COMP 4989: Lab Exercise #2   Mark: /40

Instructions:

In this lab, you will use Python to
- create line, bar, scatter and histogram plots using the Matplotlib package to visualize data used to predict the prices of an automobile
- select a feature most suited for use in a linear regression model based on the data visualization plots

Part 1: Data Visualization

[30 marks] In the first part of the lab, you will visualize data used to predict the price of an automobile. You will, for each feature, where applicable, use the Matplotlib package to create line, bar, scatter and histogram plots as well as determine the plot type that provides the most insight to each feature.

1. If you require an introduction to the Matplotlib package, https://matplotlib.org/ is a great place to start.

2. Download the preprocessed automobile price data, *AutomobilePrice_Lab2.csv*, which contains 193 rows and 25 columns (with continuous and categorical features), from BCIT Learning Hub (Content | Laboratory Material | Lab 2) and save it in your working directory.

3. Create a python script, *DataVisualization_Lab2.py*, to read data from *AutomobilePrice_Lab2.csv*. For each feature in the data, where applicable, plot a line, scatter, bar and histogram plot with the x-axis being the feature. As needed, you shall decide on the most appropriate variable for the y-axis in each plot. Your code shall plot all applicable plot types for all features. Each plot shall, at a minimum, contain a terse descriptive title, x-axis label, y-axis label and as necessary, a legend.

Deliverables:

All work submitted is subject to the standards of conduct as specified in BCIT Policy 5104. No late assignments will be accepted.

1) [20 marks] Ensure that your source code is adequately commented and submit *DataVisualization_Lab2.py* to BCIT Learning Hub (Laboratory Submission | Lab 2).
2) [10 marks] Submit a document, *PlotTypes_Lab2.txt*, to BCIT Learning Hub (Laboratory Submission | Lab 2) containing the following:
- list all applicable plots from Step (3). Clearly indicate the feature and the variable used for the x-axis and y-axis, respectively.
- based on the data visualization plots from Step (3), indicate the plot type that provides the most insight to the data for each feature.

Part 2: Data Visualization Feature Selection

[10 marks] Based solely on the data visualization plots from Part (1), select one (1) real-valued feature that you expect to be most suited for use in a linear regression model to predict the price of an automobile.

Deliverables:

All work submitted is subject to the standards of conduct as specified in BCIT Policy 5104. No late assignments will be accepted.

1) [5 marks] Append a discussion/justification of your selected feature to the document, *PlotTypes_Lab2.txt*, and submit it to BCIT Learning Hub (Laboratory Submission | Lab 2).

2) [5 marks] Append a discussion of the findings between Part (2) and COMP 4983 Lab Exercise #2 Part (2) to *PlotTypes_Lab2.txt* and submit it to BCIT Learning Hub (Laboratory Submission | Lab 2).