

Automobile Report

About this dataset:

This dataset lists different automobile makes and its design components such as the engine size, horsepower, and its overall price.

Data Cleaning:

- Removed the following column's because they did not add any value to the data story:
 - symboling;
 - aspiration;
 - length;
 - width;
 - height;
 - num-of-doors;
 - engine-location;
 - fuel-system;
 - compression-ratio and;
 - drive-wheels.
- Dropped any duplicate rows.
- Visualised all of the unique values in each column containing visible string data to ensure there were no variations of the same value.
- Checked data types.
- Changed integer columns which were 'object' to either int64 or float64.

Missing Data:

- Visualised the number of missing values.
- Removed the normalized-losses column as it had 41 missing values, and it was not key to showing the design of an automobile.
- Columns bore, stroke and price had 4 missing values each, and columns horsepower and peak-rpm had 2 missing values each.
- Replaced all missing values with 0.

Data Stories and Visualisations:

1. Comparing five most expensive automobiles to five least expensive automobiles.

Most Expensive:

Make = Mercedes-Benzes

body-style = hardtop

Cheapest:

make = Subaru

body-style = hatchback

Key differences:

- The expensive automobiles have a heavier curb-weight.

- The expensive automobiles have more cylinders.
- The expensive automobiles have 3x bigger engine size.
- The expensive automobiles have roughly double the horsepower.
- There is very little difference between the bore, stroke, and peak-rpm.
- The expensive automobiles have lower city-mpg.
- The expensive automobiles have lower highway-mpg.

2. Most popular body style produced by each make.

Created a countplot which counted how many body styles were produced by each make.

It highlighted that the hatchback and the sedan are the most popular body styles.

Volkswagen produces the most hatchbacks.

Toyota produces the most sedans.

Just over half (12) of the makes produced wagons but the amount of wagons produced is about half that of the hatchbacks and sedans.

Only four makes produce convertibles and only four makes produce the hardtop.

3. Compare the two most popular body styles to each other: hatchbacks and sedans.

On average, the sedan has a heavier curb-weight, larger engine size, higher peak-rpm and costs more than the hatchback. There is very little difference between the two regarding the other columns (wheel base, bore, stroke, horsepower, city-mpg, and highway-mpg).

The hatchback may be produced more as it is cheaper.

4. Compare the two least popular body-styles to each other: convertible and hardtop.

There is very little difference between the two body-types. The hardtop has a larger engine and more horsepower which may indicate why it is slightly more expensive than the convertible.

5. Assumptions made when comparing the two most popular body-styles to the two least popular body-styles.

The least expensive body-styles are produced the most.

The engine-size may influence the price. The bigger the engine the greater the price.

The curb-weight may also influence the price. The heavier an automobile the more expensive.

6. A line graph showing the average price of all five body-styles.

A graph is required as the price appears to be the biggest indicator of quantity of production.

7. A pie chart of the average engine-size of all five body-types.

A graph is required as this appears to be the biggest factor which influences the price.

8. A box plot comparing the body style to curb weight.

A graph is required as when comparing the body style, curb weight seems to influence the price.

By comparing all the data and not just the average it shows that the curb weight does influence the price.

The heavier the car, the more expensive.

9. A bar chart comparing the number of cylinders to the average price.

At first glance, the bar chart appears to show that the number of cylinders does affect the price. However, automobiles with two cylinders may be more expensive than automobiles with three or four cylinders. An automobile with eight cylinders may be more expensive than an automobile with twelve cylinders.

The number of cylinders is not a factor in price.

10. A bar graph to demonstrate whether or not the engine type effects the price.

There are seven engine types. Dohcv does not have an average price so only six engine types can be investigated.

Ohcv has the highest price and is the only engine type which differs from the rest.

The other five engine types average price do not differ from each other greatly.

Therefore, if an automobile has an ohcv engine then the price may be higher but otherwise engine type does not influence price.

11. Engine size

A pie chart of the average engine size of the five body styles has previously been produced (7) and suggests that engine size influences the price.

Created a histogram comparing the top five most expensive automobiles engine size to the top five cheapest automobiles engine size.

The chart showed that more expensive the automobile the bigger the engine size indicating that engine size influences the price.

12. Horsepower

Created a histogram comparing the top five most expensive automobiles horsepower size to the top five cheapest automobiles horsepower.

When you compared the five body-styles to their average horsepower it appeared it may only slightly influence the price. However, when you compare the five most expensive automobiles to the five least expensive automobiles it confirms that horsepower does influence the price.

The chart showed that more expensive the automobile the greater the horsepower indicating that horsepower is also an influencing factor when it comes to price.

13. Compare city-mpg to highway-mpg

Visualised a histogram to compare the five most expensive automobiles city-mpg and highway-mpg to the five least expensive automobiles city-mpg and highway-mpg.

Like horsepower, when you compared the five body-styles to their average city-mpg and highway-mpg it appeared they only slightly influenced the price. However, when you compare the five most expensive automobiles to the five least expensive automobiles it shows that city-mpg and highway-mpg do influence the price.

The histogram showed that the most expensive automobiles had fewer highway-mpg than the cheaper automobiles. The graph also showed that the most expensive automobiles had lower city-mpg than the cheaper automobiles.

The higher the highway-mpg, the least expensive the automobile.

The higher the city-mpg the least expensive the automobile.

14. What effects the price most?

Overall, the engine size effects the price the most.

This was shown when comparing the average engine size of the body styles and when comparing the top five most expensive automobiles and the five least expensive automobiles engine sizes.

The bigger the engine the greater the price.

This report was written by Rachel Birrell