Lab 5 – 6

Due Tuesday February 21, (Late deadline: February 28).

Your assignment is to design and conduct a survey, and then analyze 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 at least 20% of your score.

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

1. What data do you have? (Your answer)
2. What would you like to know? (Your answer)
3. Explore 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.

Use Jupyter notebooks.

What to submit:

- Present your topic.
- Answer Questions 1, 2.
- Question 3: Explain what you are computing (mean, SD, …), and then compute using Python. Create some visualizations (at least 8, and at least 6 must be of different types); use Python.
- 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.

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

To verify the correctness of each of your tests in Question 5, you need to create a toy dataset. Such a dataset should include two to three rows and the needed attributes of

the original dataset. Perform your test by hand (show your work), then test your code using this test. Attached all calculations.

Submission instructions:

1. The first submission for your project topic and questions (in Google drive) is **due Monday, January 30.** If the topic or the questions are not approved, you can edit and resubmit as many times as you want. The final version must be approved by **Friday, February 3, 11:59am.**

2. Once you completed the project, you need to submit it to canvas (ipnb) and Gradescope (pdf), 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.

Demo:

You will need to demo your project either in class or during available demo time by **Tuesday, February 21**. All group members must be present during the demo.