Owen Campbell  
August 11th, 2024  
Fuel Efficiency Prediction Project

**Fuel Efficiency Prediction Project: Learnings and Reflections**

I created a Simple Linear Regression project. I explored and analyzed fuel efficiency data using various Python libraries. My motivation for creating this project was to dip my toes into machine learning for the first time. I have watched courses and lectures on machine learning, simple linear regression, and supervised learning. However, I have yet to use this knowledge practically in a project of my own to test my comprehension and understanding.

This project was the first time I interacted with these libraries alone without the assistance or guidance of a project tutorial. This document serves as a benchmark for my learning to look back on, and I hope it will also serve as a guide for where I need to focus my attention and improve my understanding in the present moment.

This project gave me the initial insight and practical learnings of my libraries, such as JupyterThemes. While I now understand that diving deeper into this library is more about customization for aesthetics of Jupyter Notebooks rather than functionality, it gave me a fundamental understanding of calling upon it in the future for other machine learning projects. I used Pandas to perform fundamental data transformations such as CSV reading and reshaping. While I have a basic grasp of these critical functionalities, I plan to dedicate more time to learning the Pandas library as a sole focus in the future to understand better the power of importing this library. The most significant learning leap I made during this project's creation is the Matplotlib library. I called on this library heavily for creating a simple linear regression model (specifically matplotlib.pyplot), such as creating different plot elements, which include plot types, labels, and figure size. Another big learning moment in this project was interacting with sklearn.model\_selection to define test and training sets. While learning this sublibrary happened outside of the code block about the fundamentals of machine learning training and testing concepts, I am glad I was able to implement and understand how to implement these concepts in this simple linear regression project. A final benchmark of learning for this project was figuring out how to create proper documentation. Unfortunately, in my associate's program for software development, creating documentation is a topic that needs to be discussed or taught, so I had to call upon some help from the internet and peers to teach me what is necessary to document a software project. While I understand that each company has its own methods for creating documentation, and requirements might change depending on whom you ask, I have a solid yet professional starting point for creating proper documentation.

I faced a couple of challenges during the creation of this project. The biggest challenge I faced was extending the plot's linear regression line. When I created and showed the plot, the linear regression line did not extend across the entirety of the plot's data points and did not allow me to visually see if the test or train data points fit the best-fit line. The solution to this problem leads me to the second challenge of this project: a need for more understanding and experience with the Numpy library. I had to do a lot of reading and problem-solving to gap my understanding, which would help me eventually find the solution to extending the linear regression line (the solution was concatenating both X\_train and X\_test values to a single array variable, which then you would sort with Numpy which would give you the proper range to extend the linear regression line to all data points on the scatter plot). If there is any takeaway from this project, I plan to dedicate more time to understanding everything the Numpy library has to offer so I do not have to spin my wheels longer than I need to on a solution so simple.

This project contributed massively to my journey of learning simple linear regression, the various libraries that aid in making this occur, and hopefully, to more significant, more complex machine learning concepts in the future. I plan to focus on separate projects dedicated to improving my proficiency in Pandas and Numpy and exploring more advanced topics in machine learning. This project has provided a solid foundation, and I now understand where my strengths lie and the areas that require further development. Moving forward, I am excited to build on what I have learned and take on more challenging projects that will further my understanding. This document will serve as a personal benchmark to measure my growth. I hope it reflects my dedication to continuous learning and improving data science and machine learning.