Challenge para puesto Data Engieneer

Franco Balich

1. Descargar la data

Entregables

El código para descargar los datos lo pueden encontrar en src\mercadolibre\api.py del proyecto de git, al mismo lo pueden probar al ejecutar el proyecto y data\origin\productos.jsonl

Para cargar datos hay que llamar 10 veces al endpoint /getproducts

Curl de ejemplo

curl --location 'http://127.0.0.1:5000/getproducts' \

--header 'Content-Type: application/json' \

--data '{

"product":"pc"

}'

2. Modelar

Para este punto decidí analizar la respuesta y pasar a un modelo relacional todos los datos que aportan valor al modelo, por esto no cree tablas para Dominio y Categoria, que si bien tenían IDs en el producto, no tenian contenido, por lo que deje los IDs de Dominio y Categoria en Producto por si alguna vez se requiere crecer el modelo o hacer alguna consulta para completar la información con otro endpoint.

Por otro lado, no agregue otra tabla mas sobre Promocion ya que no era algo recurrente en todos los productos, ya que varios no los tenían y prefería hacer mi análisis en otras características de los productos, siendo algunas de estas:

* Hacer una comparativa de precios entre productos
* Ver que vendedor tiene más productos
* Ver que producto tiene mas atributos en sus publicaciones
* Ver que cantidad de cuotas es más recurrente
* Ver si hay mas envíos gratis o pagos

El script para generar el modelo lo pueden ver en el path

data\model\dataset-gen.sql

Reporte de Looker studio

https://lookerstudio.google.com/reporting/557dcca1-5d9e-4bf2-bc33-58e907be071d