**Module Assignment**

**Module 1**

**QMB-6304 Analytical Methods for Business**

Write a simple R script to execute the following:

**Preprocessing**

1. Load the file “6304 Module 1 Assignment Data.xlsx” into R. This file contains information on 99 used automobiles for sale.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 30 cars using the method presented in class. This will be the data frame used for your assignment.

**Analysis**

Using R calculate and report the following:

1. The structure of the data object using the str() command.
2. Mean, Median, Standard Deviation, Skewness, and Kurtosis of the price variable. Based on these descriptive measurements how closely do you think this data conforms to a theoretical normal distribution?
3. A boxplot of the mileage variable. Based on this boxplot what can you say about the symmetry/skewness of this variable?
4. Quartiles of the mileage variable. Show your quartiles running from the minimum to maximum values for the variable, incrementing by .20.
5. A simple histogram of the price variable. Color your histogram green and give it an appropriate main title. Make sure the bottom axis of your histogram covers a range from 0 to $60,000. Based on this graphical tool would you say from this histogram the distribution of price follows a symmetric distribution, or a skewed distribution?
6. Three comparative boxplots for the price variable, one for each of the three makes of automobile included in. Your boxplots should be colored red and shown side by side with an appropriate main title and labels for the makes on the bottom axis. Based on these boxplots what can you say about the similarity in price between the three makes based on your sampled data?
7. Following Part 6 above, create three comparative boxplots for the mileage variable, one for each of the three makes of automobile included in. Your boxplots should be colored purple and shown side by side with an appropriate main title and labels for the makes on the bottom axis. Based on your boxplots for Parts 6 and 7, does there appear to be any difference in price or mileage across the three makes of automobile?

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions and 2) the results of those instructions. The first line of your script file should be a “#” comment line showing your name as it appears in Canvas. Results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. **This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.**