Tala Tayebi
DSCI 510
HW3
Github Repo: https://github.com/talajune/DSCI510_FinalProject

1. What data did you collect? How many different sources did you use? How many data samples did you collect?

      I collected data from two sources, first by downloading a CSV from the "World Happiness Report" on Kaggle, and another by getting country data from the country API on API-Ninjas (https://api-ninjas.com/api/country). After merging the two datasets, some data pre-processing, and modifying the dataset to include only a subset of data from the top 100 "happiest" countries. That being said, my final dataset includes 100 rows of data and 26 columns. I chose to look closely at the top 100 "happiest" countries, and look at different attributes to see how it might influence the "happiness" of a country. I collected data from the API and Kaggle dataset and then merging them on the country name. I decided to drop some columns / attributes that I won't be looking at in my project, which you can see in the code.

For the Country API, I collected keys "name" (country name), and "limit" which I set to 1. Since my API collection happened in a for loop, I iterated over the loop using a list of the 100 different names of the top 100 happiest countries. I saved my final data collection file as a CSV.
For the CSV dataset from Kaggle of the "World Happiness Report", I read it in using Pandas and merged it with the API Dataset using the country name as the key.