Group Members: Del-Roelle Thorpe (dthorpe), Aylin Perez (aperez6), Emily Blatt (eblatt), Mindy Ng(mln1)

- Vision:
    - What is your “big idea”?
        - Can we predict happiness of a country based on the audio features of its top daily songs on Spotify?
    - What might you find at the end of your project?
        - Hopefully, by comparing happiness ratings with top songs and their song lyrics, we will be able to make a conclusion regarding if a population’s happiness level can be reflected through music
- Data:
    - What dataset do you plan on using?
        - We plan on mainly using Spotify’s Worldwide Daily Song Ranking dataset on Kaggle, and merging it with the Spotify Audio Features dataset and the World Happiness Report Dataset. The main dataset contains the daily ranking of the 200 most listened songs in 53 countries from 2017 and 2018 by Spotify users.
    - How big is it?
        - The Daily Song ranking dataset has 3.44 million rows and 7 columns so it is quite large. The Happiness dataset is much smaller and has about ~150 rows per year with about 12 columns. The Audio Features data set has abou 116k rows and 17 columns.
    - How do you plan to collect it?
        - All of the datasets are available on Kaggle, which we can collect by web scraping. We can also scrape Spotify’s API to access audio features of the tracks on Spotify.
    - How do you plan to clean it?
        - We will clean the data by finding matches between Spotify’s Worldwide Daily Song Ranking and the Audio Features Data. We will definitely need to explore this further though.
- Methodology:
    - What do you plan on doing with your data?
        - We plan on joining the data with another data set called the “World Happiness Report” which can also be found on Kaggle. After joining this data, we want to see if we can predict happiness based on the country's top spotify songs. One way to look at this data might be to find the most frequent audio attributes of the top songs in a country and see if certain attributes are more correlated with higher happiness ratings.
    - What techniques do you think you will use to analyze the data?
    - How might you visualize your results? Since we are early in the course, feel free to speak in broad terms.
        - Possibly by presenting results on a digital world map. We could color code higher correlations between positive song attributes and happiness ratings of countries.
- Getting started:

○ What do you plan to accomplish by the first TA check-in?
- By the first TA check in we will make sure all the data is clean and hopefully merged into one dataset, organized by country.

○ Likewise, what will you plan to accomplish by the midterm report?
- By the midterm report we hope to have begun analyzing the data. If we begin to see patterns, we can think of ways to visualize the data, possibly on a digital world map.

Proj requirements:
https://cs.brown.edu/courses/csci1951-a/assignments/2019_Final_Project_Requirements.pdf
Past project examples:
https://drive.google.com/drive/folders/1MLCOkMg11DZ8rx4B3ZULt9jrL72R4jt4
Data sets:
https://github.com/awesomedata/awesome-public-datasets
https://www.kaggle.com/datasets

Restaurant Health Risks

Interested? : Emily, Roelle

Potential Questions:

http://drivendata.co/case-studies/using-yelp-reviews-to-flag-restaurant-health-risks/

Tech Workspace--

Interested? : Mindy, Roelle

Potential Questions: What makes for a good work environment? What are college students looking for? How are mental health and gender or race correlated?

https://www.kaggle.com/petersunga/google-amazon-facebook-employee-reviews

https://www.kaggle.com/niyamatalmass/google-job-skills

https://www.kaggle.com/osmi/mental-health-in-tech-survey

Education:

Interested?: Mindy, Aylin

Potential Questions:

https://www.kaggle.com/passnyc/data-science-for-good

https://collegescorecard.ed.gov/data/

Interested?:

https://www.drivendata.org/competitions/46/box-plots-for-education-reboot/

Transportation:

https://github.com/BetaNYC/Bike-Share-Data-Best-Practices/wiki/Bike-Share-Data-Systems

World population data:

Interested? Mindy

Interesting aspects of the data: flight patterns, using mobile devices to produce dynamic population maps

Questions: Where are people traveling to and when? Maybe merge with data sets about violence, politics, or health of diff countries and seeing if that affects where people are traveling to and from

https://www.worldpop.org/project/list

World Happiness Report

Questions: What determines happiness? What countries are known to have the happiest people? Why?

https://www.kaggle.com/unsdsn/world-happiness

Interested? Emily

Spotify WorldWide Daily Song Ranking

Potential Question: How do the song attributes that people are listening to around the world correlate with country happiness? How do artist popularity in various countries correlate with U.S. cultural influence in those countries?

https://www.kaggle.com/edumucelli/spotifys-worldwide-daily-song-ranking

https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018

Interested? : Emily, Roelle

Potential Questions: Predicting public health risks from restaurant reviews

http://drivendata.co/case-studies/using-yelp-reviews-to-flag-restaurant-health-risks/

https://www.kaggle.com/yelp-dataset/yelp-dataset

https://www.drivendata.org/competitions/5/keeping-it-fresh-predict-restaurant-inspections/page/33/