Hyperiondev

Exploratory Data Analysis on the Automobile Data Set

Visit our website

Introduction

The data set contains data from automobiles, including their make, horsepower, body style, miles per gallon, and price to name a few.

DATA CLEANING

I removed any duplicate rows, unnecessary or redundant columns, and I removed rows that had missing values in particular columns in the data set.

DATA STORIES AND VISUALIZATIONS

### CAR MAKE WITH THE HIGHEST HORSEPOWER

I started off the analysis by looking at which car make had the highest horsepower.

While there are other cars in the original dataset that had horsepower higher than 200, unfortunately they were removed due to missing values. Interestingly, Volvo having two entries in the top ten, especially when they are considered to be one of the car manufacturers that produce the safest cars, we will look into that further later on.
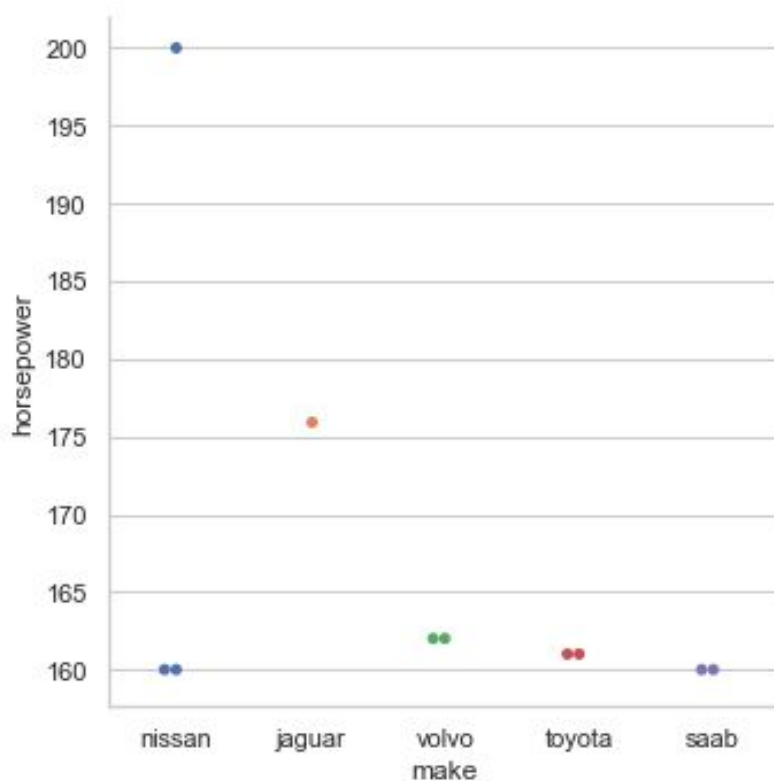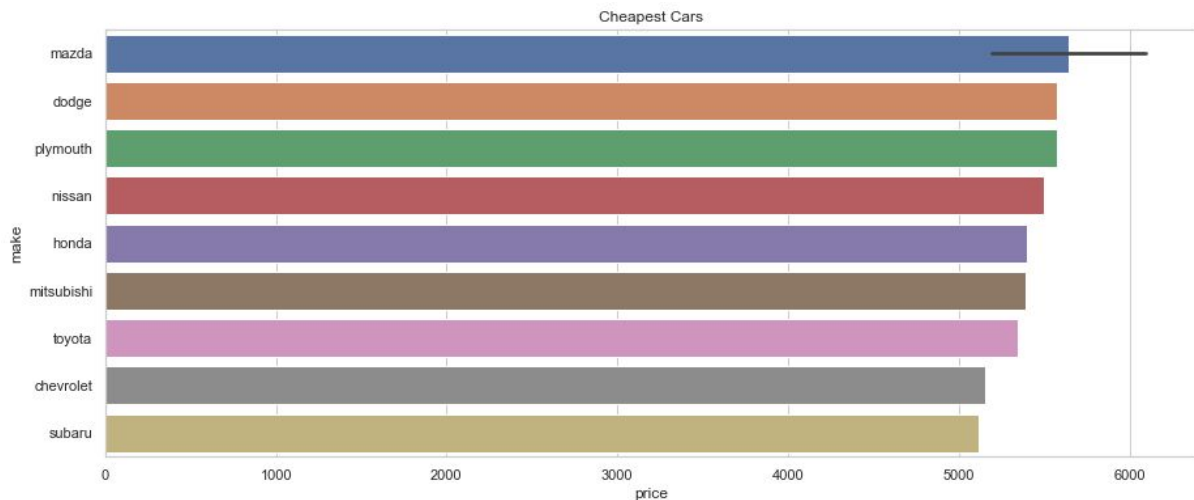
MOST EXPENSIVE CARS

There were no surprises here with the most expensive cars, except perhaps for Porsche, again with the rows that were removed, unfortunately some of the entries just did not make it into the analysis.
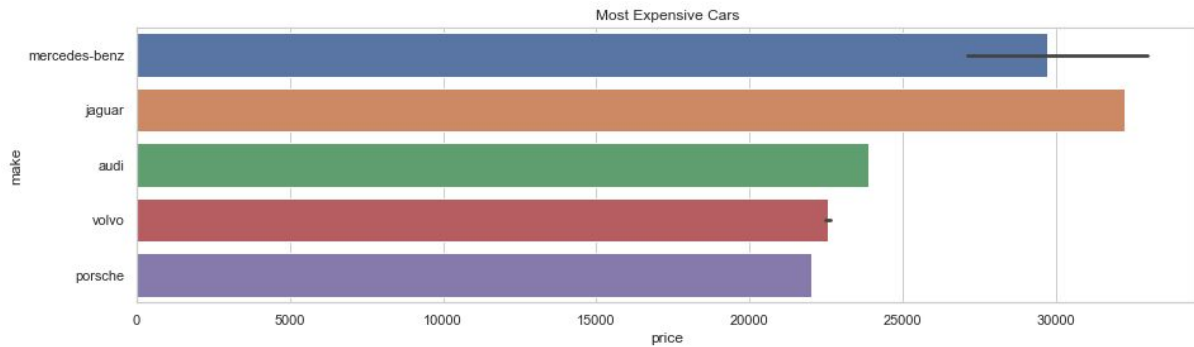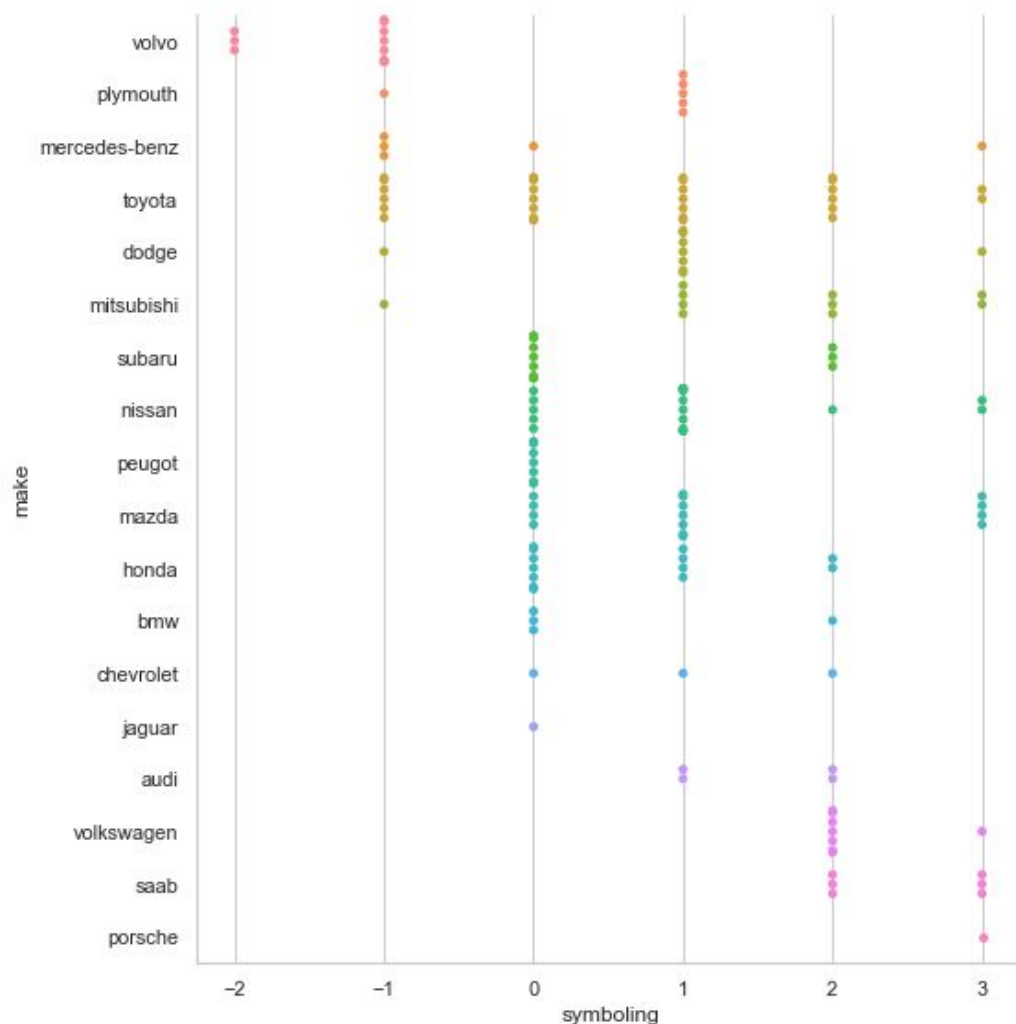

CHEAPEST CARS

Here we have the cheapest cars in the data set, seemingly much more affordable than those on the other end of the spectrum.

SAFEST CARS ACCORDING TO SYMBOLING

I then looked at the safety of the cars according to their designated symboling.

Symboling represents a car's risk factor on a scale or -3 to 3, where -3 is very safe, and 3 is very risky. Let's use this rating to see which cars are the safest, and which are not safe at all.

According to this visualization, Volvo produces the safest cars, also being the only manufacturer that achieved a symboling factor less than -1, and then Porsche making the riskiest cars.

From what we have seen so far, Volvo has been in the top listings for horsepower and safety, but also one of the most expensive. If safety is your top concern when purchasing a vehicle, then Volvo is probably your best choice.

THIS REPORT WAS WRITTEN BY : CALVIN LIANG