**SPOTIFY DATA ANALYSIS**

**BY**

**Fitsum Metaferya**

**Michel Tshilumba**

**Tyquese Taplin**

**Project1 – Group2**

**Research Questions:**

**Which songs are liked, and which songs are not?** We were given 2017 rows of data that contained songs. After cleaning the data by dropping the duplicated rows we were left with 2012 rows. Out of the 2012 songs 1039 were liked songs and 973 of the songs were disliked. Overall, the two categories were almost evenly distributed, presenting a pie chart graph with **51.6%** of songs liked and **48.4%** of songs disliked. A deeper look into songs attributes could help explain this pattern.

A pie chart of song distribution

Description automatically generated

**Who is the most popular artist?** The data presented us with a list of artists corresponding to the number of times an artist’s songs were mentioned. **Drake** is the most popular artist in the collection with **13 songs**. **Disclosure** came in a close second with a total of **12 songs** in the dataset collection.

A chart with green lines

Description automatically generated

**Does danceability make an artist more popular?** The data supplied showed the relationship between **average danceability** and **popularity**, measured as the number of liked songs by various artists. The artists with a popularity values greater than or equal to .08 tend to occur for artists with an average danceability greater than 0.8. This implies a possible trend where artists with more danceable music might have higher numbers of liked songs. While there seems to be a slight positive correlation between average danceability and popularity, the relationship is not strong.

A graph of a dance

Description automatically generated with medium confidence

**Does the duration of the song affect the likeability of a song?** The data provided was graphed using a box plot that compares the duration of liked and unliked songs. The median duration of liked songs (the central line within the box) is higher compared to unliked songs. This indicates that, on average, liked songs tend to have longer durations than unliked songs. Both liked and unliked songs have a considerable number of outliers, reflecting songs with significantly longer durations. It appears that longer songs are generally more liked, though there are exceptions as indicated by the outliers.

A graph of a comparison between a number of blue boxes

Description automatically generated with medium confidence

**How does energy and tempo affect the likeability of a song?** We used a histogram to illustrate the effects of **tempo** and **energy** has in correlation to liked and unliked songs. **Energy values** range from **0.0** (low energy) to **1.0** (high energy). Liked songs are more prominent in the energy range **0.6 to 0.9,** this explains that users tend to like songs with moderately high energy levels. Unliked songs stand out in the range **0.2 to 0.5,** showing that songs with lower energy levels may be more unliked. The **tempo ranges** from approximately 110–130 BPM in the histogram. Liked songs fall in the range of 120–125 BPM, suggesting that songs in this tempo range are generally more liked. Unliked songs are spread more evenly across tempo ranges but do not peak in any specific range, displaying no significant preference for songs with higher tempo ranges.

A graph of energy versus libra

Description automatically generatedA graph of different colored bars

Description automatically generated

**Hypotheses 1**

Null Hypothesis (H0):

H0: Danceability does not affect an artist's popularity.

Alternative Hypothesis (Ha):

Ha: Danceability affects an artist's popularity.

**Conclusion 1**

From all the results we have noticed that danceability does matter about making the singer popular. Therefore, the Alternative Hypothesis is true.

**Hypotheses 2**

**Null Hypothesis (H0):**

H0: The duration of a song does not affect the likeability of a song.

**Alternative Hypothesis (Ha):**

Ha: The duration of a song affects the likeability of a song.

**Conclusion 2**

From all the results we have noticed that the longer the duration the chance for the song to be liked is high. Therefore, the Alternative Hypothesis is true.

**Hypotheses 3**

**Null Hypothesis (H0):**

H0: The tempo of a song does not affect its likeability.

**Alternative Hypothesis (Ha):**

Ha: The tempo of a song affects its likeability.

**Conclusion 3**

From T-test and ANOVA results we have noticed that tempo does not have any effect on the likeability of a song. Therefore, the Null Hypothesis is true.

**Summary Report:**

In conclusion, this analysis provides valuable insights into the factors influencing the likeability and popularity of songs and artists. The dataset revealed a nearly even distribution of liked and unliked songs, with liked songs slightly outnumbering unliked ones. Drake emerged as the most popular artist, closely followed by Disclosure, demonstrating their strong presence in the dataset. Danceability appears to have a slight positive correlation with popularity, as artists with more danceable music often had more liked songs. Similarly, song duration plays a role, with longer songs generally being more liked, although exceptions exist. Energy and tempo also contribute to likeability, with liked songs often featuring moderately high energy (0.6–0.9) and tempos in the 120–125 BPM range. These findings emphasize how various attributes influence audience preferences, offering a foundation for deeper exploration into what makes music resonate with listeners.

**Future Exploration:**

In the future, our analysis could delve deeper into the **emotional tone of song lyrics**, using sentiment analysis to uncover how emotions conveyed in lyrics correlate with a song's popularity, genre, or listener engagement. Additionally, we could explore how **musical attributes** (such as tempo, energy, and instrumentation) have evolved across decades, revealing long-term trends and shifts in listener preferences.

To gain a sociocultural perspective, we could analyze and compare **genres across geographic locations**, uncovering how cultural influences shape regional music styles. This could include identifying distinct **music trends by region**, offering insights into how local tastes and traditions influence the global music landscape.

**Citations:**

The analysis and conclusions are based on insights derived from the dataset and methodology presented in the Kaggle notebook by Krishna Bhatt, titled *Spotify Song Attributes Analysis* ([source](https://www.kaggle.com/code/krishnabhatt4/spotify-song-attributes-analysis/notebook)).