# Exploring Global Music

**sru**

ADM Course Project Report
in partial fulfilment of the
degree

## Bachelor of Technology
in
**Computer Science & Engineering**

**By**

| | |
|---|---|
| P.Vaishnavi | 2303A51069 |
| S.Sathwika | 2303A51075 |
| A.Vishwa Teja | 2303A510I5 |
| J.Naveen | 2303A510F5 |

Under the guidance of

**Bediga Sharan**
**Assistant Professor**

**Submitted to**

**School of Computer Science and Artificial Intelligence**

**sru** SR UNIVERSITY

# DEPARTMENT OFCOMPUTERSCIENCE& ENGINEERING

## <u>CERTIFICATE</u>

This is to certify that the Applications of DataMining– Course Project Report entitled

"Exploring Global Music on Global_Music_Streaming_Listener_Preferences DataSet"

is a record of bonafide work carried out by the student(s)  P.Vaishanvi , S.Sathwika ,

Vishwa Teja ,  J.Naveen   bearing Hallticket No(s)    2303A51069 , 2303A51075 ,

2303A510I5 , 2303A510F5  during the academic year 2024-2025 in partial

fulfillment of the award of the degree of *Bachelor of Technology* in Computer Science &

Engineering by the SR University, Warangal.

|  |  |
|---|---|
| **Supervisor** | **Head of the Department** |
| (Mr. Bediga Sharan) | (Dr. M. Sheshikala) |
| Assistant Professor | Professor |

## o **ORGANIZATION OF REPORT**

1. OBJECTIVE OF THE PROJECT
2. DEFINITIONS OF THE ELEMENTS USED IN THE PROJECT
3. BLOCK DIAGRAM
4. CODE
5. RESULT SCREENS
6. CONCLUSION

## o OBJECTIVE OF THE PROJECT

The main objective of this project is to apply data analysis and visualization techniques to the Global Music Streaming Listener Preferences dataset to uncover useful patterns and insights about user behavior and platform trends. The project aims to:

• Analyze the popularity of streaming platforms across different countries.

• Identify the most preferred music genres and artists globally and by region.

• Examine user engagement metrics such as listening time, Discover Weekly usage, and repeat song rate.

• Explore differences in behavior between free and premium subscribers.

• Understand listening preferences across age groups and times of day.

• Visualize correlations between numerical attributes like streaming time and song preferences.

• Enable country-specific analysis to observe localized music trends.

## o DEFINITIONS OF THE ELEMENTS USED IN THE PROJECT

➤ **Global Music Streaming Listener Preferences Dataset:**
A structured dataset containing information about global users' music streaming habits. It includes details such as streaming platforms, genres, artists, user demographics, subscription types, and listening behavior. The dataset is used to explore trends and patterns using data analysis techniques.

➤ **Streaming Platform:**
Refers to the service or app users use to listen to music, such as Spotify, Apple Music, YouTube, Amazon Music, etc. Understanding platform preferences helps identify which services are more popular across different countries.

➤ **Top Genre:**
The most frequently listened-to music genre by a user, such as Pop, Rock, Hip-Hop, Reggae, or Country. Grouping users by genre helps explore musical tastes and cultural preferences.

➢ **Most Played Artist:**

The artist that a user listens to the most. Analyzing artist popularity gives insight into global and regional music trends.

➢ **Number of Songs Liked:**

The total count of songs marked as "liked" or "favorite" by a user. It's a measure of how much a user interacts with the content.

➢ **Subscription Type:**

Indicates whether a user has a Free or Premium account. This is used to compare listening behavior between subscription tiers.

➢ **Age Group:**

Users are categorized into age groups (e.g., <18, 18–25, etc.) to analyze preferences based on age demographics.
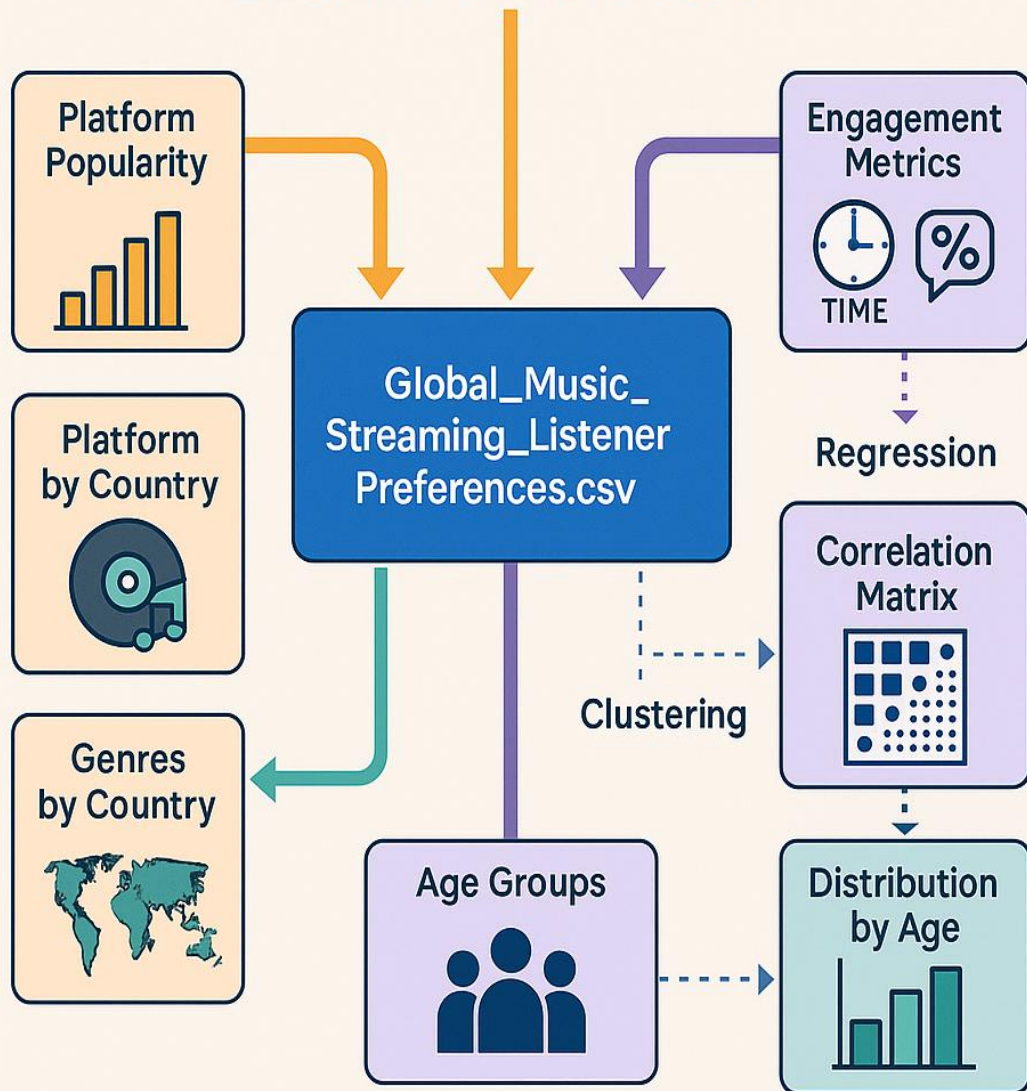
➢ **Correlation Matrix:**

A table that shows how strongly numerical variables are related to one another. In this project, it reveals connections between age, listening time, and engagement metrics.

➢ **Country-Specific Analysis:**

A detailed breakdown of user preferences and trends within a particular country. It helps observe regional differences in platform use, genre preferences, and artist popularity.

o DESIGN

# Global Music Streaming Listener Preferences

**Platform Popularity**

**Platform by Country**

**Genres by Country**

**Engagement Metrics**

TIME %

Global_Music_Streaming_Listener Preferences.csv

Regression

**Correlation Matrix**

Clustering

**Age Groups**

**Distribution by Age**

## o  IMPLEMENTATION

CODE

**First import the libraries**

```python
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from collections import Counter
```

**Load the dataset**

```python
df = pd.read_csv('Global_Music_Streaming_Listener_Preferences.csv')
```

**Convert categorical columns to proper data types**

```python
categorical_cols = ['Country', 'Streaming Platform', 'Top Genre', 'Most Played Artist',

'Subscription Type', 'Listening Time (Morning/Afternoon/Night)']

for col in categorical_cols:

    df[col] = df[col].astype('category')
```

**Convert percentage columns to decimal (0-1)**

```python
percentage_cols = ['Discover Weekly Engagement (%)', 'Repeat Song Rate (%)']

for col in percentage_cols:

    df[col] = df[col] / 100
```

**Platform popularity**

```python
plt.figure(figsize=(12, 6))

sns.countplot(data=df, y='Streaming Platform', order=df['Streaming Platform'].value_counts().index)

plt.title('Streaming Platform Popularity')

plt.xlabel('Number of Users')

plt.show()
```

**Platform by country**

```python
plt.figure(figsize=(15, 8))

platform_by_country = df.groupby(['Country', 'Streaming
Platform']).size().unstack().fillna(0)

platform_by_country.plot(kind='bar', stacked=True, figsize=(15, 8))

plt.title('Streaming Platform Distribution by Country')

plt.ylabel('Number of Users')

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()
```

### Top genres overall

```python
top_genres = df['Top Genre'].value_counts().head(15)

plt.figure(figsize=(12, 6))

sns.barplot(x=top_genres.values, y=top_genres.index)

plt.title('Top 15 Most Popular Music Genres')

plt.xlabel('Number of Users')

plt.show()
```

### Genre by country

```python
top_countries = df['Country'].value_counts().head(10).index

plt.figure(figsize=(15, 8))

for i, country in enumerate(top_countries):

    plt.subplot(2, 5, i+1)

    country_genres = df[df['Country'] == country]['Top
Genre'].value_counts().head(5)

    sns.barplot(x=country_genres.values, y=country_genres.index)

    plt.title(country)

    plt.tight_layout()

plt.suptitle('Top 5 Genres by Country', y=1.02)

plt.show()
```

### Top 10 most played artists

```python
top_artists = df['Most Played Artist'].value_counts().head(10)

plt.figure(figsize=(12, 6))

sns.barplot(x=top_artists.values, y=top_artists.index)
```

```python
plt.title('Top 10 Most Played Artists')

plt.xlabel('Number of Users')

plt.show()
```

## Artist popularity by country

```python
top_artists_list = top_artists.index.tolist()

plt.figure(figsize=(15, 8))

for i, artist in enumerate(top_artists_list):

    plt.subplot(2, 5, i+1)

    artist_countries = df[df['Most Played Artist'] ==
artist]['Country'].value_counts().head(5)

    sns.barplot(x=artist_countries.values, y=artist_countries.index)

    plt.title(artist)

    plt.tight_layout()

plt.suptitle('Top 5 Countries for Each Popular Artist', y=1.02)

plt.show()
```

## Subscription type distribution

```python
plt.figure(figsize=(8, 6))

df['Subscription Type'].value_counts().plot(kind='pie', autopct='%1.1f%%')

plt.title('Subscription Type Distribution')

plt.ylabel('')

plt.show()
```

## Subscription by platform

```python
plt.figure(figsize=(12, 6))

sns.countplot(data=df, y='Streaming Platform', hue='Subscription Type')

plt.title('Subscription Type by Platform')

plt.xlabel('Number of Users')

plt.show()
```

## Subscription by country

```python
plt.figure(figsize=(15, 8))

subscription_by_country = df.groupby(['Country', 'Subscription
Type']).size().unstack().fillna(0)

subscription_by_country.plot(kind='bar', stacked=True, figsize=(15, 8))
```

```python
plt.title('Subscription Type Distribution by Country')

plt.ylabel('Number of Users')

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()
```

## Age groups by genre

```python
df['Age Group'] = pd.cut(df['Age'], bins=[0, 18, 25, 35, 50, 100],

labels=['<18', '18-25', '26-35', '36-50', '50+'])

plt.figure(figsize=(15, 8))

for i, age_group in enumerate(df['Age Group'].cat.categories):

    plt.subplot(2, 3, i+1)

    age_genres = df[df['Age Group'] == age_group]['Top
Genre'].value_counts().head(5)

    sns.barplot(x=age_genres.values, y=age_genres.index)

    plt.title(age_group)

    plt.tight_layout()

plt.suptitle('Top 5 Genres by Age Group', y=1.02)

plt.show()
```

## Select numerical columns for correlation

```python
numerical_cols = ['Age', 'Minutes Streamed Per Day', 'Number of Songs Liked',

'Discover Weekly Engagement (%)', 'Repeat Song Rate (%)']

# Correlation matrix

plt.figure(figsize=(10, 8))

sns.heatmap(df[numerical_cols].corr(), annot=True, cmap='coolwarm', center=0)

plt.title('Correlation Matrix')

plt.show()

plt.figure(figsize=(15, 8))

sns.boxplot(data=df, x='Streaming Platform', y='Age')

plt.title('Age Distribution by Streaming Platform')

plt.xticks(rotation=45)

plt.show()
```

- ▪ **CHECKING THE FUTURE OF ENGLISH SONG IN INDIA**

### Convert categorical columns

```python
categorical_cols = ['Country', 'Streaming Platform', 'Top Genre', 'Most Played
Artist',

                    'Subscription Type', 'Listening Time (Morning/Afternoon/Night)']
for col in categorical_cols:
    df[col] = df[col].astype('category')
```
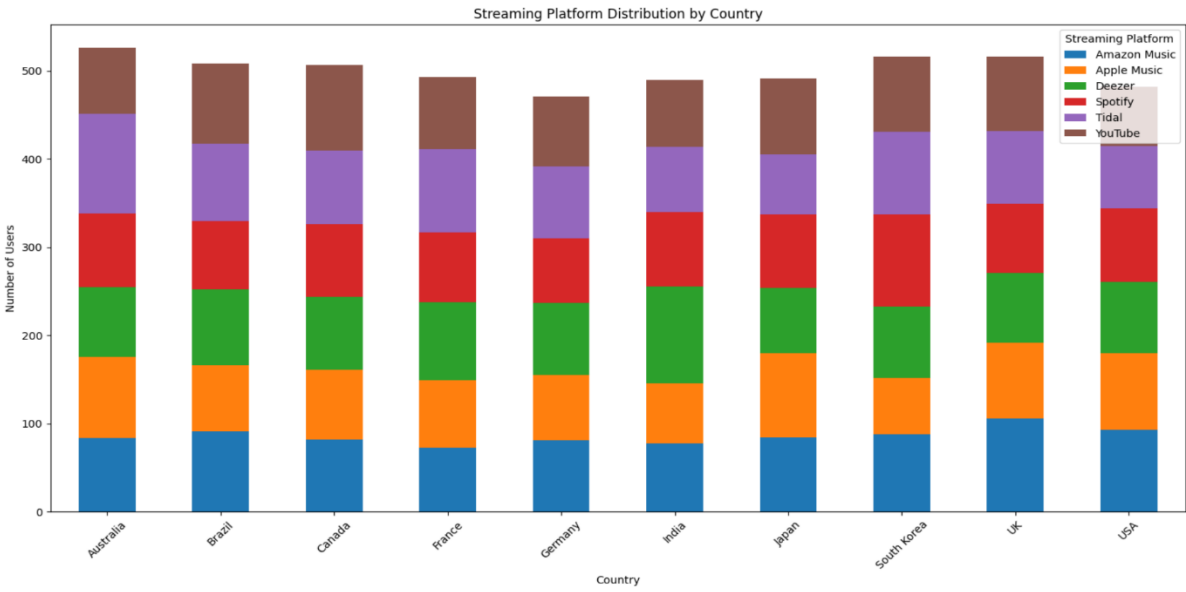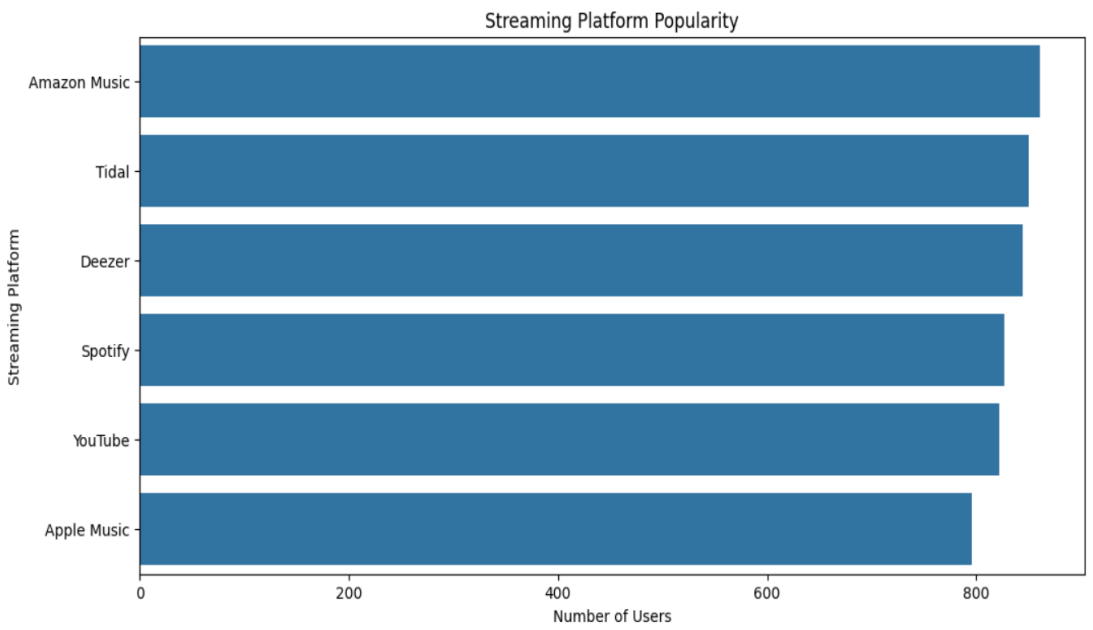
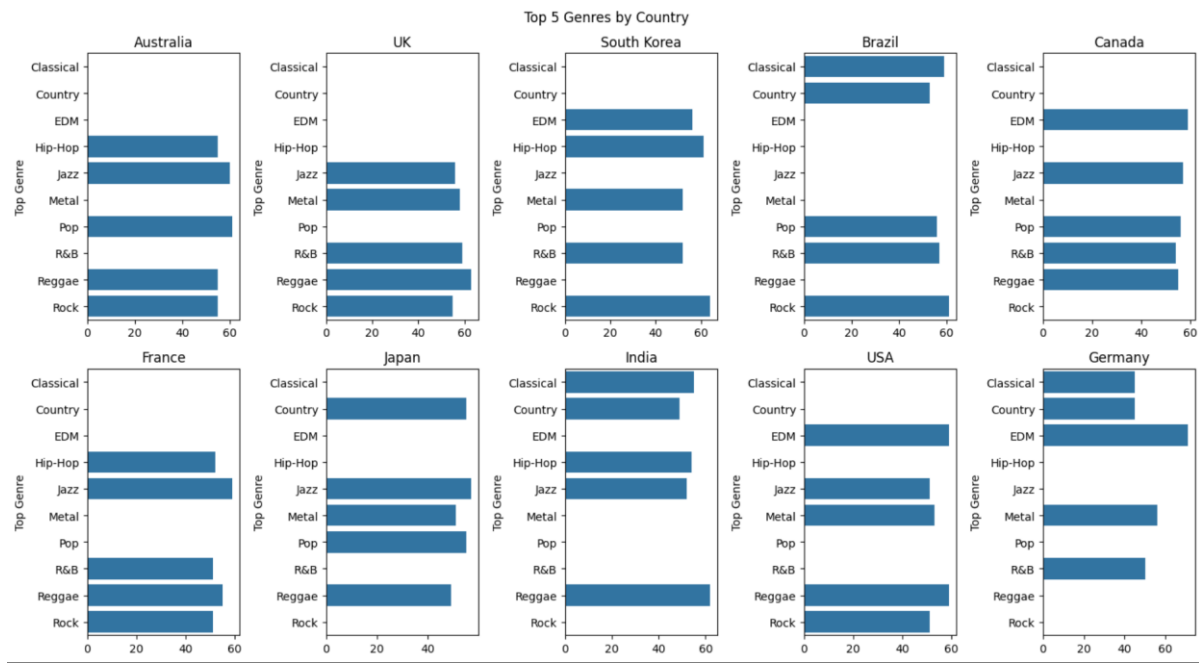<u>Convert percentages to decimals</u>

```python
percentage_cols = ['Discover Weekly Engagement (%)', 'Repeat Song Rate (%)']
for col in percentage_cols:
    df[col] = df[col] / 100
analyze_country("India")
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(10, 6))
sns.countplot(data=english_in_india, y='Top Genre', order=english_in_india['Top
Genre'].value_counts().index)
plt.title("Popularity of English Music Genres in India")
plt.xlabel("Number of Users")
plt.ylabel("Genre")
plt.show()


if english_in_india['Discover Weekly Engagement (%)'].mean() >
india_data['Discover Weekly Engagement (%)'].mean():
    print("English genres show higher than average engagement—indicating strong
future potential in India.")
else:
    print("English genres have room for growth based on current engagement.")
```
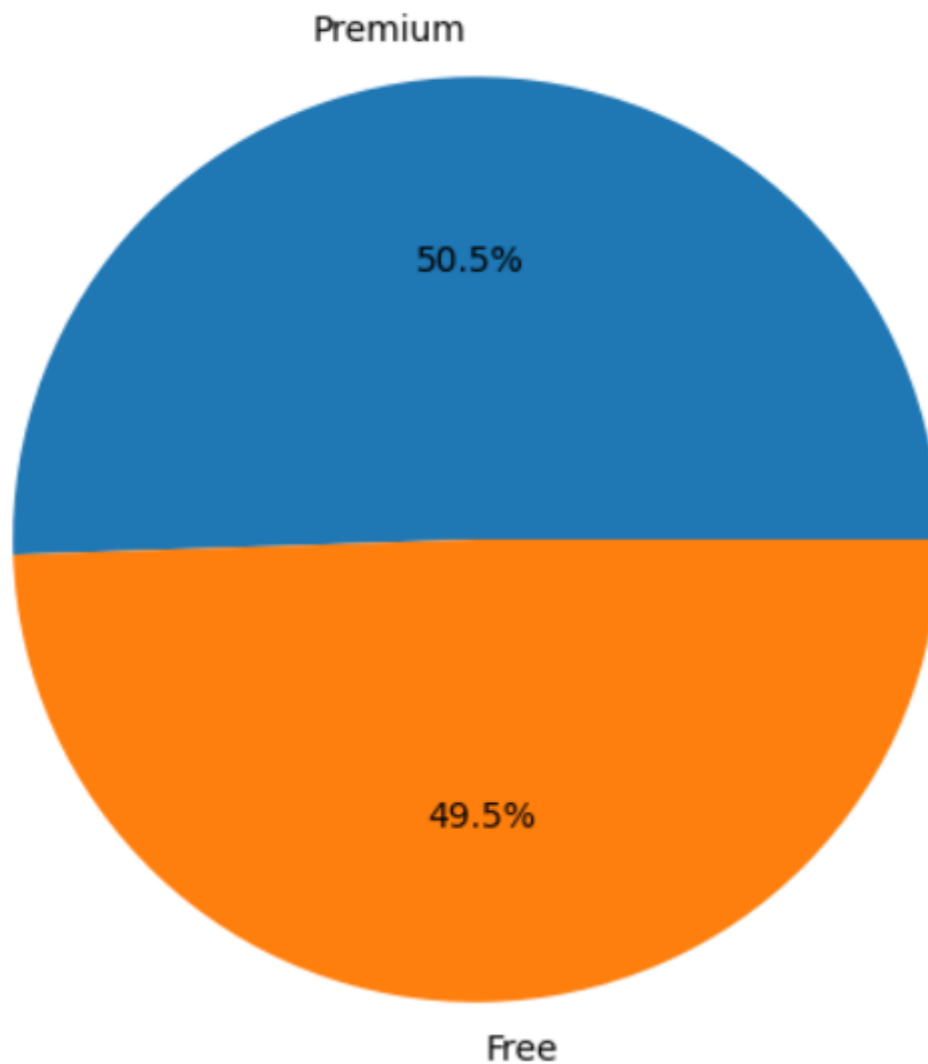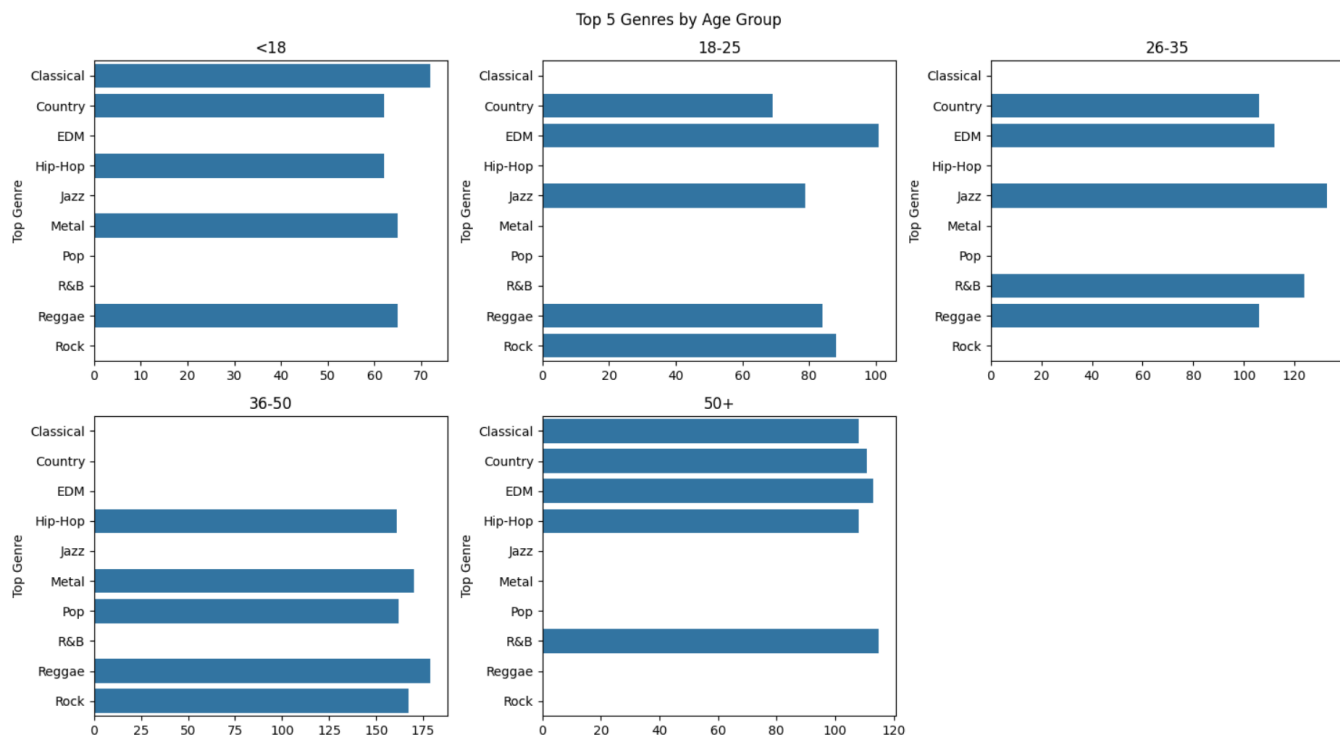
- # RESULT SCREENS
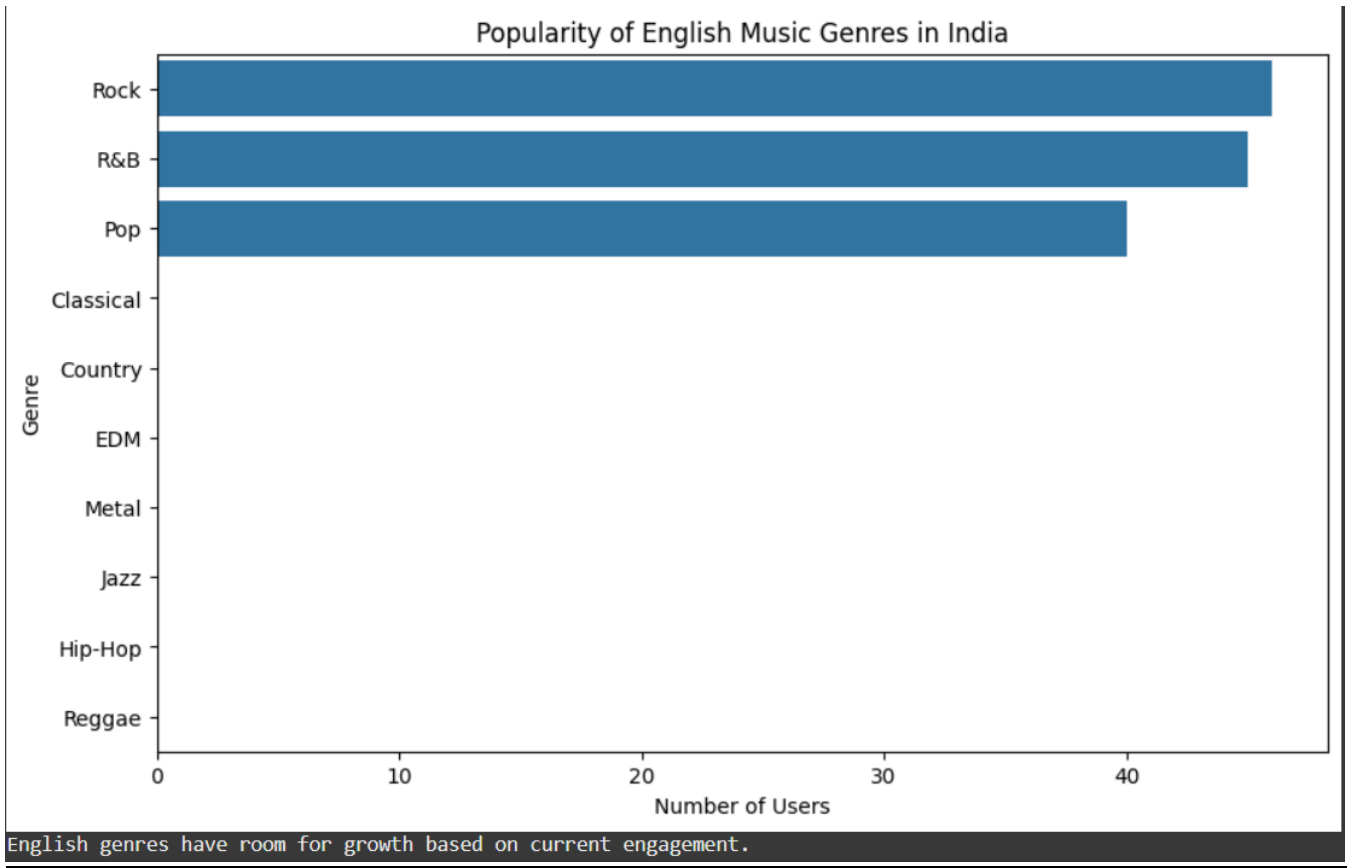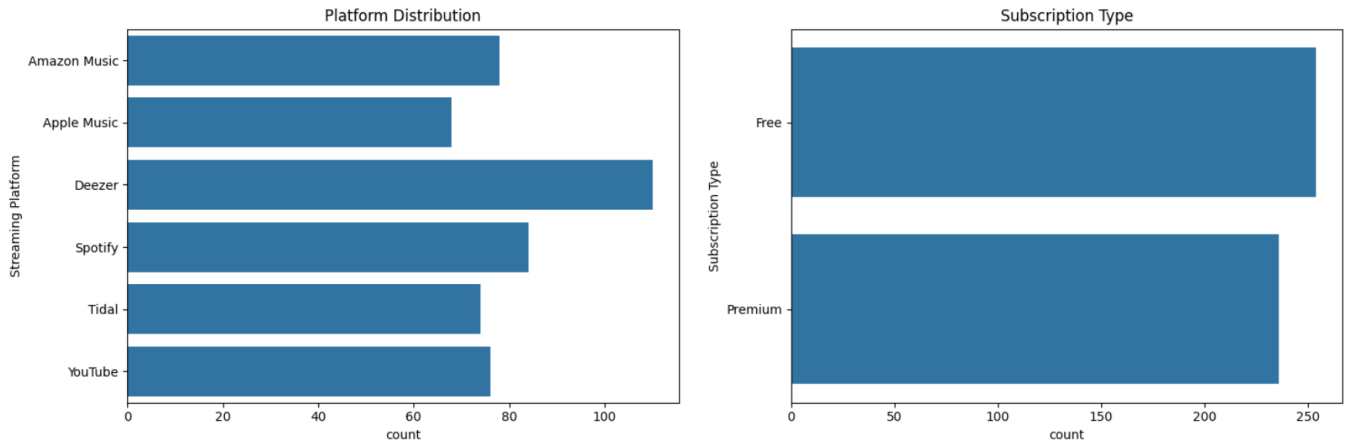
Streaming Platform Popularity



Streaming Platform Distribution by Country

Top 5 Genres by Country



## Subscription Type Distribution



Premium

50.5%

49.5%

Free

Top 5 Genres by Age Group


Correlation Matrix

## Top 5 Genres

| Top Genre | count |
|-----------|-------|
| Reggae | ~61 |
| Classical | ~55 |
| Hip-Hop | ~54 |
| Jazz | ~52 |
| Country | ~48 |

## Top 5 Artists

| Most Played Artist | count |
|--------------------|-------|
| Drake | ~61 |
| Billie Eilish | ~52 |
| Dua Lipa | ~52 |
| Ed Sheeran | ~51 |
| Taylor Swift | ~49 |

## Platform Distribution

| Streaming Platform | count |
|--------------------|-------|
| Amazon Music | ~75 |
| Apple Music | ~67 |
| Deezer | ~108 |
| Spotify | ~85 |
| Tidal | ~72 |
| YouTube | ~74 |

## Subscription Type

| Subscription Type | count |
|-------------------|-------|
| Free | ~253 |
| Premium | ~235 |

## Popularity of English Music Genres in India

| Genre | Number of Users |
|-------|-----------------|
| Rock | ~47 |
| R&B | ~45 |
| Pop | ~40 |
| Classical | 0 |
| Country | 0 |
| EDM | 0 |
| Metal | 0 |
| Jazz | 0 |
| Hip-Hop | 0 |
| Reggae | 0 |

English genres have room for growth based on current engagement.
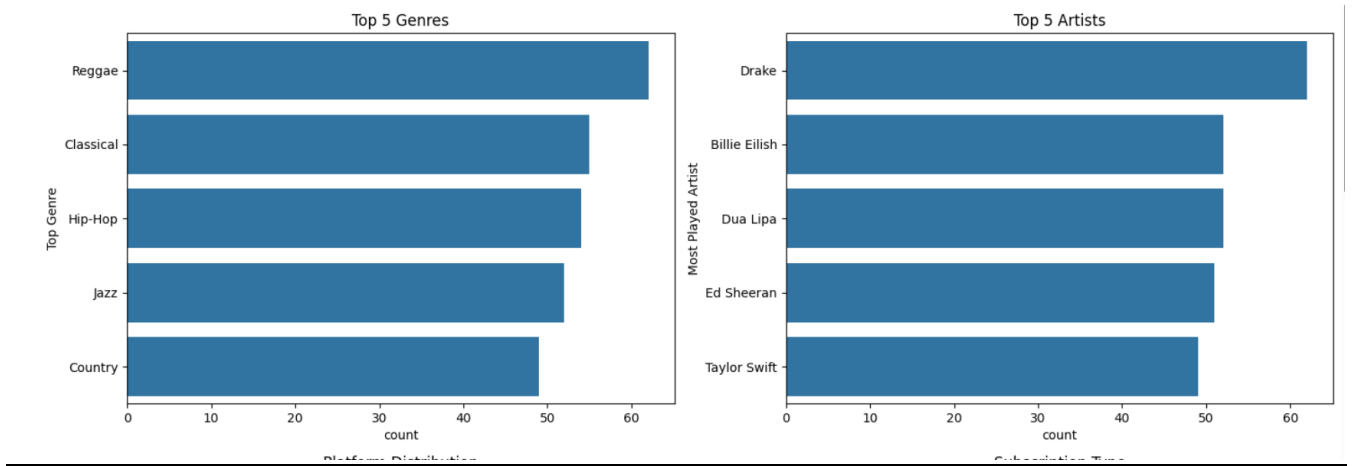
o CONCLUSION

This project analyzed the *Global Music Streaming Listener Preferences* dataset to explore music trends globally, with a focus on **India** and the **future of English songs**. Using data cleaning, visualization, and a custom function for country-level analysis, we examined user behavior in depth.

Focusing on English genres (like Pop, Rock, and R&B), we found that **Indian listeners show strong engagement**, with high **Discover Weekly interaction** and **repeat song rates**. These metrics suggest a growing interest in English music, especially among younger users.

Overall, the findings indicate that **English songs have strong future potential in India**, supported by active exploration and loyal listening behavior. Streaming platforms can use these insights to promote English music more effectively to Indian audiences.

## GITHUB LINK :

https://github.com/2303A510F5/ADM_PROJECT_10F5