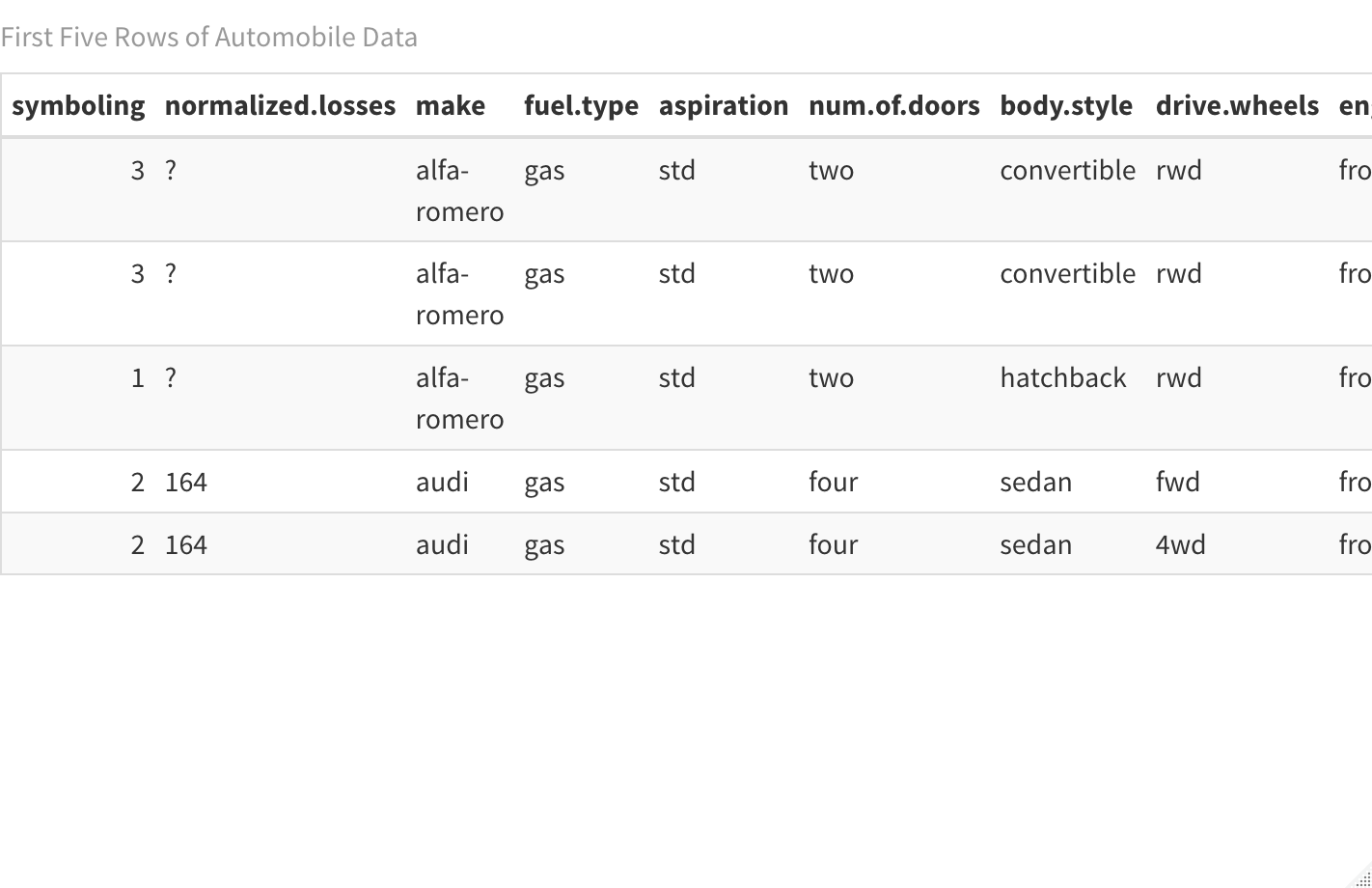
**Evaluating Automobile Statistics**

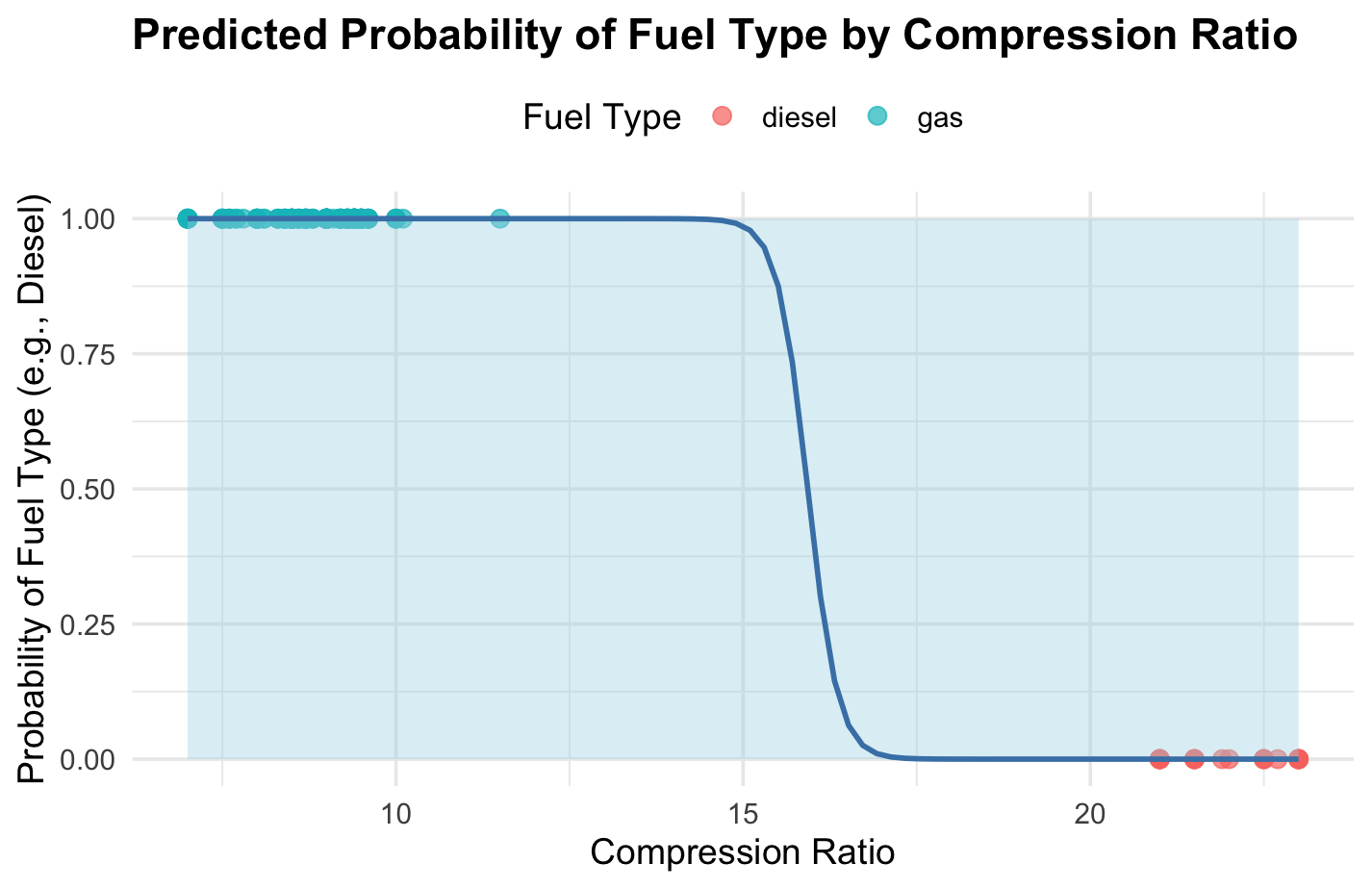
Nirav Naidu

**Introduction**

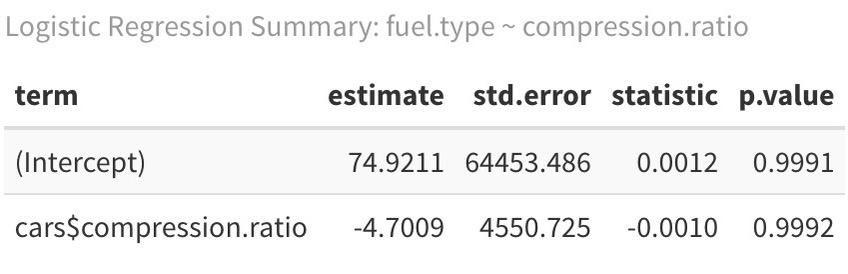
The data set that was utilized is “Automobile\_data”, a data set from Kaggle. This data set allows you to see various statistics of cars from city miles per gallon, highway miles per gallon, type of car, engine size, etc. There are multiple variables that are being assessed. First, we assessed a relationship between fuel type and compression ratio using logistic regression. Next, we assessed a multiple logistic regression model using fuel type, compression ratio, and peak revolutions per minute (RPM). Some variable types were changed, for instance, we changed fuel type from a character to a factor with two levels, since the only fuel types included are gasoline and diesel. We also did the same for peak RPM. We use this data to rebuild vehicles.

**First Five Rows Table**

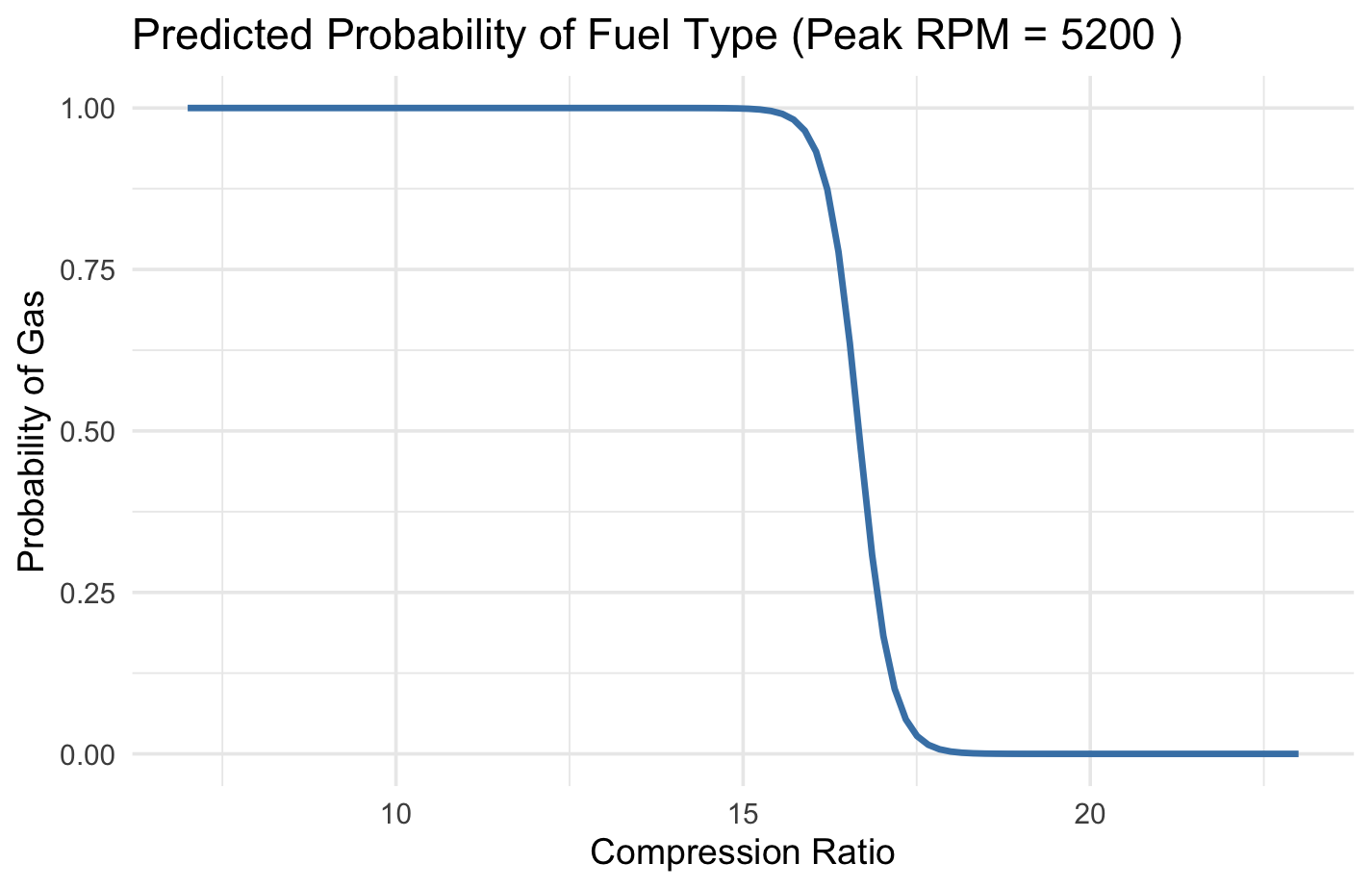
Using the table to the left, we can see the table for the first five rows of the dataset. There are more variables that we were not able to fit in the image, however, it gives us enough information.

**First Logistic Regression Model**

Using the sigmoidal model to the left, we can see that this model has a steep S-curve and is going in the negative direction. This means that vehicles that have lower compression ratios are more likely to be using gasoline, while those with higher compression ratios are less likely to use gasoline.

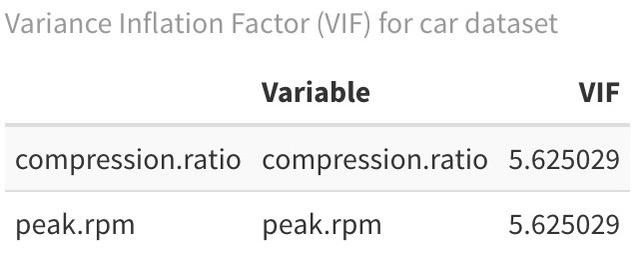
**Summary of First Logistic Regression Model**

Using the table to the left, we can see the summary statistics of the first logistic regression. Cars with higher compression ratios are less likely to use gasoline . We can see the p-values for both variables is higher than 0.05, meaning that our variables are not useful.

**Multiple Logistic Regression Model**

Using the graph to the left, we can see the multiple logistic regression. The variable that was added was peak RPM. Much like the first logistic regression model, we can see that vehicles with 5200 peak RPM, vehicles with lower compression ratios are more likely to use gasoline, while those with higher compression ratios are less likely to be using gasoline.

**Model Checks & Diagnostics**

Using the table to the right, we can see the variance inflation factor (VIF) for the multiple logistic regression model. We can see that the VIF is moderately high, meaning that there is some multicollinearity. This means that peak RPM does not bring new or useful information.

**References**

* ChatGPT was used to make the graphics look better.
* A point was earned during class for answering a question.