**Introduction**

The automobile file contains information about different vehicles, which is used to calculate the price of the vehicle.

**DATA CLEANING**

-Used missingno.matrix to plot a visual showing any missing values to determine if any columns were redundant (due to an excess of missing values) , and could be dropped.

-Used .dtype function to see the data types of all the columns.

-Used .descibe (include = 'all') to get statistical information as well as unique and most frequent values for all columns.

- With the .replace function all instances of '?' are replaced with Nan, to use the.dropna function and with thisfunctionall rows with values = Nan get dropped.

-Dropped columns: fuel-type, aspiration, num-of-doors, engine-location, wheel-base, length, width, num-of-cylinders, fuel-system and engine-type. With the drop function in pandas.

-Check for and remove duplicates.

-Change the data type of the price column which was obj to float.

MISSING DATA

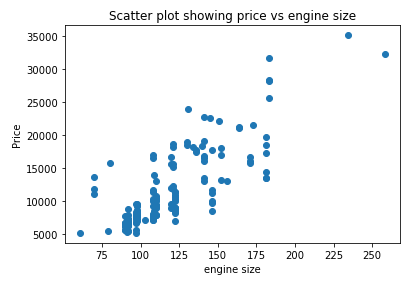
- With the .replace function all instances of '?' are replaced with Nan, to use the .dropna function.

- With .dropna function all rows with values = Nan get dropped.

DATA STORIES AND VISUALIZATIONS

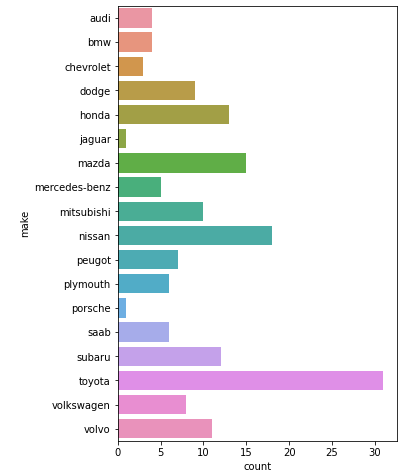
**Scatter plot: price vs. engine size**

-The scatter plot visual shows that as the size of the engine increases so does the price of the vehicle.



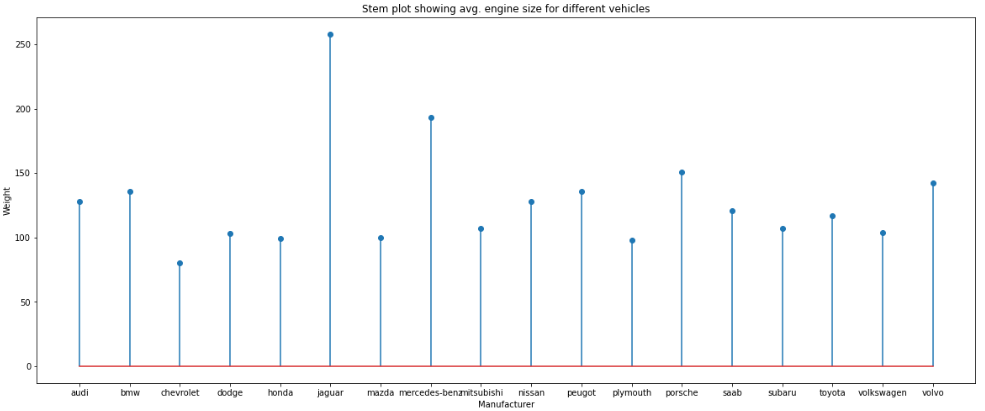
**Bar graph**

- Displays the total number of times each vehicle appears in the data table.



**Stem plot**

– Shows the average engine size for all vehicles in the data table.



-Together these two visuals show that despite the fact that the jaguar is 1 of the 2 vehicles that appear the least in the dataset it still has the highest average engine size (over 250).

-While Toyota has the highest total number of vehicles in the data table it has a standard average engine size (100-150).

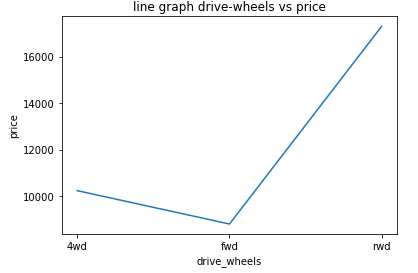
**Line graph:** Type of wheel drive vs price

-Used .value\_counts to return the number of vehicles with different wheel drive type’s (fwd 106, rwd 50, 4wd 8).

-Create drive\_wheel\_avg object to store a data table with the average prices for each of the wheel\_drive types.

-Use drive\_wheel\_avg object to plot a line graph to display drive-wheel vs price.

-Together these 2 techniques shows that even though a greater number of vehicles have fwd, the price for cars with rwd is higher.

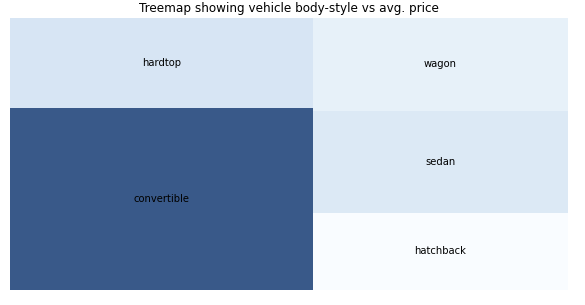


**Tree map**

- Create new object to store information on price and body-shape.

- Used this object to plot a tree map.

- Visual shows that the vehicles that are convertible have a higher average price then all the rest.



**Conclusion**

It can concluded that the jaguar vehicles have a higher price ,this is proven by the fact that a bigger engine size results in higher prices for vehicles and since the jaguars engine is the biggest it will therefore have a higher price. By viewing the vehicle make and symboling value in a data table it is clear that jaguars are a low risk vehicle having a 0 symboling value...

On the other hand vehicles with rwd and a convertible body shape have a high price, but they also have an average symboling value of 2.5 which makes these types of vehicles high risk.

Therefore when determining the price of the vehicle the symboling, engine size, drive-wheel type and body shape of a vehicle should be taken into consideration

**THIS REPORT WAS WRITTEN BY: Darren Vandeyar**