

Milestone 2 - Hit Artist Analyzer

A very broad mockup is given in annex. There can be found multiple a few plots and the overall style we wish to reach for the final website.

The goal of our Artist Analyzer is to try to visualize an artist's musical style, and compare multiple artists together visually: Frank Sinatra, The Beatles, Michael Jackson, Eminem, and Rihanna. We decided that these artists represent different musical styles during different eras of music.

For this project, we use musical data from the Spotify API as well as the Genius API. The Spotify API contains audio features, while the Genius API contains the lyrics of songs. This is how we intend to use the data to reach our goal:

1. To understand an artist musical style, we decided to start working with the artist's most representative tracks (top 5 to top 10 of their charts given their billboard score). With Spotify, we have multiple audio features describing each artist's songs. The selected ones are: **time signature** (beats per bar), **acousticness**, **danceability** (how fast, stable and regular the beat is), **energy** (how fast loud and noisy a track is), **valence** (how positive a track is). Due to the the number of features, we decided that a **radar chart** for each artist, representing the value of the features for their top 5 song, as well as the mean for their entire career to see how representative the hits are.
2. To compare artist's audio feature, a larger radio chart will show the mean (or the median) of each feature for each artists. Buttons listing the name of the artist's will be clickable to show or not a specific artist, allowing a user to compare many combinations or artists. This plot will be generated using D3. Radio Charts can be fairly simple, but this smoothed out design is what we desire to reach, and D3 seems to be the most appropriate tool.



3. After visualizing and comparing audio features, artists will be described by their lyrical features. For each artists, themes and general mood of their songs will be shown: **word clouds** seem to be appropriate here, with positive words highlighted in green, negative in red. Statistics will be visualized with simple bar plots, highlighting the size of the lyrics as well as the vocabulary of each artists. The **word clouds** will be generated using **d3-cloud**. The most used words will be used to compare the artists' top songs, and see how representative their hits are to their general lyrics.
4. Artists will be lyrically compared by scattering their most representative words against each other. A simple manner would be to select top words of each artists, and top words that both artists say, against each other, 2 by 2. This plot will be generated using **D3.js**.



5. **(TO GO FURTHER)** Once these two illustrations will be done, all tracks will be clustered given the audio features mentioned in point 1. Each cluster will be labelled given its representative audio features (example: high danceability, low acoustics). The features of the labels will be plotted using a **bar plot** showing the median of the value of each feature for each cluster. This **bar plot** will be implemented with **Chart.js**. A similar **bar plot** will be printed to show the number of songs of each artist's in each cluster.
6. A 2D representation of the songs latent space. The idea is to plot and observe how similar songs are related in the vector space. We will use a scatter plot where each point will be a song (lyrics) and we will display the change in sentiment between the songs.
7. **(TO GO FURTHER)** The final visualization of the website will come from visualizing links between artists. Here are the links to highlight:
 - * Sinatra has multiple tracks featuring **Quincy Jones**. Quincy Jones produced most of Michael Jackson's hit. How correlated are the lyrical and audio features of Sinatra's songs featuring Quincy Jones, and MJ songs produced by Quincy Jones. This will be highlighted plotting a **heat map** highlighting the correlation (or not) between all those features. This heat map will be implemented using....
 - * Sinatra sings 2 Beatles songs: Yesterday and Something. How do these songs compare musically? Re-using the radio chart from the first points seems to be appropriate here.
 - * Eminem mocks MJ in his song *Just Lose It*. How does this song (or the album containing this song) compare musically and lyrically to MJ's discography? Another **heatmap** can highlight the musical and lyrical negative correlation between the two songs.
 - * Eminem and Rihanna have multiple songs featuring each other: *Love The Way You Lie* (Part 1 and 2), *The Monster* and *Numb*. Musically, who's influenced is synced more (radio chart and scatter plot of the words contained in the songs).
 - * Rihanna samples MJ's song *Don't Stop 'Til You Get Enough* in *Please Don't Stop The Music*. How does the influence of the first song is seen in the second one (radio chart).

To implement our project, all lectures about javascript development will be needed (javascript, D3.js, SVG, Web Development...). Moreover, lectures on textviz and graph will be useful for the lyrics and the audio features respectively.