Worth it?

Predicting the Prices of Used Cars in Singapore

As a buyer, how can I quantify that my purchase is reasonable?

Data-Driven Approach:

- Linear Regression Model to develop quantifable way to benchmark prices
- Using a few typical features of a car

DATA COLLECTION

Browse by Category | Save Search | Print | « Back (1 2 3 4 5 ... 127) Next » | Sort by Date Posted | 100 results/page

12,683 vehicles | Search | Car Make / Model | Advanced Search | Search | View All

	Make	Model	Price	Depreciation	Reg Date	Eng Cap	Mileage	Veh Type	Status
Search Selection	Any	Any	Any	Any	Any	Any	Any	Any	**Available**

Kia Cerato Forte 1.6A SX | **$79,888** | $7,900 /yr | 16-Oct-2018 | 1,591 cc | 58,000 km | Sedan | **Available**

Upgrade Playable Youtube, Bluetooth Enable, 18" Rim, Custom Interior Neon Lights.

DIRECT OWNER

Posted: 29-Oct-2019 Tags: 2018 Kia Cerato, 2018 kia cerato, Kia Cerato, kia cerato, Kia, Cerato, cerato, Used Kia

Chery T11 2.0A 5DR | **$9,800** | N.A | 28-Apr-2010 | 1,971 cc | 150,000 km | SUV | **Available**

Full Loan Available, Budget SUV In The Market, Super Cheap, Point A To Point B Car, Please Call Our Sales To View And Test Drive The Car.

1axis

Posted: 29-Oct-2019 Tags: 2010 Chery T11, 2010 chery t11, Chery T11, chery t11, Chery, T11, t11, Used Chery

Kia Cerato Forte 1.6A SX

DIRECT OWNER

Price	**$79,888**		
Depreciation	$7,900 /yr View models with similar depre	**Reg Date**	16-Oct-2018 (8yrs 11mths 16days COE left)
Mileage	58,000 km (56k /yr)	**Manufactured**	2018
Road Tax	$738 /yr	**Transmission**	Auto
Dereg Value	$38,622 as of today (change)	**OMV**	$18,009
COE	$28,000	**ARF**	$18,009



Kia Cerato Forte 1.6A SX

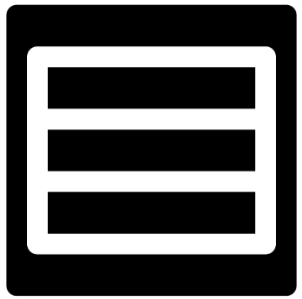
DIRECT OWNER

Price	**$79,888**		
Depreciation	$7,900 /yr View models with similar depre	**Reg Date**	16-Oct-2018 (8yrs 11mths 16days COE left)
Mileage	58,000 km (56k /yr)	**Manufactured**	2018
Road Tax	$738 /yr	**Transmission**	Auto
Dereg Value	$38,622 as of today (change)	**OMV**	$18,009
COE	$28,000	**ARF**	$18,009



DATA CLEANING

Data Cleaning



Rows:
- 2584

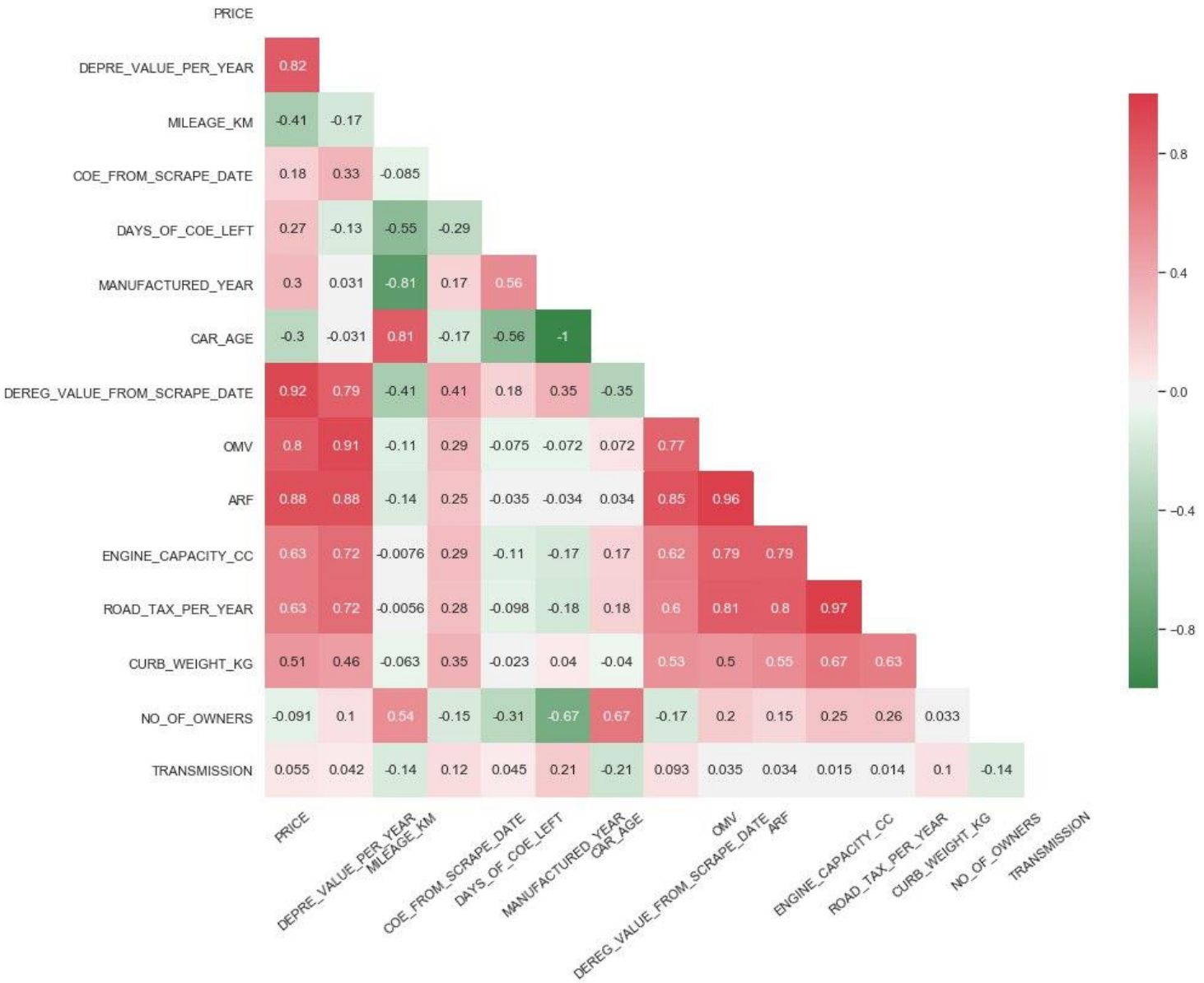
Columns:
- 17

Target Variable:
- Price

Predictor Variables:
- 16 variables

FEATURE SELECTION

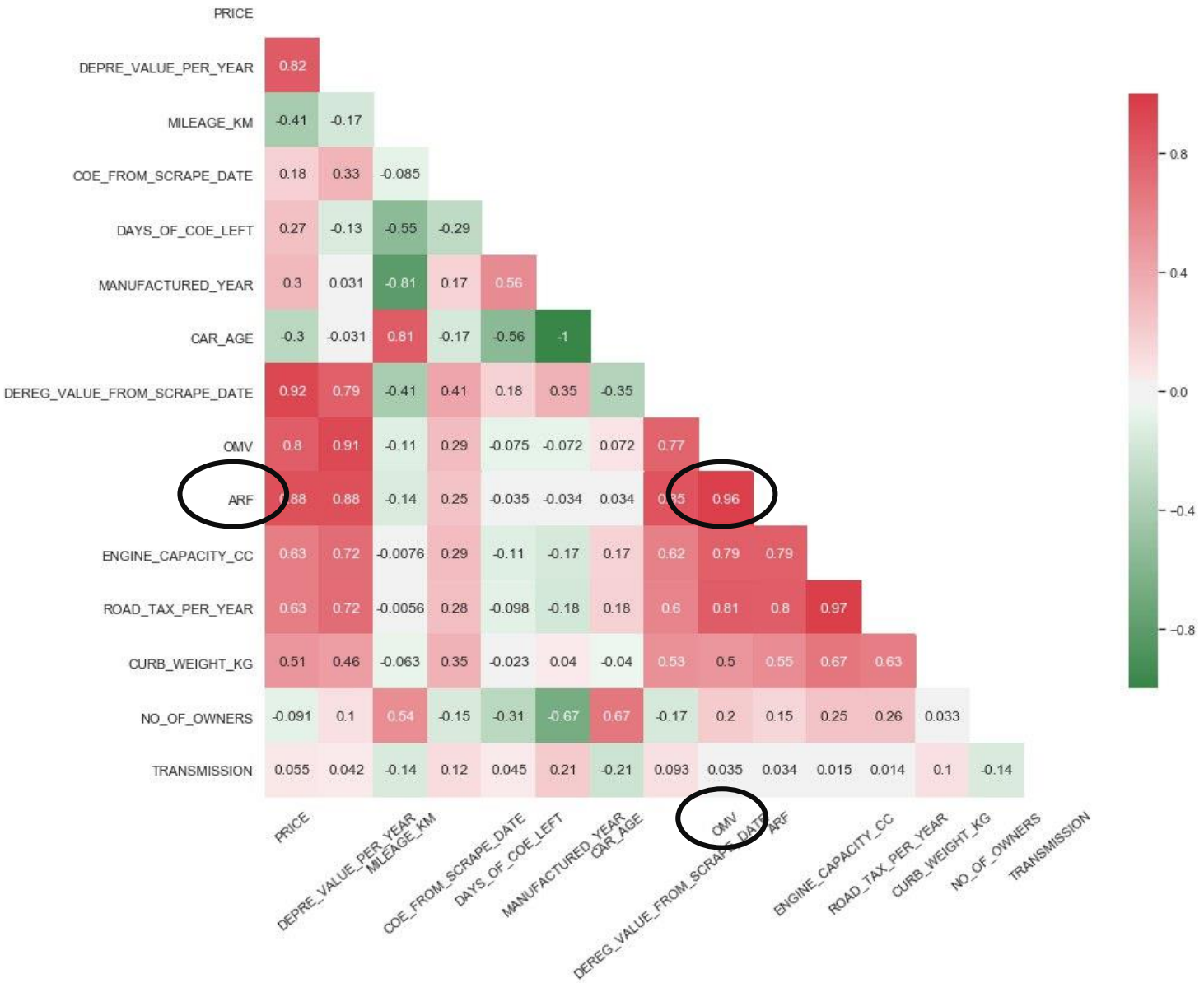
Feature Selection



Correlation Heat Map

- Correlation to price
- Predictors correlate to each other

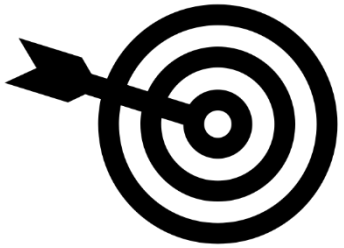
Feature Selection



Correlation Heat Map

- Some predictor various poor fit with price
- Predictor variables with high correlation

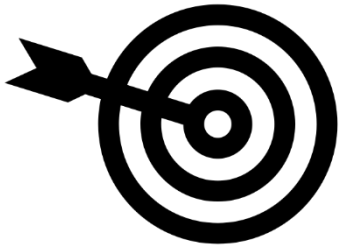
Data Used: After Dropping Features



Target Variable:
- Price

Predictor Variables:
- Mileage (km)
- Engine Capacity (CC)
- Car Age
- OMV
- COE Price
- Days of COE left
- Curb Weight (kg)
- # of Owners
- Transmission

Data Used: Finalized Features



Target Variable:

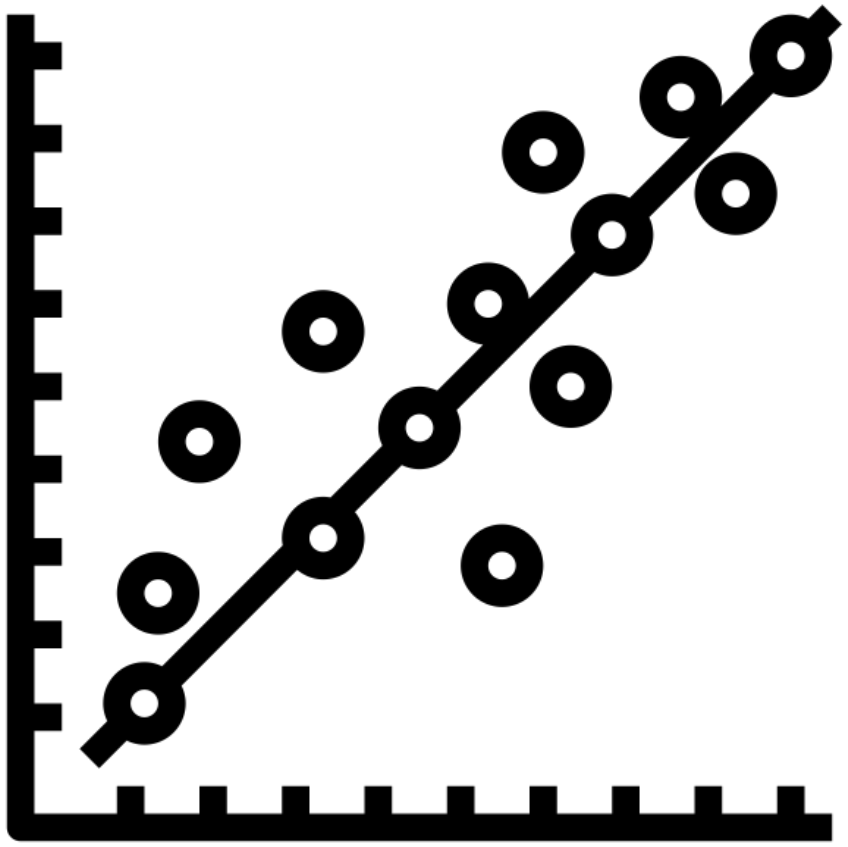
- *Log* Price

Predictor Variables:

- *Log* Mileage (km)
- Engine Capacity (CC)
- Car Age
- OMV
- COE Price
- Days of COE left
- Curb Weight (kg)
- # of Owners
- Transmission

METHODOLOGY

Designing the Regression Model for Data



- Simple Linear Regression
- Lasso
- Ridge
- ElasticNet
- Polynomial Regression

Train / Cross-Validate each Model

Train Models on 60% Data

- Simple Linear Regression
- Lasso
- Ridge
- ElasticNet
- Polynomial Regression

KFold

- 5

Rigorous Model Selection Method

R2 Scores over 5 Validations

#	Model	R2 Score
1	Simple Linear Regression	0.913 +- 0.016
2	Lasso	0.913 +- 0.015
3	Ridge	0.913 +- 0.015
4	ElasticNet	0.913 +- 0.015
5	Polynomial Regression	0.829 +- 0.105

Model Selection

#	Model	R2 Score
1	Simple Linear Regression	0.913 +- 0.016
2	Lasso	0.913 +- 0.015
3	Ridge	0.913 +- 0.015
4	ElasticNet	0.913 +- 0.015
5	Polynomial Regression	0.829 +- 0.105

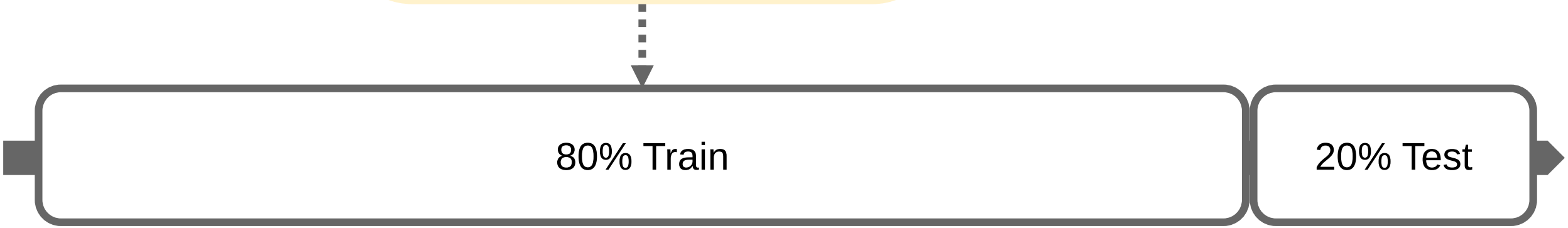
Model Selection

#	Model	R2 Score
1	Simple Linear Regression	0.913 +- 0.016

- Easy to use
- No need for hyperparameters

Train / Test Simple Linear Regression

Re-Train Simple Linear Regression Model

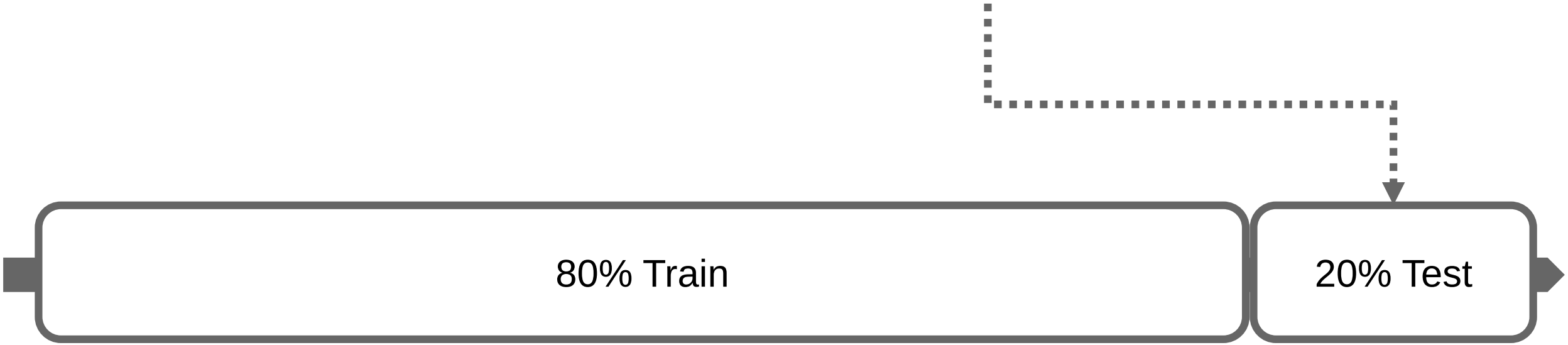


Train / Test Simple Linear Regression

Test Simple Linear Regression Model

Test Results

R2 Score: 0.913



Test Results

91%

of data variation
explained by model

9%

Possibly due to condition of car that could
not be obtained during web-scraping

RESULTS

Using the Model: Model Coefficient

Feature	Coefficient Value	Feature Effect	Price Effect
Car Age	-0.06	Every unit ***increase*** in Car Age	6% ***decrease*** in price
Engine Capacity (CC)	0.0001	Every unit ***increase*** in CC	0.01% ***increase*** in price

Using the Model: User Value

Use **typical**, easily accessible features of cars

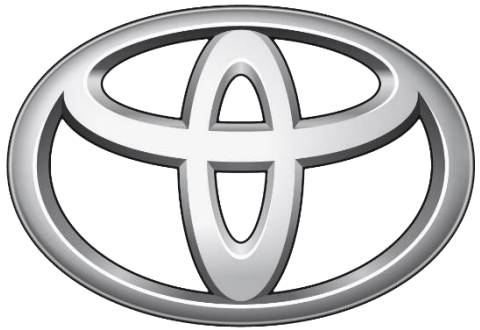
Obtain **benchmark** of base car pricing

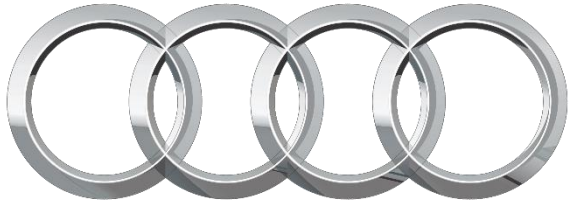
Cross-check with used car listing price and car brands


FUTURE IDEAS
Y321 MCL

Future Ideas

- Build on current features
- Include brand data into model itself
- Model also able to predict how much a used car would cost given a brand
- More customization for the user





Questions?

Appendix

Model Summary

OLS Regression Results

Dep. Variable:	PRICE_LOG	R-squared:	0.922
Model:	OLS	Adj. R-squared:	0.921
Method:	Least Squares	F-statistic:	1731.
Date:	Wed, 30 Oct 2019	Prob (F-statistic):	0.00
Time:	01:46:15	Log-Likelihood:	716.49
No. Observations:	2067	AIC:	-1403.
Df Residuals:	2052	BIC:	-1318.
Df Model:	14		
Covariance Type:	nonrobust		

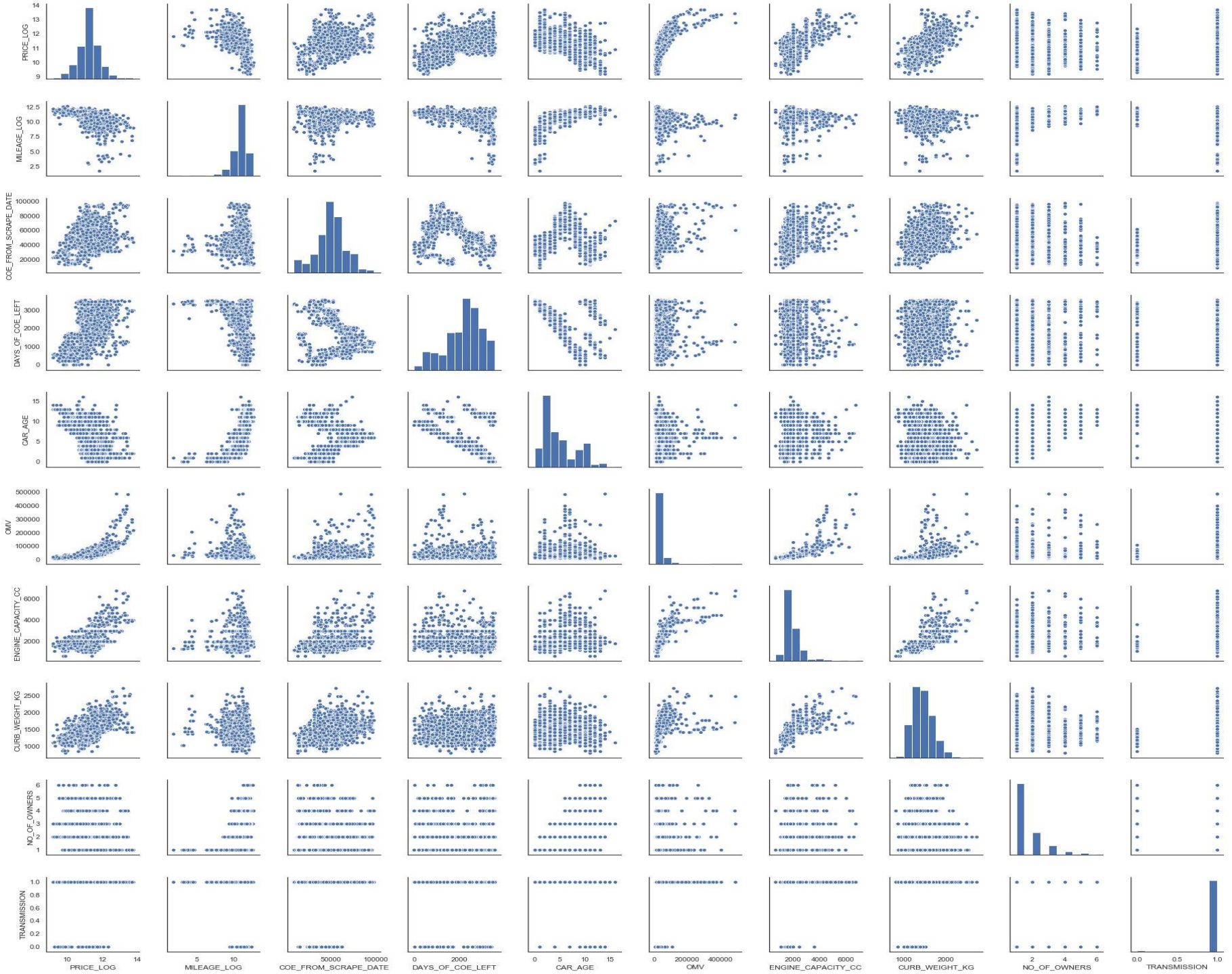
	coef	std err	t	P>|t|	[0.025	0.975]
const	9.6648	0.050	194.726	0.000	9.567	9.762
MILEAGE_LOG	-5.811e-07	1.71e-07	-3.398	0.001	-9.16e-07	-2.46e-07
COE_FROM_SCRAPE_DATE	4.446e-06	2.98e-07	14.907	0.000	3.86e-06	5.03e-06
DAYS_OF_COE_LEFT	0.0003	6.56e-06	43.785	0.000	0.000	0.000
CAR_AGE	-0.0588	0.003	-22.568	0.000	-0.064	-0.054
OMV	5.969e-06	2.49e-07	23.978	0.000	5.48e-06	6.46e-06
ENGINE_CAPACITY_CC	0.0001	1.16e-05	10.458	0.000	9.87e-05	0.000
CURB_WEIGHT_KG	0.0005	2.35e-05	19.555	0.000	0.000	0.001
NO_OF_OWNERS	-0.0139	0.006	-2.514	0.012	-0.025	-0.003
TRANSMISSION	-0.1328	0.033	-4.060	0.000	-0.197	-0.069

Omnibus:	404.838	Durbin-Watson:	1.950
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6823.058
Skew:	-0.429	Prob(JB):	0.00
Kurtosis:	11.859	Cond. No.	inf

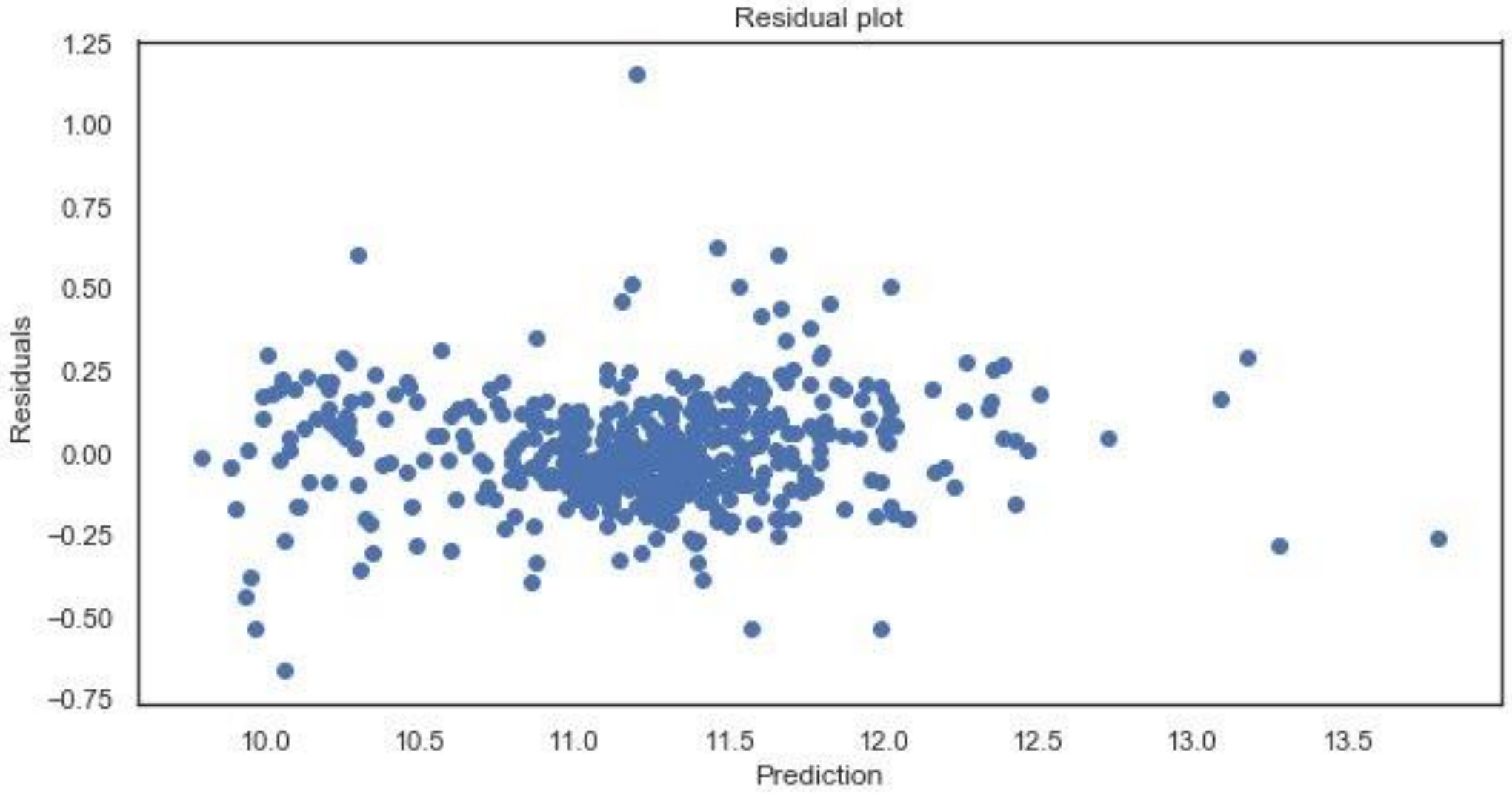
Model Scores

- MAE: 0.123
- MSE: 0.0285
- RMSE: 0.169
- R2 Score: 0.913

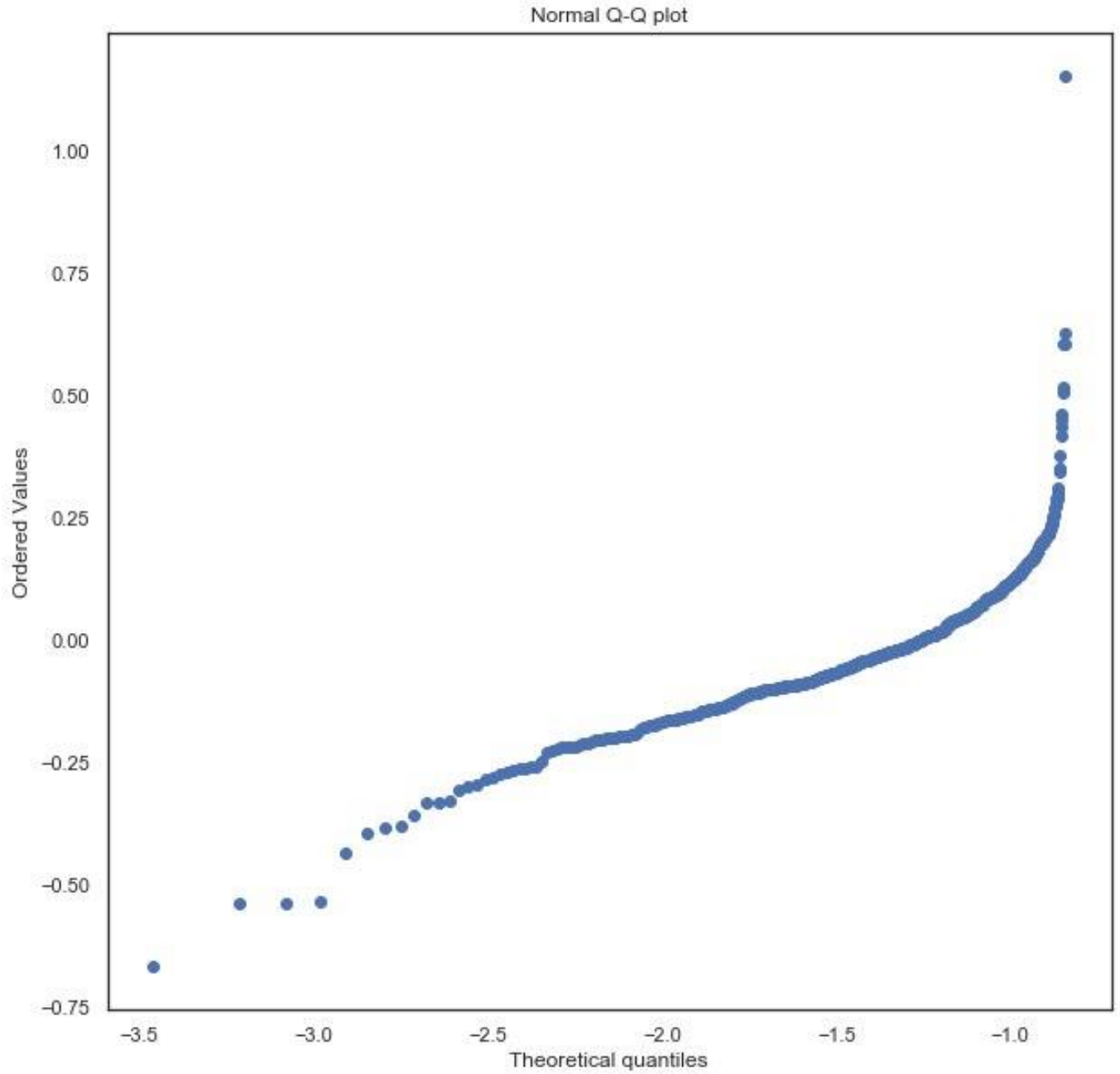
Pairplot


PRICE_LOG
MILEAGE_LOG
COE_FROM_SCRAPE_DATE
DAYS_OF_COE_LEFT
CAR_AGE
OMV
ENGINE_CAPACITY_CC
CURB_WEIGHT_KG
NO_OF_OWNERS
TRANSMISSION

Residual Plot



QQ Plot


Normal Q-Q plot
Ordered Values
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
Theoretical quantiles