AIM 5001 M8 Assignment (100 Points)

****You may work in small groups of no more than three (3) people for this project. ****

This assignment will allow you to demonstrate your ability to: (1) make use of Python's Pandas library; (2) perform basic exploratory data analysis on a provided data set; (3) create graphics using **Matplotlib** and **Seaborn** as part of your exploratory data analysis work; and (4) present your work in the form of a more "formal" research paper framework.

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

For this assignment, you will need to load the data file into your online AIM 5001 GitHub repository and then read the data from your GitHub repository into a Pandas dataframe. You will then use your Python and Pandas skills to answer and complete the content required for the outline specified below.

Your deliverable **must** include the following sections (with section headings + commentary **provided within formatted Markdown cells**):

Part 1: Data Summary (5 Points) – 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 include any Python code used as part of your Data Summary work.

Part 2: **Exploratory Data Analysis (EDA) (40 Points)** – 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 include any Python code used as part of your EDA work.

Part 3: Inferences (40 Points total) – Perform whatever analysis is necessary to answer the following questions:

1. **(4 Points)** What **proportion** of diamonds have a clarity of Sl1, Sl2, or VS2?
2. **(4 Points)** How many of the diamonds have a **length** that is less than ¾ of the mean diamond length?
3. **(4 Points)** How many of the diamonds have a **carat** value that is greater than the median carat value?
4. **(4 Points)** How many diamonds have either a '**Fair**' or a '**Premium**' cut? 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. **(8 Points)** Which diamond has the lowest **price per carat**? What is its value? Answer by providing the dataframe row index and the price per carat for that specific diamond.
6. **(8 Points)** Using both Matplolib and Seaborn, make and compare boxplots of **carat** metric for each distinct **clarity** value and discuss any conclusions you can draw from your comparison of the appearance of the boxplots.

7. **(8 Points)** Using both Matplolib and Seaborn, make a scatter plot of **carat vs. depth**. What can we say about the relationship between those two attributes?

Provide a short written narrative **in formatted Markdown cells** that explains your approach for each of these questions and tasks within your Jupyter notebook. Be sure to include any Python code used as part of your work.

Part 4: Conclusion (10 Points) – A brief, concise narrative explaining your conclusions.

References (5 Points) - Be sure to include proper citations for any references you may have relied on as part of your work.

Your Jupyter 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, and be free of spelling and grammar errors. Furthermore, your Pythion code should include succinct explanatory comments.

Save all of your work for this project within **a single Jupyter Notebook** and upload / submit it within the provided M8 Assignment Canvas submission portal. Be sure to save your Notebook using the following nomenclature : **first initial_last name_M8_assn**" (e.g., J_Smith_M8_assn). ***Small groups should identity all group members at the start of the Jupyter Notebook and each team member should submit their own copy of the team's work within Canvas.***