Project Proposal

Pooja Kabber, Dingkun Yang, Echo Chen, Andrew Kroening

2022-09-30

For part 1, you will conduct exploratory data analysis on your selected dataset. You are required to produce a report of your exploratory data analysis findings in R Markdown. The report should be at most five pages. Tables and figures should be well formatted with clear labels and descriptions. You can organize the report as follows

Data Overview

Characteristics of the dataset, sample size, number of variables. Include questions here.

Primary Relationship of Interest

Present descriptive statistics and exploratory plots in whichever format you think is best (tables, figures) for your primary relationship of interest (dependent variable and primary independent variable, if applicable). Describe your findings.

Other Characteristics

Briefly describe other variables in the data. If there are many, do not list them all. Rather, describe the types of variables that are present (e.g., "demographic information").

Potential Challenges

Describe aspects of the data that may present challenges in the modeling stage. For example, might certain categorical variables need to be collapsed? Is there a lot of missingness? Could the size of the dataset present model selection challenges?