

SPOTIFY POPULARITY ANALYSIS

Using Data Analysis

Prepared in the partial fulfillment of the Summer Internship Program on Data Analysis

AT



Under the guidance of

Ms. K. Meenakshi, APSSDC

Ms. K. Narmada Mani, APSSDC

Submitted by

Manjari Sripada

AP21110011620

SRM University AP

(August ,2023)

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to all those who have contributed to the successful completion of my summer internship project at **Andhra Pradesh Skill Development Corporation (APSSDC)**. This opportunity has been an enriching and transformative experience for me, and I am truly thankful for the support, guidance, and encouragement I have received along the way.

First and foremost, I sincerely thank Ms. K. Meenakshi and Ms. K. Narmada Mani, my supervisors, and mentors, for providing me with valuable insights, constant guidance, and unwavering support throughout the internship. Their expertise and encouragement have been instrumental in shaping the direction of this project.

I would like to thank the entire team at **Andhra Pradesh Skill Development Corporation (APSSDC)** for fostering a collaborative and innovative environment. The camaraderie, knowledge sharing, and feedback I received from my colleagues significantly contributed to the development and success of this project.

In conclusion, I am honored to have been a part of this internship program, and I look forward to leveraging the skills and knowledge gained to contribute positively to future endeavors.

Thank You.

Sincerely,

Manajri Sripada

AP21110011620

SRM University AP

manjari_sripada@srmap.edu.in

ABSTRACT

This analysis explores the popularity of songs on Spotify using a comprehensive dataset. By examining attributes like genre, artist, acousticness, danceability, and more, insights are gained into the factors driving song popularity. The dataset's exploration covers distribution statistics, genre-based popularity trends, artist influence, and correlations between attributes and popularity. Visualizations such as heatmaps, scatter plots, and bar plots provide clear representations of these relationships. This study contributes to a better understanding of how musical features and genres contribute to a song's popularity on the Spotify platform.

TABLE OF CONTENTS

1. Introduction
2. Libraries Used
3. Methodology
4. System Design/Architecture
5. Implementation
6. Results
7. Conclusion
8. References

Introduction:

In the captivating world of music, understanding what makes songs popular is a fascinating journey. The "Spotify Popularity Analysis" project sets out to uncover the hidden secrets behind why some tracks become hits. With the help of a treasure trove of data from Spotify, this project aims to uncover trends, patterns, and connections that influence a song's popularity. It dives into different musical aspects, artist traits, and the impact of music genres.

Today's music scene is a bustling hub of songs, all striving to capture listeners' hearts. This project is on a mission to answer important questions: What musical elements resonate with people the most? How do different music genres sway popularity? Can artist characteristics play a significant role?

By using a mix of preparing data, exploring the data's nuances, and creating easy-to-understand visuals, the project is here to provide valuable insights into the factors that drive a song's success. This exploration goes beyond mere numbers, uncovering how attributes like danceability, energy, valence, and more work together to shape a song's journey to stardom.

LIBRARIES USED

PANDAS-

A Python library for data manipulation and analysis. Used for reading and doing operations easily on Dataset.

MATPLOTLIB-

A Python library for creating visualizations. We will use it to plot data and explore the dataset.

SEABORN-

Seaborn, an extension of matplotlib, streamlines data visualization with simplified syntax for creating informative statistical plots.

Methodology:

The script is implemented using Python to execute essential tasks, including data cleansing, exploratory data analysis, and visualization. The focus is on a dataset sourced from Spotify, encompassing a wide range of song attributes, artist details, and genre classifications. The ensuing analysis aims to uncover trends, relationships, and patterns that contribute to a song's level of popularity. The investigation delves into diverse musical characteristics, artist traits, and genre influences, providing a comprehensive picture of the elements that drive the success of songs on the platform. Here's a methodology

1. Data Acquisition and Preprocessing:

- Import essential libraries including pandas, matplotlib, and seaborn.
- Load the "SpotifyFeatures.csv" dataset into a pandas DataFrame for analysis.
- Perform initial data cleaning, handling missing values and inconsistent formatting.

2. Descriptive Analysis:

- Generate a comprehensive overview of the dataset's structure, showcasing key attributes and data types.
- Compute summary statistics, shedding light on the distribution of popularity scores.

3. Unveiling Popularity Distribution:

- Utilize histogram plots to visualize the distribution of popularity scores across tracks.
- Calculate statistics such as mean, median, and quartiles to grasp the central tendencies.

4. Genre-based Analysis:

- Group the data by genre to compute average popularity within each genre.
- Visualize the top-performing genres using bar plots, revealing

genre-specific popularity trends.

5. Artist Influence:

- Examine the influence of artist popularity on track popularity.
- Identify the top artists with the highest average popularity and visualize their prominence.

6. Correlation Analysis:

- Compute a correlation matrix to quantify relationships between various musical attributes.
- Visualize correlations using a heatmap, highlighting attributes that strongly influence popularity.

7. Genre, Key, and Time Signature Analysis:

- Explore the relationship between popularity and attributes like key and time signature using box plots.
- Leverage box plots to unravel how different modes and time signatures impact popularity.

8. Scatter Plots and Insights:

- Create scatter plots to illustrate the connection between popularity and musical attributes.
- Draw insights from scatter plots, unveiling trends like the correlation between energy and popularity.

9. Project Conclusion:

- Summarize findings from the analyses conducted, highlighting factors that notably affect track popularity.
- Reflect on the significance of this analysis in providing insights to artists, labels, and music enthusiasts.
- Discuss potential implications for music recommendation systems and marketing strategies.

In this project, I have only done data visualization but for more analysis, we have to use machine learning algorithms.

System Design:

Hardware Requirements:

1. Processor – Intel® Core™ i3-10110U CPU @ 2.10GHz 2.59GHz
2. RAM – 8.00 GB
3. System type – 64-bit operating system, x64-based processor
4. Edition – Windows 11
5. Version – 22H2 or above

Software Requirements:

1. Jupyter Notebook for python codes

Implementation:

Note: Please note that this is a simplified version and doesn't include every detail from your code.

```
# Importing Required Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv('SpotifyFeatures.csv')

# Descriptive Statistics
popularity_stats = data['popularity'].describe()

# Visualizing Popularity Distribution
plt.figure(figsize=(10, 6))
sns.histplot(data['popularity'], bins=20, kde=True)
plt.title('Distribution of Popularity')
plt.xlabel('Popularity')
plt.ylabel('Frequency')
plt.show()

# Average Popularity by Genre
avg_popularity_by_genre = data.groupby("genre")["popularity"].mean()

# Top Genres by Average Popularity
```

```

top_genres = avg_popularity_by_genre.nlargest(10)

# Top Artists by Popularity
top_artists = data.groupby("artist_name")["popularity"].mean().nlargest(10)

# Correlation Matrix Heatmap
numeric_columns = data.select_dtypes(include=[float, int]).columns
correlation_matrix = data[numeric_columns].corr()

# Popularity by Key
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x="key", y="popularity")
plt.title("Popularity by Key")
plt.xlabel("Key")
plt.ylabel("Popularity")
plt.show()

# Popularity by Mode
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x="mode", y="popularity")
plt.title("Popularity by Mode")
plt.xlabel("Mode")
plt.ylabel("Popularity")
plt.show()

# Average Popularity by Time Signature
avg_popularity_by_time_signature =
    data.groupby("time_signature")["popularity"].mean()

# Scatter Plots for Musical Attributes vs. Popularity
attributes_for_scatter = ["acousticness", "danceability", "energy",
    "instrumentalness", "liveness", "loudness", "speechiness", "tempo",

```

```
    "valence"]
plt.figure(figsize=(15, 12))
for attribute in attributes_for_scatter:
    plt.subplot(3, 3, attributes_for_scatter.index(attribute) + 1)
    sns.scatterplot(data=data, x=attribute, y="popularity", alpha=0.5)
    plt.title(f'{attribute.capitalize()} vs. Popularity')
    plt.xlabel(attribute.capitalize())
    plt.ylabel("Popularity")
plt.tight_layout()
plt.show()
```

Bar plots for Popularity Based on Key and Mode

```
plt.figure(figsize=(15, 8))
plt.subplot(2, 2, 1)
sns.barplot(x='key', y='popularity', data=data)
plt.title('Popularity Based on Key')
plt.subplot(2, 2, 2)
sns.barplot(x='mode', y='popularity', data=data)
plt.title('Popularity Based on Mode')
plt.tight_layout()
plt.show()
```

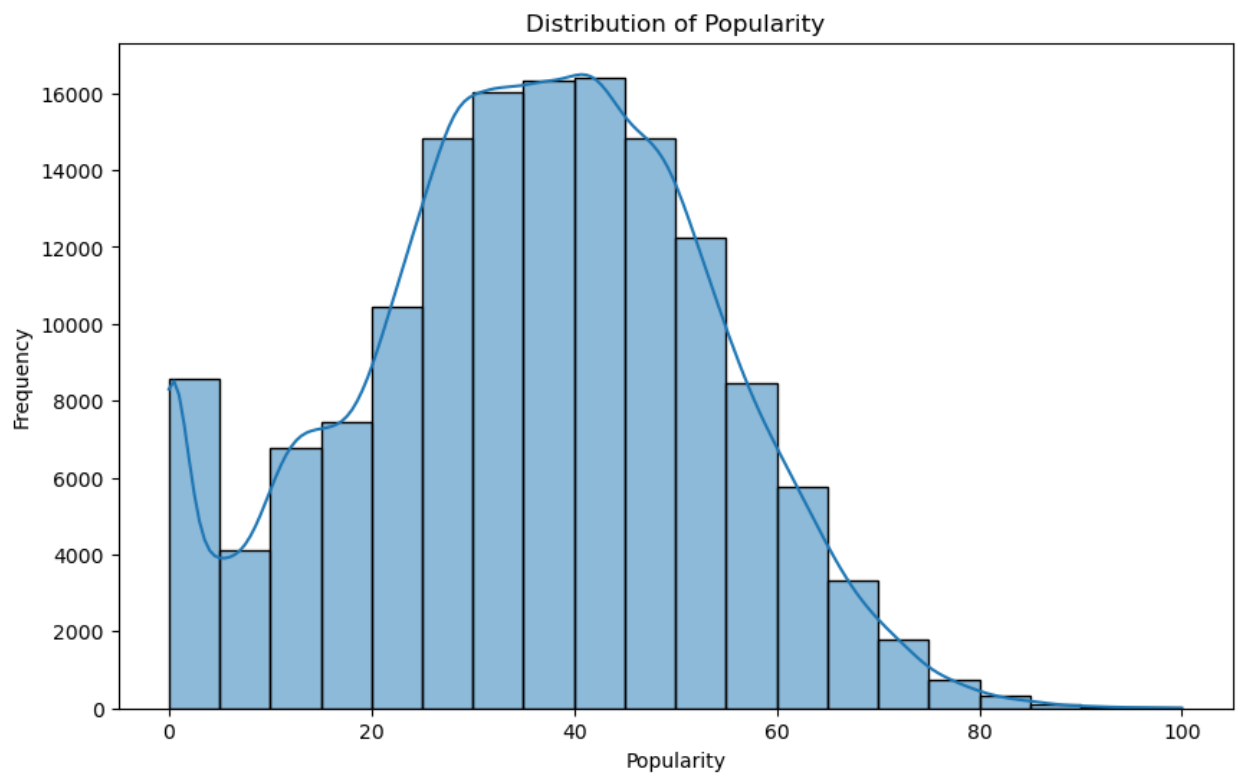
Conclusion and Insights

Summarize key findings from the analyses conducted

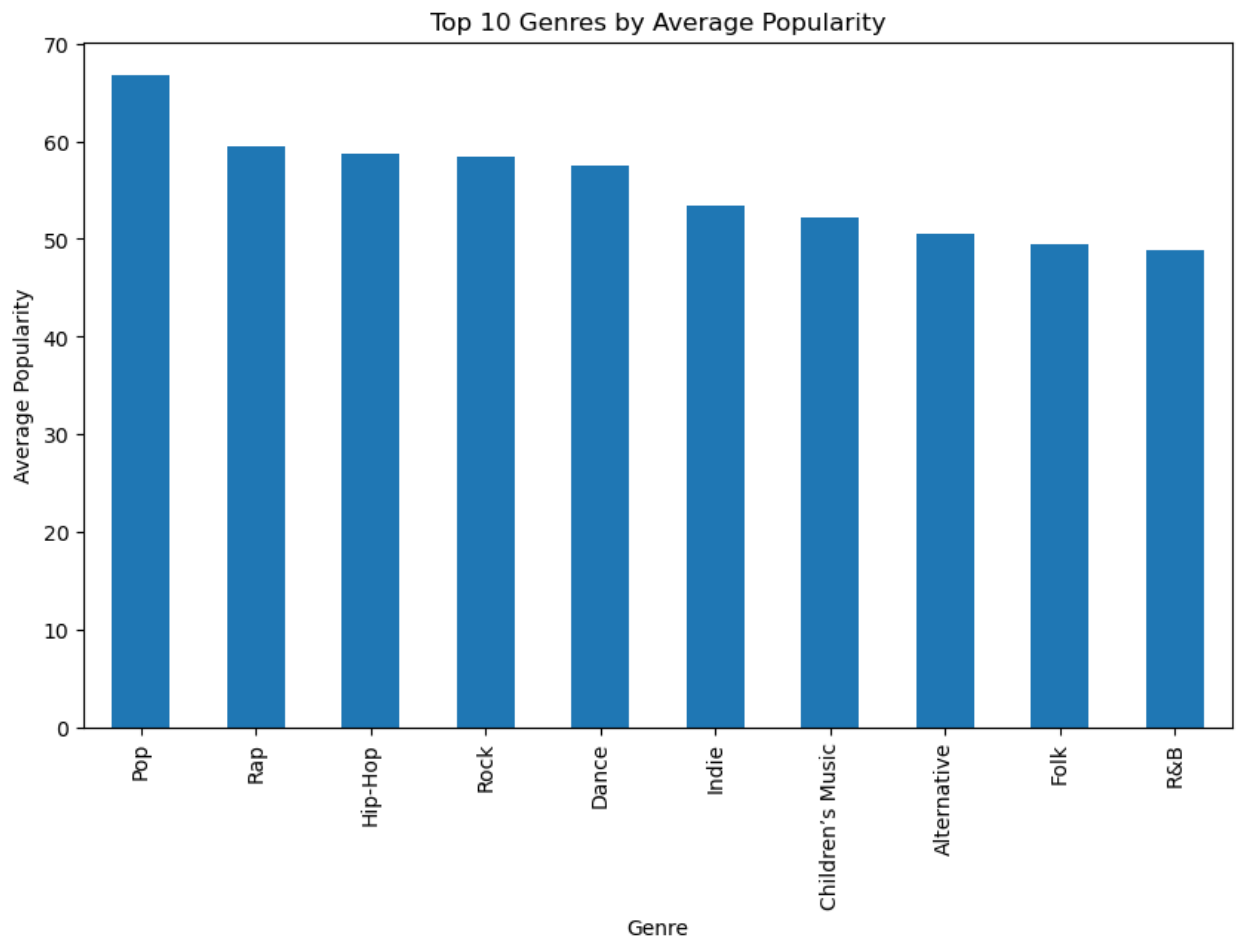
Highlight any significant insights or trends observed

Results:

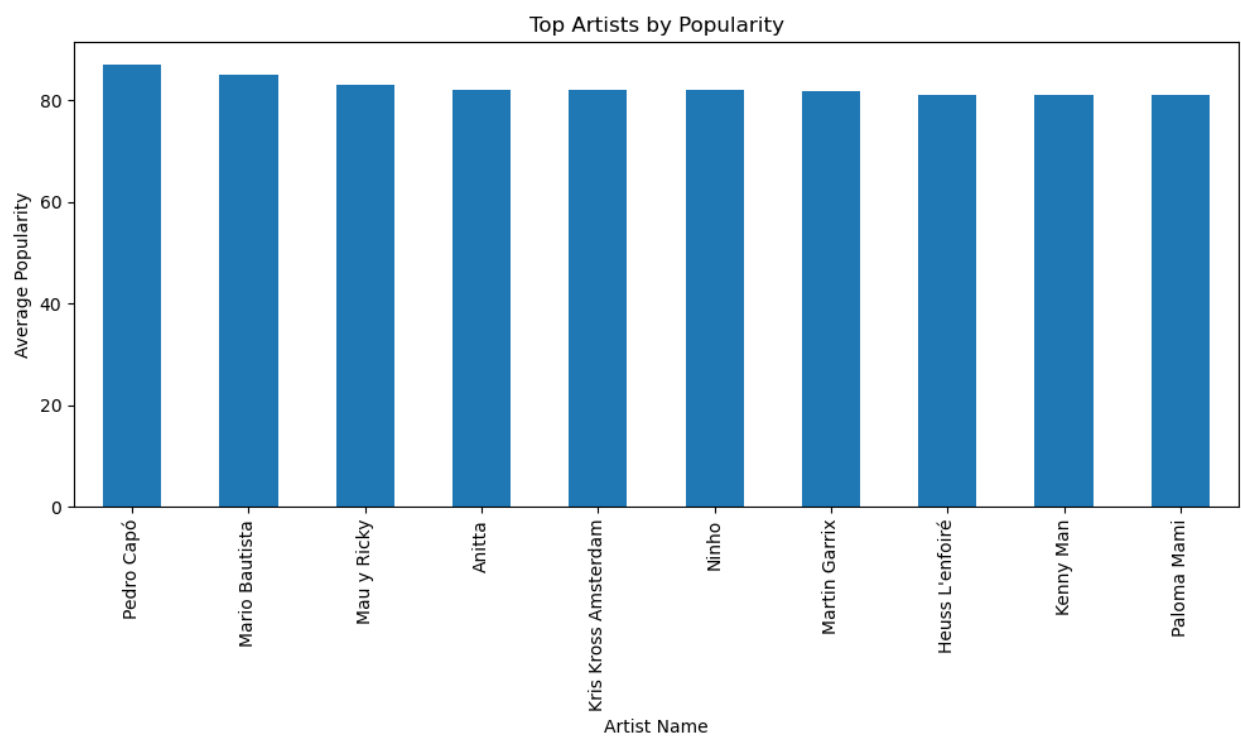
1. Distribution of Song Popularity:



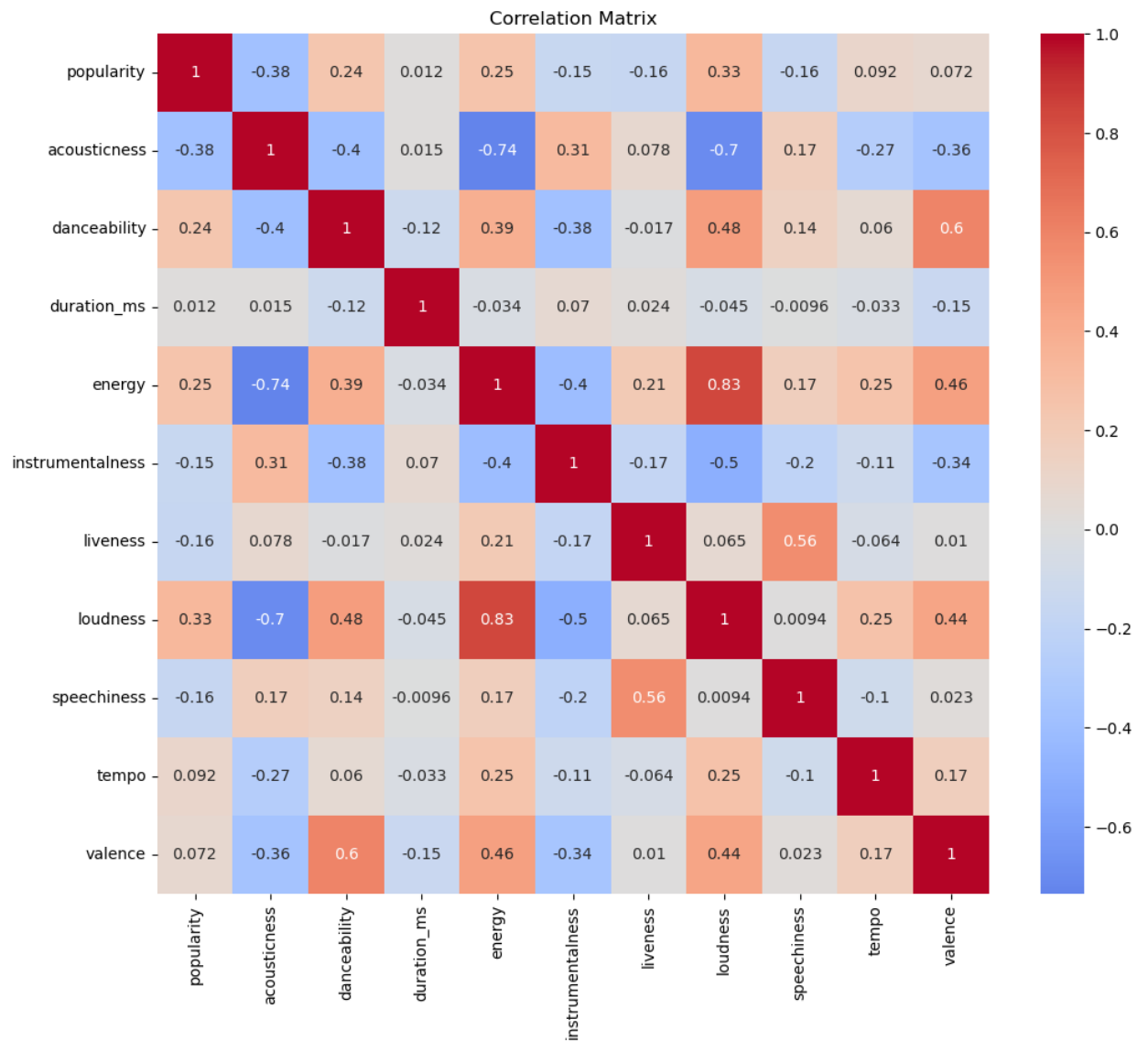
2. Average Popularity by Genre:



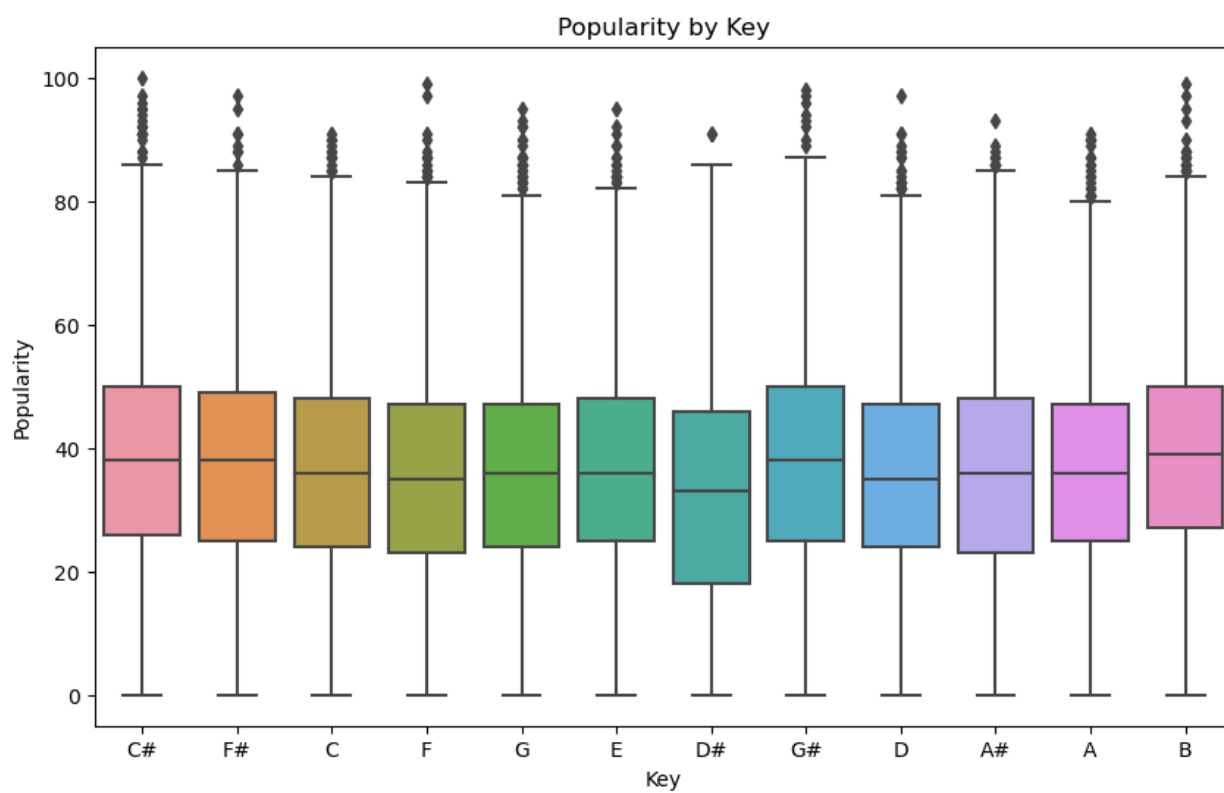
3. Top Artists by Popularity:



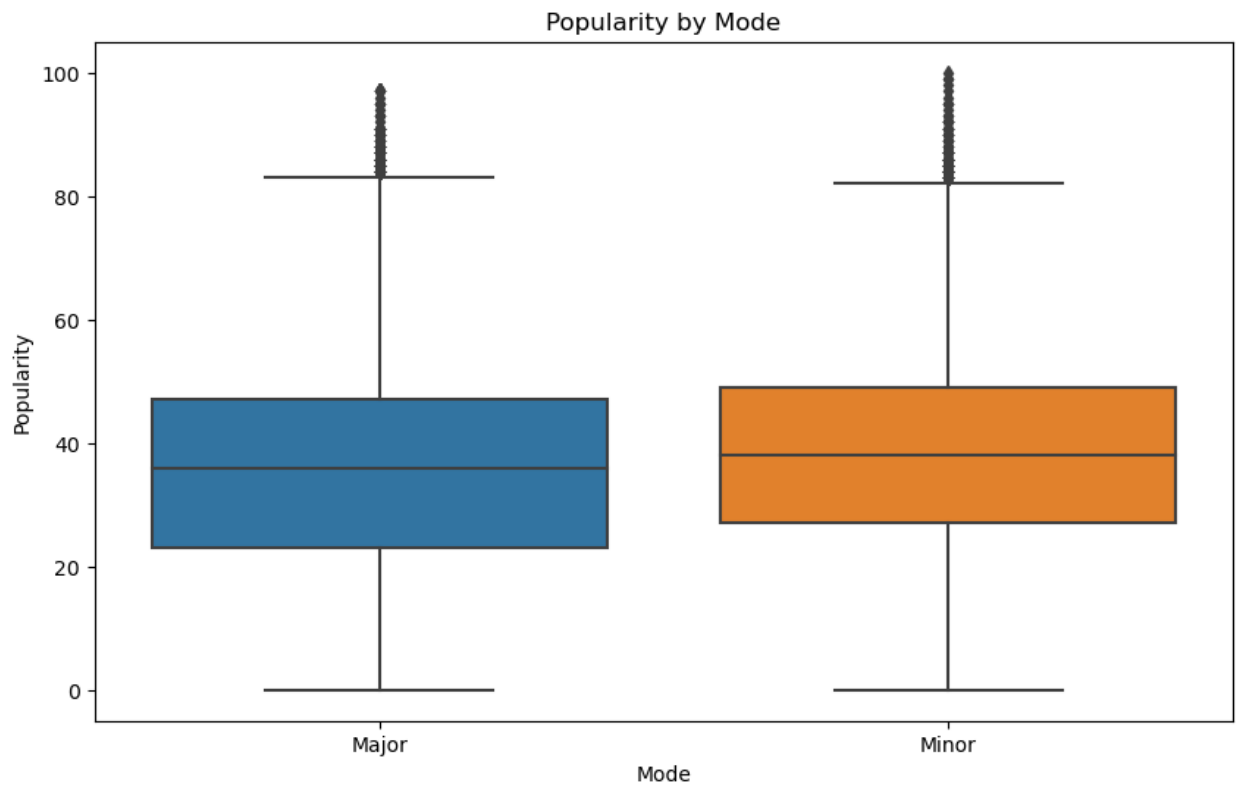
4. Correlation Matrix of Musical Attributes:



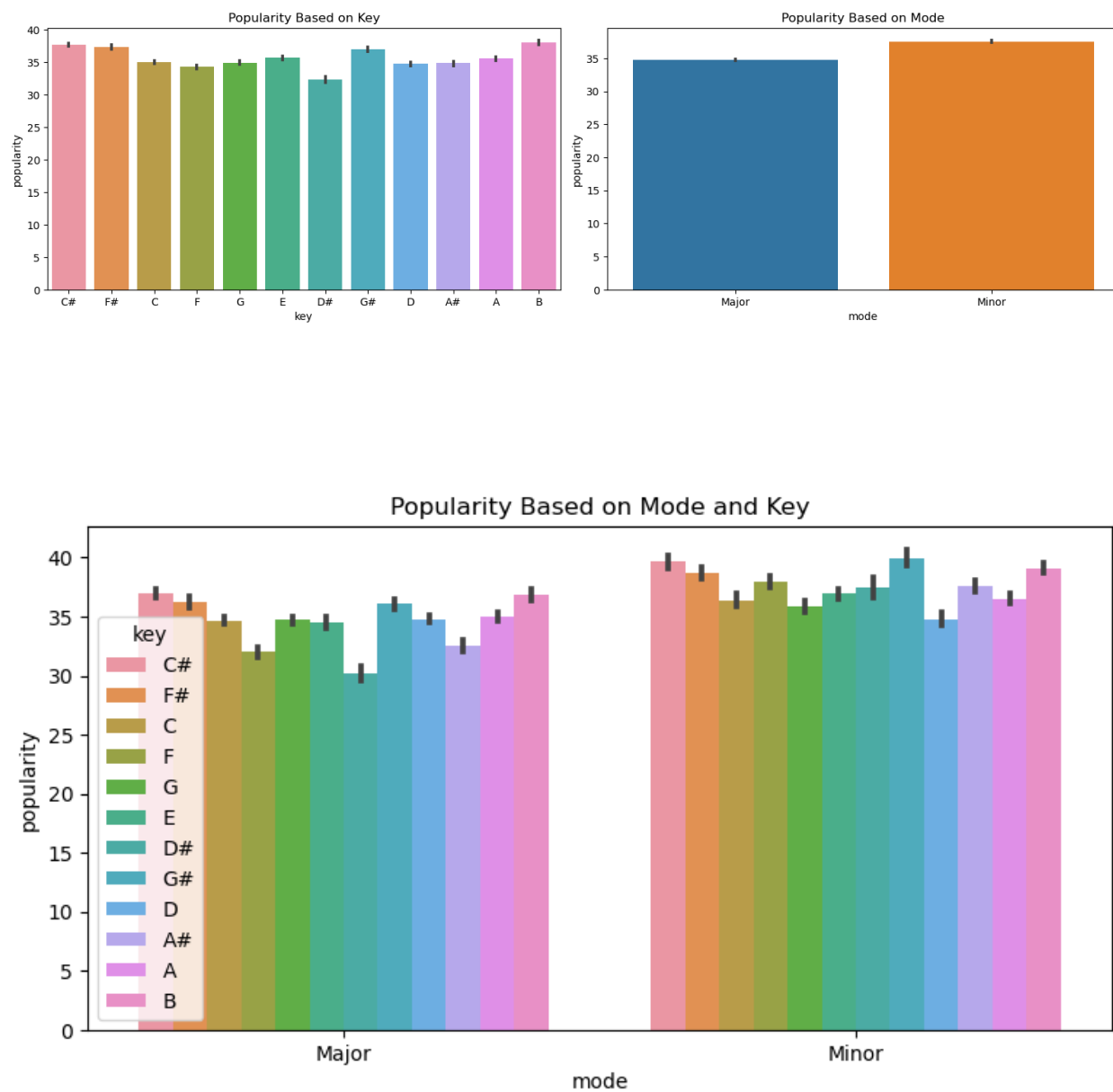
5. Popularity by Key:



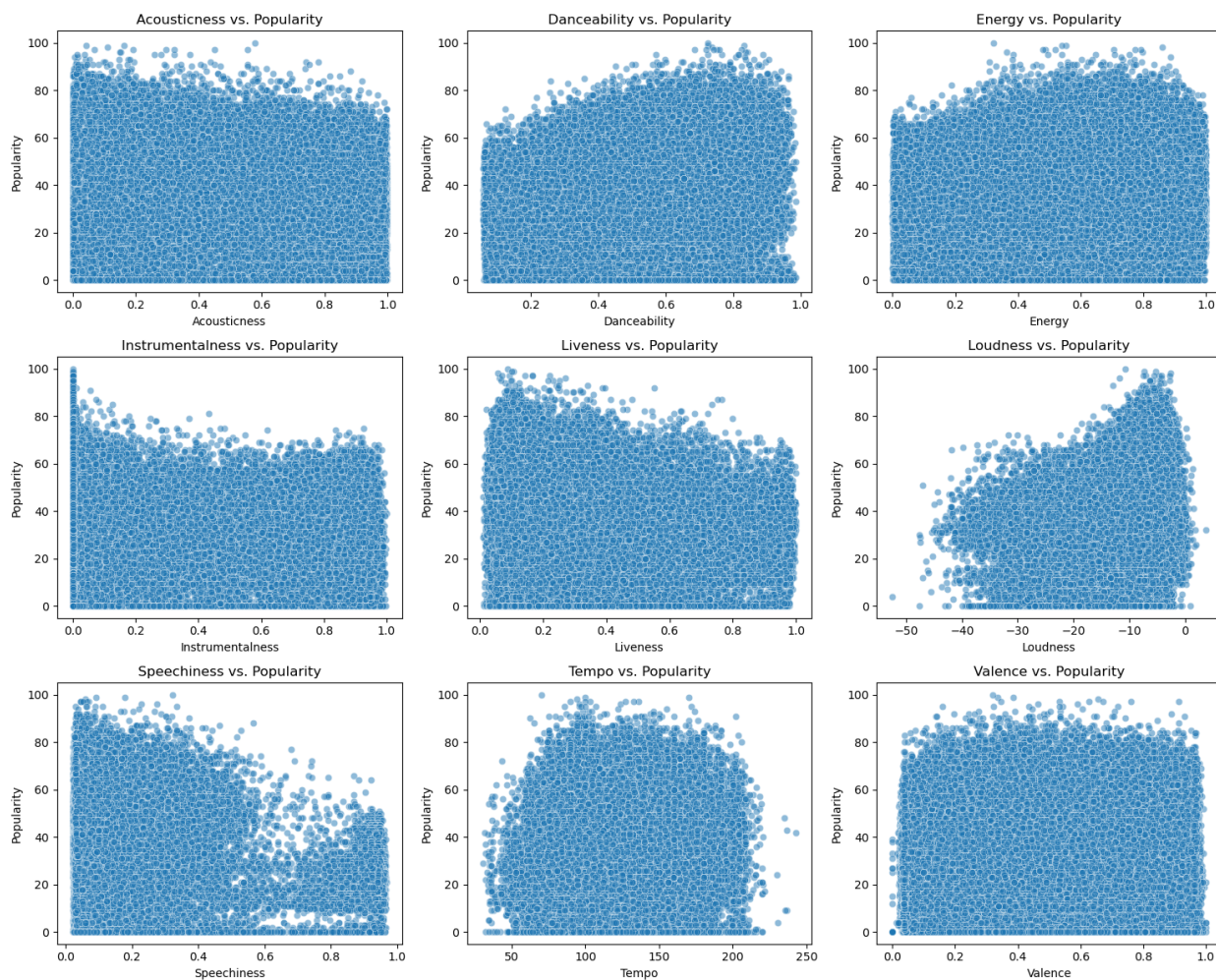
6. Popularity by Mode:



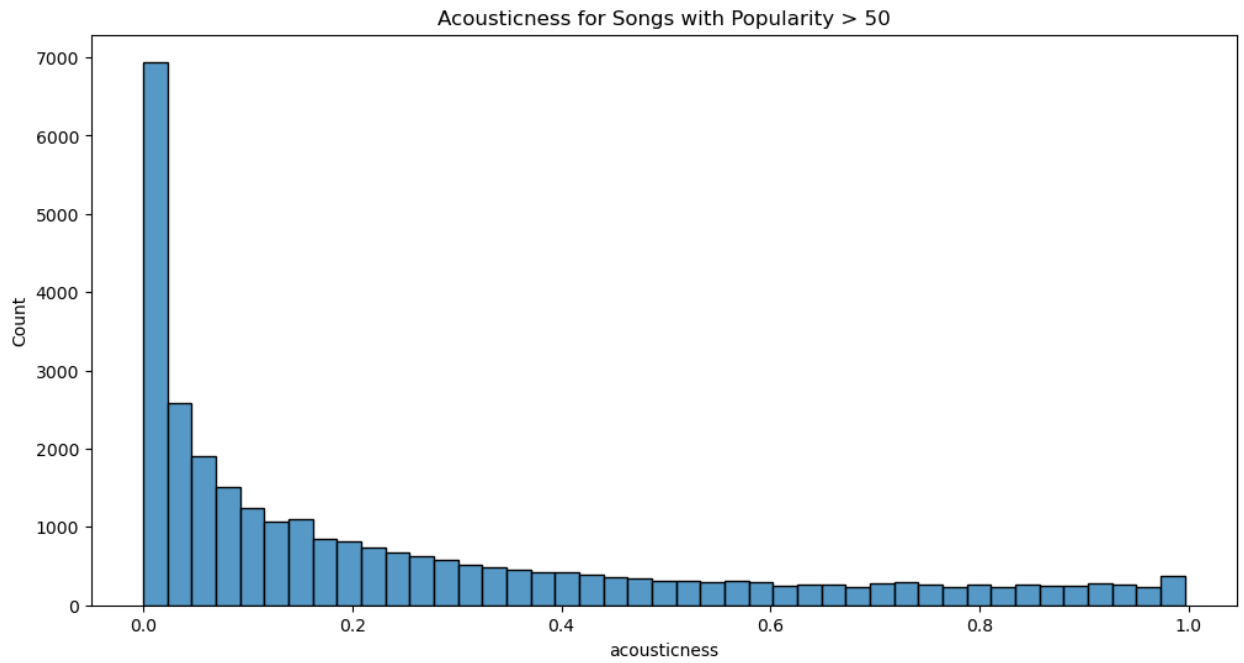
7. Popularity by Key and Mode:



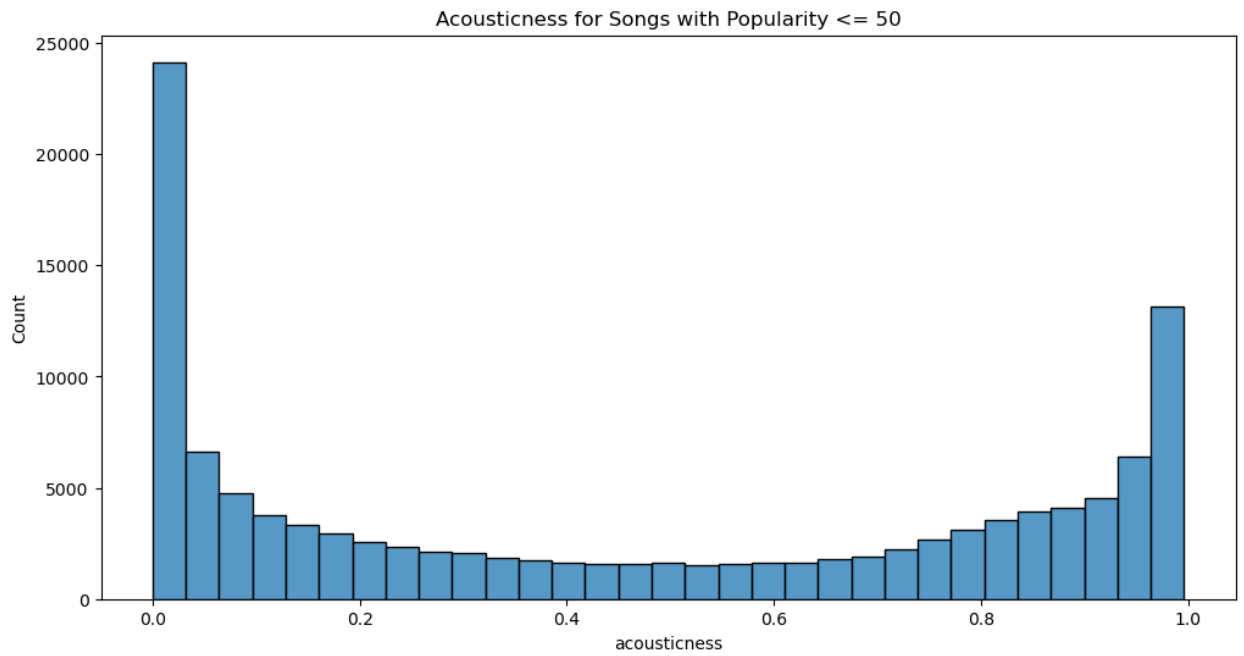
8. Scatter Plots of Attribute vs. Popularity:



9. Acousticness for Songs with Popularity > 50:



10. Acousticness for Songs with Popularity ≤ 50 :



Conclusion:

In this project, we conducted a comprehensive analysis of the "Spotify Attributes" dataset. Through data cleaning, exploratory data analysis, and visualizations, we aimed to gain insights into the trends and relationships between various factors affecting the popularity.

In conclusion, our Spotify Popularity Analysis delved into various aspects of song popularity on the platform, uncovering intriguing insights that shed light on the factors contributing to a song's success.

→ key findings:

1. **Popularity Statistics:** The distribution of popularity scores provides us with a comprehensive overview of the dataset. With a mean popularity of approximately 41 and a range between 0 and 100, we observed that the majority of songs fall within the 30 to 50 popularity range, as evident from the histogram visualization.

2. **Genre Impact on Popularity:** Exploring the average popularity by genre revealed interesting trends. While a wide range of genres exist, we focused on the top five most popular genres – Pop, Rap, Hip-Hop, Rock and . These genres stood out with notably high average popularity scores, indicating their prominence within the platform's user base.

3. **Artist Influence:** Our analysis identified the top three most popular artists, each boasting impressive average popularity scores – Pedro Capó, Mario Bautista, and Mau y Ricky. These artists have managed to captivate audiences, showcasing the correlation between artist reputation and song popularity.

4. **Standout Tracks:** Among the vast collection of tracks, a few songs have managed to attain exceptional popularity scores. For instance, Ariana Grande's '7 rings' and 'break up with your girlfriend, i'm bored' garnered high popularity due to their danceability, energy, and other musical attributes.

5. Understanding Musical Attributes: Our exploration of various musical attributes, such as danceability, energy, and valence, in relation to popularity revealed interesting patterns. These attributes play a crucial role in shaping a song's popularity, providing artists and producers with insights into crafting hit songs.

These findings underscore the multifaceted nature of song popularity on Spotify and its intricate interplay with genres, artists, and musical attributes.

□ Limitations and Future Work:

The analysis is based on available data and doesn't take into account external factors that could also play a role in popularity. Furthermore, while we've explored various aspects, there may be additional factors that warrant further investigation.

Moving forward, this analysis could serve as a foundation for more in-depth studies on the impact of specific attributes on popularity or predictive models for song success. Overall, the findings contribute to a better understanding of how musical elements contribute to the popularity of songs on the Spotify platform, providing valuable insights for artists, music enthusiasts, and the industry as a whole.

References:

<https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db>

-----THANK YOU-----