

# Prueba técnica Lulo Bank - Data engineers

Versión: 2023-06-01

Realizar esta prueba utilizando Python (Utilizar buenas prácticas de codificación).

- Debe generar un repositorio en github donde esté toda la información requerida para ejecutar el proyecto y los resultados obtenidos. (Se va a tener en cuenta los commits realizados al repositorio)
- Puntos extra por pruebas unitarias a las funciones/clases desarrolladas.

## Actividades

1. Obtener información del siguiente API Rest <http://api.tvmaze.com> trayendo todas las series que se emitieron **en diciembre del 2022**.  
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. Almacenar los datos crudos (json).
3. Con base a los Json obtenidos del API, generar diferentes dataframes con pandas (sugerido) que conserven la integridad referencial de los datos del Json.
4. Realizar profiling a los DataFrames y realizar un análisis.
  - a. Se espera el resultado del profiling (PDF o HTML) y el análisis de éste.
5. De acuerdo al profiling del punto anterior, realizar operaciones de limpieza a los datos de los dataframes en caso de ser necesario.
6. Almacenar los diferentes DataFrames en archivos parquet (con compresión snappy).
7. Leer los archivos parquet del punto anterior y almacenar esta información en una base de datos (sugerimos sqlite), en un modelo definido por ustedes que respete la integridad de los datos.
8. A partir de los archivos parquets o de la base de datos, realizar operaciones de agregación para obtener:
  - a. Runtime promedio.
  - b. Conteo de shows de tv por género.
  - c. Listar los dominios (web) del sitio oficial de los shows.

## Entregables:

Link a un repositorio **público** 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 data/ con los archivos parquet generados..
- Carpeta db/ con el archivo o export de la base de datos generada.
- Carpeta model/, con imagen del modelo de datos creado para almacenar información.