

Spotify Songs-2024 Data Analysis

Dataset:- <https://www.kaggle.com/datasets/nelqiriyewithana/most-streamed-spotify-songs-2024>

General Guidelines

- Ensure you use the "latin-1" encoding when reading the CSV file to handle special characters correctly.
 - Handle missing values thoughtfully; explain your approach and justify the chosen method (e.g., mean/median imputation, removal of missing entries, etc.).
 - Convert date columns to datetime format and extract relevant features, such as the year, to facilitate analysis.
 - Clearly document your analysis steps and the rationale behind your decisions.
 - Ensure your code is well-organized, commented, and adheres to best practices.
 - Present your findings in a clear, concise manner, using visual aids and explanations where necessary.
 - Submit Google drive Folder Link which includes both your .ipynb file along with summary pdf.
-

Easy Questions

1. **Basic Descriptive Statistics:**
 - What is the average Spotify Streams count for all tracks in the dataset?
 - Calculate the median Shazam Counts.
 - Find the mode of the All Time Rank column.
 2. **Data Filtering:**
 - List all the tracks that have an Explicit Track value of 1.
 - Calculate the average Spotify Streams count for tracks released in 2024.
-

Intermediate Questions

1. **Aggregation and Grouping:**
 - Calculate the total Spotify Streams for each artist.
 - Find the artist with the highest average Track Score.
 - Find the artist with the highest total Deezer Playlist Reach.
2. **Data Visualization:**
 - Create a bar chart showing the top 10 tracks by Spotify Streams.
 - Plot a bar chart of the top 5 tracks by Shazam Counts.
3. **Conditional Operations:**

- Identify tracks that have more than 1,000,000,000 Spotify Streams and more than 10,000,000 Shazam Counts.
-

Advanced Questions

1. Advanced Aggregation and Complex Queries:

- Which artist has the highest combined count of Spotify, Amazon, and Pandora streams?
- Calculate the correlation between Spotify Streams and Spotify Playlist Reach.

2. Scenario-Based Analysis:

- A music production company wants to promote tracks that are popular on multiple platforms. Identify tracks that are in the top 10% for both Spotify Streams and Pandora Streams.