# CAR PRICE PREDICTION MODEL REPORT

OLOKPO MERCY AKWUM

# OBJECTIVES

The objectives of this project are:

To get the features that are important in predicting the price of car

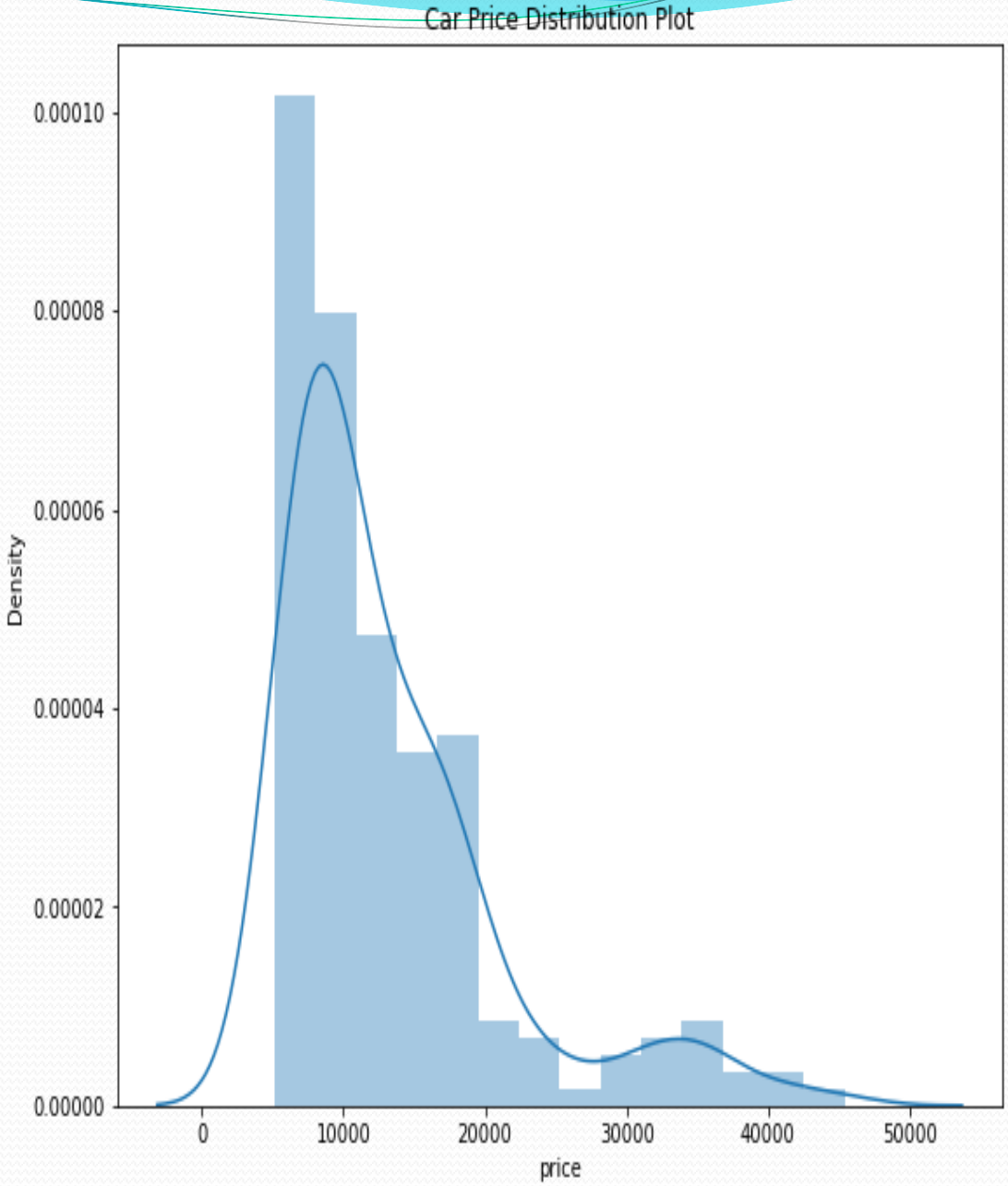
To build a model for proper understanding of the pricing dynamics

# UNIVARIATE ANALYSIS

## CAR PRICE DISTRIBUTION

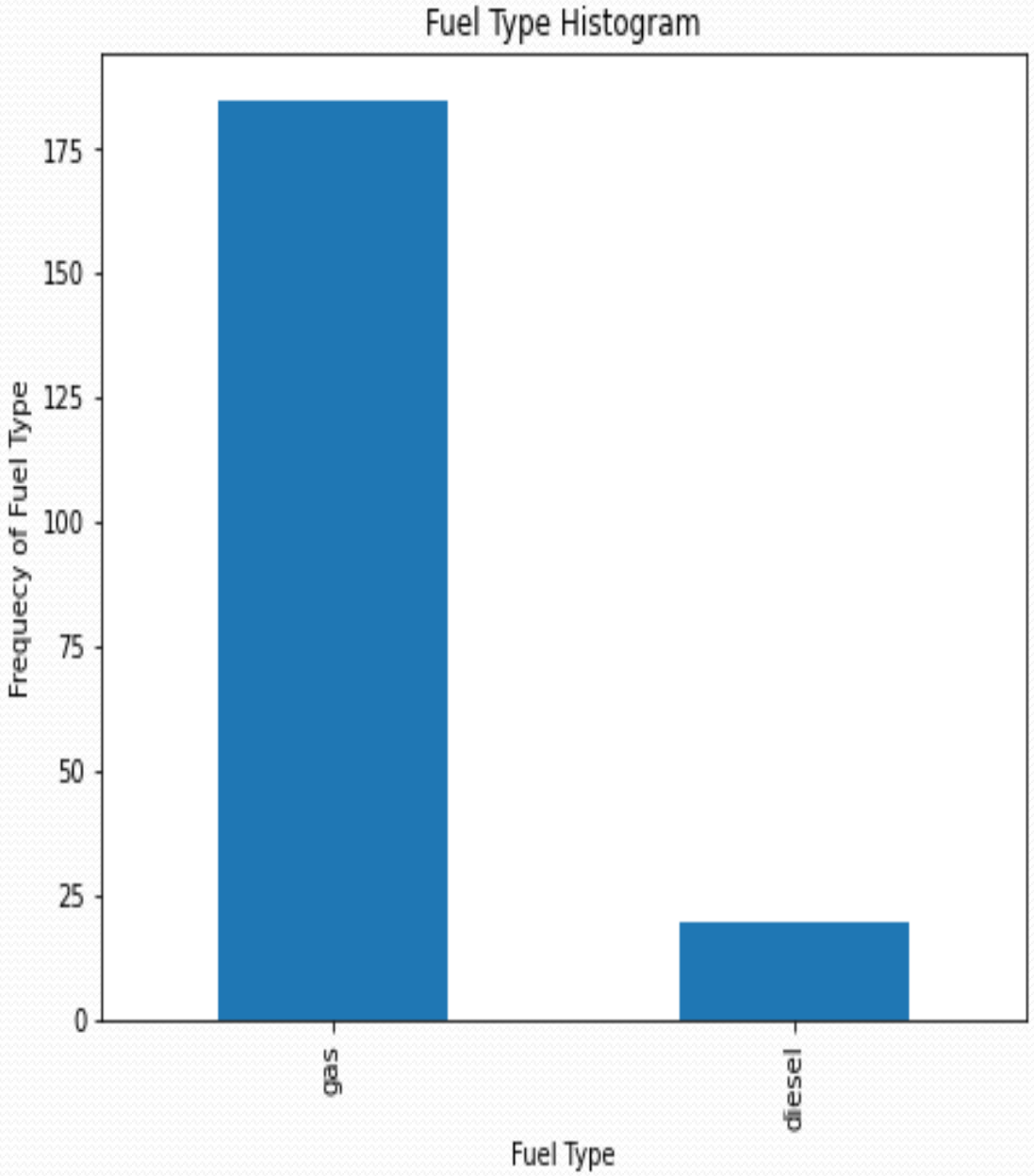
The price plot seemed to be right-skewed, meaning that the most prices in the dataset are low (Below 15,000).

There are a significant difference between the mean and the median of the price distribution. The data points are far spread out from the mean, which indicates a high variance in the car prices (85% of the prices are below 18,500, whereas the remaining 15% are betweening 18,500 and 45,400)



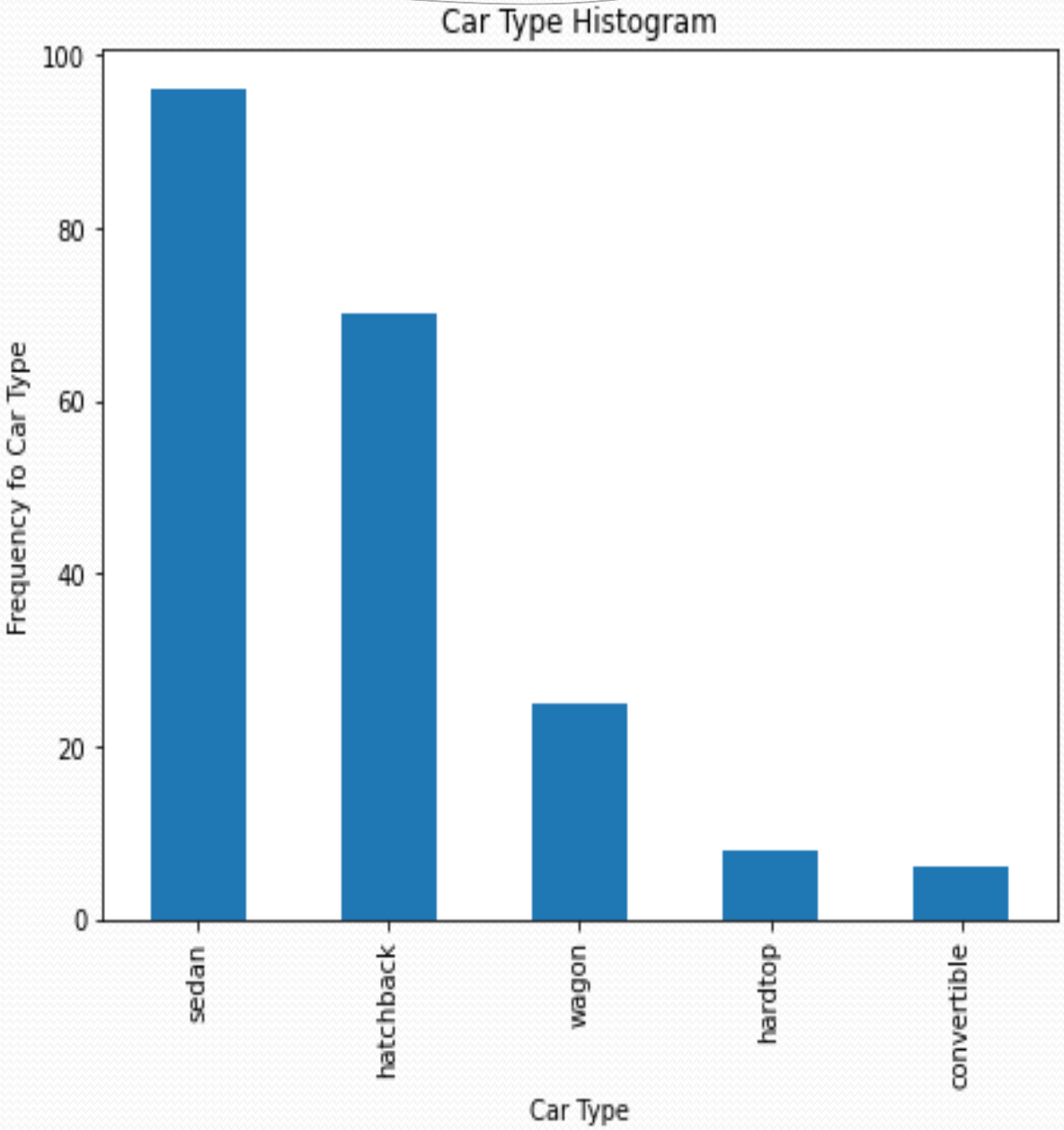
# FUEL TYPE

number of gas fuel cars are more than disesel.



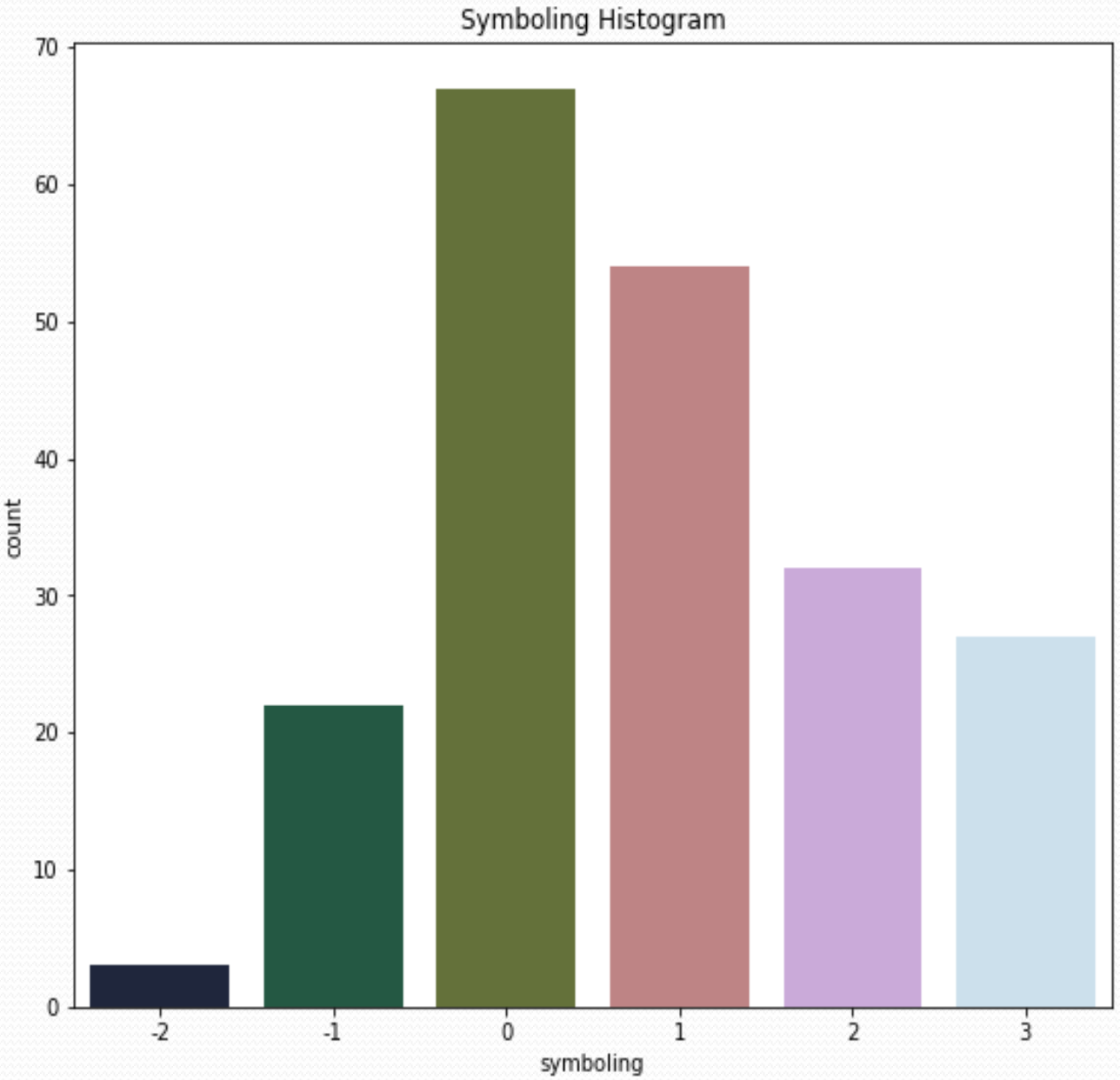
# CAR TYPE

Toyota seemed to be favoured car company and Sedan is the top cat type prefered



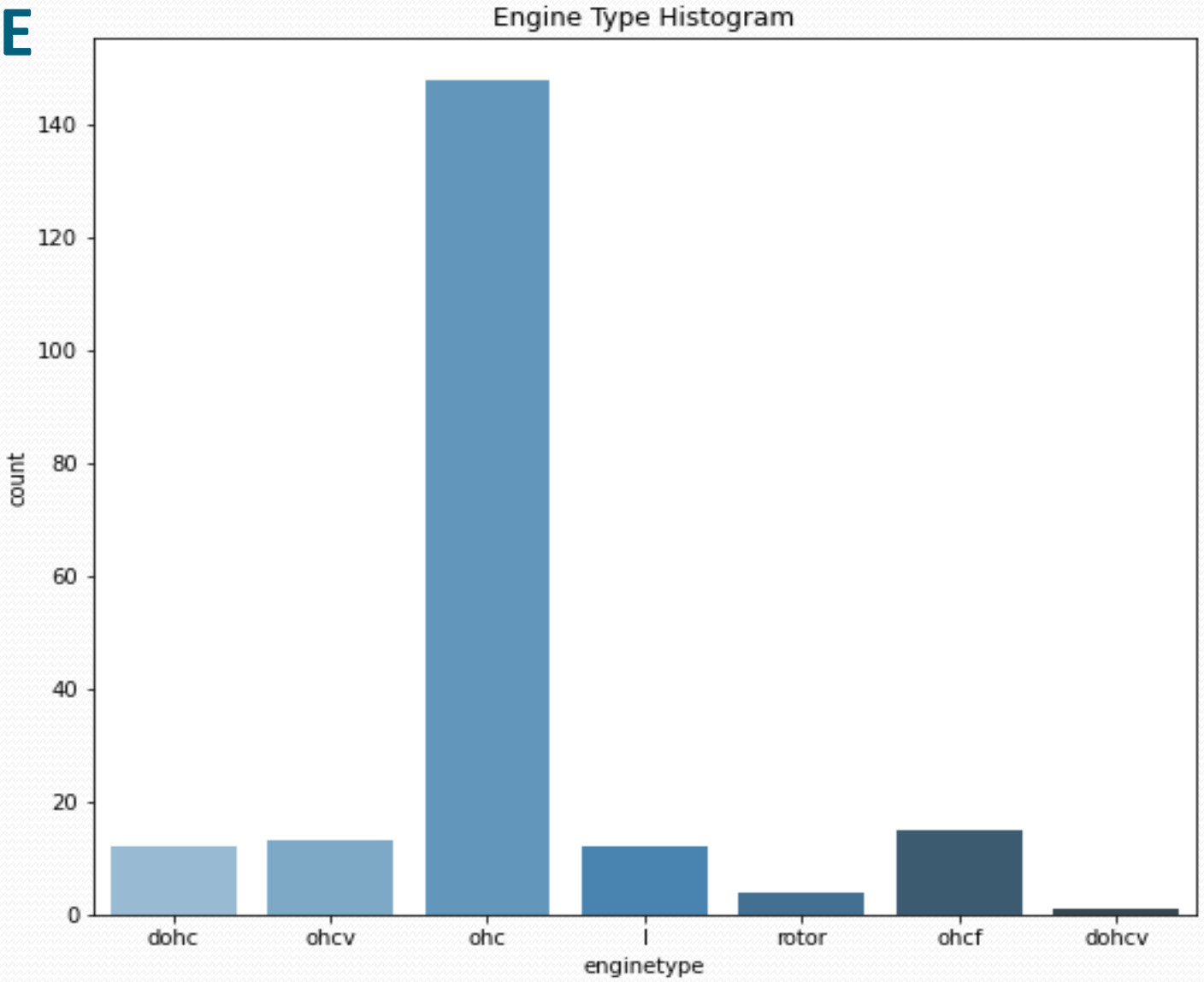
# SYMBOLING

It seems that the symboling with 0 and 1 values have high number of rows (i.e. They are most sold.)



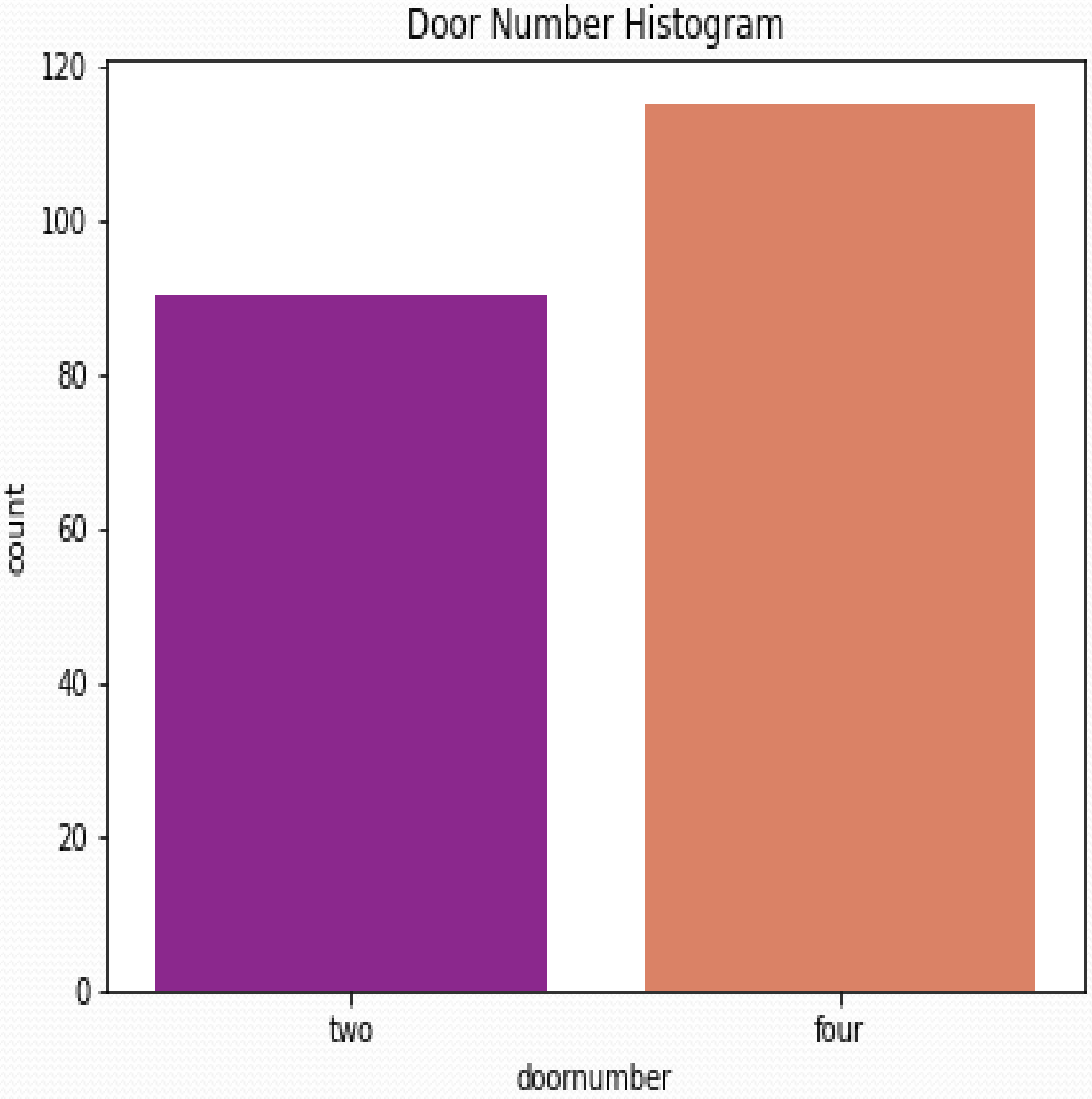
# ENGINE TYPE

Ohc Engine type seems to be most favored type.



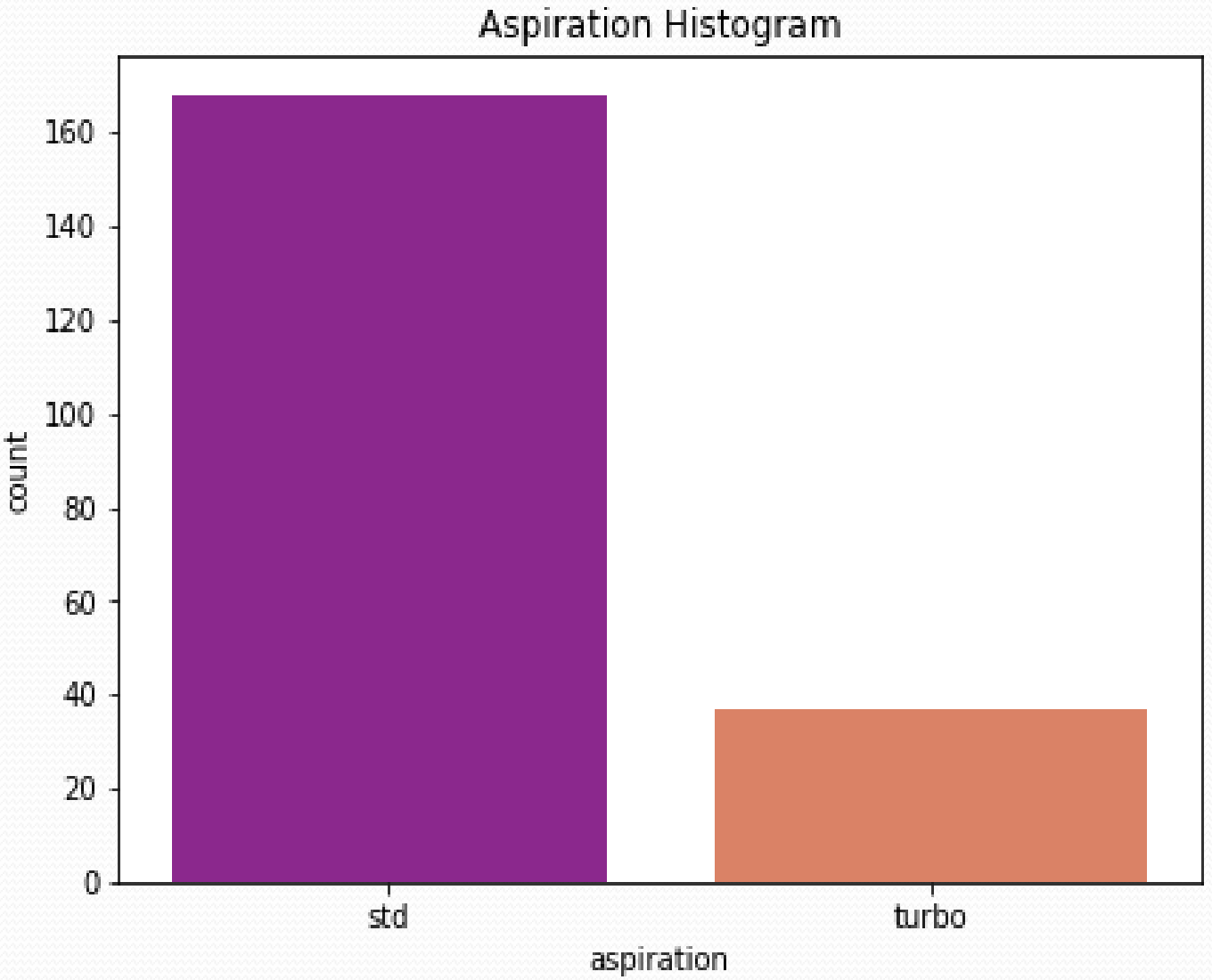
# DOOR NUMBER

cars with 4 door number are the highest



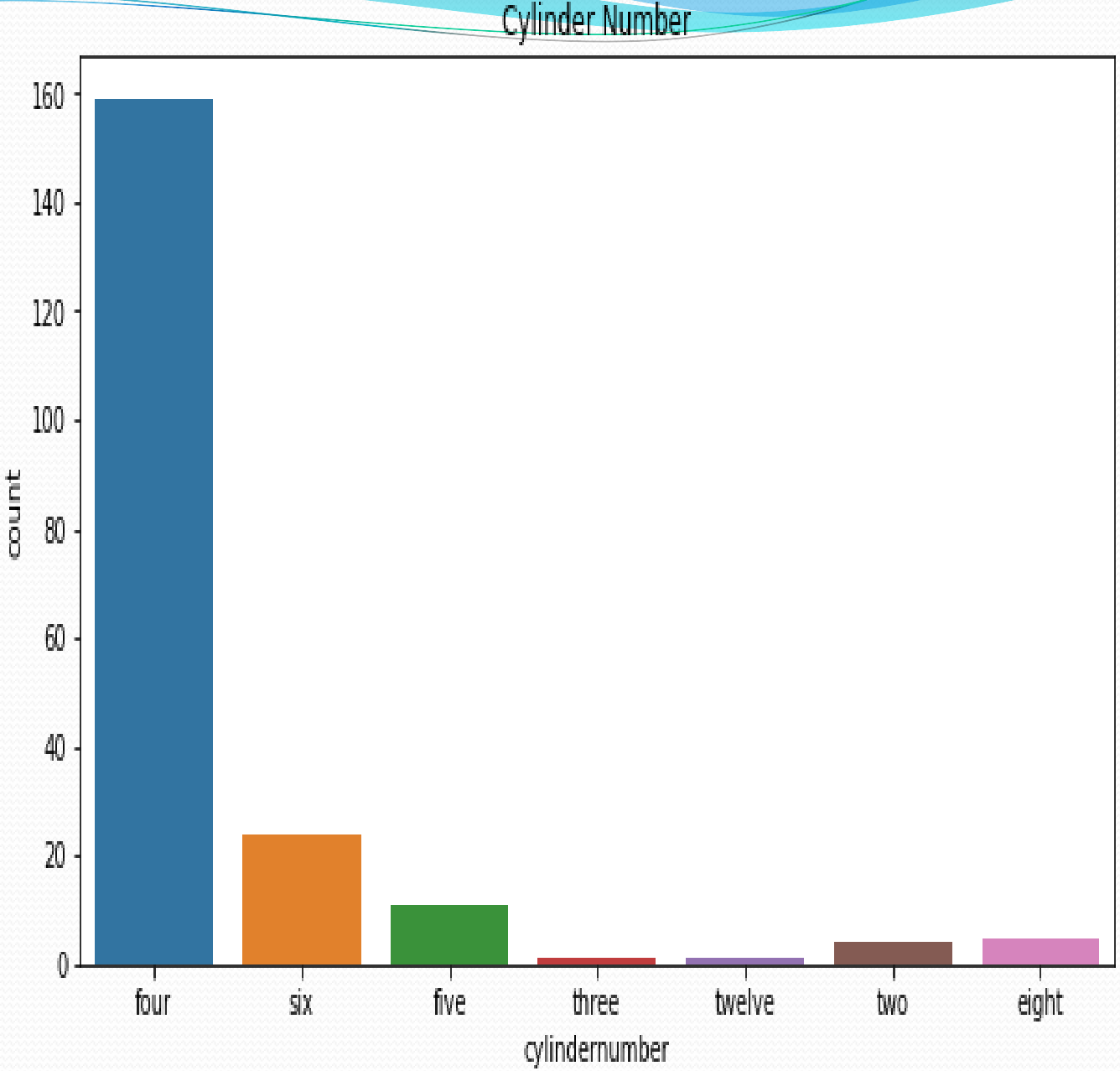
# ASPIRATION

std are the highest



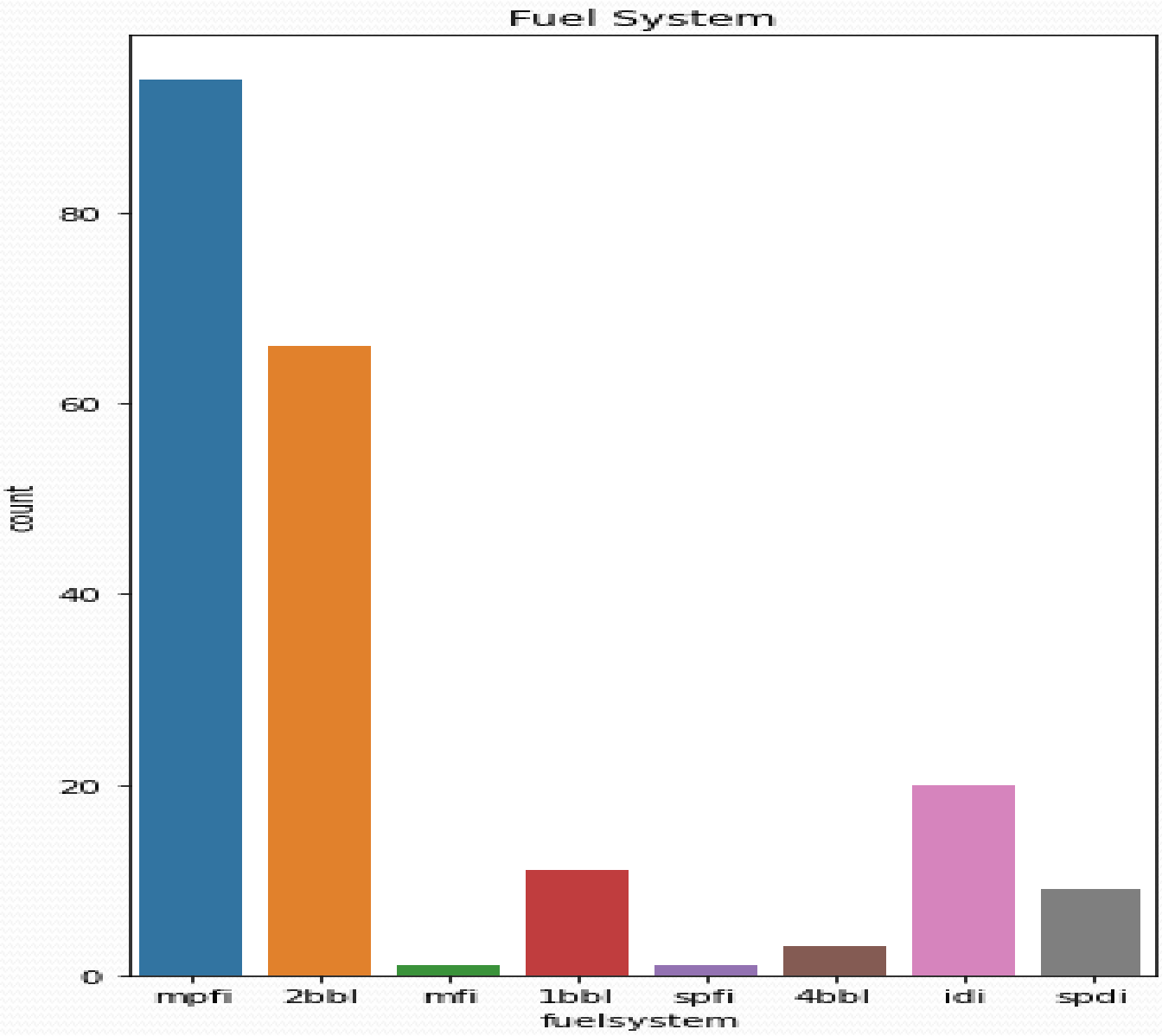
# CYLINDE NUMBER

Most common number of cylinders are four, six and five.



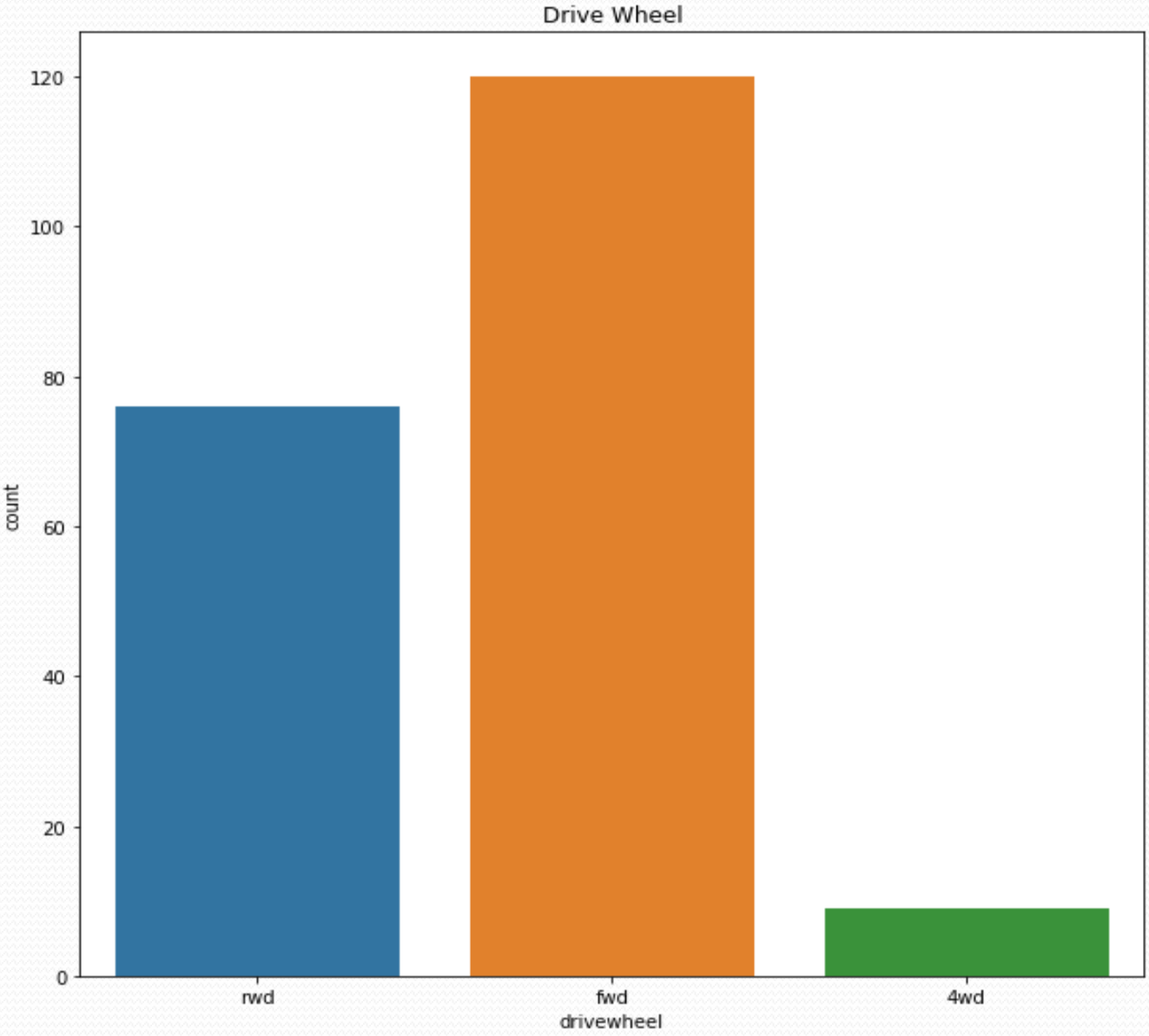
# FUEL SYSTEM

mpfi and 2bbl are most common type of fuel systems.



# DRIVE WHEEL

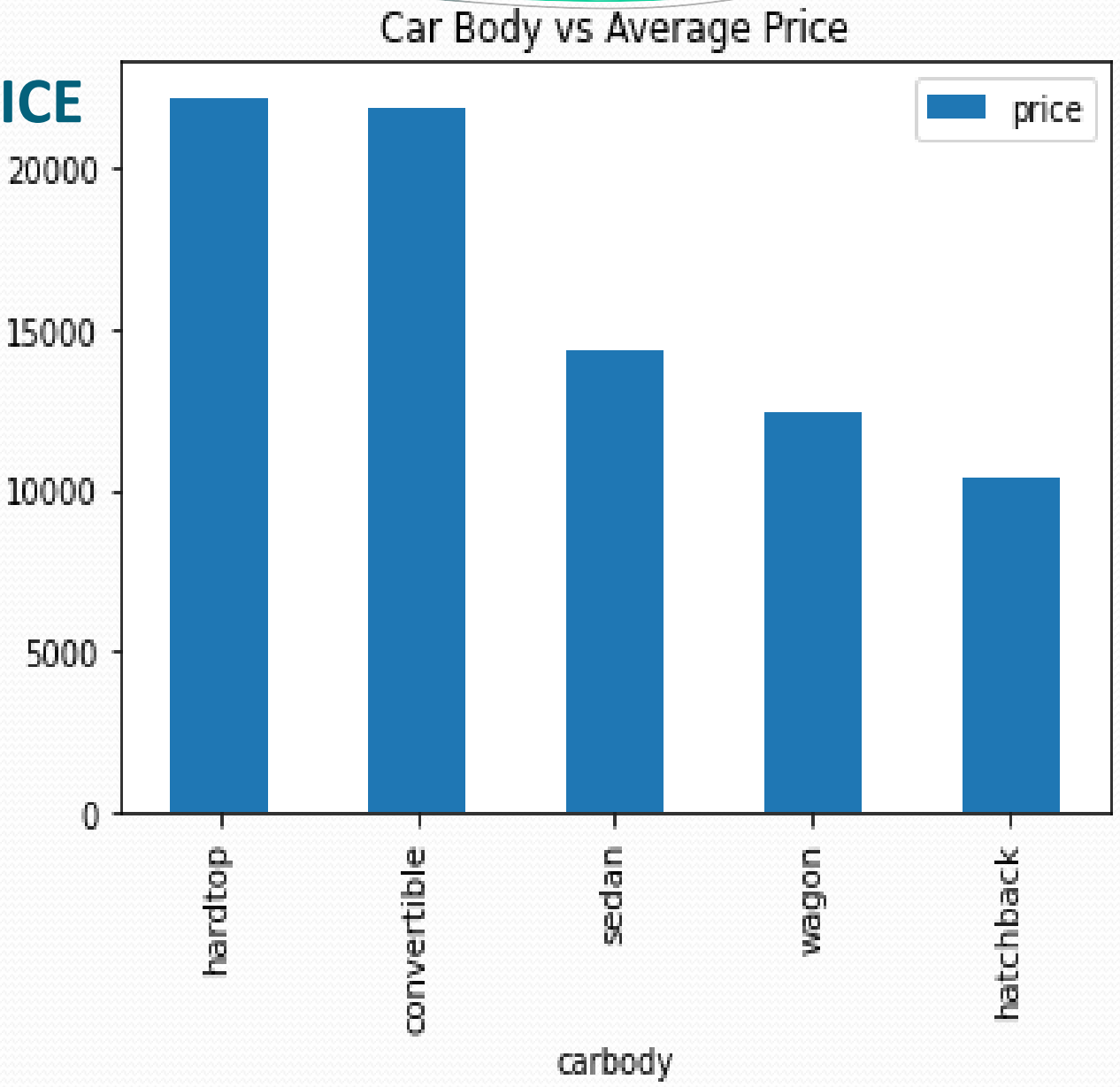
A very significant difference in drive wheel category. Most high ranged cars seems to prefer real wheel drive (rwd).



# BIVARIATE ANALYSIS

## CAR BODY AND PRICE

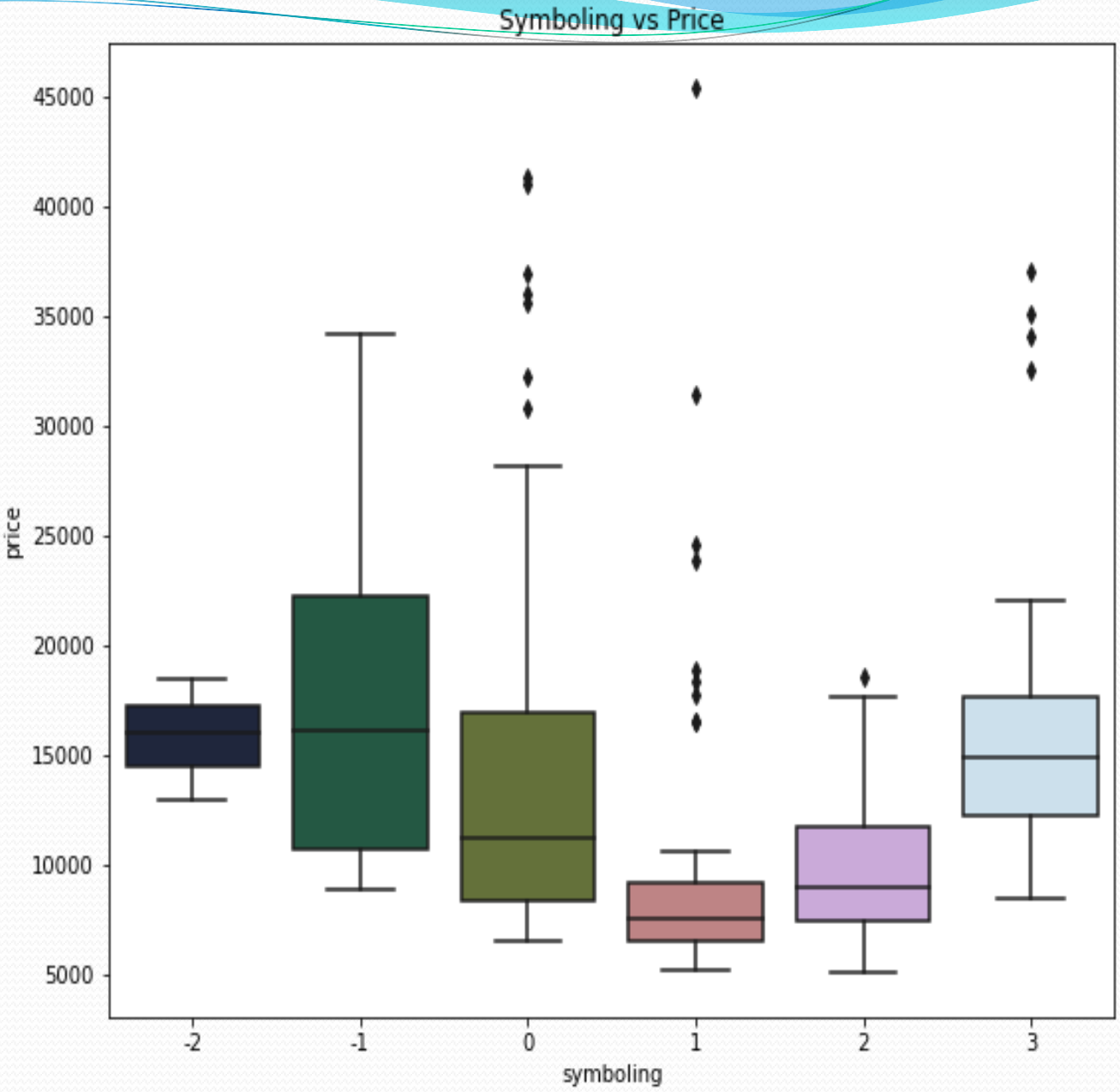
Hardtop and convertible have higher average price.



# SYMBOLING AND PRICE

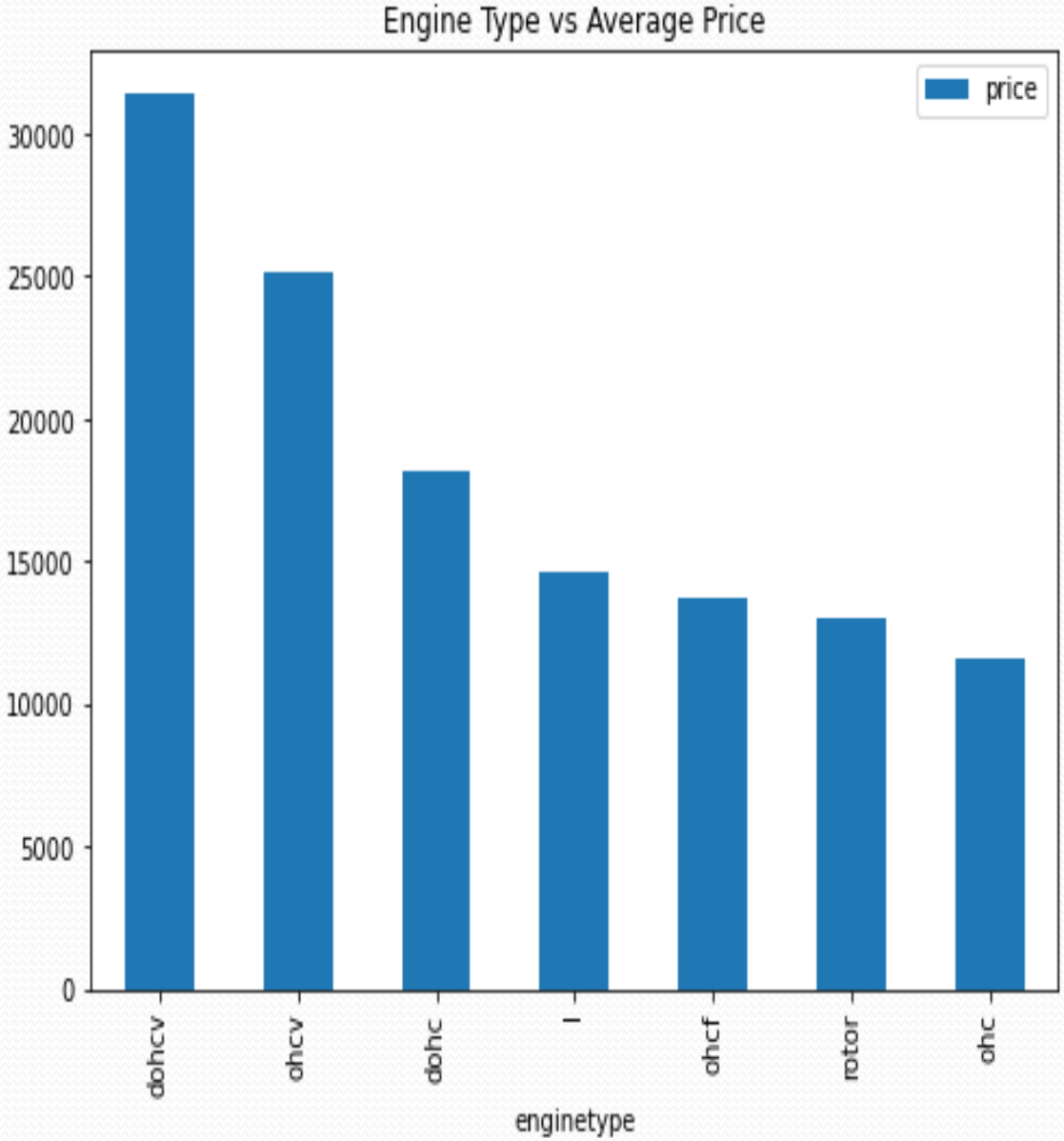
The cars with -1 symboling seems to be high priced (as it makes sense too, insurance risk rating -1 is quite good)

But it seems that symboling with 3 value has the price range similar to -2 value. There is a dip in price at symboling 1.



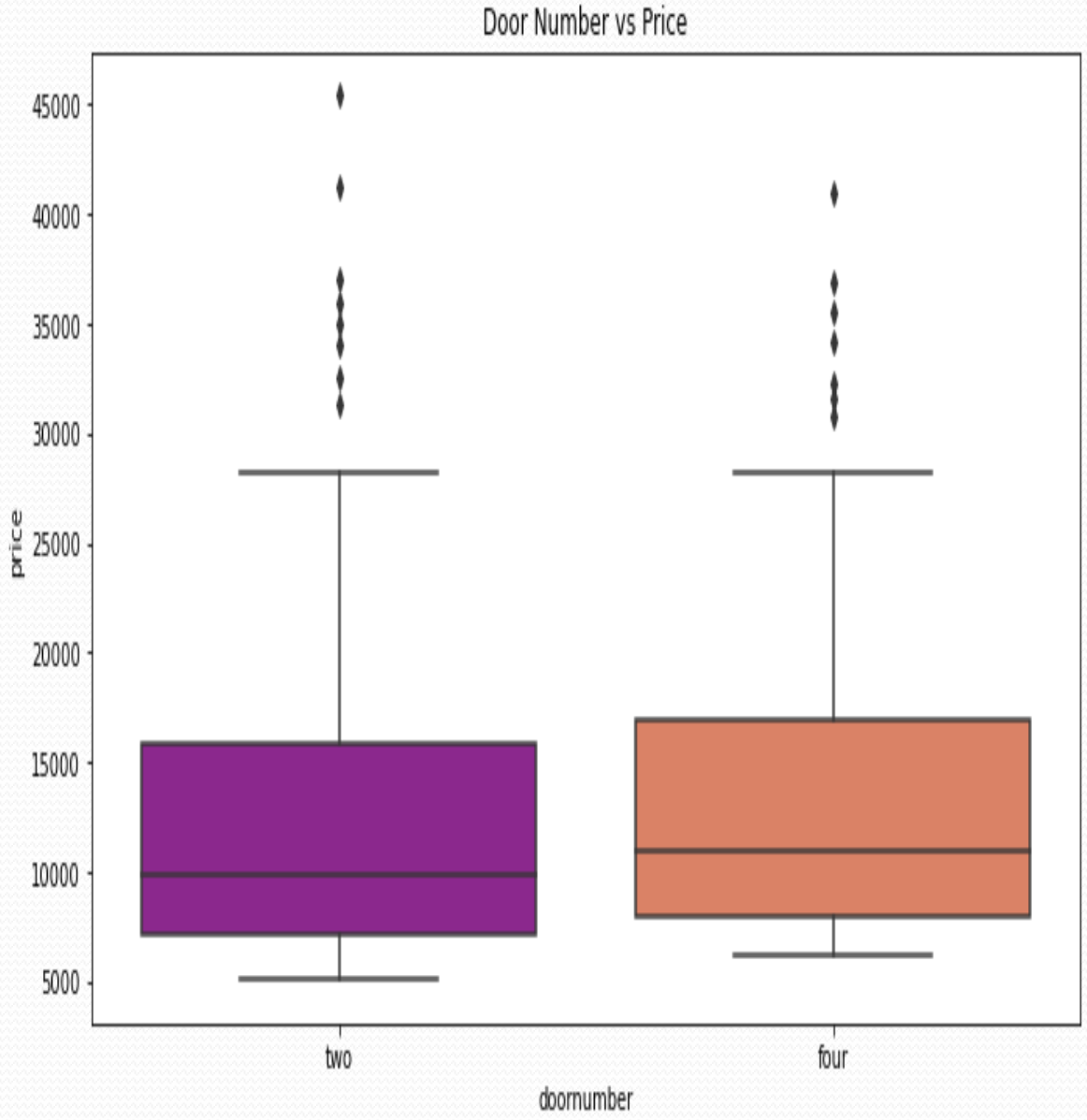
# ENGINE TYPE AND PRICE

ohcv has the highest price range (While dohcv has only one row), ohc and ohcf have the low price range.



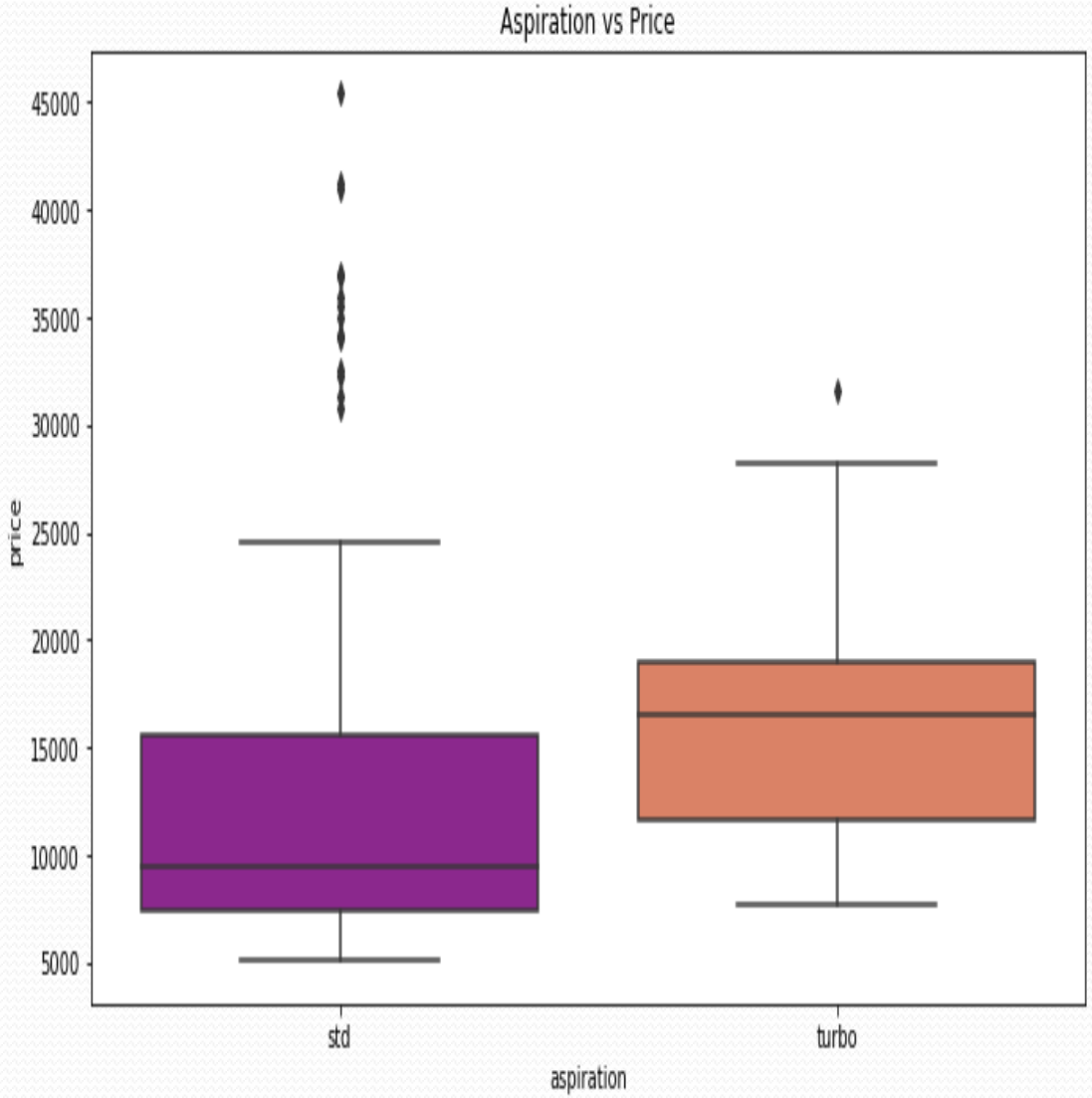
## DOOR NUMBER AND PRICE

Doornumber variable is not affacting the price much. There is no sugnificant difference between the categories in it.



# ASPIRATION

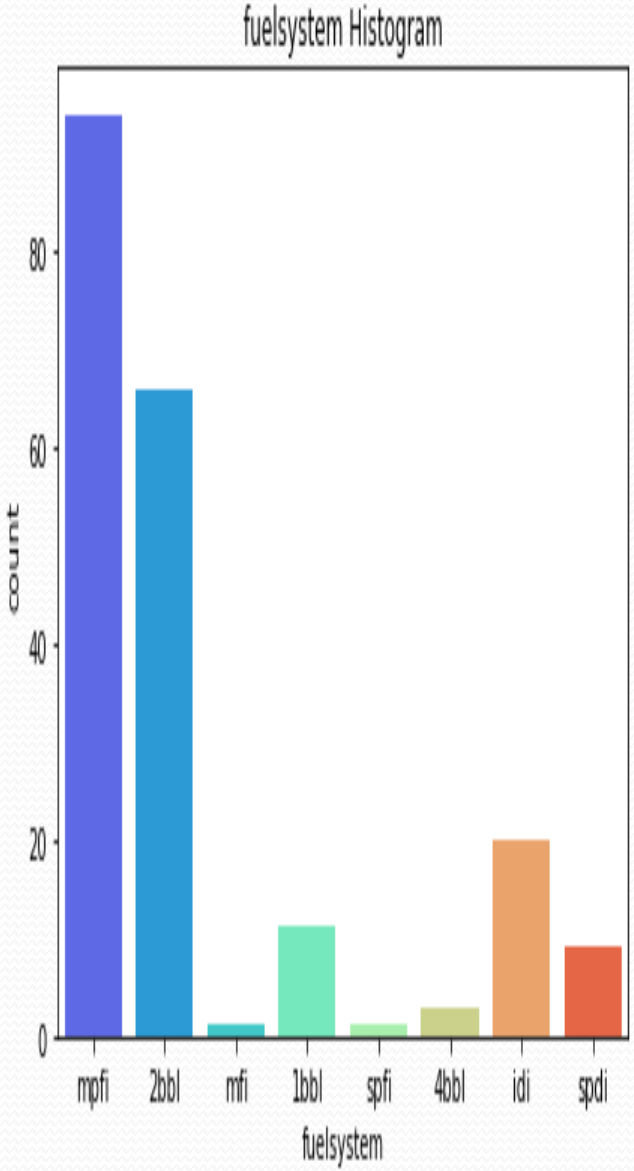
It seems aspiration with turbo have higher price range than the std(though it has some high values outside the whiskers.)

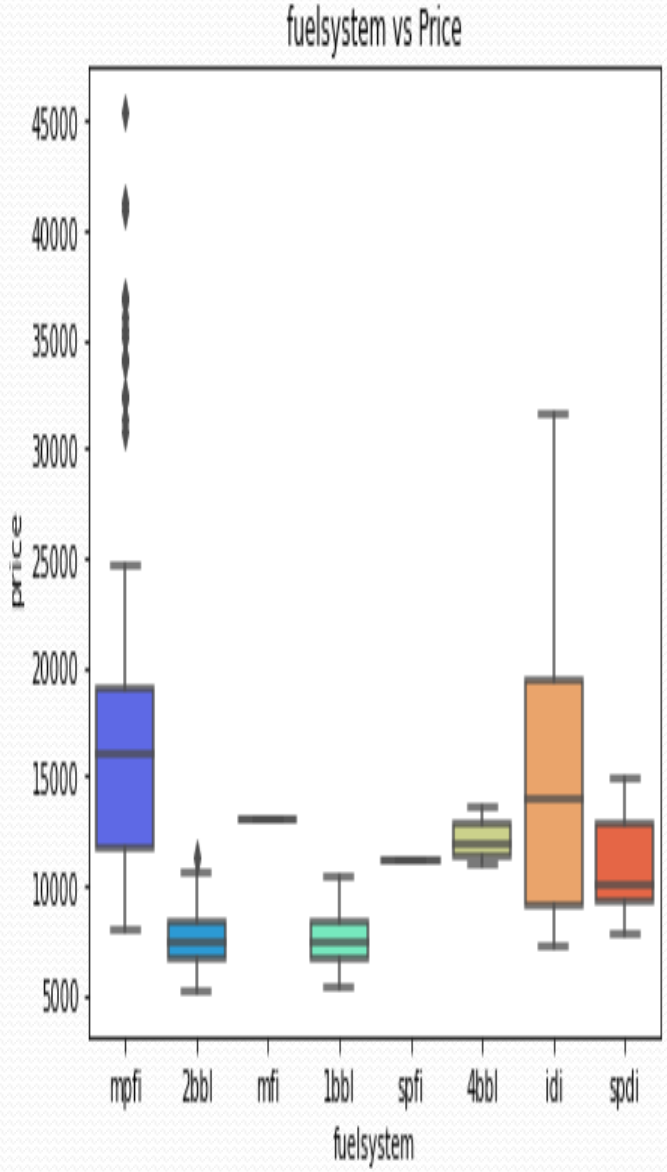


## FUEL SYSTEM AND PRICE

mpfi and idi having the highest price range.

But there are few data for other categories to derive any meaningful inference

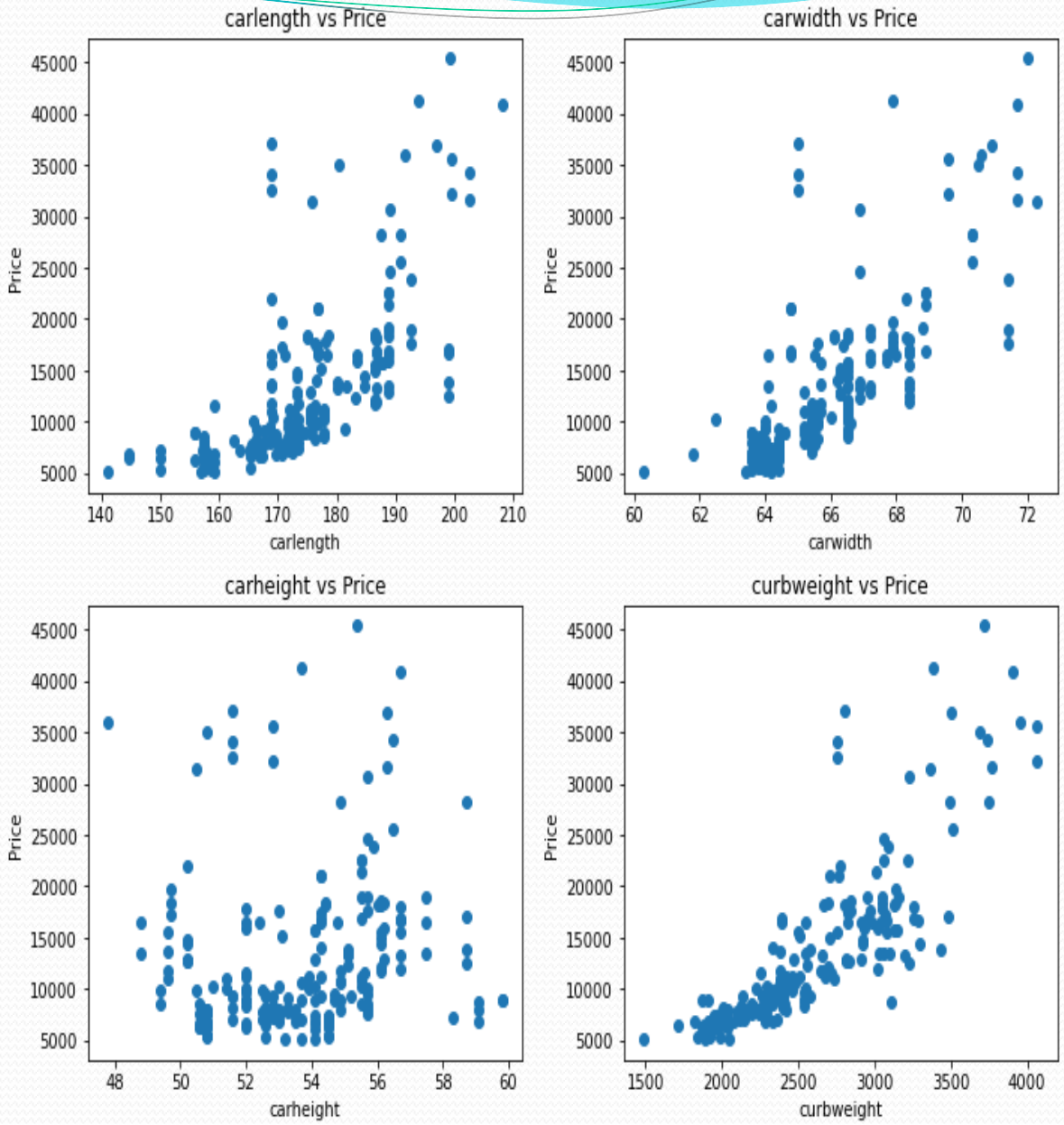


## CAR LENGTH, CAR WIDTH, CAR HEIGHT, CURB WEIGHT VS PRICE

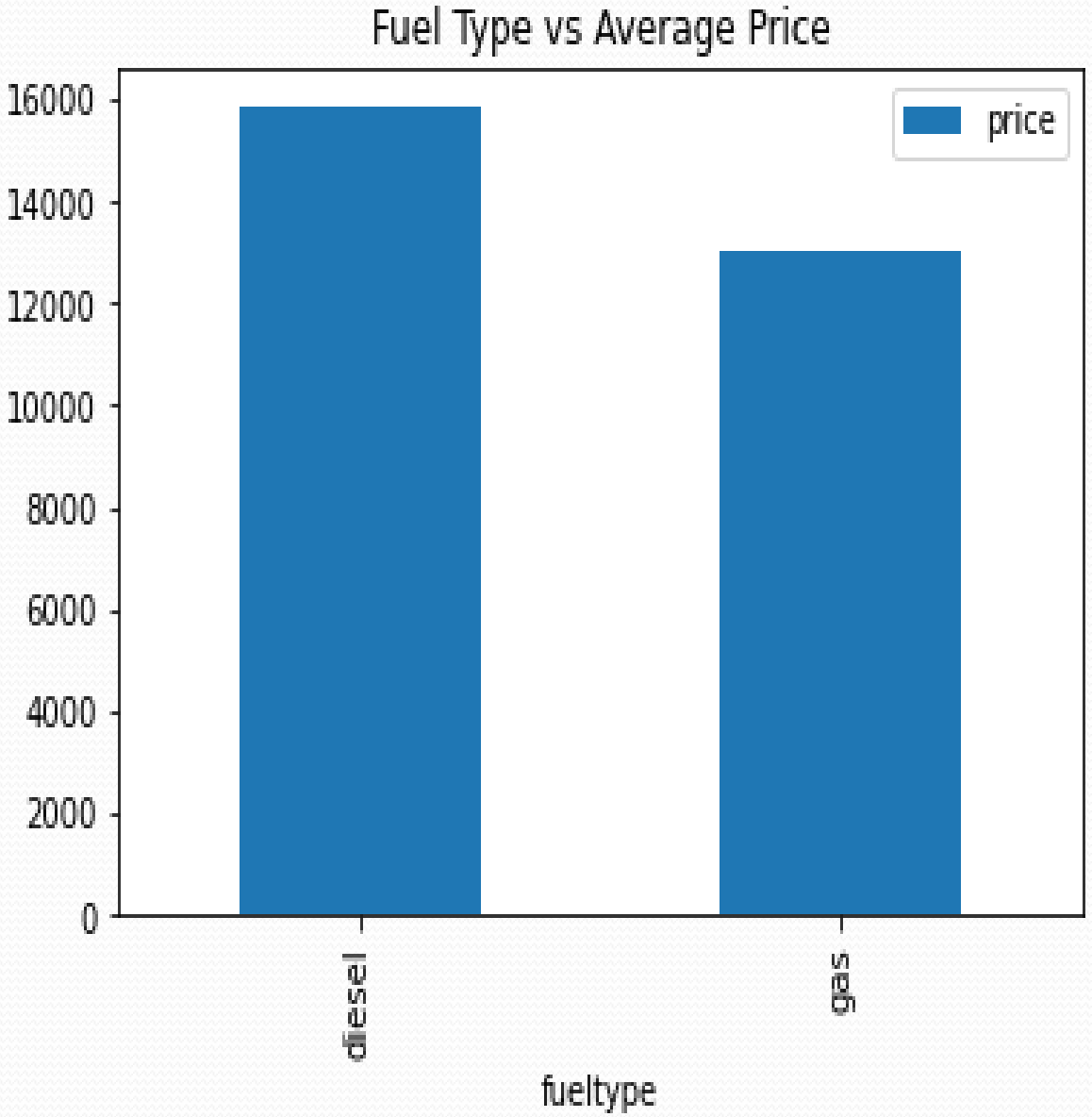
carwidth, carlength and curbweight seems to have a poitive correlation with price.

carheight doesn't show any significant trend with price.



## FUEL TYPE AND PRICE

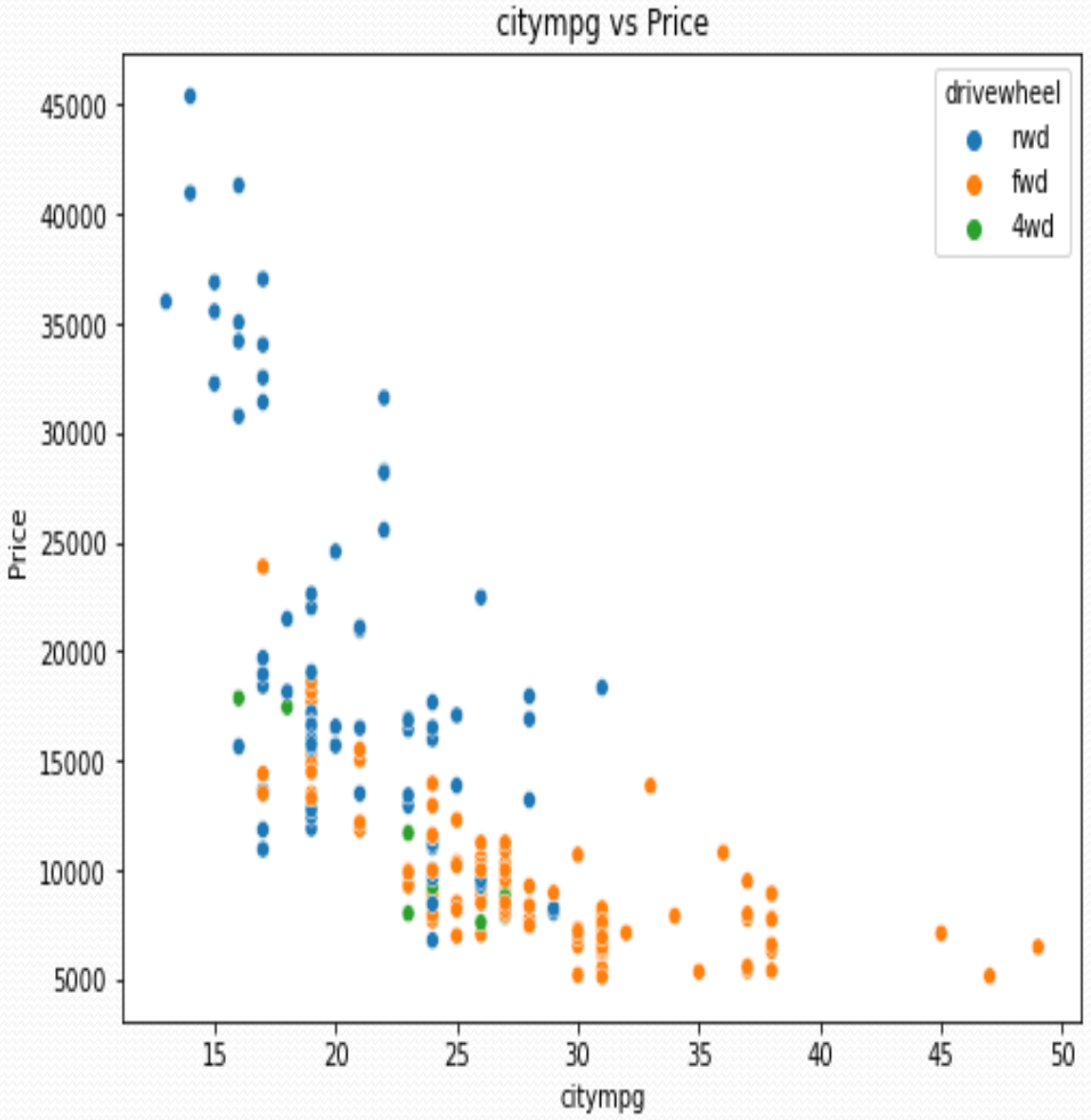
Diesel has higher average price than gas.



# MULTIVARIATE ANALYSIS

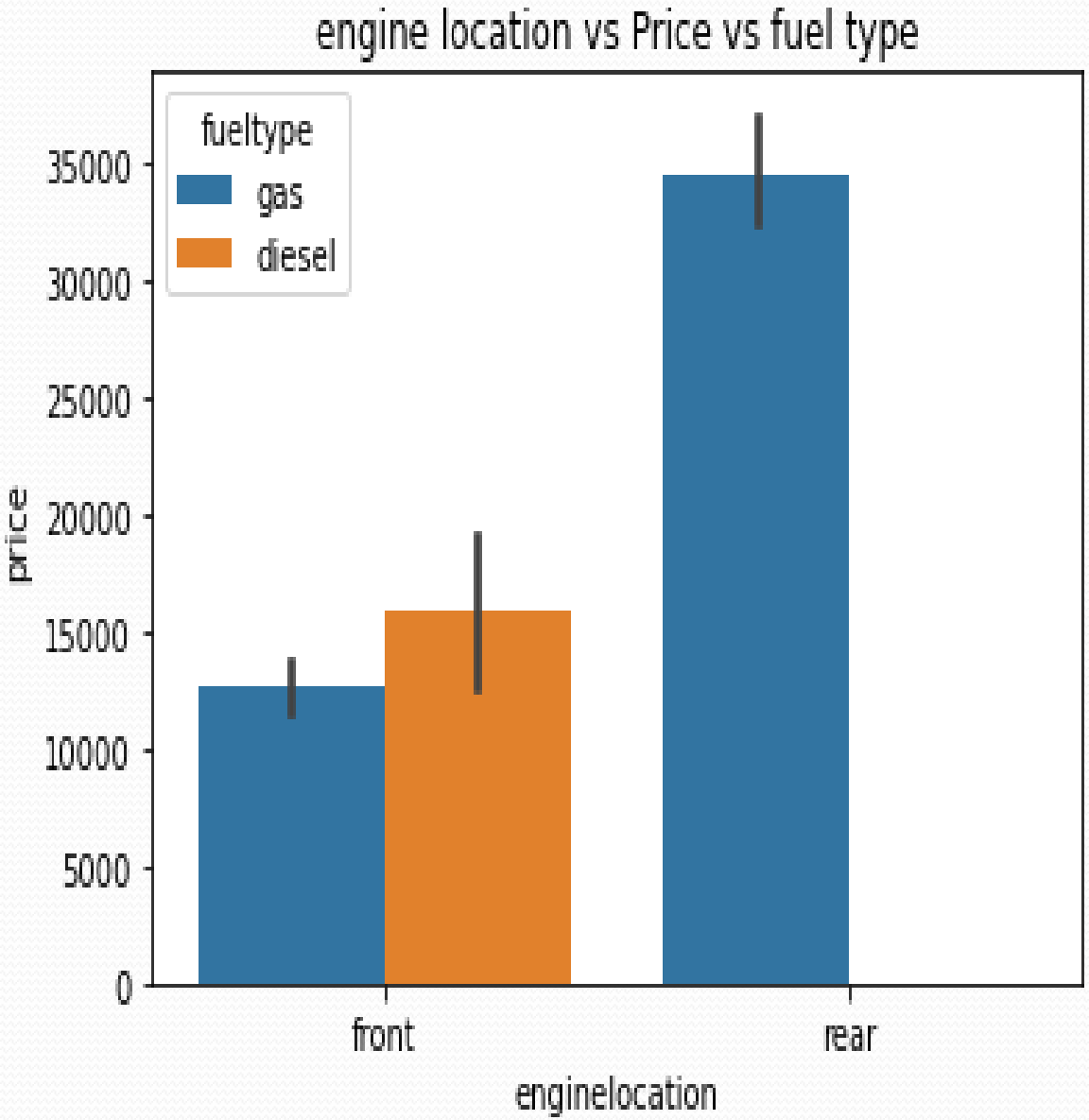
## CITY MPG VS DRIVE WHEEL VS PRICE

City mpg has an obvious negative correlation with price and is significant.



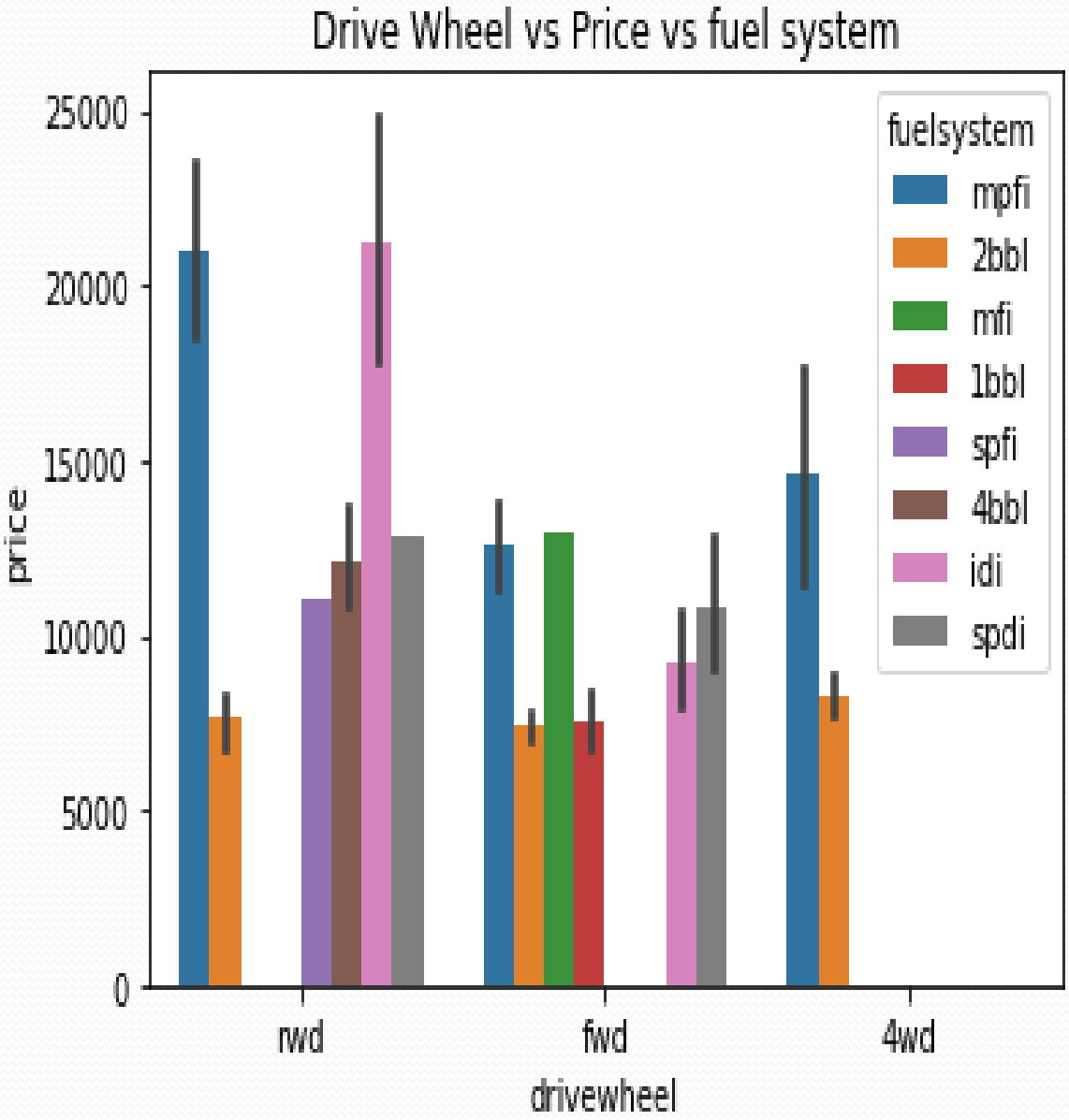
## ENGINE LOCATION VS PRICE VS FUEL TYPE

Rear engine location cars prefer rwd drive wheel with idi or mpfi fuel system



## DRIVE WHEEL VS PRICE VS FUEL SYSTEM

Rear engine location cars prefer rwd drive wheel with idi or mpfi fuel system

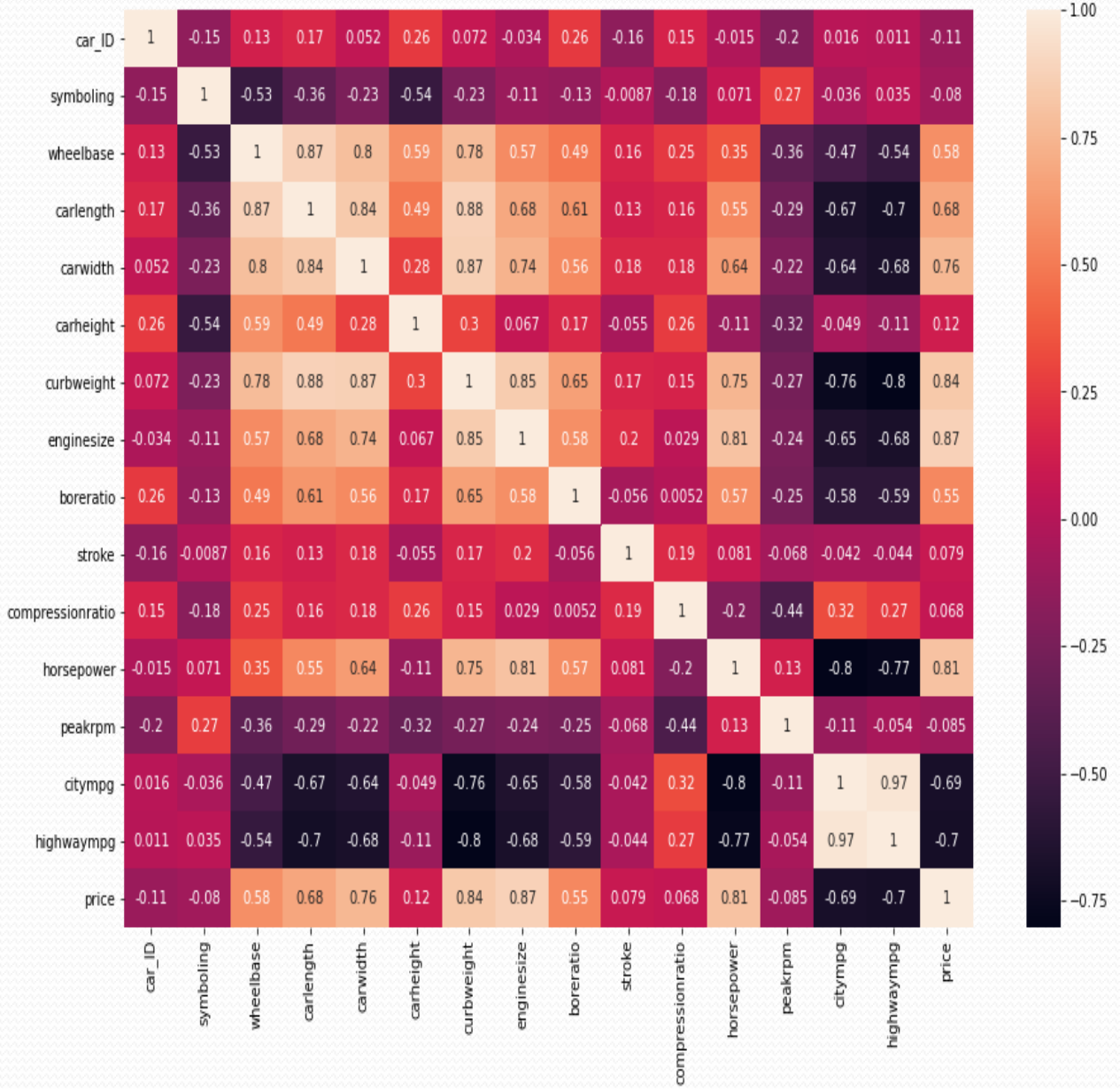


# HEATMAP CORRELATION PLOT

Engine size, bore ratio, horsepower, wheelbase seems to have a significant positive correlation with price.

City mpg, highway mpg seems to have a significant negative correlation with price.

There is a significant correlation level between all variables



# List of significant variables after Visual analysis:

These features are important models in predicting the price of car

- Engine Type
- Fuel type
- Car Body
- Aspiration
- Cylinder Number
- Drive wheel
- Curb weight
- Car Length
- Car width
- Engine Size
- Bore ratio
- Horse Power
- Wheel base

# MACHINE LEARNING

Regression models employed in the prediction are:

- Linear Regression
- Elastic Net

# Model performance

- Linear regression model was employed for the prediction of the price. The model gave a coefficient of determination score of 0.56 which is low, so we explored other regression models like the elastic Net.

- Elastic Net gave a better R- score (0.82) than the Linear Regression which

# THANK YOU