



# SPOTIFY ANALYSIS

DATA 557 | WINTER 2022

GROUP MEMBERS ▾

- 👤 Aniket Fadia
- 👤 Hannah Luebbering
- 👤 Rebecca Klein
- 👤 Teju Kandula

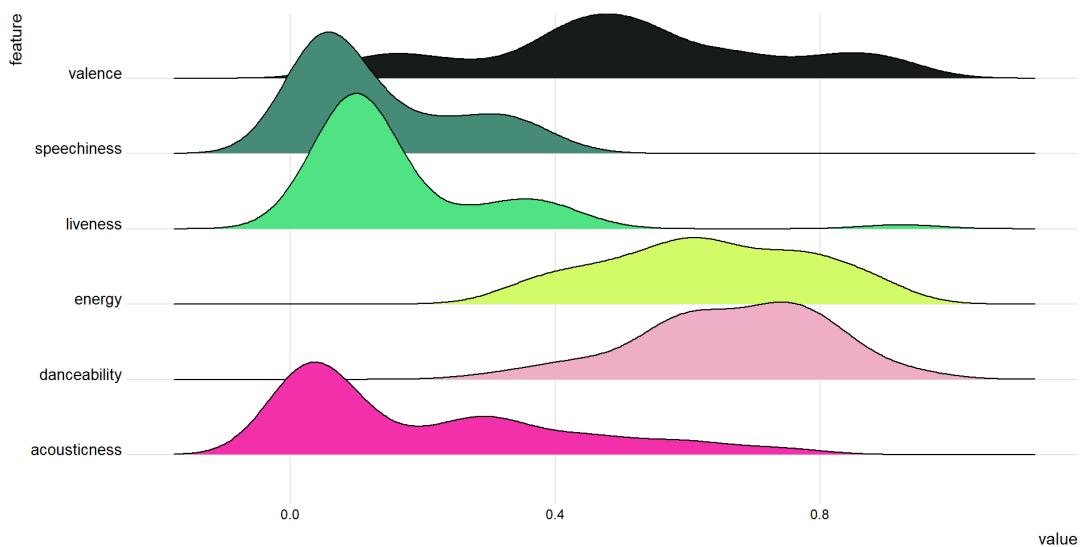
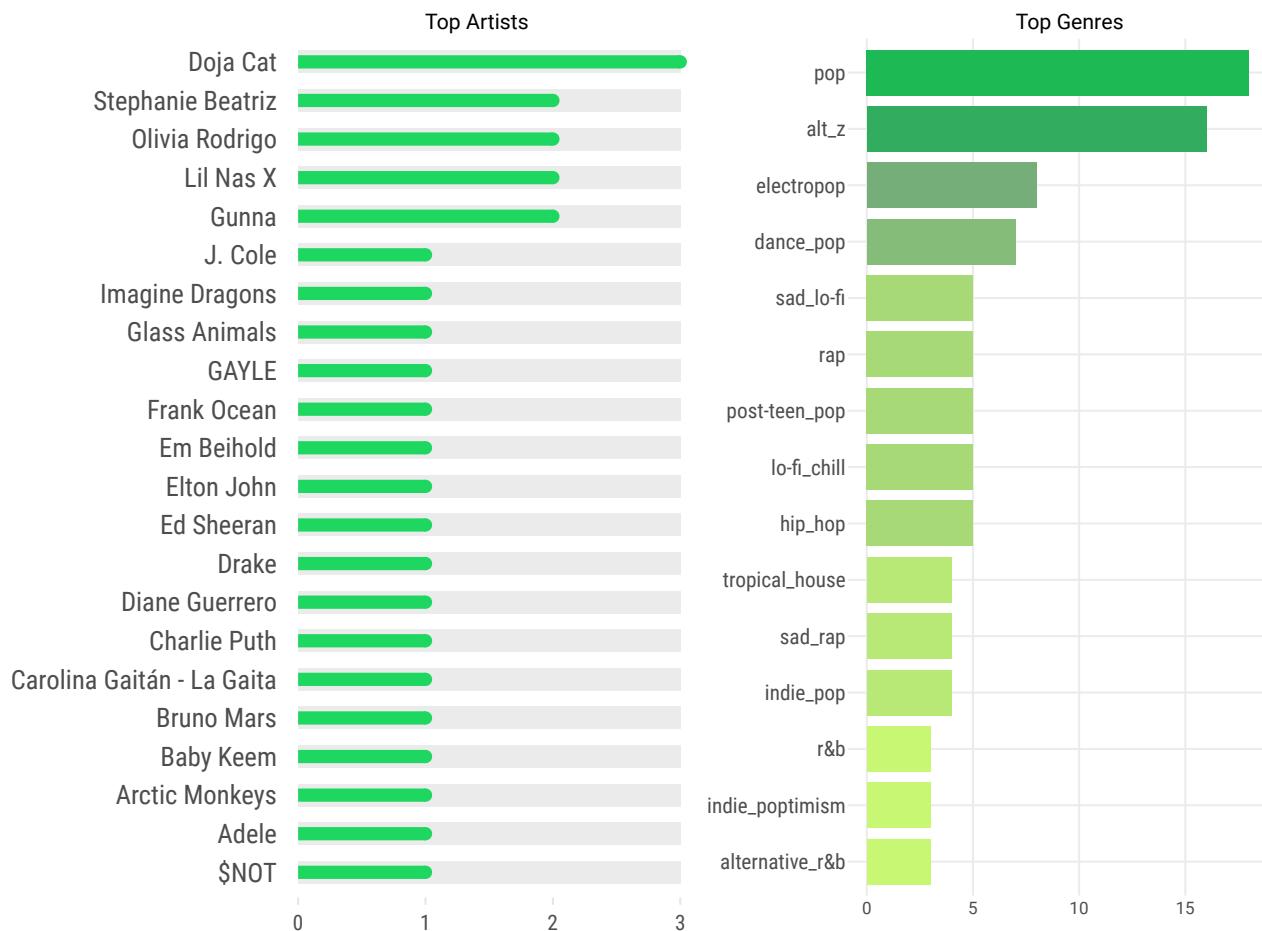
## DATA SOURCE:

### 1. Spotify Daily Top Tracks (SPOTIFY API)

- **Source:** Spotify API
- **Description:** shows the top 200 tracks per United States Spotify Charts each day from January 1, 2017 to December 31, 2021
- **Variables:** day , track\_id , rank , title , main\_artist , main\_artist\_id , streams , all\_artists , all\_artists\_ids , acousticness , danceability , duration , energy , instrumentalness , key , liveness , loudness , mode , speechiness , tempo , explicit , time\_signature , valence , genre .

The screenshot shows the Spotify mobile application interface. At the top, it displays "Top 50 - USA". Below this, there's a play button icon with "Top 50" and "USA" text. The main area shows a list of three songs with their titles, artists, and durations:

Rank	Title	Artist(s)	Duration
1	We Don't Talk About Bruno	Carolina Gaitán - La Gaita, Mauro Castillo, Adassa, Rhenzy Feliz, Diane Guerrero, Stephan...	3:36
2	Heat Waves	Glass Animals	3:58
3	pushin P (feat. Young Thug)	Gunna, Future, Young Thug	2:16



## QUESTIONS/HYPOTHESES:

- 1 Does time period change the popularity (proportion of streams) of pop, rap, hip hop, r&b, and rock genres in the Spotify US daily charts?
  - a Whether or not it is a weekday or weekend
  - b Whether or not it is a holiday
  - c Season (summer, winter, spring, fall)
- 2 Did the popularity of happy songs (mean valence) in the top 200 Spotify US daily streams change during Covid?
  - a Before March 13, 2020 (exclusive)
  - b After March 13, 2020 (inclusive)
- 3 What parameters are the most important in predicting the popularity on Spotify in the US?
  - a Popularity is calculated by summing  $201 - \text{Rank}$  (i.e. rank 1 has score 200, rank 200 has score 1) over all the days they are on the Spotify Top 200 streams

## METHODOLOGY:

### Question 1 (TIME PERIOD AND PROPORTION OF GENRES PLAYED)

- Use Spotify API to pull 2017-2021 US daily top track data
- Join Spotify daily tracks data with Artist, Genre, and Tracks tables
- Use two-sided two-sample large sample Z-test to answer subquestions a and b
  - Perform the above test for each of the following genres: pop, rap, hip hop, r&b, and rock
- Use ANOVA to answer subquestion c to test the difference in proportion of a genre depending on the seasons
  - Perform the above test for each of the following genres: pop, rap, hip hop, r&b, and rock

### Question 2 (HAPPINESS OF SONGS DURING COVID)

- Same dataset used in Question 2
- Use two-sided two-sample large sample Z-test to answer question

### Question 3 (PARAMETERS FOR POPULARITY)

- Popularity is calculated by summing  $201 - \text{Rank}$  (i.e. rank 1 has score 200, rank 200 has score 1) over all the days they are on the Spotify Top 200 streams
  - Use Linear Regression with all the features as a baseline model
    - Possible features: acousticness , danceability , duration , energy , instrumentalness , key , liveness , loudness , mode , speechiness , tempo , explicit , time\_signature , valence , genre
    - Predicted value: calculated popularity
  - Remove features one by one to improve the  $R^2$  score
  - Consider adding interaction effects and check the performance of the linear regression model. However, adding interaction effect will complicate the interpretability of the model
-