

Assignment 1

Group 15:
Slowak Sebastian
Zimball Markus
Kuznetsova Ekaterina
Alla Lundrim





Topic and data selection



- 30000 Spotify Songs:

https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs?select=spotify_songs.csv

- Spotify Charts:

<https://www.kaggle.com/datasets/dhruvildave/spotify-charts>

- Spotify Playlists:

https://www.kaggle.com/datasets/andrewmvd/spotify-playlists?select=spotify_dataset.csv



Topic and data selection: reasoning

Why work with music data?

- Music is an integral part of our lives. Knowing what kind of music to use to appeal to the largest amount of people is quite powerful. Regional and temporal differences make music selection not trivial.

Why choose these spotify datasets?

- largest streaming platform in the world
- large and detailed datasets
- 2 out of 3 datasets work with information formatted like Spotify's public API responses, meaning it is quite close to current real world data



Analytical questions

1. What are the most popular songs on Spotify? And what are their key features?

Benefits: By identifying the key features of popular songs, we can understand what resonates with listeners, helping artists and record labels tailor their future releases.

2. Is there a correlation between certain song attributes (e.g., danceability, energy, tempo) and their performance on the Viral 50 charts?

Benefits: Investigating correlations between song attributes and chart performance yields valuable insights into the drivers of viral success. This data can empower artists and producers to refine their music production process, boosting their chances of chart success. Additionally, analyzing viral songs by region offers insights tailored to specific geographic markets, enhancing strategic decision-making for targeted regional promotions and engagements.

3. How does the genre classification of a playlist relate to the song that's included, and what are the key features of the song that are relevant for its genre classification within that playlist?

Benefits: Understanding the relationship between song features and genre classification within playlists can improve playlist curation algorithms. This can enhance user satisfaction by offering more personalized recommendations, leading to longer user's session durations on music streaming platform.



Data Modeling

Three data sets

- 30000 Spotify Songs
- Spotify Charts
- Spotify Playlists

turned into:

- Facts Table
- Dimensions (Artists, Spotify Playlists, User Playlist, Songs, Top 200, Viral 50)



ER-model of the original data

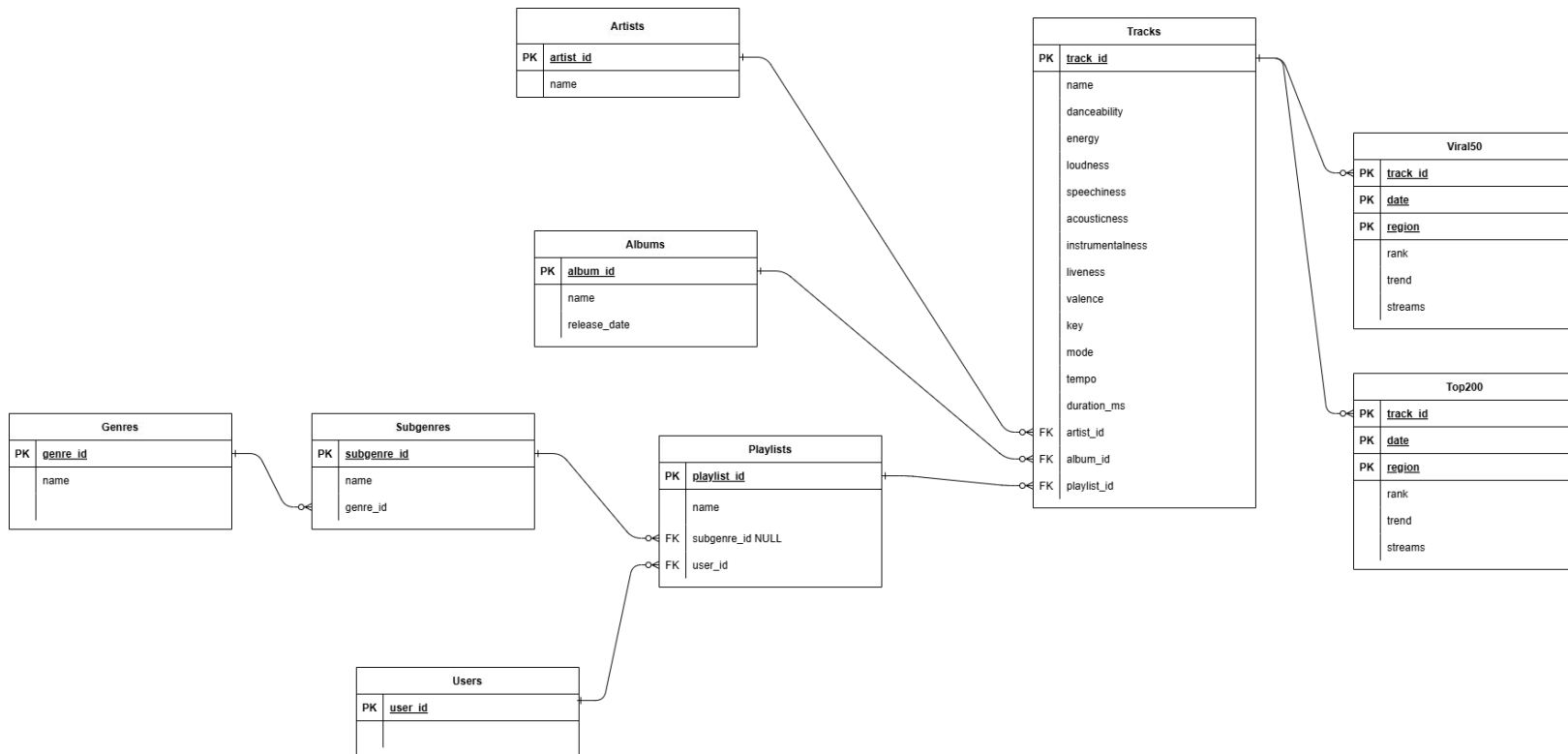
| Playlists | |
|-----------|---------------|
| | artist_name |
| | track_name |
| PK | playlist_name |
| PK | user_id |

| Songs30000 | |
|------------|--------------------------|
| PK | track_id |
| | track_name |
| | track_artist |
| | track_popularity |
| | track_album_id |
| | track_album_name |
| | track_album_release_date |
| | playlist_id |
| | playlist_genre |
| | playlist_subgenre |
| | danceability |
| | energy |
| | key |
| | loudness |
| | mode |
| | speechiness |
| | acousticness |
| | instrumentalness |
| | liveness |
| | valence |
| | tempo |
| | duration_ms |

| Charts | |
|--------|---------|
| PK | title |
| | rank |
| PK | date |
| | artist |
| | url |
| | chart |
| | trend |
| | streams |
| | region |

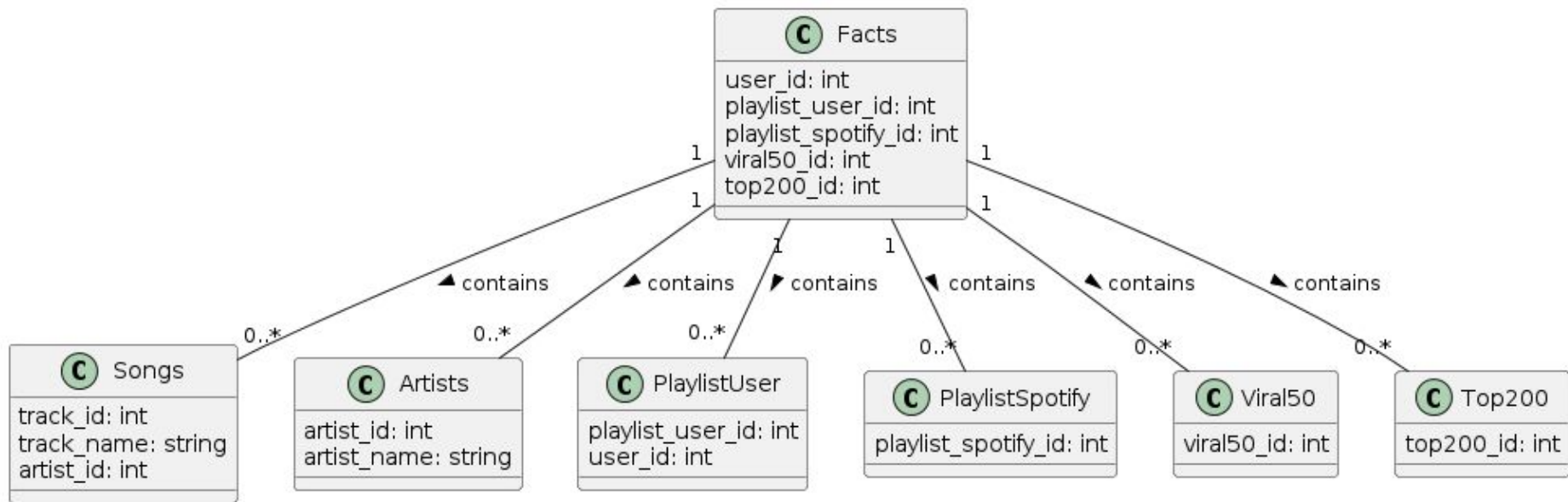


The normalized ER-model





Star schema



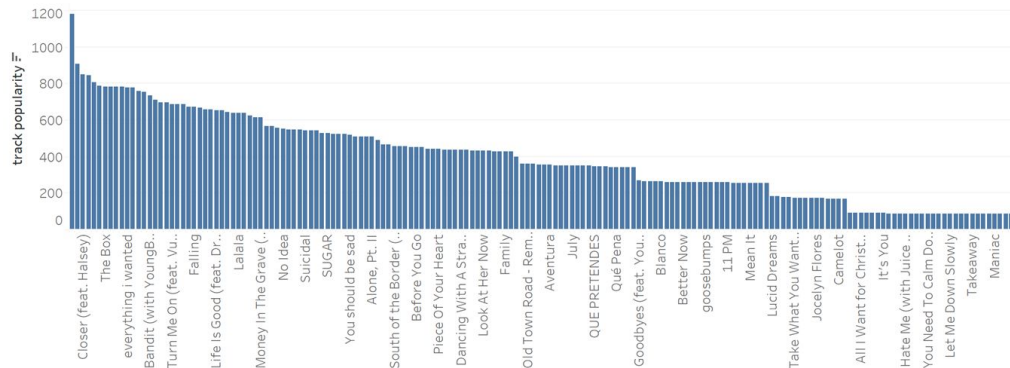


Data processing

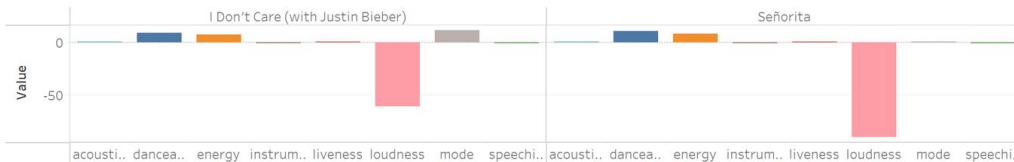
- The data processing is done with the three raw datasets.
- Columns are uniformly renamed.
- The Playlist and Chart datasets are filtered for songs existing in 30000-songs dataset by track name and artist name.
- Artist feature table is extracted.
- To reduce the data volume a chart snapshot every six months is maintained.
- The Playlist table represents playlists generated by Spotify and is extended with track identifiers (track_id).
- Charts data is split into 'Top200' and 'Viral50'.
- Facts table is created by considering valid feature combinations.

Tableau Dashboard

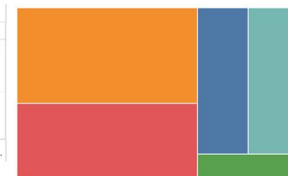
Track Popularity



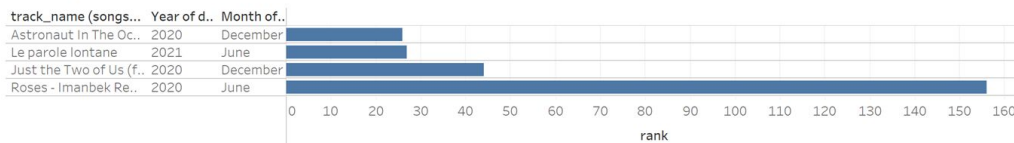
Track features



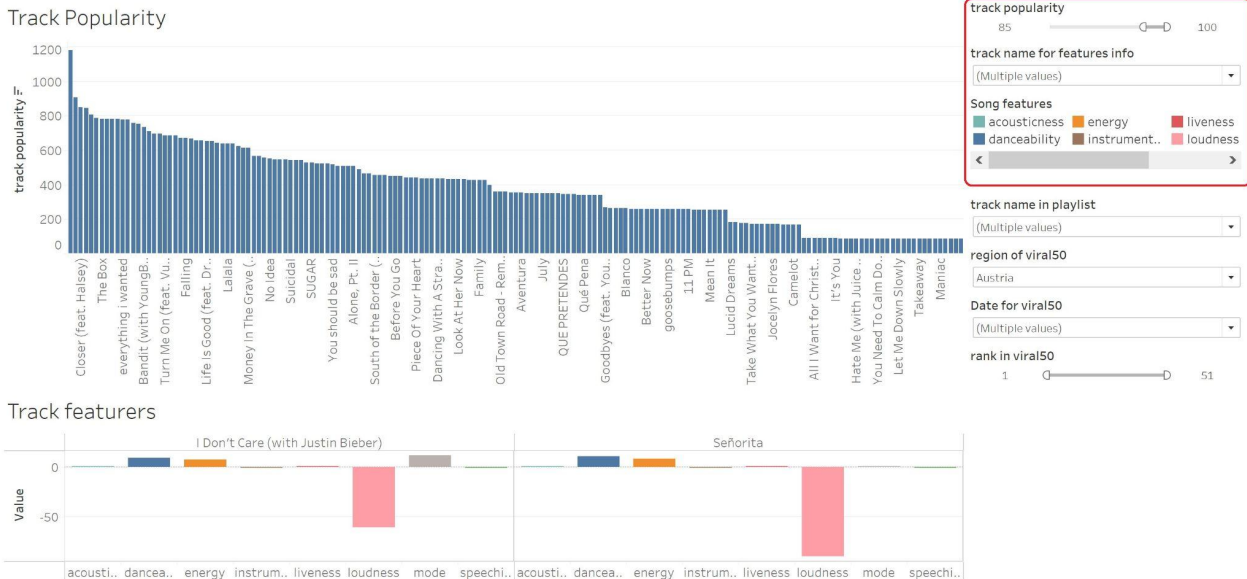
Playlist genres



Charts



Using the Track Popularity bar chart, you can obtain information about the most popular songs based on Spotify's popularity metric. The data can be filtered by popularity value. Then, you can select popular tracks in the filter to gather information about their features such as acousticness, danceability, energy, instrumentalness, liveness, loudness, mode, and speechiness.

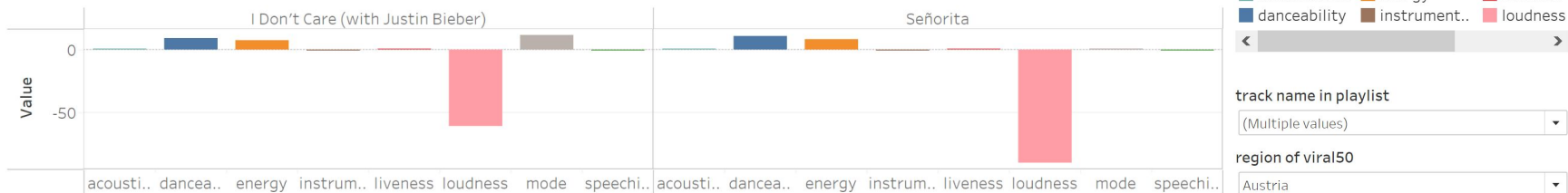




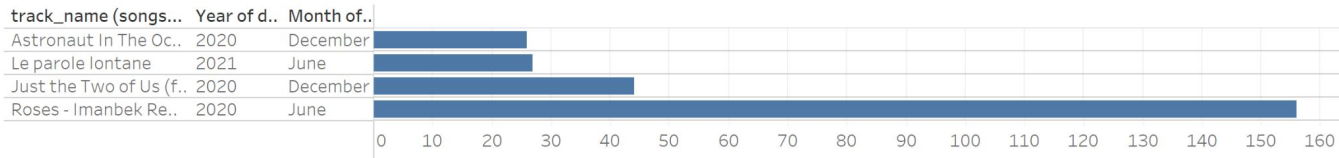
Data analytics. Question 2

Using the Charts bar charts, you can discover information about the most viral songs in specific regions and years. From there, you can select popular tracks using the filter to gather detailed insights about their features, including acousticness, danceability, energy, instrumentalness, liveness, loudness, mode, and speechiness.

Track featurers



Charts



Song features

track name in playlist
(Multiple values) ▼

region of viral50
Austria ▼

Date for viral50
(Multiple values) ▼

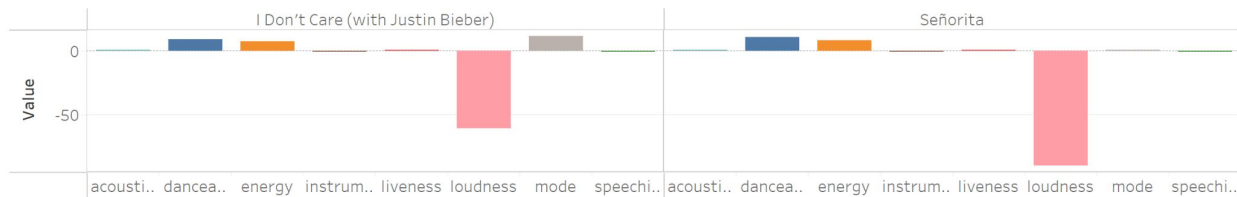
rank in viral50
1 ————— 51



Data analytics. Question 3

Using the Playlist genres treemap, you can access information about the genres of playlists that include the song. This can help identify the main genre of the song. Additionally, you can delve deeper into details using the Track features bar chart to understand the song's key features.

Track features



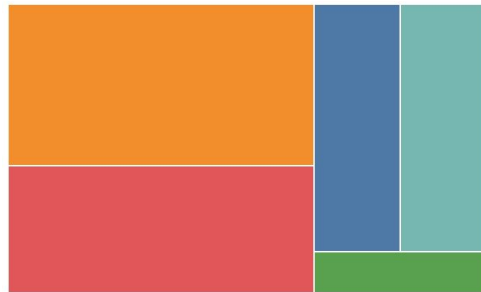
track name for features info

Song features

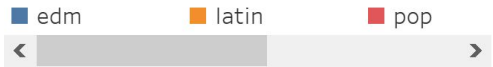


track name in playlist

Playlist genres



playlist genre





Reflection

- Dashboard covers all analytical questions.
- More powerful machines could have been advantageous as the data has high resolution.
- Drawing real time data can be interesting as the popularity score changes
- The dashboard could be improved by connecting several sheets so that when you choose an element (for example, a song) on one sheet, the other chart will update automatically.



Work distribution

- Data selection: Sebastian, Markus, Ekaterina, Lundrim
- Create analytical questions (AQ): Sebastian, Markus, Ekaterina, Lundrim
- Design ER-models: Markus, Ekaterina
- Star Schema: Lundrim
- Data processing: Sebastian, Markus, Ekaterina, Lundrim
- Tableau visualization: Ekaterina, Sebastian
- Data analytics (link dashboard and AQ): Sebastian, Markus, Ekaterina, Lundrim
- Report: Sebastian, Markus, Ekaterina, Lundrim