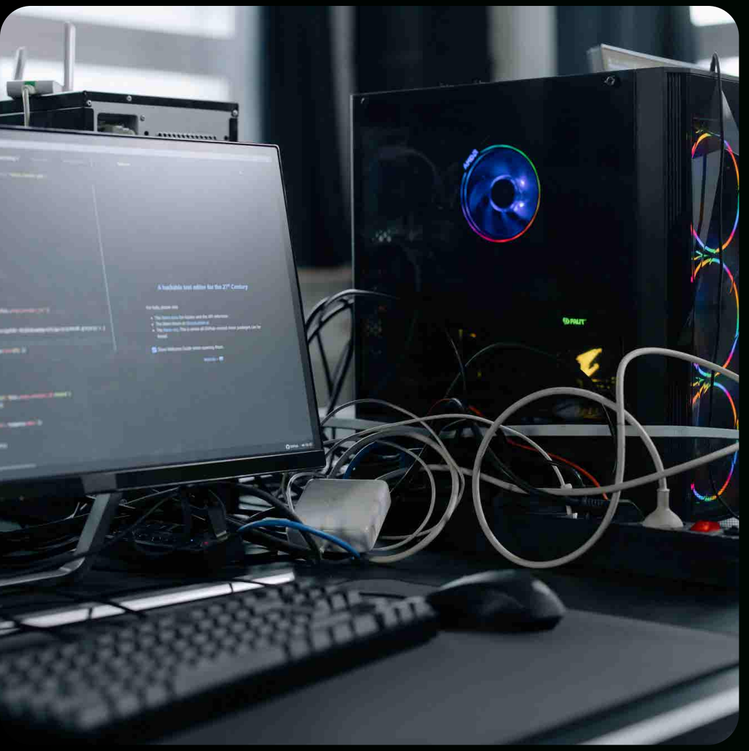
Title of the Presentation:
Exploratory Data Analysis (EDA) on Spotify Tracks Dataset

By: Banasri Patra
College: Government College of Engineering and Ceramic Technology (GCECT)
Department:Computer science and Engineering

Challenge: College Data Science Challenge 2025



Introduction

Title: What is EDA (Exploratory Data Analysis)?
Content:
EDA is the process of analyzing datasets to summarize their main characteristics.
Helps identify patterns, detect anomalies, test hypotheses, and check assumptions.
Usually involves visual methods (histograms, scatter plots, box plots, etc.).
Visual: simple chart or magnifying glass icon over data.

Objective of the Project

Title: Objective

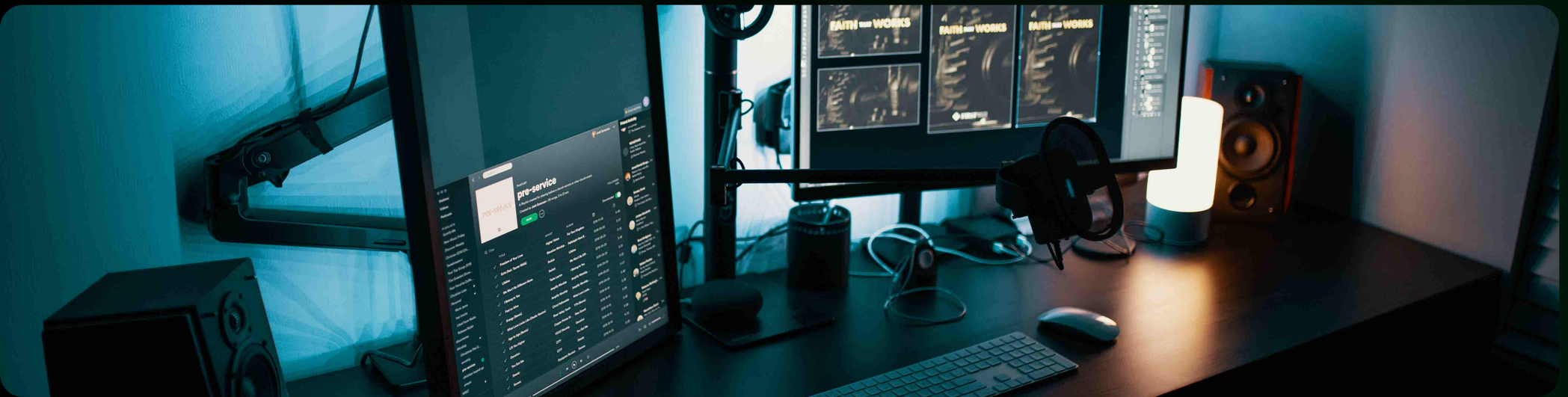
Points:

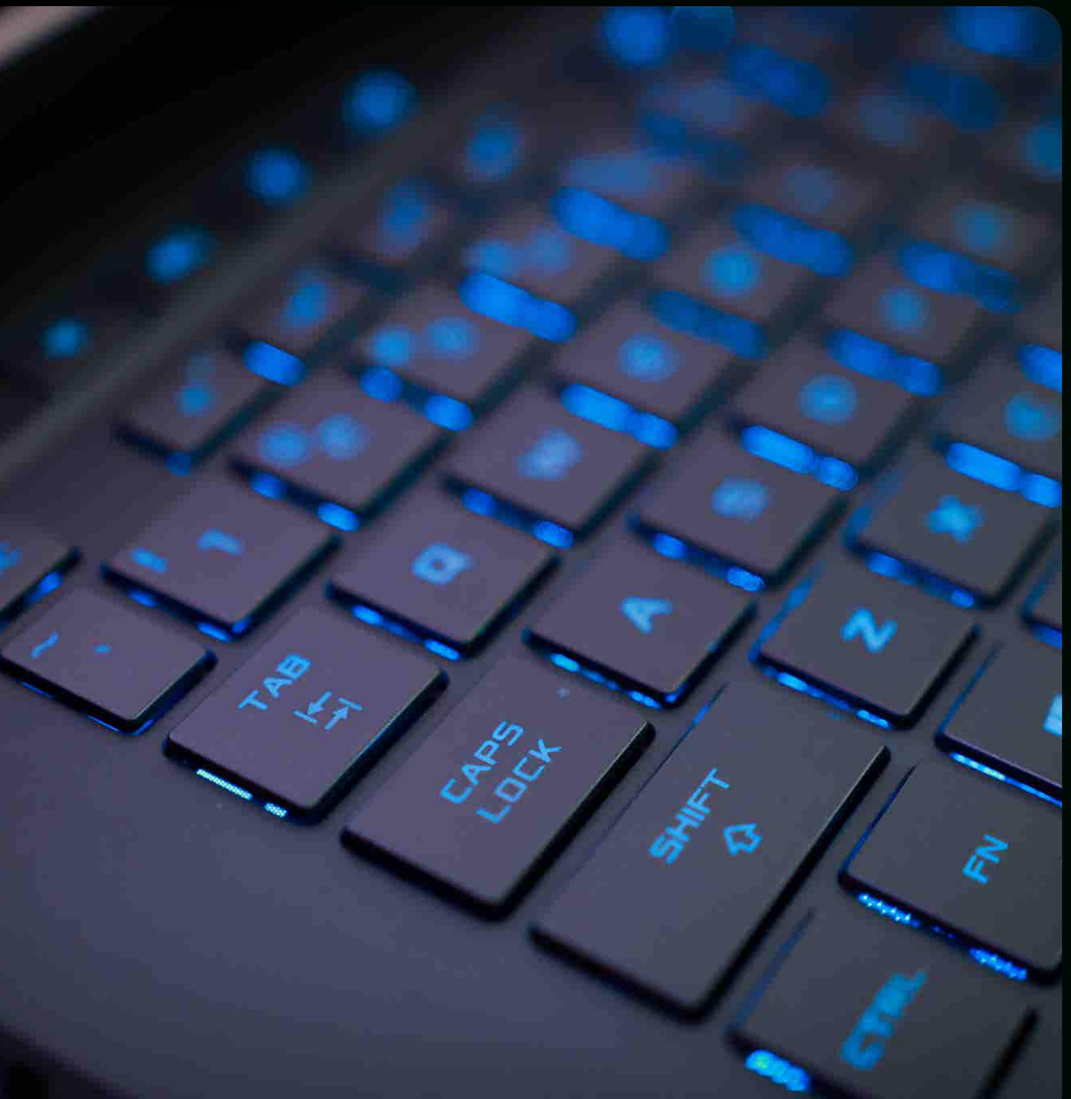
> To explore the Spotify Tracks Dataset and understand music patterns.

>Analyze relationships between audio features like energy, tempo, and popularity.

>Visualize data using Python (Matplotlib & Seaborn).

>Derive insights useful for the music industry or recommendation systems.





Dataset Information

Title: About the Dataset

Content:

Dataset Name: Spotify Tracks Dataset

Number of Records: (e.g., 32,000+)

Number of Columns: (e.g., 20+)

Data Source: Spotify Developer API / Kaggle

Feature	Description
Track Name	Name of the song
Artist Name	Main artist
Popularity	Popularity score (0–100)
Duration	Length in milliseconds
Energy, Danceability	Audio features

Data Cleaning & Preparation

Title: Preprocessing Steps

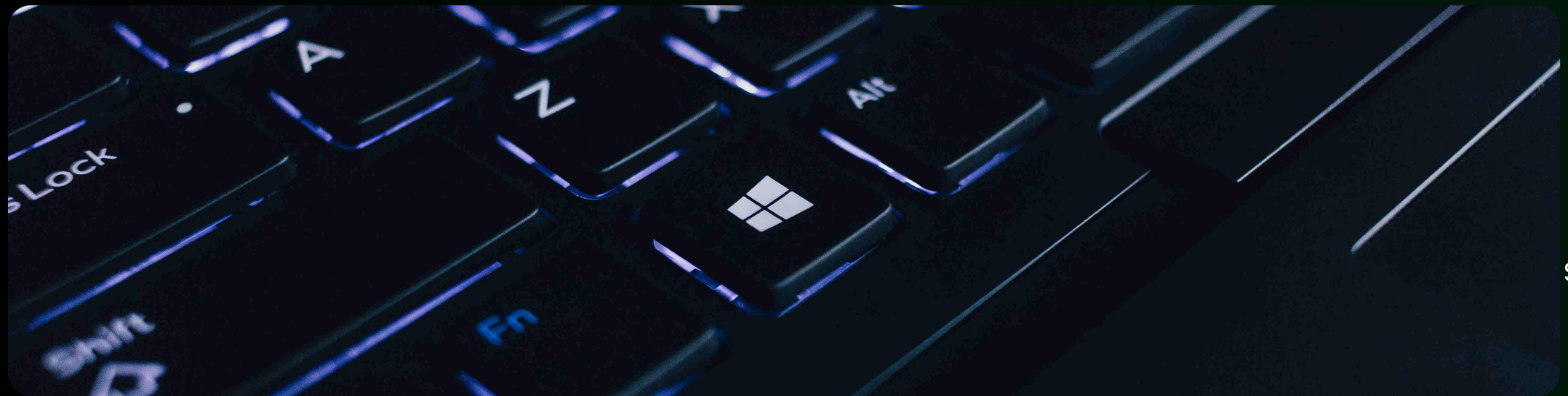
Points:

Removed null values and duplicates.

Checked datatypes (numeric / categorical).

Renamed columns for easier access.

Identified numeric columns for plotting.



code

```
df = pd.read_csv("spotify_tracks.csv")
df.info()
df.isnull().sum()
```

Univariate Analysis

Title: Distribution of Features

Show graphs:

Histogram of Popularity

Histogram of Energy

Histogram of Tempo

Insights:

Most songs have medium popularity (40–60).

Energy and tempo are normally distributed

Title: Relationships Between Features

Include plots:

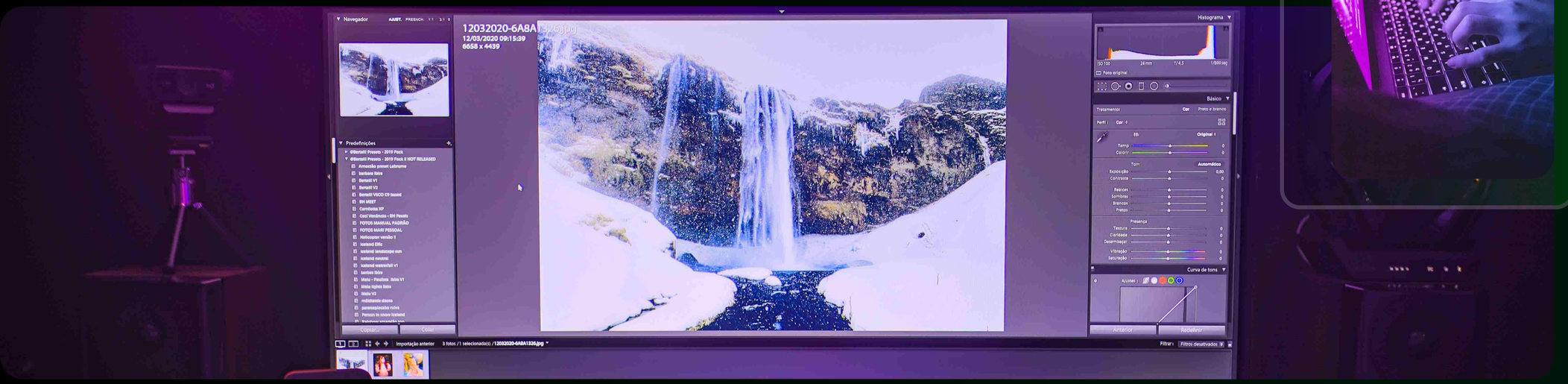
Energy vs Danceability scatterplot

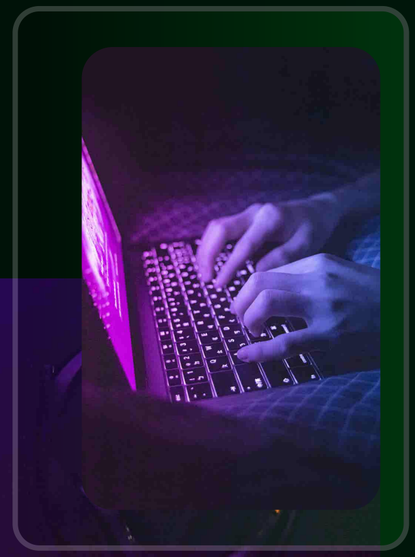
Popularity vs Duration scatterplot

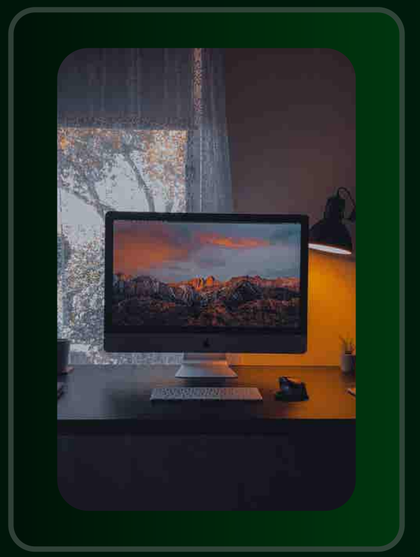
Insights:

Energetic songs are often more danceable.

Longer songs do not necessarily have higher popularity.







Key Insights

Title: Key Findings

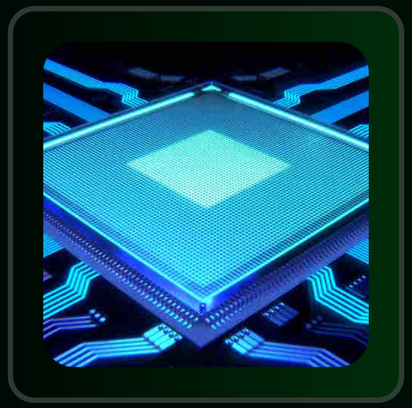
Points:

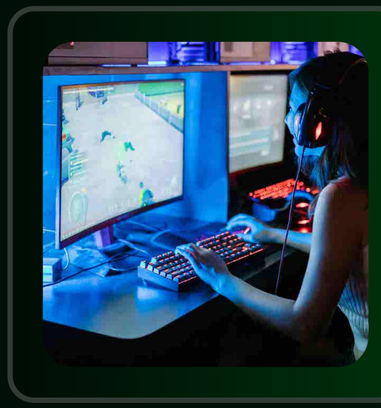
Popular songs are generally energetic and danceable.

Pop and Hip-hop dominate the dataset.

No strong correlation between track duration and popularity.

Dataset is suitable for building recommendation systems.



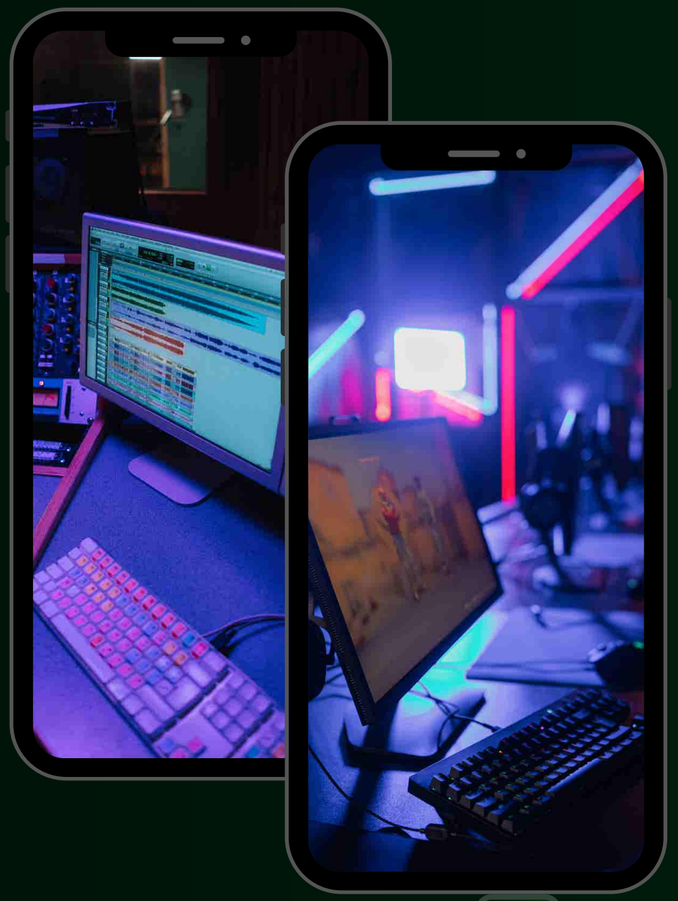


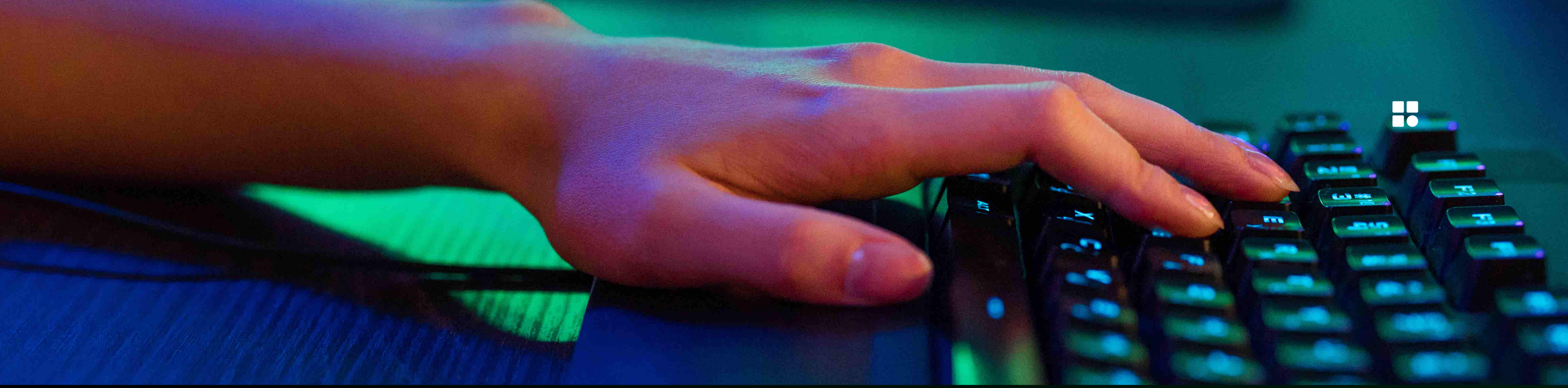
Conclusion

The EDA helped us understand musical patterns in Spotify tracks.

Insights can guide music recommendation or trend analysis.

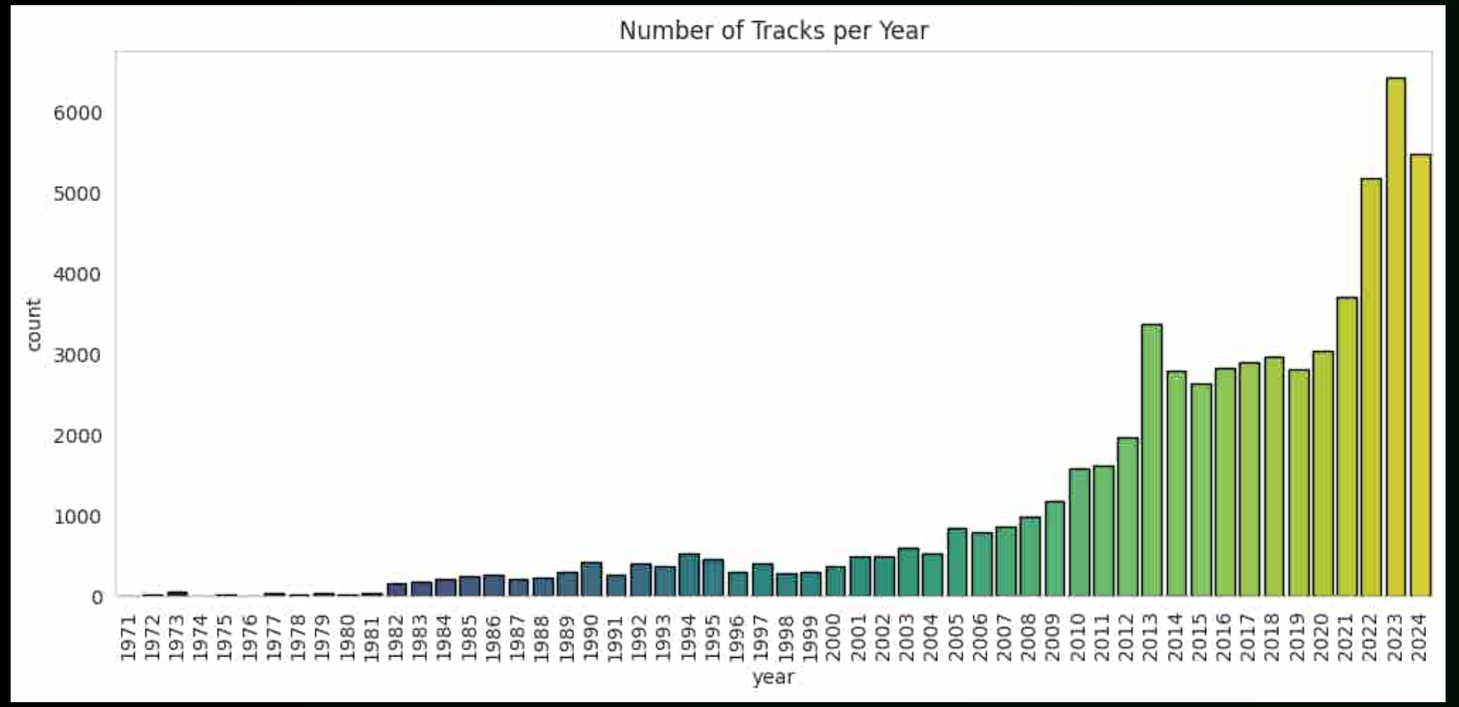
Future work: Apply machine learning to predict song popularity or cluster genres

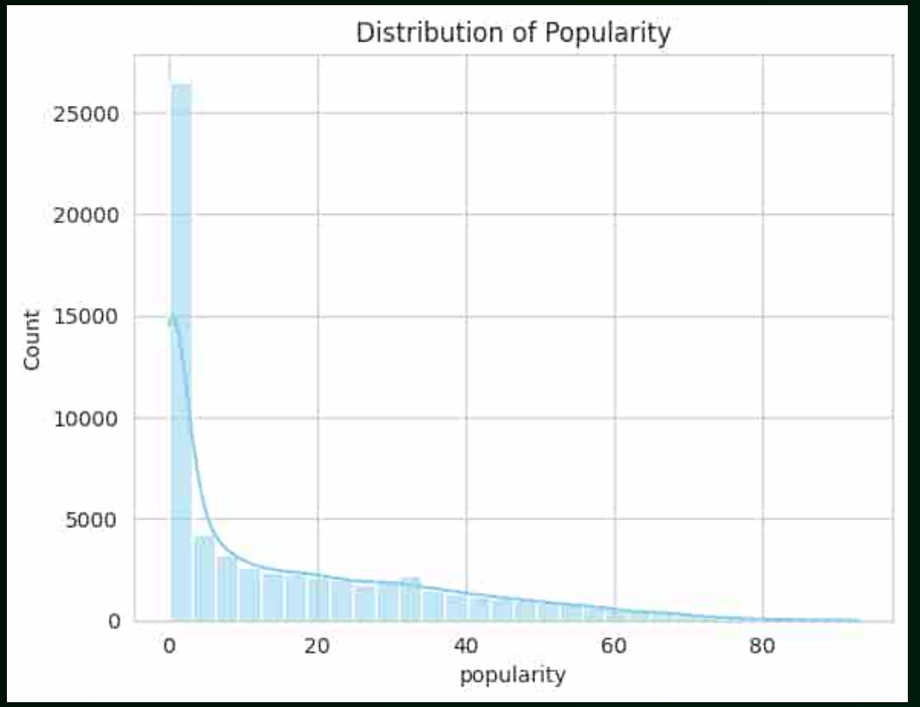


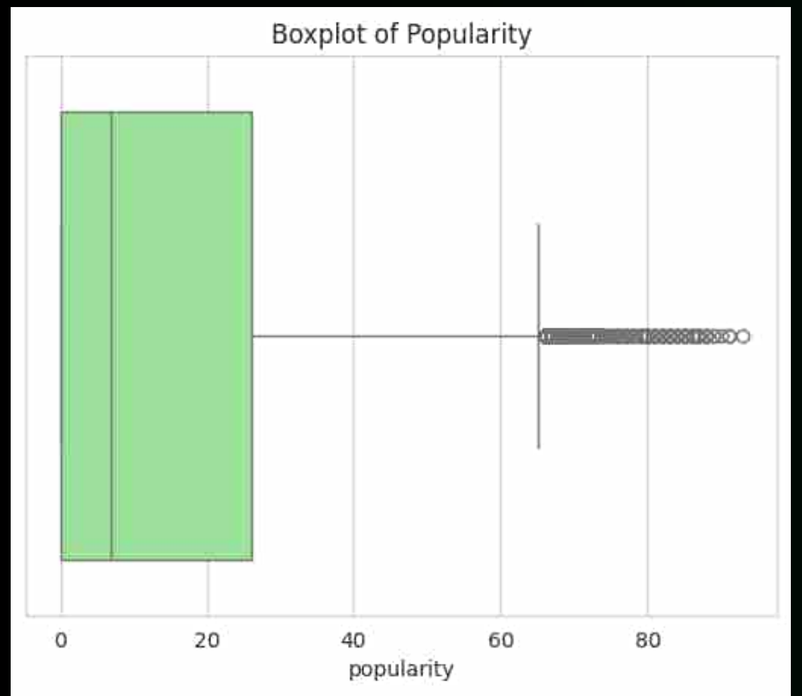


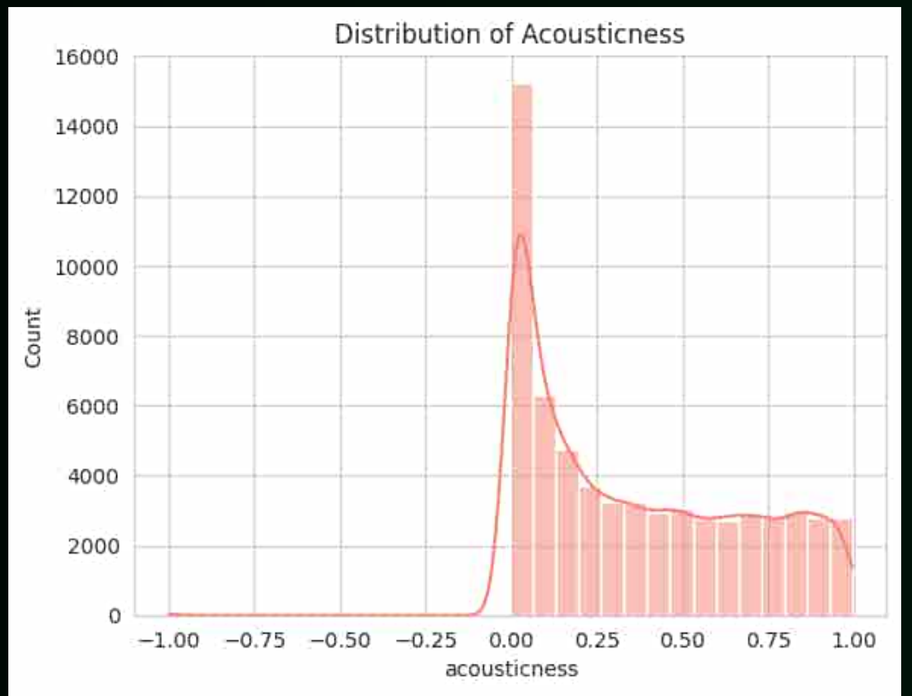
THANK YOU

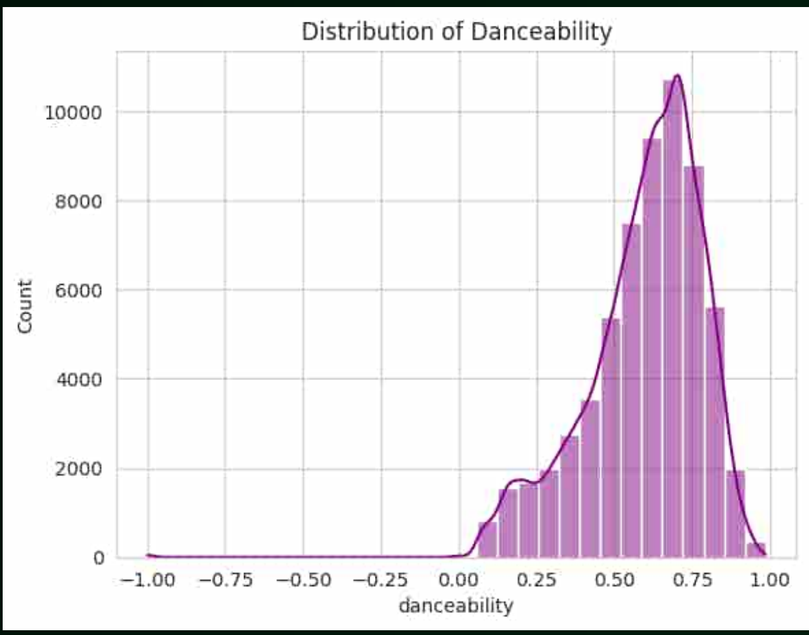
-Banasri Patra

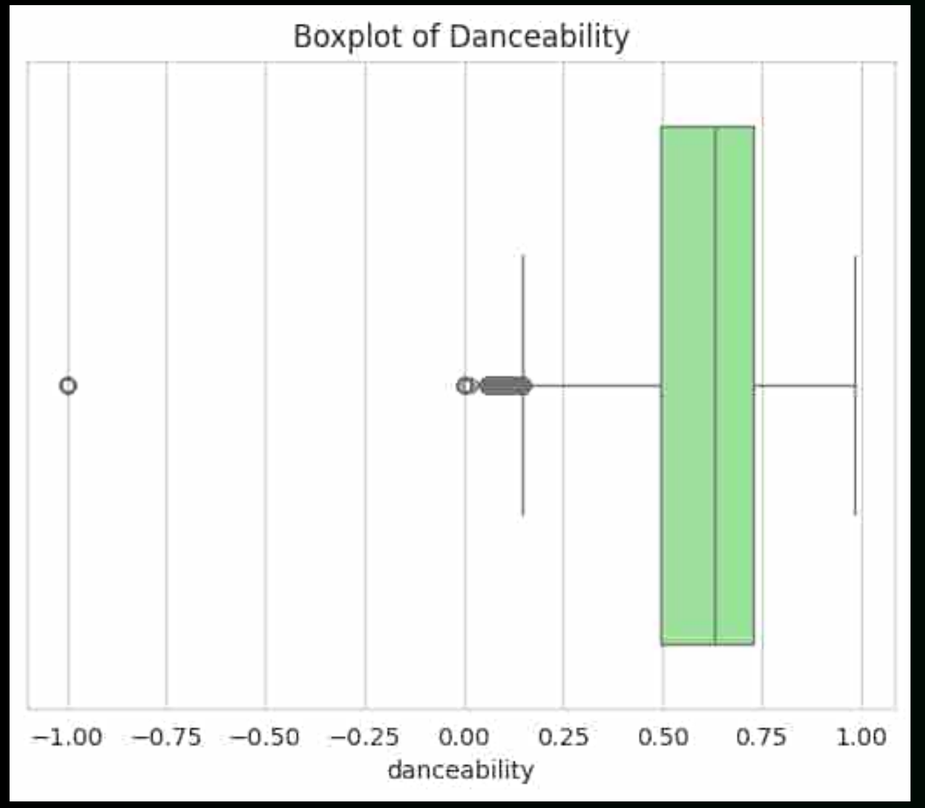

Number of Tracks per Year
count
year
0
1000
2000
3000
4000
5000
6000
1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

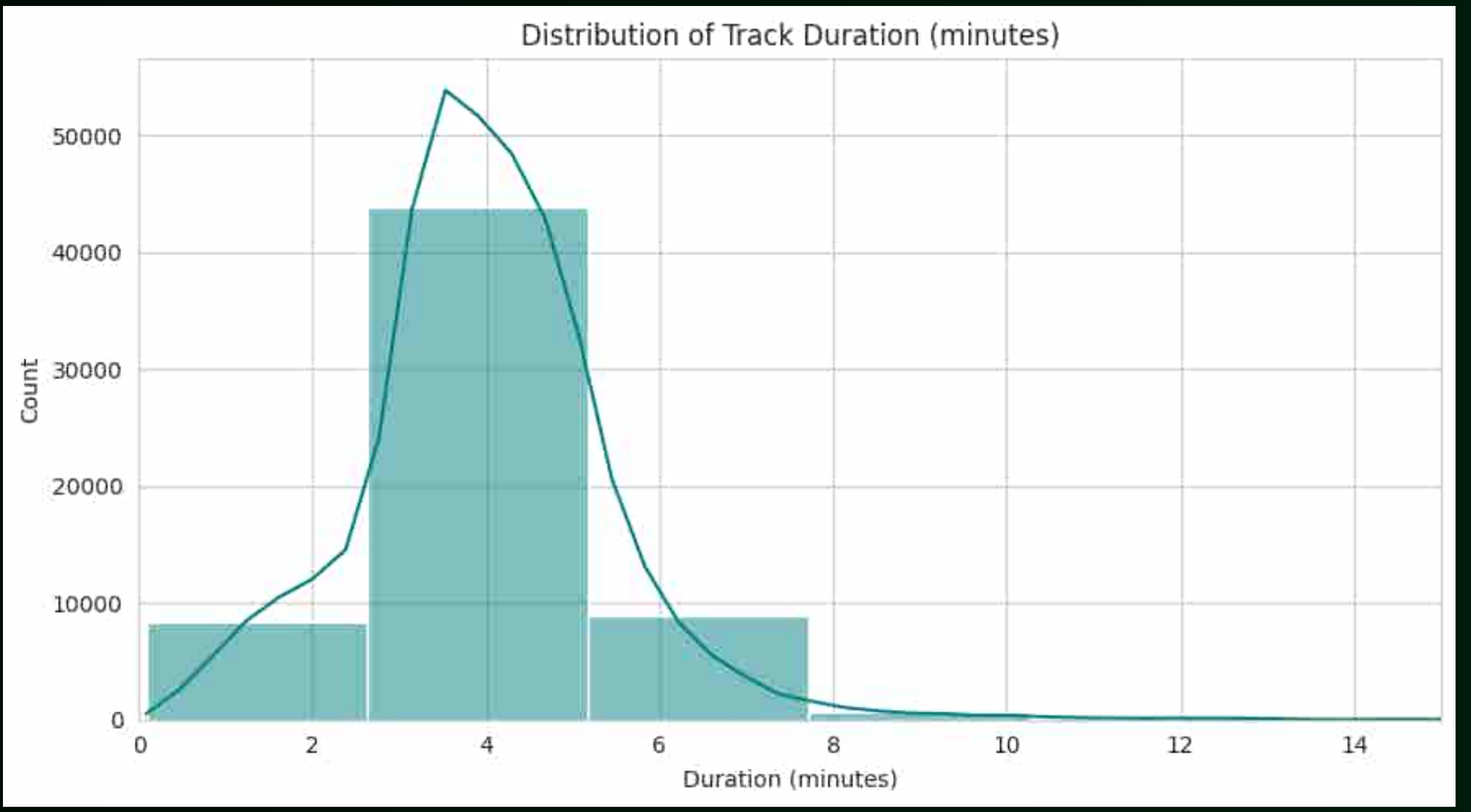

Distribution of Popularity
Count
popularity
0
5000
10000
15000
20000
25000
0 20 40 60 80

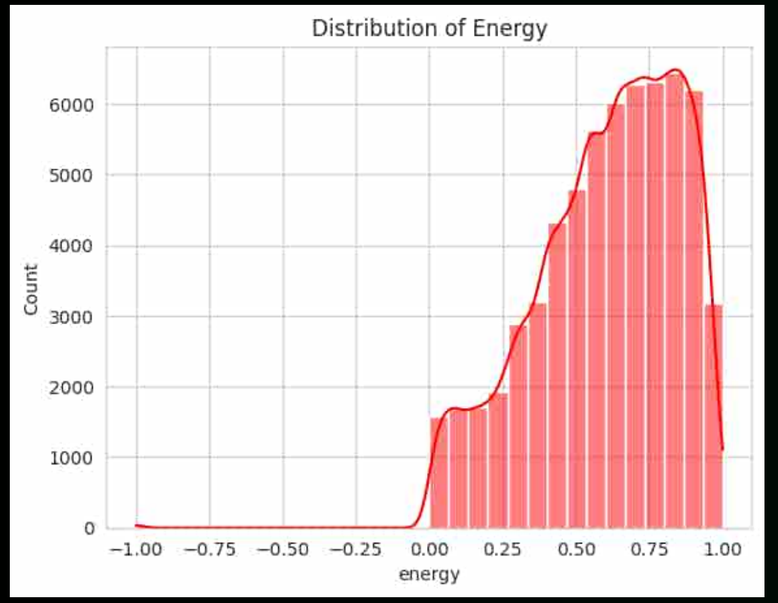

Boxplot of Popularity
popularity
0 20 40 60 80

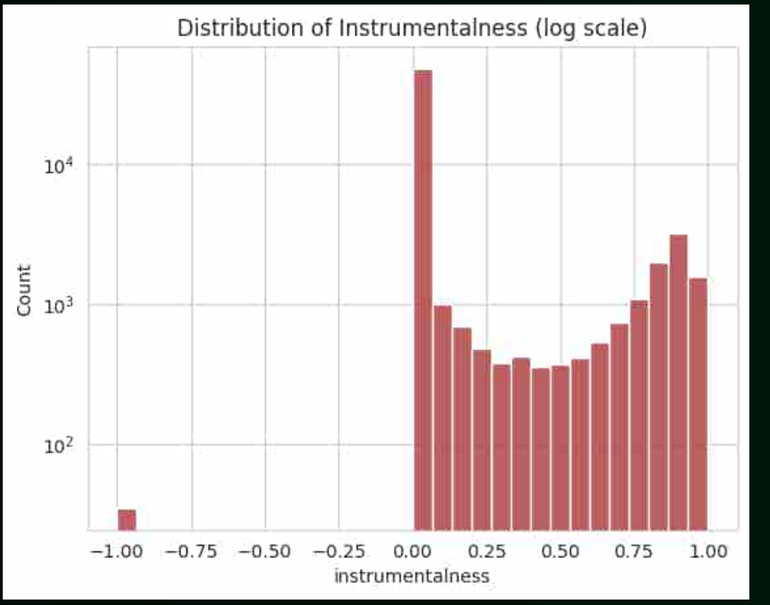

Distribution of Acousticness
Count
acousticness
0
2000
4000
6000
8000
10000
12000
14000
16000
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

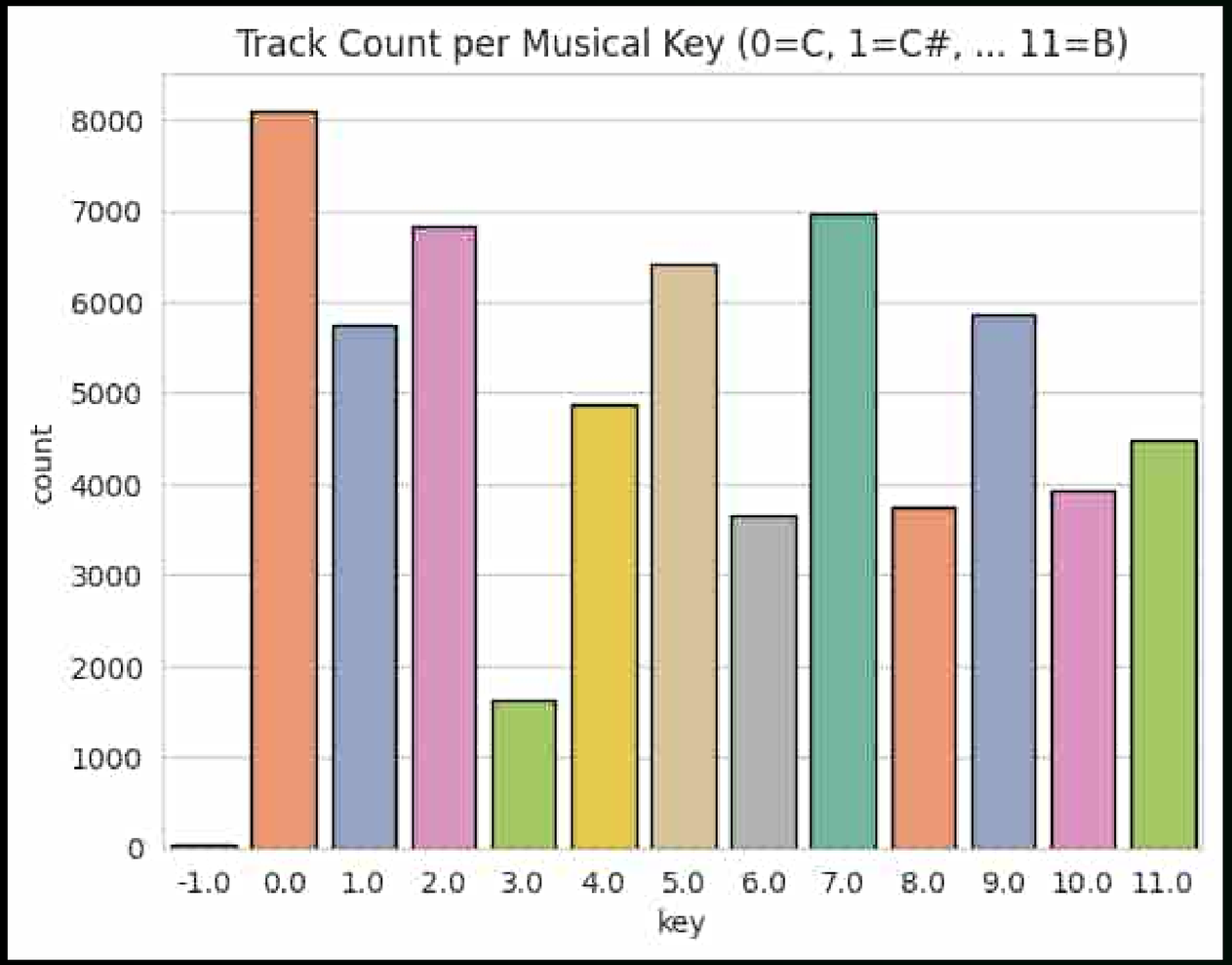

Distribution of Danceability
Count
danceability
0
2000
4000
6000
8000
10000
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

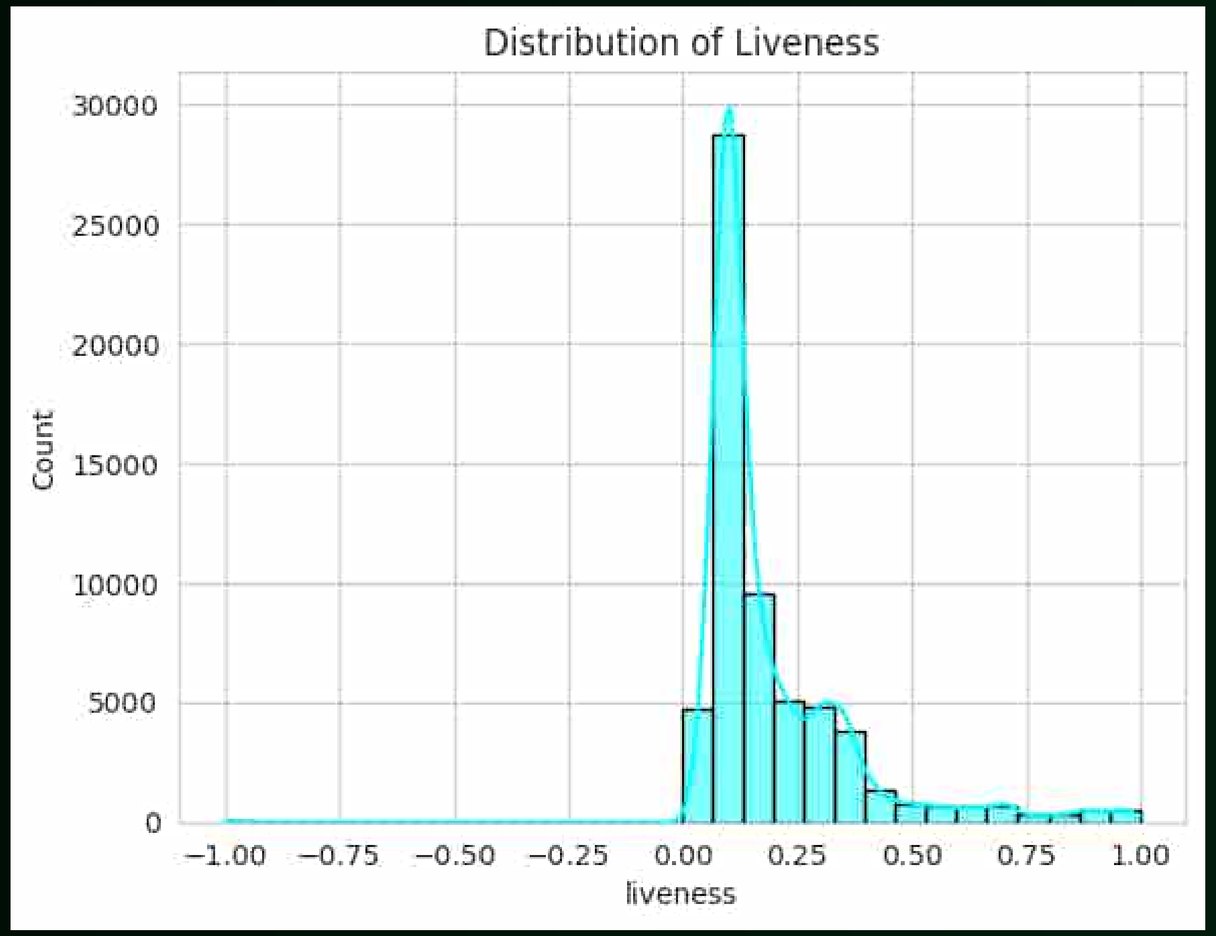

Boxplot of Danceability
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00
danceability

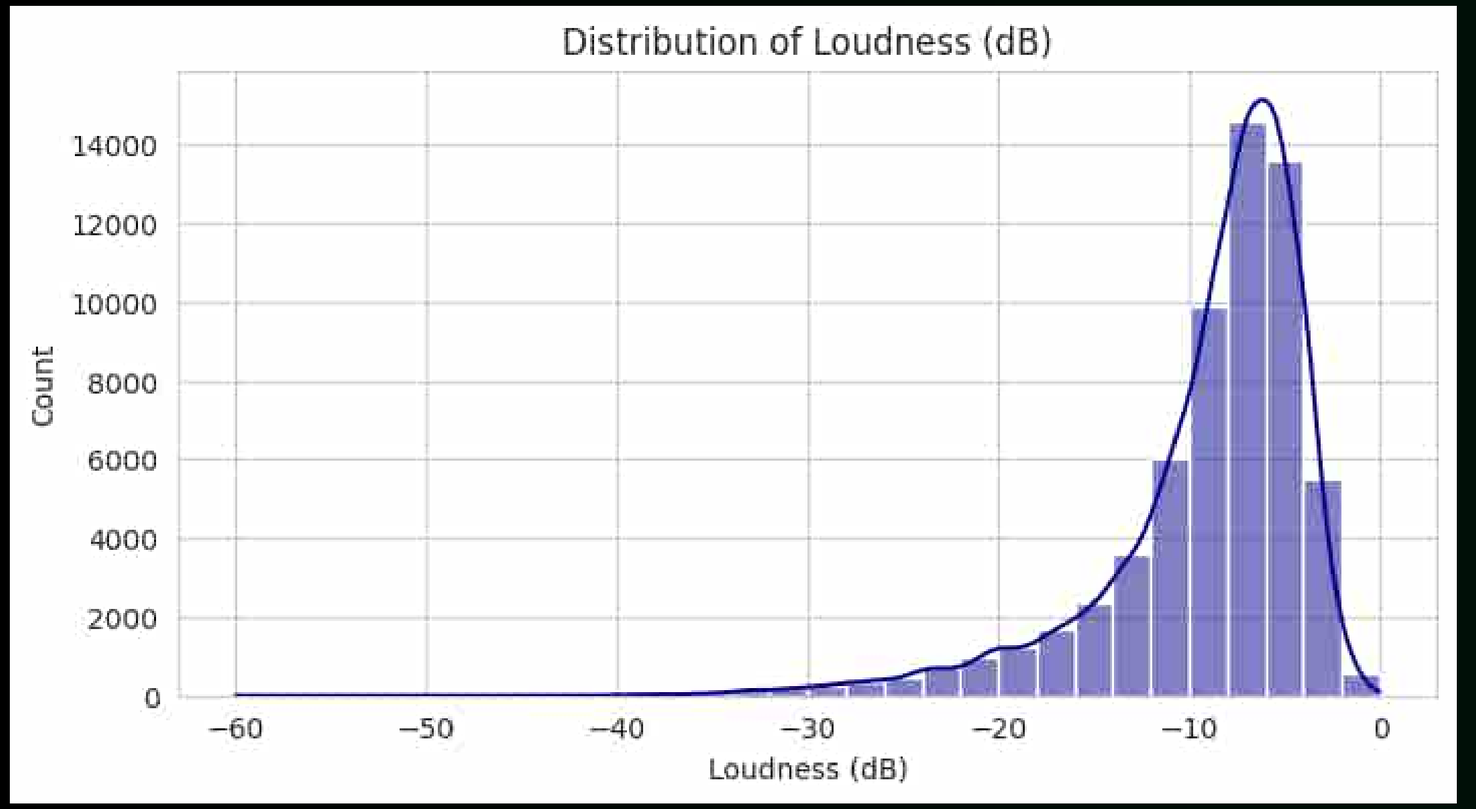

Distribution of Track Duration (minutes)
Count
50000
40000
30000
20000
10000
0
0 2 4 6 8 10 12 14
Duration (minutes)

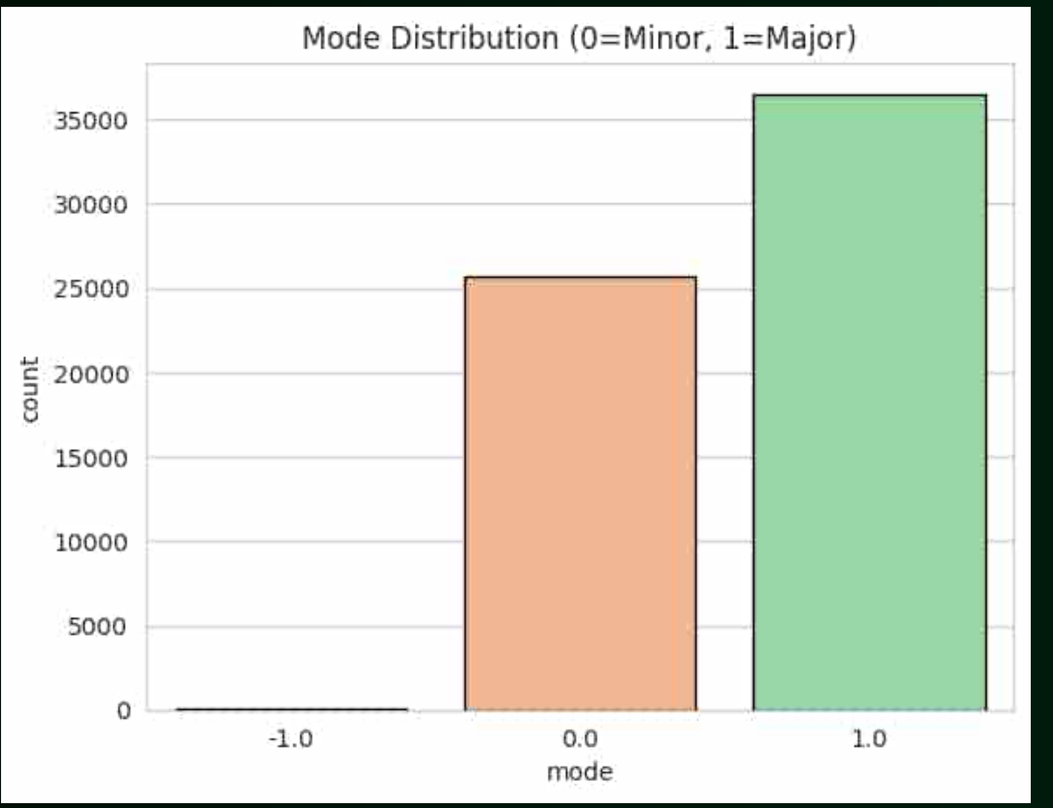

Distribution of Energy
Count
6000
5000
4000
3000
2000
1000
0
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00
energy

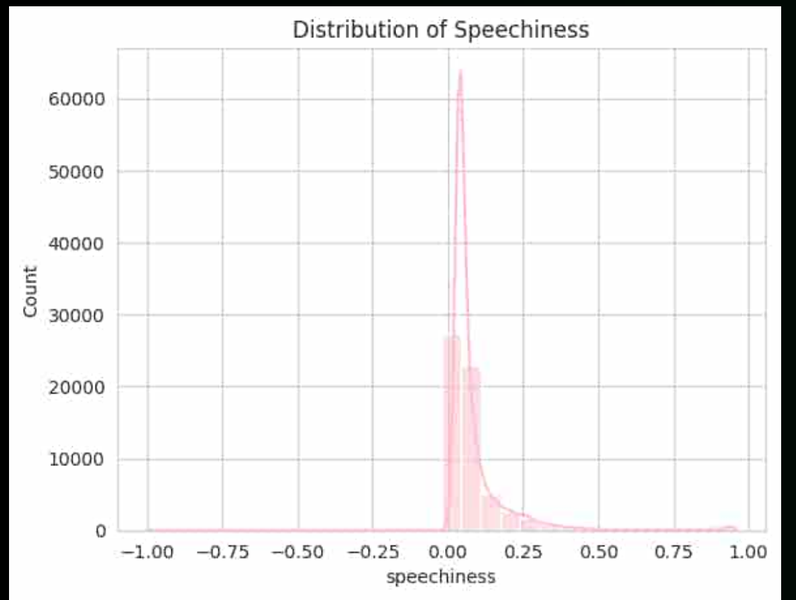

Distribution of Instrumentalness (log scale)
Count
10^4
10^3
10^2
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00
instrumentalness

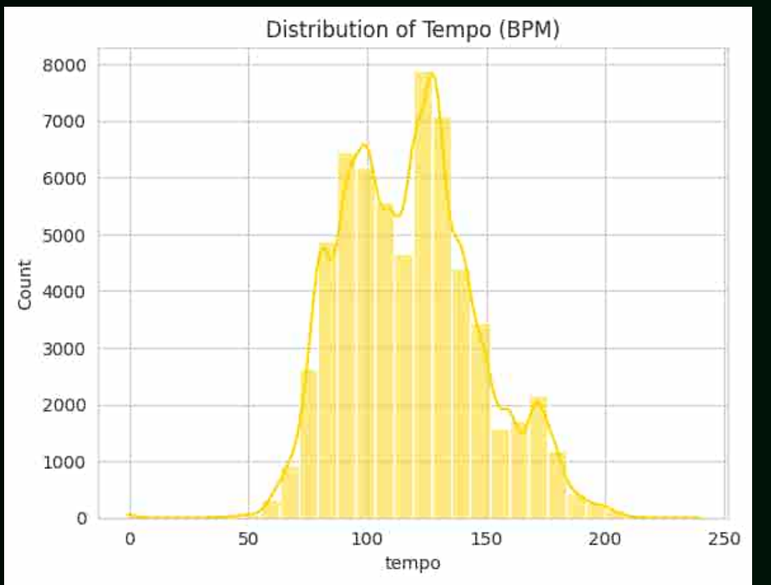

Track Count per Musical Key (0=C, 1=C#, ... 11=B)
count
key
0
1000
2000
3000
4000
5000
6000
7000
8000
-1.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0

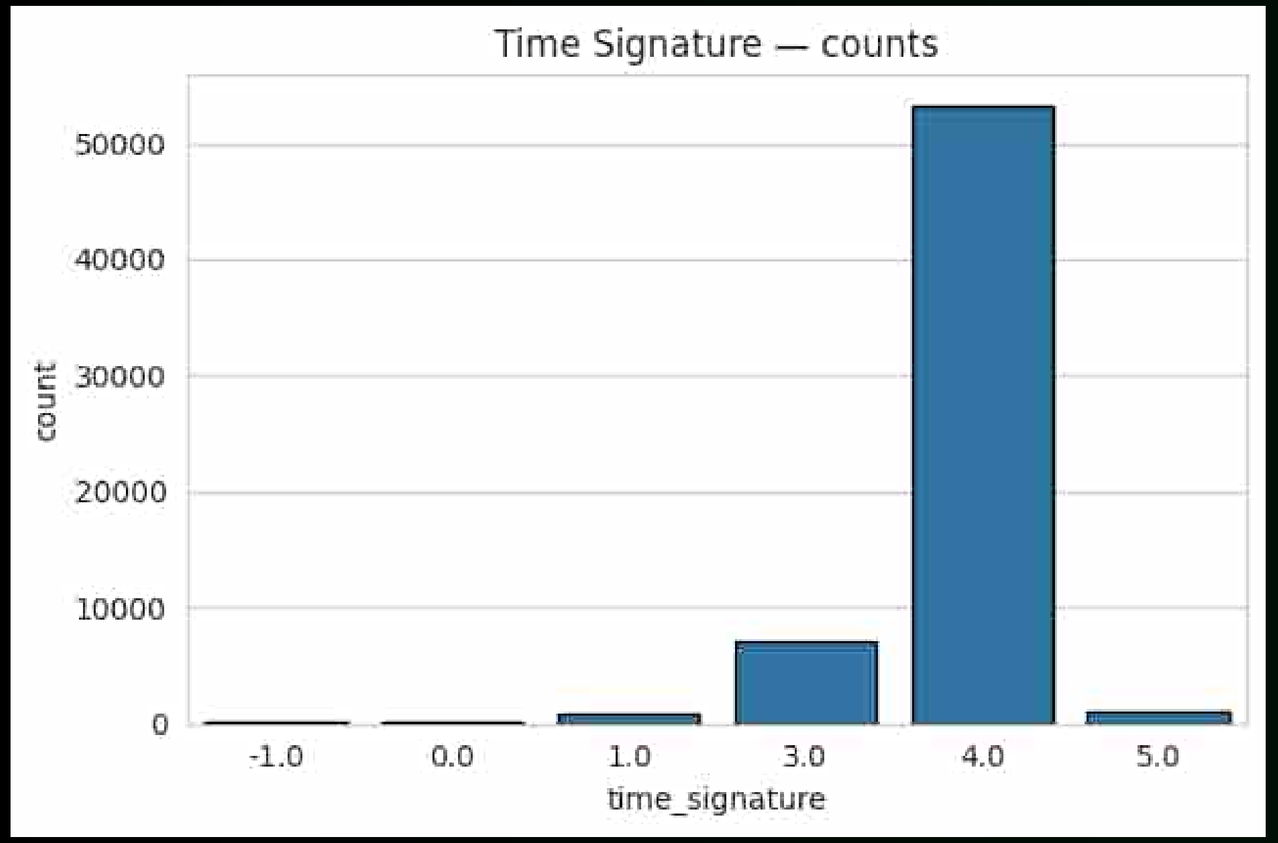

Distribution of Liveness
Count
liveness
0
5000
10000
15000
20000
25000
30000
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


Distribution of Loudness (dB)
Count
14000
12000
10000
8000
6000
4000
2000
0
−60
−50
−40
−30
−20
−10
0
Loudness (dB)


Mode Distribution (0=Minor, 1=Major)
count
35000
30000
25000
20000
15000
10000
5000
0
-1.0
0.0
1.0
mode


Distribution of Speechiness
Count
60000
50000
40000
30000
20000
10000
0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
speechiness


Distribution of Tempo (BPM)
Count
8000
7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
200
250
tempo


Time Signature — counts
50000
40000
30000
20000
10000
0
count
-1.0
0.0
1.0
3.0
4.0
5.0
time_signature