GEELY AUTOS PRICE PREDICTION MODEL

Cheshi Emmanuel

PROBLEM STATEMENT

A Chinese automobile company Geely Auto aspires to enter the Nigerian market by setting up its manufacturing unit and producing cars locally to compete with their Nigerian, US and European counterparts.

They have contacted you, a Data Scientist to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the Nigerian market, since those may be very different from the Chinese market.

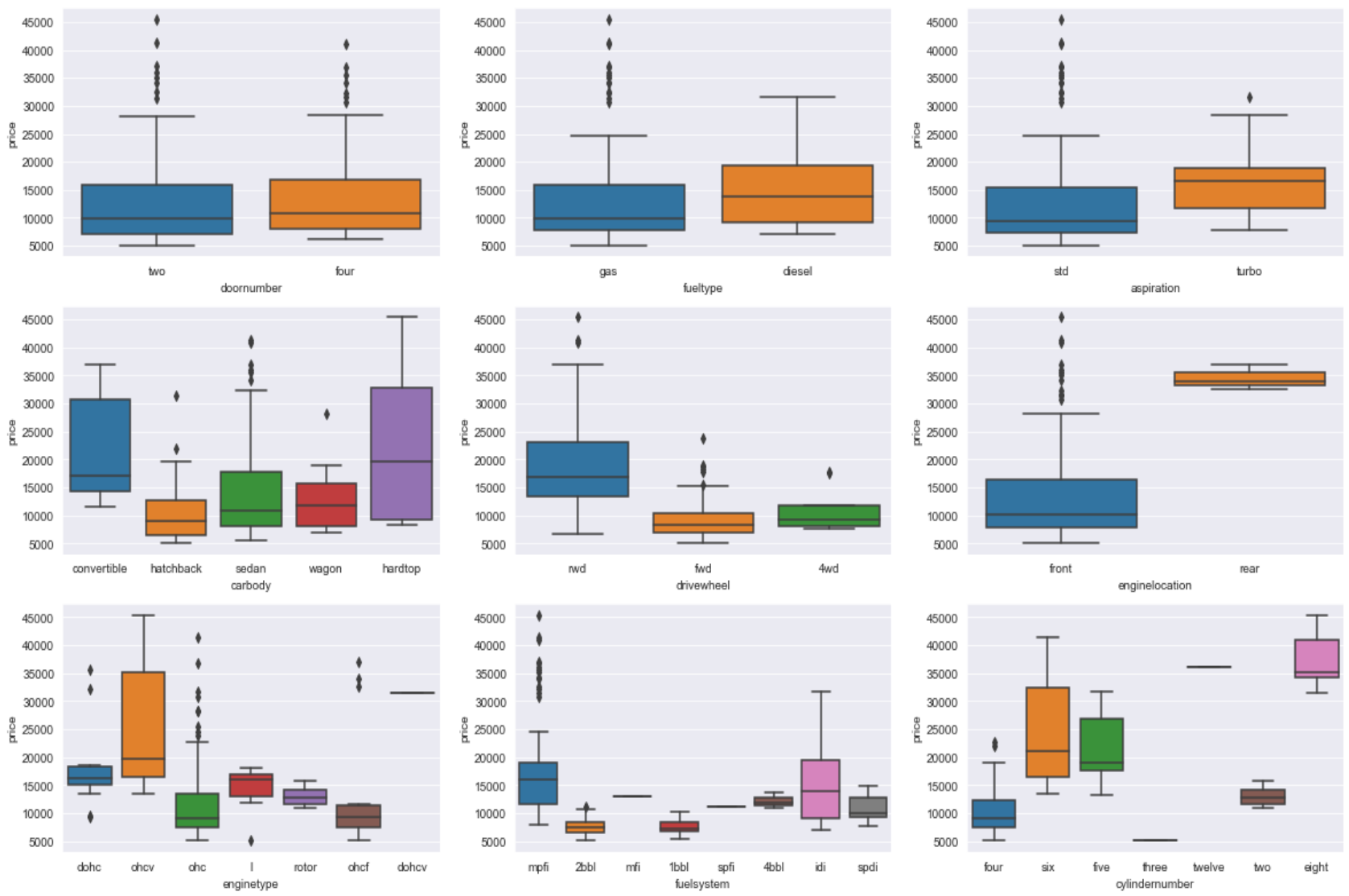
The company wants to know:

- Which variables are significant in predicting the price of a car
- How well do those variables describe the price of a car
- Based on various market surveys, you have gathered a large data set of different types of cars across the Nigerian market.

BUSINESS GOAL/OBJECTIVE

To model the price of cars with the available independent variables for management to use it to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

BOXPLOT INDICATING FEATURES OF CATEGORICAL VARIABLES



- Fuel type have an effect on the pricing of the cars.
- Engine location and aspiration surely has a visible effect on the pricing of the car
- The price of real wheel drive is significantly higher than other drive wheel options.
- Cylinder number and engine type also seem to regulate the price of cars.
- Hardtop and convertible cars are definitely priced higher than other body types available. Also car body is contributing in determining the price.

WHEELBASE CORRELATION WITH PRICE

From the visualization it can be seen that Wheelbase has a strong positive correlation (0.57)with price. This Indicates that the longer the wheelbase, the higher the price because longer wheel bases tend to have a better ride quality than those with shorter ones.

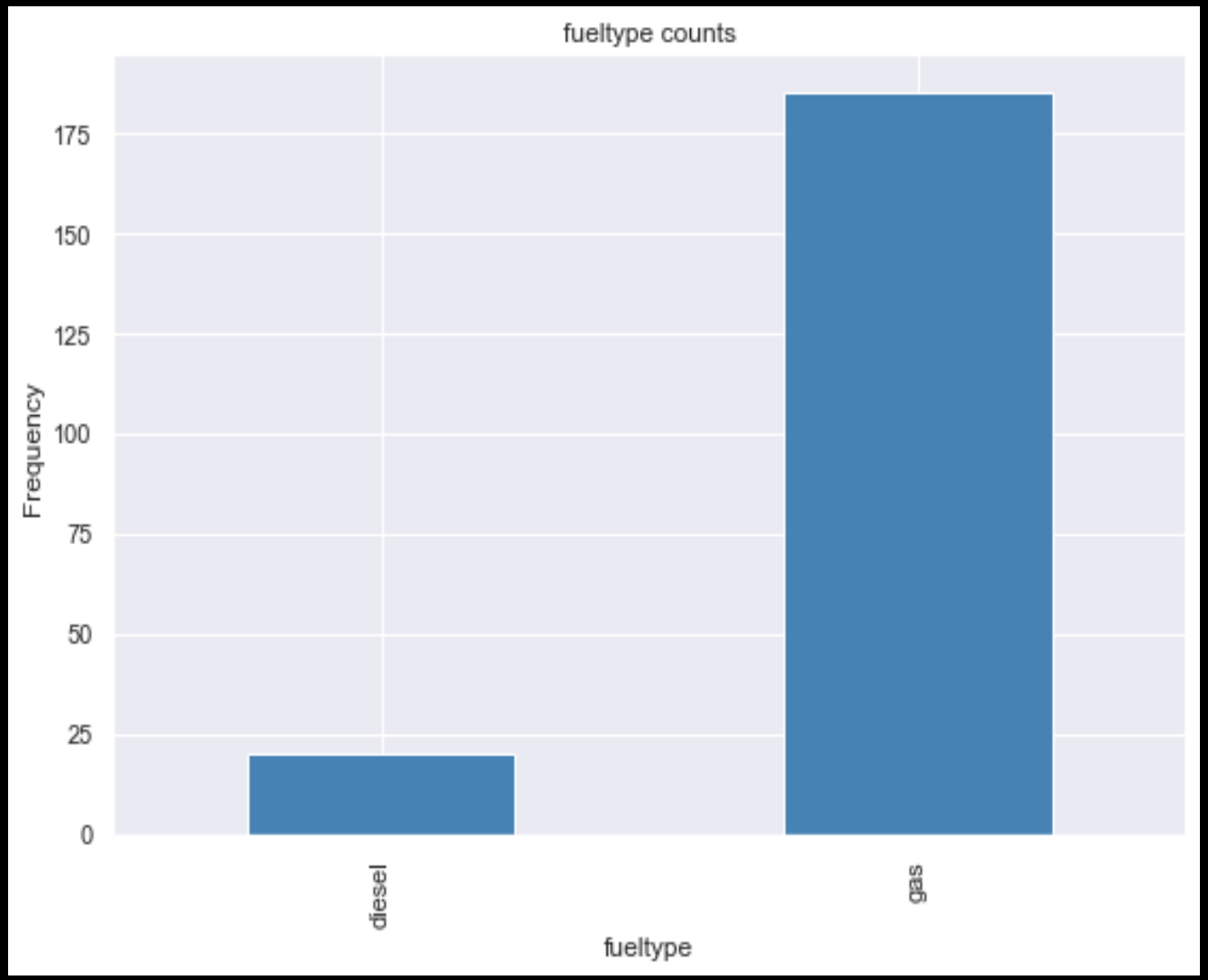
CAR LENGTH CORRELATION WITH PRICE

From the visualization it can be seen that car length has a strong positive correlation (0.68)with price.

ENGINE SIZE CORRELATION WITH PRICE

From the visualization it can be seen that engine size has a strong positive correlation (0.8) with price. This Indicates that the bigger the engine in size, the higher the price because they possess higher turbo charge and have around 150hp making them more expensive than cars with smaller ones.

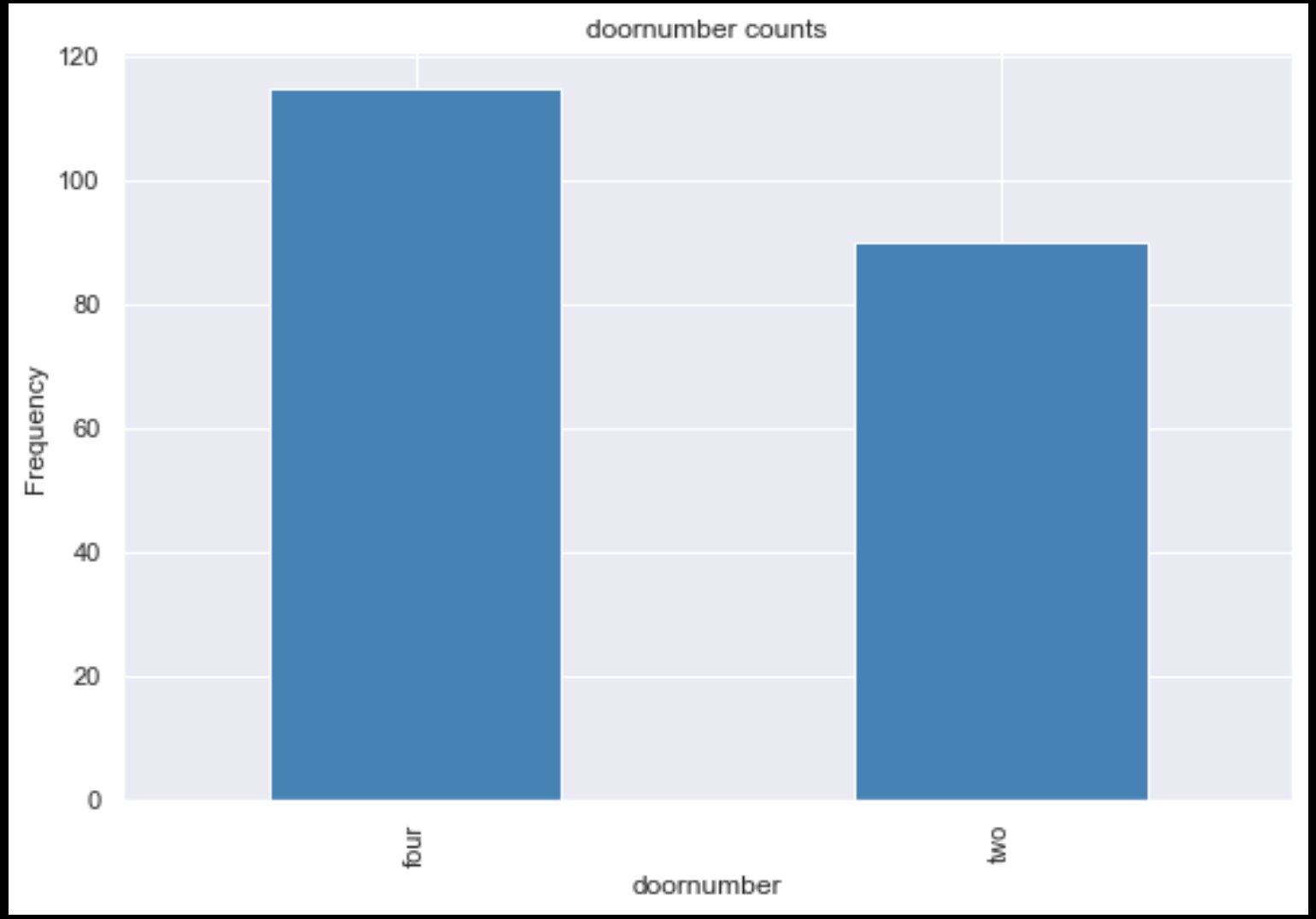
FUEL TYPE PREFERENCE



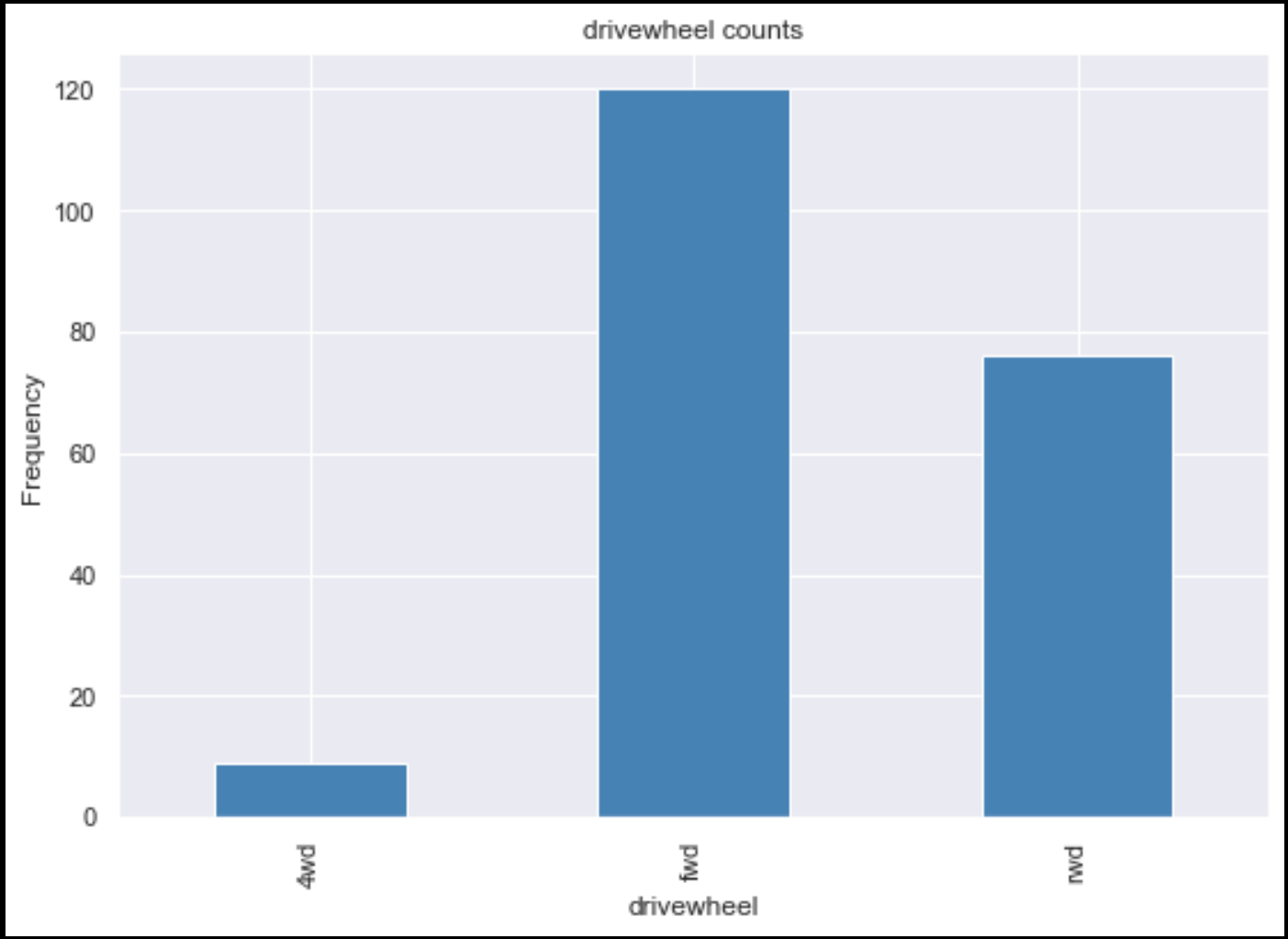
From the visualization, it can be seen that majority of the cars preferred are the ones that run on gasoline instead of diesel

CARS CHOICE BASED ON NUMBER OF DOORS

The visualization indicates that majority of the cars are those with 4 doors against the ones with 2 doors.



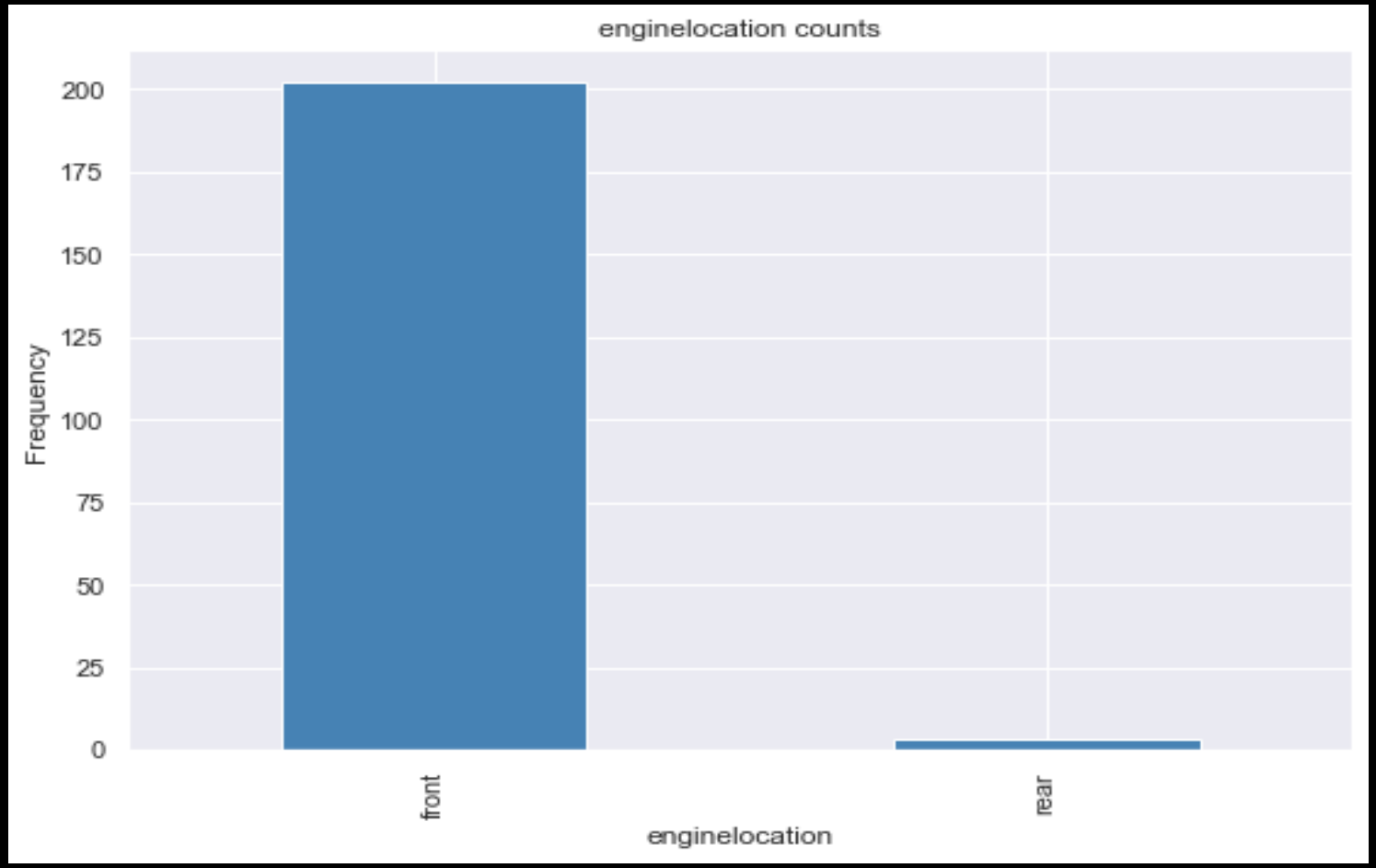
CAR CHOICES BASED ON WHEEL DRIVE



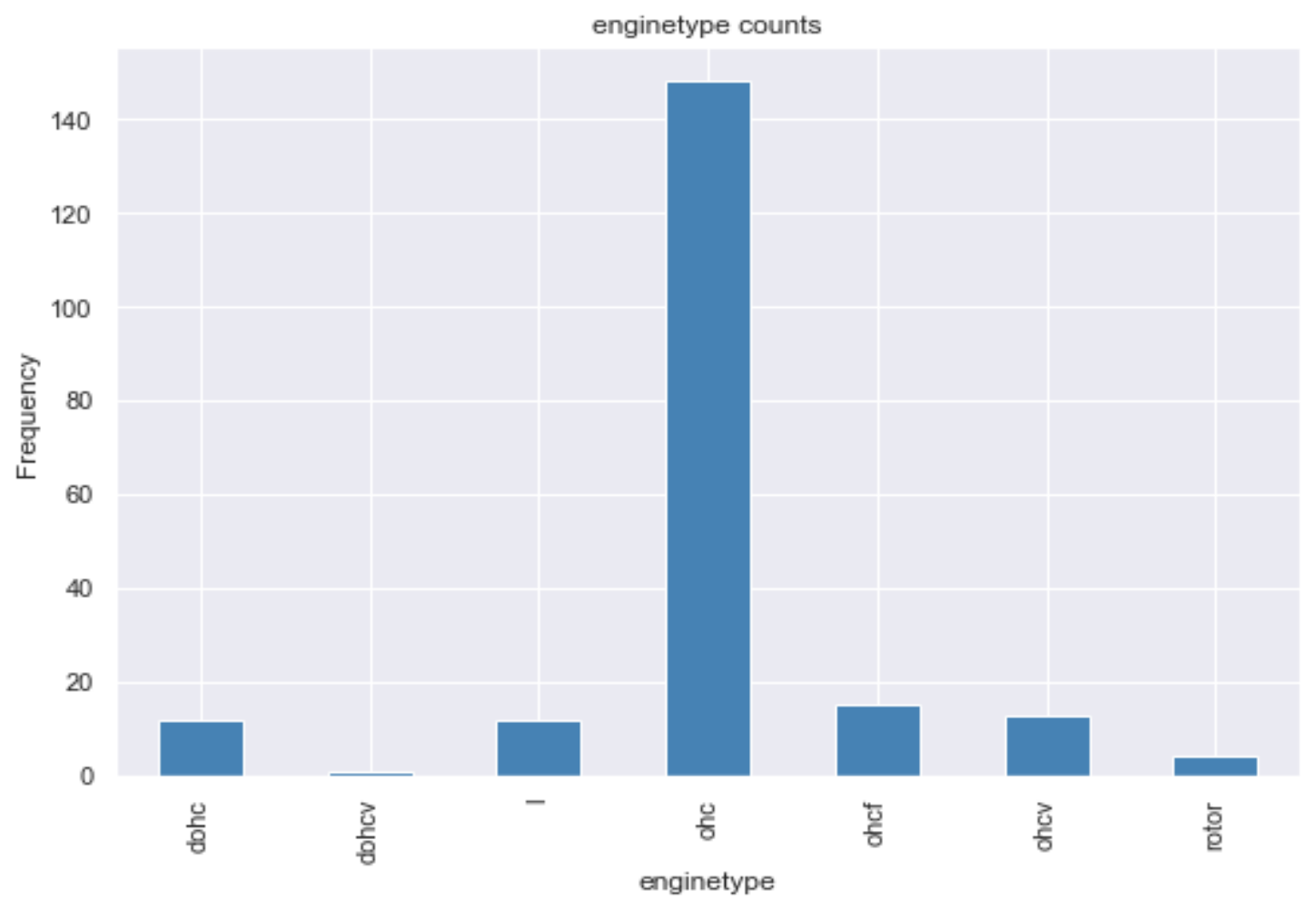
From the visualization the most preferred cars are the ones with the forward wheel drive (FWD) feature.

CARS CHOICES BASED ON ENGINE LOCATION

The Visualization indicates that majority of the cars have their engines in front



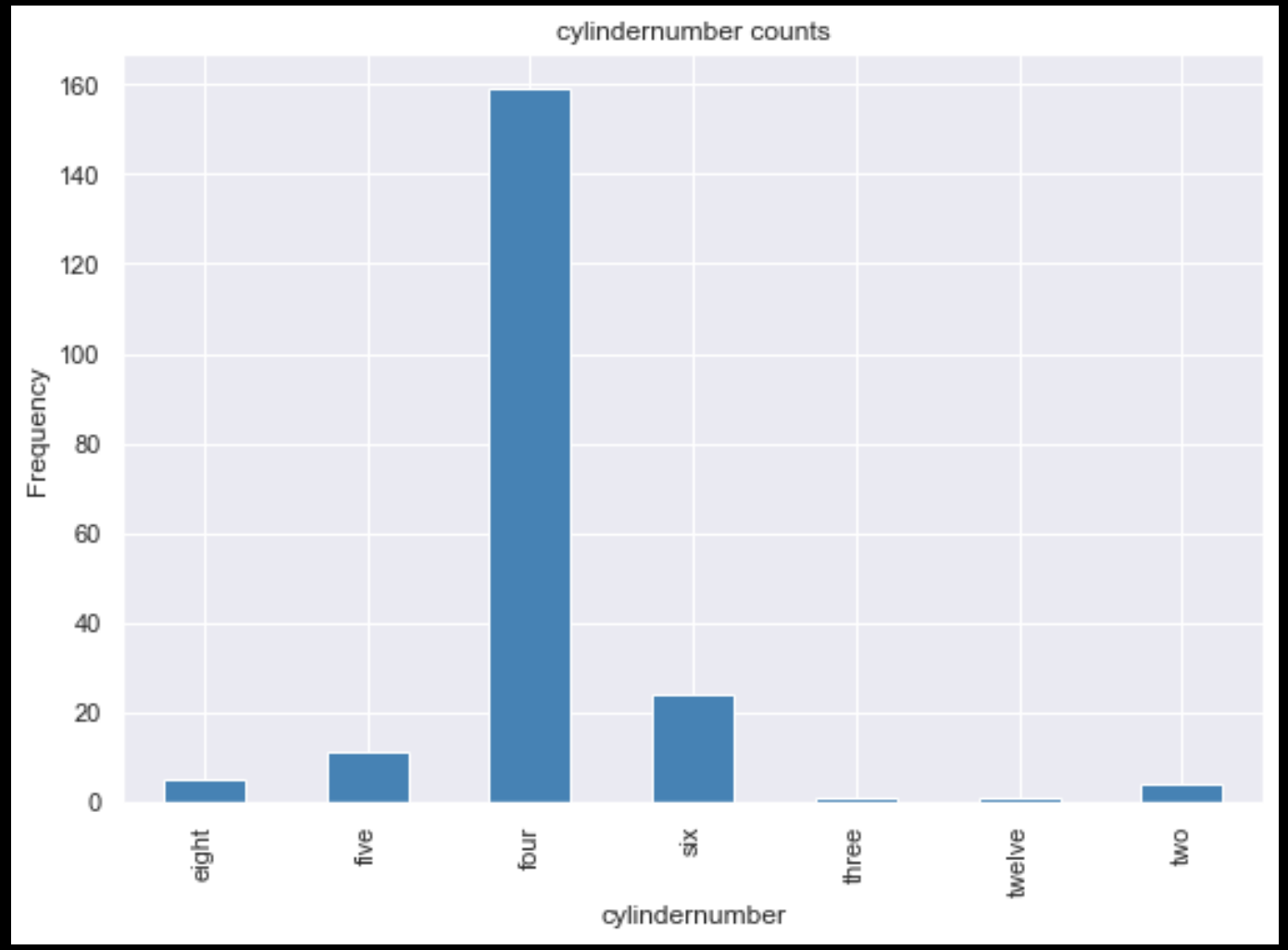
CARS CHOICES BASED ON ENGINE COUNT



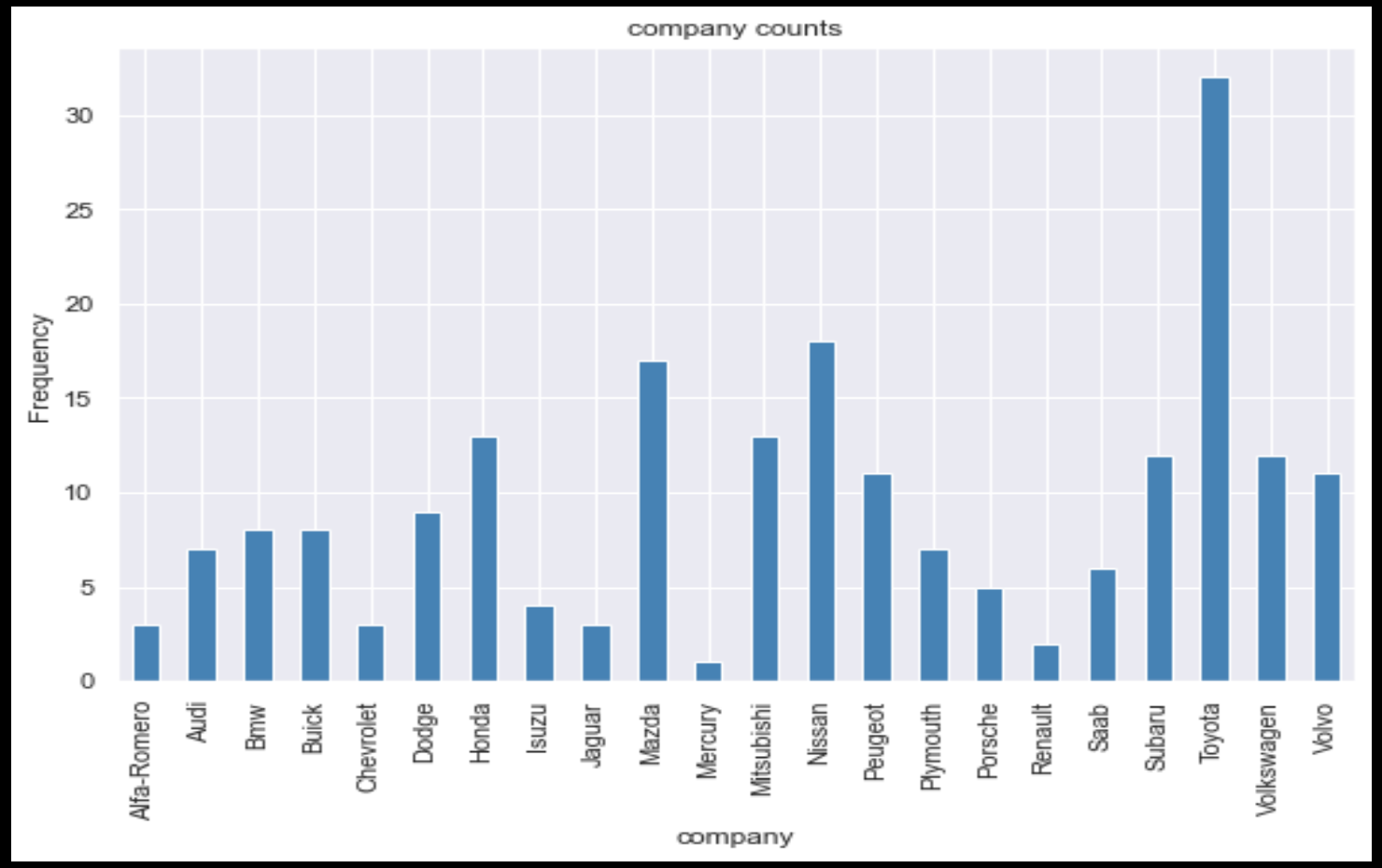
This visualization indicates that Overhead Car Engines Were The Most Preferred Engine Type Count

CARS CHOICES BASED ON CYLINDER NUMBER

From the visualization, it can be seen that cars with four cylinders were the most preferred.



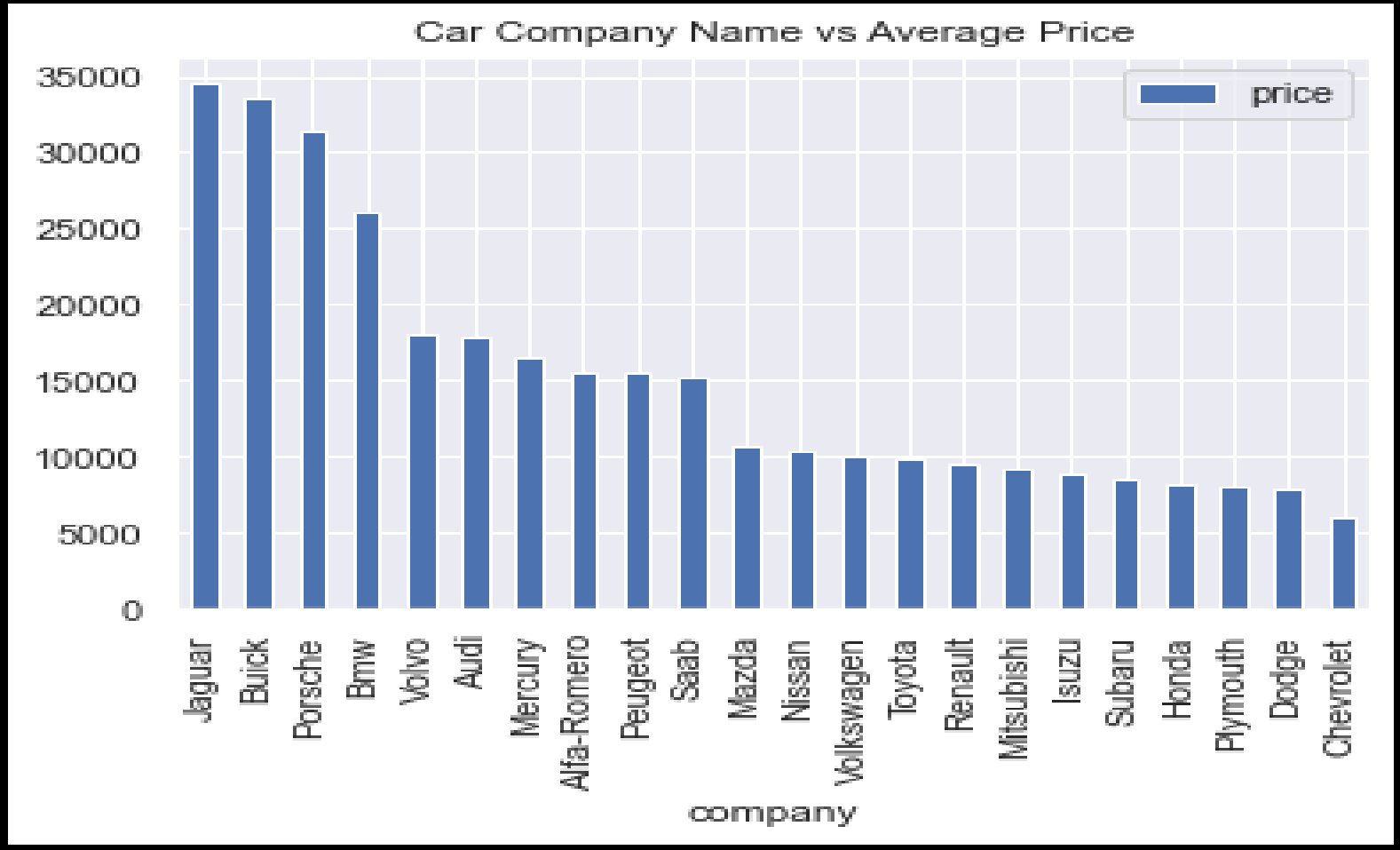
CAR PREFERENCE BASED ON COMPANY



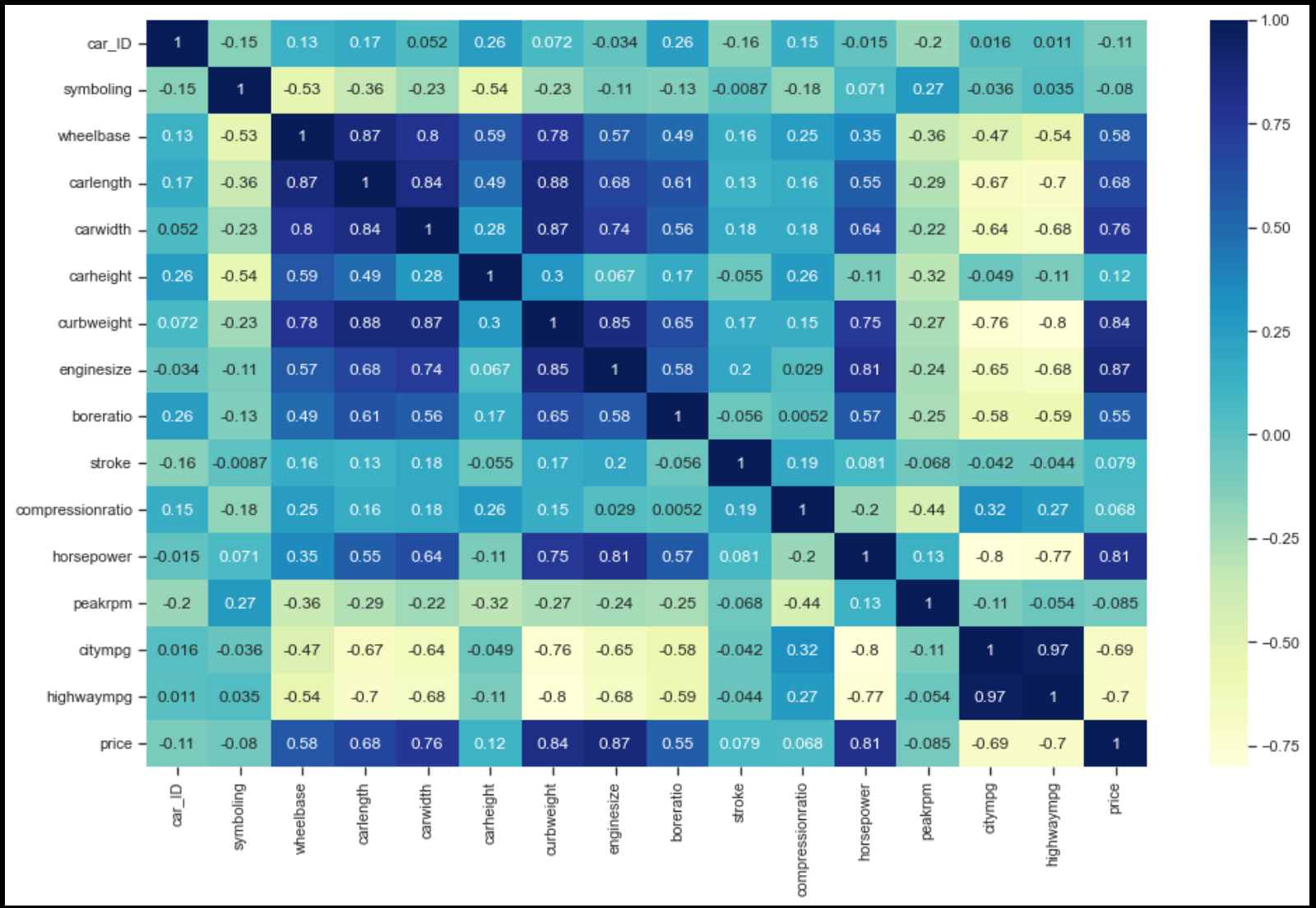
This visualization indicated that the Toyota company has the most users with a count of 32, followed by Nissan 18 and the Mazda with a count of 17.

AVERAGE CAR PRICE OF COMPANIES

This visualization indicates that Jaguar company by average has the average price.

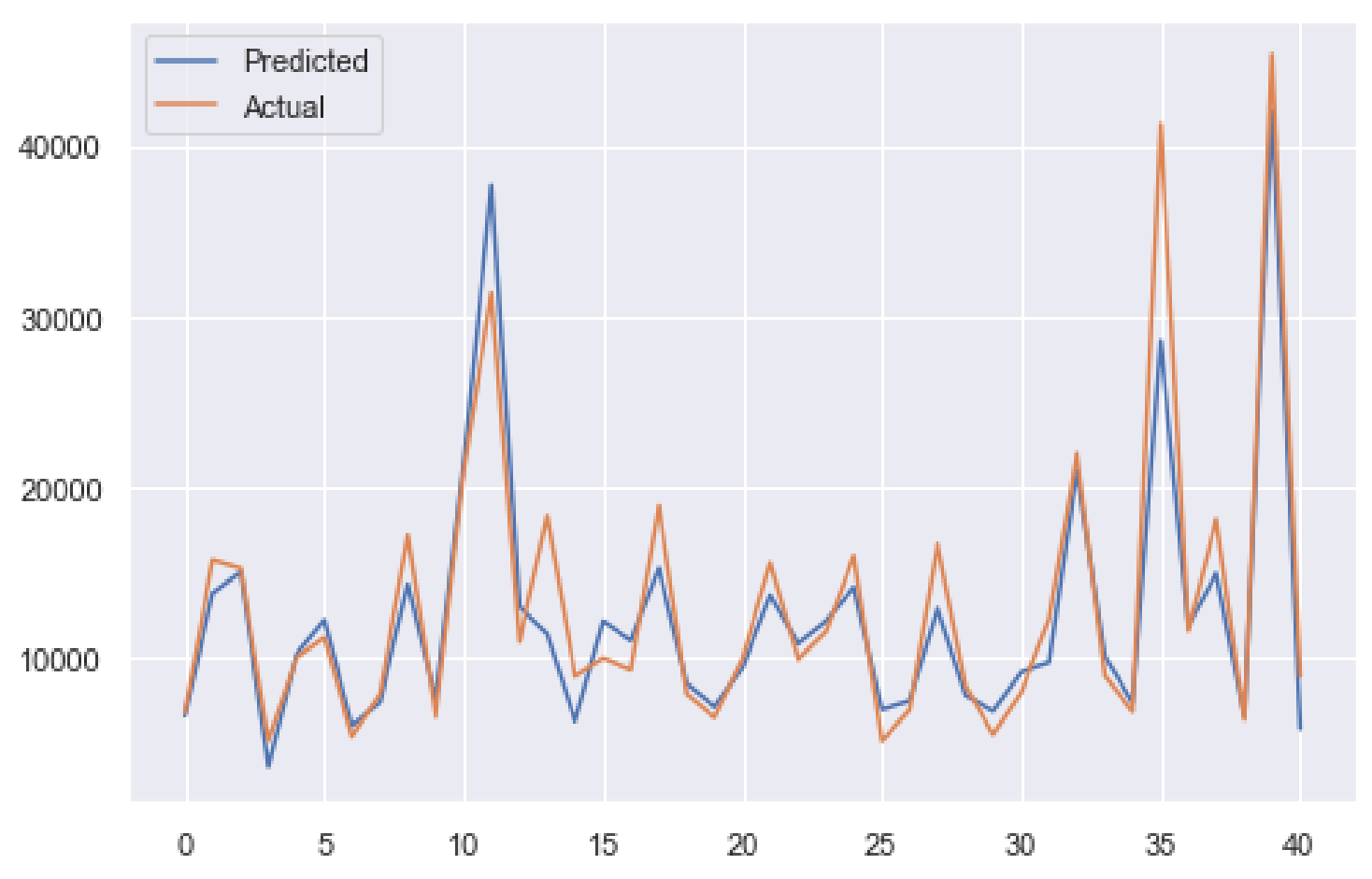


A HEATMAP INDICATING THE RELATIONSHIPS BETWEEN INDIVIDUAL VARIABLES



The Visualization indicates that car price has very strong correlation with engine size, car width, engine type, wheel base, car length, horse power, curb weight, bore ratio, stroke, compression ratio.

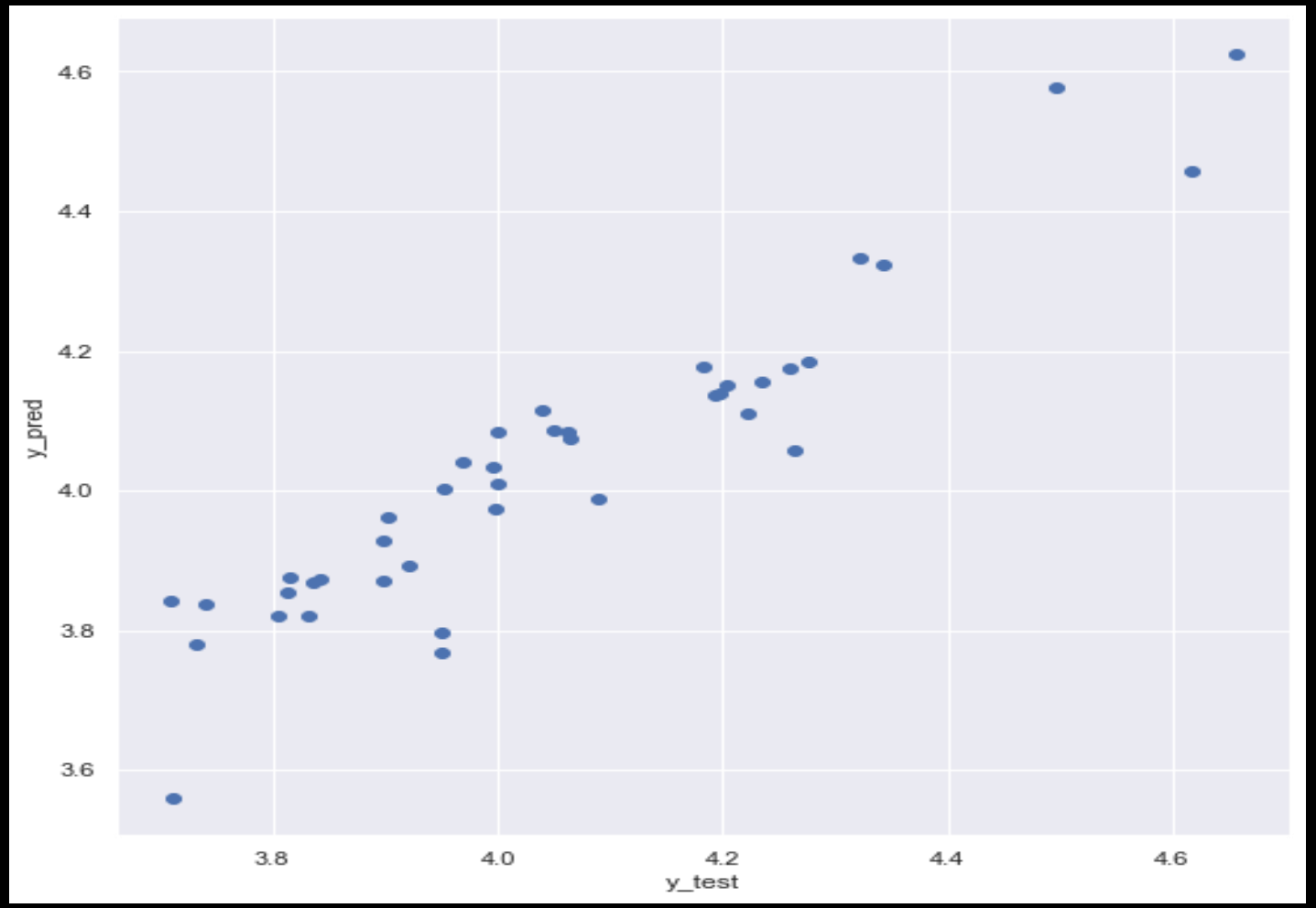
ACTUAL PRICE VERSUS PREDICTED PRICE GRAPGH



This visualization indicates the actual prices against the model prediction price with their individual represented legends

SCATTERPLOT INDICATING NEW PREDICTED PRICE VERSUS INITIAL PRICE

This visualization indicates that the is a strong positive correlation between our new predicting price model. This implies that with an R-squared value of 0.8834 and adjusted R-squared value of 0.708 the model perfectly predicted our prices



OBSERVATIONS AND CONCLUSION

From the model predictions and visualizations, the following observations were made:

- Fuel type had more effect on the pricing of the cars.
- Engine location and aspiration also has a visible effect on the pricing of the car
- The price of rear wheel drive is significantly higher than other drive wheel options.
- Cylinder number and engine type also seem to regulate the price of cars.
- Hardtop and convertible cars are definitely priced higher than other body types available. Also car body is contributing in determining the price.
- Diesel Cars had comparatively more car prices than gas car.

OBSERVATIONS AND CONCLUSION

- Turbo aspiration has more prices
- Hardtop and convertible carbody has higher prices than others
- RWD (Rear Wheel Drive) has higher average prices.
- Rear engine location has very high price comparatively than front location.
- Prices are affected by numbers of cylinder number.
- Door numbers doesn have much affect the prices of a car

OBSERVATIONS AND CONCLUSION

The variables/features which are significant in predicting the price of a car are:

- Engine Size
- Car Width
- Engine Type
- Wheel Base
- Car Length
- Horse Power
- Curb Weight
- Bore Ratio
- Stroke
- Compression Ratio