Exploratory Data Analysis on the Automobile Dataset

Report

Introduction

- The dataset contains technical specifications, performance measures, and pricing information for a variety of cars from different manufacturers. It includes details such as body style, engine size, horsepower, fuel efficiency (city and highway MPG), and price, making it useful for exploring relationships between vehicle characteristics and market value.

Data Cleaning

- The way I cleaned the data I removed the "normailized-losses" and the "symboling" columns from the dataset because it was unnecessary to have it there. I also removed all of the rows that were duplicated.

```python
# Drops columns normalized-losses',and 'symboling'
automobiles_df.drop(['normalized-losses','symboling'], inplace = True, axis = 1)

✓ [3] 10ms
```

Remove any duplicate rows

```python
# Code here
automobiles_df.drop_duplicates(inplace=True)

✓ [4] 11ms
```

Missing Data

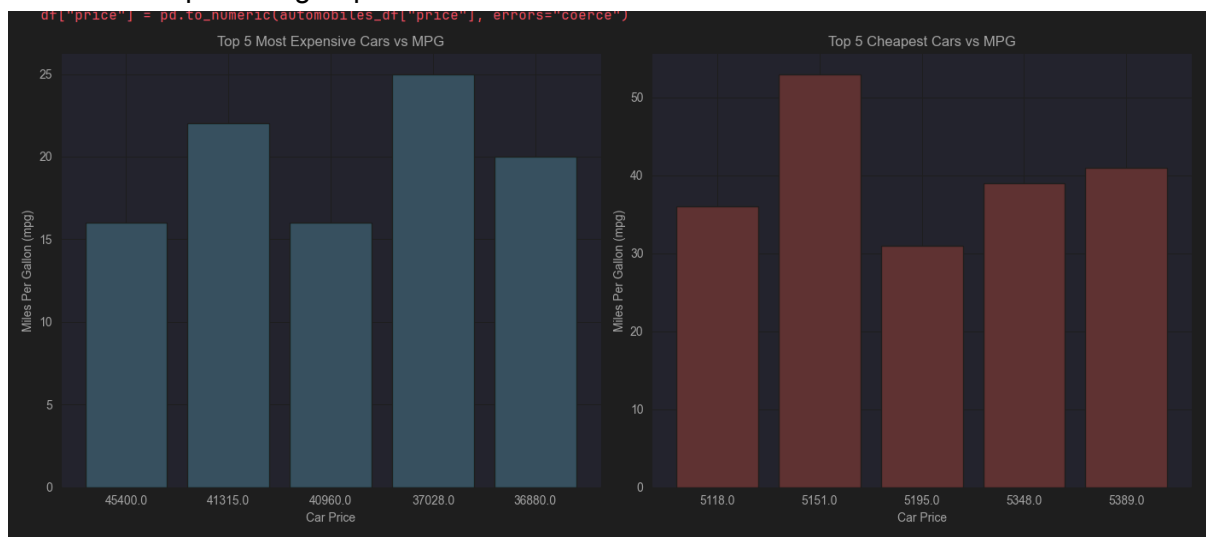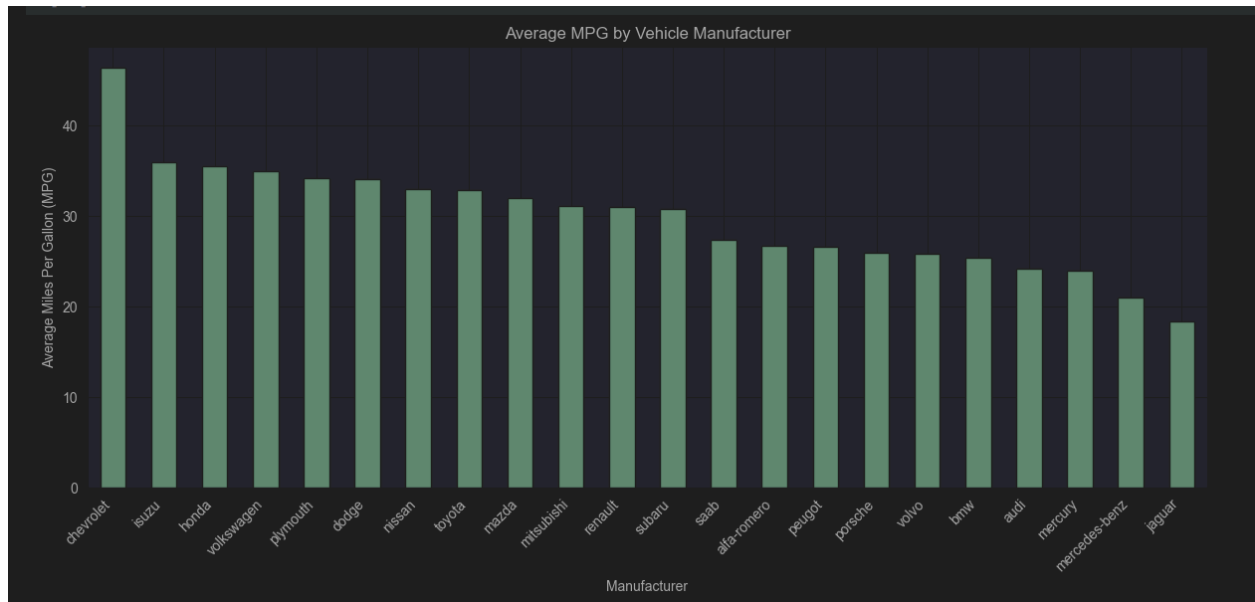- The way I dealt with missing data, and I removed rows that had missing data.

Data stories and visualisations:

- By comparing both graphs, we can clearly see that the cheapest cars have significantly better fuel efficiency than the most expensive ones. In fact, even the lowest fuel-efficient car among the cheapest group has a higher MPG than the most fuel-efficient car in the most expensive group.

- In this graph, Chevrolet is clearly the make that is the most fuel efficient with over 40 MPG. Next is isuzu, honda, and volkswagen with average of 34 MPG.
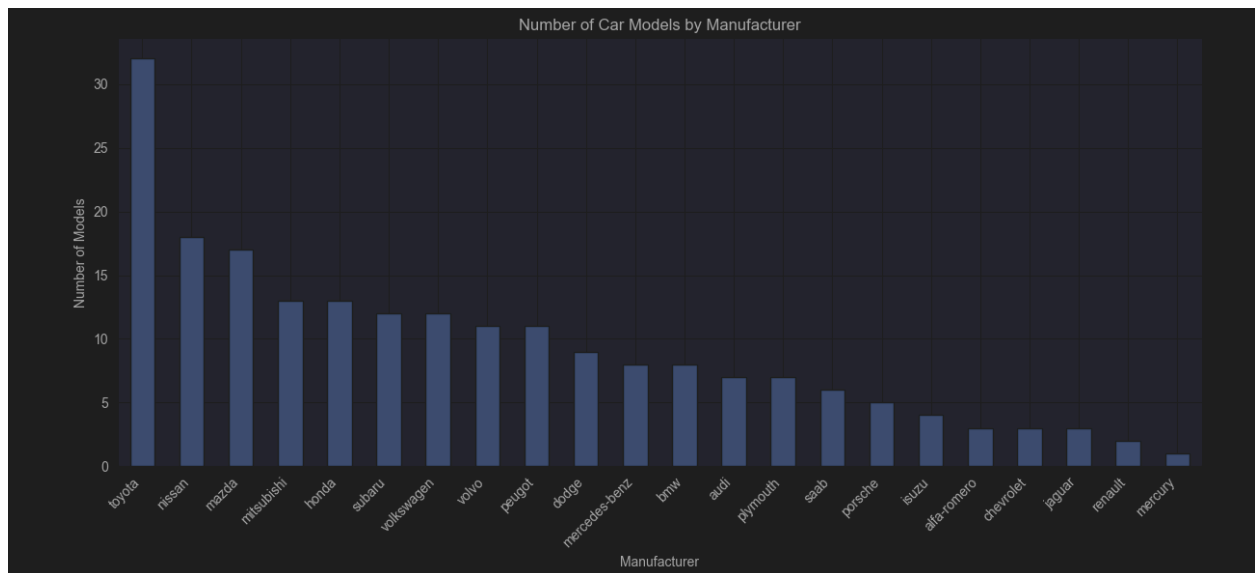
Average MPG by Vehicle Manufacturer

- This chart shows the top cars with the largest engine capacity. The three car companies with the largest engines are Jaguar, Mercedes-Benz, and BMW.

```
1   # Group by manufacturer and get the max engine size per make
2   make_max_engine = automobiles_df.groupby("make")["engine-size"].max().reset_index()
3
4   # Sort by engine size (largest to smallest)
5   make_max_engine_sorted = make_max_engine.sort_values(by="engine-size", ascending=False)
6
7   # Show top 10 manufacturers with largest engines
8   make_max_engine_sorted.head(10)
9
    ✓ [23] 22ms
```
-

	⊕ make	⊞ engine-size
7	jaguar	326
9	mercedes-benz	308
2	bmw	209
15	porsche	203
12	nissan	181
21	volvo	173
19	toyota	171
4	dodge	156
14	plymouth	156
11	mitsubishi	156

- This final graph shows the number of car models by manufacturer. Toyota has by far the most number of models shown in the data with over 30 times, second is nissan and third is mazda.

This report was written by : Leandro Mendez