Exploratory Data Analysis on the Automobile Dataset

Report

Introduction

The automobile dataset contains spec data regarding mulitple automobile manufacturers. There were a few missing values and duplicate entries that had to cleaned. Throughout this documents I will discuss what insights the automabile dataset generated.

Data cleaning

Normalized-losses and symbolizing columns were removed as it did not contain data that was useful to this analysis. There were also multiple duplicate rows that had to be removed using the pandas functions.

Missing data

In identifying any missing fields I used a heatmap to visualize where the missing fields are. I had to remove rows that contained missing data. Fields like price, highway and city mpg and peak-rpm, contained numerical values and were changed to integer data. In those fields I also saw multiple fields containing “?” that also had to be changed to mean() values.

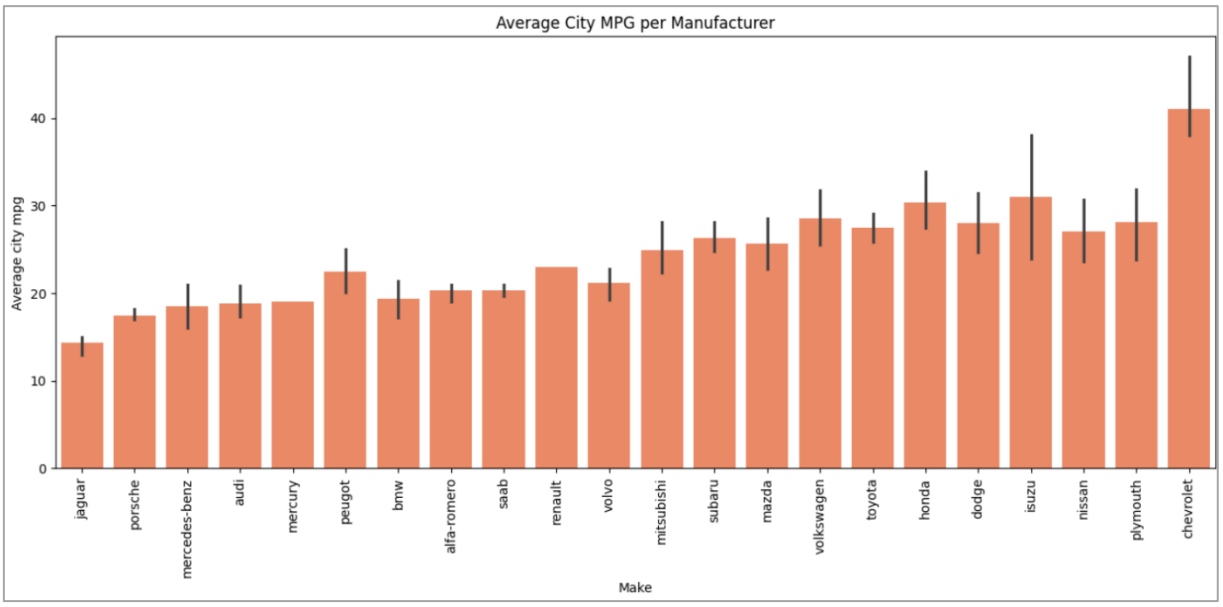
Data stories and visualisations

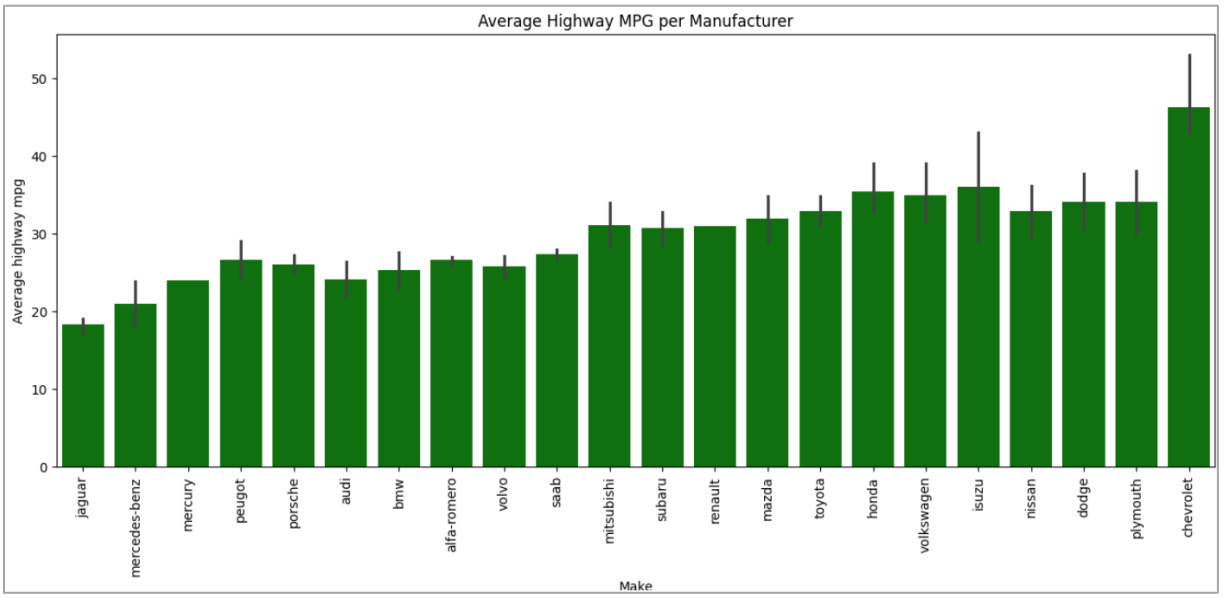
In the following scatterplots I tried to compare miles per gallon (mpg) vs price per vehicle make. The goal was to highlight the top five most expensive vehicles from the data based on mpg and price. I created two separate scatterplots to show city vs highway mpg.

In the above scatterplot the five most expensive cars throughout remain Mercedes-Benz, Jaguar, BMW, Porsche and Audi, when taking into account, price, highway and city mpg.

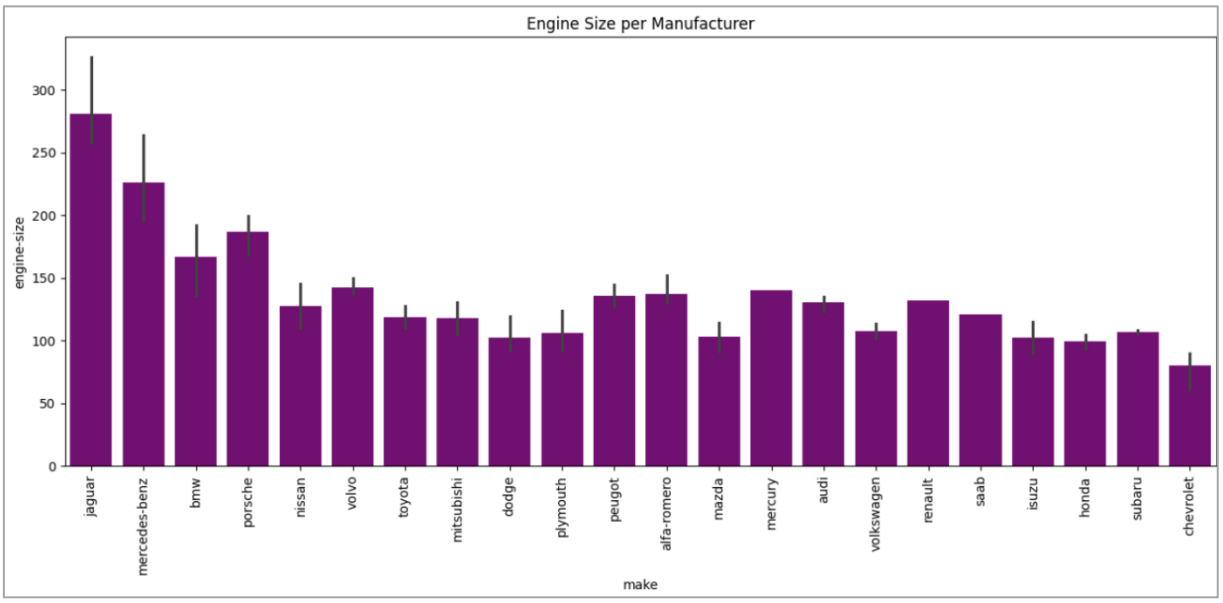
What are the most fuel-efficient vehicles?

In the below barplots I took the make versus the mpg in the city versus the highway. In the city and on the highway Chevrolet, Nissan, Isuzu and Plymouth stay the most fuel efficient vehicles throughout.

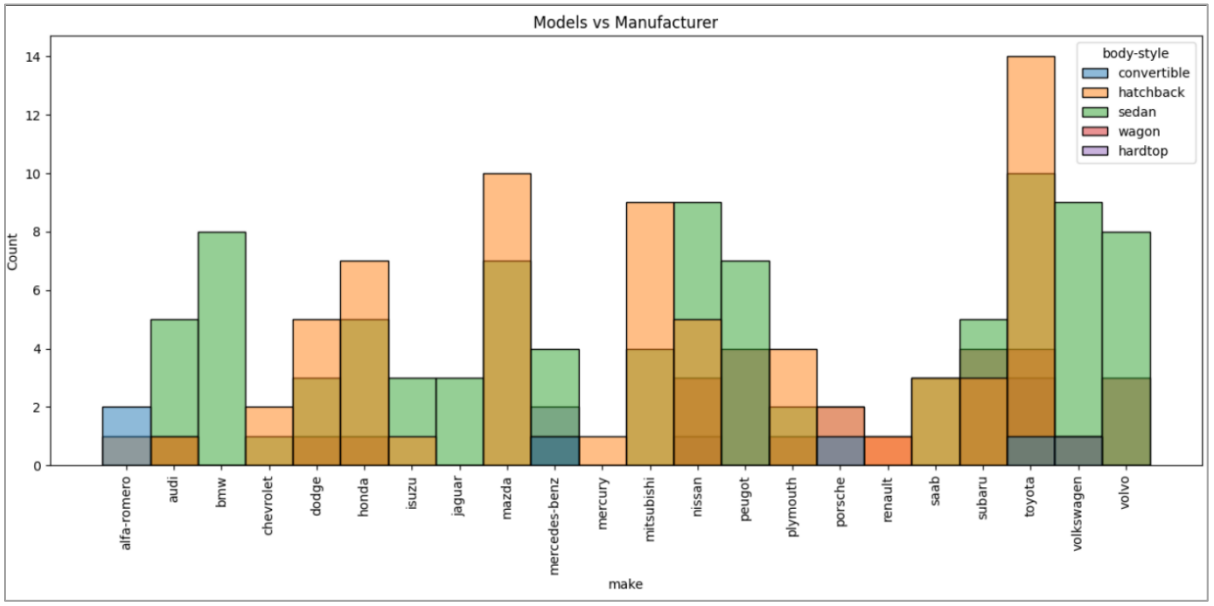




Which vehicles have the largest engine capacity?
The below barplot indicates the engine-size vs the vehicle manufacturer (make). Jaguar, followed by Mercedes-Benz and Porshe has the greatest engine-size compared to all the other vehicles make in the dataset.



Which vehicle manufacturer has the most car models in the dataset?
The histogram below displays the make of the vehicle vs the number of vehicles. I added a legend on the top right to show what body-styles a vehicle makes have. From the below figure we can see that "Toyota" has the most cars and the most body styles compared to all the other vehicle makes.



This report was written by : Carli Burger