Data Science Group Project

The project represents an opportunity to examine a topic of interest with your new found tools. Work together to develop a plan to study the data in a systematic manner as outlined below. Be ready to enthusiastically present your completed Jupyter notebook with the class on the last class meeting day.
Example and suggested datasets can be found at: https://temple.2i2c.cloud/hub/user-redirect/lab/tree/datascience/Fall%202023/Group-project/dataset.ipynb
For this project you will work in a group of 3-5 which will be assigned at random. Discuss a common topic and design a plan to work collaboratively. The methodologies we will focus should be drawn from module 3 and module 2.

Project goals and milestones

1. Develop a project title, a short description of the data source and motivation for studying dataset. Submit this title, description, names of group members, the name/url of the data set(s) you will use, a brief data science plan, and how work will be distributed across members to Canvas.
2. Carry out exploratory data analysis. Create initial data tables and plots to explore the nature of the dataset.
3. Formulate a research question(s). Create a detailed markdown cell to detail this research question and related hypotheses. Support with research papers or links where possible.
4. Formulate a data science plan to study the research questions which may include testing hypotheses.
5. Implement your methodology in necessary and annotated code including functions.
6. Use appropriate statistical tests to indicate the level of uncertainty or accuracy (in the case of training and test data). Create markdown cells to annotate your work.
7. Organize your sheet and visualizations to maximize interest and impact for submission to Canvas, posting for class, and final presentation.
8. Develop an abstract with less than 200 words which gives an overview of the project and highlights the findings. This should appear in a markdown cell under the title and a list of names of group members at the top of the Jupyter notebook.
9. Carefully organize a presentation which provides 1) background and overview of selected data 2) overview of data in tables and through visualization 3) Data science plan 4) code discussion explaining strategy and algorithms 5) Conclusions and their statistical significance.

Deliverables

1. Title and description of project submitted to Canvas (See 1 above)
2. In class presentation of project by complete group
3. Project notebook and slide submission submitted to Canvas
4. Statement of personal contribution to group project submitted to Canvas