TOP 100 SPOTIFY SONGS

TABLE OF CONTENT
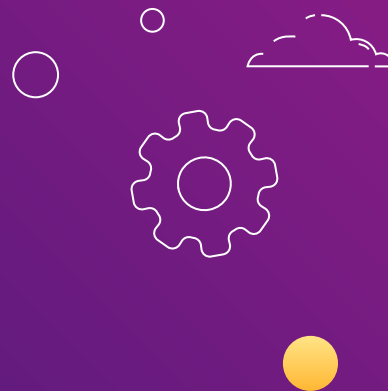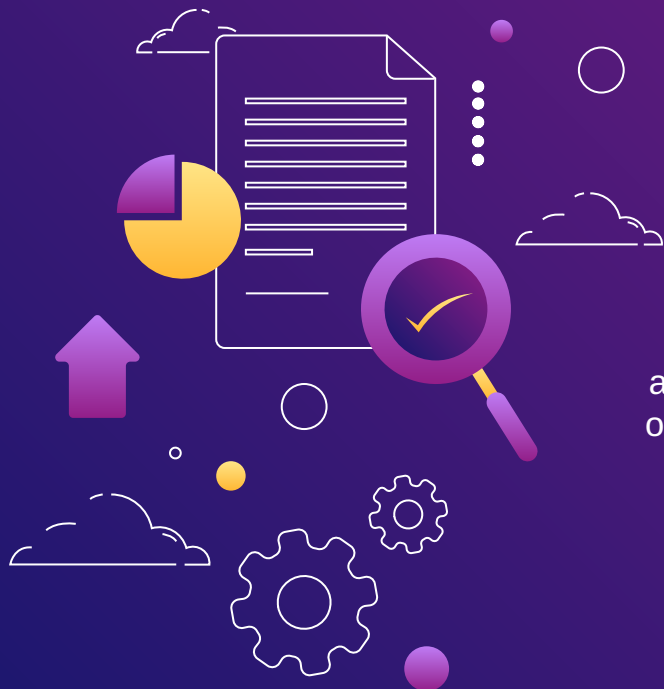
01 INTRODUCTION

Research project aims to investigate the audio features of popular songs across four consecutive years from 2017 to 2020. Specifically, it seeks to analyze the correlation among audio features and to identify patterns that emerge from this analysis. The project also aims to examine the motivation behind the high streaming numbers of these particular songs and whether this can be attributed to their audio features. Finally, the study is set to determine whether the available audio features can be used to predict one feature based on the others, and which features exhibit the strongest correlation with one another.

BACKGROUND

The data used in this study were obtained from four different sources, namely Top Spotify Tracks of 2017, Top Spotify Tracks of 2018, Top Spotify Tracks of 2019, and Top 50 Spotify Tracks of 2020. These datasets were sourced from Kaggle, a popular platform for data science competitions, and were made available by Nadine Tamer and Atilla Colak. The datasets contain details such as the Spotify URI of each song, song title, name(s) of the artist(s) who created the song, and audio features of the song. These audio features may include information on factors such as tempo, key, danceability, among others.

QUESTIONS

- How do the audio features of these songs correlate, and what patterns emerge from this analysis?
- What motivates people to stream these particular songs the most, and can this be attributed to their audio features?
- Can we use the available audio features to predict one feature based on the others, and which features exhibit the strongest correlation with one another?
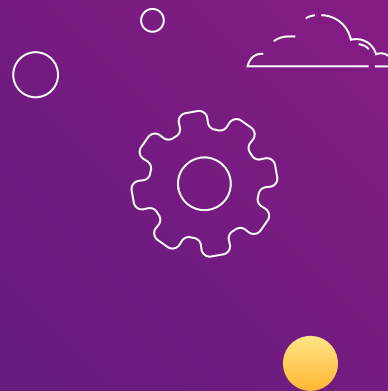- What are the commonalities among these popular songs, and what qualities make them appealing to listeners?
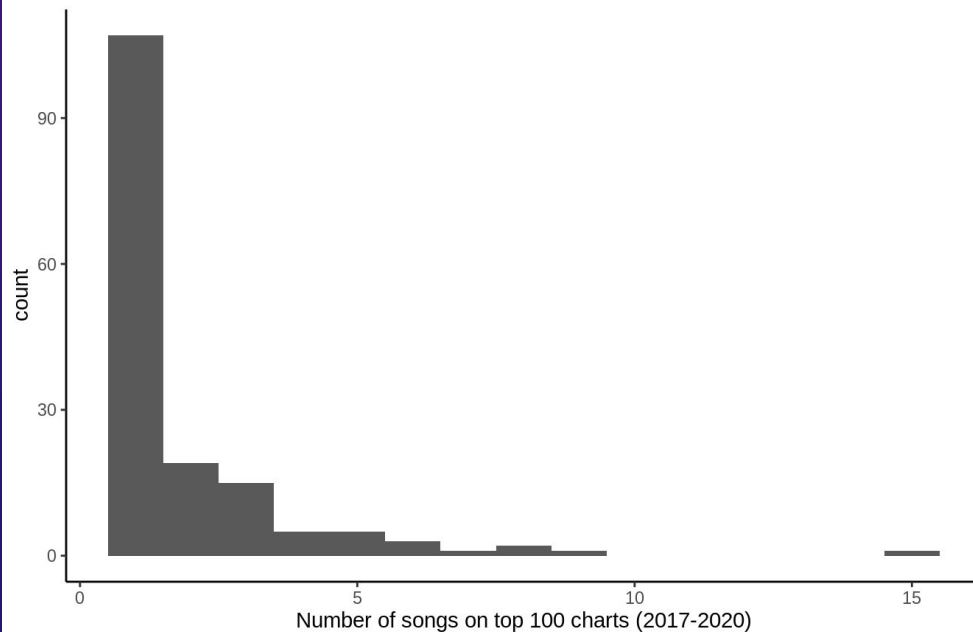
02 DATA ANALYSIS
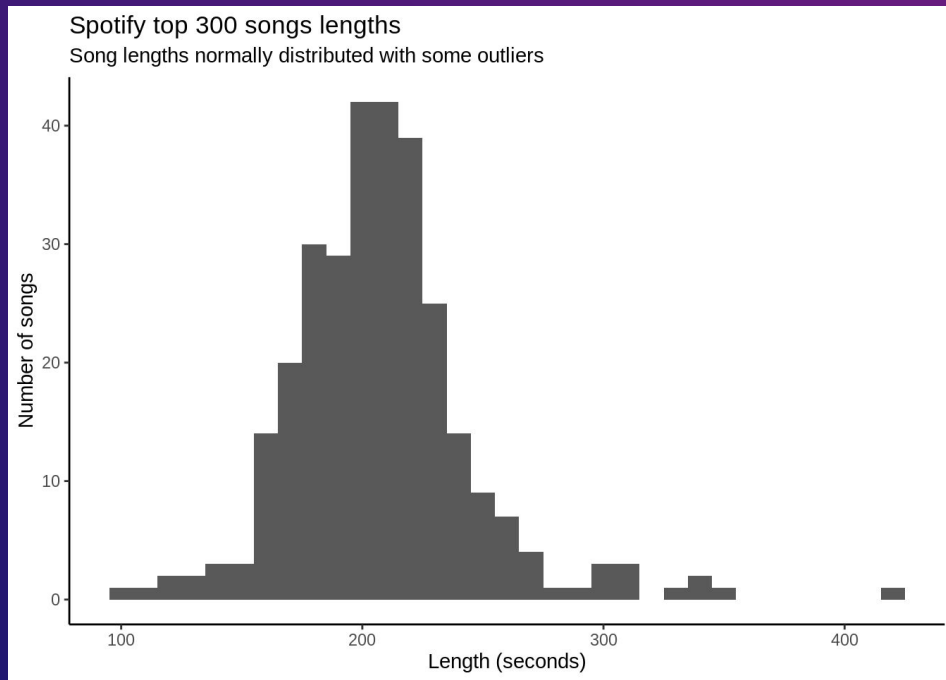
How many hit songs have artists made?
The majority of artists only have a single hit from 2017 to 2020

-Most artists only have a single hit song.

-Post Malone stands out with 15 over 4 years!

Spotify top 300 songs lengths
Song lengths normally distributed with some outliers

The average song length is just over 3 minutes with the shortest song being "Everybody Dies In Their Nightmares" by XXXTENTACION with a duration of 1:35 and the longest being "Te Bot? - Remix" by Nio Garcia at a duration of 6:57.

Most popular keys and modes
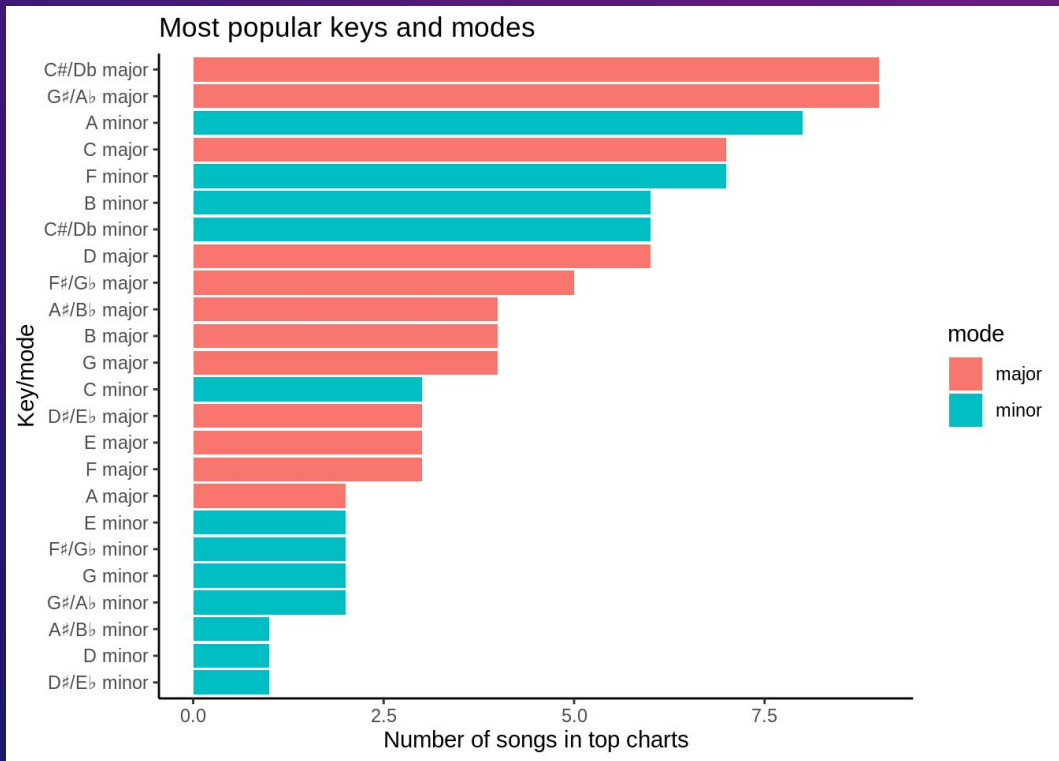
There is a large distribution of keys and modes within the data set, with C sharp and G sharp major being the most popular. Major and minor modes are also distributed fairly evenly throughout the data set.

Most popular keys and modes
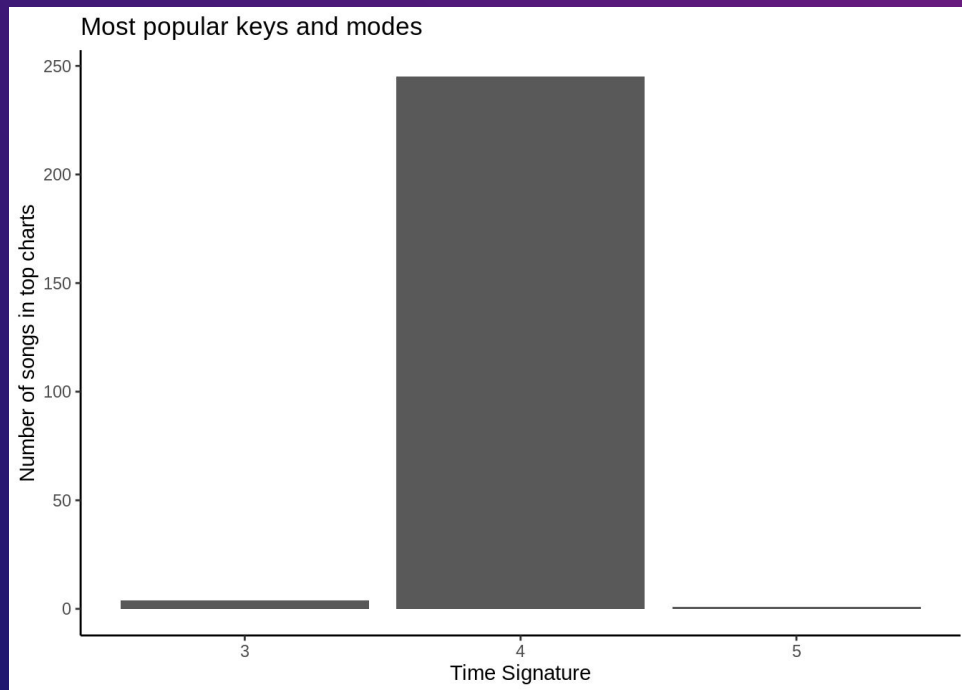
-4/4 time is by far the most popular, possibly as 4/4 time is easy to dance to and generally the most common time signature

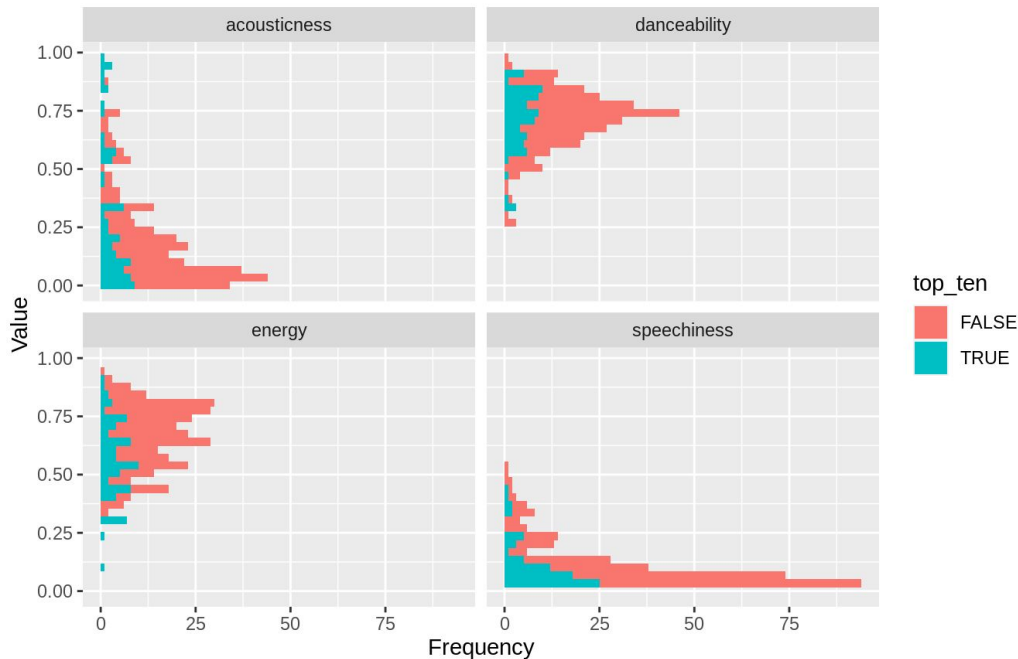-With only a few songs using alternate time signatures

-Assumes constant time signature

Top 100 Songs Characteristics
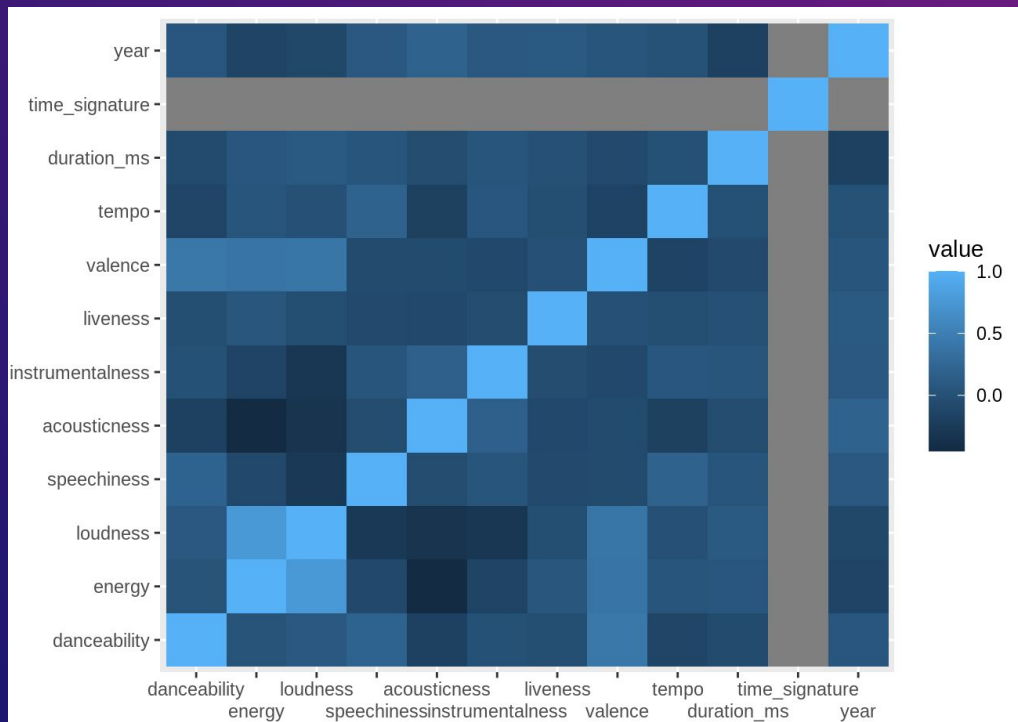Split between top 10 artists by number of songs on Spotify charts

-Lower speechiness and acousticness with some outliers.

-Higher danceability and energy.

-No obvious differences between top 10 artists by number of songs and song qualities
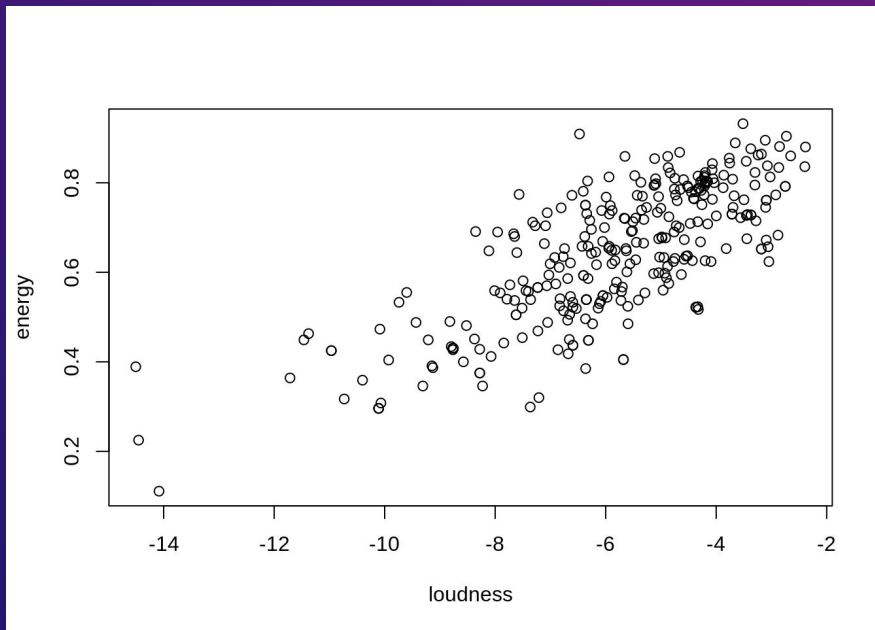
-Loudness and energy strongly correlated.

-Valence (song sentiment) is correlated with loudness, energy and danceability.

-As top songs have higher energy they are also mostly positive in sentiment

-A 75% correlation is found between loudness and energy of songs, allowing for a visualization to be created an a linear model.

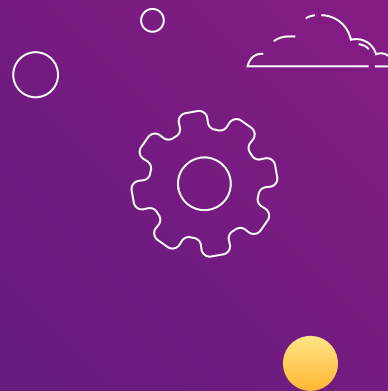-Louder songs are more likely to be energetic (fast/intense)

03 RESULTS

FINDINGS

-C#/Db major, G#/Ab major and A minor were the most popular keys. Though it should be noted that an artist has who has a high popularity must to be taken into consideration, as songs alone won't just make it into the top 100 Spotify list.

-Artists in the top 10 have songs with higher energy, danceability, and valence.
Often, the higher energy the song is, the more the intense the song is.
On average, speechiness was low compared to other audio features.

-When taking correlation the one with the highest pairing would be songs that are loud in decibels and have high energy. Of course, if you have a song with high energy it tends to get your attention. We can see that there is a 75% correlation in the data for loudness and energy between the top 100 songs in Spotify charts between 2017-2020.

CONCLUSION

-Top songs tended to be high in energy, loud and contain little speech.

-Though a comparison with a larger sample size of spotify songs may be needed, top trending songs had danceability and energy which leads to the conclusion that songs with those factors are more likely to trend, while of course bearing in mind outliers.

-Following these observations, it makes sense that top streamed songs would be high energy and generally easy to play in many different settings. A larger dataset would be needed, however it could be hypothesized that including songs outside the top list would lower the mean energy, valence, and danceability.
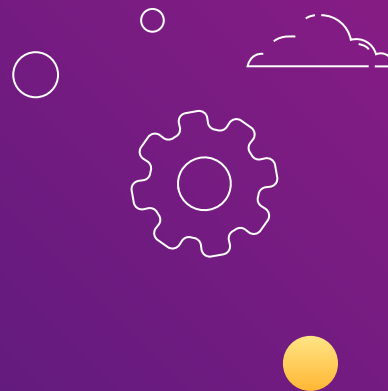
04 THE TEAM

OUR TEAM

HUNG-YU KO(ERIC)

JOSUE SANCHEZ

CARTER NIELSEN

THANKS!