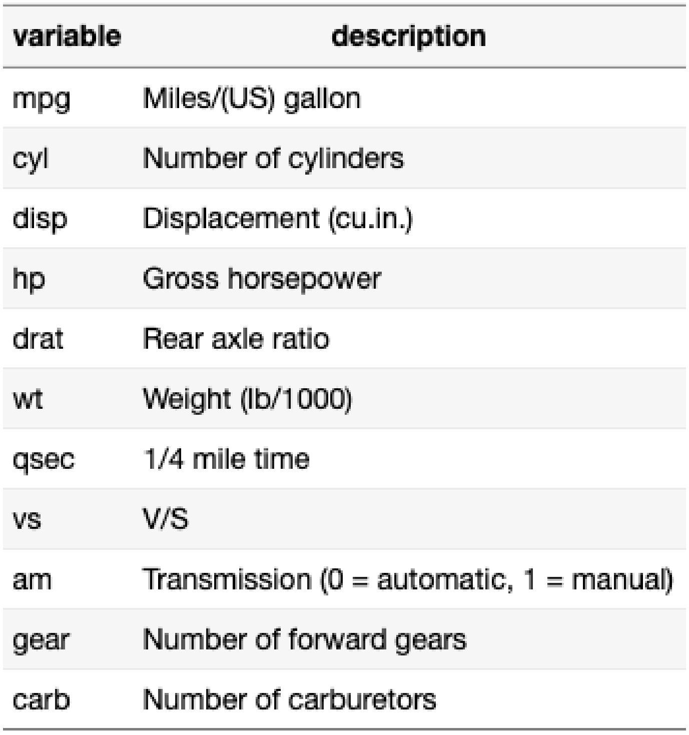
Step 1: Loading mtcars dataset.
There are contains 32 obs. of 11 variables.
Notably all the variables are Numerical.



variable	description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

Step 2: Loading required known library
Ridge - Linear and logistic ridge regression functions.
Glmnet - Lasso and Elastic-Net Regularized Generalized Linear Models

Step 3: Perform the exploratory data analysis.
Str
Summary
Missing values
Empty values
Duplicate

Step 4: Density plot
Slightly Right Skewed, which implies most of the values are positive in nature.
Mode > Median > Mean values
Skewness value is > 0, so data values are less than mean

Step 5: Correlation Heatmap
Darker shades denotes less or -ve correlations
Lighter shades denotes high or +ve correlations
Assuming multicollinearity is present.

Step 6: Checking for outliers in highly positive correlated values.
Clearly outlier are there qsec(1/4 mile time), Gross horsepower, Weight (lb/1000).

Step 7: Initial Linear Regression Model

Random Sample with 70:30, Train and test data ratio.
Variance Inflation Factor for multicollinearity check.
Multicollinearity is identified through VIF

Step 8: Multi Linear Regression Model

Predict and Compare Response variable Millage
Accuracy – 0.86 => 86 %
RMSE – 3.242

Step 9: Optimum lamba value

lambda_seq is created
Cross validation is impleted by nfolds value 5
best_lam is found to be 2.29 and ridge_model1

Extract the model using k-cross validation

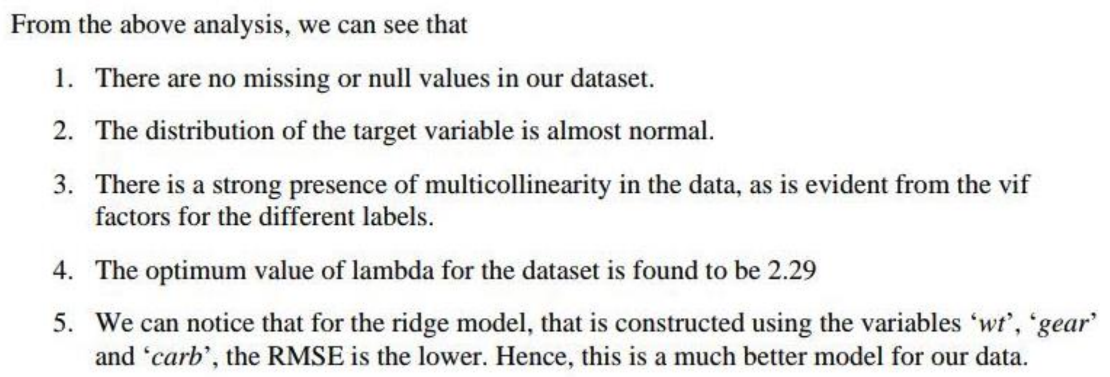
Step 10: Build the final model

With all variables

Accuracy – 0.90 => 90 %
RMSE – 2.532

With significant variables

Accuracy – 0.94 => 94 %
RMSE – 1.389



From the above analysis, we can see that

1. There are no missing or null values in our dataset.
2. The distribution of the target variable is almost normal.
3. There is a strong presence of multicollinearity in the data, as is evident from the vif factors for the different labels.
4. The optimum value of lambda for the dataset is found to be 2.29
5. We can notice that for the ridge model, that is constructed using the variables '*wt*', '*gear*' and '*carb*', the RMSE is the lower. Hence, this is a much better model for our data.