

# Exploring Spotify and Music Trends

## Overview

This report explores the **top artists on Spotify and how they compare** using the **Spotify API**, and **Music Trends** over time, which describes how things like *audio features*, *duration* or *tempo* have changed over various decades.

## Datasets and Obtaining Data

- We scraped monthly listeners' data from a table in *ChartMasters* website.
  - It contains a table of 1000 top artists and their monthly listeners ranked from most listeners to least.
- We used the *spotifyr* package from CRAN to access the **Spotify API**.
  - We took the names of top 100 artists obtained from the monthly listeners' data from the ChartMasters website and then using the functions available in the *spotifyr* package, we scraped all of their image urls, similar artists, detailed info about their top tracks and info about the audio features of their top tracks.
  - For a part of the app we also directly used the API accessing utility of the *spotifyr* package to compare audio features of the top tracks of two different artists at once.
- We obtained the dataset of over 600k+ tracks available on Spotify from Kaggle as a *.csv* file
  - We cleaned all columns appropriately
  - We removed the first 11 rows of the data as it had data missing for a few years between 1921 to 1946.
  - We grouped the rows by year and took the mean of all audio features for the songs released in that year to obtain year-wise data for music features.

## Data components

- **Danceability:** Reflects how suitable a track is for dancing based on tempo, rhythm stability, and beat strength.
- **Valence:** Represents the emotional tone, with higher values feeling more positive or happy.
- **Energy:** Measures intensity and activity; energetic tracks feel fast, loud, and dynamic.
- **Tempo:** The speed of the song, measured in beats per minute (BPM).
- **Instrumentalness:** Indicates the likelihood of the track being instrumental; higher values mean fewer or no vocals.
- **Acousticness:** Assesses the probability of a track being acoustic, with higher values indicating a more natural sound.
- **Speechiness:** Captures the presence of spoken words; higher values suggest more spoken content (e.g., a podcast).
- **Popularity:** Reflects how well-liked the track is, based on plays and recent listeners (0–100 scale).
- **Monthly listeners:** Number of unique listeners who listen to the artist at least once in a span of 28 days.

## Biases in the Data

- When looking at an artist's style of music based on their top tracks and the audio features of those top tracks, it may not accurately represent the kind of music that the artist regularly creates, it just shows the selected few tracks which are really popular among the listeners.
- When analyzing how audio features, duration or tempo of songs have evolved year-wise, every year does not contain the same number of songs; some years which contain very few number of songs, the mean of audio features for that year may not represent the actual music trend in that year due to very small sample size.

## Interesting Questions to ask from the data

- What are the most popular tracks that the top artists from across the world have made? What are their audio features which make them so popular among listeners?
- How do two different artists compare in terms of audio features of their top tracks? For example, for two artists of the same genre, one of them might make happier songs, or faster or more energetic songs.
- How have audio features of songs on an average evolved over various years and decades?

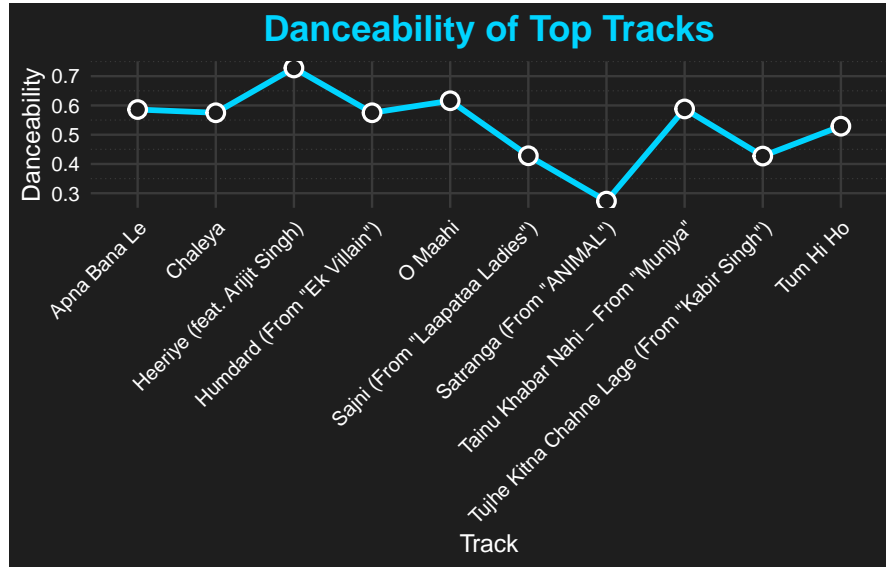
- Are songs getting shorter over time?

## Visualizations

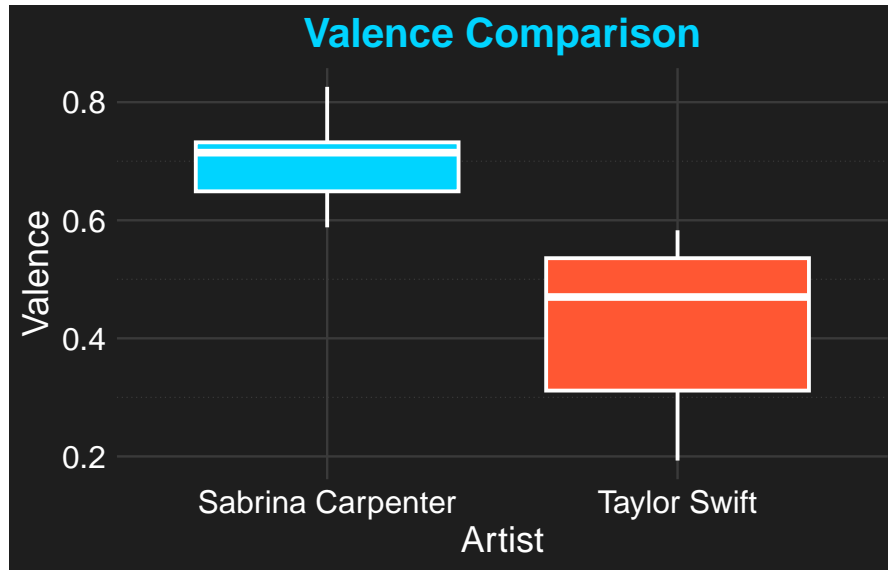
We can see the most popular tracks of an artist, any collaborating artists who worked on that track, the duration of the track, release dates of the tracks, and *popularity*, which is a spotify metric on a 0-100 scale.

Track	Artists	Duration	Release Date	Popularity
Sajni (From "Laapataa Ladies")	Ram Sampath, Arijit Singh, Prashant Pandey	2:50	2024-02-12	79
Tujhe Kitna Chahne Lage (From "Kabir Singh")	Arijit Singh, Mithoon	4:44	2019-05-31	77
Apna Bana Le	Sachin-Jigar, Arijit Singh, Amitabh Bhattacharya	4:21	2022-12-06	73
Satrange (From "ANIMAL")	Arijit Singh, Shreyas Puranik, Siddharth - Garima	4:31	2023-10-27	77
Chaleya	Anirudh Ravichander, Arijit Singh, Shilpa Rao, Kumaar	3:20	2023-09-05	72
Heeriye (feat. Arijit Singh)	Jasleen Royal, Arijit Singh, Dulquer Salmaan	3:14	2023-07-25	76
Tum Hi Ho	Mithoon, Arijit Singh	4:21	2013-04-06	75
O Maahi	Pritam, Arijit Singh, Irshad Kamil	3:53	2023-12-20	70
Tainu Khabar Nahi - From "Munjya"	Sachin-Jigar, Arijit Singh	3:08	2024-05-31	74
Humdard (From "Ek Villain")	Arijit Singh	4:20	2020-05-19	72

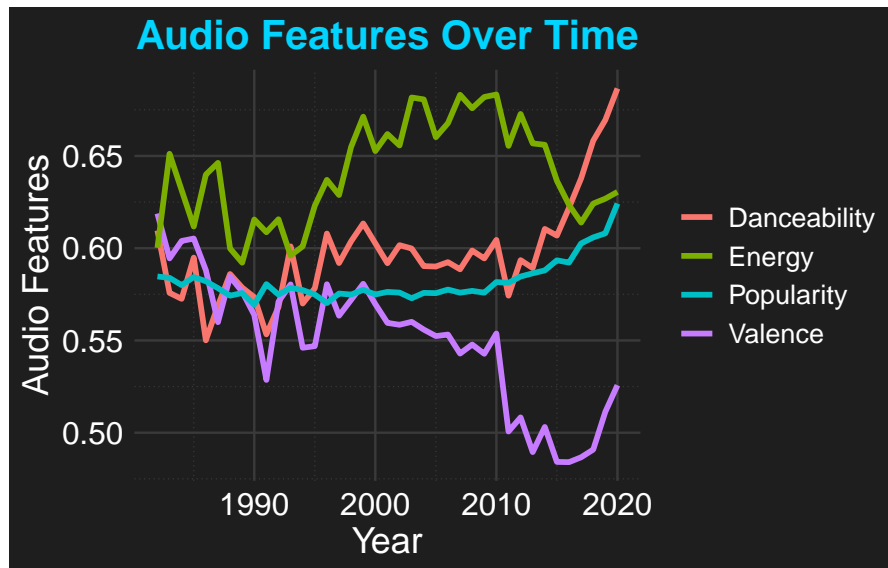
We can also get to know about what the audio features for each of these tracks. Here, we can see that *Heeriye* and *O Maahi* are more suitable for dancing than tracks like *Satrange*. We can observe that Arijit Singh makes diverse songs in terms of danceability.



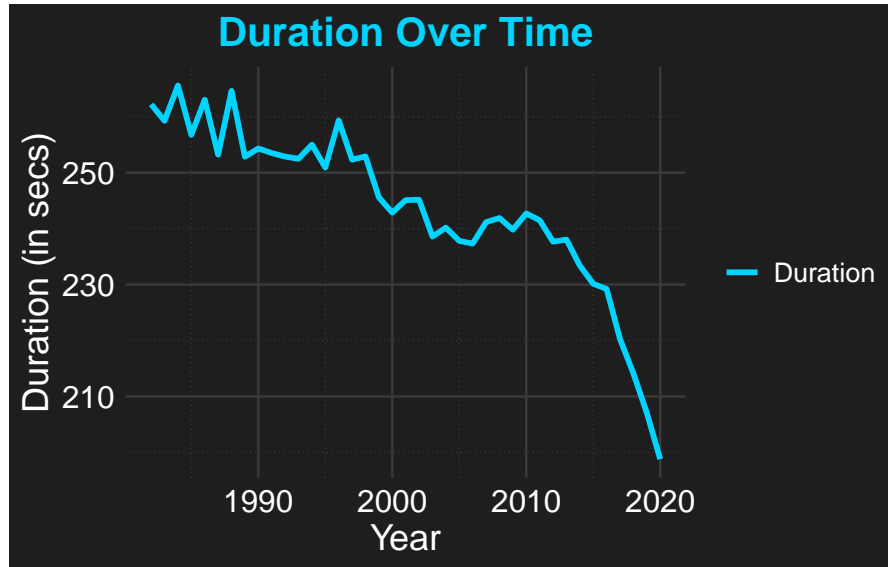
We can compare different artists on the basis of audio features of their top 10 tracks. Here on seeing this visualization we can observe that Sabrina Carpenter makes more cheerful tracks (with higher valence) than Taylor Swift.



We can see how audio features have evolved over the years and various decades. We observe that *valence* has decreased over the years, which means songs have got a sadder tone as compared to the '80s, but we can also see that *danceability* has slightly gone up since the '90s.



Now we can answer the question that we raised earlier, i.e., are songs getting shorter? We observe that indeed, there is a sharp decline in the duration of songs as we move from the '80s to the late 2010s. We suspect that this might be due to people's reduced attention spans with the advent of platforms like *Instagram Reels* or *YouTube Shorts*.



## Final Thoughts

This analysis of Spotify offers valuable insights about the kind of music people prefer from their favourite artists and how have music tastes of the common public changed over the years.

- Most popular tracks of top artists share some common features like high danceability, upbeat tempos etc. which might be the driving factor of the artist's success.
- By comparing artists we could see how various artists of the same genres can be so different in terms of the audio features of their top hits. Two artists of the same genres can vary in valence, energy, tempo, etc. which tells us about their unique sound.
- Over the years, we can observe how danceability and energy of songs have seen an increase but features like valence have seen a decline, from which we can understand the kind of songs the new generation likes in comparison to the previous generations.
- The decreasing trend in duration of songs suggests higher preference for shorter, quick-to-end songs among the current fans, possibly due to the rise of social media platforms which focus more on quick consumption content.

## References

- <https://chartmasters.org/most-monthly-listeners-on-spotify/> : Table for artists with most monthly listeners.
- <https://github.com/charlie86/spotifyr> : Link to the *spotifyr* package, credits to *charlie86* on GitHub.
- <https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks> : Link to the Kaggle dataset.