Istanbul Medipol University

School of Engineering and Natural Sciences

Advanced Programming Laboratory Instructions

- 2 exercises will be held during laboratory sections.
- Second exercise will be done individually and grading will be done according to that performance.
- Exercises will be about the topics which have been held by Prof. Dr. Selim Akyokuş in that week.
- **Submissions after Friday 6:30 PM will not be accepted.**
- **For second exercise, any type of plagiarism is not allowed.**
- Please submit your exercises as YourDept_StudentID_Lab#.py and zip it as YourDept_StudentID_Lab#.zip
- Your codes should have comments. Codes with no comments will not receive full credit.
- For any questions,please contact me via Teams.

Exercise 1 (Quiz Question)

Perform the following tasks to study and analyze the diamonds dataset:

a) Download diamonds.csv from the given assignment.

b) Load the dataset into a pandas DataFrame with the following statement, which uses the first column of each record as the row index:
df = pd.read_csv('diamonds.csv', index_col=0)

c) Display the first seven rows of the DataFrame.

d) Display the last seven rows of the DataFrame.

e) Use the DataFrame method describe (which looks only at the numerical columns) to calculate the descriptive statistics for the numerical columns—carat, depth, table, price, x, y and z.

f) Use Series method describe to calculate the descriptive statistics for the categorical data (text) columns—cut, color and clarity.

g) What are the unique category values (use the Series method unique)?

h) Pandas has many built-in graphing capabilities. Execute the %matplotlib magic to enable Matplotlib support in IPython. Then, to view histograms of each numerical data column, call your DataFrame's hist method. The following figure;