**Assignment 5**

Weeks 8 & 9 – Pandas

In this homework assignment, you will explore and analyze a public dataset of your choosing. Since this assignment is “open-ended” in nature, you are free to expand upon the **requirements** below. The preferred method for this analysis is in a .ipynb file.

Please use comments or heading to indicate what you’re coding!

Some data examples:

* <https://www.data.gov/>
* <https://opendata.cityofnewyork.us/>
* <https://datasetsearch.research.google.com/>
* <https://archive.ics.uci.edu/ml/index.php>

Resources:

* <https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html>
* <https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Data Exploration**

Import your dataset into your environment and explore your data. You should include any summary statistics, means, medians, quartiles, or any other relevant information about the dataset.

**Data Wrangling**

Create a subset of your original data and perform the following.

* **Use .rename() on multiple columns.** Take note of inplace=True.
* Look at the structure of your data – are any variables improperly coded? Such as strings or characters? **Convert to correct structure** if needed.
* **Create a new column based on existing columns or calculations.** For example, you have two columns: city and state. You create a new column called location that consists of city and state.
* **Drop column(s) from your dataset.**
* **Sort your data based on multiple variables.**
* **Filter your data based on some condition.**
* **Apply multiple aggregation functions to your data (count, mean, min, max).** Use the .groupby() and .agg() or .apply() functions if needed.

**Data Visualization**

Use pandas to create a data visualization of your choosing on your dataset.

**Conclusions**

Provide some conclusions or insight gained from looking at your data.