

## Project 1 Outline

Goal: Better understand the data science lifecycle by completing a data driven project.

1. Development of a research driven question focused on a dataset from either inside R or one of your choosing. This should include the development of a null and alternative hypothesis.
2. Provide the follow information in an Rmarkdown document including explanations of the steps being taken and rationale for each (Comment heavily, more context is better. Also make sure to include references for sampled code as needed.)
  - a. Conduct Exploratory Data Analysis that will begin to answer your question, this can include for example:
    - i. Summary of the dataset/Descriptive Statistics
    - ii. Graphical representations of the data
    - iii. Measures of Variance
    - iv. Initial correlation
  - b. Select and develop an appropriate model (Chi-square/ANOVA/Regression) to address your research question allowing you to either accept or reject the null hypothesis. This should include, as appropriate, model evaluation techniques.
3. Write a minimum 5 page summary report of your project in APA format. This should include the following sections:
  - a. Introduction and background research (citations).
  - b. Overview of the research question, hypotheses and methods (variable explanations).
  - c. Overview of the findings including EDA and model results.
  - d. A discussion of what conclusions can be drawn as a result of your project.
4. Develop a 10 minute presentation that effectively communicates the results of your data science project to be presented during class.

Grading:

1. 25%
2. 25%
3. 25%
4. 25%