

SQL PROJECT - SPOTIFY DATASET

Problem Statement -

This project focuses on analyzing the Spotify dataset from Kaggle. It has various attributes about a song such as tracks, albums, etc. It covers and solves a set of SQL queries.

About Dataset

Introduction to the [Spotify Dataset](#) -

- Here we are specifying the datatype as we will be uploading the dataset by creating a table in PostgreSQL.

Column name	Datatype	Description
artist	VARCHAR(255)	Name of the artist
track	VARCHAR(255)	Name of the track
album	VARCHAR(255)	Name of the album
album_type	VARCHAR(50)	Album of single
danceability	FLOAT	A measure value
energy	FLOAT	A measure value
loudness	FLOAT	A measure value
speechiness	FLOAT	A measure value
acousticness	FLOAT	A measure value
instrumentalness	FLOAT	A measure value
liveness	FLOAT	A measure value
valence	FLOAT	A measure value

tempo	FLOAT	A measure value
duration_min	FLOAT	Time of the song
title	VARCHAR(255)	Title of the song
channel	VARCHAR(255)	Name of the channel
views	FLOAT	Number of views received
likes	BIGINT	Number of likes received
comments	BIGINT	Number of comments received
licensed	BOOLEAN	True of not
official_video	BOOLEAN	True of not
stream	BIGINT	
energy_liveness	FLOAT	
most_played_on	VARCHAR(50)	Where is it played - youtube or spotify

Data Cleaning -

The dataset is already cleaned.

- No missing values
- No duplicates.

STEPS FOR SOLVING -

1. In PostgreSQL create a database
2. Create a table called Spotify and insert the data.
3. Import the values in the table with the import dataset function
4. If there is any data type issue either fix it in excel or in the postgresSQL.
5. Perform EDA and get to know your data better
6. Solve some easy to hard SQL queries to answer questions about the data.