



Data Quality Engineer – Prueba Técnica


Reto: Spotify API

Primera parte:

El siguiente archivo fue extraído de la API de Spotify  `taylor_swift_spotify.json` y usted deberá crear un script en Python que lo procese y lo deje con el siguiente formato  `dataset.csv`. Para esta primera parte de la prueba solo podrá usar Python y las librerías Pandas y JSON.

Entregable: Archivo de Python con la solución del problema.

Segunda parte:

Luego de procesar el archivo descargado de la API de Spotify del punto anterior, deberá hacer un análisis de calidad de datos sobre ese  `dataset.csv` y generar un reporte de los resultados. Para hacer este análisis puede usar Python, SQL, cualquier librería, framework o servicio que considere necesario. En esta parte de la prueba se evaluará su atención al detalle, así que lo importante es encontrar **todas** las anomalías de calidad de datos del dataset. No deberá corregir las anomalías que encuentre, solo encontrarlas y justificarlas en el reporte.

Entregable: Reporte de calidad de datos del dataset, adicionalmente puede incluir el código que muestre cómo encontró las anomalías de calidad de datos. También puede incluir vídeos o cualquier material que considere necesario para explicar el proceso que realizó y los resultados de su análisis. El reporte puede estar en PDF, HTML o el formato que considere necesario.

Documentación de referencia:

[The Six Primary Dimensions for Data Quality Assessment](#)

[Spotify Web API – Track](#)

[Spotify Web API – Album](#)

[Spotify Web API – Artist](#)

[Spotify Web API – Track Audio Features](#)