

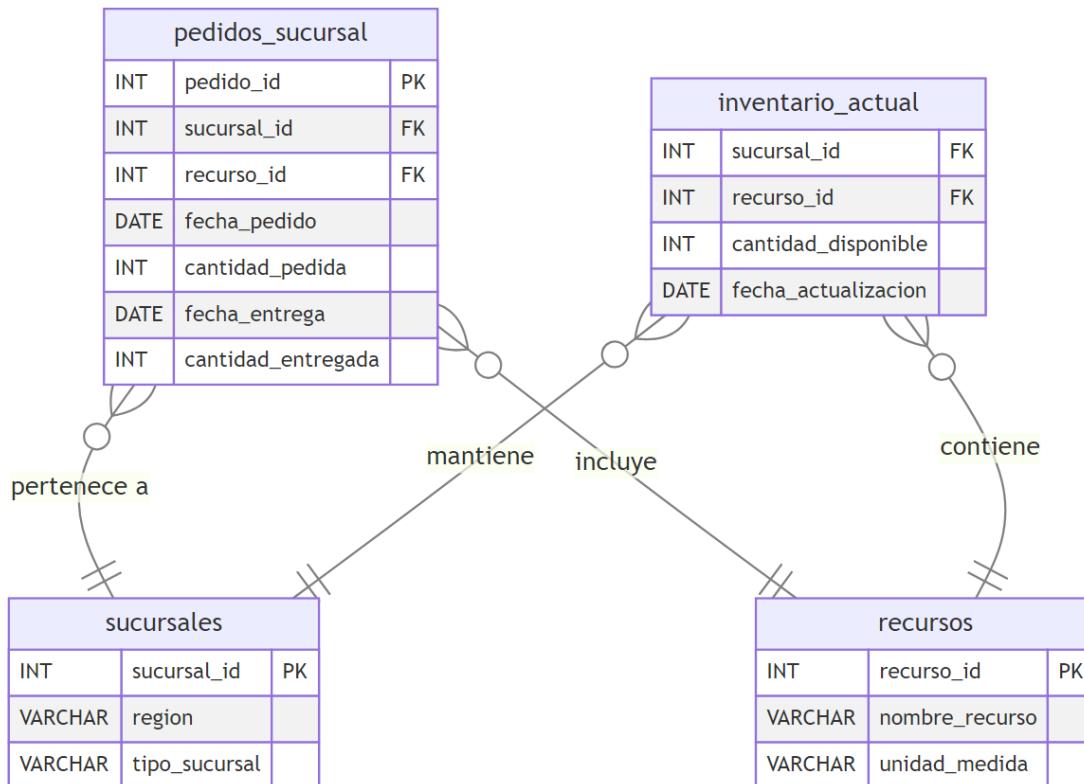
Prueba Técnica para Analítico

Gerencia de Servicios Logísticos

Punto 1

El banco cuenta con una red extensa de sucursales en todo el país, cada una de las cuales realiza mensualmente pedidos de recursos necesarios para sus operaciones diarias (papel higiénico, papel de impresión, café, sellos, desechables, entre otros). La eficiencia en la gestión de estos recursos es clave para reducir costos y asegurar que cada sucursal siempre tenga los materiales necesarios.

A continuación, se presentan las tablas de datos relacionadas:



- **pedidos_sucursal:**
 - o **pedido_id** (INT): ID del pedido.
 - o **sucursal_id** (INT): ID de la sucursal.
 - o **recurso_id** (INT): ID del recurso solicitado.
 - o **fecha_pedido** (DATE): Fecha en que se realizó el pedido.
 - o **cantidad_pedida** (INT): Cantidad de recursos solicitada.

- **fecha_entrega** (DATE): Fecha en que se entregó el pedido.
- **cantidad_entregada** (INT): Cantidad de recursos entregados (puede diferir de la solicitada en casos de falta de stock).
- **recursos:**
 - **recurso_id** (INT): ID del recurso.
 - **nombre_recurso** (VARCHAR): Nombre del recurso (ej. "Papel Higiénico", "Café", "Papel de Impresión").
 - **unidad_medida** (VARCHAR): Unidad de medida del recurso (ej. "Paquetes", "Cajas").
- **sucursales:**
 - **sucursal_id** (INT): ID de la sucursal.
 - **region** (VARCHAR): Región en la que está ubicada la sucursal (ej. "Norte", "Sur", "Centro").
 - **tipo_sucursal** (VARCHAR): Tipo de sucursal (ej. "Pequeña", "Mediana", "Grande").
- **inventario_actual:**
 - **sucursal_id** (INT): ID de la sucursal.
 - **recurso_id** (INT): ID del recurso.
 - **cantidad_disponible** (INT): Cantidad actual del recurso en la sucursal.
 - **fecha_actualizacion** (DATE): Fecha de la última actualización de la cantidad disponible.

Tareas:

1. Escribe una consulta SQL para identificar los cinco recursos más solicitados en el último trimestre por las sucursales del banco.
 - a. ¿Qué recomendaciones harías si observas que ciertos recursos son recurrentemente pedidos en grandes cantidades?
 - b. ¿Qué factores externos o internos podrían explicar esta demanda?
2. Escribe una consulta SQL para encontrar los pedidos en los que la cantidad entregada es menor a la cantidad pedida en el último semestre.
 - a. Si descubres que ciertas sucursales o tipos de recursos están experimentando problemas recurrentes de abastecimiento, ¿qué medidas propondrías para mejorar la eficiencia de suministro?
3. Escribe una consulta SQL para identificar los recursos cuyo inventario en cada sucursal es menor a un umbral de cinco unidades.
 - a. ¿Cómo propondrías gestionar estos casos de inventario bajo para evitar que las sucursales se queden sin recursos?

- b. ¿Crees que establecer una política de reabastecimiento automatizada podría ser beneficiosa?
4. Diseña una propuesta de modelo predictivo para anticipar la cantidad de cada recurso que cada sucursal necesitará en los próximos meses. La propuesta debe ser clara y abarcar los siguientes aspectos:
 - a. ¿Cómo plantearías el problema de predicción? ¿Qué tipo de modelo consideras adecuado para estimar el consumo futuro de recursos?
 - b. ¿Qué variables incluirías en el modelo? ¿Qué factores específicos del negocio podrían influir en la demanda de recursos?
 - c. ¿Cómo sugerirías evaluar el modelo? ¿Qué métricas de rendimiento recomendarías?
 - d. Si el modelo muestra baja precisión en las predicciones para ciertos recursos o sucursales, ¿Cómo abordarías el problema? ¿Qué técnicas de mejora aplicarías?
 - e. ¿Cómo propondrías implementar el modelo en la operación de las sucursales?

Nota: NO es necesario construir código; solo se evaluará la propuesta.

Punto 2

Objetivo: Utilizar un conjunto de datos de libros y reseñas para extraer insights mediante técnicas de procesamiento de lenguaje natural (NLP) y análisis exploratorio de datos, seguido de una visualización de los resultados. El código debe estar estructurado como un producto profesional, modularizado y utilizando principios de buenas prácticas.

Datos: <https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews>

El conjunto de datos contiene dos tablas:

- Book Details: Incluye detalles sobre los libros, como título, descripción, autores, imagen, fecha de publicación, categorías y conteo de valoraciones.
- Reviews: Contiene información sobre las reseñas, como el ID de la reseña, título del libro al que pertenece, precio, usuario, puntaje, tiempo de reseña, resumen y texto de la reseña.

Instrucciones

Estructura del Código:

- Desarrolla el análisis en Python aplicando buenas prácticas de desarrollo.
- El código debe estar modularizado en archivos separados (no en un solo notebook), estructurado de manera que pueda ejecutarse de forma independiente.
- Usa un enfoque orientado a clases y funciones, organizando el código en módulos específicos para diferentes tareas, como carga de datos, procesamiento de texto, análisis de sentimiento, generación de reportes y visualización.
- Asegúrate de que cada módulo y función tenga documentación y docstrings claros.
- Utiliza un entorno virtual para gestionar las dependencias y proporciona un archivo requirements.txt con las bibliotecas necesarias.

Carga y Exploración de Datos:

Crea un módulo específico para cargar y limpiar los datos. Este módulo debe verificar la existencia de valores nulos y anomalías en los datos.

Proporciona una función de inicialización para cargar y limpiar los datos, de modo que sea reutilizable.

Análisis Exploratorio:

En un módulo dedicado al análisis, incluye funciones para:

- Calcular las valoraciones promedio por libro.
- Determinar el número total de reseñas y el número total de valoraciones.
- Identificar los autores más populares y los géneros/categorías más reseñadas.

Análisis de Sentimiento (NLP):

Crea un módulo para el análisis de sentimientos. Dentro de este módulo:

- Incluye funciones para preprocesar el texto de las reseñas (limpieza, tokenización, etc.).
- Implementa una función que aplique un modelo de análisis de sentimiento preentrenado (puede ser TextBlob, VADER o NLTK).
- Asegúrate de que el análisis de sentimientos sea reproducible y fácil de actualizar si se decide cambiar el modelo en el futuro.

- Genera funciones para calcular el sentimiento promedio por libro y por categoría.

Identificación de Mejores Libros:

Crea un módulo para identificar los mejores libros, con funciones que generen listas de los 10 libros mejor valorados en función de:

- Número de reseñas.
- Promedio de puntaje.
- Sentimiento promedio de las reseñas.

Automatización y Ejecución:

Desarrolla un script de ejecución principal (main.py) que combine todos los módulos y ejecute el análisis completo de manera automática.

Incluir opciones de configuración (usando un archivo .env o config.json) para manejar parámetros clave, como la ubicación del conjunto de datos y opciones de análisis.

Entrega del Proyecto:

Subir todo el código en un repositorio de Git estructurado profesionalmente. El repositorio debe incluir:

- Un archivo README.md explicando el propósito del proyecto, instrucciones para ejecutar el análisis, y una guía para configurar y desplegar el entorno.
- El archivo requirements.txt para la gestión de dependencias.
- Documentación de cada módulo y un resumen de los hallazgos principales en un archivo de reporte (report.md o report.pdf).

Criterios de Evaluación

Calidad del Código:

- Modularización, claridad en el diseño.
- Documentación y cumplimiento de buenas prácticas de Python.

Calidad del Análisis:

- Nivel de detalle y profundidad del análisis exploratorio y de sentimientos.
- Identificación y análisis de tendencias relevantes y patrones clave.

Interpretación de Hallazgos:

- Capacidad para comunicar de manera clara los hallazgos clave y tendencias.

Visualización y Presentación:

- Claridad y relevancia de las visualizaciones.

Automatización y Flexibilidad del Proyecto:

- Facilidad para ejecutar el análisis de manera programada.
- Capacidad del proyecto para ser reutilizable y extensible.

Pública