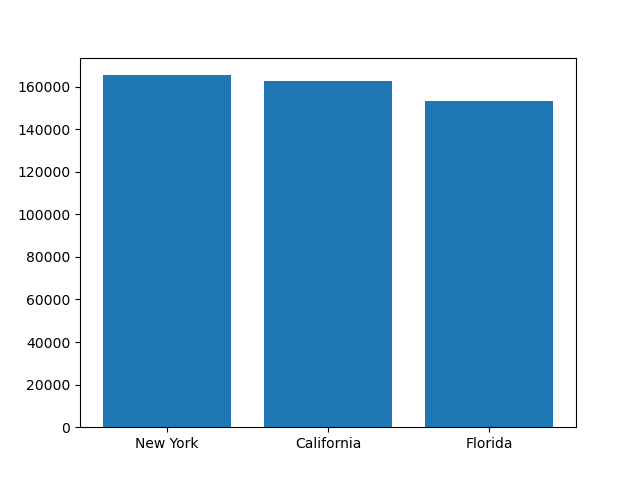
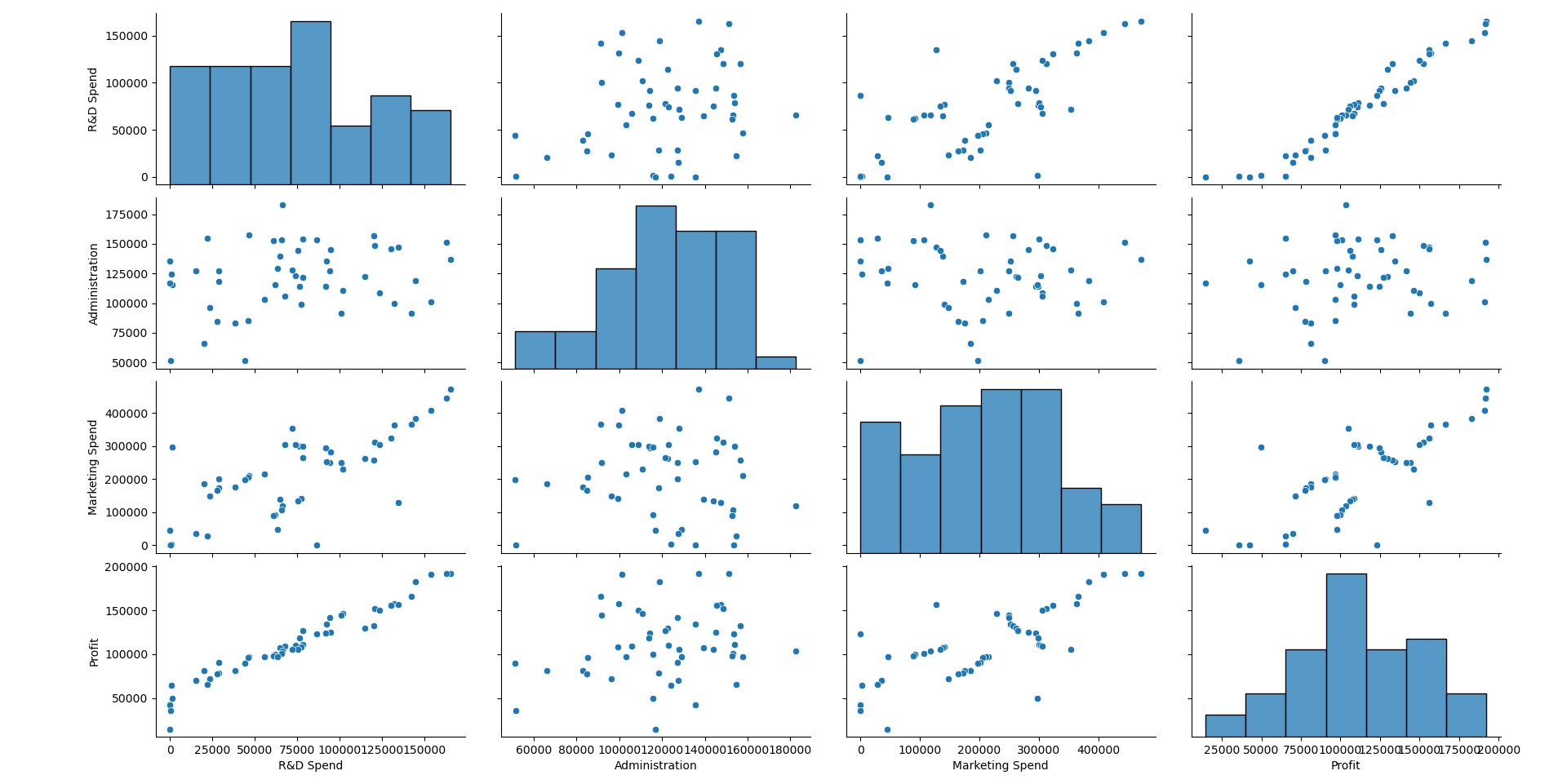
LRR REGRESSION

1. 50 START UPS

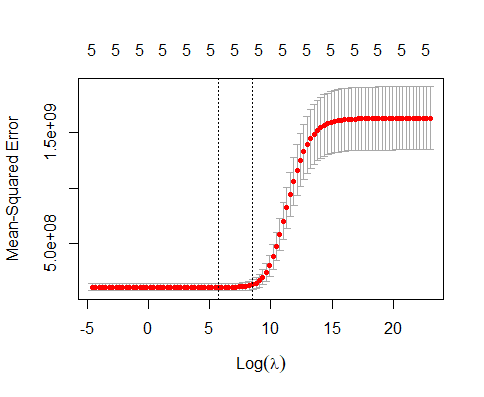


For all 3 states New York has the highest Marketing and RnD.



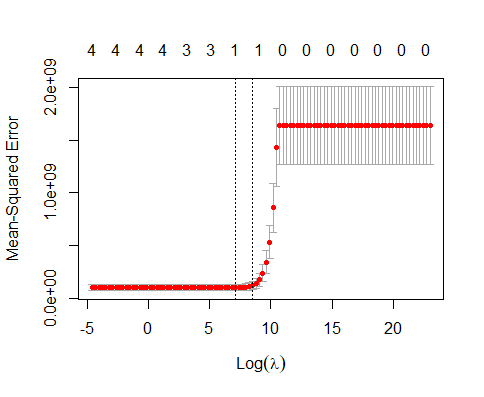
Also with Rndand Profit there is a good correlation. That is with every unit increase in RndSpend there is also 1 unit increase in Profit earned (that is oour predictor) .

* Rnd and Profit
* Marketing Spend and profit
* Marketing and Rnd (multicollinearity)
* Meaning these explain the variances in our dataset.



The Ridge Plot:

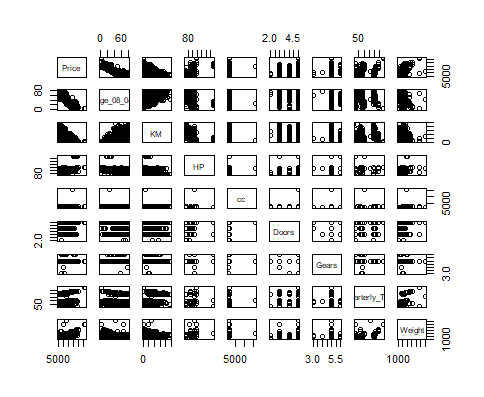
The MSE and Log(lambda) is responsible for our slope steepness and sensitivity of X to y.



The Lasso plot rather have significant stability after 1 and remains constant after 0 around when log(lambda) is 10.

1. Toyota Corolla

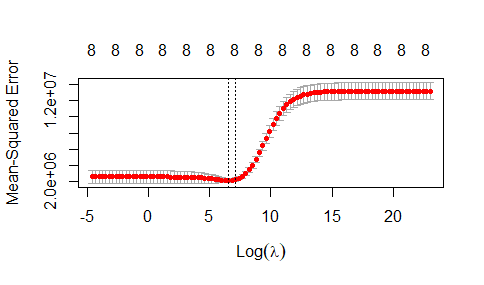
Pairplot



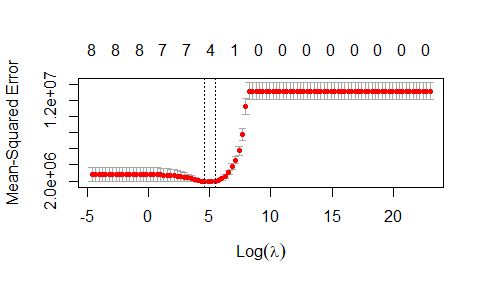
KM and Age has a corr and its hetroscedasticity

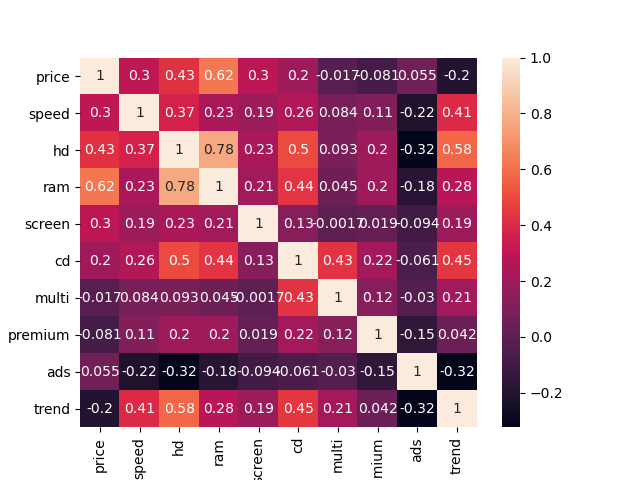
Using Lasso we can shrink our parameter till 0 where as the slope in Ridge will be closer to 0.

Cv plot using Ridge for Corolla Dataset:



Cv Plot using Lasso for Corolla Dataset

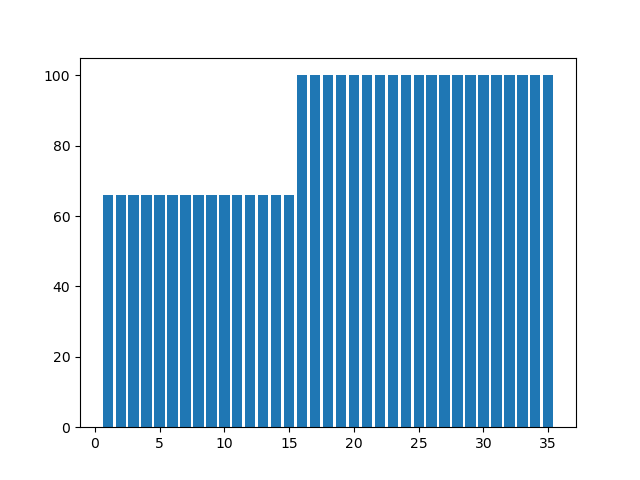


3.)Computer Data : 

Heatmap for correlation between each variable

Higher corr observed for RAM and HD

Multi and Speed , Screen and ads



This barplot is for Speed v Trend. As clearly seen. Most trend is for Higher Speed might be RAM, CPU, or GPU.

# we need to find the optimal value of alpha (hyper tuning parameter ) for our dataset

#-----------------------------------------------------------------------------------------------------

from sklearn.model\_selection import GridSearchCV

from sklearn.model\_selection import RepeatedKFold

# LASSO ADJUSTED R SCORE after using K Fold and gridSearch to find the god value for Alpha as 0.11

lasso.score(X,y) #0.7715882298605266 -->0.7715882298605266

# Ridge ADJUSTED R SCORE

ridge\_Adj.score(X,y) #0.5350889766460389

# RMSE

np.sqrt(np.mean((pred\_ridge - y)\*\*2)) #395.9859628078518

# Ridge model performed worse than Lasso did even with GridSearch where we evaluated the perfect value for alpha