Patrick O'Boyle
CSC 346 Introduction to Data Science
Project 5
4/5/2022

Problem 1:

1. The purpose of this problem was to recreate the python notebook into the R programming language. In it, we examined linear regression with our data and built a regression model.
2. Methodology: We started by importing our data and checking for missing data. Then we made boxplots of the data and examined the distributions. We plotted the linear regression models in a scatter plot. We had to split the data into training and test cases and plot the models. Lastly, we obtained the MSE values and R2 values from the columns.
3. Conclusion: From the feature distributions you can see that the sales seem the most normal distribution while the newspaper seems to be right-skewed heavily. The TV and Radio seem to be normally distributed but not in good form.
4. References: Linear regression model slides.
   https://www.rdocumentation.org/packages/caTools/versions/1.17.1/topics/sample.split
   https://www.statology.org/how-to-calculate-mse-in-r/
   https://www.journaldev.com/47888/r-squared-in-r-programming#:~:text=R%20squared%20(R2)%20is%20a,for%20the%20response%2Ftarget%20variable.