The data set you will be using contains prices and other attributes of nearly 54,000 diamonds. The data set is provided via a separate file (‘diamonds.csv’) which you will need to download from Canvas. A description of the attributes contained within the data set can be found here: <https://ggplot2.tidyverse.org/reference/diamonds.html>

Part 1: Data Summary – Explain how many use cases your data set provides; how many attributes are in each use case; what the data types are for each of the attributes; etc. Be sure to include any Python code used as part of your Data Summary work.

Part 2: Exploratory Data Analysis (EDA) – Provide summary statistics for each attribute; provide appropriate graphical analysis for each attribute using both Matplotlib and Seaborn. For example, if you believe it is appropriate to generate a histogram for a particular variable as part of your EDA, create it first using Matplotlib and then once again using Seaborn. Include a narrative describing your EDA findings. Be sure to include any Python code used as part of your EDA work.

Part 3: Inferences – Perform whatever analysis is necessary to answer the following questions:

1. What proportion of diamonds are between 0.30 and 1.08 carats?

2. How many of the diamonds have equal x and y dimensions?

3. How many of the diamonds have a carat value that is less than the mean carat value?

4. How many diamonds have a Premium cut or better? Note that the possible values for the quality of a cut are ranked in ascending order as follows: Fair / Good / Very Good / Premium / Ideal

5. Which diamond has the highest price per carat? What is its value?

6. Using both Matplotlib and Seaborn, make boxplots of the diamond price for each distinct cut value and discuss any conclusions you can draw from the appearance of the boxplots.

7. Using both Matplolib and Seaborn, make a scatter plot of price vs. carat. What can we say about the

relationship between those two attributes?

Provide a short-written narrative that explains your approach for each of these questions and tasks using formatted

Markdown cells in your Jupyter notebook. Be sure to include any Python code used as part of your work.

Part 4: Conclusion – A brief, concise narrative explaining your conclusions.

References- Be sure to include proper citations for any references you may have relied on as part of your

work.

Your Notebook deliverable should be similar to that of a publication-quality / professional caliber document and

should include clearly labeled graphics, high-quality formatting, clearly defined section and sub-section headers,

Python code should include succinct explanatory

comments.