

**TASK**

**Exploratory Data Analysis on the Automobile Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

When considering buying a new automobile, many people try to remember the practical suggestions from fellow internet inhabitants when looking for a new automobile. While these are good practical points to look at, we also need to know the risk of the automobile we are looking at and if it will be a good investment for the next few years of use.

In this data report on automobiles, we will be looking at the impact of risk on various components of an automobile and what would be a safe buy for a car owner and what they need to look out for when buying a car.

The data set which we used contains the data of 205 different automobiles and attributes such as the symboling, price, make, fuel type, body style etc.

**DATA CLEANING**

To work with the data, we downloaded numpy, matlibplot, pandas, seaborn and datetime using Jupyter. After reading the automobile.txt in as mobile\_df, I looked at the first few lines and column names using ‘.head()’. There were no commas, out of place spaces or signs that worried me, but we are able to see that there are often ‘?’ marks indicating missing values.

I replaced the question marks with missing values which will makes the missing data easier to count and work with. The ‘engine-location’ column was removed, because the location of the engines were all at the front. The ‘normalized-losses’ column was removed, it is the relative average loss for insured vehicles per year, I did not need it to answer any of my questions and removed it.

The ‘aspiration’ column was removed because information of the internal combustion engine was not going to help with my questions. The ‘fuel-system’ column was removed as well as well as the ‘bore’ column. It I have the ‘stroke’ column and the ‘rpm’ column, which will make the ‘bore’ column unnecessary, due to the relationship between the bore and stoke that creates power (rpm) (Silvestro, 2020).

All duplicates were then dropped as well.

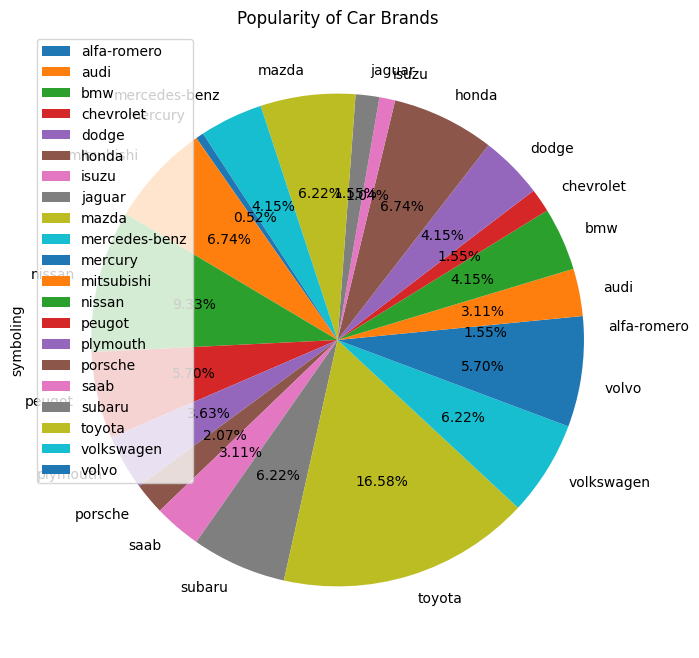
**MISSING DATA**

Missing values were counted using .isnull().sum() and the percentage determined using mobile\_df.shape[0] which showed that the highest missing value percentages were for ‘stroke’ (1.95%) and ‘price’ (1.95%) which is so low, that missing values were dropped using mobile\_df.dropna(inplace=True).

**DATA STORIES AND VISUALISATIONS**

**1.1 Popularity of Car Brands**

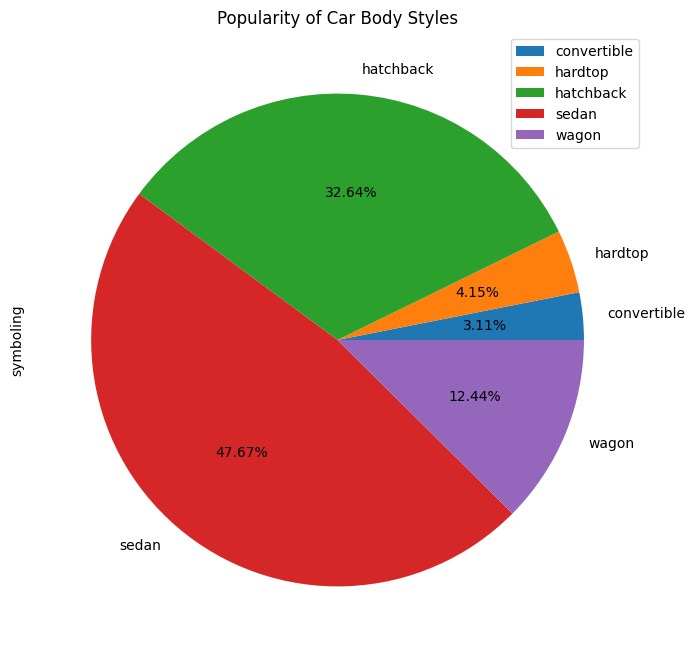
Firstly we look at the most popular makes that are sold of the automobiles (Popularity of Car Brands).

****

Toyota is by far the most popular brand (16.58%), followed by Nissan (9.33%) and Honda (6.74%) and Mitsubishi (6.74%). Toyota tends to be a popular brand due to the reliability, variety of genre and its focus on families. It is also a brand that is affordable in servicing and for parts (Enterprise Motor Group, 2022).

**1.2 Popularity of Car Body Types**

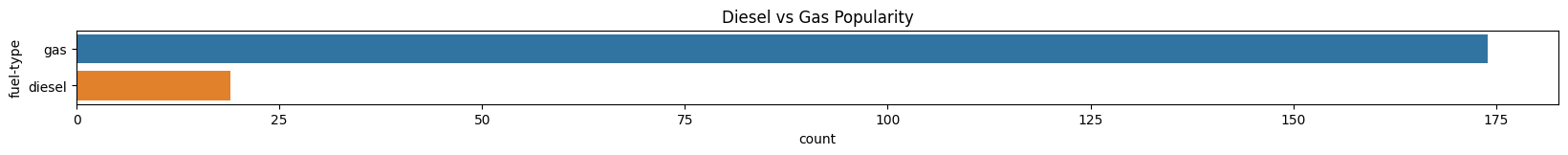
Choosing the body style of a car mainly depends on your needs and what you will be using the vehicle for, though there are styles that are widely preferred as can be seen in the graph below (Popularity of Body Styles).



The Sadan is by far the most popular body style (47.67%) for cars. It is the most popular body type for a car because it can easily fit five people and has space for cargo, which works well for families. These cars are better around corners, fuel-efficient and are stable on the road (Devan, 2021).

**1.3 Diesel vs. Popularity**

Let us look at the difference in popularity between cars that use gas and cars that use diesel.

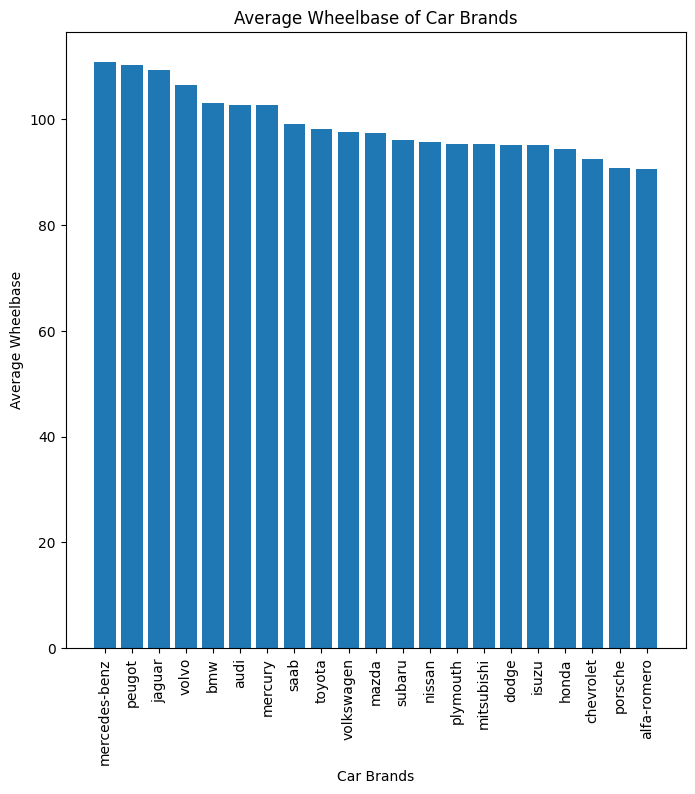


Petrol is clearly the most popular option, this can be because diesel cars are more expensive than petrol cars. You would think that because petrol is more expensive than diesel, more people would think of diesel cars as a good buy in the long run, but diesel is only more fuel-efficient for long distance and petrol would be more fuel-efficient over short distances and with stop-and-go (BusinessTech, 2016).

**1.4 Average Wheelbase of Cars**

Wheelbase is the distance between the middle of the front wheels and the middle of the back wheels. Wheelbase may differ from type to type of car, but you would generally want the wheelbase as long as possible, because it gives more interior space and a smoother drive (Carwow, 2022).

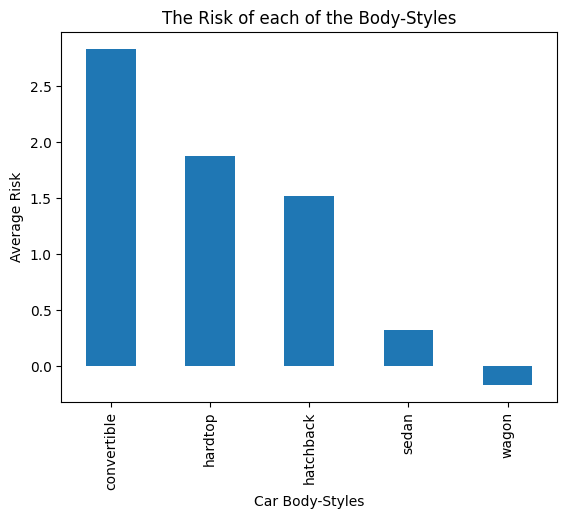
By grouping the makes together and using the average wheelbase, I created a bar graph which will help us determine what brands have a longer wheelbase.



The car brands with the highest average wheelbase are the Mercedes-benz, Peugeot and Jaguar brands. These brands are known as luxury brands, which is why the extra legroom with the bigger interior would be needed. Vehicles with the longer wheelbase generally are also used as high-end taxis (Carsblog, 2023).

**1.5 The Risk of Each Car Body Type**

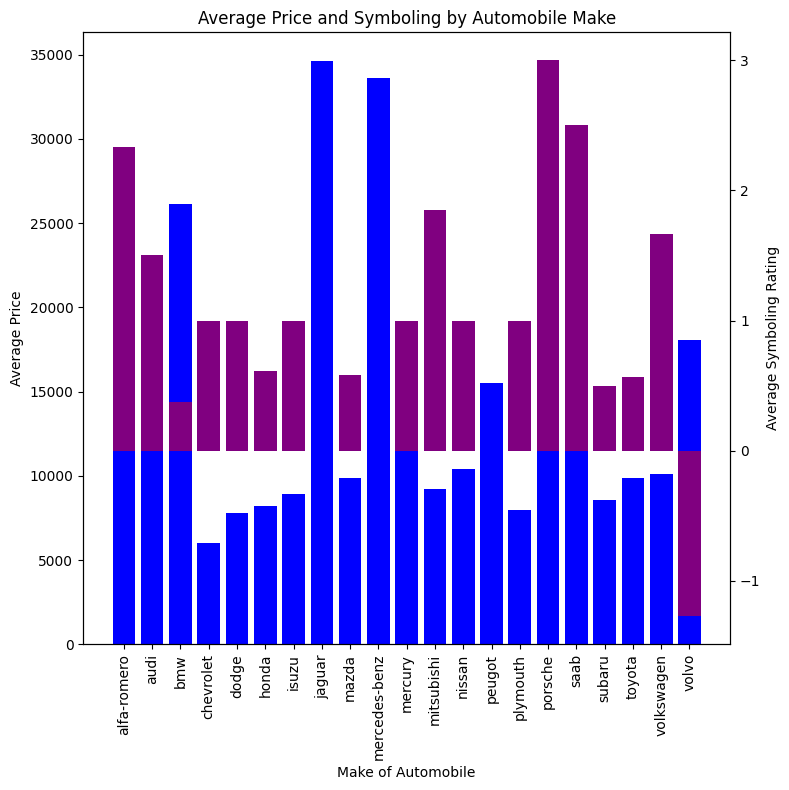
The risk is indicated by symboling, which was all indicated in the ‘symboling’ column. Symboling is an indication of the insurance risk level of the car. This can increase for a car if it is dangerous, so we want cars with smaller risk.



The car body types with the lowest insurance risk is the sedan and the wagon, which makes these cars safer. There are not many automakers that still make wagons, but if you are able to find one, it is a durable car and good on rough terrain. The sedan is considered a safe car because of its safety and handling (Auto Simple, 2017).

**1.6 The Average Price and Symboling by Automoble Make**

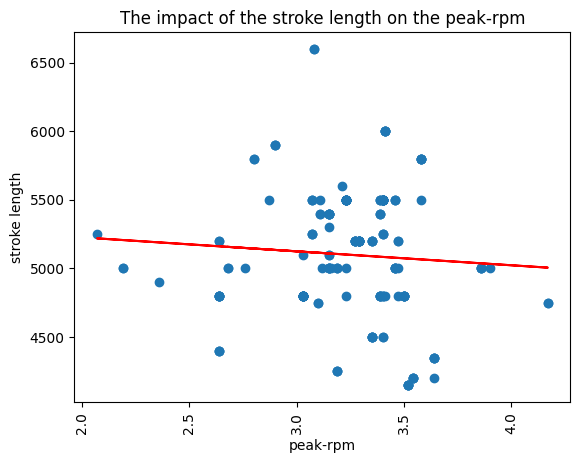
The data was grouped by make for this graph and the average price and symboling for each added with twinx(). A good combination would be a make with a low risk and low price.



We can see that Toyota, Saab, Mazda and Honda have low risk, but still low prices, which will make it a good buy. Chevrolet and Plymouth may cost little, but the risk for these are higher. Toyota is known for its solid risk profile due to its worldwide market share and strong focus on R&D for safer vehicles (GlobalData, 2023).

**1.7 The impact of the stroke on peak-rpm**

The stroke is the distance of the cylinder in the engine which the piston must travel, a lower stroke will give a higher peak-rpm (resolutions per minute) and will indicate how fast the engine operates (Cars.com, 2022).



In the graph it can be seen that the lower the stroke length is, the higher the peak-rpm will be which will produce more power in the engine. This is because a stroke which is short does not have far to travel and can cover greater distance in a short time, cars like this will also rev higher (Silvestro, 2020).

**CONCLUSION**

By looking at the results, we can see that the ideal vehicle with the least risk will be a Toyota sedan that uses petrol. Even if Toyota is not one of the brands with the longest wheelbase, we can ignore that, because these tend to be less agile around corners (Carsblog, 2023). The sedan is not just the most popular choice in car types, but also the second safest. It is also important to look for a car with a higher peak-rpm for engine power.

**REFERENCES**

* Auto Simple. (2017). Car Body Styles Explained. Reviewed on 16 July 2023. https://www.autosimple.com/blog/car-body-styles/#:~:text=Sedans%20tend%20to%20provide%20better,sedan%20is%20the%20classic%20choice.
* BusinessTech. (2016). Diesel vs Petrol Car- How to Choose Which is Best. Reviewed on 16 July 2023. https://businesstech.co.za/news/motoring/137027/diesel-vs-petrol-car-how-to-choose-which-is-best/
* Carwow. (2022). What is a Car Wheelbase? Reviewed on 16 July 2023. https://www.carwow.co.uk/guides/glossary/what-is-a-car-wheelbase-0282
* Carsblog. (2023). Importance of Car Wheelbase. Reviewed on 16 July 2023. https://www.dubizzle.com/blog/cars/car-wheelbase/#:~:text=The%20longer%20the%20wheelbase%20the,taxis%20also%20have%20larger%20wheelbase.
* Cars.com. (2022). What Does RPM Mean in Cars? Reviewed on 16 July 2023. https://www.cars.com/articles/what-does-rpm-mean-in-cars-1420697442798/
* Devan, S. (2021) Types of Cars- Different Car Body Types Explained. Reviewed on 16 July 2023. https://www.carsome.my/news/item/types-of-cars-different-car-body-styles-explained#:~:text=A%20sedan%2C%20also%20known%20as,bonnet%2C%20cabin%2C%20and%20boot.
* Enterprise Motor Group. (2022). Why Toyota is Such a Popular Brand. Reviewed on 14 July 2023. https://www.enterprisecars.co.nz/more/news/buying-a-used-car/why-toyota-is-such-a-popular-brand.html
* GlobalData. (2023). Toyota Motor Corp-Risk Profile. Reviewed on 16 July 2023. https://www.globaldata.com/data-insights/automotive/toyota-motor-corp-risk-profile/
* Silvestro, B. (2020). Why the Most Powerful Engines Have Short Strokes and Big Bores. Reviewed on 14 July 2023. https://www.roadandtrack.com/car-culture/a30443334/engine-stroke-vs-bore-explained/

**THIS REPORT WAS WRITTEN BY : Carla Hills**

