Mini - Project (Lab 5 – 6)

Due: submission - February 18 (Needs to be uploaded to Canvas and Gradescope),
demo - February 21.
Late submission - February 25, demo - February 28 (20% off), no late passes!

Your assignment is to design and conduct a survey, and then analyze and present the data collected from the students of CS105 and CS111 classes. The dataset will be available in Google Drive.

Decide on a topic of your study and create a survey that will allow you to collect the appropriate data (as discussed in class). (If you don't submit your questions by the deadline, you can use the available dataset, but you will lose 20% of your score.)

When the master dataset becomes available, download it and create your own dataset. (Remove all columns that you won't need, clean the data that you plan to analyze.) Now perform exploratory data analysis (EDA).

Questions to answer and steps to go through:
1. What data do you have? (Your answer)
2. What would you like to know? (Your answer)
3. Explore and present the data. (Generate statistics, perform visualizations)
4. Can you state any hypotheses or make predictions? Which tests can you apply to verify your hypothesis? (Your answer).
5. Test your hypotheses.
6. Draw the overall conclusion.

Use Jupyter notebooks.

What to submit:
- Present your topic.
- Answer Questions 1, 2.
- Question 3: Explain what you are computing, and then compute using Python. Create some visualizations (at least 8, and at least 6 must be of different types); use Python. **Analyze.**
- Question 4. State clearly each of your hypotheses (at least 3).
- Question 5: Test your hypotheses and predictions (use at least 2 different tests). For each: i) describe the test you are using; ii) perform it; iii) analyze the results and draw the conclusion.
- Draw the overall conclusion.

Note. You must perform correlation analysis and chi-squared test.

Submission instructions:

1. The first submission for your project topic and questions (in Google drive) is **due Wednesday, January 31, 11am.** If the topic or the questions are not approved, you can edit and resubmit. The final version must be approved by **Friday, February2, 11am.**

2. Once you completed the project, you need to submit it to canvas (ipnb) and Gradescope (pdf) by the deadline (**February 18**), one per group.

To submit to Gradescope:
Go to File > Export Notebook As >PDF.
Double check that the entire notebook is in this PDF file. If not, try first exporting the notebook to HTML, and then printing to PDF.
Upload the PDF to Gradescope.

3. Demo: You will need to demo your project in class or during the available demo time. All group members must be present during the demo.