

Assignment 4: Data Analysis

During the first lecture, you all answered a simple questionnaire, resulting in a messy CSV file which includes information such as: your year at UCSD, major, age, gender, height, weight, eye color, and favorite ice cream flavor. In this assignment, you will use this dataset to work through a simple data science project - using a dataset to come to an informed conclusion about a question of interest - using several analysis techniques along the way.

Tasks / Learning Goals

This project has two main objectives:

- To work through a template of a full project: going from background work and hypotheses, exploring and checking a relevant dataset, doing hypothesis driven data analysis, exploring potential confounds and/or alternative explanations and ultimately coming to an informed conclusion regarding the research question.

Due Date

11:59 pm on Sunday, February 25th, submitted on TritonED.

Submitting Assignments

You must submit the provided Jupyter notebook file (.ipynb) to TritonED. Make sure that the file you submit has the following filename (filled in with your course ID number - first letter of your last name, followed by the last 4 numbers of your student ID number):

'A4_#####.ipynb'

Grading Rubric

This assignment is worth 12% of your grade (12 points).

There are 6 parts to this assignment, with the following point values:

Part 1	Loading & Cleaning	3.5 points
Part 2	Data Visualization	1 point
Part 3	Exploring the Data	1 point
Part 4	Distribution Testing	1 point
Part 5	Data Analysis	5 point
Part 6	Conclusions	0.5 points

Note that the assignment also contains optional parts 7 & 8, which serve to let you explore clustering and dimensionality reduction - but these sections are completely OPTIONAL and are UNGRADED.

Associated Data Files

- COGS108_IntroQuestionnaireData.csv