

Trabajo integrador individual Python + Pandas

Proyecto integrador

Situación inicial

El equipo de Ingeniería de Datos de la Entidad, tiene como misión construir y mantener la plataforma de datos, que será utilizada para la toma de decisiones estratégicas basadas en análisis de datos e interpretación. Esta plataforma permite que la empresa examine y organice sus datos con el fin de atender mejor a sus clientes y consumidores.

Al utilizar datos para impulsar acciones, la Entidad puede contextualizar y personalizar sus mensajes a sus clientes y posibles clientes, generando públicos objetivo adecuados para sus distintas campañas.

El área de Marketing solicitó nuestra ayuda para entender qué públicos objetivo, podrían ser contactados durante el transcurso del presente Q.

Como miembro del equipo y aplicando los conocimientos adquiridos en la unidad de SQL y Python, te enfocarás en las siguientes tareas detalladas para cumplir con los requisitos establecidos:

Nuestro objetivo

Nuestra tarea consiste en crear y ejecutar sentencias de SQL, para generar a futuro los públicos objetivo requeridos.

Los datos de entrada son archivos csv disponibles en el siguiente [repositorio](#).

Utilizaremos las librerías **pandas** y **sqlite3** en un entorno python para poder trabajar con la práctica integral.

Se deberán generar los archivos que den cumplimiento a lo requerido. Todo el código generado será resguardado, utilizando control de versiones con GIT.

Metodología de trabajo

A lo largo de las próximas 2 semanas deberán trabajar individualmente, en el desarrollo de las sentencias de SQL, asegurando la calidad del código producido.

Requerimientos

- Utilizar la herramienta para acceder a una base de datos SQLite en línea.
- Analizar los archivos proporcionados para comprender su estructura y los datos que contienen:
 - ecommerce_customers_dataset.csv
 - ecommerce_orders_dataset.csv
 - ecommerce_order_items_dataset.csv
 - ecommerce_products_dataset.csv
 - ecommerce_order_payments_dataset.csv
- Realizar las siguientes actividades con pandas:
 - Cargar los archivos CSV en DataFrames de Pandas
 - Establecer la columna índices de las tablas como la clave primaria en el DataFrame, ejemplo para customers la clave será customer_id.
 - Obtener el número total de clientes únicos en el conjunto de datos
 - Calcular el promedio de valor de pago por pedido
 - Determinar la categoría de producto más vendida
 - Calcular el número total de pedidos realizados
- Realizar las siguientes actividades con SQLite
 - Crear una base de datos de SQLite llamada ecommerce.db.
 - Guardar los DataFrames en la base de datos como tablas, incluyendo la clave primaria.
 - Ejecutar una consulta **PRAGMA** para obtener la estructura de las tablas y guardar la info a un archivo con extensión .info.
 - Obtener los nombres de los índices de las tablas y guardar la info a un archivo con extensión .info.
 - Usando consultas **SELECT** que involucren joins y agrupaciones para obtener información valiosa respondiendo a las siguientes necesidades:
 - Analizar el comportamiento de compra de los clientes:
 - Utilizar los datos de los customers, orders y order_items.
 - Calcular el valor promedio de los pedidos realizados por cada cliente.
 - Identificar a los clientes con los pedidos de mayor valor.
 - Calcular la frecuencia promedio de compra por cliente, es decir, la cantidad de pedidos realizados en un período de tiempo determinado.

- Generar un informe que muestre los clientes más frecuentes y su promedio de frecuencia de compra.
 - Realizar un análisis de popularidad de productos por categoría:
 - Utilizar datos de los archivos products y order_items.
 - Agrupar los productos por categoría y calcula la cantidad total de unidades vendidas para cada categoría.
 - Generar un informe que muestre las categorías de productos más populares en términos de ventas.
 - Realizar un análisis de la satisfacción del cliente basado en la velocidad de entrega:
 - Utiliza datos de los archivos customers y orders .
 - Calcula el tiempo promedio de entrega de los pedidos por cliente.
 - Genera un informe que muestre los clientes con los tiempos de entrega más rápidos y más lentos.
 - Analizar el rendimiento de los métodos de pago utilizados por los clientes:
 - Utilizar datos de los archivos customers y order_payments.
 - Agrupar los pagos por método de pago y calcula el total de pagos realizados con cada método.
 - Generar un informe que muestre los métodos de pago más utilizados por los clientes.
-
- Exportar los resultados de las consultas a archivos CSV para su posterior análisis y presentación.

Entregables

Archivos fuentes.

Se debe generar y compartir un repositorio en Github/Gitlab con el código fuente de la solución.

Archivos de salida.

Para cada análisis, genera un informe con la siguiente estructura:

- Título del análisis y su objetivo.
- Descripción de los datos utilizados y las tablas involucradas.

BBVA *by* M

