**Project 2**

**Team Members:**

* Andrew Sooklall
* Mikey Esteban
* Pratik Pathak
* Zane Christopher Lynch

**Proposal Date: 12/12/19**

**Due Date: 01/04/20**

**GitHub Repo**: https://github.com/ASooklall/project-2-spotify.git

**Proposed Project:**

* **Option 1: Custom “creative” D3.js project (nonstandard graph/chart)**
  + With multiple simpler charts for specific portions of our data.
* Using Datasets from Spotify we aim to show users popular artists for each year and trends within songs and artists.
* Upcoming artists will also be able to see trends between years and predicted trends for upcoming years to adapt their music to match the growing interests of the population.
* We plan to clean and load CSVs into databases to load within a flask app. Using D3/JS and HTML, we’ll be creating webpages to host our graphical representations of our data and explain our findings.
  + Users will be able to search through our representations and view different comparisons by interacting with the charts and graph in real time.

**Purpose:**

* Show the top songs/artists per year in a graphically pleasing manner.
* Show trends between years for duration, tempo, genre, etc. to analyze changes in popularity among different factors.
  + These trends could be used by record labels or new upcoming solo artists to show what direction their songs should head towards in 2020/2021 etc.
* Provide comparisons between artists and songs and allow people to look into similarities between artists they might not normally consider or allow users to see other similar artists they might like.

**Tasks:**

* Load CSV from Kaggle
* Webscrape wiki/spotify for artist bio
* Scrape spotify for album image for each song
* Create init.py file to load data, clean data, and push into a database
* Use Flask to load routes for HTML pages.
  + Index page, dataset page which search function?
* Flask server should include loading databases and output JSON based on translated data.
* D3/Javascript to create plots and charts being used for the project.
  + Interactive Concept Map / Network
    - Order: (Years -> Genre -> Artist -> Album -> Song)
  + Sunburst chart that shows top songs in genre per year
  + Searchable data chart
* HTML to load the Javascript into a template that the flask app will pull from and load onto routes.

**Data sources:**

* Wikipedia
  + Scraping pages for artist bio and pictures
* Spotify
  + Scraping data for album pictures and song lyrics?
* Kaggle Datasets
  + <https://www.kaggle.com/nadintamer/top-tracks-of-2017>
  + <https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018>
  + <https://www.kaggle.com/leonardopena/top50spotify2019>