

July 1, 2022

• • •

THE SOUND OF THE FUTURE MUSIC POPULARITY PREDICTION MODEL SPOTIFY TOP 2000 SONGS

• • •

MUSIC ON DEMAND, ANYTIME, ANYWHERE

• • •

Is it possible to predict the future of Music using linear regression?

How is Big Data
Revolutionizing the Music
Industry?

• • •

DISCUSSION FLOW

Is it possible to predict a song's placement on Spotify Top 100 list based on a given set of variables?

What trend can be observed by visualising the patterns over the years?

DEFINITION & OBJECTIVE

...

USING A DATA SET OF TOP SPOTIFY HITS FROM 2000-2019, FINDING THE MAJOR FACTORS IN DETERMINING WHAT MAKES ONE SONG MORE POPULAR THAN ANOTHER IN ORDER TO PREDICT HOW POPULAR NEW SONGS CAN BE



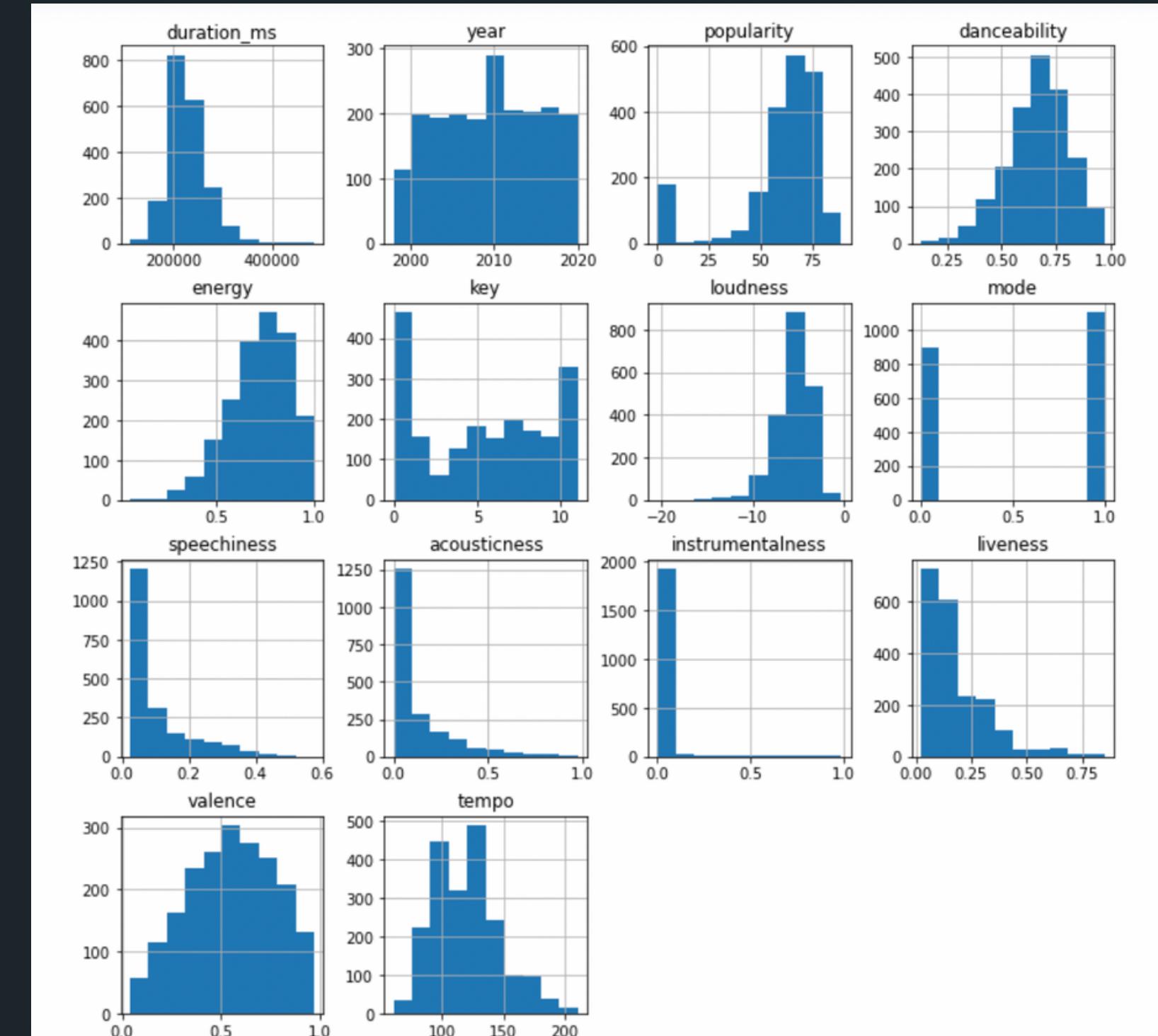
FEATURES

2000 observations
18 continuous variables

- artist
- song
- duration_ms
- explicit
- year
- popularity
- danceability
- energy
- key
- loudness
- mode
- speechiness
- acousticness
- instrumentalness
- liveness
- valence
- tempo
- genre

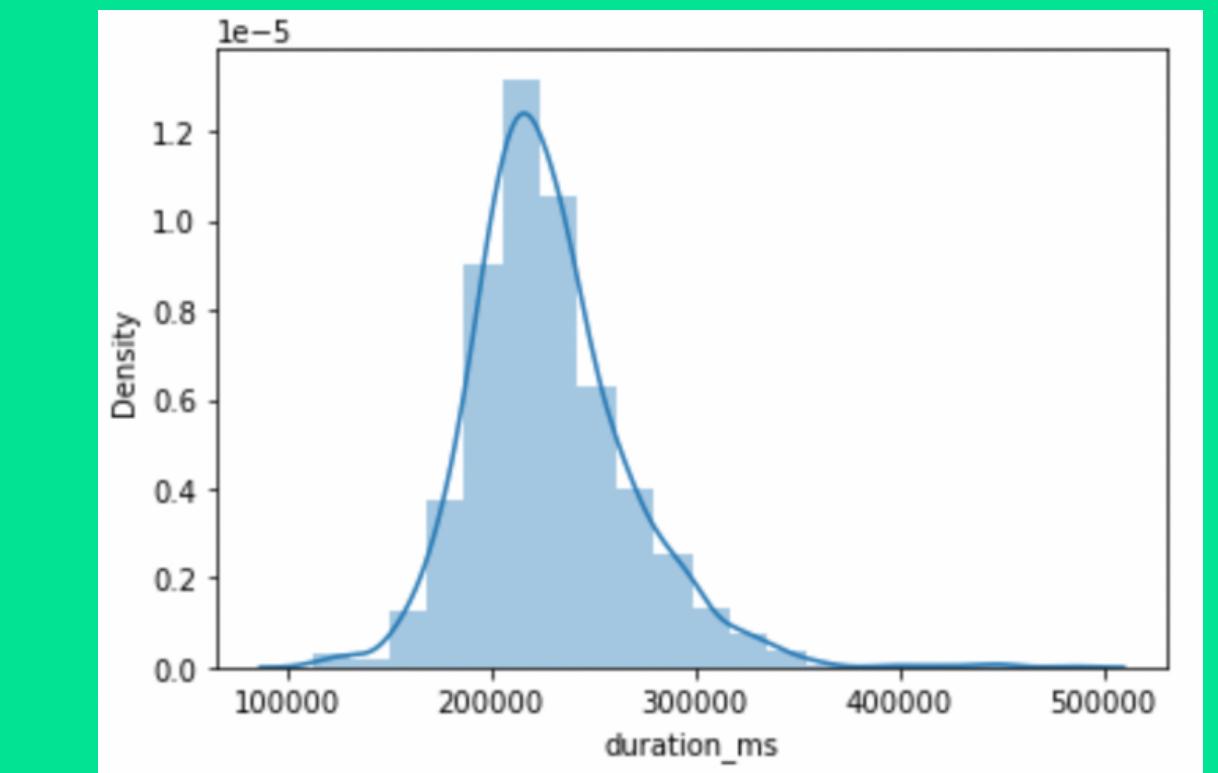
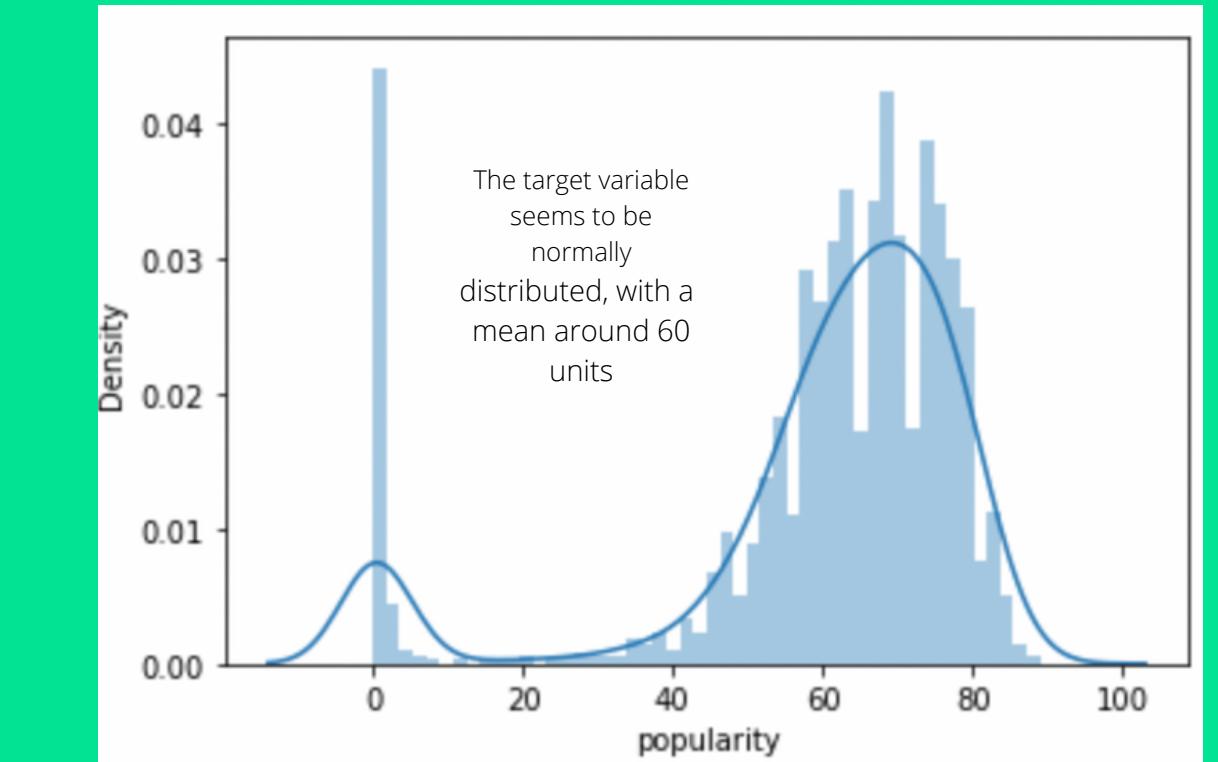
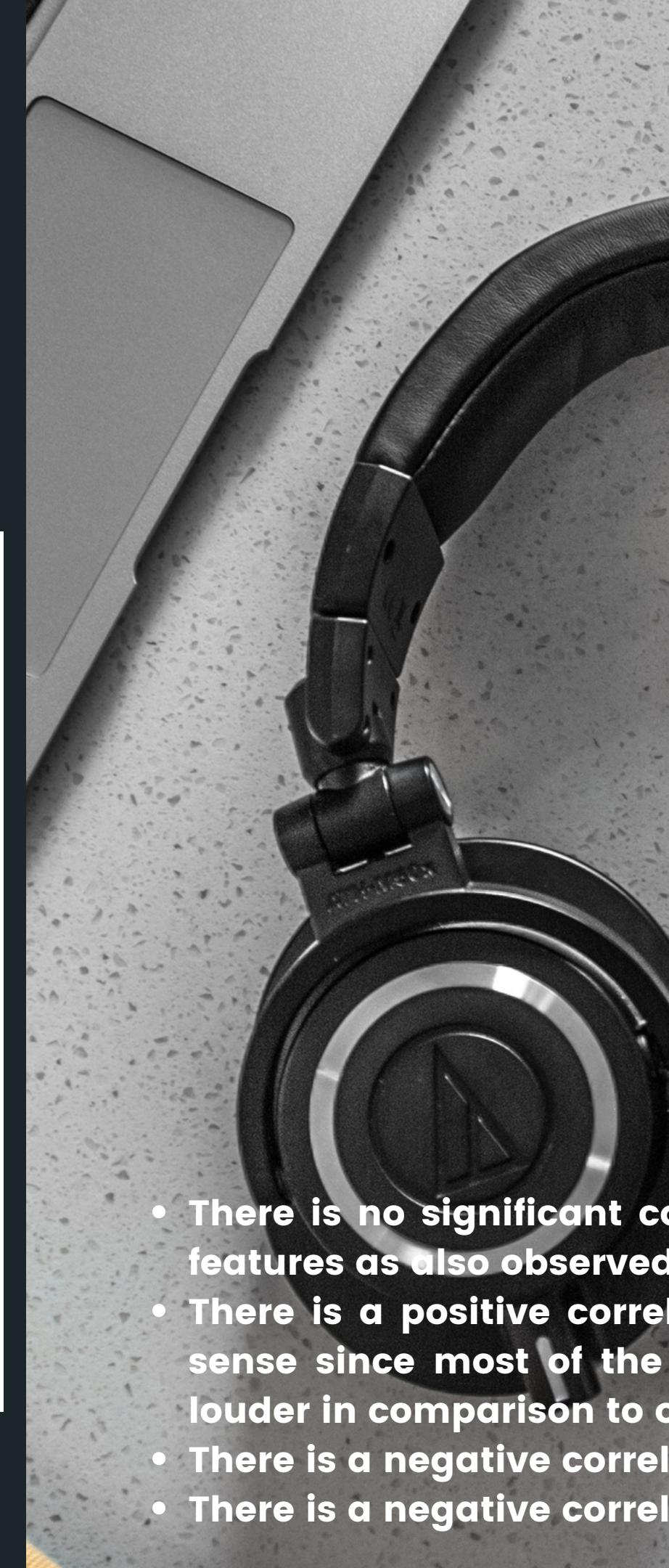
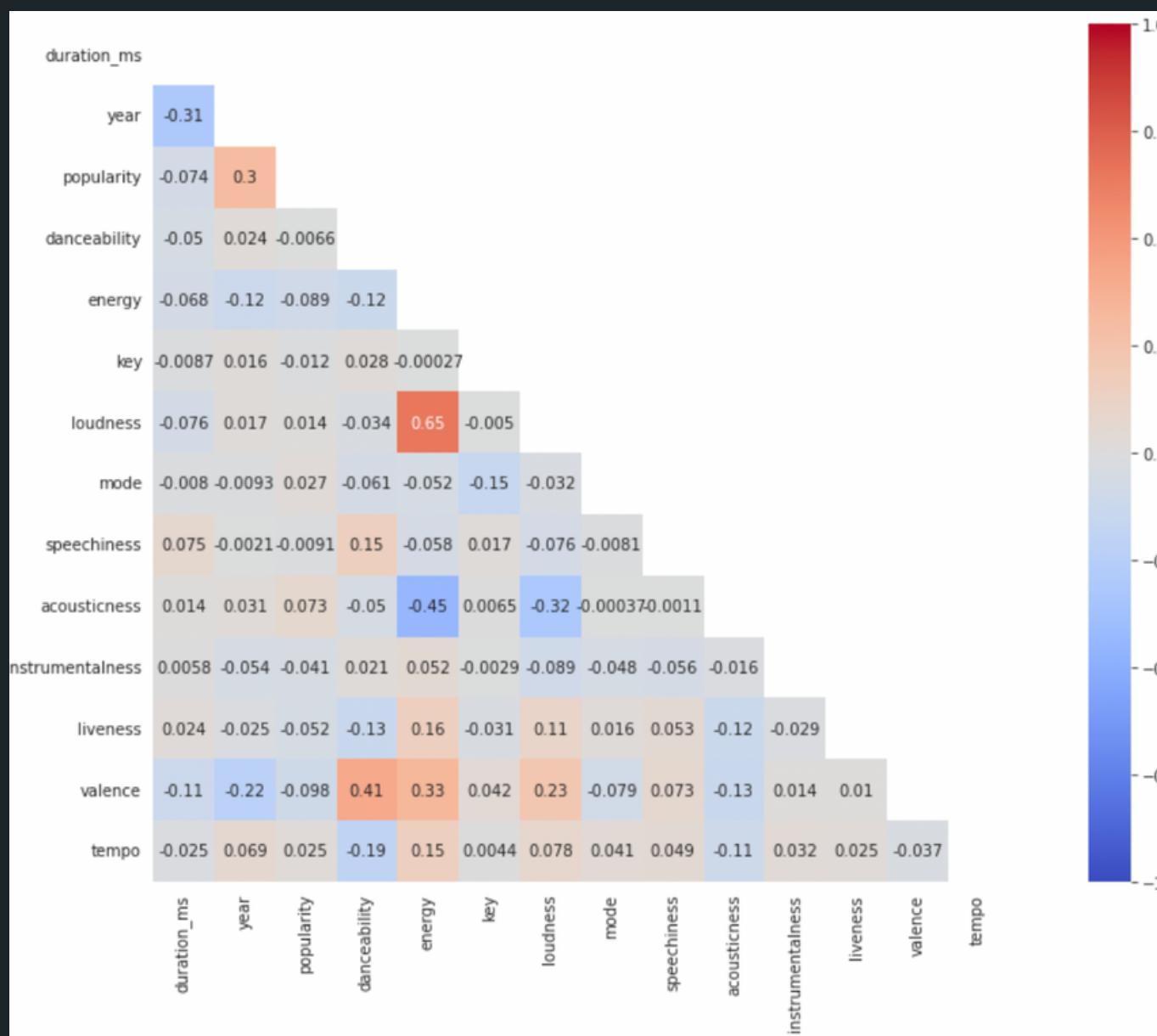
POPULARITY

Dependent variable | 1-100
The higher the value the more popular the song
Placement on Top 100 List

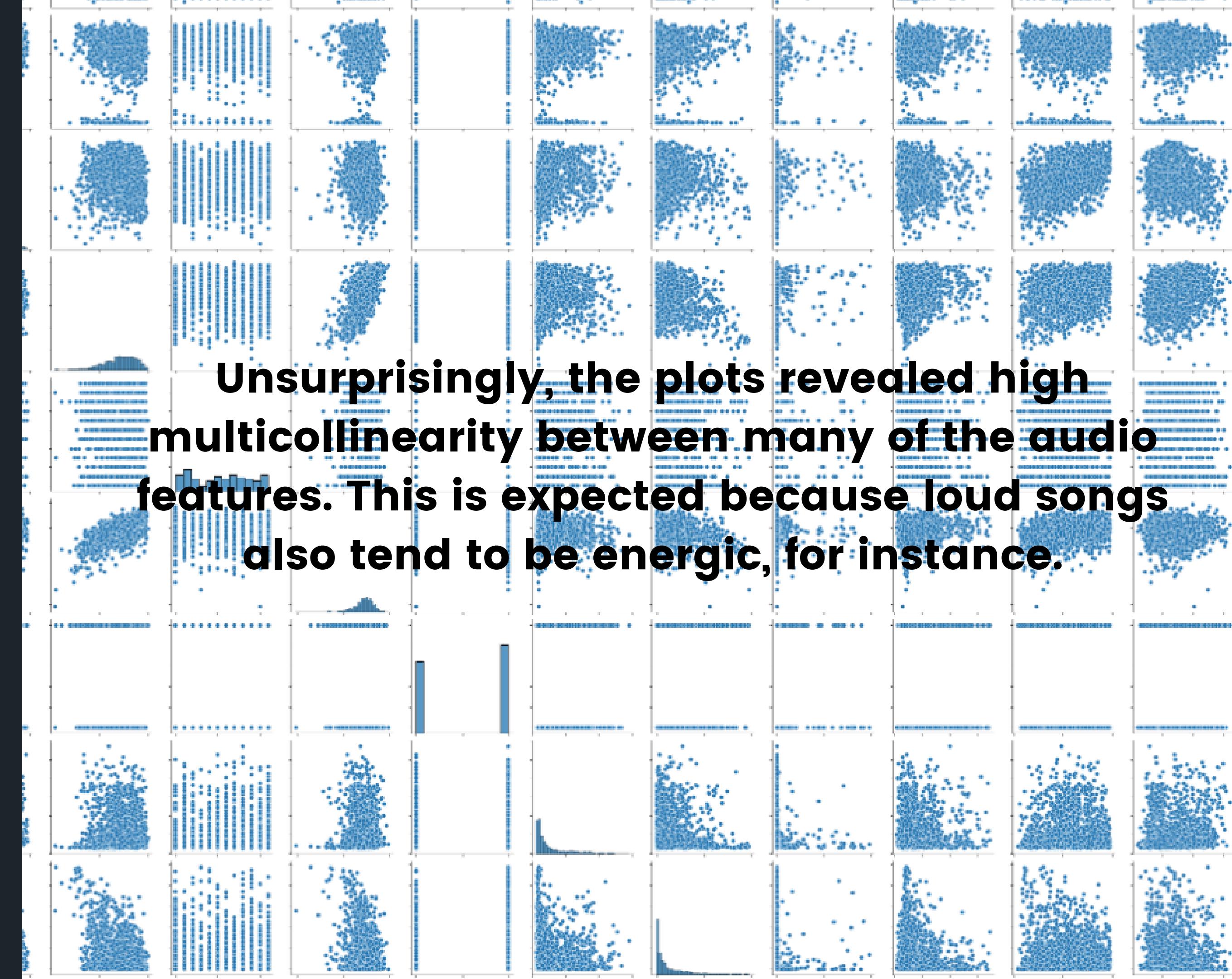
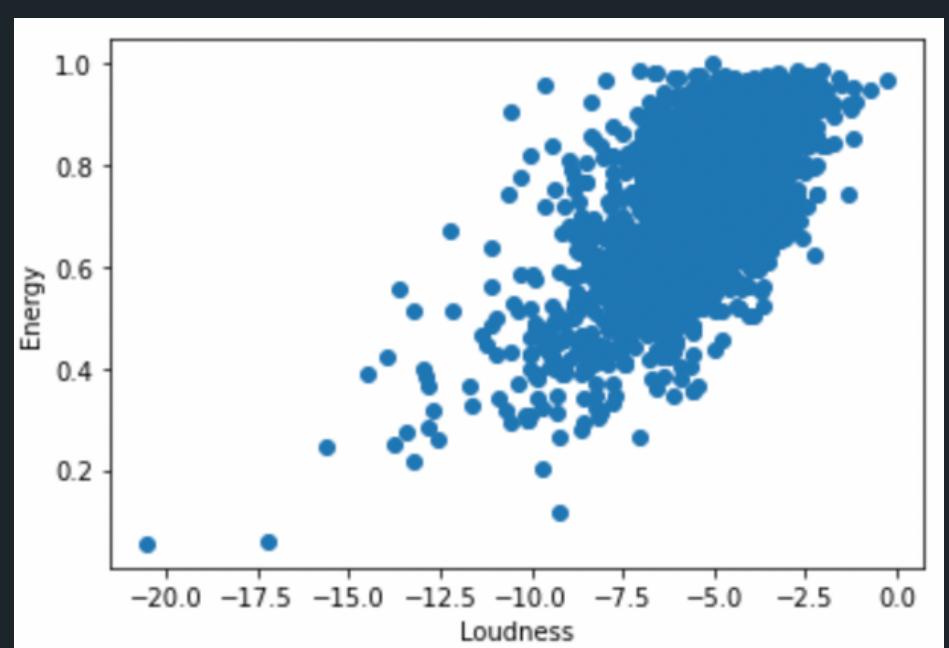
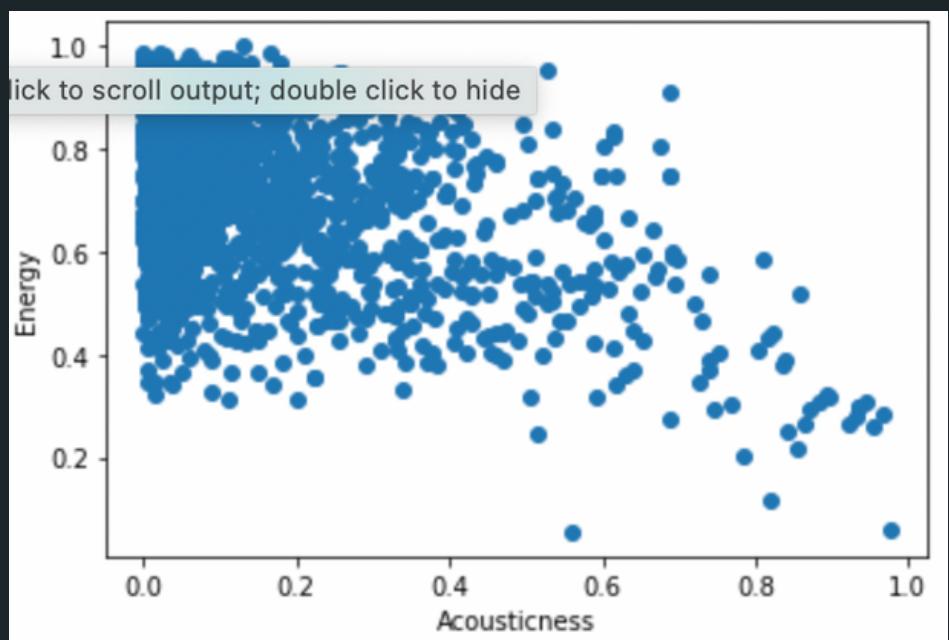
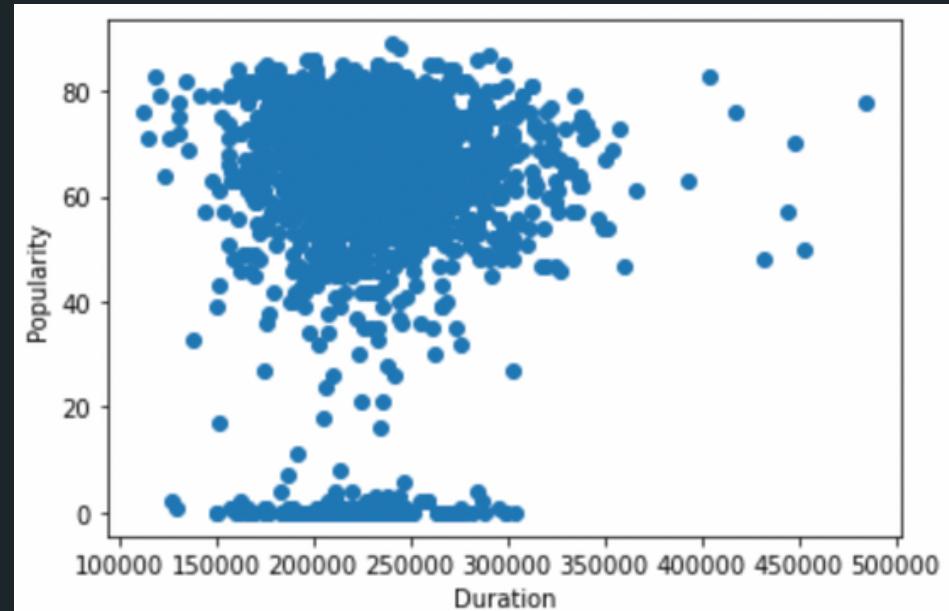


EDA & Plotting

- Visualising patterns over the years
- Filter genre: how is the trend
- Exploratory analysis



- There is no significant correlation between song popularity and the input features as also observed by the distributions plots above
- There is a positive correlation between energy and loudness. This makes sense since most of the songs which feel energetic (i.e. Rock songs) are louder in comparison to other songs which feel less energetic (i.e. ballads)
- There is a negative correlation between energy and acousticness
- There is a negative correlation between loudness and acousticness



Looking at the data, there are not obvious trends in any of the variables.

No Linea Regression was calculated

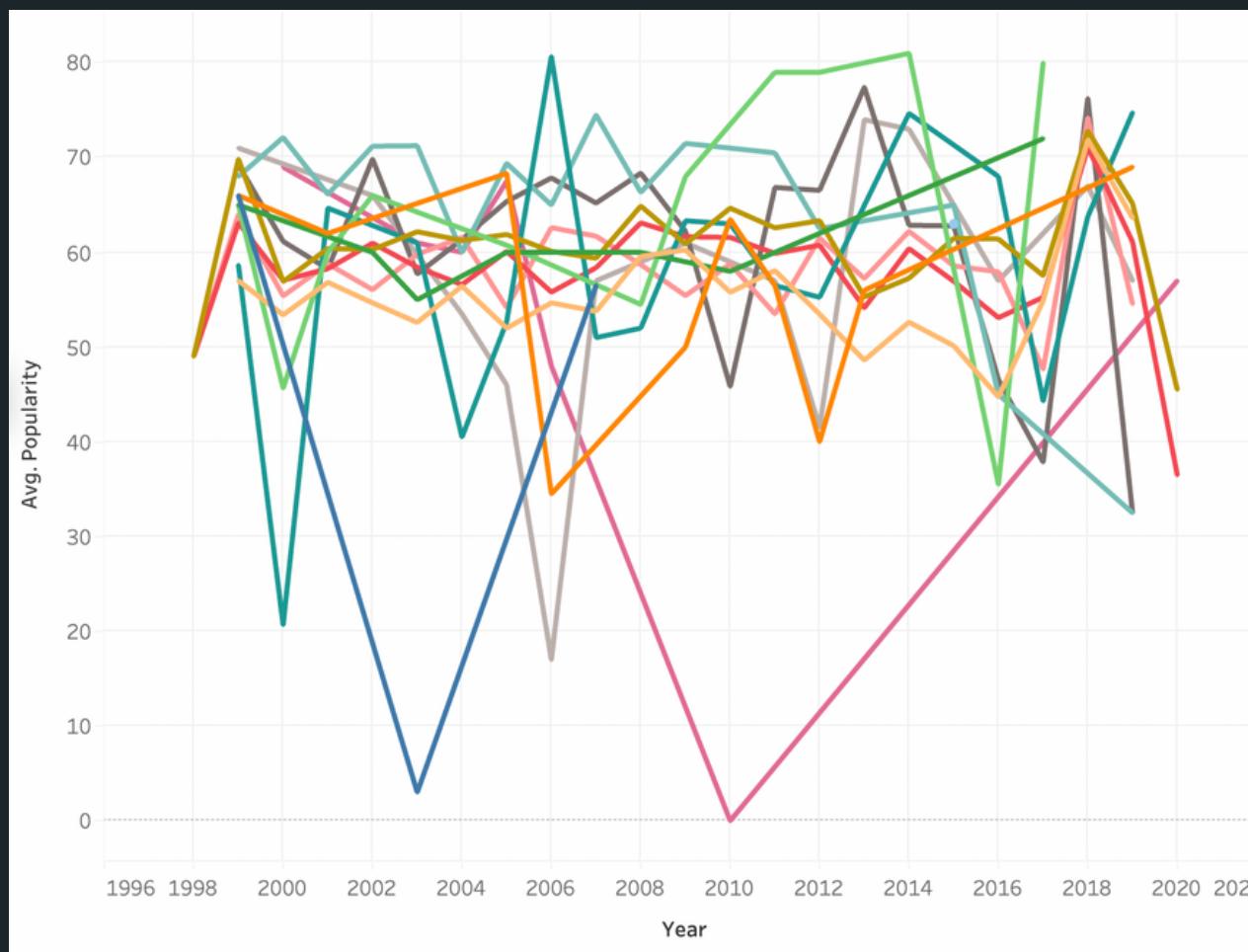
Every parameter is significant, but loudness behave different, because of good correlation with some parameters

Popularity is a complex concept and other variables can have an impact beyond the random variables such as the artist's reputation, lyrical meaning, release date

Tracks have become more Energetic and Danceable in the recent years. The loudness and tempo has also increased. Meanwhile the tracks have become less "Acoustic"

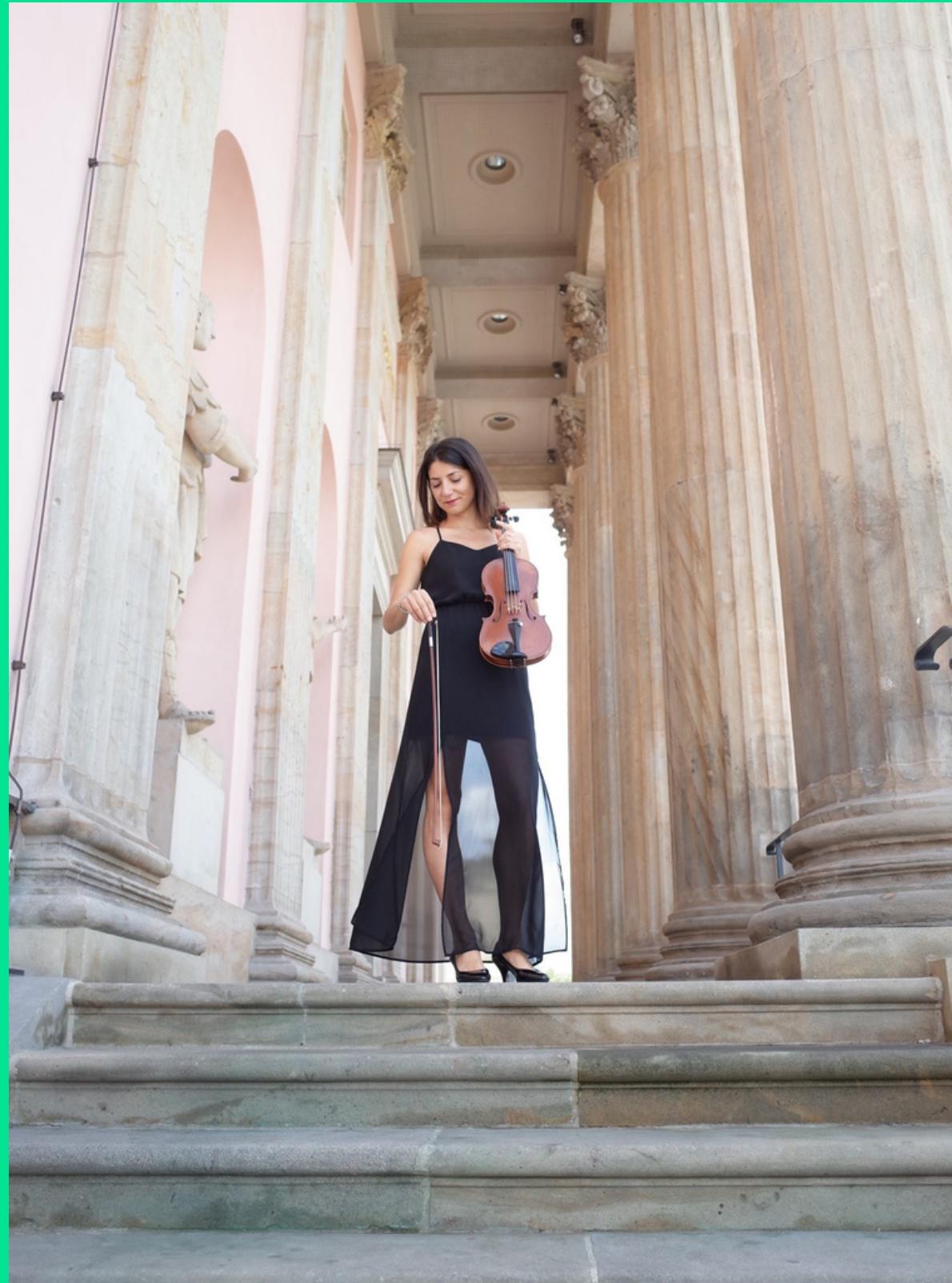
EDA showed that the distribution of the dependent variables is unbalanced and makes it difficult for a model to predict the very popular songs.

Next step: filtering the genre and observe the trend



BI tools
Genre had many subgenres and classifications
Difficult to filter and divide into only one categories
Melt function in Pandas

MUSIC LAW DATA MUSIC LAW DATA MUSIC LAW DATA MUSIC LAW DATA



Costanza Brusutti

• • •

MUSICIAN

Classical music expertise, Violin player

CAREER HIGHLIGHTS

Copyright Music Law professional
CWR, Income Tracking, Audio Fingerprinting,
Data Analytics