



TASK

Exploratory Data Analysis on the Automobile Dataset

[Visit our website](#)

Introduction

DATA CLEANING

During the data cleansing phase, several steps were taken to prepare the dataset for analysis.

1. Column Removal:

Unessential columns such as 'wheel-base,' 'body-style,' 'drive-wheels,' 'bore,' and 'peak-rpm' were identified and removed from the dataset. This decision was driven by the focus on specific visualisations and the relevance of information related to travel aspects. For future projects, excluding these columns might prove advantageous.

2. Remove Duplicate Rows:

Duplicate rows were removed to maintain data integrity and ensure that each entry in the dataset is unique.

3. Remove Rows with Missing Data:

Rows with missing or zero values in crucial columns, such as profit were filtered out. This step helps in focusing the analysis on complete and meaningful data.

4. Filling missing values:

Missing data can lead to gaps in analysis and potentially distort the results. Filling in missing values helps to maintain the integrity of the dataset, ensuring that every record has complete information.

5. Change Data Types:

- Data types were adjusted to facilitate easier manipulation.
- 'profit' columns were converted to the integer data type (int64) using NumPy this was shown when checking the datatypes of all the columns.

6. Flatten Columns:

- were flattened, making the data more accessible for analysis.
- Inspecting flattened data involves reviewing the structure, content, and quality of the data in its flattened form.

Data Stories and visualisations

Data Exploration:

Identify Relationships Between Variables:

In the exploration phase, the goal is to identify interesting relationships or patterns within the dataset. Specific questions or relationships can be defined for further investigation.

Which are the 5 most expensive automobile?

Mercedes-Benz and Porsche are among the top 5 most expensive cars. Luxury brands like Mercedes-Benz and BMW are known for their high-priced vehicles and this shows in their price.

Make	Price
Mercedes-benz	45400
BMW	41315
Mercedes-benz	40960
Porsche	37028
BMW	36880

What are the top automobiles with high mpg ?

Honda has the highest highway miles per gallon (mpg), indicating good fuel efficiency. Japanese brands such as Toyota and Nissan also appear, suggesting a focus on fuel efficiency.

Make	Highway-mpg
Honda	54
Chevrolet	53
Nissan	50
Toyota	47
Toyota	47

Top 7 makes with the most diesel cars?

Peugeot is the leading brand with the highest count of diesel cars.

German luxury brands (Mercedes-Benz, Volkswagen) are present, suggesting a mix of fuel types in their lineup.

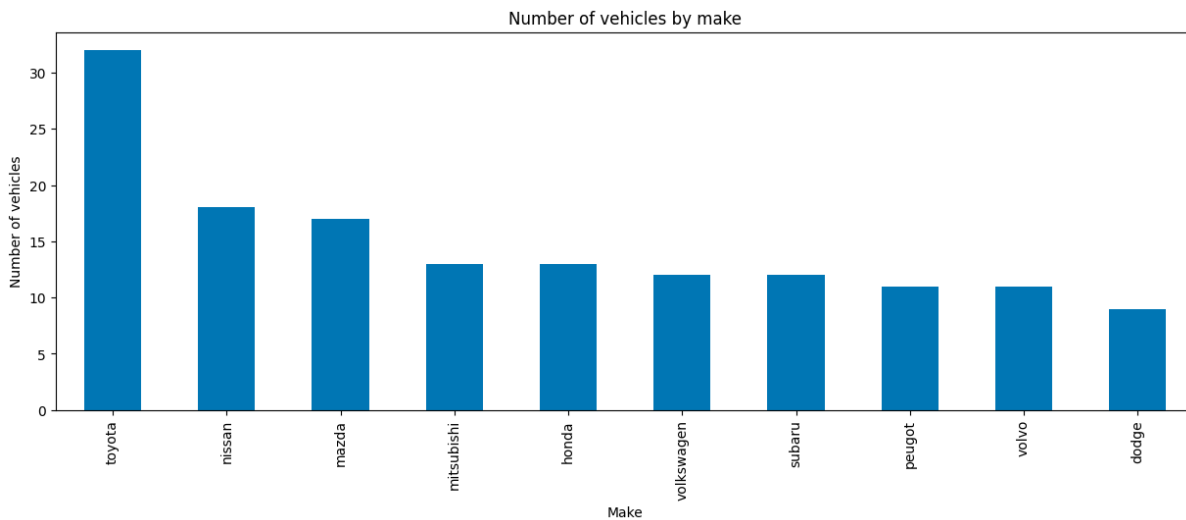
Japanese brands (Toyota, Mazda, Nissan) also have a presence, but with fewer diesel cars.

Volvo is the only non-German luxury brand in the list.

Make	Count
Peugot	5
Mercedes-benz	4
Volkswagen	4
Toyota	3
Mazda	2
Nissan	1
Volvo	1

- **Luxury Brands:** The absence of specific BMW models in the top 5 most expensive cars reinforces that the analysis considers the brand holistically. Luxury brands, including Mercedes-Benz and Porsche, maintain a strong presence in high-priced vehicles.
- **Fuel Efficiency:** Japanese brands dominate the list of cars with the best highway mileage, with Honda, Toyota, and Nissan emphasising their commitment to fuel-efficient designs.
- **Diesel Variants:** Diesel cars are not limited to a specific market segment. Peugeot's leadership in diesel cars, alongside the presence of both luxury and non-luxury brands, underscores the diversity of the diesel car market.

Create a bar plot explaining the frequency of car models



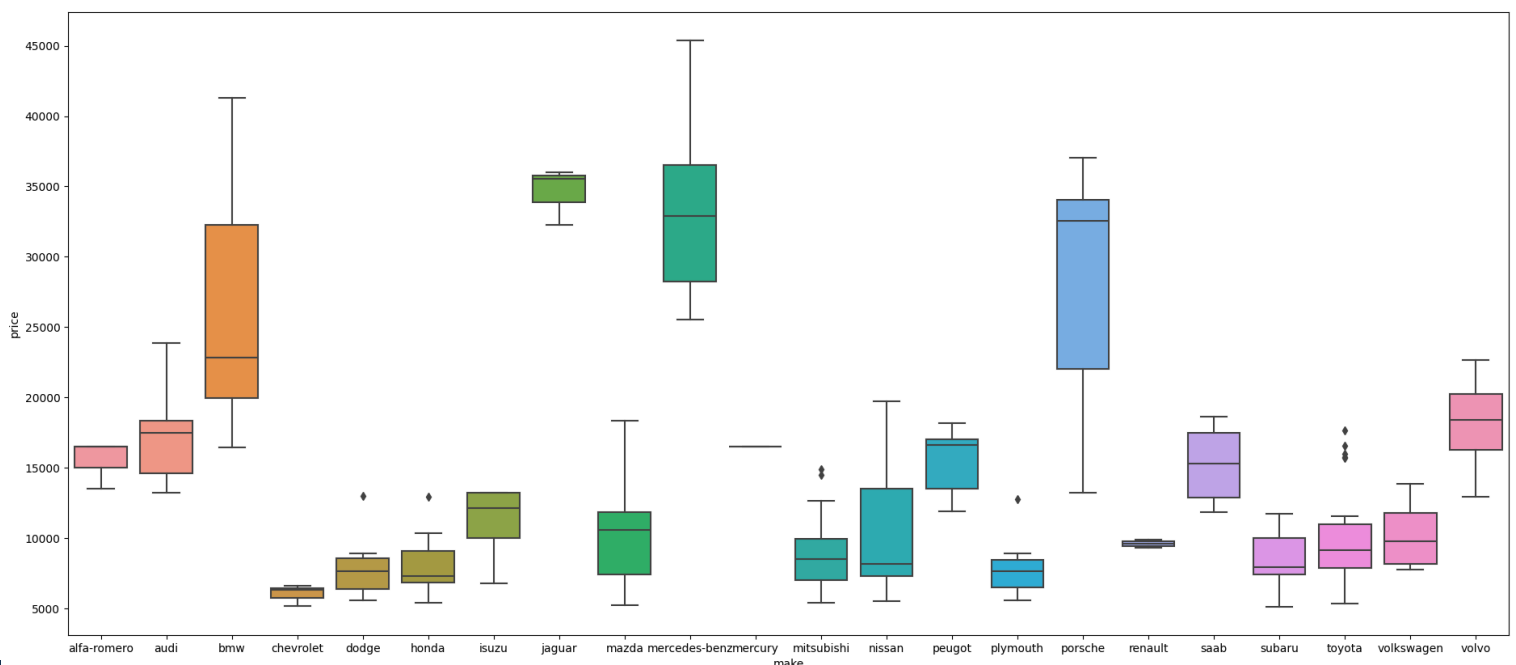
As shown in the bar graph Toyota and Nissan has a hight number of vehicles this might be due to.

Global Presence: Toyota and Nissan are renowned multinational automakers with a global presence. They operate in numerous countries and cater to diverse markets, contributing to a higher overall production volume.

Diverse Model Range: Both Toyota and Nissan offer a wide range of vehicle models, spanning from compact cars to SUVs, trucks, and hybrids. This diversity allows them to address various consumer preferences and market segments, appealing to a broad customer base.

While doge has the lowest

Market Niche: Dodge may focus on specific market niches or vehicle types that have lower overall demand compared to more mainstream models. Specialised or performance-oriented vehicles may have a smaller market share.



Prices of BMW, Mercedes-Benz and Porsche are very high but Porsche offers car models in a lot of price ranges. From premium pricing to pricing comparable to top models of other car manufacturers while BMW, Mercedes-Benz have only premium segment cars.

This comprehensive data cleansing and exploration process set the foundation for further in-depth analysis and storytelling. The cleaned dataset now provides a solid basis for extracting meaningful insights from the world of automobiles.

THIS REPORT WAS WRITTEN BY :Elsy Theledi
