

Prueba técnica Lulo Bank - Analytics engineers

Versión: 2022-06-15

Actividades

Realizar esta prueba utilizando Python (Utilizar buenas prácticas de codificación, se dará puntos extra pruebas unitarias).

1. Obtener información del siguiente API Rest <http://api.tvmaze.com> trayendo todas las series que se emitieron en diciembre del 2020.
Ayuda: para obtener las series emitidas el 29 de mayo del 2020 se utilizó el siguiente llamado <http://api.tvmaze.com/schedule/web?date=2020-05-29>
2. Con base a los Json obtenidos del API generar diferentes dataframes (en cualquier tecnología: pandas, dask, pyspark) que conserven la integridad referencial de los datos del Json.
3. Realizar profiling a los DataFrames obtenidos en el punto anterior y realizar un análisis de los datos obtenidos.
 - a. Se espera el resultado del profiling (documento en PDF o HTML) y el análisis de éste.
4. Realizar operaciones de limpieza (si las considera necesarias) de los datos que están en los dataframes.
5. Almacenar los diferentes DataFrames en una base de datos (sqlite) con integridad referencial entre las tablas creadas.
6. El área de negocio necesita generar gráficos con la siguiente información
 - a. averageRuntime por tipo (type) por mes.
 - b. Cantidad de series emitidas por género por mes.
 - c. Porcentaje por país, con respecto al total de series emitidas por mes.
 - d. Rating promedio de series por país y género por mes.

Generar ETL(s) (en python o en sql) que lleve los datos a un modelo de datos nuevo que permita responder a las preguntas del área de negocio (puede utilizar como fuente de datos de los dataframes o la base de datos creada).

Entregables:

Link a un repositorio de github que contenga:

- README.md (pasos de instalación/ejecución como minimo)
- Carpeta src/ con el proyecto de python que desarrolló el ejercicio (notebooks o scripts .py).

- Carpeta json/, con los json obtenidos de las consultas al API.
- Carpeta profiling/, con el archivo del profiling y un archivo adicional del análisis de éste.
- Carpeta db/ con el archivo de la base de datos SQLite.
- Carpeta model/, con imagen del modelo de datos creado para almacenar información y para almacenar los resultados de las etls.