

Project: Investigate Spotify 2000-2019 Dataset

- **Introduction**

This project was done using a Spotify dataset from Kaggle at:

<https://www.kaggle.com/datasets/paradisejoy/top-hits-spotify-from-20002019/code>.

I used Univariate, Bivariate and Multivariate exploration to make my observations.

The libraries used to make the analysis include;

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Plotly

Information On The Dataset

Danceability: describes how suitable a song is for dancing

Energy: is a measure from 0.0 to 1.0 and represents perpetual measure of intensity and activity

Speechiness: detects the presence of spoken words in a song

Acousticness: This measures from 0.0 to 1.0 whether the song is acoustic

Liveness: Higher liveness values represent an increased probability that the track was performed live

Valence: This describes the musical positiveness conveyed by a track from 0.0 to 1.0

Cleaning Observations

All these observations were fixed.

- Some column names can be more descriptive.
- The duplicated rows are from artists with multiple songs, they therefore do not need to be cleaned.
- First alphabets of columns are not capitalized.
- Columns are not arranged alphabetically.

Summary Of Findings

This dataset had a lot of data, who would have known that Sweater Weather by The Neighbourhood is the most popular song on Spotify between 2000-2019? or that pop is the most recurrent genre? Or that most songs do not have explicit content?.

There are some observations that are definitely clear, like the fact that the duration of songs vary track by track.

The following are the questions I came up with for this data and my observations.

Data Analysis

Some questions asked for this EDA include;

Question 1

What are the top 5 most popular songs on Spotify between 2000-2019 and artist ?

Observations 1

The top 5 most popular song on Spotify between 2000-2019 are:

- The Neighbourhood-Sweater Weather
- Tom Odell-Another Love
- Eminem-Without Me
- WILLOW-Wait a Minute!
- Eminem-The Real Slim Shady

The Neighbourhood is an Indie Rock Pop(Rock,pop is a sub-genre of Pop) band, which makes sense because Rock Pop is the most popular genre, the other artists on this list however are Pop artist.

Question 2.

What are the most popular music genres ?

Observation

Rock, pop is the most popular genre

Question 3.

What are the most repetitive music genres ?

Observations 3

Pop is the most repetitive genre.

This observation was made using a pie-chart.

Pop music is probably the best choice for an artist, it's popular, however it is also a very repeated genre already.

Question 4.

What is the most popular energy level of Songs ?

Observation 4

0.8 has the highest frequency

Question 5.

Who are the most frequent artist ?

Observations 5

Rihanna, Drake and Eminem are among the most frequent artists is the most frequent artist

Question 6.

Least Popular Artists By Song

Observations 6

Camille Jones has the least popular song

Question 7.

Is the duration of songs the same ?

Observations 7

The duration of songs varies from song to song .

Question 8.

Does the energy and acousticness correlate ?

Observations 8

There is some correlation between energy and acousticness

Question 9

How popularity is correlated to danceability ?

Observation 9

As the popularity of the song increases, the danceability score for that song also increases.

Question 10.

What songs belong to what artist, and what is the popularity of each ?

Observation 10

The artist with the most songs is Rihanna and the lowest is Camille Jones

Question 11.

Whats the yearly song collection ?

Observation 11.

The year with the highest song collection is 2012

Question 12.

What decade had the most songs?

Observation 12.

The 2010's had the highest number of songs

Question 13.

What decade had the most explicit content ?

Observation 13.

The 2010s had the songs with the highest number of explicit content

Question 14.

Are songs with no explicit content more popular ?

Observation 14.

Songs with explicit content are less popular than songs without explicit content .

Question 15.

What amount of songs have explicit content ?

Observation 15.

Only 27.6% of songs have explicit content