

Propuesta Solucion - Desafio NUBIMETRICS

Marx Morales Tello

July 14, 2021

1 Introduction

Se presenta una propuesta de solución para el desafío brindado por la empresa NUBIMETRICS, para el postulante al área de datos, perfil Semi Senior.

Dicha propuesta muestra la utilización de herramientas para manejo de datos masivos, y la integración de los mismos; además de scripts para consumir datos de la Mercado Libre

2 Defición del Problema

El desafío fue dividido en 3 partes para su mejor análisis y desarrollo

2.1 Interacción con la API de Mercado Libre

Para plantear la solución a este problema utilizamos Python como lenguaje de programación, esta decisión se tomó por la sencillez, potencia y documentación que tiene esta herramienta.

- Se definió un archivo: **"utils.py"** el cual contiene la función: **"def almacenamiento(namecarpeta,namefile)"**. La función retorna una ruta donde se almacenarán los .json que se genera. El motivo principal de crear este archivo es la reutilización en procesos posteriores.
- Se declaró una clase: **"Consulta_Api_ML"**, es la encargada de realizar la CONSULTA y EXTRACCIÓN de la API de ML, utilizando la librería **'requests'**. Se pensó en una clase que consulte cualquier URL de la API de Mercado Libre, y con un valor del OFFSET ingresado como parámetro.
- El proceso termina almacenando un archivo JSON en la ruta solicitada en el desafío.

2.2 Integración de Datos

Contemplamos aquí los desafíos: 3, 4,5,6,7. Para la realización de estos se utilizó la librería PANDAS de python.

2.2.1 Ventaja

- Permite la lectura de archivos JSON sin complicación ni muchas líneas de código.
- El tipo de datos que maneja son DataFrame, y facilita la manipulación de datos con indexación integrada.
- Nos facilita la extracción de algunas columnas para la creación de nuevos DataFrame

2.2.2 Desventaja

- La principal desventaja que encontramos es el tiempo en procesamiento para archivos pesados. Llevándonos a pensar en otras opciones de manejo

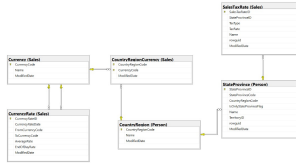


Figure 1:

2.2.3 Desición

Sujeto al requerimiento del desafío concluimos que pandas es una buena herramienta para el desarrollo de la solución.

2.2.4 Dasafio 4,5,6,7

El desafío se resolvió utilizando una notebook de Jupyter, el problema encontrado que nos llevo a tomar la desicion de utilizar esta herramienta sobre un editor de codigo fue la necesidad de visualizar los dataframe resultantes de cada desafio sin necesidad de almacenar en disco(Desafio 5, requeria de un dataframe resultante del desafio4, desafio 6requeria del desafio 5, y asi sucesivamente)

2.3 Desarrollo de SCRIPTS

Corresponde al desafío 8 y desafío 9. Utilizamos la libreria "datetime" de python: `from datetime import datetime, timedelta` De esta forma pudimos transformar eficientemente string a fechas y realizar las operaciones sobre las mismas.(sumar y restar días)

2.4 SQL

Se realizó la descarga de AdventureWorks2014.bak, y la restauración siguiendo la documentación de Microsoft. Para el desafío 12, se genero el diseño de la base de datos para visualizar la relacion existentes entre las tablas.

Figure 1