Hyperiondev

TASK

Exploratory Data Analysis on the Automobile Data Set

Visit our website

Introduction

This dataset is about various characteristics of automobiles from different manufacturers with different body-styles, engines sizes, horsepower and more.

This data was analysed based on factors of interest that could be of use to manufacturers and potential customers.

DATA CLEANING

The automobile data from the text file was read into Jupyter notebook using a Pandas method. The data was looked at to get an idea of what was contained in the data frame.
Columns that were not necessary or redundant were removed. These columns include: 'symboling', 'normalized-losses', 'aspiration', 'num-of-doors', 'engine-location', 'length', 'width', 'height', 'curb-weight', 'fuel-system', 'bore', 'stroke', 'compression-ratio', 'num-of-cylinders', 'fuel-system'.

Using the .info() method, the data types for all columns was assessed to ensure they were all in a form that could be easily manipulated. Columns with incorrect data type were changed.

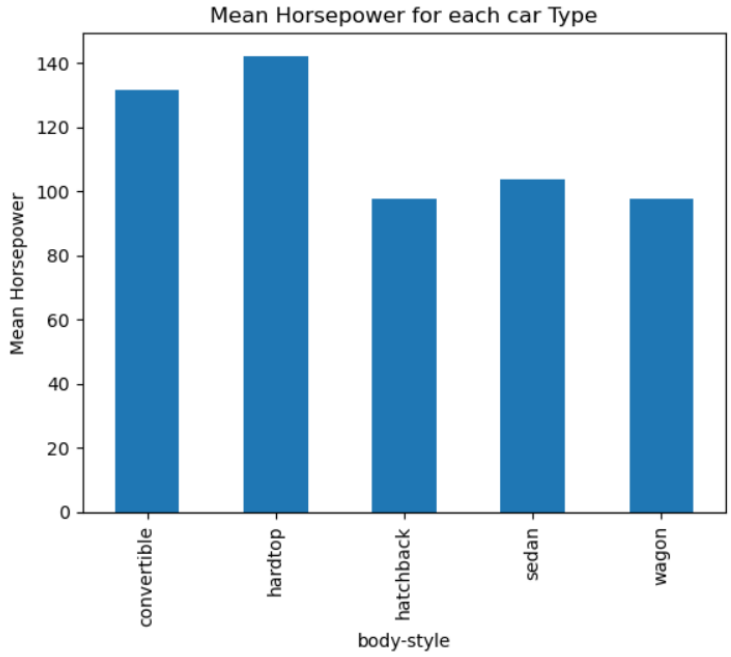
Checks were done for duplicate rows. No duplicate rows were found.

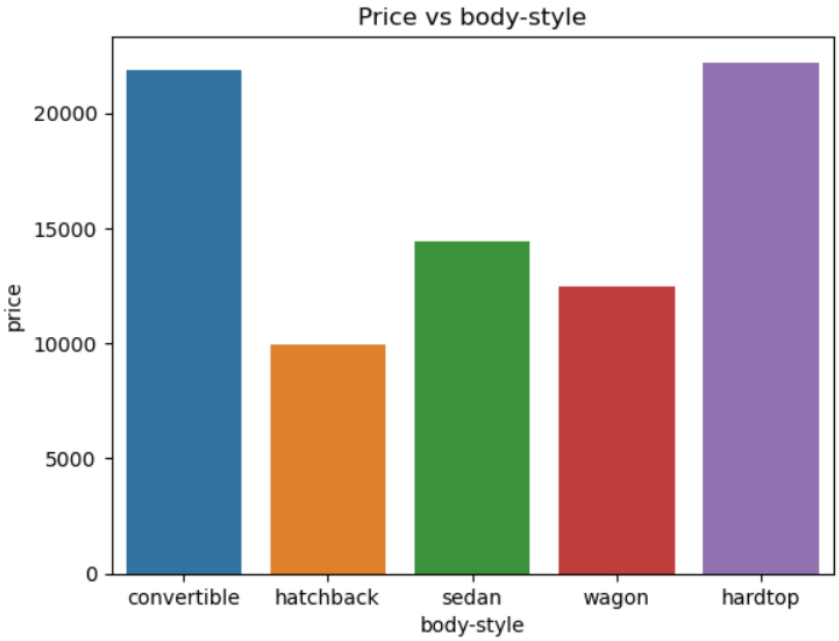
MISSING DATA

Once columns that were not relevant for the analysis were removed. There were still some missing data which were expressed as question marks.
In order to determine how many missing data each column had, I first had to convert the question marks to NaN values to be picked up by Pandas '.isna() method.
It was determined that 2 data entries were missing for horsepower and 4 entries were missing for price. The data set had a total of 205 row. The amount of missing data was considered insignificant, so the six rows were deleted.

DATA STORIES AND VISUALISATIONS

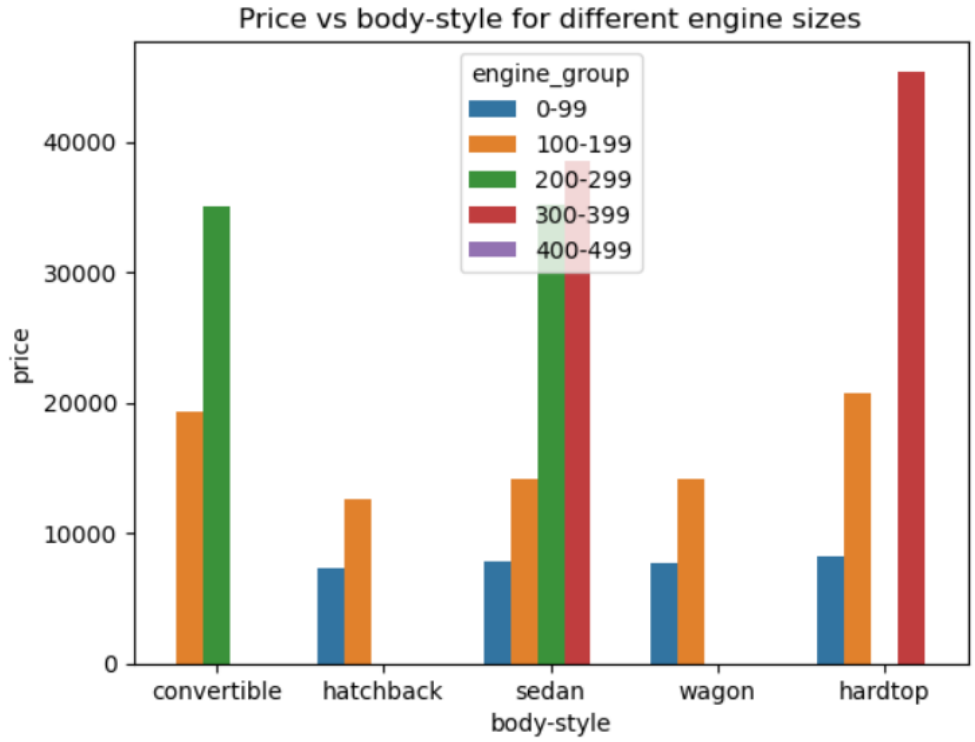
From the automobile data analysed. It was determined that hardtops and convertibles on average had the highest horsepower. The data also shows the price of a vehicle has a positive relationship to horsepower, as both the convertible and hardtop ranked highest for price.



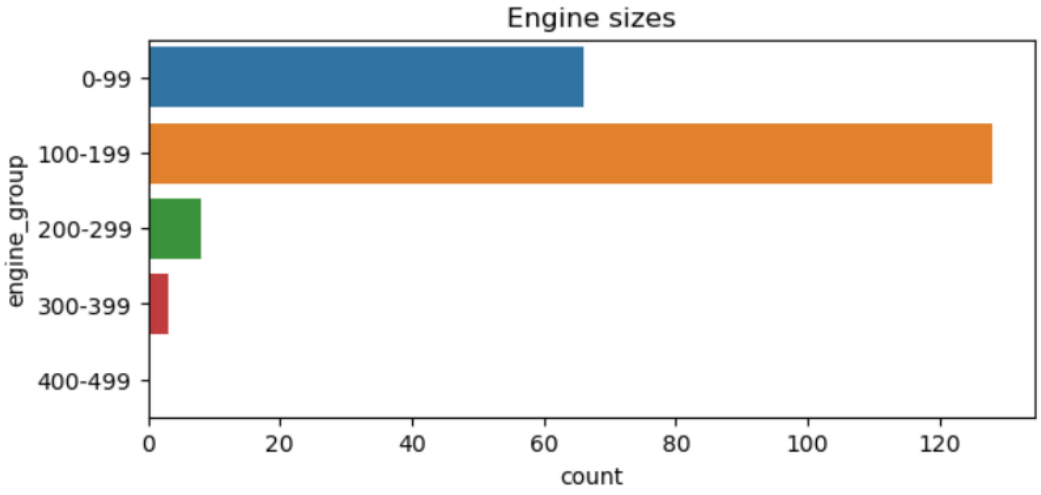


Hatchbacks are the most affordable and still provide decent horsepower. This will be a good car to buy for a first-time buyer.

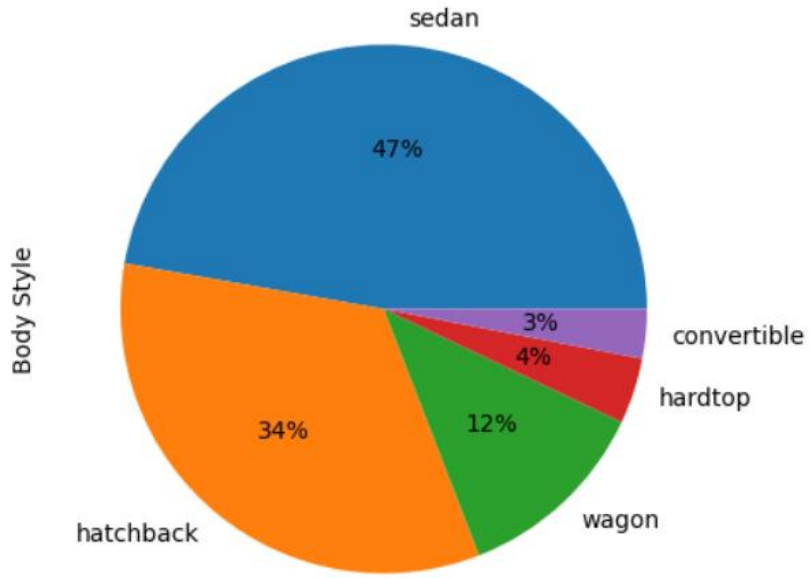
From the image of the plot below, you can also see that the engine size plays a role in the price of the car, which is also why customers tend to stay within 100-200cc engine sizes.



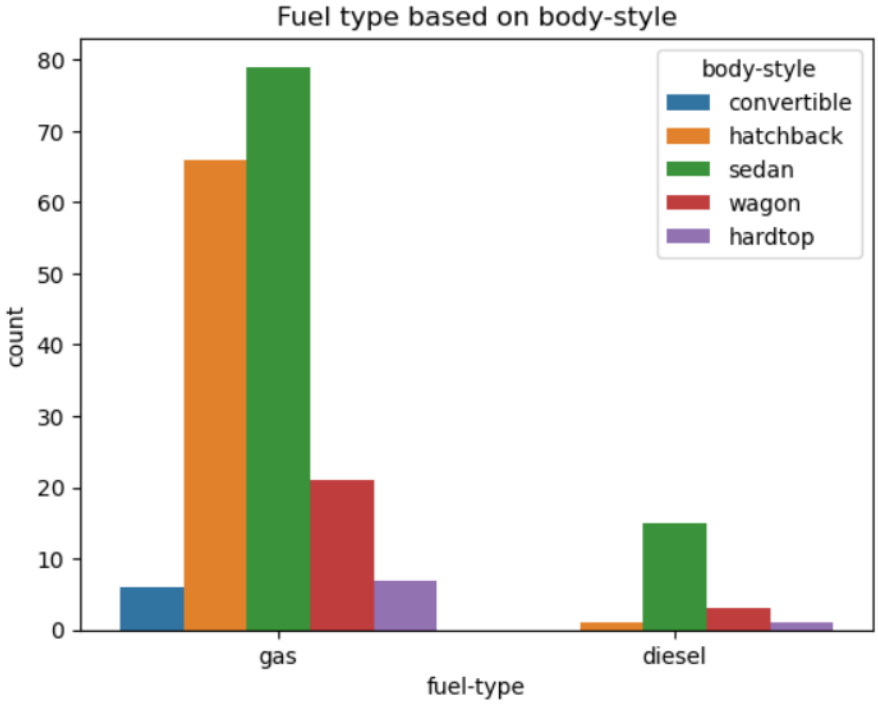
The most popular engine sizes in the cars bought range between 100-200cc (or 1-2litres). Car manufactures should consider making fewer models with more than 300cc engine size.



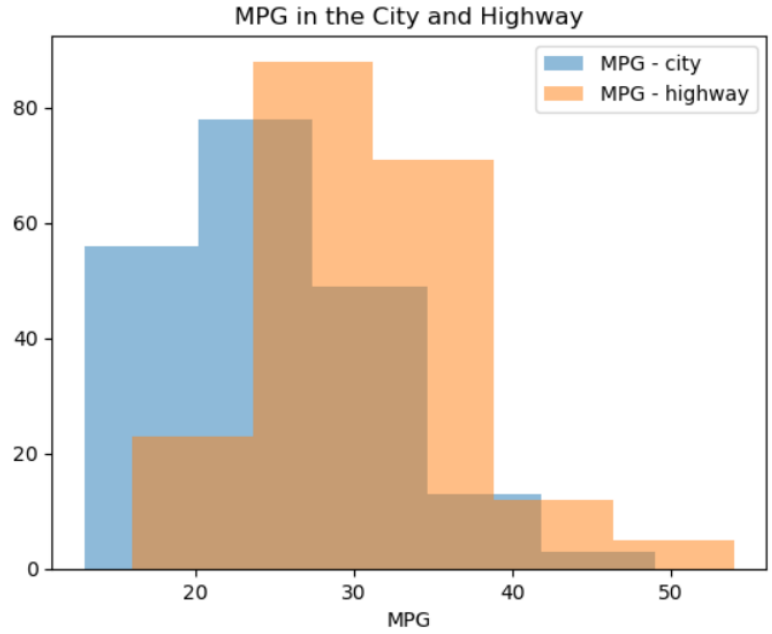
It appears that most customers prefer sedans (47%) to hatchbacks (34%). In addition to this, gas(petrol) based cars are preferred to diesel vehicles.



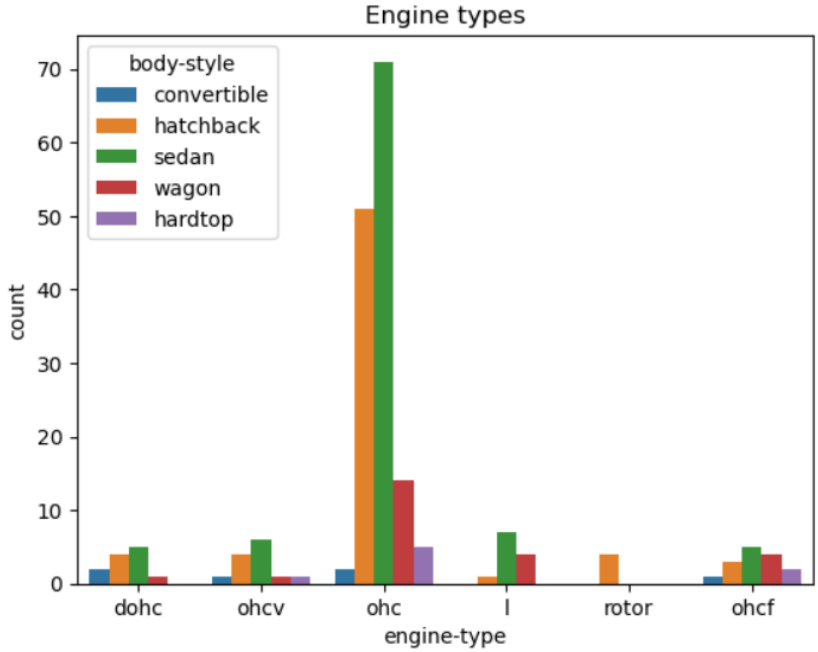
It would be advisable for manufacturers to consider phasing out the production of diesel vehicles as this option is not very popular with the customers. In the current market, diesel is also more expensive than gas(petrol) per litre.



The data also revealed that it is more fuel-efficient to drive on the highway than in the city. The city is congested and shorter trips appear to consume more fuel. If you want to save on fuel costs, it will be best to use your vehicle for longer trips on open roads such as a highway. Or prepare to increase your fuel budget for the month if you only drive in the city.

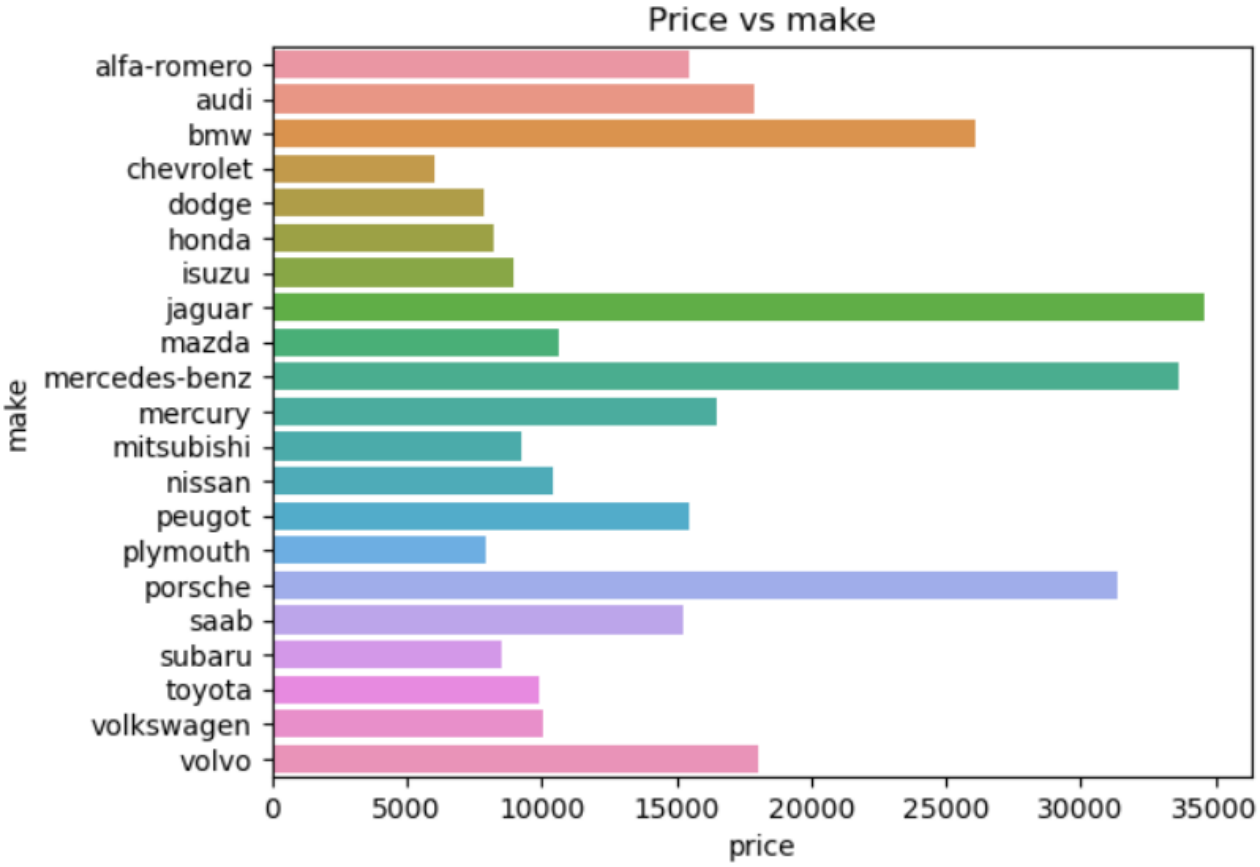


It was also noted that manufacturers seem to prefer using Overhead Camshaft (OHC) engines regardless of body-style. These engines are believed to be lighter and stronger, hence its popularity among manufacturers (Honda.com, 2022).



Mercedes Benz, Jaguar and Porsche, are considered to be luxury car brands. It is no surprise that these car makes rank at the top of the list when it comes to prices. If you are considering one of these car brands but don't want to break the bank, you may want to go for an entry level car or look in the second-hand market for a good deal.

If you do not care about luxury brands, the cheapest car brand is Chevrolet. Again, as a first-time customer or a parent looking for a car for your child, the best pairing would be Chevrolet Hatchback or Sedan with an engine size between 100-200cc.



Manufacturers can make use of this information to assist them in their decision-making processes, as well as potential clients in the market for a new vehicle based on their budget.

THIS REPORT WAS WRITTEN BY: KP USEH

References

Honda.com, 2022, https://engines.honda.com/why/overhead-cam-design#:~:text=Overhead%20Camshaft%20(OHC)%20Engines,an%20OHV%20design%20and%20more.