Predicting Vehicle Popularity by Features

By
Douglas Hundley

Problem Statement

- Need to identify the features that make a car popular
- Need to create a model to predict the popularity of new cars.
- Success metrics are having and R2 score of 80 and having a better RMSE score than null model

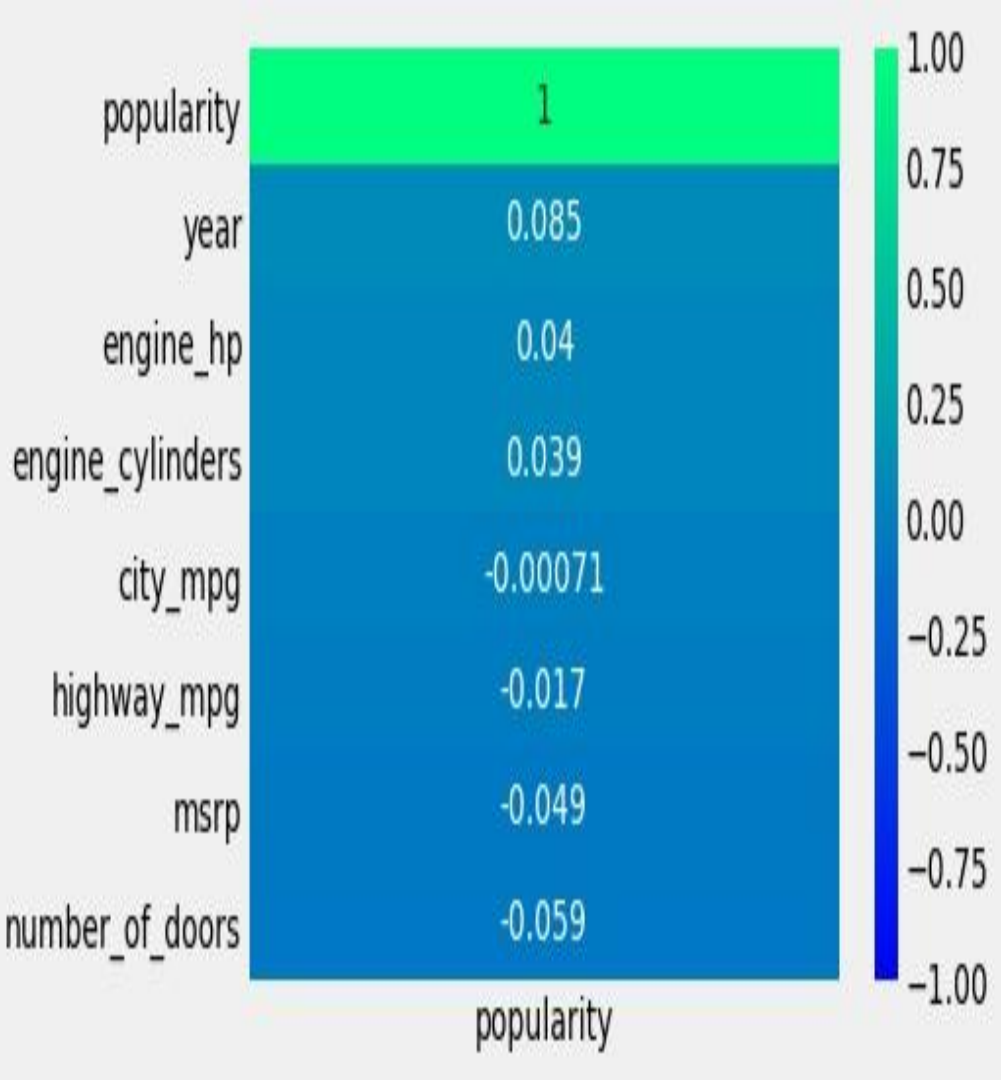
Data Dictionary

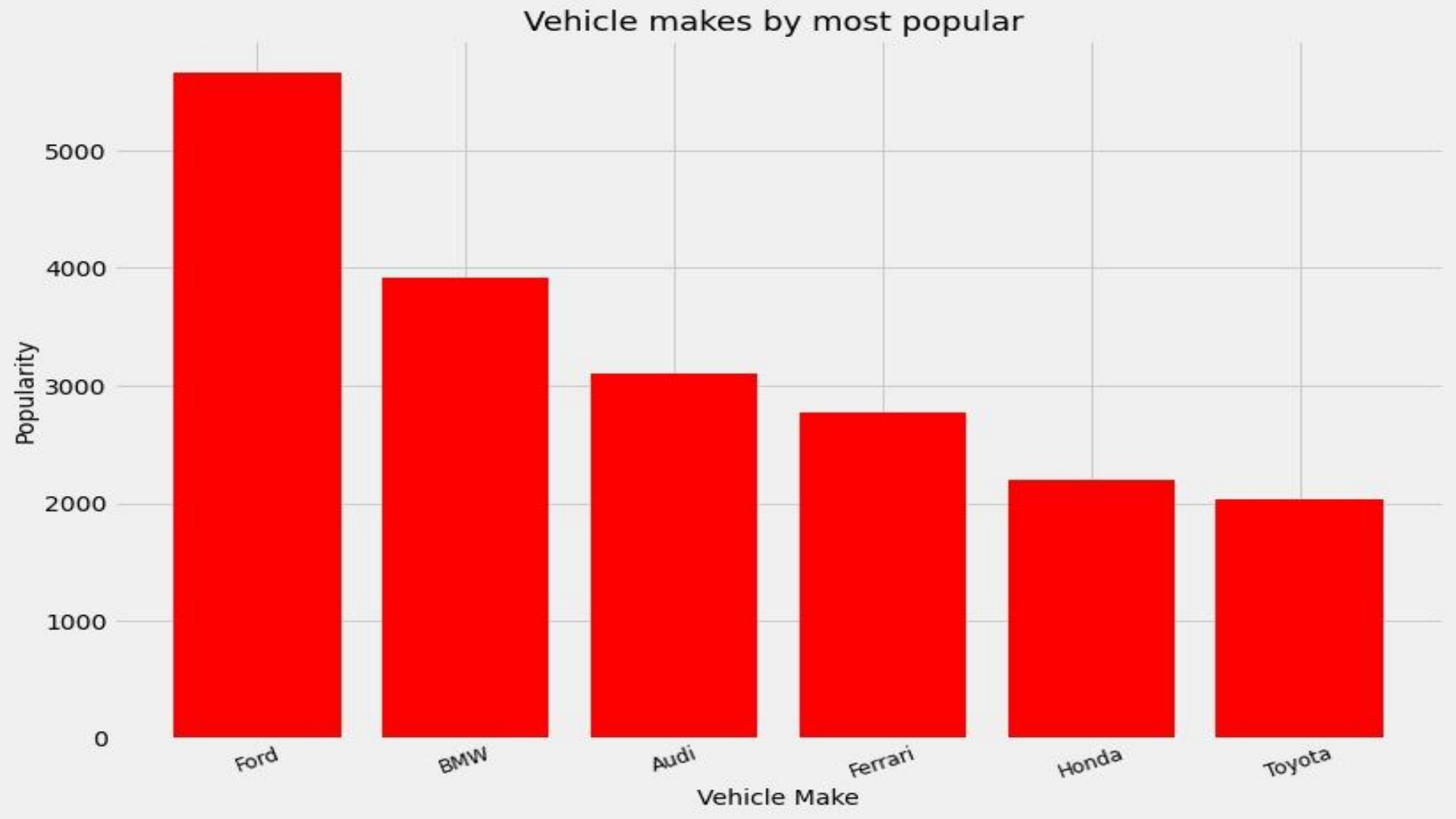
Feature	Type	Dataset	Description
make	*object*	cars.csv	The car manufacturer
model	*object*	cars.csv	Model of the car
engine_fuel_type	*object*	cars.csv	Type of fuel the car takes
engine_hp	*float*	cars.csv	Car Horsepower
engine_cylinders	*integer*	cars.csv	Number of cylinders
transmission_type	*object*	cars.csv	Transmission type of car
driven_wheels	*object*	cars.csv	Drive type of car
number_of_doors	*float*	cars.csv	Total count of words in title column
market_category	*object*	cars.csv	Marketing type of car
vehicle_size	*object*	cars.csv	Size type of car
vehicle_style	*object*	cars.csv	Style of car
highway_mpg	*integer*	cars.csv	Highway miles per gallon
city_mpg	*integer*	cars.csv	City miles per gallon
popularity	*integer*	cars.csv	Popularity rating of car
msrp	*integer*	cars.csv	Manufacturer suggested retail price

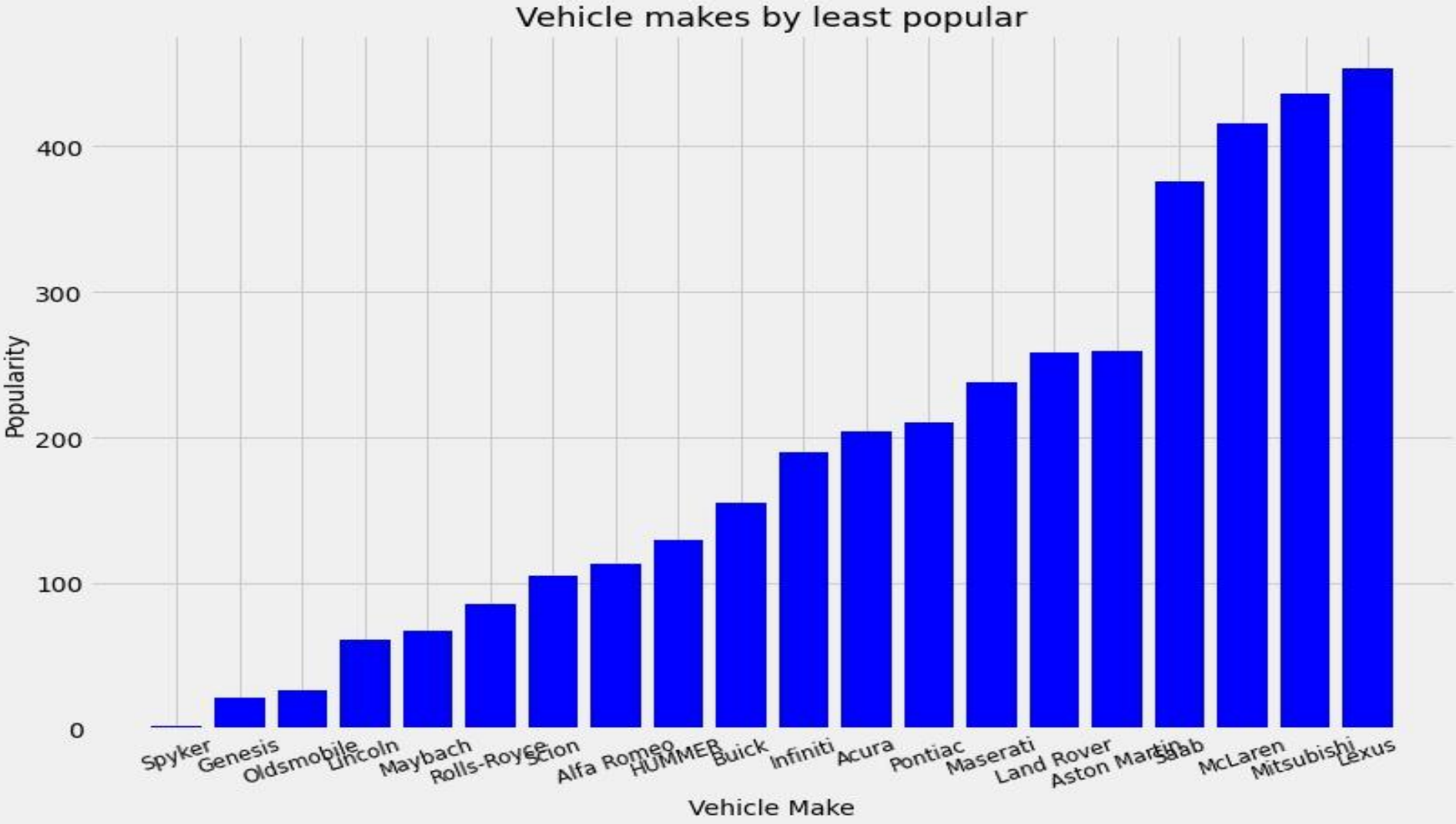
Prepping The Data

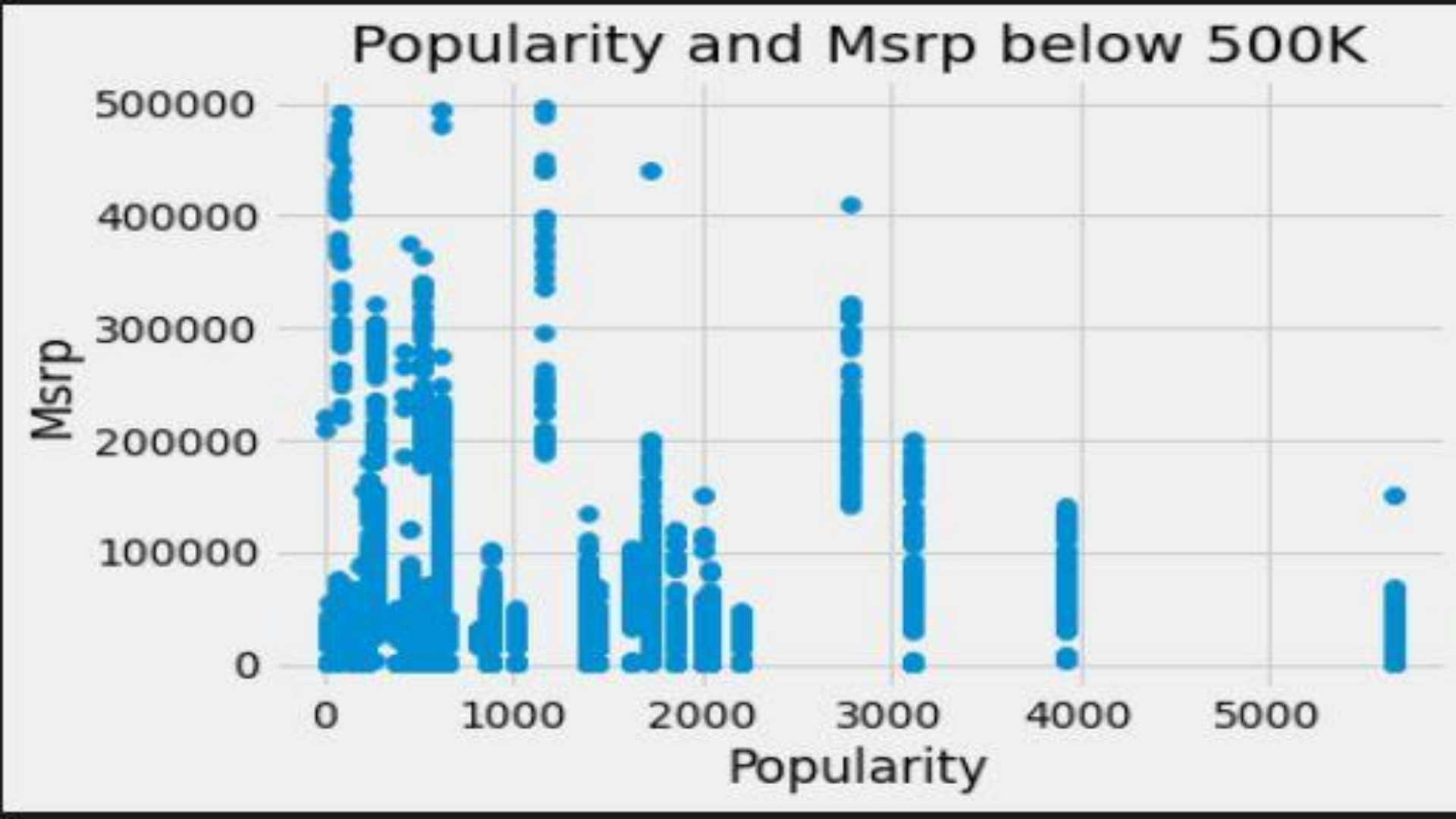
- Clean the data
- Impute the nulls
- Select the features

Correlations




Vehicle makes by most popular
Popularity
5000
4000
3000
2000
1000
0
Ford
BMW
Audi
Ferrari
Honda
Toyota
Vehicle Make


Vehicle makes by least popular
Popularity
0
100
200
300
400
Spyker
Genesis
Oldsmobile
Lincoln
Maybach
Rolls-Royce
Scion
Alfa Romeo
HUMMER
Buick
Infiniti
Acura
Pontiac
Maserati
Land Rover
Aston Martin
Saab
McLaren
Mitsubishi
Lexus
Vehicle Make


Popularity and Msrp below 500K
500000
400000
300000
200000
100000
0
Msrp
0
1000
2000
3000
4000
5000
Popularity

The Models

Model	Train Score	Test Score	RMSE
Null	NA	NA	1445.6989
1st Lasso Model	0.9999	0.9999	4.1932
2nd Lasso Model	0.1347	0.1331	1385.5655
Linear Model	0.1335	0.1350	1385.9820
Ridge Model	0.1336	0.1353	1385.8341
Regression Tree Model	0.9809	0.8764	524.0349

Conclusion And Recommendations

- Brand affects popularity the most
- Keep your car affordable to the general public
- Regression tree is the winning model