

DAV 5400 Fall 2019 Week 6 Assignment (30 Points)

This week's assignment continues our work with the Pandas library. The data set we'll 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 DAV5400 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 following questions and tasks:

1. What proportion of diamonds are between .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 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. Make a bar plot of the 'cut' data. What does it tell us about the frequencies of the individual cut values within the data set?
7. Make boxplots of the diamond price for each cut and discuss any conclusions you can draw from the appearance of the boxplots.
8. Make histograms of the length, width, depth, and total depth attributes and discuss any conclusions you can draw from the shape of the resulting distributions.
9. 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 the questions and tasks **using formatted Markdown cells in your Jupyter notebook**.

Save all of your work for this assignment within **a single Jupyter Notebook** and upload it to your online DAV5400 GitHub directory. Be sure to save your Notebook using the following nomenclature : **first initial_last name_w6_assn**" (e.g., J_Smith_w6_assn).

As a reminder, this assignment is due no later than 11.59pm on Sunday Oct 6.