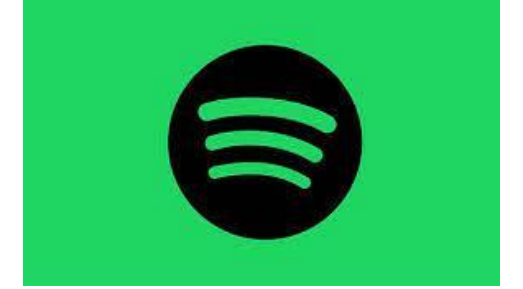Spotify

Agenda

1. About Spotify
2. Primary Goals
3. Research Question
4. Analysis - Summary
5. Limitations and Recommendation

ABOUT SPOTIFY

Spotify is a proprietary audio streaming application for millions of music and other content from all over the world. It is a digital music, podcast, and video service that gives you access to millions of songs and other content from creators all over the world.

One can find millions of songs from a variety of genres and artists: from obscure indie rock to top 40 pop, to movie soundtracks and classical music. It also has a complex algorithm to recommend music based on your listening history, as well as curated playlists and internet radio stations
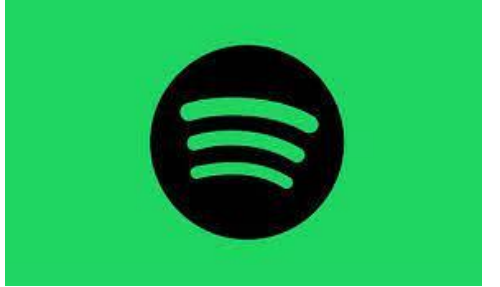
PRIMARY GOALS

Determine the factors that influence:

- **The ranking of top artists**

- **The recommendation of top artists on Spotify**
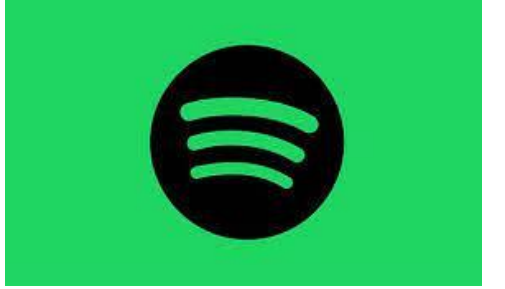
DATASET

- The data source we are analyzing is through Kaggle and the data is originally from Spotify web API.

- The dataset includes 169k songs taken from the year 1921 to year 2020 and is organized into 19 columns.

- This data set contains from basic data such as track id, artist name to sophisticated data such as tempo and energy.

We are focusing primarily on these important data sets such as

- Artist (Artist Name)

- Name (Song Name)

- Popularity (popularity index)

- Energy

- Danceability

- Duration_MS

- Liveliness

- Speechiness

- Key

- Valence

METHODOLOGY

DATA PREPROCESSING

• We first transformed the data by altering the main data set, data.csv file, by deleting unnecessary columns that we would not end up using.

• Artist's data column had invalid values and those values have been replaced/ removed

• 2159 duplicates were found and removed based id and we have 172230 unique values

DATA SELECTION & TRANSFORMATION

• Data selection and transformation process was done post the cleaning by creating popularity bin, ranking, correlation as specific columns and are converted from dimensions to measures.

• Normalization of data at this point is not required of analysis based on research questions.

• We used Excel as well as Tableau to filter, cleanse, and visualize our data to draw conclusions

METHODOLOGY - FORMULAS

Key Formulas Used

- Ranking Function (To rank Artists) = RANK(AVG([Popularity]))

To find Correlations between a measure and Poularity
- Danceability CORR = CORR([Popularity],[Danceability])
- Duration(ms) CORR = CORR([Popularity],[Duration Ms])
- Instrumentalness CORR= CORR([Popularity],[Instrumentalness])
- Liveness CORR = CORR([Popularity],[Liveness])
- Loudness CORR = CORR([Popularity],[Loudness])
- Key CORR = CORR([Popularity],[Key])
- Speechiness CORR = CORR([Popularity],[Speechiness])
- Valence CORR = CORR([Popularity], [Valence])

Bins Used

- Popularity Bin (Values taken from 0 to 10)

RESEARCH QUESTIONS

Based on Popularity of Artists

- How does Popularity of the songs affect the Artists?

- How do we define what artists fall under what range of the Popularity Index?

- What artists are ranked in the top based on popularity index?

Based on Correlation:

- Is there a correlation between valence, Danceability, Energy, Key, Liveness, duration_ms, Loundness, Speechiness, Instrumentalness with Popularity?

- Which factors have the highest influence on popularity?

RESEARCH QUESTIONS

Based on Danceability

- How does danceability to the song Varies with Valence (Musical positivity conveyed by a track)?

- Does danceability alone is a measure to track changes in Popularity of the songs?

Based on Past Trends:

- How do we rank the top 10 artists based on popularity in 2021? Do the top trends differ from the past years? How does the ranking of top artists differ for the past 10 years?

- Is an artist repeatedly seen on the top charts for two consecutive years? How can we recommend top artists based on the top artists trends from past years?
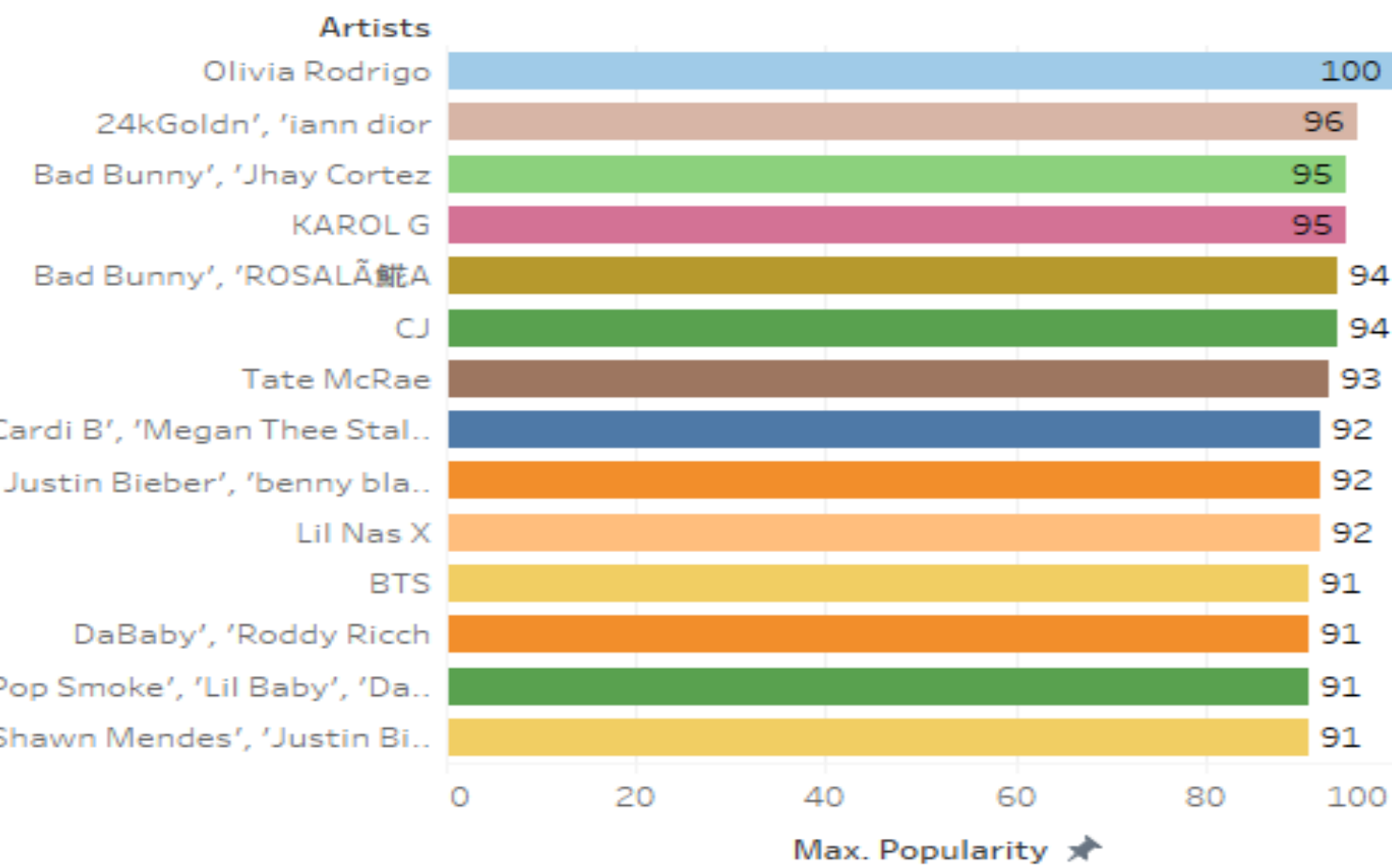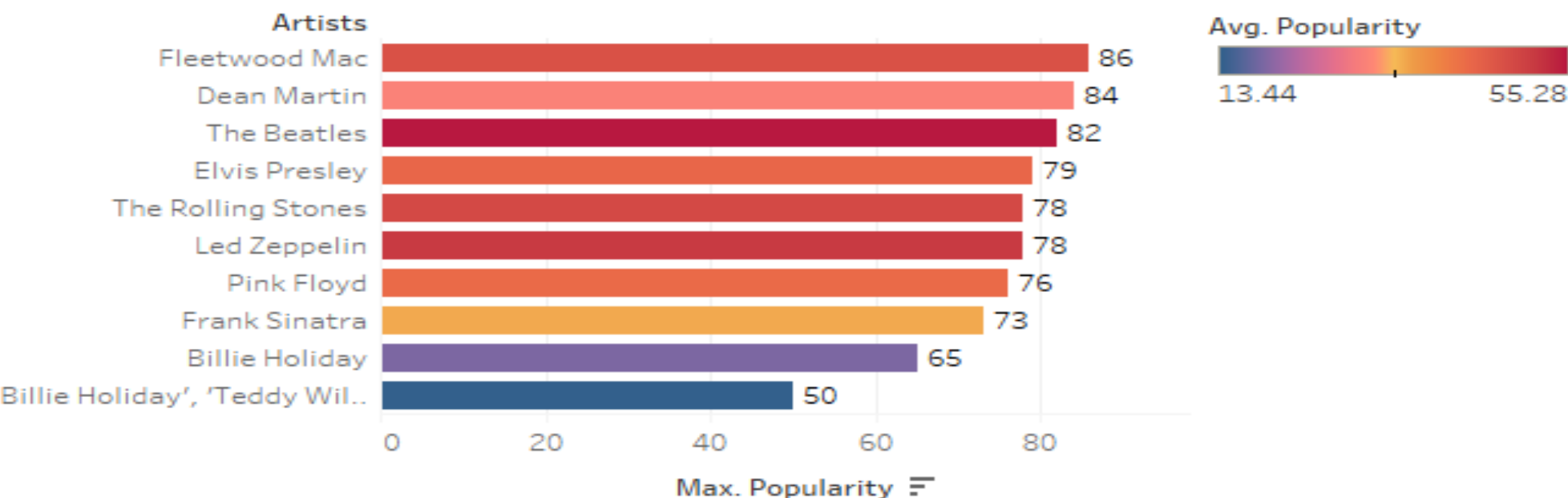
ANALYSIS – BASED ON POPULARITY

Top Artist List

Popularity (..	Artists	
90	Ariana Grande	96
	24kGoldn', 'iann dior	96
	KAROL G	95
	Bad Bunny', 'Jhay Cortez	95
	The Kid LAROI	94
	CJ	94
	Billie Eilish	94
	Bad Bunny', 'ROSALÃ鱿A	94
	Tate McRae	93
	Pop Smoke	93
	Lil Nas X	92
	Justin Bieber', 'benny bla..	92
	Justin Bieber	92
	Cardi B', 'Megan Thee Stal..	92
	SZA	91
	Sia	91
	Shawn Mendes', 'Justin Bi..	91
	Pop Smoke', 'Lil Baby', 'D..	91
	Harry Styles	91
	DaBaby', 'Roddy Ricch	91
	Conan Gray	91
	BTS	91
	MEDUZA', 'Dermot Kenne..	90
	Mariah Carey	90
	Lewis Capaldi	90
	Justin Bieber', 'Chance th..	90
	Internet Money', 'Gunna', ..	90
	Ed Sheeran	90
	Dua Lipa', 'DaBaby	90
	Bad Bunny	90
100	Olivia Rodrigo	100

Popular artists

Artists	Max. Popularity
Olivia Rodrigo	100
24kGoldn', 'iann dior	96
Bad Bunny', 'Jhay Cortez	95
KAROL G	95
Bad Bunny', 'ROSALÃ鱿A	94
CJ	94
Tate McRae	93
Cardi B', 'Megan Thee Stal..	92
Justin Bieber', 'benny bla..	92
Lil Nas X	92
BTS	91
DaBaby', 'Roddy Ricch	91
Pop Smoke', 'Lil Baby', 'Da..	91
Shawn Mendes', 'Justin Bi..	91

Artists
- Olivia Rodrigo
- 24kGoldn', 'iann dior
- Bad Bunny', 'Jhay Cor...
- KAROL G
- Bad Bunny', 'ROSALÃ...
- CJ
- Tate McRae
- Cardi B', 'Megan Thee..
- Justin Bieber', 'benny..
- Lil Nas X
- BTS
- DaBaby', 'Roddy Ricch
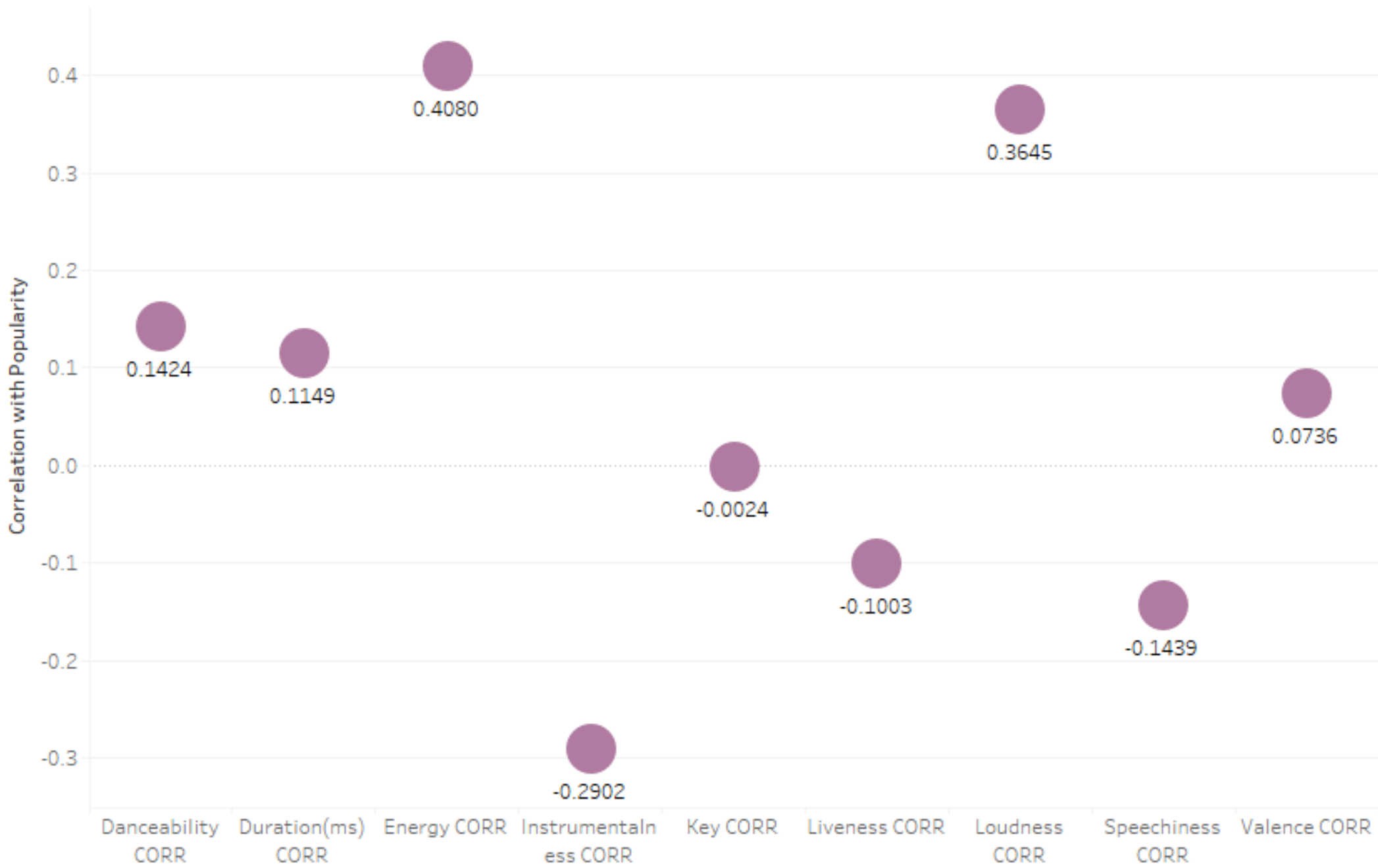- Pop Smoke', 'Lil Baby'..
- Shawn Mendes', 'Just...

Top artists of all time

Artists	Max. Popularity
Fleetwood Mac	86
Dean Martin	84
The Beatles	82
Elvis Presley	79
The Rolling Stones	78
Led Zeppelin	78
Pink Floyd	76
Frank Sinatra	73
Billie Holiday	65
Billie Holiday', 'Teddy Wil..	50

Avg. Popularity
13.44 — 55.28

ANALYSIS – BASED ON CORRELATION

Correleation between Popularity & other measures

Factors have the highest influence on "Popularity"

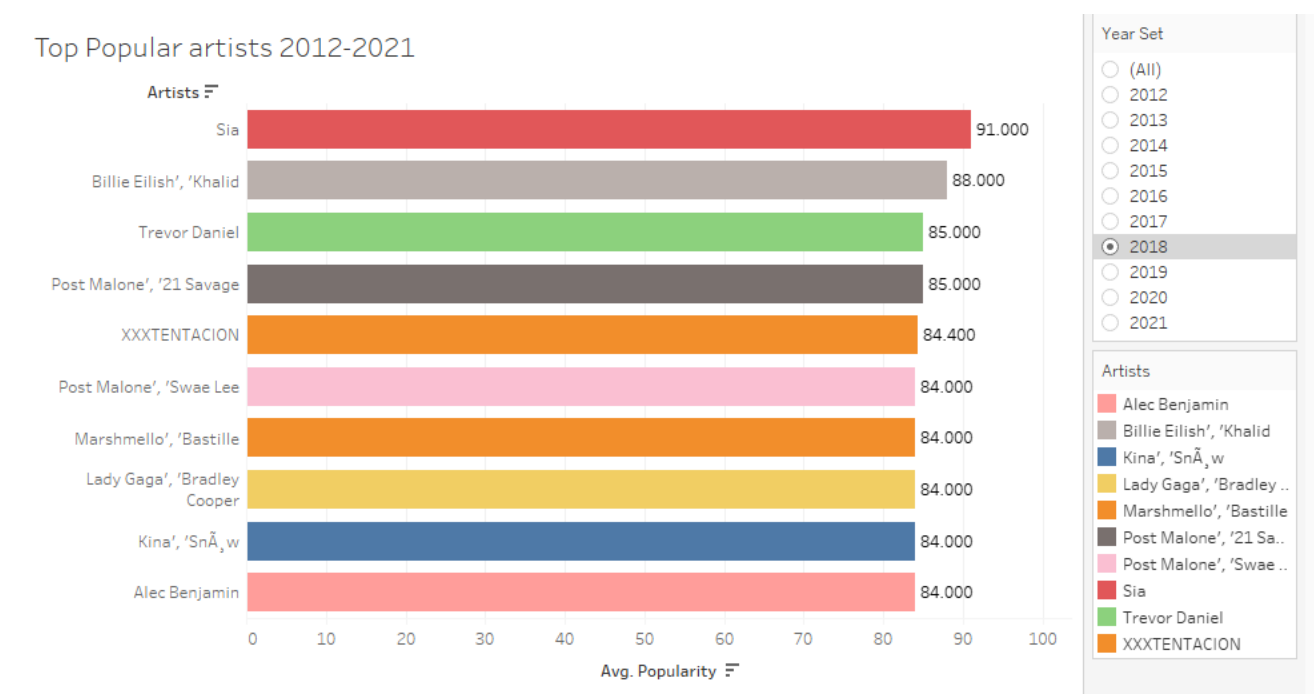|	Correlati..
Danceability CORR	0.1424
Duration(ms) CORR	0.1149
Energy CORR	0.4080
Loudness CORR	0.3645

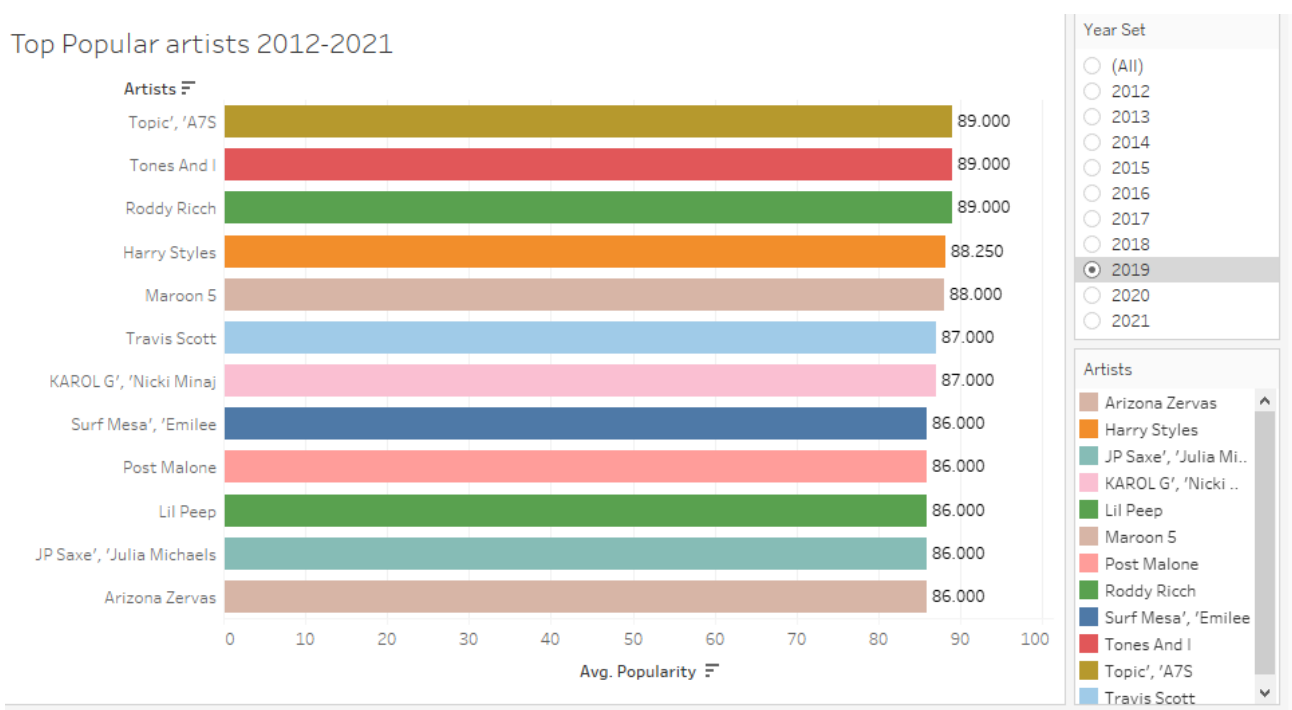# ANALYSIS – BASED ON DANCEABILITY
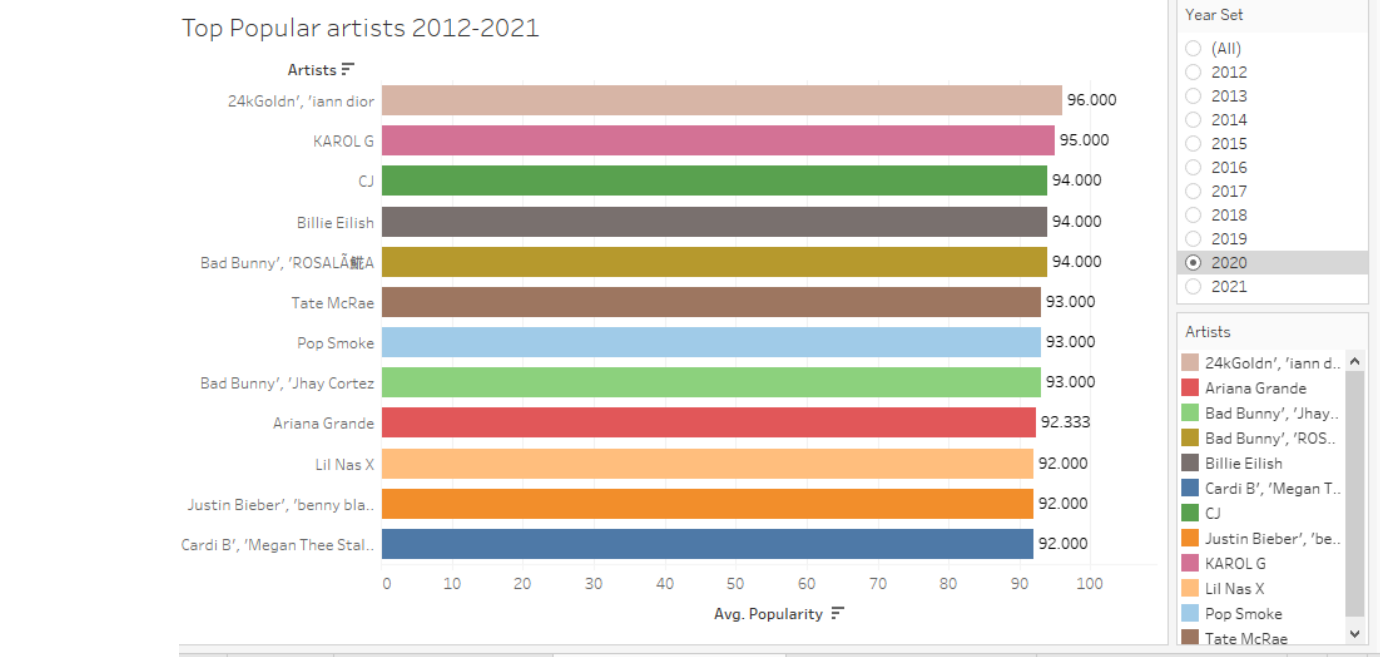
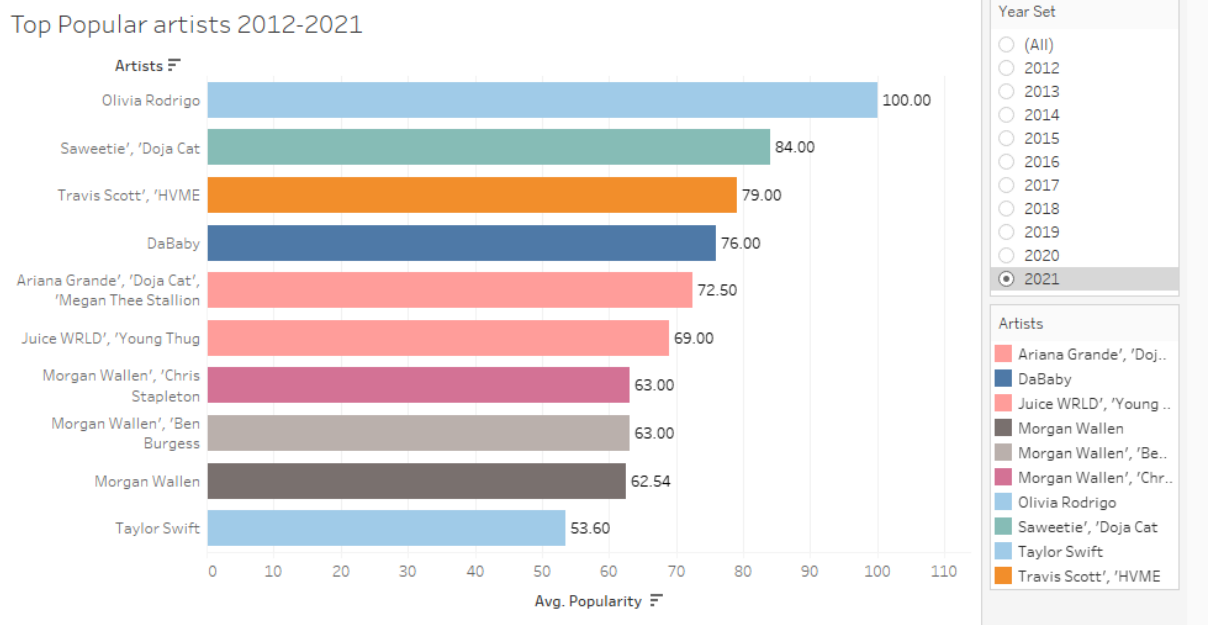# ANALYSIS – BASED ON PAST TRENDS

Top Artists for 2018

Top Artists for 2019

Top Artists for 2020

Top Artists for 2021

SUMMARY

Based on Popularity of Artists

- The analysis shows top 10 artists based on popularity. From the analysis it is evident that Popularity as a measure affects the artists significantly as it drives the artists to top the charts in Spotify.

- Artists can be grouped into a distribution bins from 0-100 with the interval of 10. As we have taken account of Top 10 artists , we can all the top artists with popularity index from 90 to 100.

- Based on the popularity, artist Olivia Rodrigo is considered as the top-most ranked artist. Based on all time popularity, bands and artists like Fleetwood Mac, Dean Martin and The Beatles are on top of the charts.

Based on Correlation:

- There is correlation between popularity and other measures. Some of the measures have positive correlation such as Danceability, Loudness, Duration_MS, Energy. There is negative correlation between variables such as Key, Liveliness, Speechiness and Instrumentalness with Popularity.

- Danceability, Loudness, Duration_MS, Energy are some of the measures or factors that have the highest influence on the Popularity of artists.

SUMMARY

Based on Danceability

- Danceability does have a correlation on Valence. Higher the Danceability, Higher the Valence. A valence score of 1 describes happiness whereas a valence score of 0 describes sadness or depressing mood of the song

- Danceability alone is not a good indicator for changes that affect the popularity index as its R-Squared value is very low. We need more measures to see the correlation.

Based on Past Trends:

- Top artists are ranked based on the popularity (Popularity index) as the measure. On analysing the past trends of Spotify data from 2012 to 2021, we can say that the top artists change every year. The popularity index of the top artists has increased from 2012 to 2021.

- One Direction (2012 -2015), Justin Bieber(2015,2016 and 2020) and Ed Sheeran (2014, 2017) are some of the top artists charting repeatedly twice and more in the past decade.
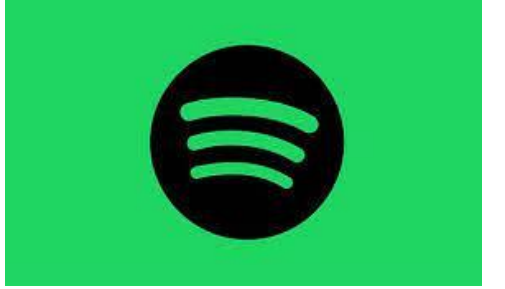
LIMITATIONS

- Our project had its limitations in terms of the dataset we used. We did have the data for the number or percentage of listeners in the data and hence we cannot comment on how the popularity can influence the listenership of the songs.

- The data set provided is from 1920-2021 and the historical data is complex to analyze all the artists from the yester years to determine the top artists in the recent years considering there the trend of the top artists is changing every year and many upcoming artists are grabbing the top charts. It becomes difficult to give a series or a trendline for the past years.

- There can be other factors too that take account for Popularity Index of the songs and hence there is not just one measure enough to track changes in Popularity.
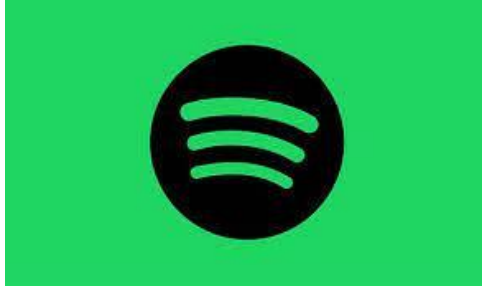
RECOMMENDATION

Based on our analysis from our project, we would like to recommend Spotify to help their Artists improve certain Key factors that influence their artist ranking. Based on our analyses, we can say that the artists / band should have songs with high valence, more danceability to their tracks and their songs are for longer duration as compared to short ones. The energy of their songs also play a key role in affecting the popularity of the artists. By bringing in more young and budding artists whose energy levels are great such as that of Olivio Rodirgo, we can see these artists get ranked on the top and since popularity index has been increasing every year for artists and has reached 100, it is possible that artists can get great platform on Spotify to brand themselves. Ranking the artists definitely helps in deciding the quality of top artists and also help increase their audience. This in turn helps Spotify gain traction and more net promoter score as a brand before they become a completely premium audio streaming service like Apple Music.