

Exploratory Data Analysis on the Automobile Data Set

Introduction

This is a dataset made of 205 rows and 26 columns that talk about cars.

The columns in the dataset are symboling, normalized losses, make, fuel type, aspiration, num of doors, body style, drive wheels, engine location, wheelbase, length, width, height, curb weight, engine type, num of cylinders, engine size, fuel system, bore, stroke, compression ratio, horsepower, peak rpm, city mpg, highway mpg, and price.

DATA CLEANING

In the Automobile file, I decided to remove the normalized-losses column because I didn't need it in my data analysis.

I checked the duplicate rows, but I didn't find any.

I found the values contained in the rows: price, horsepower, peak-rpm, bore, and stroke.

They were in string format and I had to convert them to float numbers for retrieving the value that I need for my analysis.

MISSING DATA

I checked for any missing value equal to Nan, but I didn't find any.

I checked for each row the unique element of that row and I found something interesting: there were some values equal to “?” that didn't make sense in the context.

So I understood that those were the missing value of the dataset.

I decided to convert all the “?” values to Nan values.

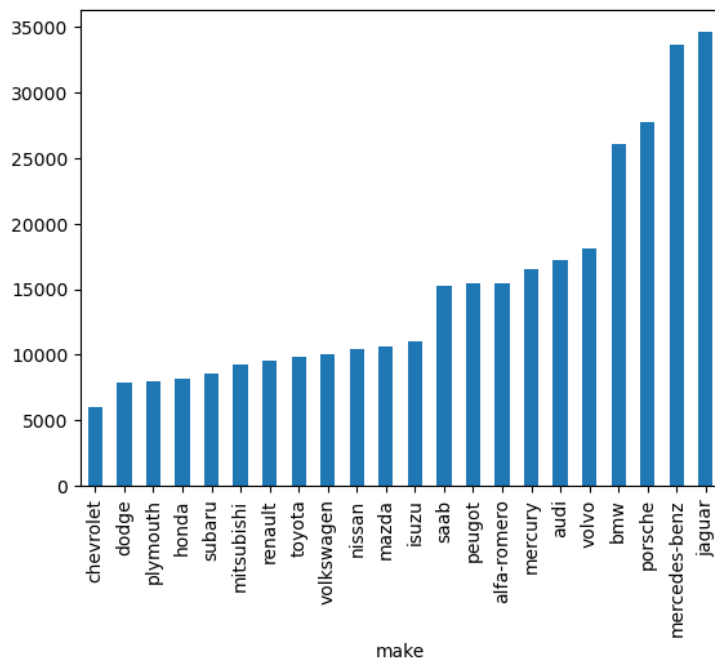
After doing this, I was able to handle the Nan value with pandas' methods.

There were missing values in the columns: price, horsepower, peak-rpm, bore, stroke, and num-of-doors. I decided to modify the missing values to the average for each value, except for the missing values in num-of-doors which I decided instead to replace with the most common value.

DATA STORIES AND VISUALISATIONS

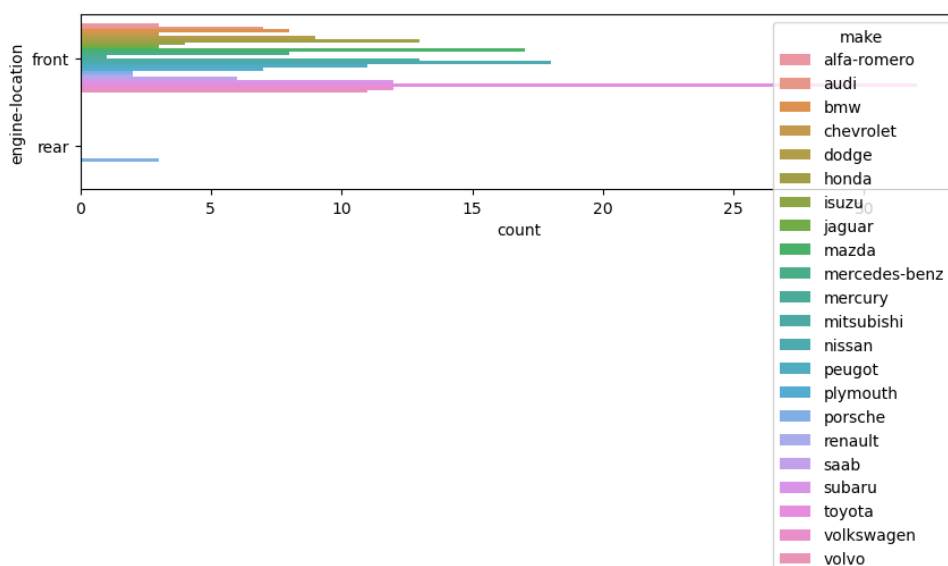
Average price vs model car:

Through this bar graph, you can visualize the average price for each car model from the cheapest to the most expensive where I found that the Jaguar cars are the most expensive on average.



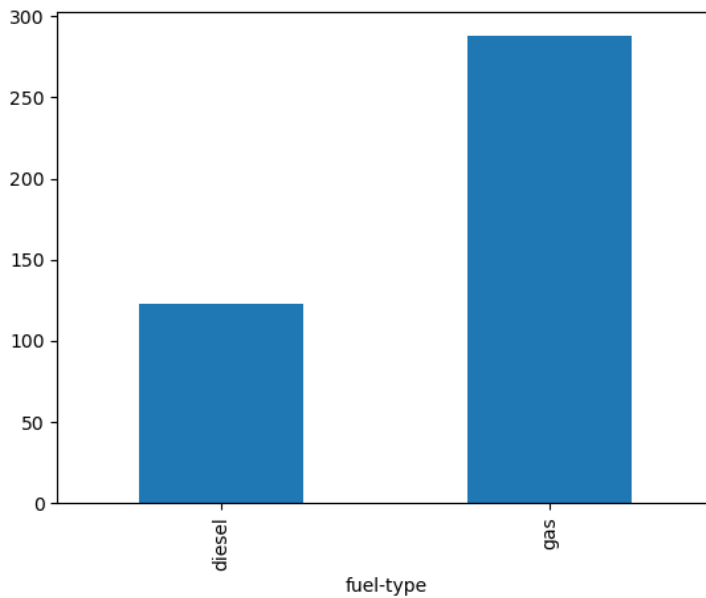
Engine location vs model car:

I made a count plot graph to show that every car model had the engine located to the front with only the exception of the model Porsche that which had to the rear.



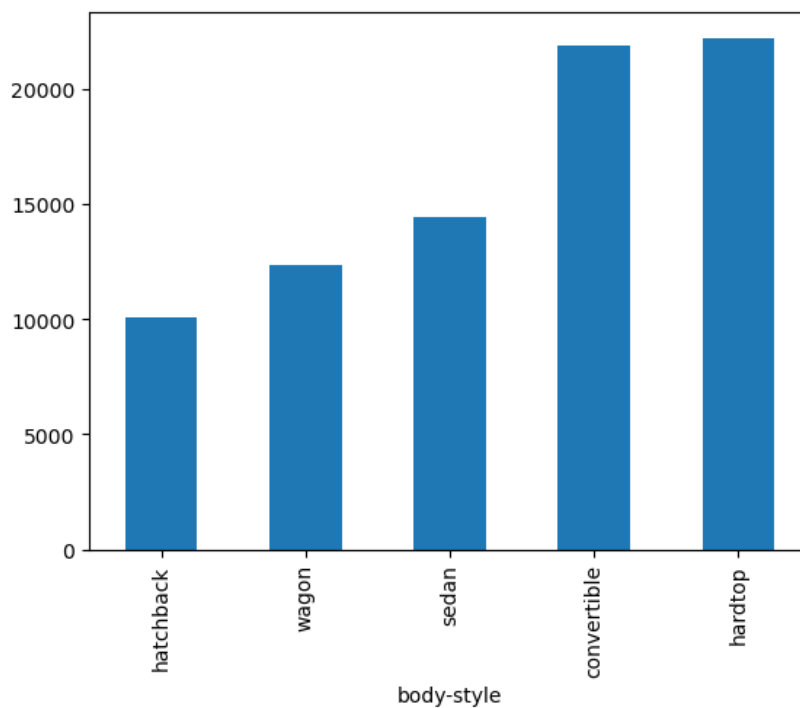
Horsepower vs fuel type:

I made a bar graph to show the gas fuel type has much more horsepower compared to the fuel type diesel.



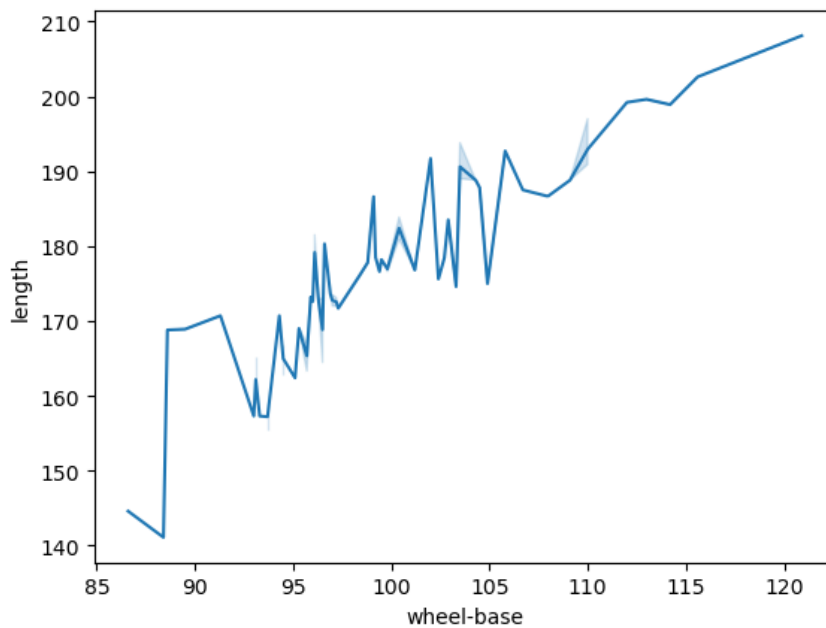
Body style vs average price:

I made a bar graph to visualize the relationship between the body style and the price:
I found the hardtop body style to be the most expensive.



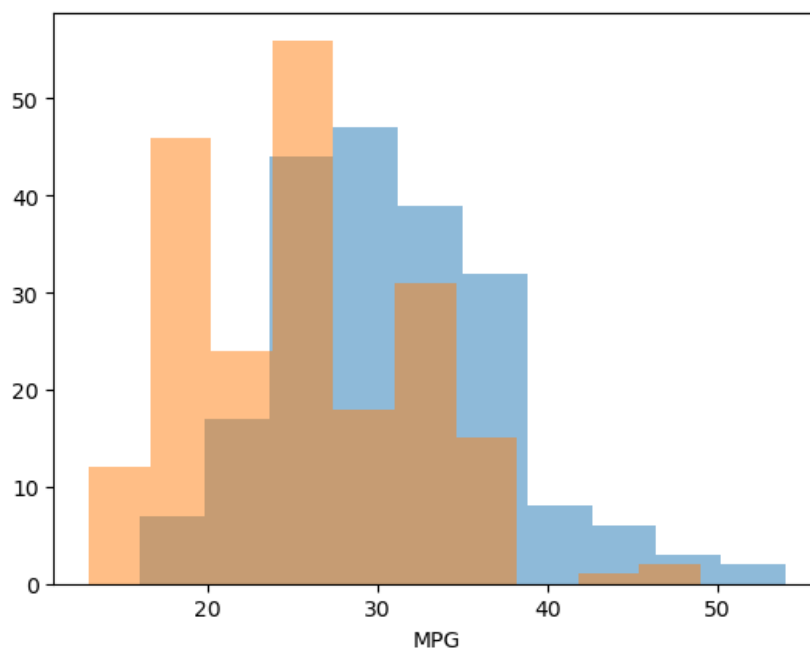
Wheelbase vs length:

I made a line plot graph to show that the wheelbase and length are directly proportional.



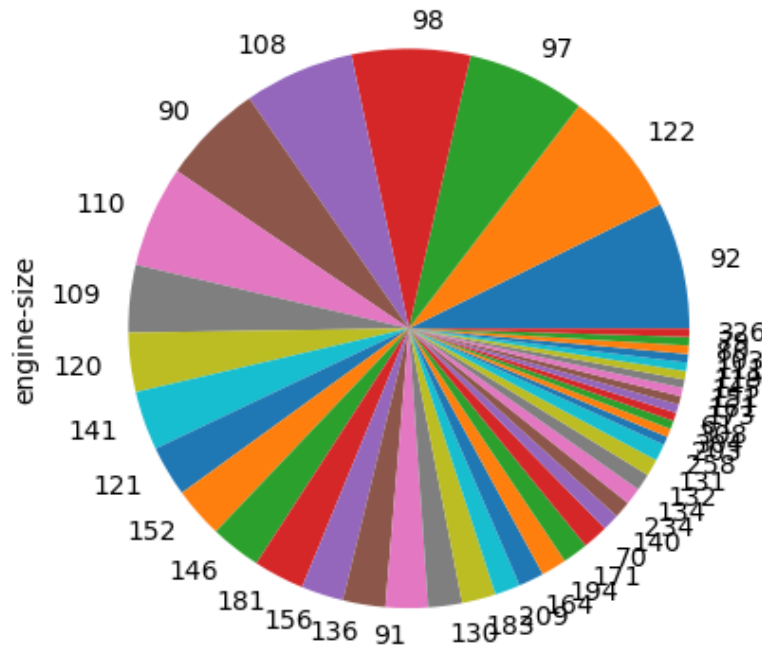
Fuel efficient City vs Highway:

I made a histogram graph to visualize that is generally more fuel efficient to drive in City compared to Highway.



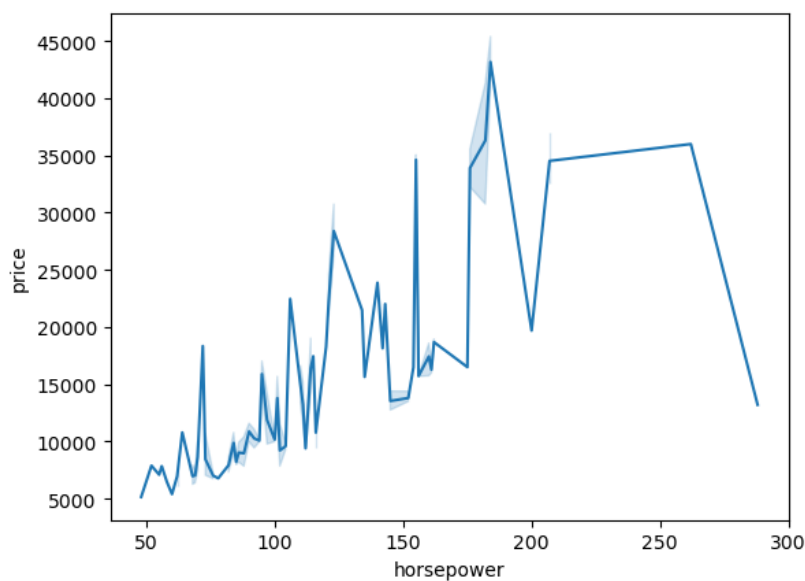
Common engine size:

Through this graph, you visualize that the most common engine sizes are 122,92,97 and 98.



Horsepower vs price:

We can see that horsepower affects the price, but there are some exceptions: the car with the most horsepower is a Porsche, however, his price is relatively cheap



THIS REPORT WAS WRITTEN BY: GAETANO LOPEZ

DATE: 20/12/2022

