

Challenge para puesto Data Engineer

Franco Balich

Repositorio de GitHub

<https://github.com/francobalich/DataEngineer-Meli-Challenge>

Reporte de Looker Studio

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

1. Descargar la data

El código para descargar los datos lo pueden encontrar en:

src\mercadolibre\get_products.py

Al mismo lo pueden probar y obtener los productos en un jsonl en:

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"  
}'
```

Esto puede ser con el mismo producto o con distintos.

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 tenían 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.

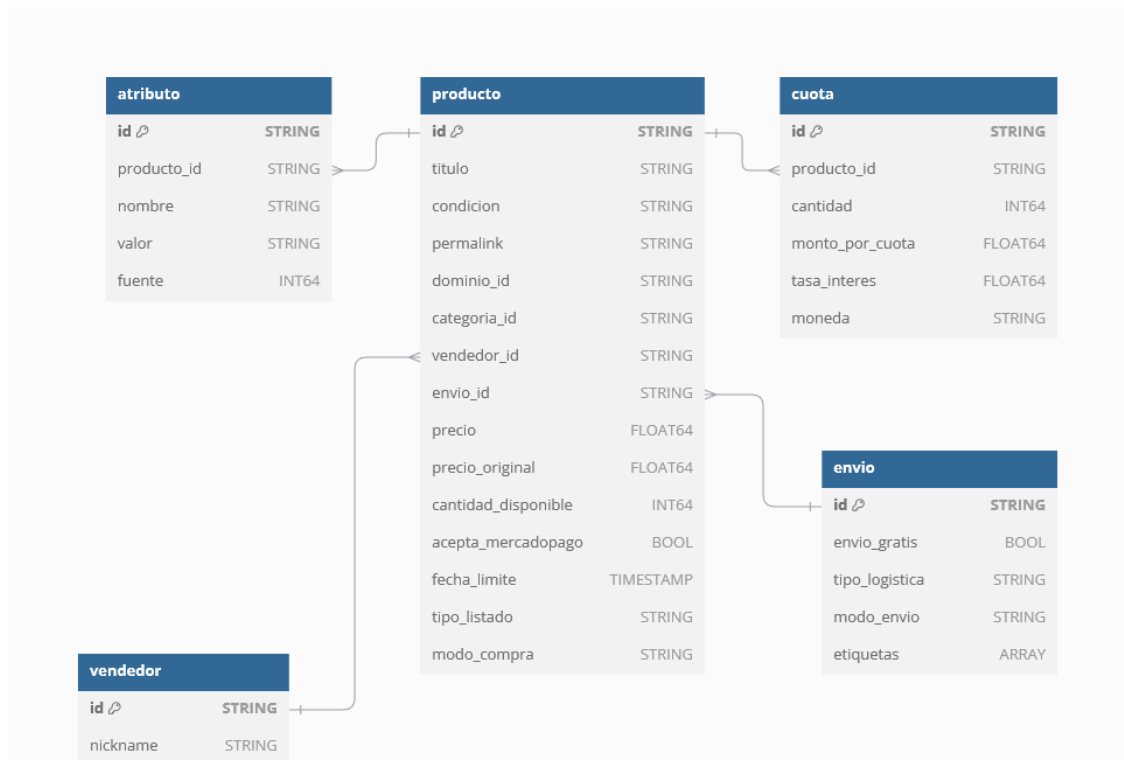
Por otro lado, no agregue otra tabla más sobre Promocion ya que no era algo recurrente en todos los productos, ya que varios no los tenían y prefería hacer mi análisis en otras características de los productos, siendo algunas de estas:

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

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

data\model\dataset-gen.sql

Diagrama entidad relación



3.Parsear

El código del parseo lo pueden encontrar en `src\mercadolibre\product_processor.py` donde tomo los valores que me interesan el json `productos.jsonl` y genero los jsonl para cada tabla para luego importarlas en bigquery

Dichos archivos los pueden encontrar en los paths

- `data\conversions\atributo.jsonl`
- `data\conversions\cuota.jsonl`
- `data\conversions\envio.jsonl`
- `data\conversions\producto.jsonl`
- `data\conversions\vendedor.jsonl`

Para realizar la conversión deben ejecutar el siguiente curl:

```
curl --location --request POST 'http://localhost:5000/convertdata' \  
--data "
```

4.Looker

Para este punto el código que use para generar los views y los explorer en LookML se puede encontrar en el path

- `src\mercadolibre\lookml_generator.py`

y pueden ver los archivos convertidos en los paths:

- `data\lookml\atributo.view.lkml`
- `data\lookml\envio.view.lkml`
- `data\lookml\producto.view.lkml`

- data\lookml\vendedor.view.lkml
- data\lookml\producto.explore.lkml

Para realizar la conversión deben ejecutar el siguiente curl:

```
curl --location --request POST 'http://localhost:5000/genlookml' \  
--data "
```