

Final Project Requirements

The following guidelines should be followed and will be used to grade your project:

- The code for each part should be implemented using Jupyter Notebook/Google Colab. All
- the code should be included in one single notebook (ipynb file). Submit 1 zipped folder
- that contains: the pickle (.dat) file used, 1 ipynb file containing your python code & documentation, and powerpoint presentation slides.
- Only 1 zipped file per team to be uploaded to the Final Project folder on eLearning. You must name your zipped file using 'Group[No].Section[No].FP' notation. You will
- get zero points if your code has syntax errors or is presented in other analytical tools besides Python.

Project Rubric

In this project, you will use a provided dataset to tell a story about your data, and present visualizations to the audience explaining data trends.

Use the techniques introduced in class to read your data, and show the following types of visualizations in your report/presentations:

1. Bar Charts
2. Boxplots
3. Line Plots
4. Scatter Plots
5. Histograms
6. Pie Charts
7. Heat Maps (not covered in class)

Your Deliverables:

1. **Dataset:** You are provided with the project instruction file on eLearning in a folder titled 'Project Assets - Group X', where X is your project group number (under Final Project folder). This file contains:
 - a. An OOP problem, and its corresponding instructions.
 - b. A link to a pickle file with a hypothetical dataset that aligns with the problem.
2. **Python Code (14.5 pts):**
 - a. OOP code (6 pts):
 - i. Build and test OOP classes given in the instruction set.
 - ii. Load the given pickle file to retrieve the objects.
 - iii. Generate a csv file from the loaded objects.
 - b. **Visualization code (8.5 pts):**
 - i. Shows how the data is read, parsed, cleaned, and prepared.
 - ii. Shows the code for every visualization.
 - iii. The code must be included in only 1 ipynb file (Jupyter Notebook/Google Colab) and presented along with the pickle file.

3. **Documentation (5.5 pts):** Along with the code, add documentation and data trends to your ipynb file (Jupyter Notebook/Google Colab) containing the following:
 - a. Introductory section on the data set (types of data)
 - b. How you dealt with missing values in the data. Use the techniques covered in class (mode for categorical variables, median or mean for numerical variables).
 - c. Description of trends depicted by all 7 visualization types mentioned above (make sure every graph has a paragraph explaining the data trend)
 - d. Conclusion section (summarize the story that your data set is telling you overall)
 - e. Recommendation section (suggest ways to improve your data set to draw better insights in the future)
4. **PowerPoint Presentation (5 pts):** shown in class to the audience where you:
 - a. Introduce the data (1 slide)
 - b. Dive into visualization / trend discussion (Maximum 7 visualizations, 1 slide/vis)
 - c. Conclude (1 slide)
 - d. Propose a recommendation (1 slide)
 - e. Total presentation time: 12 minutes/team.