

DATABASE

SPOTIFY WITH PYTHON AND POWER BI



JONATAN AGUSTIN CAMACHO SANCHEZ

JAIR ALEJANDRO GUTIERREZ GONZALEZ

DIEGO CASTILLO SALAZAR

Universidad Cuauhtémoc de Querétaro
Big Data - TANIA DANIELA GARCIA MUNOZ

INSTRUCCIONES

- ☐ Descarga una base de datos en xls, csv, etc.
- ☐ Carga los datos en Python.
- ☐ Elabora una limpieza de datos.
- ☐ Elimina nulos repetidos.
- ☐ Eliminaciones de datos innecesarios

INTRODUCCIÓN

Este proyecto tiene como objetivo la visualización de la base de datos descargada con la aplicación de la librería Spotpy, al igual que la integración de Power BI y Python en Visual Studio Code.





Para el tratamiento de limpieza de datos y la recolección de información musical.

El análisis se centra en la popularidad, así como en los géneros musicales de la época y del año a nivel global con 10k registros.

OBJETIVO

El objetivo principal de este proyecto es integrar habilidades técnicas en el manejo de bases de datos, limpieza de información y representación visual para comprender la popularidad de las canciones en Spotify. Así mismo, se busca explorar patrones de consumo musical con base en métricas como el año de lanzamiento, la popularidad y el género.

Spotify

-  1. Descargar una BD en xls, csv, etc.
-  2. Cargar los datos en Python
-  3. Hacer una limpieza de datos, eliminar nulos, repetidos, hacer casteos de datos si es necesario, etc.
-  4. Hacer una visualización de datos que nos muestre:

HERRAMIENTAS UTILIZADAS

- Python: para extracción de datos desde la API de Spotify mediante la librería Spotipy, limpieza y transformación de datos.
- Power BI: para la carga, modelado y visualización de datos.
- Spotify API: para obtener información actualizada y real de canciones, artistas y popularidad.
- Pandas / Openpyxl: para manipulación y exportación de archivos CSV y Excel.

¿Qué se elaboró en este proyecto?

Se utilizaron diferentes herramientas en las cuales utilizamos algunas librerías y APIs desarrolladas. Se utilizó Spotipy para la extracción de la Database a Python con más de 10k de canciones agrupadas por género musical como el pop, reggaeton, etc. Estos registros fueron optimizados para la eliminación de datos innecesarios y duplicados.

A lo cual se generó un archivo CSV; posteriormente se realizó la estructura de estos datos, con el siguiente formato:

Nombre de la canción
Artista
Popularidad
Año de lanzamiento
Género (cuando fue posible tenerlo)

Posteriormente, para el análisis de estos datos limpios, se cargaron los 2 datasets a Power BI.

Al igual que también se realizaron relaciones, creación de medidas y columnas calculadas para facilitar el análisis.

Algunos dashboards implementados fueron

1. Top 10 canciones por popularidad

Gráfico de barras mostrando las canciones más populares según el campo "Popularidad".

2. Popularidad a lo largo del tiempo

Gráfico de líneas donde se visualiza el promedio de popularidad por año de lanzamiento.

3. Distribución de canciones por artista

Gráfico de barras verticales mostrando cuántas canciones tiene cada artista en el dataset.

4. Comparación de canciones por nombre y artista

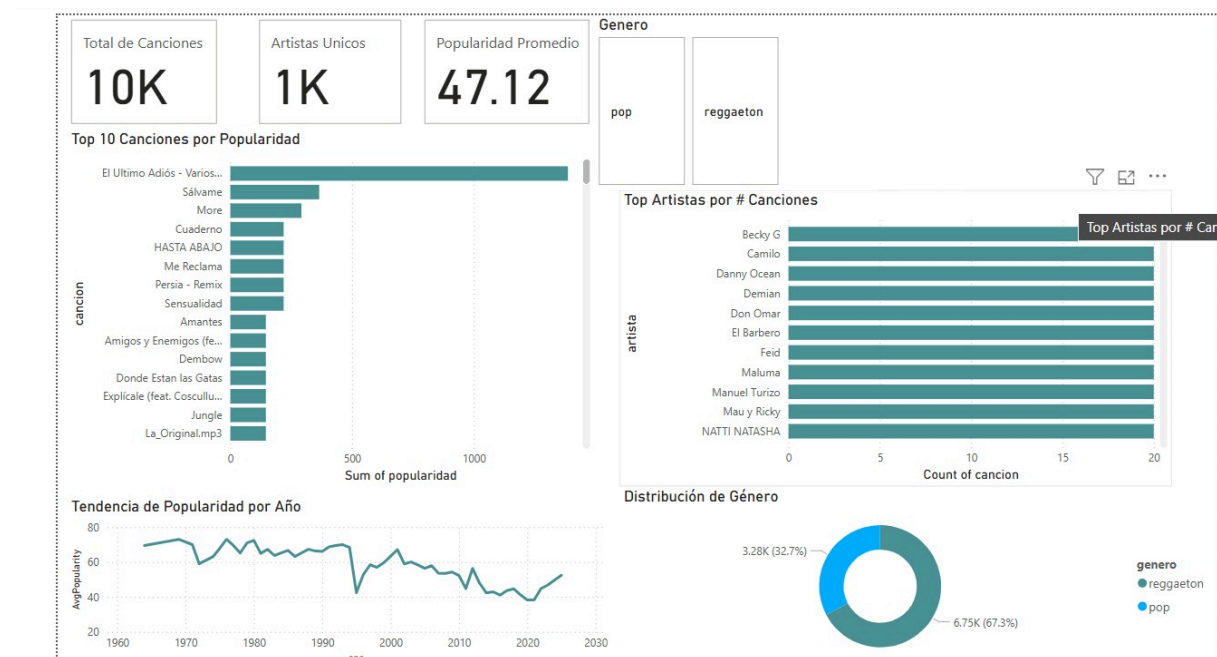
Tabla interactiva para buscar y comparar canciones de un mismo artista.

5. Dashboard general de top 10 globales

Se elaboró un dashboard separado con el archivo spotify_top_tracks_global.csv donde se destacan las canciones de tendencia actual.

6. Filtro por género musical (cuando aplica)

Aunque la variedad de géneros fue limitada por la API, se incluyó un slicer para explorar los datos segmentados por género.



Conclusiones

Este proyecto permitió consolidar habilidades clave en el manejo de datos musicales reales, integrando Python y Power BI en un flujo de trabajo que simula un caso profesional.

Este análisis fue muy bueno, ya que pudimos ver claramente los cambios a través de los años en cómo la sociedad fue evolucionando e integrando algunos nuevos géneros musicales, al igual que la evolución de algunos artistas al ver a grandes rasgos cómo algunas de sus canciones se mantenían de manera positiva ante el público por varios años.

Estas visualizaciones en Power BI expresan de manera gráfica y ordenada cómo la música en el entorno de Spotify va en aumento y crecimiento al utilizar con más constancia este tipo de aplicaciones y sus visualizaciones como tal y como esta puede ser comparada con grandes empresas y competencias como lo son Apple Music y YouTube, entre otras.

Aunque hubo ciertas limitaciones en los datos (como el acceso restringido a géneros o top 1000 por consulta), en pocas palabras, el resultado final ofrece una vista clara y atractiva del panorama musical actual.

Algunas fuentes y documentaciones en las que nos basamos para hacer este proyecto posible fueron las siguientes.

<https://spotipy.readthedocs.io/en/2.22.1/#redirect-uri>

<https://developer.spotify.com/documentation/web-api/tutorials/getting-started>

<https://www.youtube.com/watch?v=j4J7B1C3G0U>

<https://www.youtube.com/watch?v=6NqYy0RrLiY>