

# **Spotify Track Analysis**

Yoatam Gebremicael

## Background Of The Problem

In this analysis, I was given two major questions to address about the dynamics of track popularity and musical attributes on Spotify:

- *How does the presence of one or more highly popular songs in an album affect the popularity and listener engagement of other tracks within the same album?*
- *How do the musical attributes of tracks change when artists collaborate compared to when they work solo?*

To explore these questions, I divided them into four sub-questions:

- Do high-popularity albums show higher listener retention for less popular tracks?*
- Do features that make tracks popular, like high energy or danceability, lead to more interaction with other tracks on the same album that have those qualities?*
- Do collaborations cause noticeable changes in musical qualities like tempo, loudness, and energy versus solo tracks?*
- Does the popularity of collaborations vary by genre, possibly because collaboration tracks are more popular than solo performances in certain genres?*

These sub-questions can help in finding patterns of listeners' behavior, track popularity, and how collaboration might affect musical attributes.

## Data Analysis Steps

### ***Data Cleaning and Transformation:***

- **Handling Missing Values:** The missing values have been cleaned, especially in columns related to artist, album name, and track name.
- **Duration Conversion:** Duration values are in milliseconds and were converted to minutes for better understanding.
- **Popularity Threshold:** A popularity threshold was considered to define highly popular tracks: tracks above the 80th percentile. This threshold was then used to differentiate albums containing highly popular tracks from those without, addressing my first research question.

## ***Exploratory Data Analysis (EDA):***

- **Visualizations:**

- **Box Plots:** Box plots were created comparing the distribution of popularity for albums with and without highly popular tracks; this helped me in answering the first sub-question, which was to determine if high-popularity albums show higher listener retention for less popular tracks. Another box plot compared musical attributes between solo and collaborative tracks. This visualization helped to identify the second research question: how musical attributes change in collaborative versus solo tracks. A third box plot compared these attributes across genres for solo and collaborative tracks, which helped to analyze how collaboration influences musical qualities across different genres.
- **Scatter Plot:** A scatter plot was created to show how popular tracks influence the popularity of other tracks within the same album, especially if the tracks share common musical attributes or features such as energy or danceability. This addressed my sub-question of whether high-energy or danceable tracks encourage more interaction with other tracks in the album.

- **Summary Statistics:** Summary statistics on important attributes have been calculated to understand central tendency and variability in these data. An estimate of the mean has also been done to mark the highly popular tracks and, at the same time, to set a threshold to select those albums with popular tracks. In comparing attributes of solo versus collaboration across track genres, this statistic also comes in handy.

- **Hypothesis Testing (T-tests):**

- **T-test for Album Popularity:** A t-test was conducted to test whether the observed difference in popularity between albums that contain highly popular tracks and those that don't was statistically significant. The test confirmed albums with highly popular tracks would have notably higher average popularity than those without ( $p\text{-value} < 0.0001$ ).
- **T-test for Musical Attributes by Genre:** Another t-test was done to see how musical attributes vary across genres for solo versus collaborative tracks. This test indeed confirmed that in some genres, like Jazz, the collaborations resulted in higher energy and more danceable tracks, while for Blues, the collaborations resulted in more acoustic and slower tracks.

## Main Results/Conclusions

- Albums with at least one highly popular track have higher overall popularity. This suggests that highly popular tracks might positively affect listener engagement for other tracks of the same album. This was further justified through statistical analysis, as the t-test result showed high statistical significance,  $p\text{-value} < 0.0001$ , hence proving that the presence of popular tracks increases overall album popularity.
- This analysis shows that the shared characteristics between tracks within an album and the highly popular track do not drive more interaction or popularity. In fact, it seems that variety across different musical attributes is what drives higher engagement. This would, in turn, suggest that listener interest may be more about the overall diversity of tracks within an album rather than having multiple tracks with similar attributes to the highly popular one.
- Overall, solo tracks are more energetic and faster, and people can dance to them. Solo artists tend to create music that is loud and upbeat, but these collaborative tracks are generally quieter and more acoustic in nature, with a more relaxed and natural sound. That would say collaboration often does tend to shift the style of music to subtlety and less on high energy and speed.
- Collaborations affect the sound of tracks differently across genres:
  - **Blues:** It's really acoustic and slower in collaborative tracks, giving it more of a reflective vibe.
  - **Disco:** This genre has more danceable and upbeat collaborations, likely to engage a listener through rhythm and tempo.
  - **Emo:** The collaborative tracks are slightly more energetic and with a more positive feel, though often slower tempos.
  - **Jazz:** Collaborations show higher energy and danceability but tend to be less acoustic than solo tracks, which might mean that the style in collaborative Jazz is more dynamic and groove-oriented.

**Conclusion:** The collaborations do tend to change the mood and style of the music, making it more varied and sometimes more accessible to a wider audience; significant changes can be observed across genres.

## **Drawbacks and Concerns**

While the analysis was instructive, certain limitations have to be taken into account:

### **Dataset Limitations:**

- The dataset is missing some key attributes, such as artist gender and release year, that could have provided more insight into how collaborations vary by artist characteristics or over time.

### **Outliers in the Dataset:**

- Outliers in key musical attributes were examined for energy, danceability, loudness, tempo, valence, and acousticness. However, for most of these attributes, their influence was less than 1%. For loudness, this number is slightly higher, at about 2%, but since loudness was not core to the analysis, their retention did not significantly affect the results. Hence, outliers were not removed to ensure data completeness and to avoid the addition of unnecessary complexity.

### **Generalization Issues:**

- The analysis was done based on a certain sample of tracks; this may not be representative of all music genres or collaborations. For example, other genres, such as Classical or Hip-Hop, might have different dynamics in collaboration, which were not explored here.
- External factors such as marketing, social media trends were not accounted for in the analysis.

### **Statistical Power:**

- While t-tests were used in the analysis of differences between solo and collaborative tracks, regression models could have been used that better quantify the exact effect of collaboration on musical attributes, with controls for genre and track length.