EDA Used Cars Dataset

Presented by: Rehab AL zaidi

Contents

Introduction

When purchasing a car, there are many factors that go into making that decision. It’s important that the car fits into the customer’s budget as well as their lifestyle.

While purchasing a pre-owned vehicle is often a more economical option, the buyer may also consider the car’s mileage, condition, manufacturer, etc. Using a dataset containing information about vehicles listed on Craigslist.

Dataset Description

Feature	Description
id	Vehicle Identification Number
region_url	Link to region page
region	Craigslist region in which this listing was posted
url	Link to listing
price	Price of vehicle
year	Year of vehicle
manufacturer	Manufacturer of vehicle
model	Model of vehicle

Dataset Description

Feature	Description
cylinders	Number of cylinders of vehicle
size	Size of vehicle
type	Vehicle type
paint_color	Color of vehicle
image_url	Link to image of vehicle
description	Listing description provided by owner
county	County of Vehicle
state	State of Vehicle
lat	Latitude of vehicle (not precise but very close)

Dataset Description

Feature	Description
drive	Drive of vehicle
condition	Condition of vehicle
fuel	taken by vehicle
VIN	Vehicle Identification Number
odometer	Vehicle type
paint_color	Color of vehicle
image_url	Link to image of vehicle
long	Longitude of vehicle (not precise but very close)
posting_date	Date posting of Vehicle

- The dataset contains 26 features, 426880 observation.

Preprocessing

01 Read Data

02 Check if there is missing values

03 Check if there is outliers

Clean Data

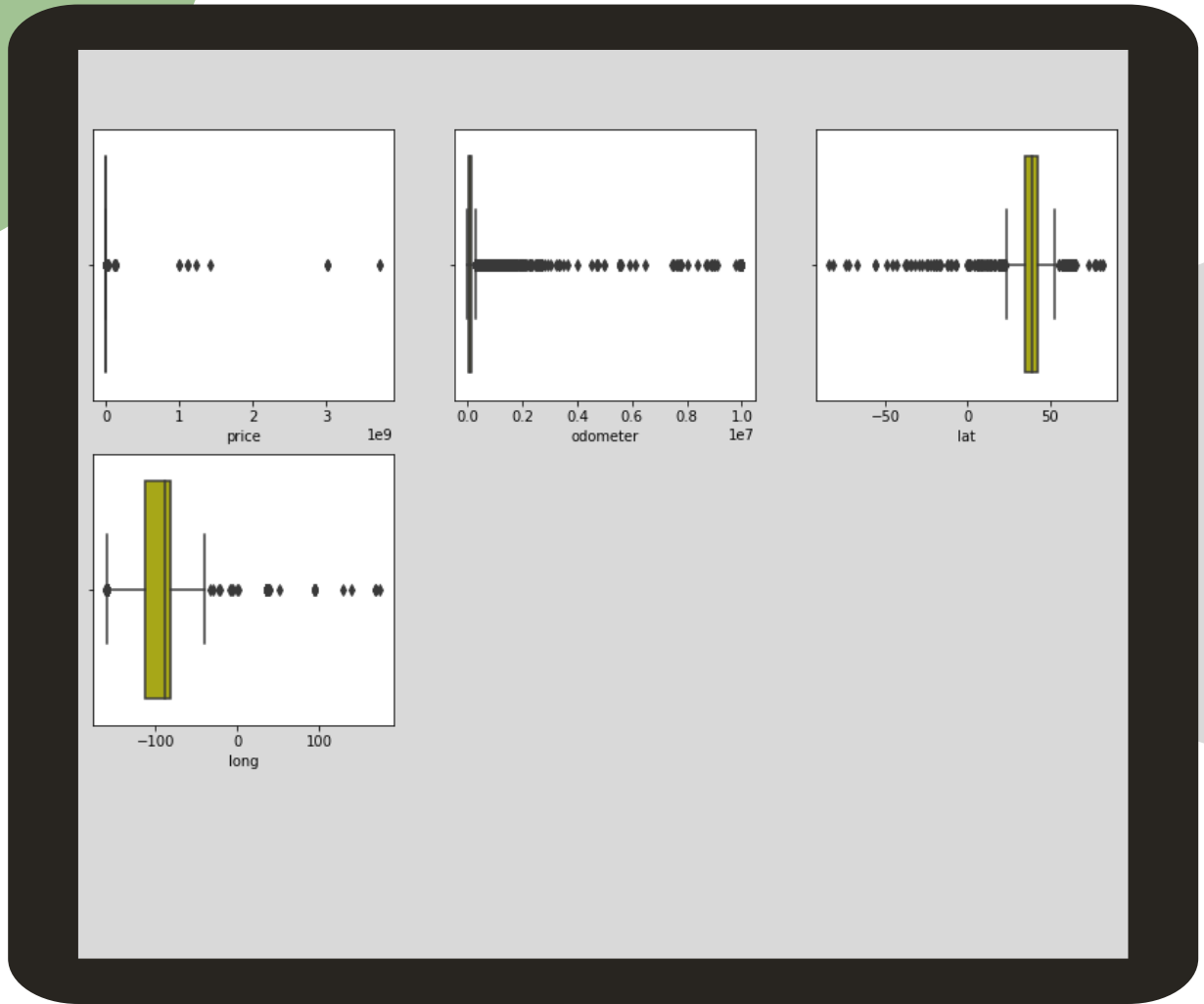
After reading Data then check missing values, and outliers.

Handling missing data by :

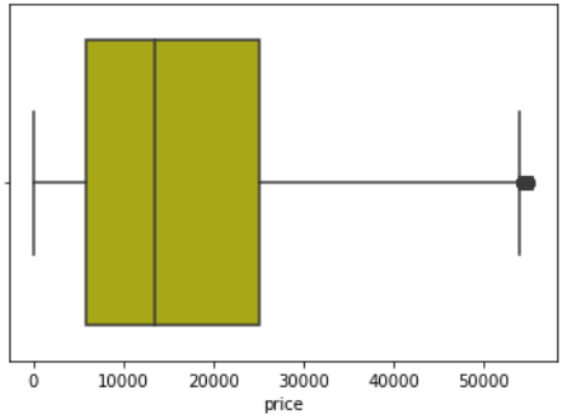
- drop 5 columns we didn’t need it at all.
- separate data to Numerical and Categorical.

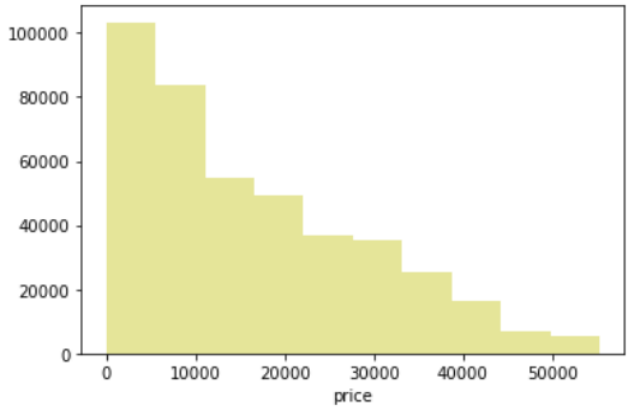
fill value in numerical with mean , fill value in categorical with mode

outlier detection
in numerical



After dropping outliers, let’s check the boxplot and histogram of our data.

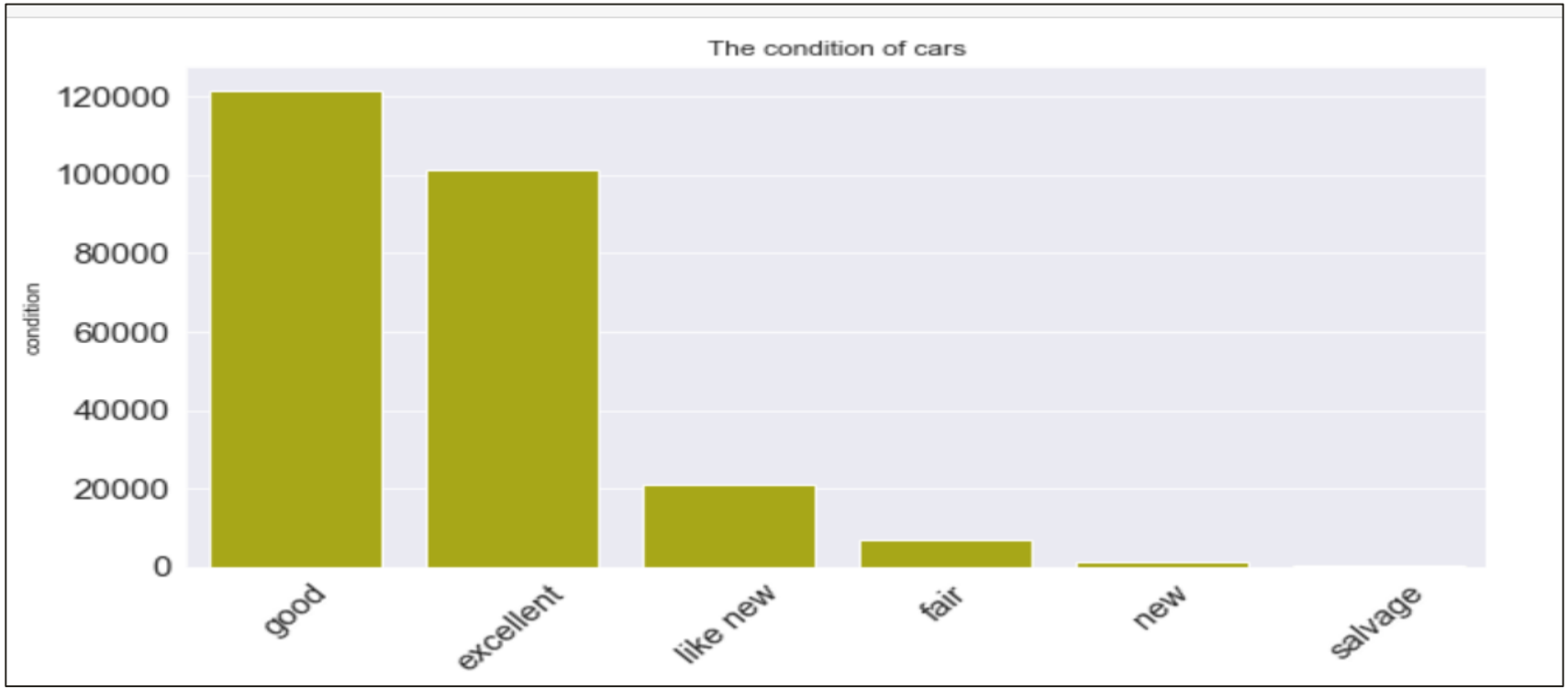




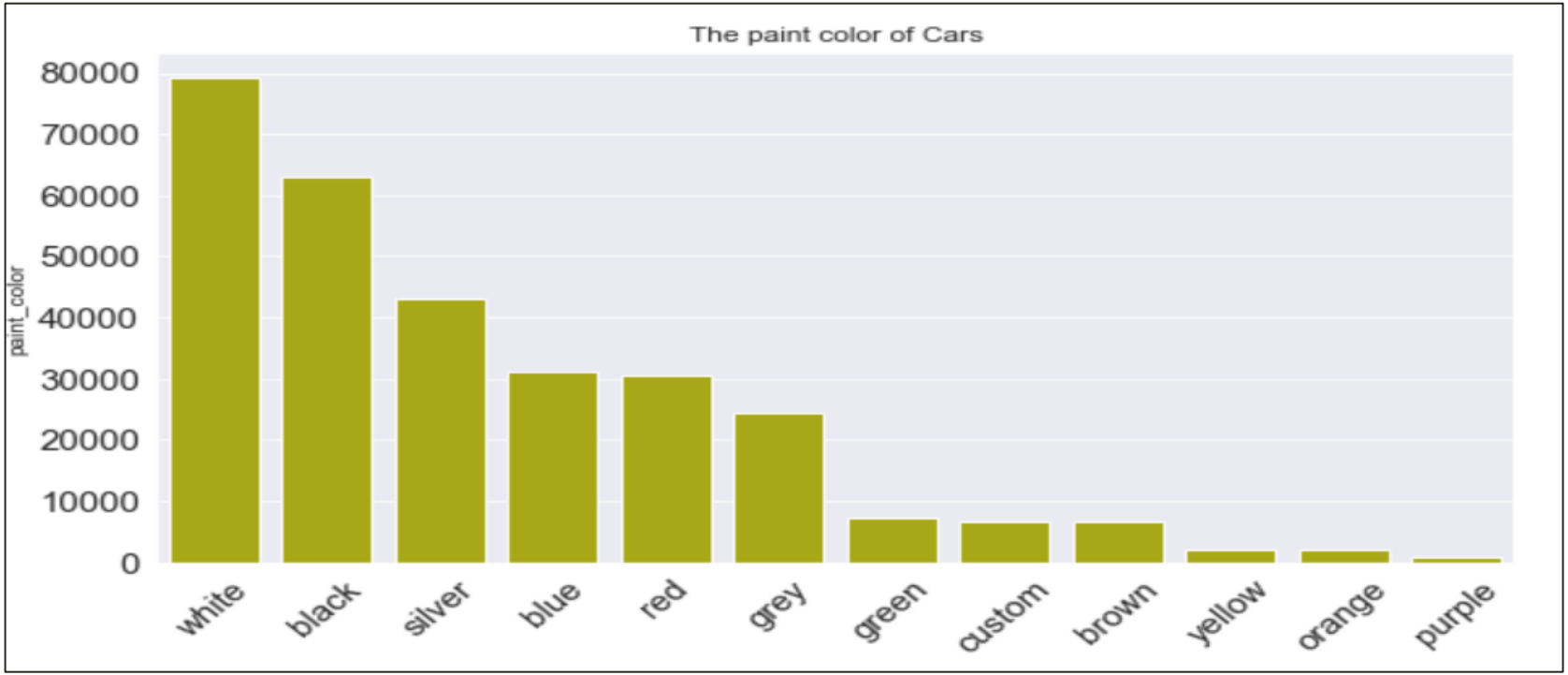
Results

For Categorical:

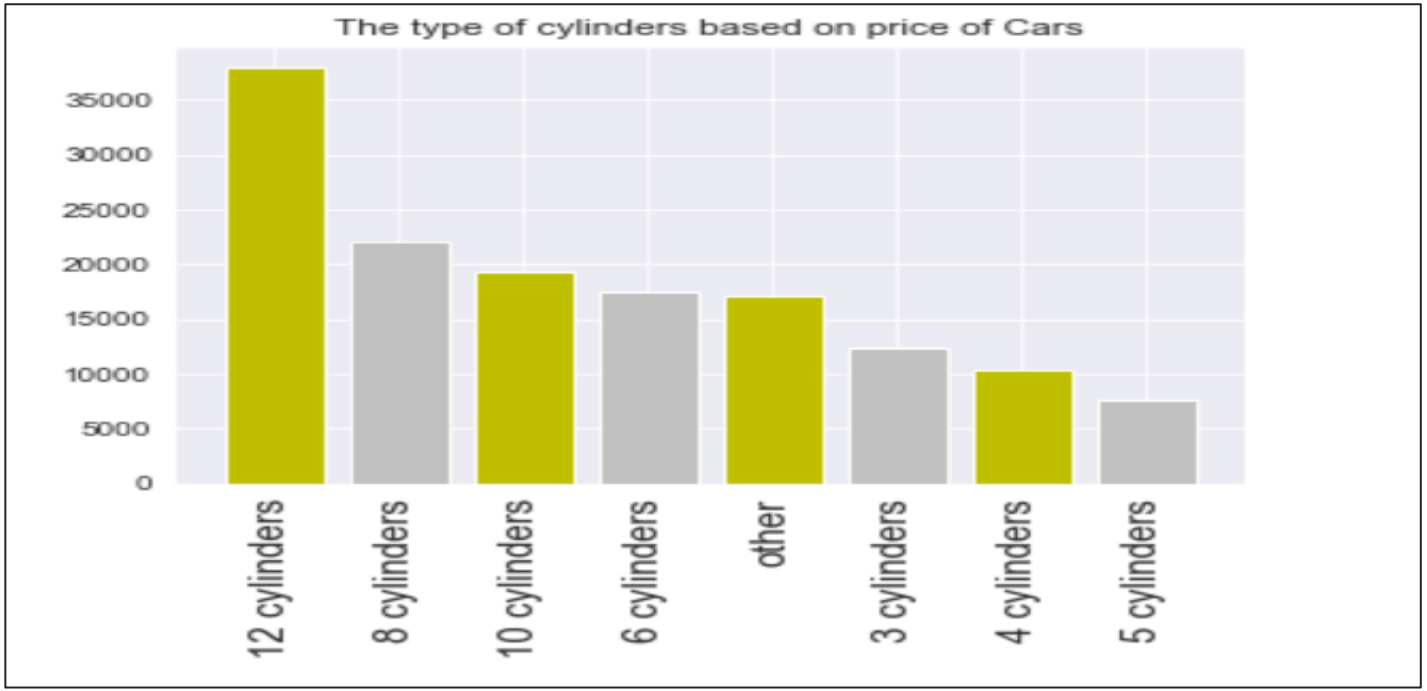
1- The condition of cars.



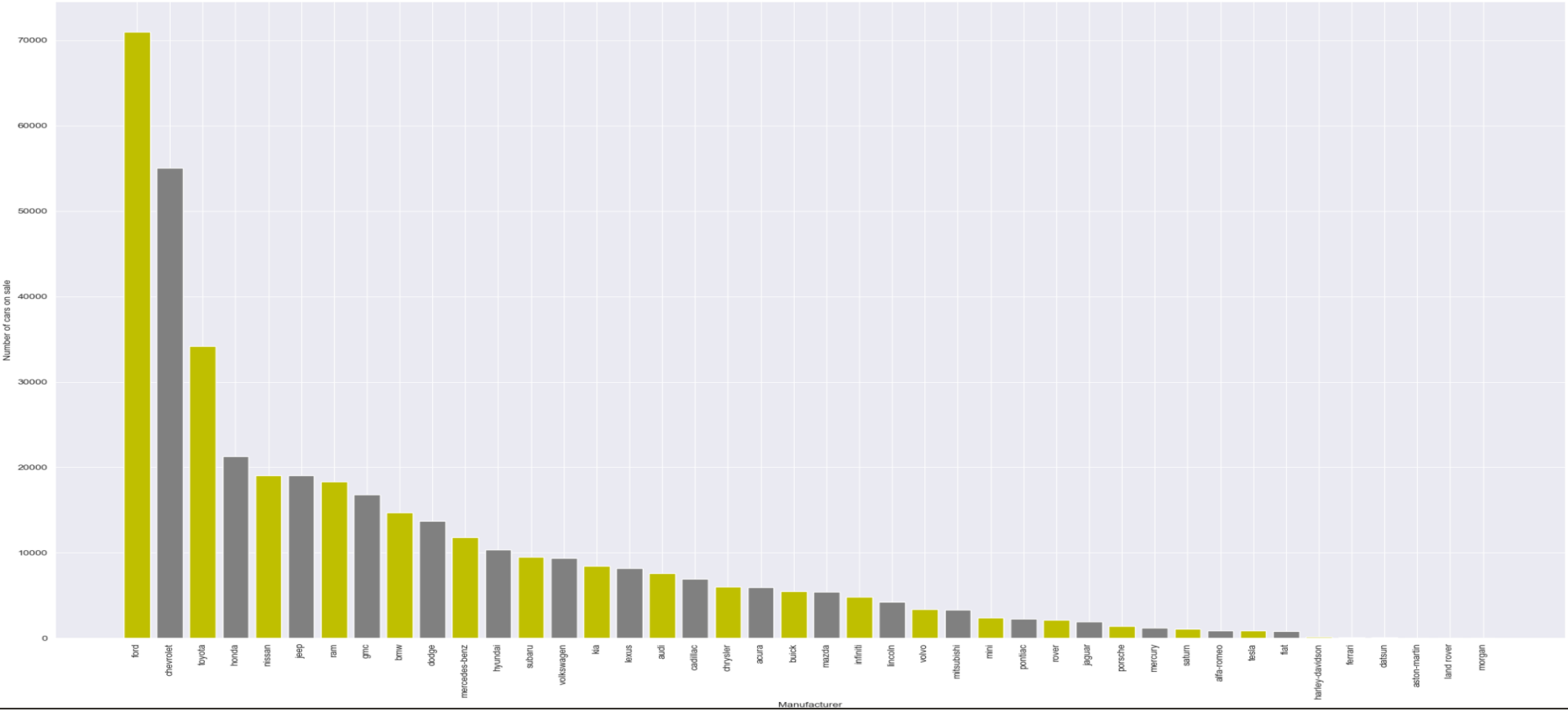
2- The paint color of cars.



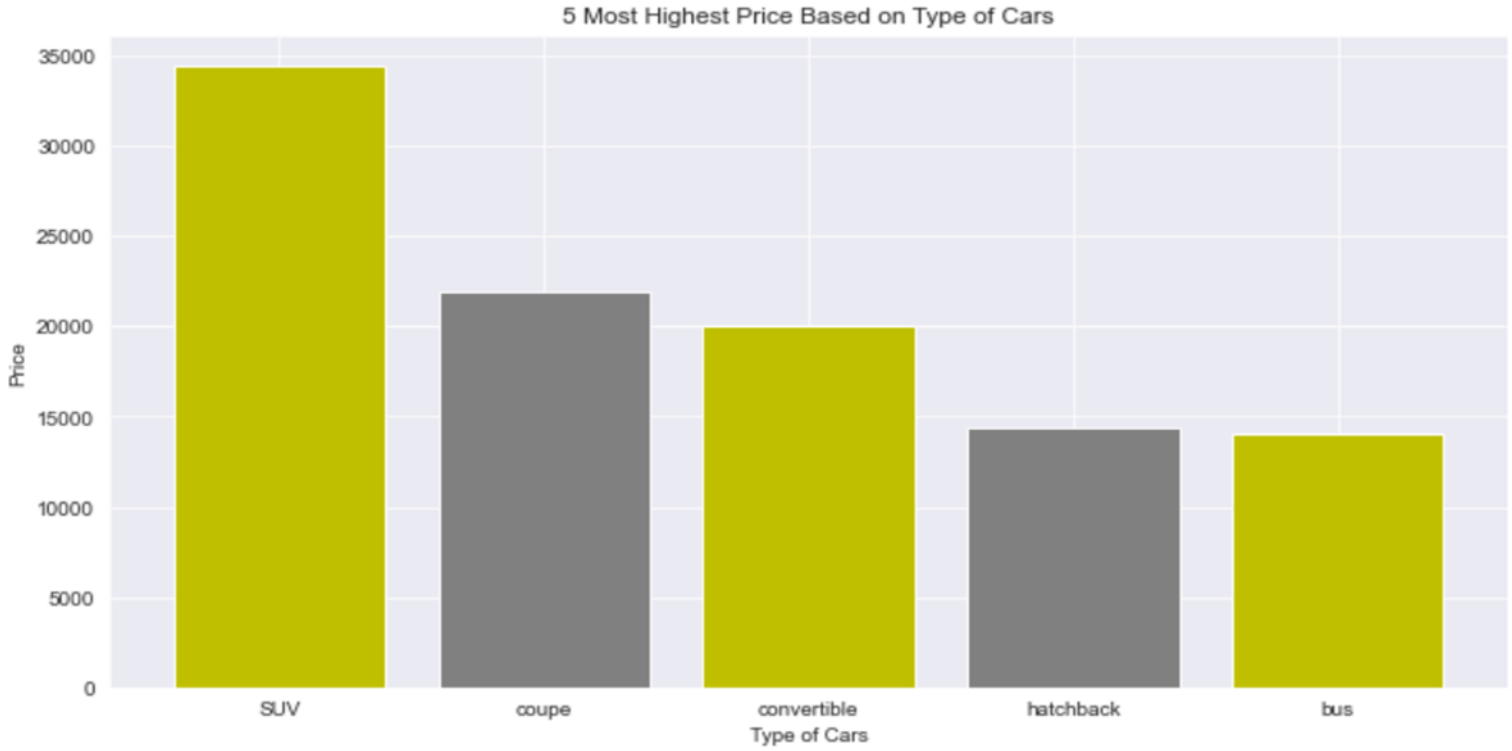
3- The cylinders of cars.



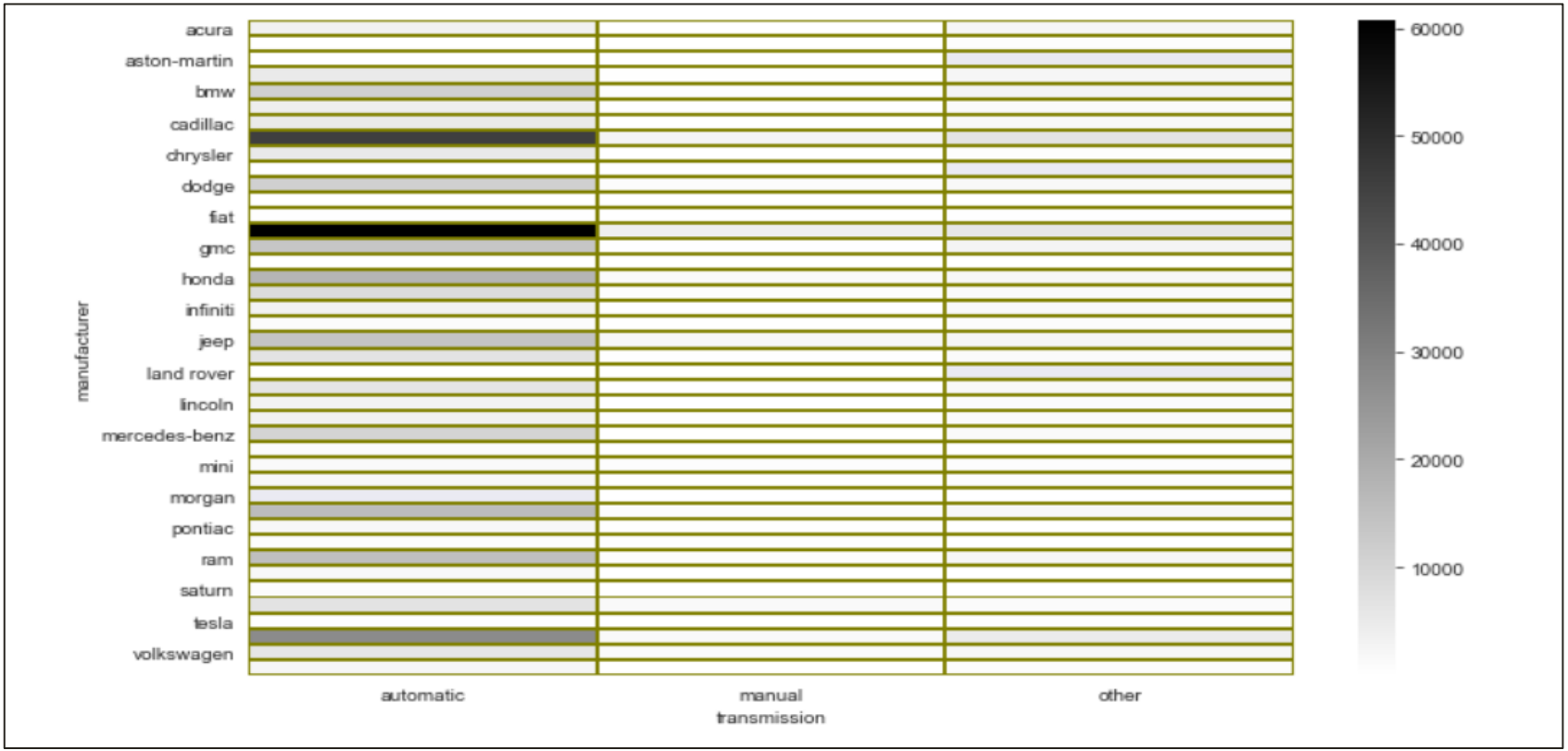
1- The most common manufacturers in dataset based on number of cars?



2-The Highest Price Based on Type Car?

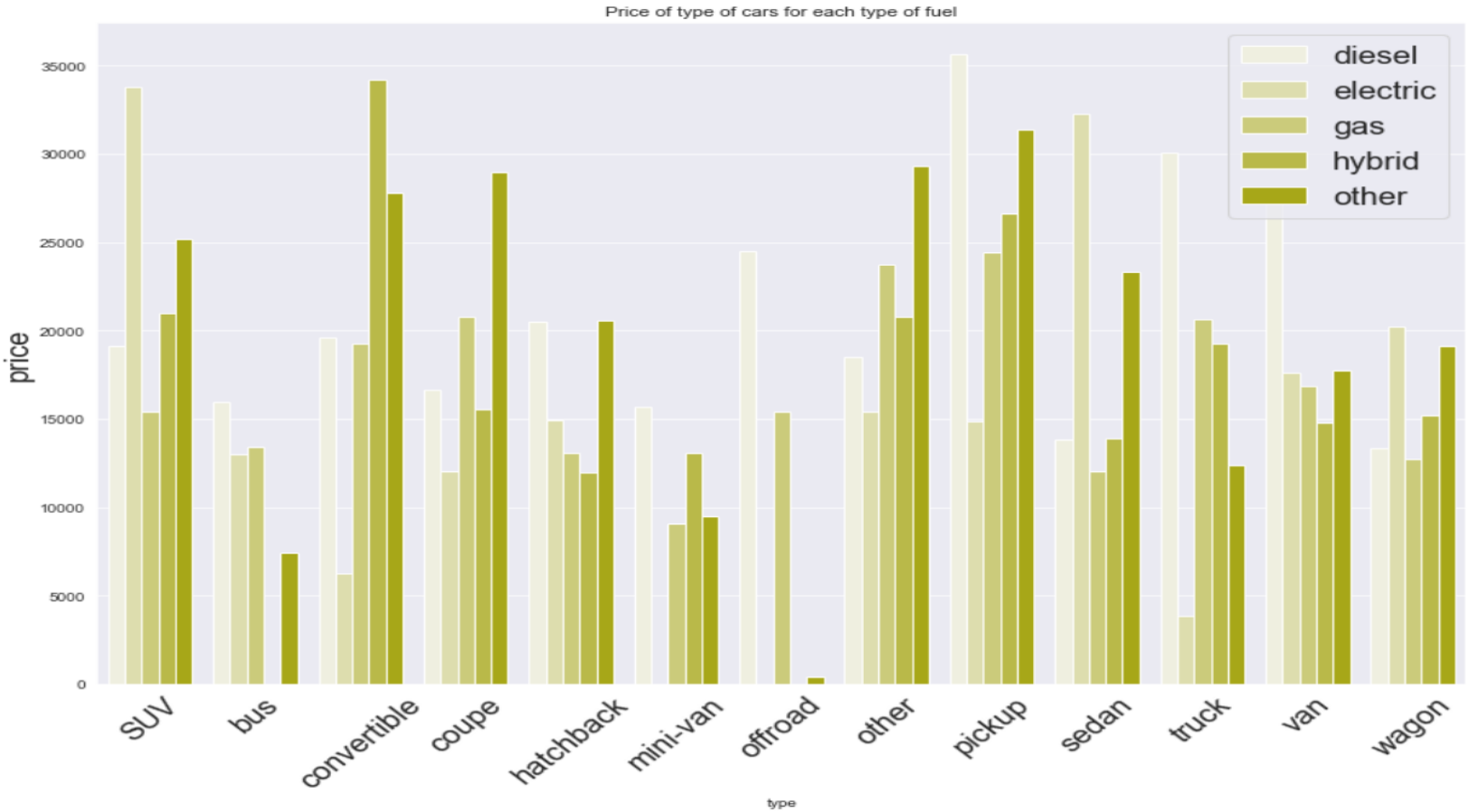


3-The most transmission type used based on manufacturer?



It was no surprise that automatic used cars dominated the used cars market.

- Price of type of cars for each type of fuel



CONCLUSIONS

If you want to buy a certain type of used car, chart above could help you to decide what manufacturer to go for or type of cylinders .

So, those are some information we can obtain by doing this EDA.

Thanks...