5.1 Introduction

In this sprint you will be required to do three things:

1. Clean up any coding issues identified during grading.

2. Fix any outstanding issues in your initial report.

3. Put forth an initial analysis around your original questions.

Fixing your code

Please update and resubmit your code. You should make sure that your code is clean, in smart functions and follows the instructions outlined in the previous assignment. Make sure that the following are true about your code:

1. The code can be run from top to bottom.

2. All plots are generated from *functions*: e.g. `def plot_XXX()` and not in the mainline of the code.

3. Your plots functions should start from SQL, *not* from the loading your text or other files.

4. The code should be documented.

5. Variables should be easy to understand (no df_1, df_2, df_3, etc.).

Report

Please update your report according to both the comments in the report as well as the requirements below:

1. Are all plots easy to read and the correct size for the page?

2. Do your plot axis have *human* readable descriptions?

3. There should be no code in your write-up.

4. For each graph do you describe why it is interesting and what insights it is giving you?

Remember that the goal of this report is to be put online under your name, so quality matters! I saw a lot of low effort write-ups, please fix that.

Analyze away

Please go back to your initial Sprint #1 questions and answer each question! In class we have covered the basics of Classification and a few methods of doing it, we will also dive a bit into Linear Regression and Clustering over the next few classes. Using whatever method makes sense, please provide a short (2-3 pages) initial analysis as well as provide the code used to complete the analysis. To complete this assignment:

1. Write a brief sentence over why you choose a particular method.

2. Explain what result you got (numerically).

3. Explain (specifically) how it answered your Sprint #1 initial question.

SUBMISSION

Please submit TWO things:

1. A python file (not notebook), which contains:

   - Well written code (all in functions).
   - Your initial analysis (also all in functions).
   - Good comments explaining what you are doing and easy to read code.

2. A PDF file which has two sections:

   - The EDA section (Sprint #3) fixed according to the the grading comments and the requirements above.
   - A 2-3 page initial analysis section.