A5: Data Analysis

During the first lecture, you all answered a simple questionnaire, resulting in a messy CSV file containing the following information: your year at UCSD, major, age, places lived, gender, height, weight, favorite ice cream flavor, and favorite color. 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.
- To explore and apply data analysis methods, including linear modelling, clustering and dimensionality reduction.

Due Date

11:59 pm Friday, May 26th, 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 student ID number):

'A5_A########.ipynb'

Grading Rubric

This assignment is worth 15% of your grade (15 points). There are 8 parts to this assignment, with the following point values:

Part 1	Loading & Cleaing	1 point(s)
Part 2	Exploratory Data Viz	2 point(s)
Part 3	Exploring the Data	2 point(s)
Part 4	Data Analysis	6 point(s)
Part 5	Discussion & Conclusions	1 point(s)
Part 6	Testing Distributions	1 point(s)
Part 7	Clustering	1 point(s)
Part 8	Dimensionality Reduction	1 point(s)

Detailed Instructions

Detailed instructions are written into the assignment notebook.