

Alexa Batino
Omar Abdul-Rahim
Sheetal Athrey

Project 1: Spotify Data Comparison

A musical sentiment comparison between each of the three group members and the US Top 100

Data Retrieval

Using the Spotify Web API and spotipy Python library, each group member used a Python notebook to extract the audio features of each of their 2018 Top 100 tracks and compile the data into a JSON file. We found the existing data for the USA Top 100 Songs of 2018 online and uploaded this dataset as a JSON. All of the datasets were rather clean, meaning each track had each field completed. Thus, they did not need to be filtered.

The US Top 100 included more audio features than the three displayed in our visualization, i.e. acousticness and tempo. Our group, however, decided to focus on three audio features of each track, specifically valence, energy, and danceability. We evaluated these audio features of our most played songs to compare them to the most popular songs in the country. According to Spotify, valence is “a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.” A higher valence value means the track is more positive, while a lower value means the track conveys a more negative feeling. Energy is “a measure from 0.0 to 1.0 that represents a perceptual measure of intensity and activity”, thus higher energy values correspond to fast and loud tracks. Danceability “describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.” If a song has a danceability value of 1.0, it is the most danceable.

Additionally, we also compared the mean values for each data group while comparing each group member’s different features to the US Top 100.

Design Rationale

We decided to display our data for each audio feature in a scatterplot of the data for the US Top 100 songs from 2018 as well as our own data of our top 100 songs of 2018. The x-axis in every graph is the list rank for every song on the list, and the y-axis refers to the values for the different audio features: Valence, Energy and Danceability ordered from 0.0 to 1.0, where 0.0 is the least measure of the attribute and 1.0 is the most. The x-axis is ordered from the least popular of the 100 songs to the most popular.

The reason we used the scatter plot is to show how diverse the top 100 songs for both the group members as well as that for US would be with respect to every attribute. We used one consistent color to show the attributes for US. We included a legend at the top of the visualization which shows what the different colors refer to. At the top of the visualization, we have included the description about our visualization. The picture on the visualization is an illustration created by Humaans of the three of us listening to music signifying Spotify.

Our project consists of the plots for the three elements Valence, Energy, and Danceability, with each feature contained within a box, so that we could have the data appear to pop out. Within each box, we described what each feature was essentially describing, followed by the three scatter plots for each member. Finally, these plots were followed by an analysis of what the data means.

The plot is further enhanced by including a dashed line which conveys the mean feature value for each of the group members as well as the one for the US top 2018. The reason we used a dashed line is because a solid line would obscure some of the data points. Under each plot, we also included in text the mean value for the features for each member as well as that for the US top 100, because this would further explain the difference between the values.

We also attempted to make the webpage responsive to change in window size which was successful for the most part, except for really small resized windows. We also positioned the svgs and the elements of the project in the center of the screen so that it would be easier for the user to view the visualization better. We used a linear scale for the different plots because that best represented the kind of data which worked with a top 100 list, with ranks.

Analysis

The class analysis for the Valence measure tells us that for all the three group members we see that their means lay lower than that of the US Top 2018 mean. Omar's mean lays a lot lower, with a greater difference between his mean and that of the US Top 2018, whereas Alexa and Sheetal are super close to the mean of the US Top 2018.

The class analysis for the Energy measure contains an interesting behavior because Omar and Alexa both have their mean energy values lying a lot lower than that of the US Top 2018, but Sheetal's mean energy value is slightly above that of the US Top 2018.

The class analysis for the Danceability measure tells us that for all the three group members we see that their means lay lower than that of the US Top 2018 mean. All there group members have nearly the same difference between their Mean Danceability Values and that of the US Top 2018.

Responsibilities:

- Alexa:
 - Found initial 2018 US Top 100 dataset on Kaggle and picked out the three audio features to focus on
 - Looked into analysis features, i.e. mean, median, range, to display on viz
 - Calculated average value for each dataset + variable and added the average value line to each visualization after plotting my Top 100 data
- Omar:
 - Writing script to pull individual Spotify data
 - Building the first version of the visualization
 - Styles and illustration
- Sheetal:
 - Helped with first version to pinpoint key analysis features of the visualization
 - Included System description and Analyzed the different features displayed on viz
 - Styles and illustration of different elements of the visualization