**Analyze dataset for any specific domain. Perform detailed EDA and state the insights and inferences**

**Name:** Pravin B  
**Reg No:** 23127035  
**Course:** BSc CS with DA (2nd year)

**1. State the Hypothesis or Assumptions**

* **Hypothesis 1:** A few top artists dominate the most streamed songs on Spotify.
* **Hypothesis 2:** There is a positive correlation between the number of playlists a song is included in and its number of streams.
* **Hypothesis 3:** Songs with higher danceability scores are among the most streamed.

**2. Subset the Data to Verify and Gather Insights**

* **Data Cleaning:**
  + Combined released\_year, released\_month, and released\_day into a single release\_date column.
  + Dropped unnecessary columns (bpm, key, mode).
  + Removed duplicates based on track\_name.
  + Converted streams and in\_deezer\_playlists to numeric formats.
  + Removed rows with null values.
* **Top 10 Artists:**
  + The analysis showed the top 10 artists with the most hits in 2023.
* **Top 10 Streamed Songs:**
  + Identified the top 10 streamed songs in 2023.
* **Platform Totals:**
  + Calculated the total number of songs in playlists for Spotify, Apple, and Deezer platforms.
* **Top 10 Songs vs. Danceability:**
  + Determined the top 10 songs with the highest danceability scores.

**3. Describe the Insights and Assumptions**

* **Insight 1:** A few artists had a significant number of hits, indicating their dominance in 2023. This supports the hypothesis that a few top artists dominate the most streamed songs.
* **Insight 2:** There is a strong positive correlation between the number of playlists a song is included in and its streams. Songs featured in more playlists tend to have higher streams.
* **Insight 3:** Songs with higher danceability scores are among the most streamed, suggesting that danceability is a significant factor contributing to a song's popularity.

**4. Description of the Problem with Insights:**

**Objective:**  
Perform an in-depth exploratory data analysis (EDA) on the Spotify dataset for 2023 to uncover factors contributing to a song's popularity on the platform.

**Problem Statement:**

1. **Identify Top Artists:**
   * A few artists had a significant number of hits, indicating their dominance in 2023.
2. **Correlation Between Playlists and Streams:**
   * There is a strong positive correlation between the number of playlists a song is included in and its streams.
3. **Danceability and Popularity:**
   * Songs with higher danceability scores tend to be among the most streamed.

**Conclusion:**

The analysis reveals that top artists, playlist inclusion, and danceability are key factors influencing a song's popularity on Spotify. These insights can guide artists, producers, and marketers in optimizing their strategies for success on the platform.

**GitHub Repository :** https://github.com/pravinboopathi/spotify-analysis