OMIS 30 - Fall 2020 - Project 4 ASYNC

Logistics:

Assigned: Thursday, November 19, 2020
Due: Thursday, December 12, 2020

Objective:

Perform a data analysis on a data.gov dataset using Jupyter Notebooks.

The requirements for this project are:

- Select two or more related datasets from data.gov (https://catalog.data.gov/dataset)
- Choose a datasets which no one else has (including the other section) and record it on Google Groups
- Use Jupyter Notebooks and submit the .ipynb file
- Combine 2+ datasets into a single dataframe using a merge, join, or concat function correctly. (You may use the dataframe method or the pandas function.)
- Find 4 interesting facts/patterns about the new dataset and present your findings with the use of graphics - at least one table and at least three charts/graphs. (You may use matplotlib or seaborn.)
- Tell a story with the data

Guidelines for judging 'interesting':

There are multiple ways things get to be 'interesting'. Here's two of the best heuristics we know:

- This fact/pattern is so interesting you would go to a party and say: "Guess what I found out about xyz!"
- This fact/pattern is crucial to understanding the topic: e.g. For gerrymandering, that would be something like 'There are x many districts, that are up for debate every y years, and z are the decision-makers. If *a* many districts shift to red/blue, then the odds of the election swaying one way is *b%* higher.' By the way, this would be one of the three sections - not all three in one.

Storytelling & visualizations:

- Present the data in a manner which draws people in and keeps them engaged
- Be concise, clear, concrete, correct, coherent, complete, and courteous (7 C's of communication)
- Use comments for code, and Jupyter elements for storytelling
- Pictures are worth a thousand words. Use them to distill complicated data into an easily graspable chart or table.

Resources:

- https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html
- https://catalog.data.gov/dataset
- http://jupyter.org/
- https://matplotlib.org/
- https://seaborn.pydata.org/
- https://pandas.pydata.org/
- https://www.mindtools.com/pages/article/newCS_85.htm
- https://datavizblog.com/2013/05/26/dataviz-history-charles-minards-flow-map-of-napoleo

  ns-russian-campaign-of-1812-part-5/

Collaboration:

You will work in PAIRS on the assignment.

Submission:

- Name your final file <your_username>_project4_fall2020_ASYNC.ipynb (mine would look like dvrdoljak_project4_fall2020_ASYNC.ipynb).
- Make sure it runs completely and correctly on your computer
- Submit it via Camino
- (We will run your program on our computer to test your answers)

Grading Rubric:

Section	**Grade**	**Criteria**
Interesting Fact 1	20%	Interestingness, factfullness, analysis, presentation
Interesting Fact 2	20%	Interestingness, factfullness, analysis, presentation
Interesting Fact 3	20%	Interestingness, factfullness, analysis, presentation
Merge/Join/Concat	20%	Correct use of merge/join/concat between 2 dataframes
Use of comments & Readability	15%	Documentation of author & dates; Explanation of steps Use of whitespace; Use of new lines; Naming convention of variables; Sequencing of code and outputs
General & Submission	5%	Directions followed correctly