Data Science Group Project

The project represents an opportunity to carry out a detailed study of a topic and demonstrate your prowess in using the statistical and coding tools you have learned during this course. 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 and/or slides with the class on the last class meeting day.

- Datasets and suggested data cleaning and questions to test can found at:
  https://temple.2i2c.cloud/hub/user-redirect/lab/tree/DataSTEM/STEM%20UP/Group-project/Group_Project_Description.pdf

- Example of a finished final project:
  https://temple.2i2c.cloud/hub/user-redirect/lab/tree/DataSTEM/STEM%20UP/Group-project/example_group_project-indego_bikes/Indego.ipynb

- Dr. Jonathan Smith's Data Science Primer:
  https://laserchemist.github.io/dprimer/
  Particularly the following section when data wrangling is needed:
  Data wrangling: strings to numbers — Data Science Primer

For this project you will work in a group. Devise a plan to work collaboratively. Your group will apply statistical methodologies from the course to test a supported hypothesis. Careful research into the dataset topical area should guide and support hypothesis development. This background research should be evident in the presentation and final project submission.

Project goals and milestones

1. Provide a project title, a short description of the data source and motivation for studying dataset. Include research questions, statistical tests planned, and the roles for each team member.
2. Research the topic of study carefully producing 1-2 paragraphs of text supported be references to research papers and authoritative websites. This research should inform hypothesis development. Images can also enhance the understanding of the topic.
3. Formulate a research question(s). Create a detailed markdown cell to detail this research question and related hypotheses where appropriate. This should build from support presented in (2.) above.
4. Carry out exploratory data analysis. Create initial data tables and plots to explore the nature of the dataset.
5. Formulate a data science plan to study the research questions which may include testing hypotheses if appropriate.
6. Implement your methodology in necessary and annotated code including functions.
7. Required: Use appropriate statistical tests to test your hypothesis and indicate the level

of uncertainty or accuracy (in the case of training and test data). Create markdown cells to annotate your work.

8. Organize your sheet and visualizations to maximize interest and impact for submission to Canvas, posting for class, and final presentation.

9. 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. Annotate this abstract with a single figure which best illustrates your findings from the statistical tests.

10. 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.