

$$\text{Total} = 1 + 7 + 11 + 8 + 6.5 + 10 + 10.5 = 55/80$$

Good effort.

With only 13 features remaining in your dataset after cleaning, I am concerned that you will not be able to complete the second part of the project. It's not clear how you will be able to make 3 different meaningful models to answer your question. It is likely that you will need to get more data.

0.Title (1) 1 Remove the word "Title"

1.(Brief) Introduction (10): 7

- Problem Context and Motivation (3) 3 Good.
- Describe Questions (4) 2 Can also include some specific subsidiary questions.
- Overview of Methodology (2) 1 The presented methodology just contains tasks but does not indicate why these tasks will help answer the question.
- Section Contents Overview (1) 1

2.Data Description (20): 11

- Origin (1) 1
- Contents (Who, What, Where) (3) 3
- Date Collected (When) (1) 1
- Collection Method (How) (2) 0
- Dataset Size (rows and columns) (1) 1
- Date Downloaded (1) 1
- Validations (3) 0 Need to compare key metrics against some external source to verify accuracy of data.
- Aspects of Data Quality (see Class Notes) (5) 3
- Ability to Answer Question (3): 1 How do the aspects of the data affect the outputs of your analysis – will be biased in any direction? Must consider if sample is representative or not.

3.Data Wrangling (10): 8

- Tidyness of Data (4) 4
- Techniques: Joins, removing duplicates, cleaning variables, changing types, aggregations (4) 4
- Sense checks after cleaning (raw vs cleaned) (2) 0 how has the cleaning process affected the dataset?

4.Exploratory Analyses (10): 6.5

- Choice of Initial Visualisations and Statistics (3) 2 Would like more commentary on the plots, especially on univariate plots. Further consideration of the scatterplot can reveal why correlations are low.
- Appropriate Measures (2) 2
- Relevance to Questions (3) 1.5 Why did you make these plots, and calculate these statistics? Need to tie it back to questions.
- Consideration of Accuracy (2) 1
- Overall, need to specify why you are plotting / calculating stats, relating it back to the questions. After plotting / calculating, you need to summarise the main

phenomena with respect to the questions you posed. You should also point out other interesting phenomena.

Presentation (14): 10

- appropriate use of tables and figures 1
- all figures and tables labelled with captions indicating items of interest 1
- plots are well presented: number of graphs on a single plot, use of multiple panes, etc. 1
- irrelevant information not presented 1.5
- long code blocks are hidden 1.5
- consistency of format 2
- appropriate use of colour 2

Data wrangling can be presented in more compact form.

Avoid long printed outputs.

Large percentage of the notebook consists of data wrangling, which is important, but the insights from the data are more important. Focus should be placed on analysis of data.

Many dataframes are printed without any heading or indication of what the purpose it.

Writing Style (16): 10.5

- coherent argument in clear english 2
- logical flow of argument 1
- main points and findings are clear 1
- all assumptions clearly indicated 1
- conciseness (no irrelevant points) 2
- appropriate use of subsections 1.5
- sections and figures labelled and referenced throughout text 1
- sentences and paragraphs are appropriate length 1

The exploratory analysis seems quite disjointed – its not clear what the aim of each plot / calculation is. Some more annotation is required.

Lack of discussion on outputs / tables / statistics / plots.

Dont use rubric contents as subsection headings.

Some very long sentences.