar un tablero de Business Intelligence (BI) que permita visualizar información actualizada cada 15 minutos, proveniente de distintas fuentes como el ERP, CRM, catálogos de información, Google Analytics, Google Ads, Shopify, Facebook Ads, entre otros, para brindar al director comercial una visión integral y en tiempo real del desempeño de la empresa.

Tiempos y Costos:

1. Análisis de Requerimientos (1 semana):

- Reuniones con el director comercial para comprender los requisitos y objetivos del tablero.

- Evaluación de las fuentes de datos y sus formatos.

- Definición de KPIs clave a visualizar.

2. Diseño de Arquitectura y Planificación (2 semanas):

- Diseño de la arquitectura del sistema, incluyendo la integración de las fuentes de datos.

- Planificación de los procesos de extracción, transformación y carga (ETL).

- Selección de herramientas y tecnologías adecuadas para cada fuente de datos.

3. Desarrollo e Implementación (4 semanas):

- Desarrollo de los scripts ETL para extraer datos de las fuentes y cargarlos en un data warehouse o base de datos centralizada.

- Implementación de la lógica de negocio y cálculos de KPIs en el data warehouse.

- Desarrollo del tablero de BI utilizando herramientas como Tableau, Power BI o Qlik Sense.

- Configuración de la actualización automática cada 15 minutos.

4. Pruebas y Ajustes (1 semana):

- Pruebas exhaustivas del tablero y los procesos ETL para garantizar la precisión y consistencia de los datos.

- Ajustes según los comentarios del director comercial y los usuarios finales.

5. Entrenamiento y Entrega (1 semana):

- Capacitación del personal para el uso adecuado del tablero de BI.

- Entrega final del tablero con documentación detallada sobre su funcionamiento y mantenimiento.

Herramientas a Utilizar:

-ETL: Utilizar herramientas como Apache NiFi, Talend o AWS Glue para la extracción, transformación y carga de datos desde las distintas fuentes.

- Data Warehouse/Database: Emplear una base de datos centralizada como PostgreSQL o MySQL para almacenar y procesar los datos de todas las fuentes.

- Tablero de BI: Desarrollar el tablero utilizando herramientas como Tableau, Power BI o Qlik Sense para la visualización y análisis de datos.

- Automatización: Implementar herramientas de automatización como Airflow o cron jobs para programar las actualizaciones automáticas cada 15 minutos.

Supuestos:

- Se cuenta con acceso y permisos adecuados para extraer datos de todas las fuentes mencionadas.

- Las APIs necesarias para la integración con el CRM y otras fuentes externas están documentadas y disponibles.

- Se dispone de recursos técnicos capacitados para el desarrollo e implementación del proyecto.

- El director comercial proporcionará la información necesaria y la retroalimentación requerida durante el proceso de desarrollo.

- Los costos incluyen licencias de software, infraestructura de nube (si es aplicable), y honorarios profesionales para el equipo de desarrollo.