Final Project
DATA 602

The purpose of the data project is for you to conduct an analysis with a dataset of your choosing. You will use skills learned from the course to assist in your analysis.

There are three components to the project, the proposal, which will be graded on a pass/fail basis, the final project (your analysis) and a video recorded presentation.

You may choose to work in a group of 2-3 classmates or individually.

Proposal Information:

In the appropriate BlackBoard assignment post, please submit the below information regarding your project. Your proposal can be an informal document using bullet points and you may include any python code or output. At the minimum, you must include:

1. Research Question
2. Justification - why is this relevant to you or industry?
3. Data Sources - did you find this data online or collect yourself? Provide links.
4. Libraries potentially being used.
5. EDA and summary statistics.

You may upload this information as a .pdf or .doc file.

Final Project Information

You should include the following sections in your project:

1. Abstract of no more than 300 words.
2. Introduction - research question clearly stated, can be answered by the data
3. EDA - well-labeled graphs/plots that are appropriate to your analysis. Each visualization includes a description of what is being shown.
4. Data Wrangling - tidying of your data, manipulating missing values, structuring of your data, etc.
5. Data Analysis.
6. Conclusions

You may use any freely available data on the internet, your own personal data, or data from your company with their permission. Websites with freely available data are listed below.

Data Resources:

- https://www.data.gov/
- https://opendata.cityofnewyork.us/
- https://datasetsearch.research.google.com/
- https://archive.ics.uci.edu/ml/index.php
- https://www.kaggle.com/datasets

The format for submission is flexible. For instance, you could complete one of the following:

1. Submit a .pdf with your responses, visualizations, analysis and conclusion, along with your python code in an appendix.
2. Use Google Colab to create a document showcasing the information and python code using the various headings and code blocks.
3. Create a PowerPoint presentation of slides with required information and python code.

Database Connection (optional)

You may also store your data in a local SQL database. You can then access your data by importing directly from your database into your python environment.

Resources for database connection:

- https://dev.mysql.com/doc/workbench/en/wb-admin-export-import-table.html
- https://medium.com/@tattwei46/how-to-use-python-with-mysql-79304bee8753
- https://downloads.mysql.com/docs/connector-python-en.pdf
- https://datatofish.com/how-to-connect-python-to-sql-server-using-pyodbc/

Presentation

The final portion of the project is to submit a recorded video of your or if working in a group, your group's presentation. A slidedeck is recommended, however, you may present in the format of your choice. You will submit your presentation to a BlackBoard assignment post, as well as post your presentation in a discussion form (to be released)

so classmates can view your work. There is no set time requirement for the presentation - you may take as much time as you need.