Data Science

Project Goals

Data: Data retrievable from Spotify API. This includes things like popular artists and playlists, as well as information on an individual like the artists they follow and tracks in their playlist.

Data: One part of the project will utilize a csv containing all the artists and their popularity score

Methodology: Our data is retrievable in json form, which can be processed and put into a csv. As of now, we don’t have a single large dataset.

Popularity based on artist

Currently, with a user input, we gather the artists they listen to. Each artist has a ‘popularity score’ based on how popular they are on Spotify. Each artist also has related artists. What we have so far, is taking an artist and calculating their aggregate popularity by averaging their popularity with the popularity of related artists. This helps reduce artist outliers. The user’s popularity is then calculated as an average of the artists they listen to. We can expand on this in a number of ways. First, we can look at the variance between similar artists to see if it’s a good tool to balance each artist. We can catalogue artists based on popularity and suggest new ones to a user with similar popularity. Finally, with multiple people, we can do machine learning to see if we can guess if two people are friends based on popularity. We would use test cases where we know two people’s popularity and if they are friends or not.

Featured playlists indicating music preferences over time or location

Want to look at featured playlists on Spotify. A request returns popular playlists, including a link to included tracks. We can categorize the tracks and assign a genre or some grouping classification for these popular playlists. We can then run this multiple times based on location or time to see differences. Additionally, we can look at say US from when Spotify started in 2008.

Popularity based on artist

* Catalogue artists based on popularity and suggest new artists to a user with a similar popularity score. 75% goal
  + Data: csv containing all the artists on Spotify and their popularity score
  + Need to gather a list of artists from Spotify, collecting their popularity score, then writing it to a file. After that will need to accept a user’s input and output suggested artists within some range.
* Implement a machine learning program on the closeness of two people’s calculated popularity and whether they’re friends or not. Then we can use this to predict if someone is friends with someone else based on the closeness of their popularity. We can visualize this as circles where a user can drag their circle onto a person and the program will tell the user if they are friends or could be friends based on popularity score. 100% goal
  + Data:
* Want to look at featured playlists on Spotify. A request returns popular playlists, including a link to included tracks. We can categorize the tracks and assign a genre or some grouping classification for these popular playlists. We can then run this multiple times from when Spotify started in 2008 to look at what sort of music has been popular between then and now.

A future investigation will look to catalogue artists based on popularity and suggest new artists to a user with a similar popularity score. We will first need to gather a list of all the artists on Spotify and their popularity score. After that we can offer new artists to a user. Similarly, we can take a user’s favorite artist and recommend other works by them.

Another future investigation is to look at featured playlists on Spotify. A request can return popular playlists, including a link to included tracks. We can categorize the tracks and assign a genre or some grouping classification for these popular playlists. We can then run this multiple times from when Spotify started in 2008 to look at what sort of music has been popular between then and now.