Project 1/31/2022

**Statement of Work**

* Create a reusable Jupyter notebook that’s purpose is to clean datasets and perform some initial exploratory data analysis (EDA)
* The notebook you create should have the flexibility to be applied to different datasets (don’t build it just to fit one dataset)
* Apply the notebook to a dataset of your choosing as an example of your process at work

**Project Goal**

* To create a useful, repeatable process/script that will help clean datasets
* Create a product that you can have for personal and professional use in the future
* Create a product that you can include in your personal Github repository to show to potential employers and assist in the interview process

**Project Requirements**

* Research and find a dataset that needs cleaning processes applied to it
* At the end of your process, summarize the difference between original and clean datasets and how you improved the dataset
* Your process needs to implement the pandas library, and the result of process should be a pandas dataframe
* All sub-processes created to clean data should be implemented via functions (one sub-process per function)
* Visualize your dataset in at least 2 different ways
* Prove that your dataset is either normally distributed, right-skewed, or left-skewed
* Calculate the 3 different measures of central tendency we discussed in class (mean, median, mode)