**Machine Learning: Used Car Price Prediction**

**Introduction**

There are Many factors that might influence predicting the price of the car based on the features of the car. Our analysis aimed to define which features might have the strongest statistical correlation with the price.

1. **Goal:** **predict car prices using Python Machine Learning**
2. **Data used cleanup, analysis, processing:**
3. We used scraped data form Kaggle’s that scrapped form Craigslist.

The data was challenging and contains many issues.

1. We used pandas profiling

A screenshot of a cell phone

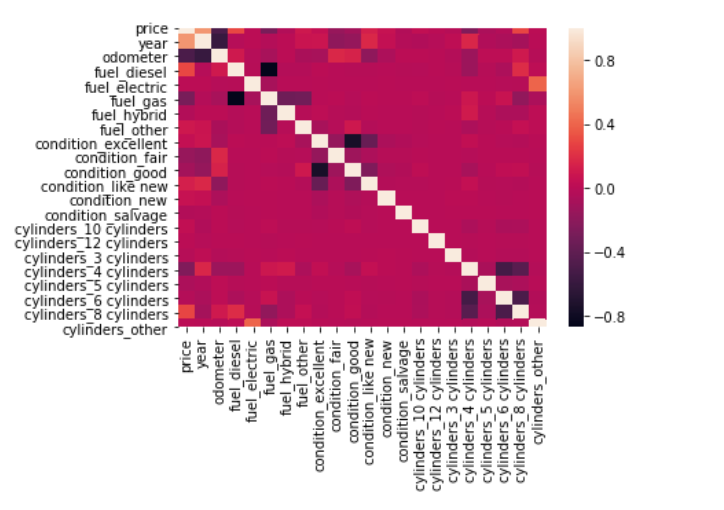
Description automatically generated A screenshot of a cell phone

Description automatically generated

1. Cleaning: We fixed or remove the anomalies and outliers discovered
2. Verification: After cleaning, the results were inspected to check the improvement
3. Reporting on the changes made and the quality of the currently stored data is recorded. This include summary statistic about the data

A picture containing screenshot, monitor

Description automatically generated



1. **Feature engineering**

* Created features through hot encoding then Checked how the features work with the model.

Hot encoded features: Fuel, condition, cylinders

|  |
| --- |
| fuel\_diesel |
| fuel\_electric |
| fuel\_gas |
| fuel\_hybrid |
| fuel\_other |
| condition\_excellent |
| condition\_fair |
| condition\_good |
| condition\_like new |
| condition\_new |
| condition\_salvage |
| cylinders\_10 cylinders |
| cylinders\_12 cylinders |
| cylinders\_3 cylinders |
| cylinders\_4 cylinders |
| cylinders\_5 cylinders |
| cylinders\_6 cylinders |
| cylinders\_8 cylinders |
| cylinders\_other |

1. **Running and choosing the models**
2. Multivariate Linear Regression:  R2: 0.65% MSE: 0.35
3. Ridge Regression R2 = 0.65% MSE :0.35
4. Lasso Regression: R² score 0.852544. MSE: 0.35, R2: 0.65
5. Elastic Net model MSE: 0.35 R2: 0.65
6. K-Nearest Neighbor KNN: did not show good results:

* k: 1, Train/Test Score: 0.837/0.303
* k: 3, Train/Test Score: 0.501/0.220
* k: 5, Train/Test Score: 0.397/0.195
* k: 7, Train/Test Score: 0.337/0.177
* k: 9, Train/Test Score: 0.298/0.161
* k: 11, Train/Test Score: 0.270/0.149
* k: 13, Train/Test Score: 0.247/0.139
* k: 15, Train/Test Score: 0.229/0.131
* k: 17, Train/Test Score: 0.214/0.124
* k: 19, Train/Test Score: 0.200/0.118

1. **Training evaluation and Predictions**

We chose the multi variate linear regression with highest R² score and the lowest Root Mean Squared Error.

A close up of a map

Description automatically generated