Data Science 311 Final Project
Catchy and Descriptive Project Title

Group n
Member One, Member2, ..., Member$k$

Here is a template to use for your final project report. As a rule, avoid vague statements, include exact numbers, include and reference figures and tables, and also reference supporting files with code to reproduce results.

All supporting code and materials must be included in the final submission for your project. Supporting notebooks should not include commented out code and all code should run without producing any errors. The notebooks should be run and saved in the executed state to confirm the absence of errors.

1 Project Overview

At least one paragraph describing project goals, motivation, and plan.

2 Datasets

In this section include:

1. Reason for selecting the dataset.
2. Source of data. Who originally collected the data and why. Be precise here including the url where data is located as well as any special instructions or considerations when acquiring the data, e.g. (long download time, accounts needed, requirement to sign an agreement).
3. Explanation of data contents, e.g. relevant CSV fields and what they mean, missing values, and other data quirks.

3 Data curation

In this section discuss any steps you took to clean, process, merge or otherwise curate your respective datasets. Make sure to reference the relevant sections of your notebook used to for initial data processing.

4 Exploratory Data Analysis

This section will include the statement of data analysis questions, approaches to analyses, and resulting findings. There should be prose and reference to a relevant figure for each data analysis question. At the beginning of this section include some high level sentences discussing motivation and approach.

4.1 Question 1

- Data analysis question
- Figure
- Findings

4.2 Question 2

- Data analysis question
- Figure
- Findings

4.3 Question N

- Data analysis question
- Figure
- Findings

4.4 Conclusions

Discuss overall conclusions from exploratory data analysis.

5 Machine Learning

5.1 Approach

5.1.1 Machine learning problem

Inputs and outputs x and y

Loss function

Metrics Discuss the metrics you will use to assess the ML portion of the project.

5.1.2 Models

Baselines Describe the baseline/s you will use to compare the performance of your machine learning model.

Prospective algorithms Here describe your main approach to the ML problem. Be precise here citing the libraries you will use as well as the modeling approach.

5.1.3 Data splits

Explain your method for splitting the data. Be exact here citing the total number of data points in train/validation/test splits and any special considerations such as removal of outlier data points or balancing splits on some categorical features.

5.2 Results and analysis

Describe the outcome of your machine learning approach. This should include discussion of results according to your loss function on training, validation, and test sets with an accompanying figure and/or table. Some discussion on model behavior beyond reporting of metrics is needed here.

5.3 Concluding remarks

Make some general statements about the findings of the project, issues involved, and interesting next steps for future research and analysis.