



What's popular on Spotify?

Group 2

- Maya Zhao
- Kelsey Liu
- Mariam Adeyemo
- Suenkei Chan
- Han Yang



Executive summary

- As more and more people switch away from traditional music platforms (such as radio and CD player), Spotify, the digital music streaming giant, is looking to grow their subscriber base by understanding and predicting what kind of content is going to be popular.
- This project will analyze the music available on Spotify, including artist, albums, tracks, market, genre, lyrics, etc.





Business Use Case



- **Why your project/analysis is needed**

Spotify wants to find the next favorite music. Therefore, our analysis will help Spotify understand the various music trend around the world and the track popularity. This insight will help in deciding what to invest on next.

- **What stakeholders want to know.**

Stakeholders want to know which genre has the highest popularity, which music are similar, which artist has the highest followers, and which music has the highest market availability.

- **Brief description of proposed analyses/reports to accomplish the business case.**

The business case will be accomplished through the analysis of music popularity, artist with the highest popularity, the average market availability per track/album.

Data Source

- **Data Source:**

<https://www.kaggle.com/datasets/saurabhshahane/spotgen-music-dataset>

- **Spotify database:**

SpotGenTrack Popularity Dataset

The data was collected from Spotify and Genius throughout their respective API's.



Database & Platforms

❑ Spotify database:

- **773.84 MB:** Separated csv files with the corresponding information of the following elements: Tracks, Artists and Albums.
- Album ranges from 0 to 75,510, without gaps. Artist ranges from Juliano Cezar to Something For Kate.
- Attributes including artist, album type, music title, available markets, country, release date, track popularity, etc.

❑ Tools:

- Python
 - data cleaning: treating the nulls, treating the duplicates
 - pre-processing: dropping, feature engineering
- MySQL
 - normalization until third normal form (3NF)
 - denormalization for populating desired tables
- Tableau
 - exploratory data analysis
 - visualization



Data Preparation & Cleaning

❑ Missing Data

- Drop rows with missing values for artist names

❑ Drop Columns

- Drop columns: 'external_urls', 'href', 'images', 'track_name_prev', 'uri' etc.

❑ Feature Extraction

- 'release_date' to 'release_year'

❑ Normalization

- 3rd degree of normalization
- 8 tables

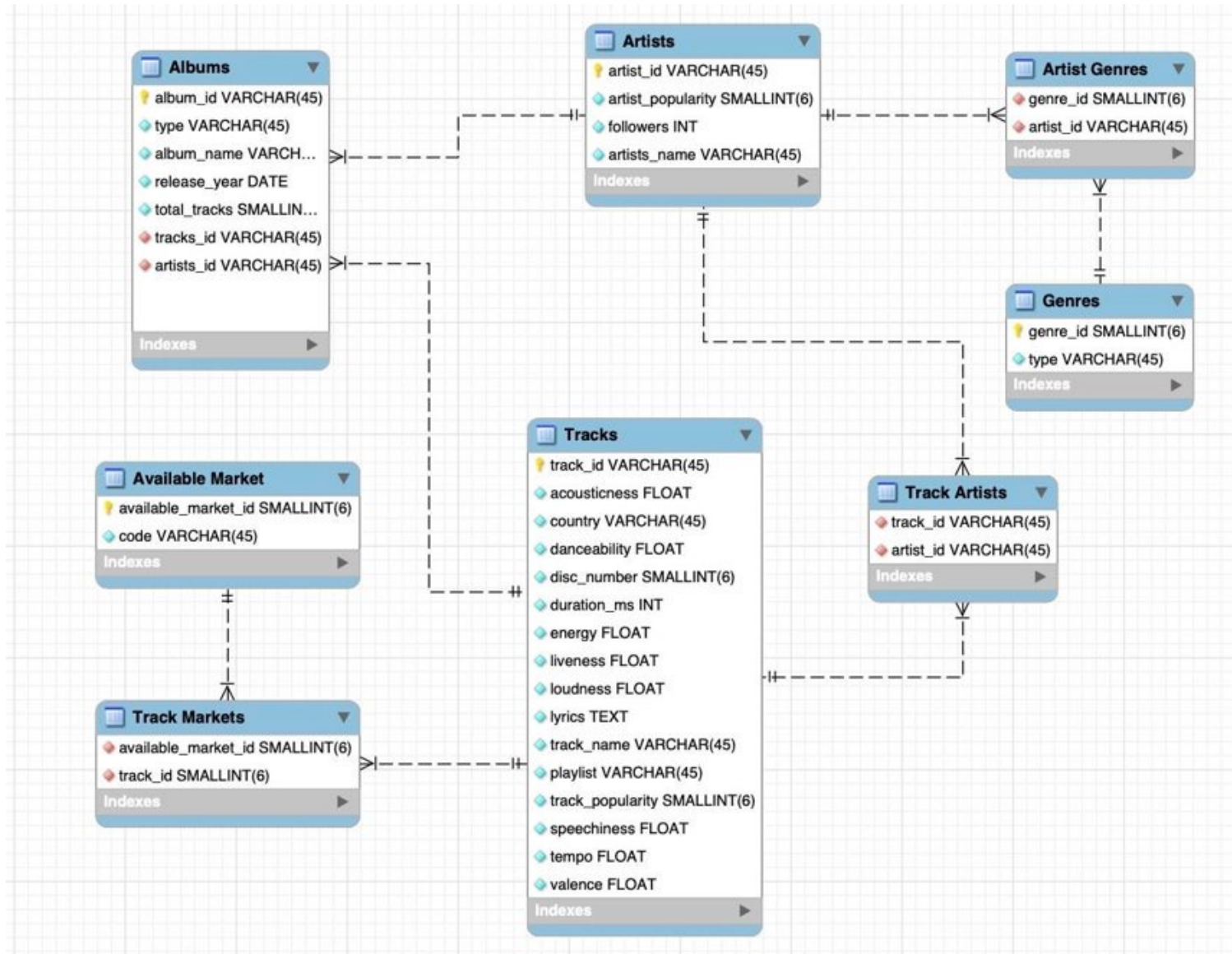


Assumptions

- ☐ An artist can release many tracks
- ☐ A track can feature many artists.
- ☐ A track can be available in many markets.
- ☐ A market can provide many tracks.
- ☐ An artist can be categorized in many different genres
- ☐ A genre can have many artists.
- ☐ An album can have only one track or many tracks
- ☐ A track can only be in one album
- ☐ An artist can have many albums
- ☐ An album can only be owned by one artist



EER Diagram



Descriptive Stats & Data Visualization



Artists
56,119



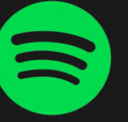
Tracks
101,734



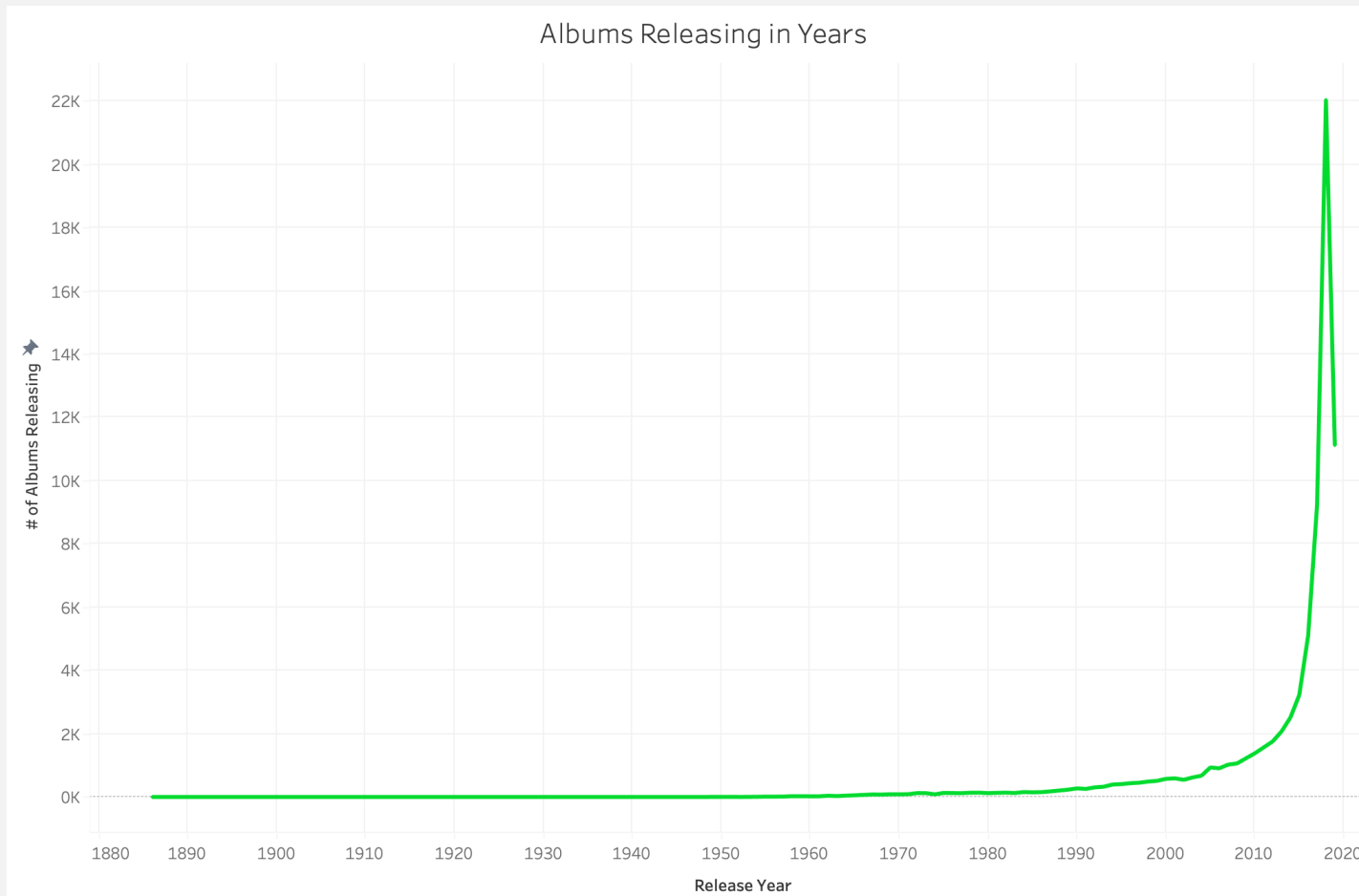
Albums
75,510



Albums Releasing in Years



Albums releasing reached peak in 2018 and dropped dramatically in 2019

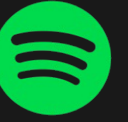


Insights & Recommendations

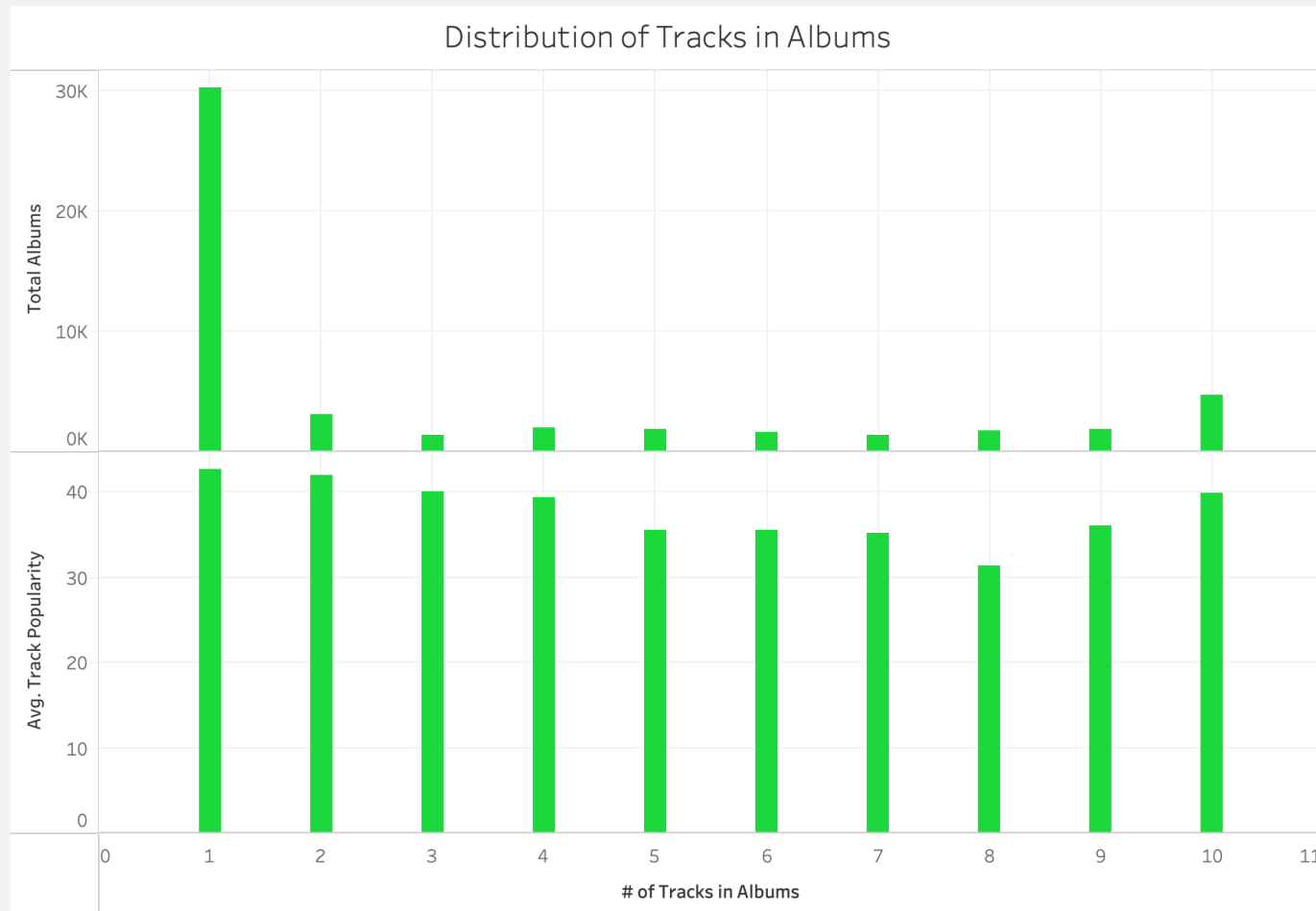
- From year 2000, albums releasing start to grow
- Year 2018 is a golden year for artists to release music
- The dramatic drop is probably due to Covid time
- In post-Covid era, a great potential for music industry to continue growing



Distribution of Tracks in Albums



No strong correlation between popularity and # of tracks in albums



Insights & Recommendations

- Most popular album type for artist to lease is single
- Musical platforms like Spotify makes it easy for artists to release single album
- Number of tracks in albums doesn't affect track popularity that much



- Songs about love have huge potential audience base
- Understanding is another key to consider when writing lyrics
- 'oh' is the most used modal particle in lyrics showing the emotion with tone, or helping with the transitions between bars



- High, tired, love & talk are the heated lines
- People today love songs expressing their feelings
- Using aggressive words is one of the way for people to relieve emotional stress

Top 10 English Tracks - Popularity

Tracks Name	Popularity
Sweet but Psycho	97.000
Talk	95.000
High Hopes	95.000
MIA (feat. Drake)	95.000
i'm so tired...	95.000
Going Bad (feat. Drake)	95.000
Giant (with Rag'n'Bone M.)	95.000
SICKO MODE	94.000
Murder On My Mind	94.000
ZEZE (feat. Travis Scott & ...)	93.000

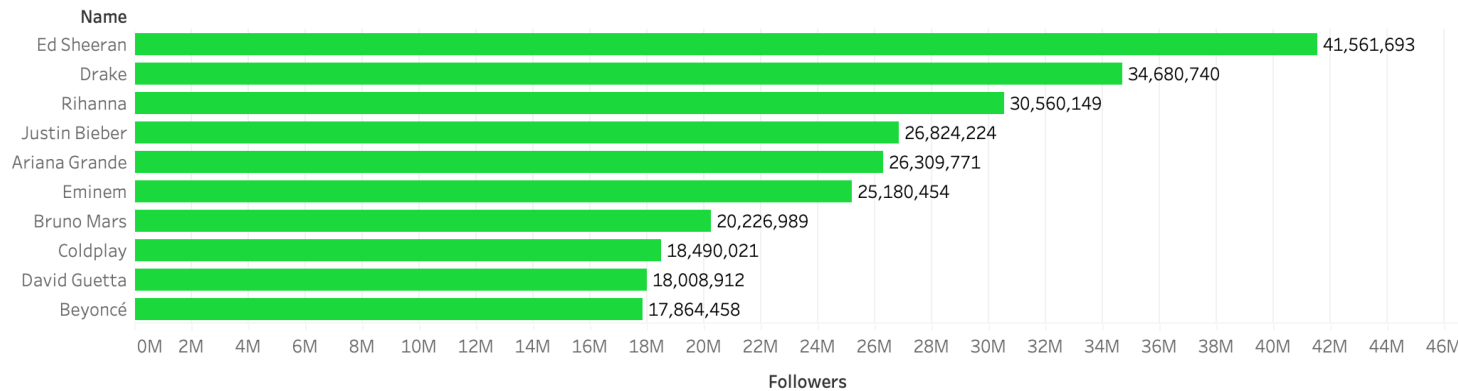


Most popular artists

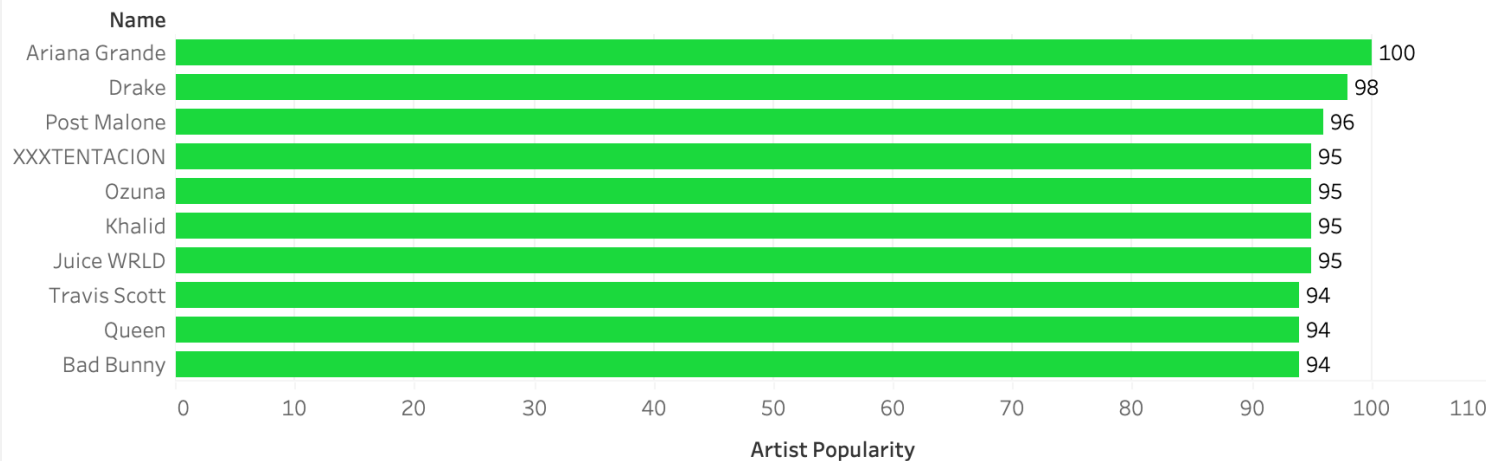


Most popular artists by followers differ from those by popularity.

Most Popular Artists - Followers



Most Popular Artists - Popularity

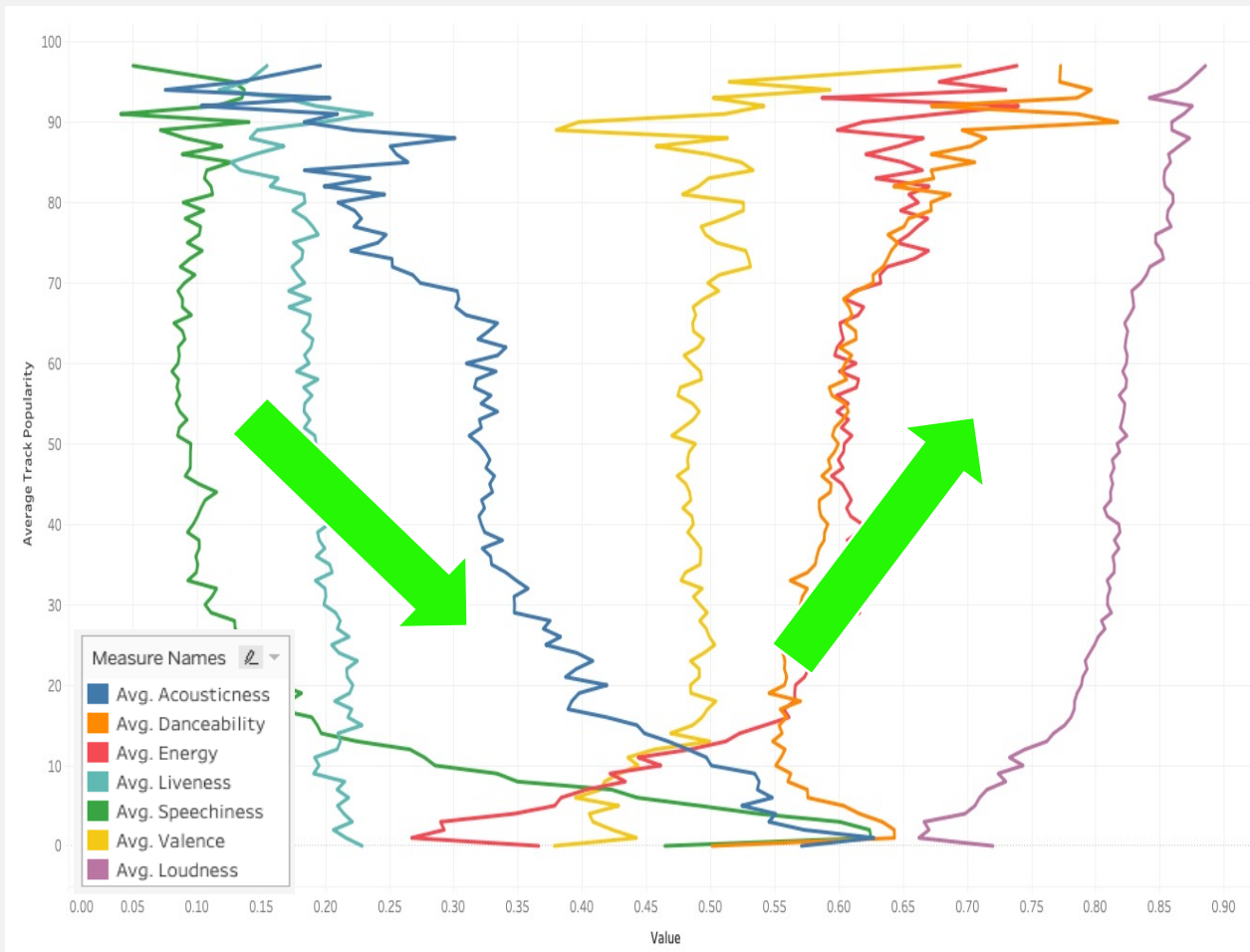


Insights & Recommendations

- Ed Sheeran is not on the list of most popular artists by popularity (how many listeners plays their music) but gets most followers.

- Increasing the exposure of Artists with less followers but more popularity since they have high potential to grow.

People prefer to listen more positive and energetic tracks



Plot illustration

1. X-axis ranges from 0 to 100
2. Y-axis ranges from 0 to 1
3. Danceability, energy and valence have positive relationships with popularity.
4. Acousticness, speechiness have negative relationships with popularity.



Insight

According to the relationships, there is a high probability that the next big thing will be positive and optimistic with little acousticness and speechiness.

- *Speechiness: the presence of spoken words in a track.*
- *Valence: the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)."*



Genres with Most Artists

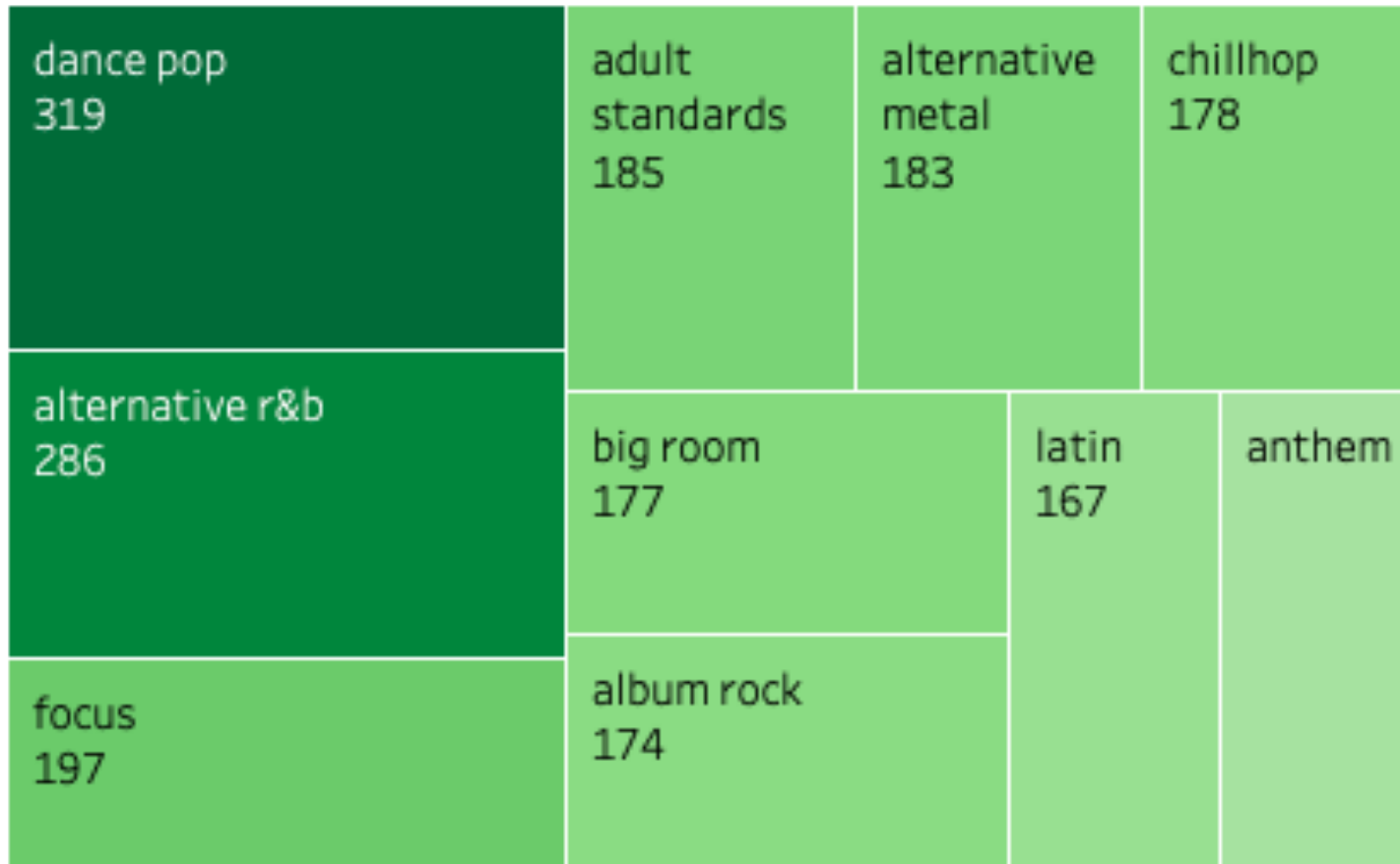


Artists tend to release songs that are energetic and rhythmic.



Insights

- Major artists genre: Dance Pop, Alternative R&B.
- Artists also like to release tracks that help audiences stay focused on their tasks.
- This can be resulting from audience preference, since it aligns with our analysis on tracks popularity.



Recommendation



Artists

Instead of just relying on popularity of artists, Spotify can take advantage of the number of followers of artists as well and use an algorithm to recommend the artist songs to their fans to better keep them engaged.



Tracks

Spotify can segment tracks based on their characteristics (such as valence, energy etc.) and promote more tracks that are positive and energetic as it tends to be the most listened songs.



Albums

Single album will be a good starting point for upcoming musicians as audience preference is not affected by album type.



Lessons Learned



- Getting the library of the dataset at the onset really helps to better understand the various attribute of the different tables and their relationship.
- It was a bit challenging to import the dataset into MySQL due to the encoding style of our personal system.
- Working on this project helped us to put together what we have learnt and apply it to the exploration of the data.



Thanks!

Q&A