OYC company

for selling cars

By:

Shahad Almubki

Nada Alqabbani

Asma Alsulami

❑ Backstory:

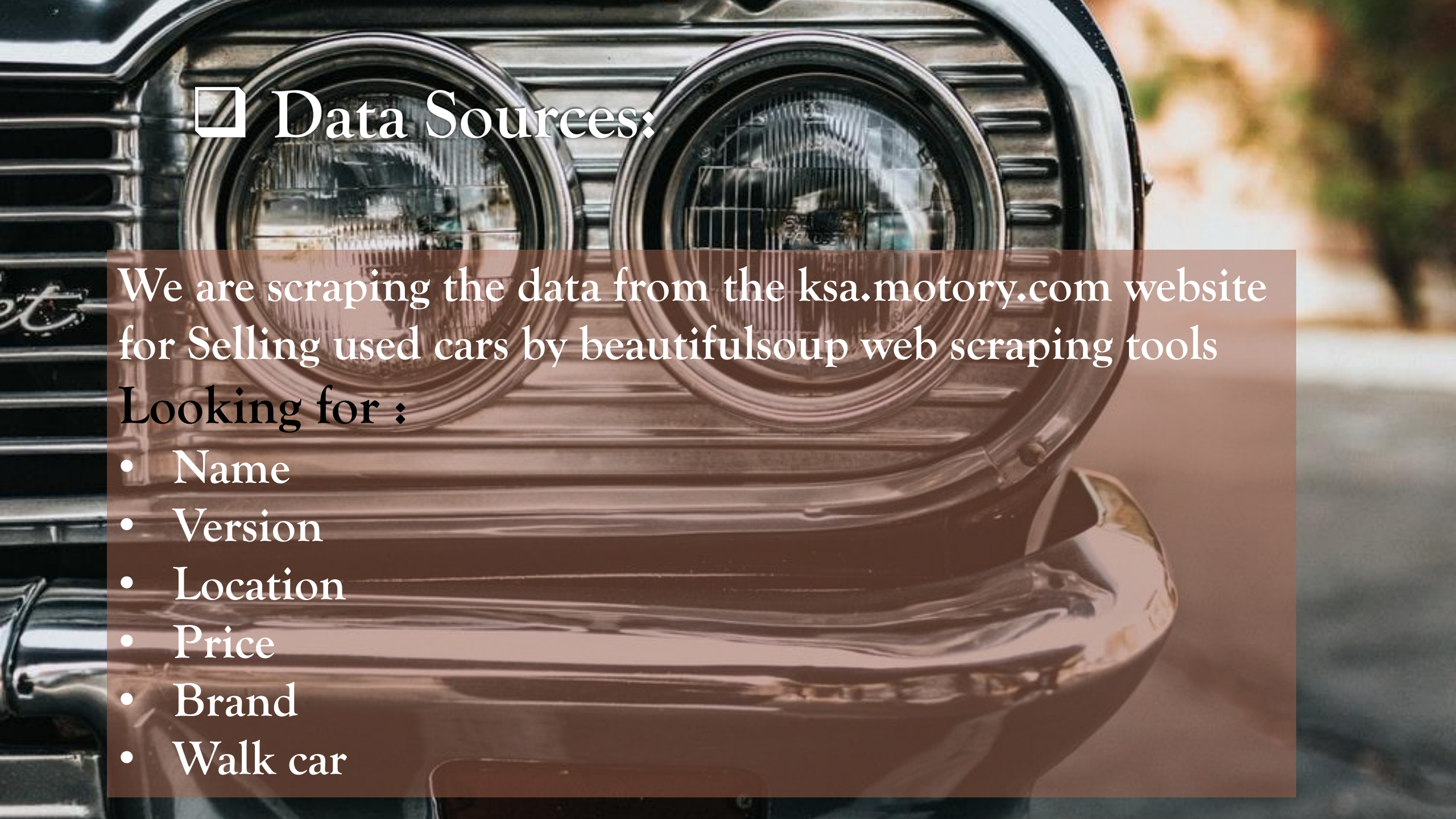
We are an OYC company that serves customers to buy the cars used they want with the required features. We work to meet all customer requirements and provide them according to his specific budget and desired features.

Tools:

- Beautifulsoup
- Pandas
- Matplot
- Sklearn.linear and sklearn.preprocessing
- Matplotlib library
- Requests
- Statsmodels
- Seaborn
- re

❑ Performing the Regression project:

- Web Scraping
- EDA
- Features Engineering
- Modeling



❑ Data Sources:

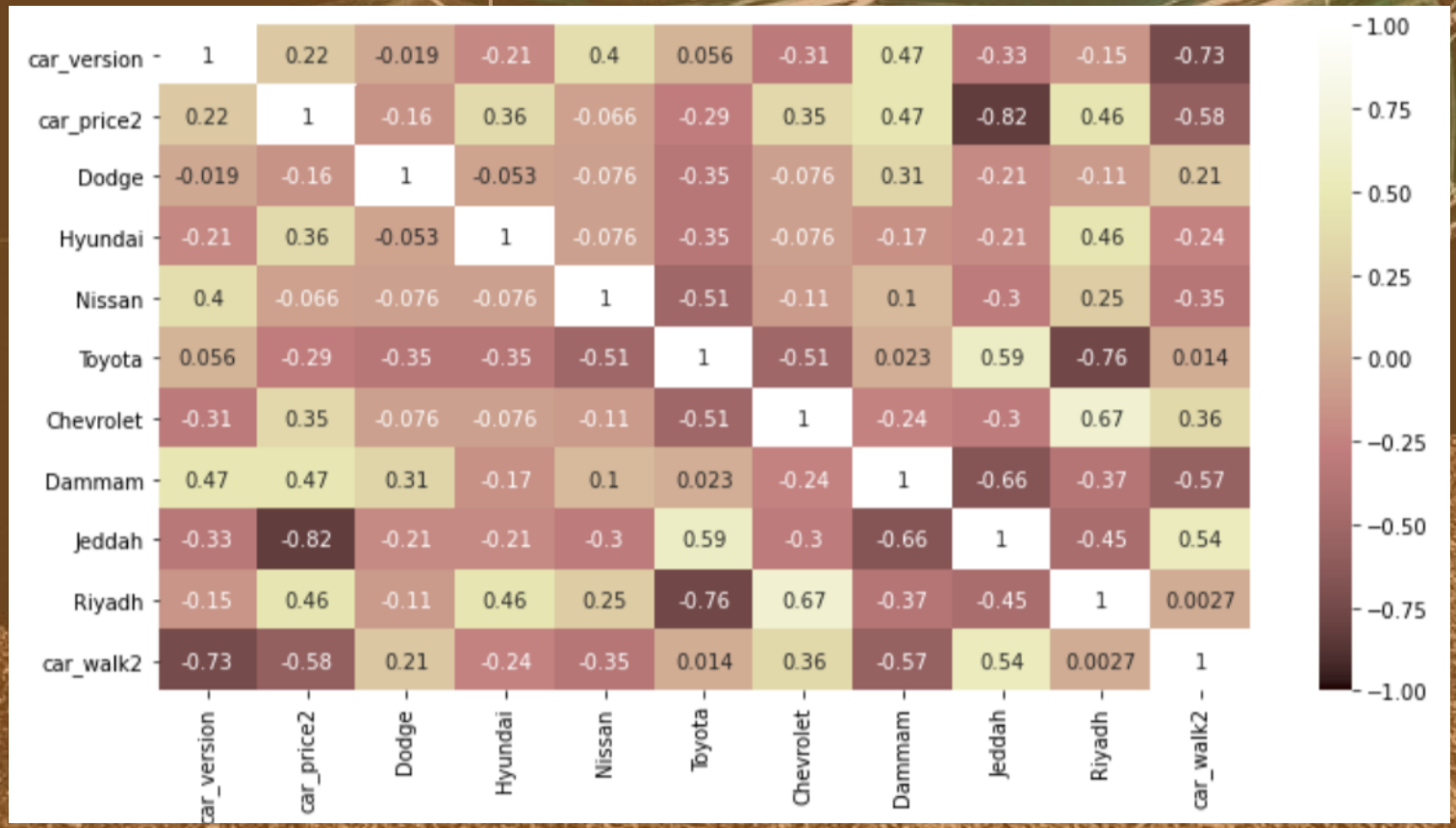
We are scraping the data from the ksa.motory.com website for Selling used cars by beautifulsoup web scraping tools

Looking for :

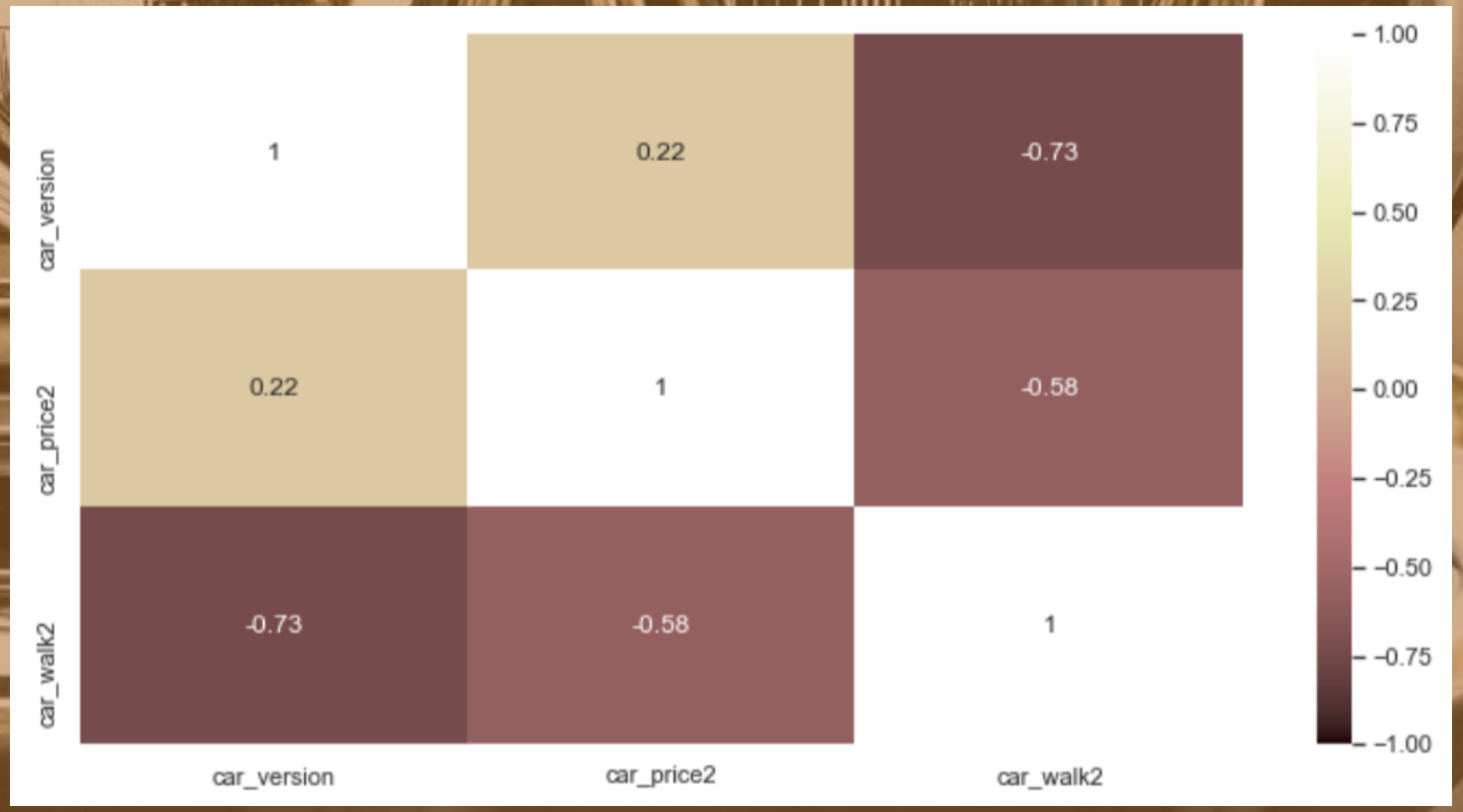
- Name
- Version
- Location
- Price
- Brand
- Walk car

The Results:

❑ Heatmap

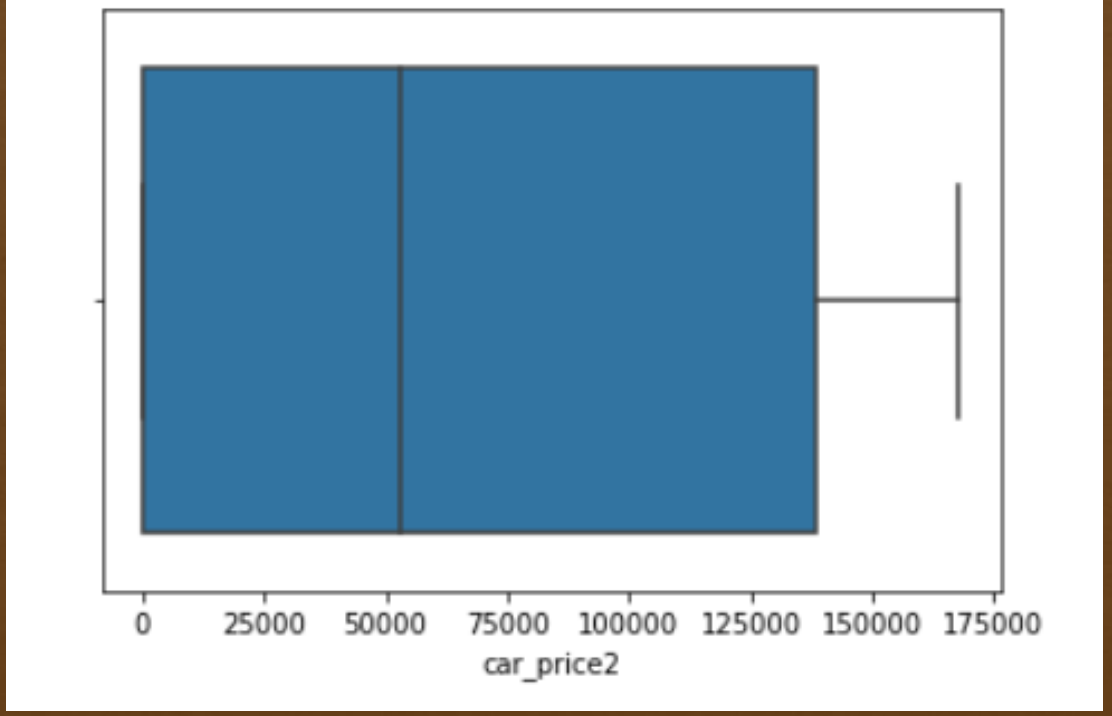

car_version 1 0.22 -0.019 -0.21 0.4 0.056 -0.31 0.47 -0.33 -0.15 -0.73
car_price2 0.22 1 -0.16 0.36 -0.066 -0.29 0.35 0.47 -0.82 0.46 -0.58
Dodge -0.019 -0.16 1 -0.053 -0.076 -0.35 -0.076 0.31 -0.21 -0.11 0.21
Hyundai -0.21 0.36 -0.053 1 -0.076 -0.35 -0.076 -0.17 -0.21 0.46 -0.24
Nissan 0.4 -0.066 -0.076 -0.076 1 -0.51 -0.11 0.1 -0.3 0.25 -0.35
Toyota 0.056 -0.29 -0.35 -0.35 -0.51 1 -0.51 0.023 0.59 -0.76 0.014
Chevrolet -0.31 0.35 -0.076 -0.076 -0.11 -0.51 1 -0.24 -0.3 0.67 0.36
Dammam 0.47 0.47 0.31 -0.17 0.1 0.023 -0.24 1 -0.66 -0.37 -0.57
Jeddah -0.33 -0.82 -0.21 -0.21 -0.3 0.59 -0.3 -0.66 1 -0.45 0.54
Riyadh -0.15 0.46 -0.11 0.46 0.25 -0.76 0.67 -0.37 -0.45 1 0.0027
car_walk2 -0.73 -0.58 0.21 -0.24 -0.35 0.014 0.36 -0.57 0.54 0.0027 1
car_version
car_price2
Dodge
Hyundai
Nissan
Toyota
Chevrolet
Dammam
Jeddah
Riyadh
car_walk2
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Heatmap "After Selecting the Feature"



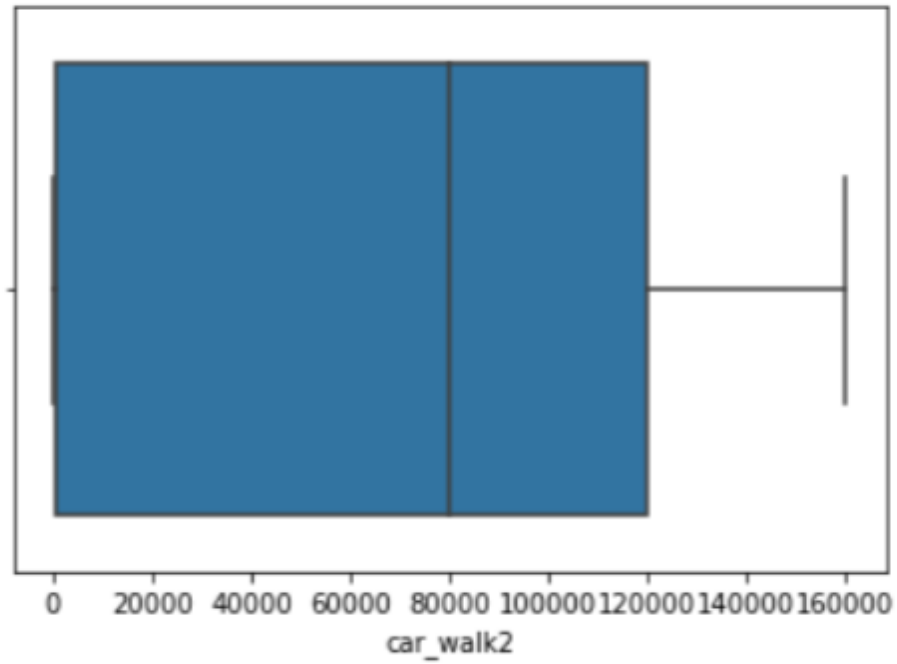
- BoxPlot: “Price”

```
1 sns.boxplot(x = cars_df1['car_price2']);
```



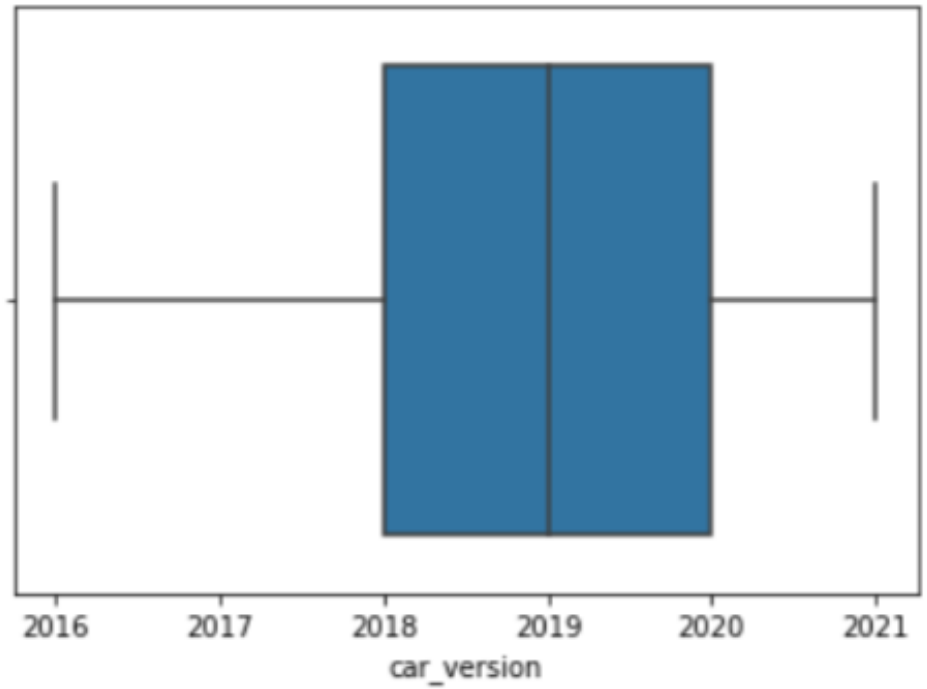
- BoxPlot: “Walk”

```
1 sns.boxplot(x = cars_df1['car_walk2']);
```

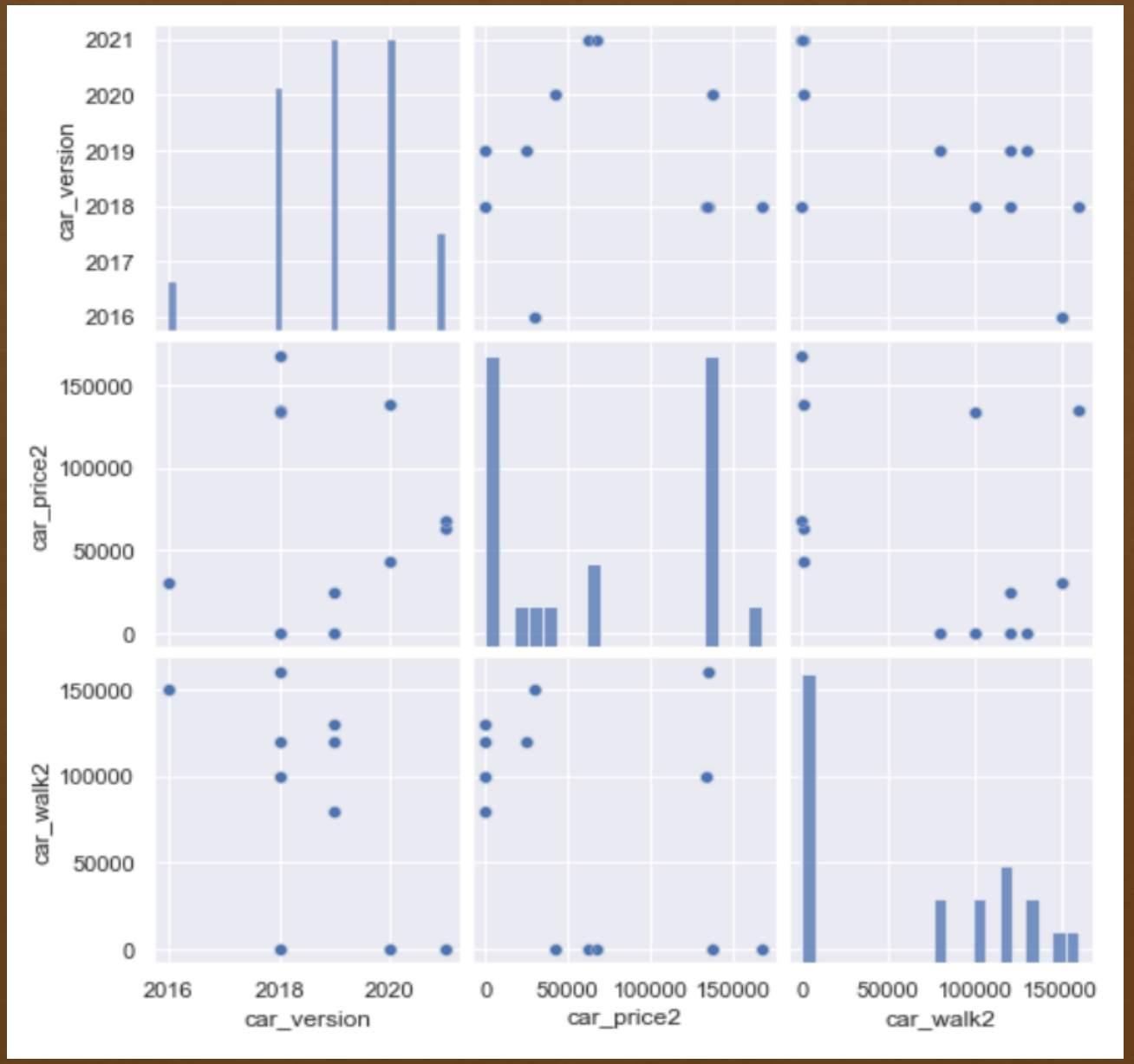


- BoxPlot: “Version”

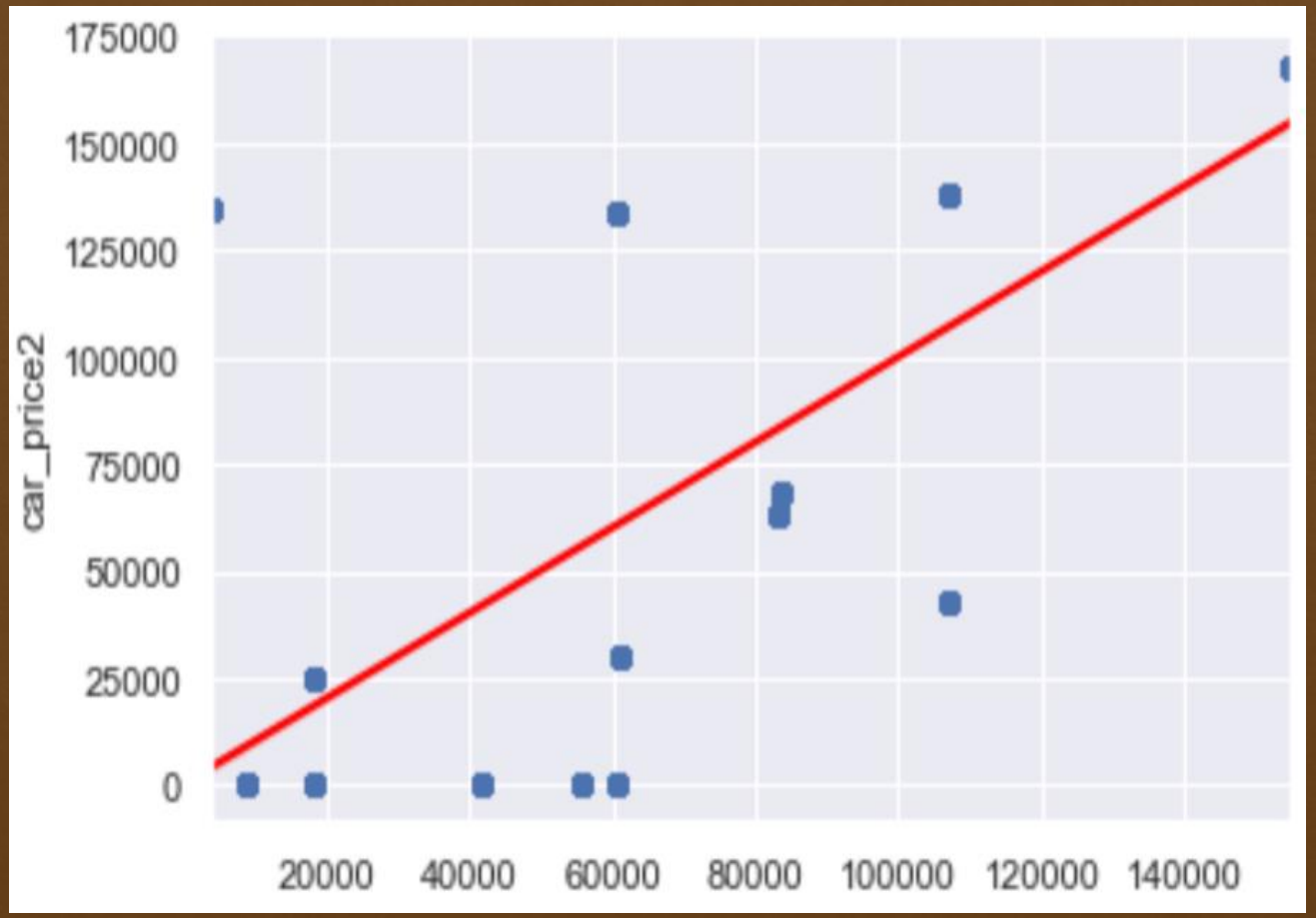
```
1 sns.boxplot(x = cars_df1['car_version']);
```



- PairPlot:


car_version
car_price2
car_walk2
2021
2020
2019
2018
2017
2016
150000
100000
50000
0
2016
2018
2020
0
50000
100000
150000

- RegPlot:


175000
150000
125000
100000
75000
50000
25000
0
car_price2
20000
40000
60000
80000
100000
120000
140000

❑ The Score of the Train and Validation

Different matrix :
The number of scores is few

```
Train score 0.43047057477992623
Val score 0.4325509645490544
MAE: 37853.48437102655
MSE: 2247757725.1515584
RMSE: 47410.52335876032
```

```
Linear Regression test R^2: 0.437

CV for Linear model score was: 0.42800439874495544

Ridge Regression test R^2: 0.437

CV for Ridge model was score was: 0.4280044066203529
```

❑ Conclusion and Future Work :

- We suggest that we collect features from another database because the current features were not enough for prediction, we need to provide new features to the dataset so that we can predict and improve a good model.

- add new columns such as (fuel type, engine size, horsepower, ...)


Thank you
Group 10