FACTORS AFFECTING THE PRICING OF CARS IN THE NIGERIAN MARKET

OBJECTIVES

This is to build a model that will predict the price of cars with the available independent variables that will use it to understand how exactly the prices vary with the independent variables.

DATA OVERVIEW

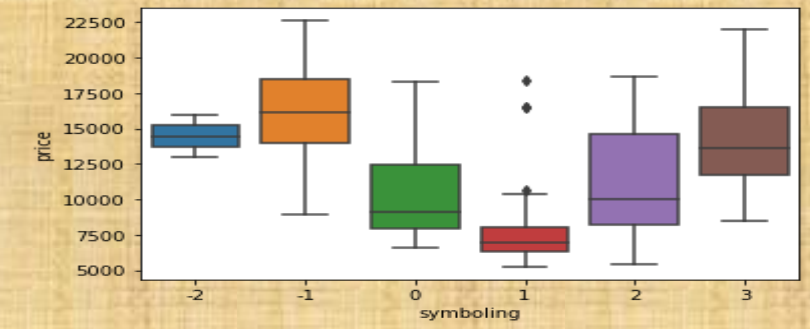
The prediction factors are:

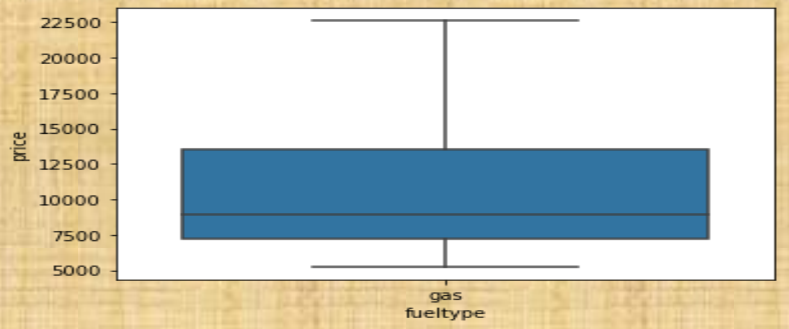
Symbolling

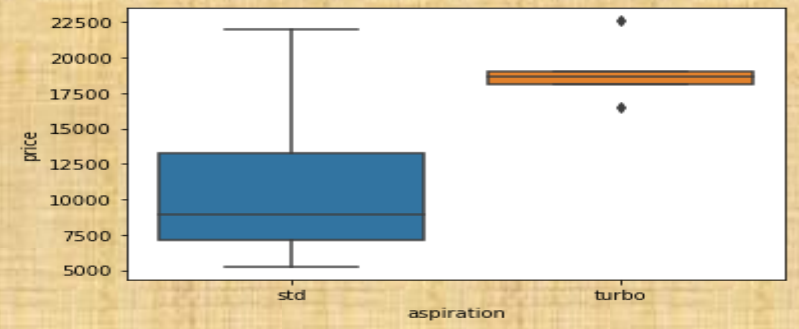
CAR COMPANY
Fuel type

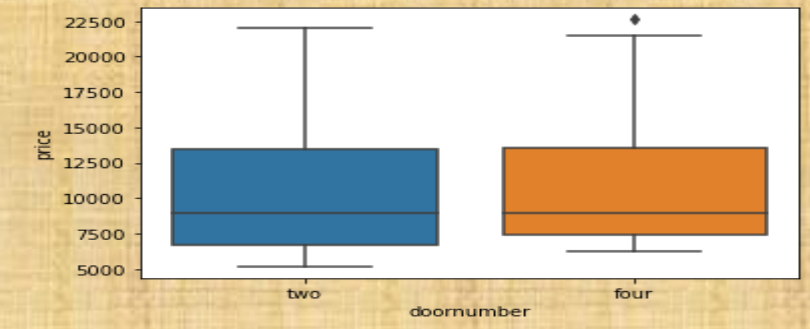
Aspiration
Door number
Car body
Drive wheel
Engine location
Wheel base
Car length
Car width
Car height

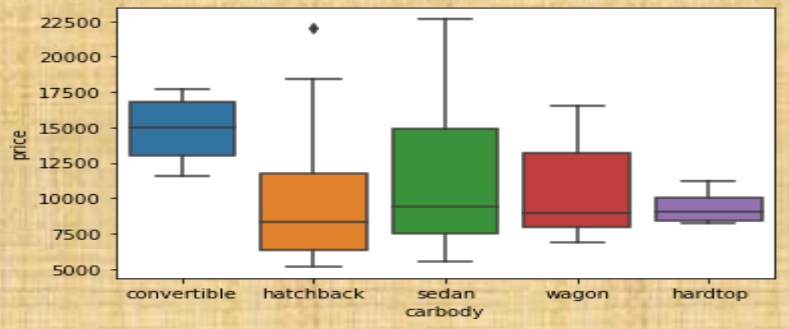
- **Curb weight**
  Engine type
  Cylinder number
  Engine size
  Fuel system
  Bore ratio
  Stroke
  Compression ratio
  Horsepower
  Peak rpm
  City mpg
  High way mpg
  price(Dependent variable)

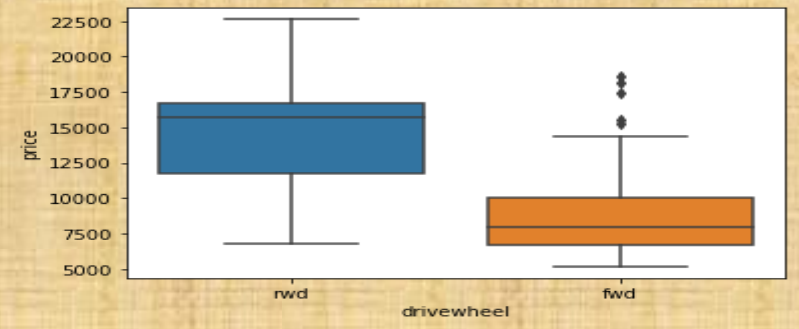

price
22500
20000
17500
15000
12500
10000
7500
5000
-2
-1
0
1
2
3
symboling

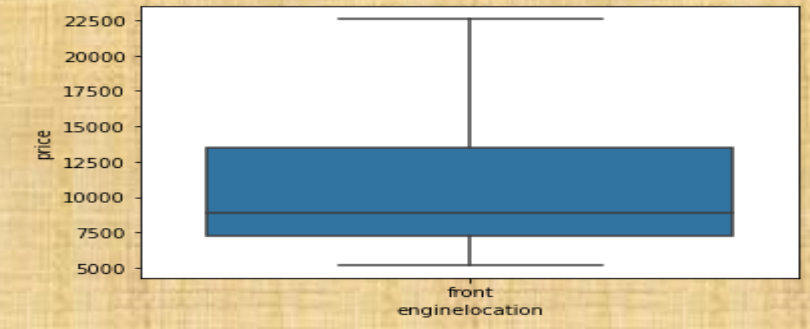

price
22500
20000
17500
15000
12500
10000
7500
5000
gas
fueltype

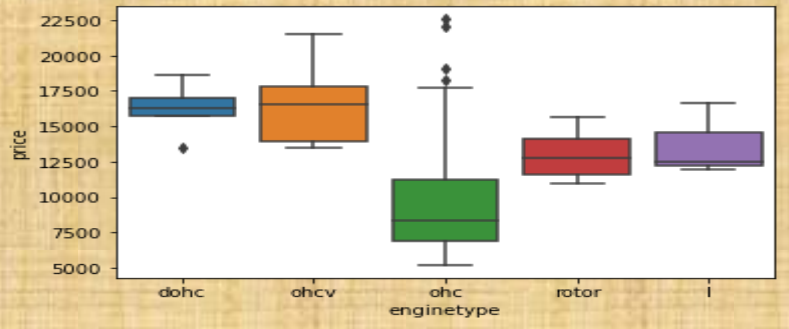

price
22500
20000
17500
15000
12500
10000
7500
5000
std
turbo
aspiration

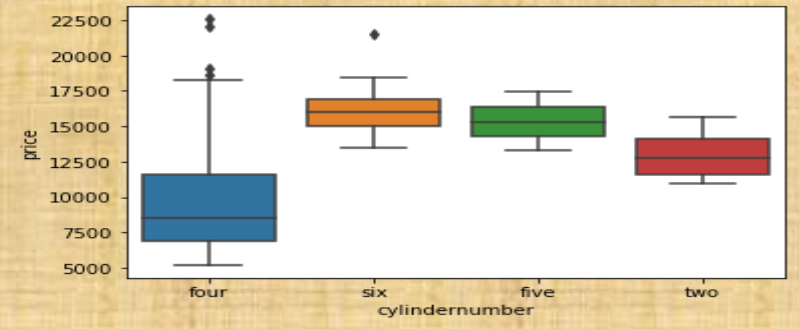

price
22500
20000
17500
15000
12500
10000
7500
5000
two
four
doornumber

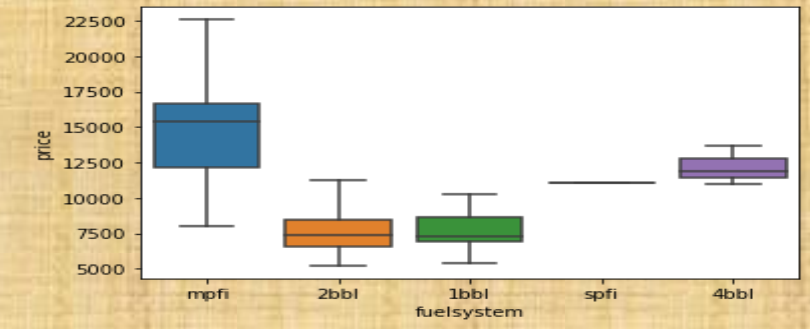

price
22500
20000
17500
15000
12500
10000
7500
5000
convertible
hatchback
sedan
wagon
hardtop
carbody

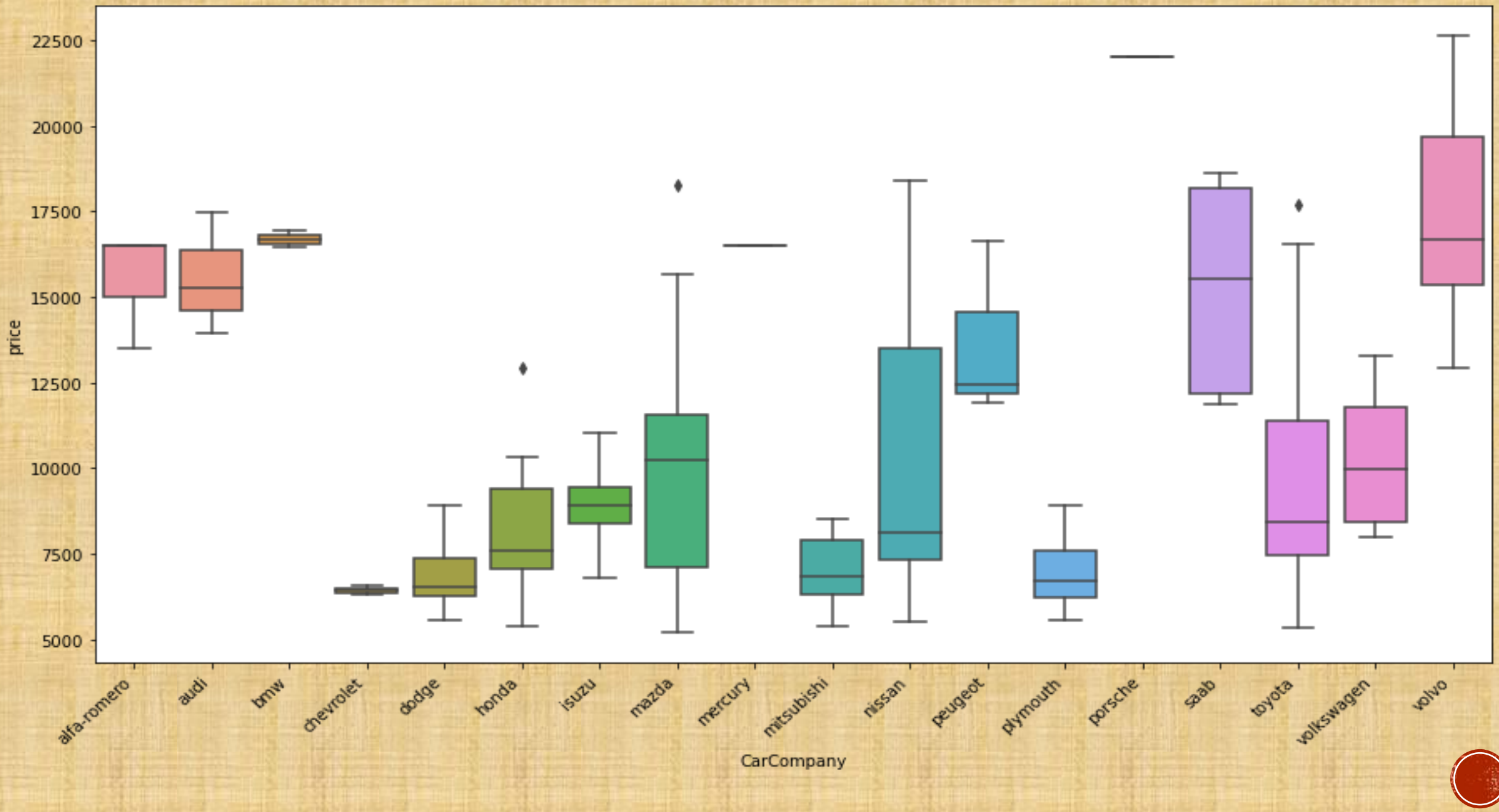

price
22500
20000
17500
15000
12500
10000
7500
5000
rwd
fwd
drivewheel


price
22500
20000
17500
15000
12500
10000
7500
5000
front
enginelocation


price
22500
20000
17500
15000
12500
10000
7500
5000
dohc
ohcv
ohc
rotor
l
enginetype


price
22500
20000
17500
15000
12500
10000
7500
5000
four
six
five
two
cylindernumber

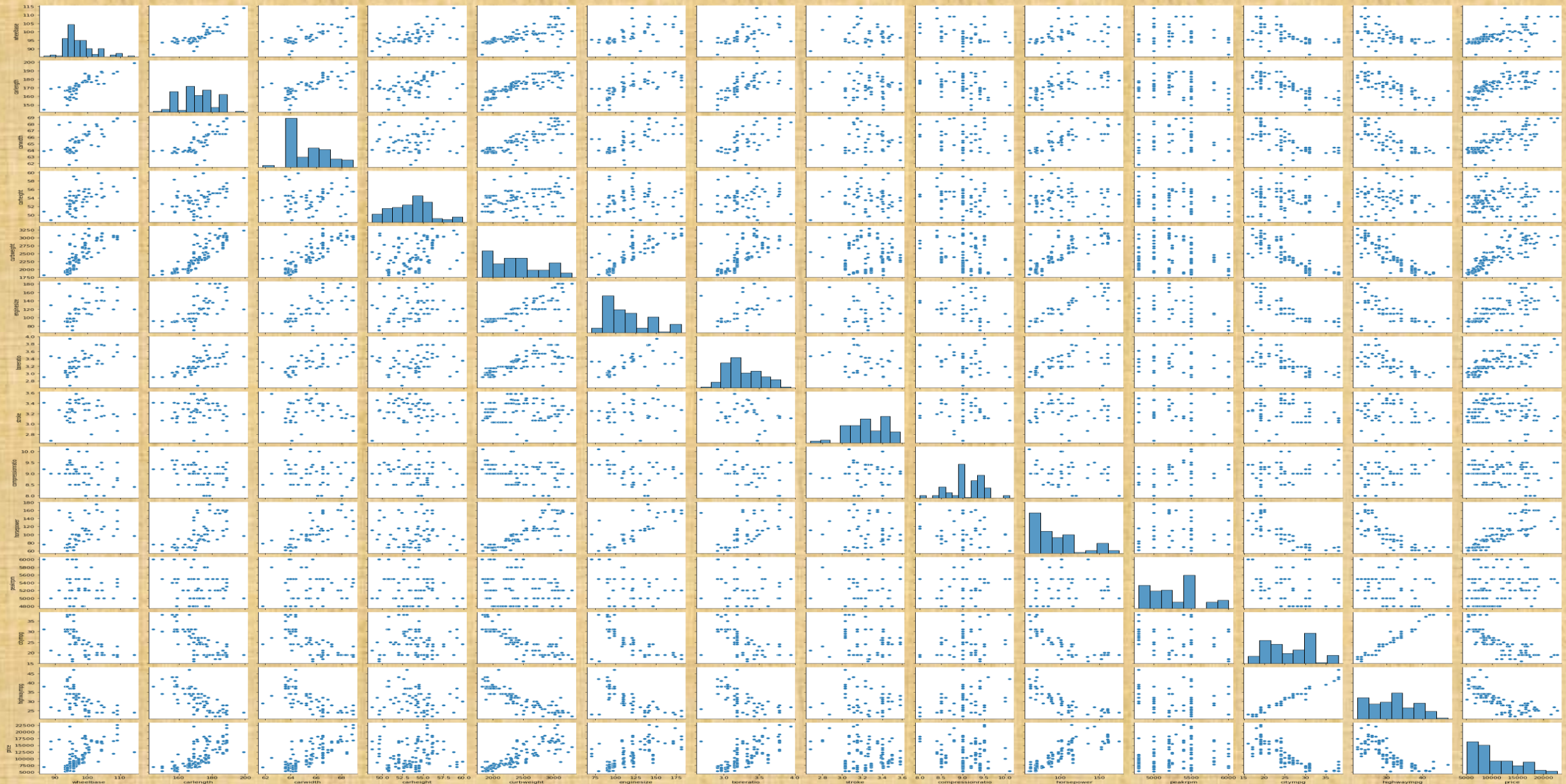

price
22500
20000
17500
15000
12500
10000
7500
5000
mpfi
2bbl
1bbl
spfi
4bbl
fuelsystem


22500
20000
17500
15000
12500
10000
7500
5000
price
alfa-romero
audi
bmw
chevrolet
dodge
honda
isuzu
mazda
mercury
mitsubishi
nissan
peugeot
plymouth
porsche
saab
toyota
volkswagen
volvo
CarCompany

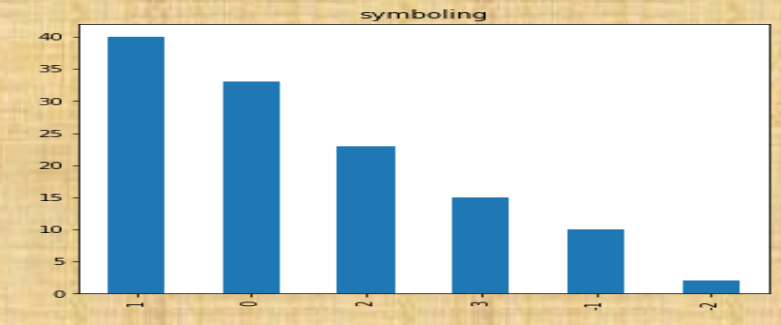
OBSERVATIONS

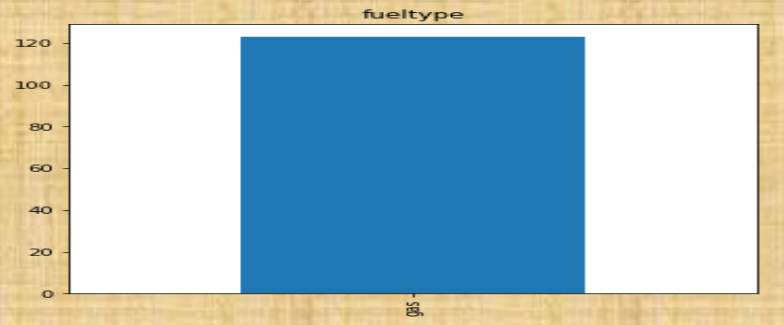
- Car Company - Porsche has very high median price compared to other cars, though the number of Porsche cars is very less.
- Volvo, alfa-romero, Audi and BMW are also high median price than others. Saab has wide rage of price, with high median price. aspiration - std has lower median than turbo.
- Car body - convertible has higher median that others. symbolling - -2 and -1 have higher median price than others.
- Engine location - rear has very high median price than front.
- Cylinder number - Four has lower median than others.
- Fuel system - 1bbl and 2bbl have lower median price than others.
- Now at least we know that what are the variables have impact on the price. So as which variables are important for the model building.

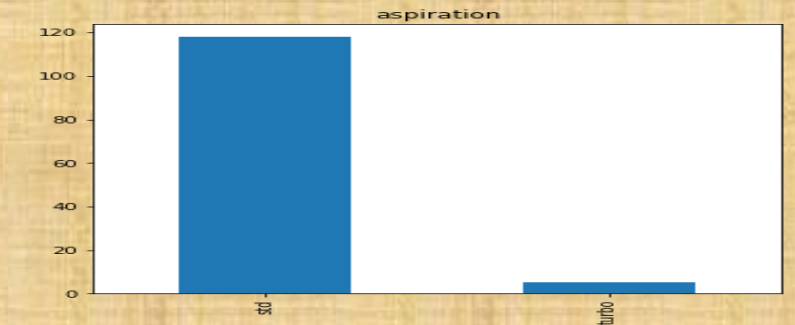
PAIR PLOT

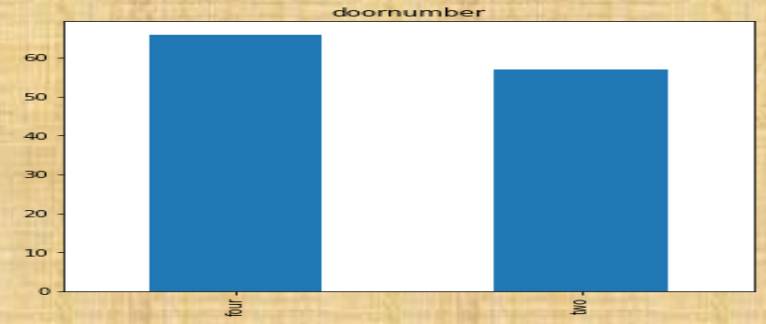

wheelbase
carlength
carwidth
carheight
curbweight
enginesize
boreratio
stroke
compressionratio
horsepower
peakrpm
citympg
highwaympg
price

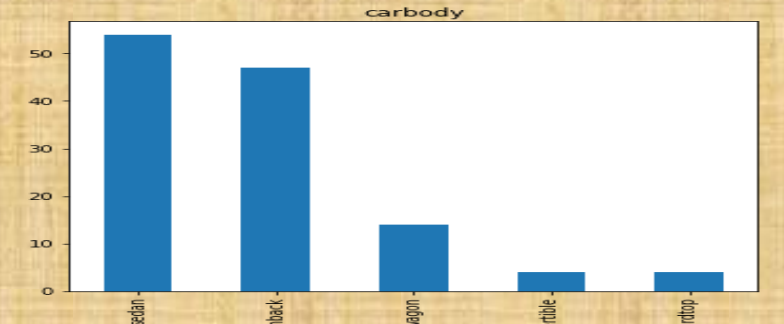
We can see From the pair plot that there are few columns that have linear relationship with the target variable "price"..

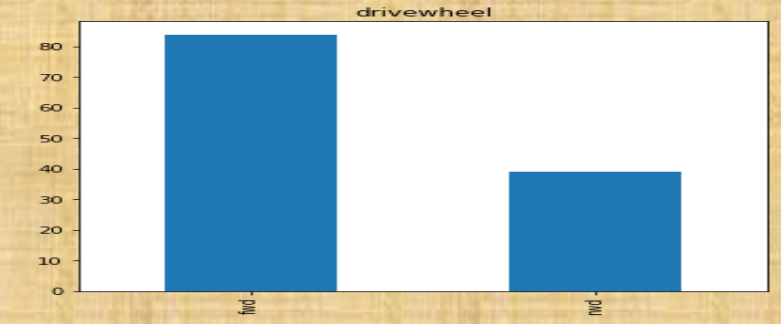

symboling
1
0
2
3
-1
-2

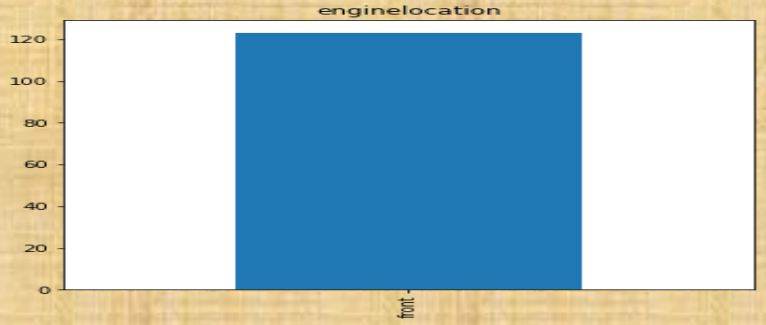

fueltype
gas

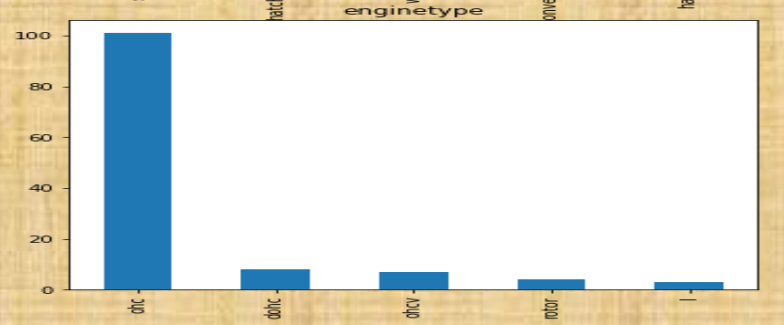

aspiration
std
turbo

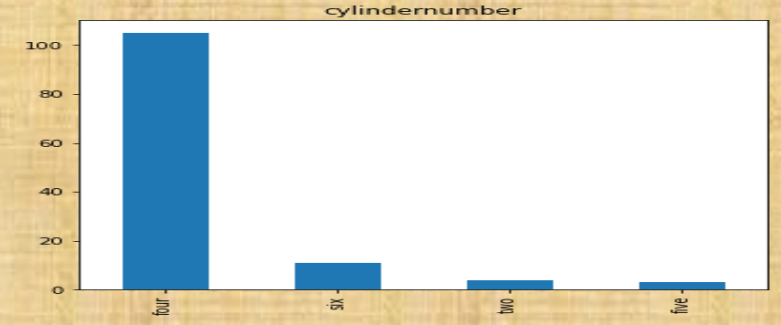

doornumber
four
two

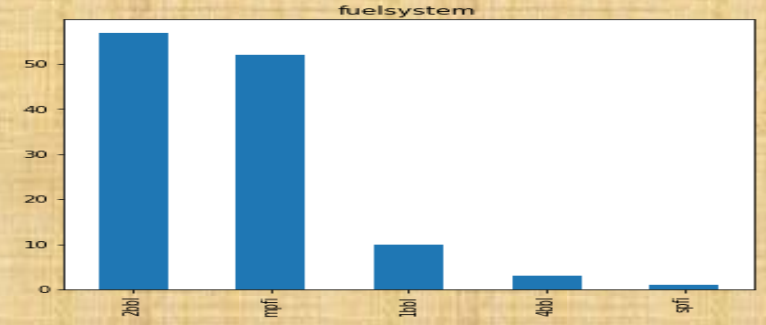

carbody
sedan
hatchback
wagon
convertible
hardtop

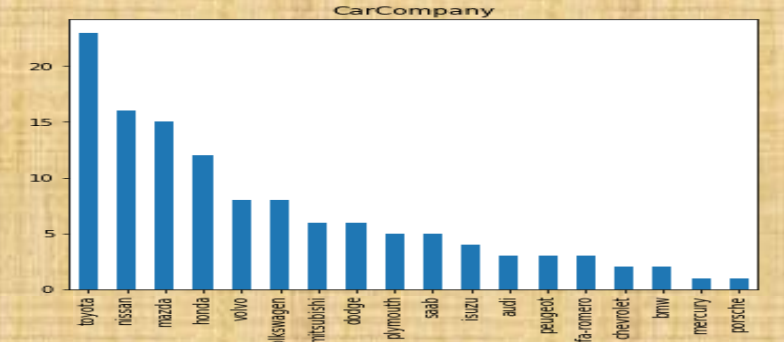

drivewheel
fwd
rwd


enginelocation
front

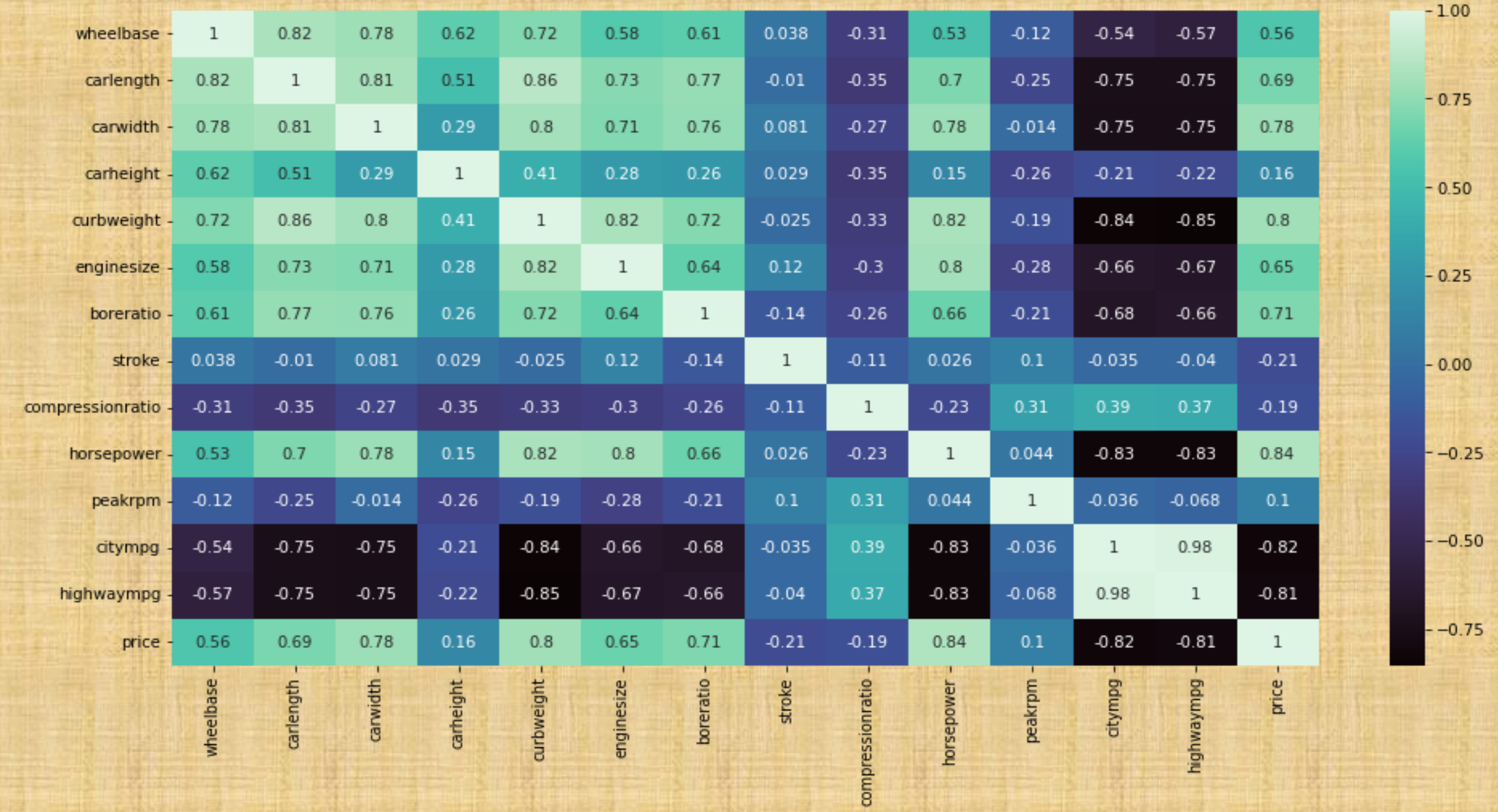

enginetype
ohc
ohcf
ohcv
rotor
l


cylindernumber
four
six
two
five


fuelsystem
2bbl
mpfi
1bbl
4bbl
spfi


CarCompany
toyota
nissan
mazda
honda
volvo
volkswagen
mitsubishi
dodge
plymouth
saab
isuzu
audi
peugeot
alfa-romero
chevrolet
bmw
mercury
porsche

- **Observations**
- We can see that there is data imbalance in below columns:-
- aspiration - Lesser number of turbo than std.
- Engine loaction - All the engine location is in front, as all the rear engine cars were removed while removing outliers.
- Cylinder number - Large number of four cylinders than others.
- Fuel system - mpfi and 2bbl fuel system cars are more comparatively others.
- Car Company - Most of the Toyata company cars were surveyed.


wheelbase 1 0.82 0.78 0.62 0.72 0.58 0.61 0.038 -0.31 0.53 -0.12 -0.54 -0.57 0.56
carlength 0.82 1 0.81 0.51 0.86 0.73 0.77 -0.01 -0.35 0.7 -0.25 -0.75 -0.75 0.69
carwidth 0.78 0.81 1 0.29 0.8 0.71 0.76 0.081 -0.27 0.78 -0.014 -0.75 -0.75 0.78
carheight 0.62 0.51 0.29 1 0.41 0.28 0.26 0.029 -0.35 0.15 -0.26 -0.21 -0.22 0.16
curbweight 0.72 0.86 0.8 0.41 1 0.82 0.72 -0.025 -0.33 0.82 -0.19 -0.84 -0.85 0.8
enginesize 0.58 0.73 0.71 0.28 0.82 1 0.64 0.12 -0.3 0.8 -0.28 -0.66 -0.67 0.65
boreratio 0.61 0.77 0.76 0.26 0.72 0.64 1 -0.14 -0.26 0.66 -0.21 -0.68 -0.66 0.71
stroke 0.038 -0.01 0.081 0.029 -0.025 0.12 -0.14 1 -0.11 0.026 0.1 -0.035 -0.04 -0.21
compressionratio -0.31 -0.35 -0.27 -0.35 -0.33 -0.3 -0.26 -0.11 1 -0.23 0.31 0.39 0.37 -0.19
horsepower 0.53 0.7 0.78 0.15 0.82 0.8 0.66 0.026 -0.23 1 0.044 -0.83 -0.83 0.84
peakrpm -0.12 -0.25 -0.014 -0.26 -0.19 -0.28 -0.21 0.1 0.31 0.044 1 -0.036 -0.068 0.1
citympg -0.54 -0.75 -0.75 -0.21 -0.84 -0.66 -0.68 -0.035 0.39 -0.83 -0.036 1 0.98 -0.82
highwaympg -0.57 -0.75 -0.75 -0.22 -0.85 -0.67 -0.66 -0.04 0.37 -0.83 -0.068 0.98 1 -0.81
price 0.56 0.69 0.78 0.16 0.8 0.65 0.71 -0.21 -0.19 0.84 0.1 -0.82 -0.81 1
wheelbase
carlength
carwidth
carheight
curbweight
enginesize
boreratio
stroke
compressionratio
horsepower
peakrpm
citympg
highwaympg
price
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

- The heatmap shows some useful insights:
- Correlation of **target variable "Price"** with **independent variables**:

- Price is highly (positively) correlated with wheelbase, car length, car width, curb weight, engine size, horsepower (notice how all of these variables represent the size/weight/engine power of the car)
- Price is negatively correlated to 'city mpg' and 'highway mpg' (-0.70 approximately). This suggest that cars having high mileage may fall in the 'economy' cars category, and are priced lower (think Maruti Alto/Swift type of cars, which are designed to be affordable by the middle class, who value mileage more than horsepower/size of car etc.)

- Correlation among independent variables:

  - Many independent variables are highly correlated (look at the top-left part of matrix): wheelbase, car length, curb weight, engine size etc. are all measures of ‘size/weight’, and are positively correlated

Conclussion

- Variables that are useful to describe the variances in car prices are present prices, Year, kms_driven, fuel type, seller type, and transmission. Our final model has satisfied the classical assumptions. The R-squared of the model is high, with 97,86% of the variables can explain the variances in the car price. The accuracy of the model in predicting the car price is measured with RMSE, with model has RMSE of 0.798 and prediction data has RMSE of 0.818, suggesting that our model may fit the prediction dataset.