BAIS:3020 Computational Thinking
Project

Instructions:

The project requires you to **define your own problem, based on data you identify for analysis, and the solution(s)** that you solve by using the programming and concepts learned in class. You can consider things from class for inspiration; however, you should be creative and pick a problem that is different and of interest to your group. Problems can be from a multitude of domains, including, yet not limited to,

- data sources, like https://www.searchenginejournal.com/free-data-sources/302601/
- business analytics
- personal or small organization
- text or document analysis
- social network analysis
- etc.

Implementation:

You are required to implement at least **four functions** in a module (file) called ***utilities.py***. The functions may provide simulation, plotting, text analytics, processing, or any other useful code. Remember to provide clear comments (docstrings) for your functions. The functions are to be used in a separate python file called ***main.py*** to create a larger application complete solution.

Make sure your functions and code use appropriate error handling and are appropriately commented and organized with good use of whitespace. You should separate code into separate modules and functions as needed for good code organization. You are not limited to the two modules designated herein. Add modules and functions as needed.

You must use at least two or more of the following Python packages that are available as part of Anaconda installation:

1. pandas (https://pandas.pydata.org/) – Data Analysis
2. matplotlib (https://matplotlib.org/) – Plotting

3. numpy (https://numpy.org/devdocs/user/quickstart.html) – Numeric Computing
4. scipy (https://docs.scipy.org/doc/scipy/reference/) – Scientific Computing

Note: Be sure that you do the processing of your data in Python code and not another application like Microsoft Excel. If all you do is read in data to display that has been pre-processed, you will not receive a good score. It is okay to do some preparation, if you have a large data file and part of it is not relevant. If you are unsure, please check with me.

Presentation:

Each group will record a 5-minute presentation and submit together with other deliverables by May 3rd. No late submissions will be accepted. The recording should be submitted in MP4 or other widely acceptable video format as a link. No other deliverable is permitted on this project. The presentation should have a target length of 5 minutes and not exceed 7 minutes in length. It must be easy to comprehend, and free of acronyms and jargon. All group members must take part in the presentation itself.

You should create a Microsoft PowerPoint for your presentation. First, introduce your group, then clearly state the problem, raw data sources, and a detailed plan of your solution.
Your presentation should include:

- Brief introduction of the group members. All group members must be present, be introduced at the beginning of the presentation, and have their cameras on during the presentation.
- A clear statement of the problem.
- Overview of data source(s), a brief summary of the data and how you processed them.
- Detailed plan of your solution and how you determined your solution approach.
- Overview of solution modules and functions, and Python libraries employed in your code solution.
- Output, i.e., the results and how you present them.
- Takeaways and conclusion

A good presentation will meet the 4-to-6-minute timing target, provide clear opening that introduces the subject and presenters, thorough development of topic and summary conclusion, and makes good use of graphics. You do not need to go over your code, but a demonstration is fine. In either case, do include output displays in your PowerPoint. In addition, an excellent video is very close to 5 minutes in length, is very specific about a practical problem, and is delivered in a stimulating and interesting manner.

Deliverables (submit as a single zip file):

You should submit the following deliverables (read carefully):

1. **Short report:** A Microsoft Word document that briefly explains your problem and data, then describes the functions that were implemented (input, outputs, and what they do – not how they do it).
2. **Code:** *utilities.py* module with the 4 functions, the *main.py* file, and any other modules used in your solution.
3. **Data files:** Include all data files used in your analysis. We should be able to reproduce your results based on these data files by running your code.
4. **Presentation:** Microsoft PowerPoint presentation (*group_<number>.pptx*) with screenshots of code output.
5. **Recording:** A video recording of the presentation in MP4 (*group_<number>.mp4*) or other widely acceptable video format as a link (GDrive, YouTube, ⋯). If you choose to share a link:
    a. Be sure that the TA and I have proper access to view your recording, and
    b. include the link in a text file (*group_<number>.txt*) and upload it together with other deliverables.

You are to work on the project only with your group and submit your zip file on ICON Assignment dropbox only one solution per group. Be sure to zip all deliverables. Do not discuss the project with any other group. Late submissions will not be accepted.