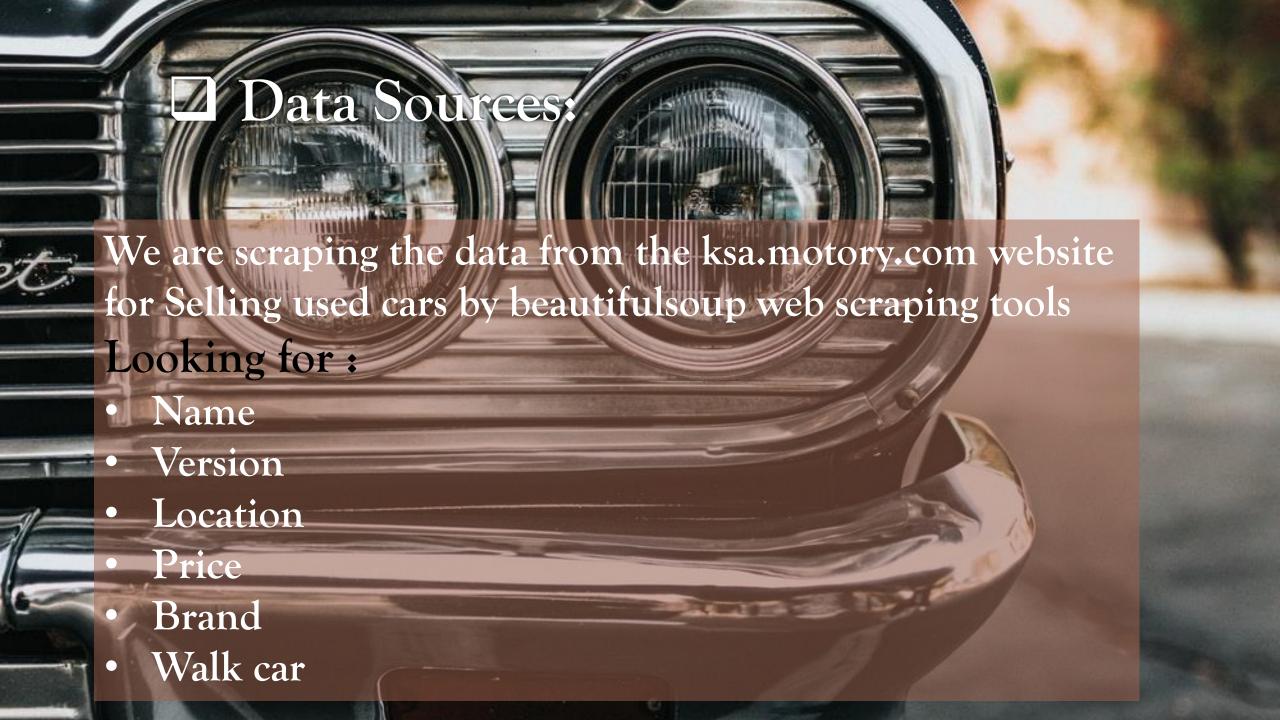






- ♦ Beautifulsoup
- ♦ Pandas
- Sklearn.linear and sklearn.preprocessing
- ♦ Matplotlib library
- ♦ Statsmodels
- ♦ Seaborn
- ♦ re



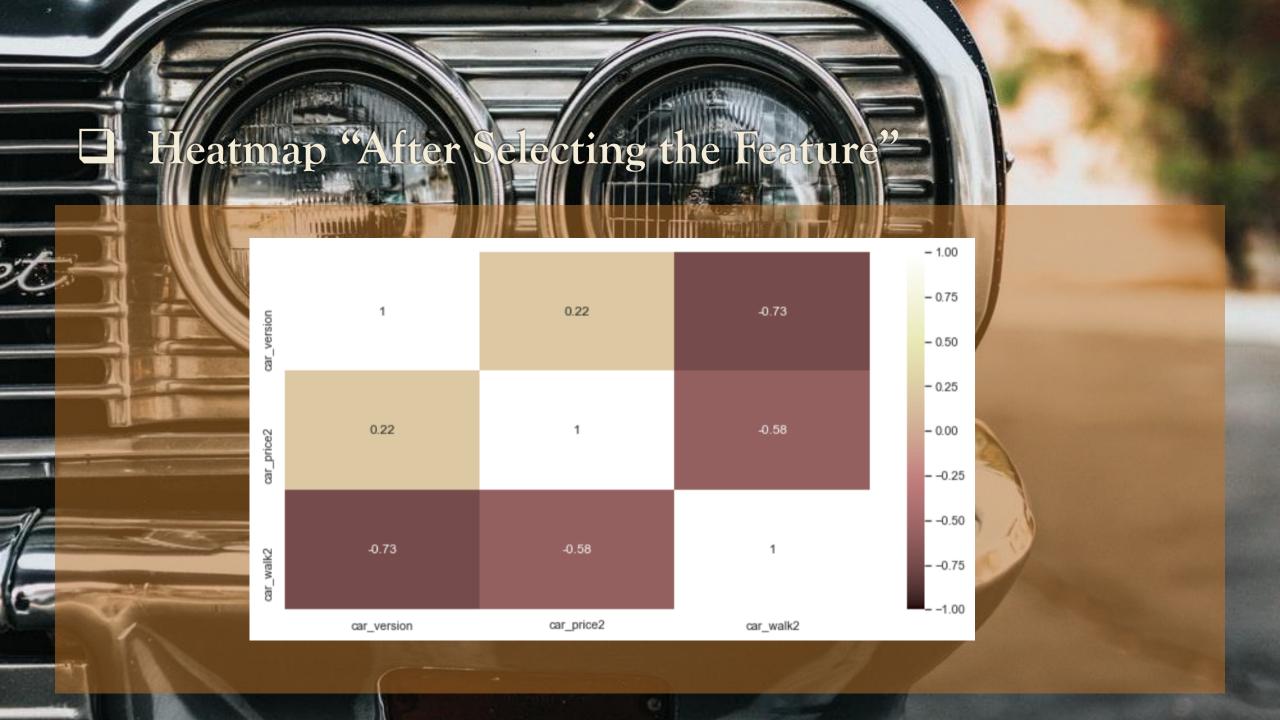


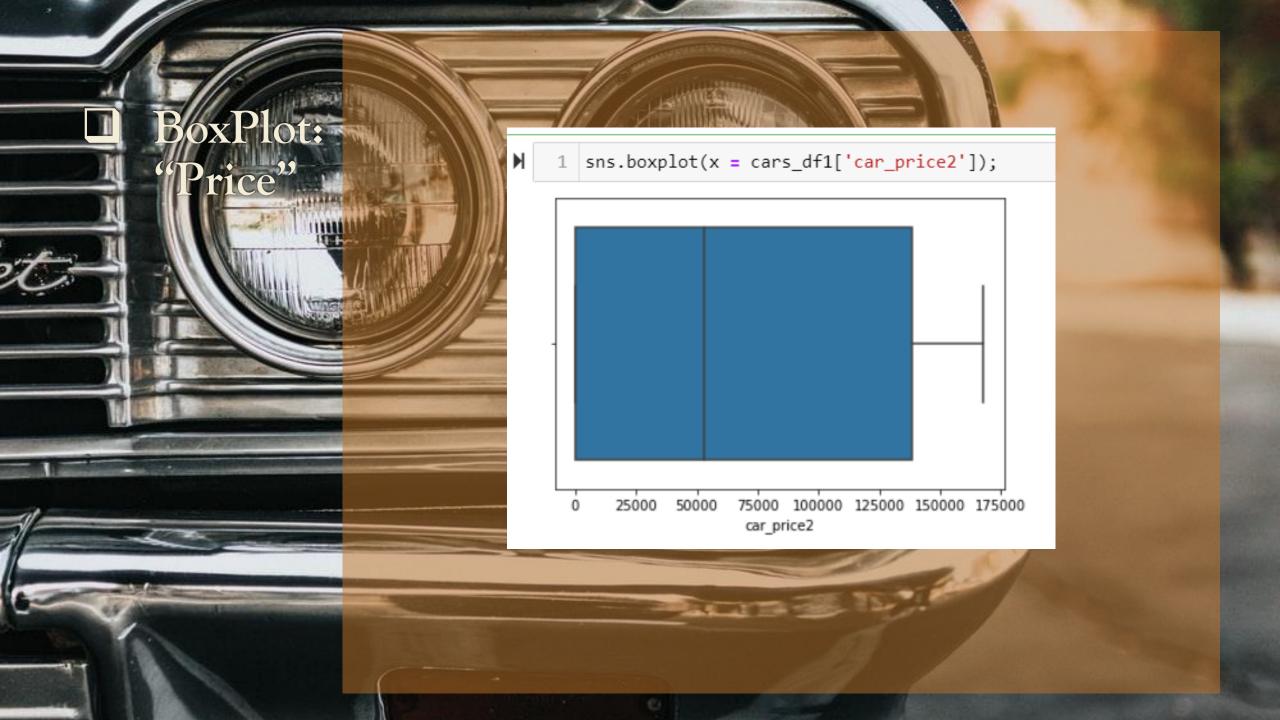


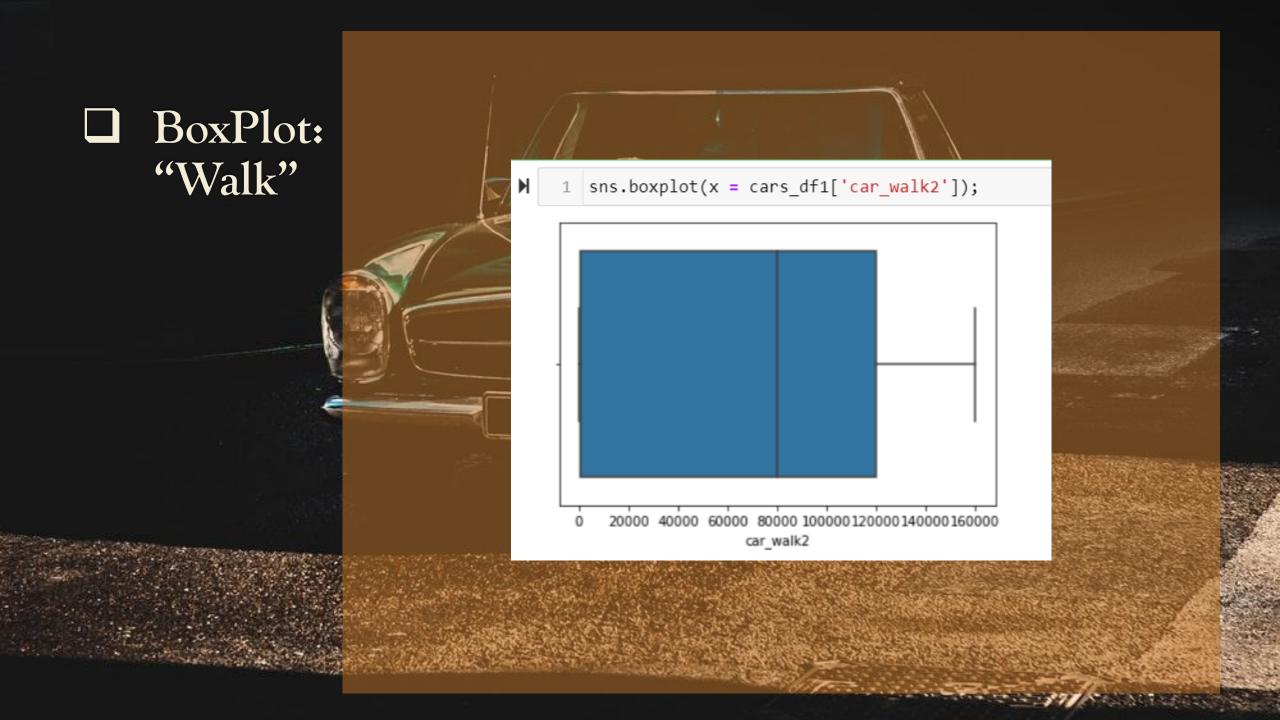
Heatmap

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		0.21
Hyundai -	-0.21	0.36	-0.053	1	-0.076	-0.35	-0.076		-0.21	0.46	-0.24
Nissan -	0.4	-0.066	-0.076	-0.076	1	-0.51		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		-0.51	1	-0.24	-0.3	0.67	0.36
Dammam -	0.47	0.47	0.31		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		0.46	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

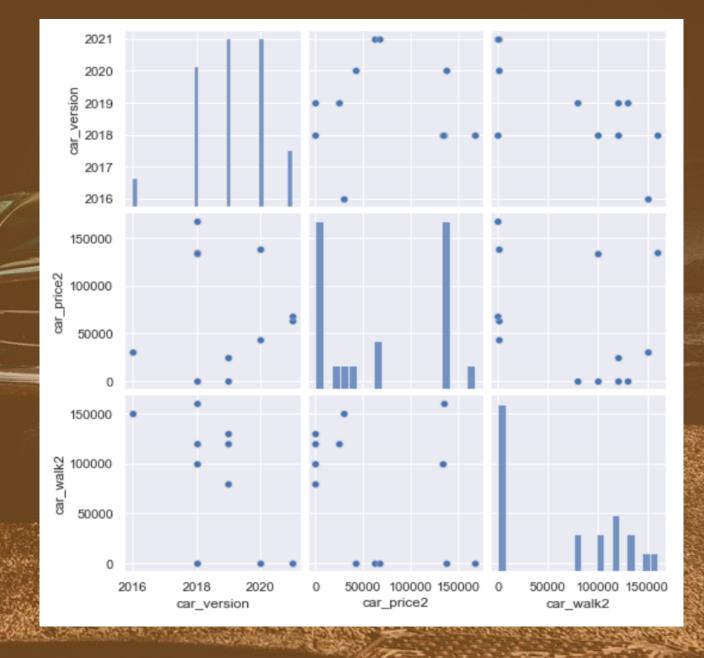






☐ BoxPlot: "Version" sns.boxplot(x = cars_df1['car_version']); 2016 2017 2019 2020 2021 2018 car_version

☐ PairPlot:



☐ RegPlot: car_price2



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, ...)

