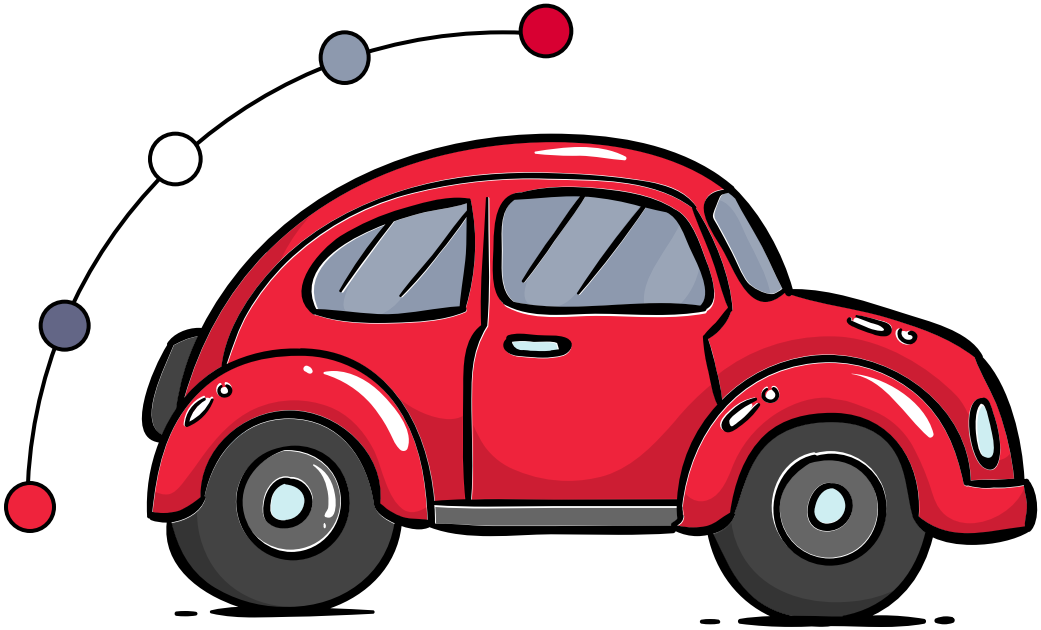
The prediction of a car price

Razan Aljuhani

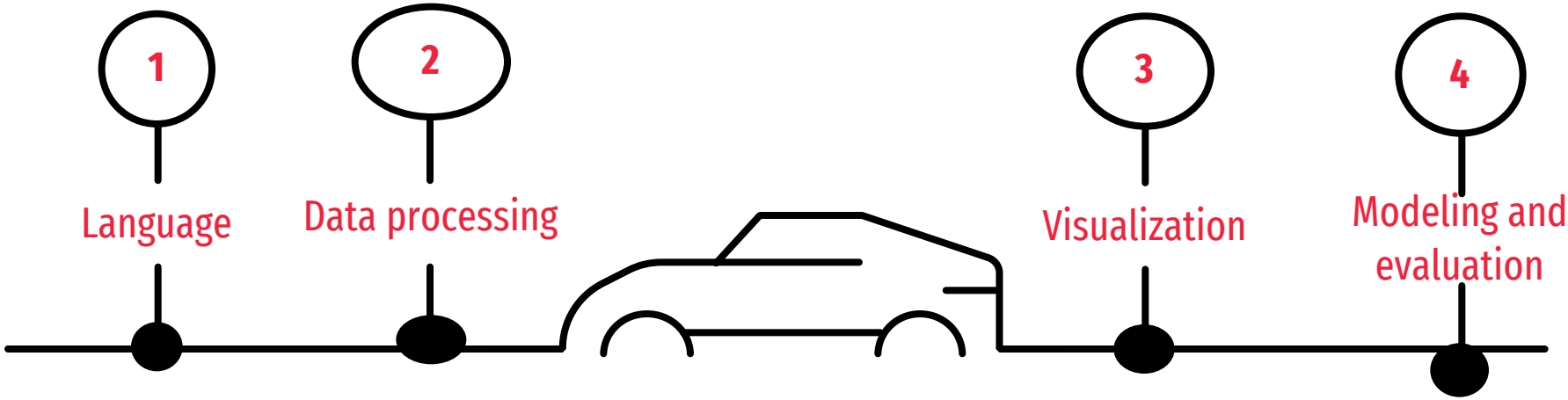


Introduction

Companies try to produce cars with special features to take attention in the market and compete with other competitors. They need to understand the dynamic changes of pricing in the market through the years, and the factors affecting the pricing

The project will help in clarify the features has a direct impact on the target variable (Selling Price), by build an effective model using Linear regression.

Tools



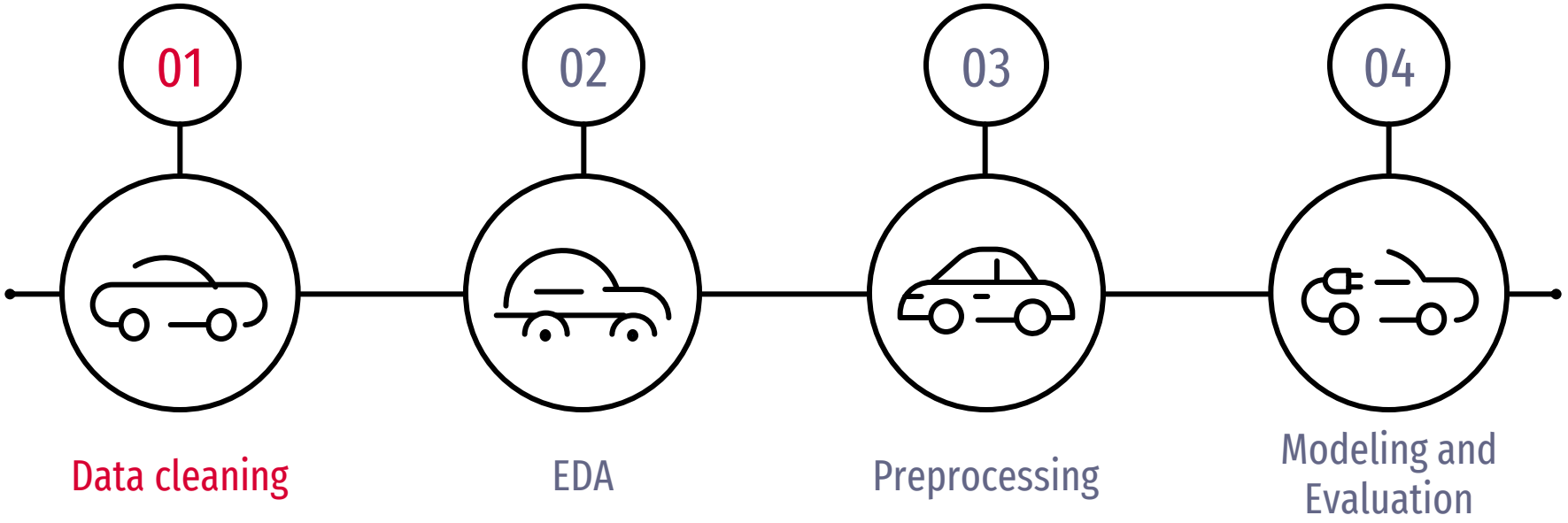
Python

1. Numpy
2. Pandas

1. Seaborn
2. Matplotlib
3. Plotly

Scikit-learn

Stages


01
02
03
04
Data cleaning
EDA
Preprocessing
Modeling and Evaluation

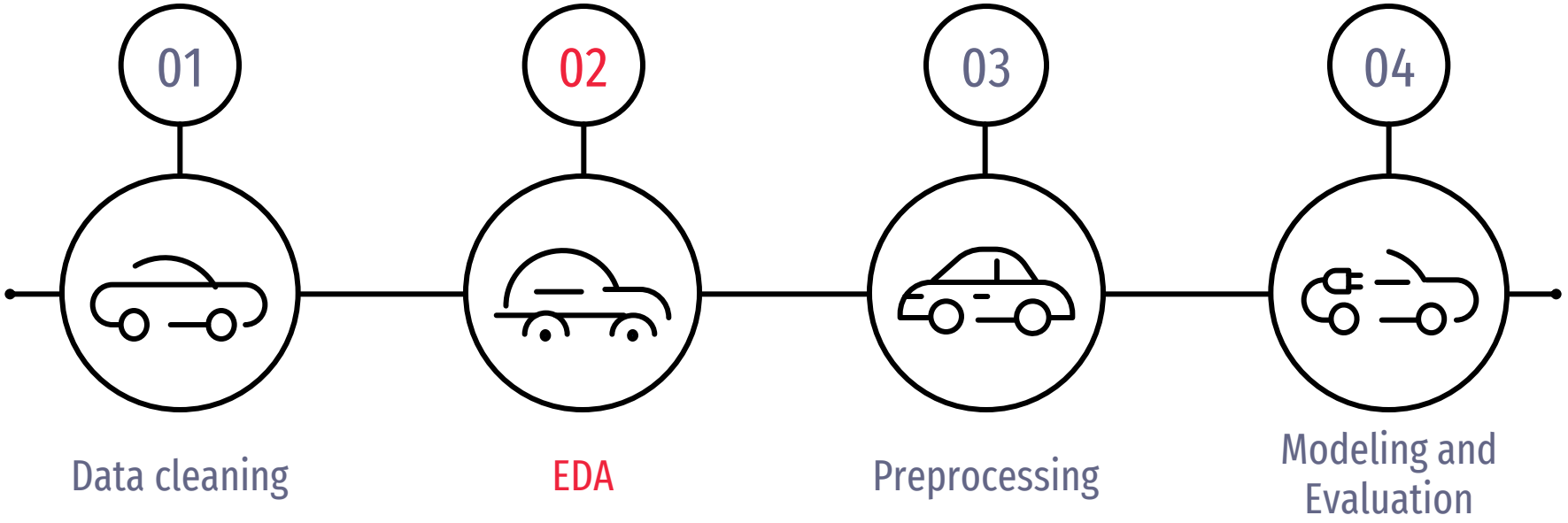
Original Data

	Name	Year	Selling_ price	Km_ driven	Fuel	Seller_ Type	transmis sion	Owner	Mileage	Engine	Max_ Power	Torque	Seats
0	Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individu al	Manual	First Owner	23.4 kmpl	1248 CC	74 bhp	190Nm @ 2000rpm	5.0
1	Skoda Rapid 1.5 TDI Ambitio n	2014	370000	120000	Diesel	Individu al	Manual	Second Owner	21.14 kmpl	1498 CC	103.52 bhp	250Nm @ 1500- 2500rpm	5.0

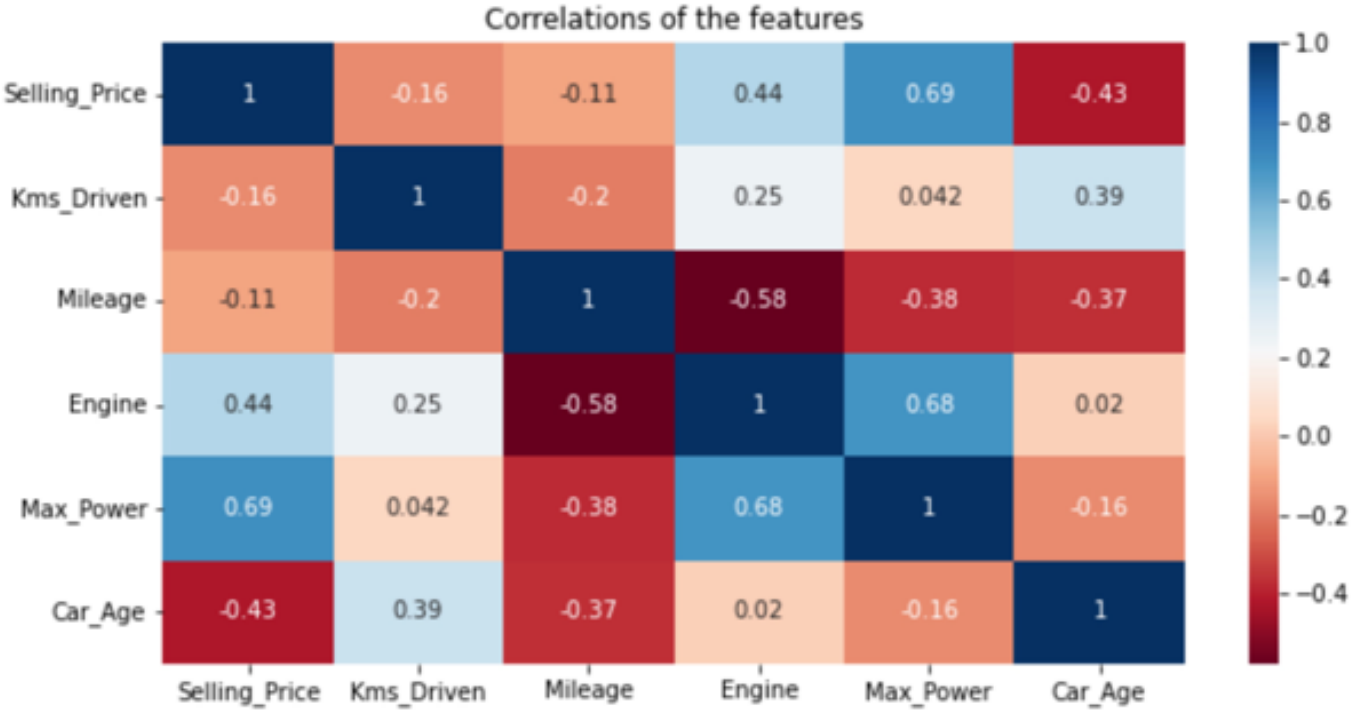
Data after cleaning

	Selling_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission_Type	Owner_Number	Mileage	Engine	Max_Power	Seats_Number	Car_Age	Car_Brand
0	13.017003	145500	Diesel	Individual	Manual	First Owner	23.40	1248.0	4.304065	5.0	1.945910	Maruti
1	12.821258	120000	Diesel	Individual	Manual	Second Owner	21.14	1498.0	4.639765	5.0	1.945910	Skoda

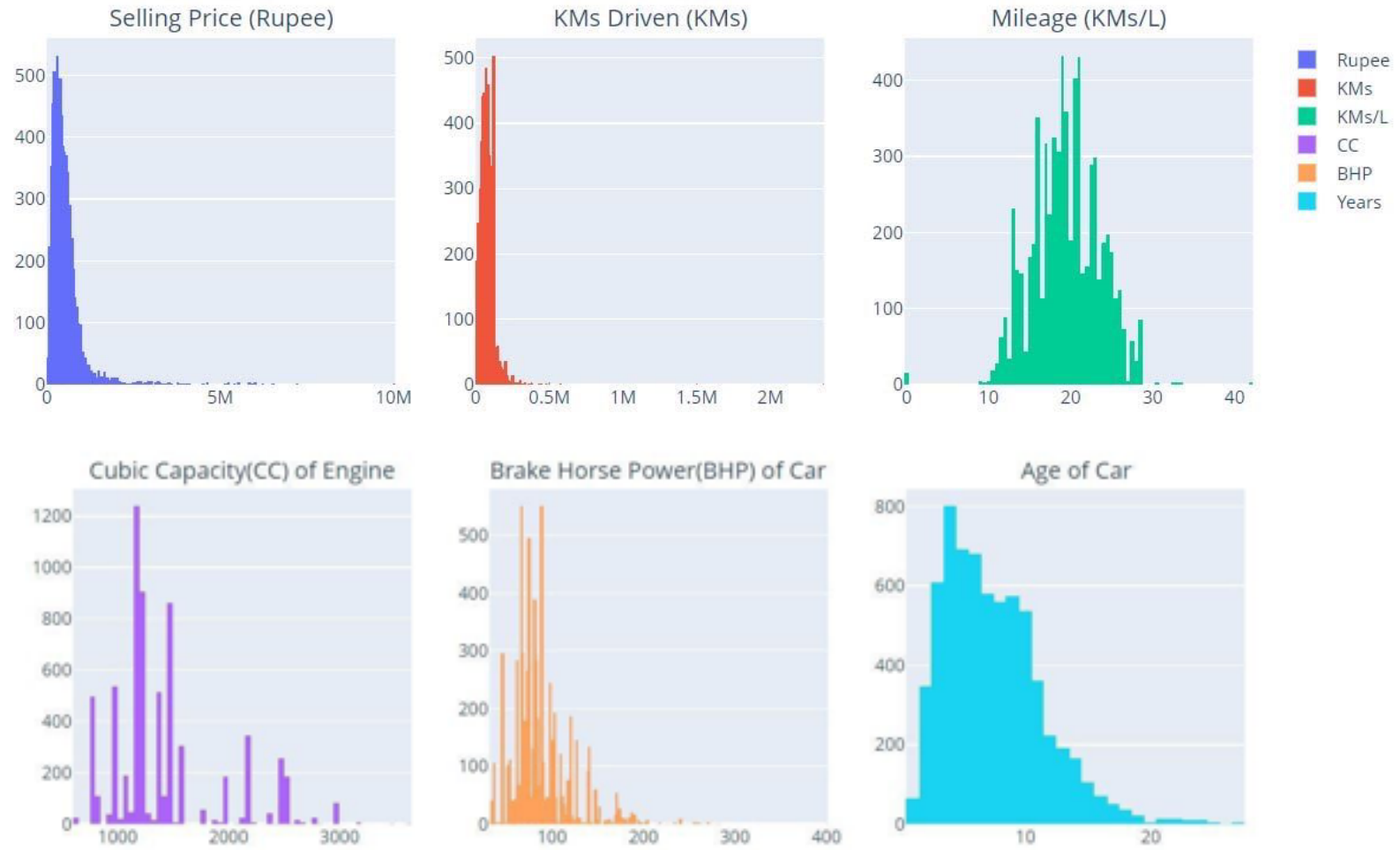
Stages

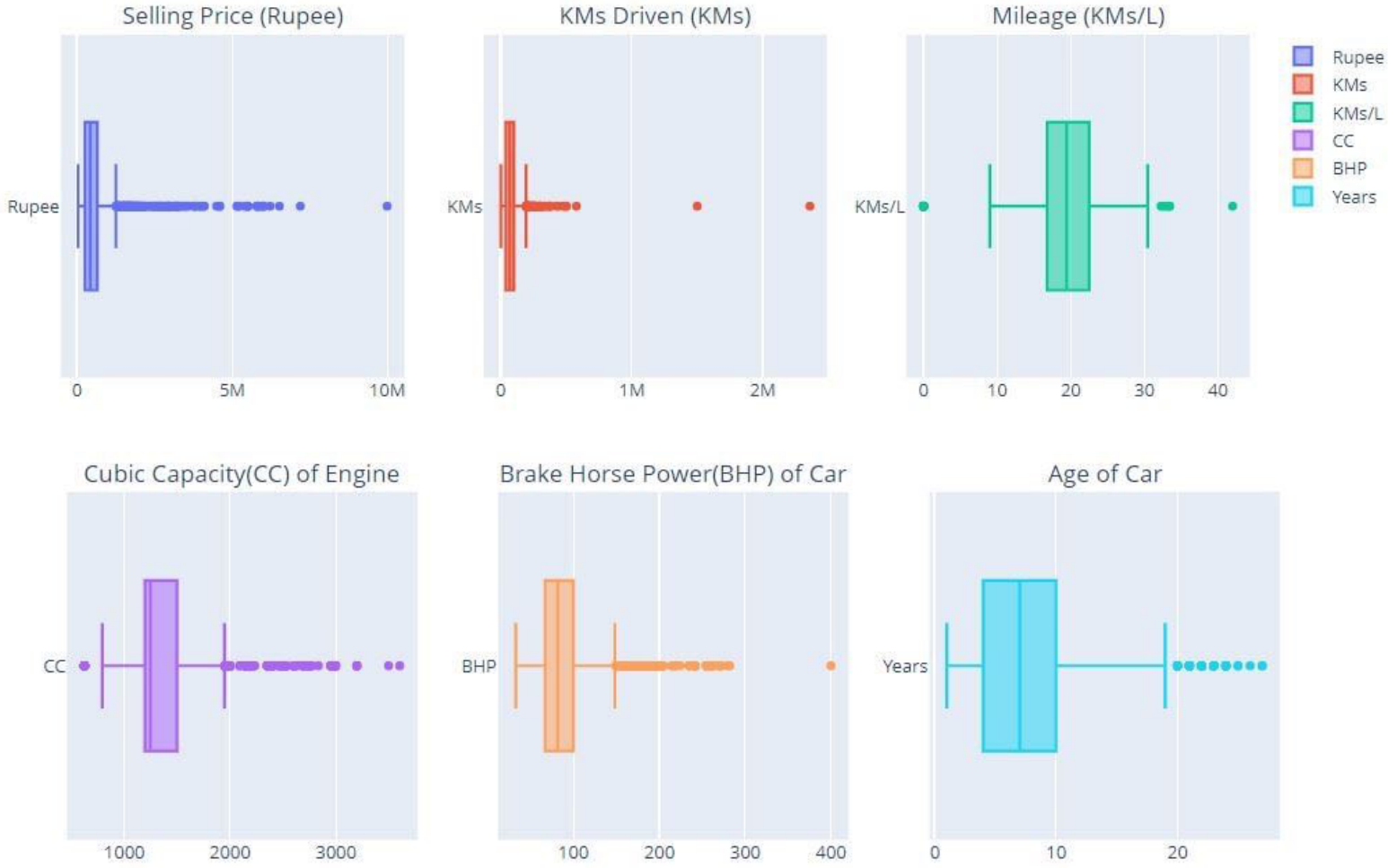

01
02
03
04
Data cleaning
EDA
Preprocessing
Modeling and Evaluation

Correlation of the features

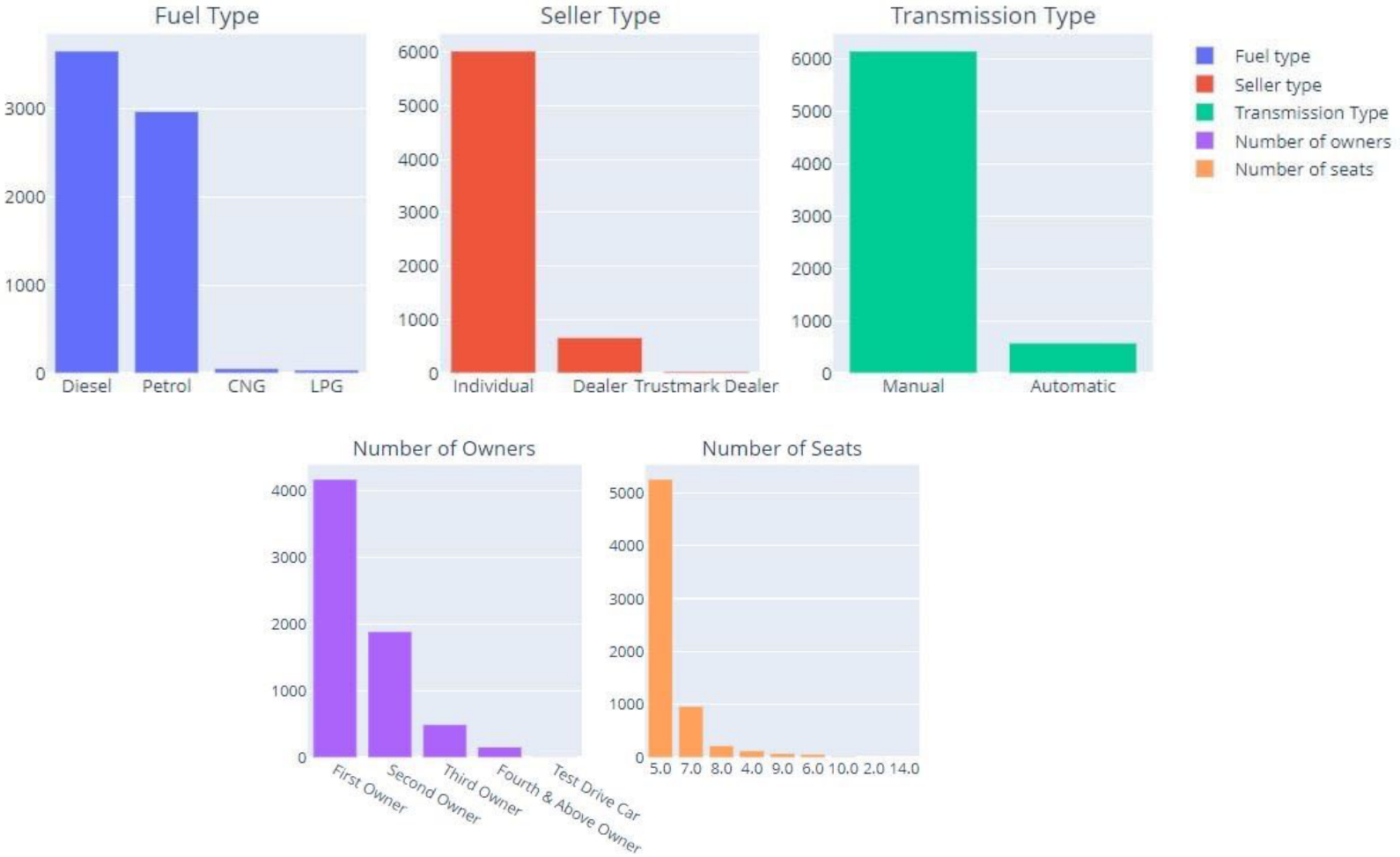


Distribution of the numerical data

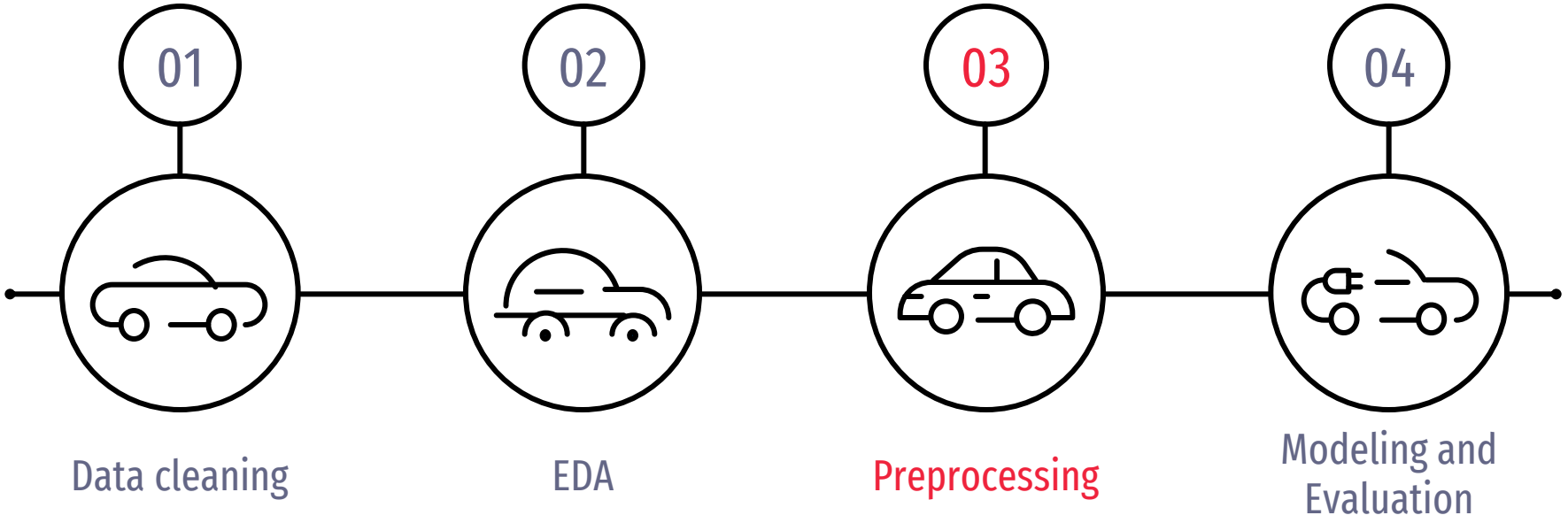

Selling Price (Rupee)
KMs Driven (KMs)
Mileage (KMs/L)
Rupee
KMs
KMs/L
CC
BHP
Years
Cubic Capacity(CC) of Engine
Brake Horse Power(BHP) of Car
Age of Car


Selling Price (Rupee)
Rupee
0
5M
10M
KMs Driven (KMs)
KMs
0
1M
2M
Mileage (KMs/L)
KMs/L
0
10
20
30
40
Rupee
KMs
KMs/L
CC
BHP
Years
Cubic Capacity(CC) of Engine
CC
1000
2000
3000
Brake Horse Power(BHP) of Car
BHP
100
200
300
400
Age of Car
Years
0
10
20

Distribution of the categorical data


Fuel Type
3000
2000
1000
0
Diesel
Petrol
CNG
LPG
Seller Type
6000
5000
4000
3000
2000
1000
0
Individual
Dealer
Trustmark Dealer
Transmission Type
6000
5000
4000
3000
2000
1000
0
Manual
Automatic
Fuel type
Seller type
Transmission Type
Number of owners
Number of seats
Number of Owners
4000
3000
2000
1000
0
First Owner
Second Owner
Third Owner
Fourth & Above Owner
Test Drive Car
Number of Seats
5000
4000
3000
2000
1000
0
5.0
7.0
8.0
4.0
9.0
6.0
10.0
2.0
14.0

Stages

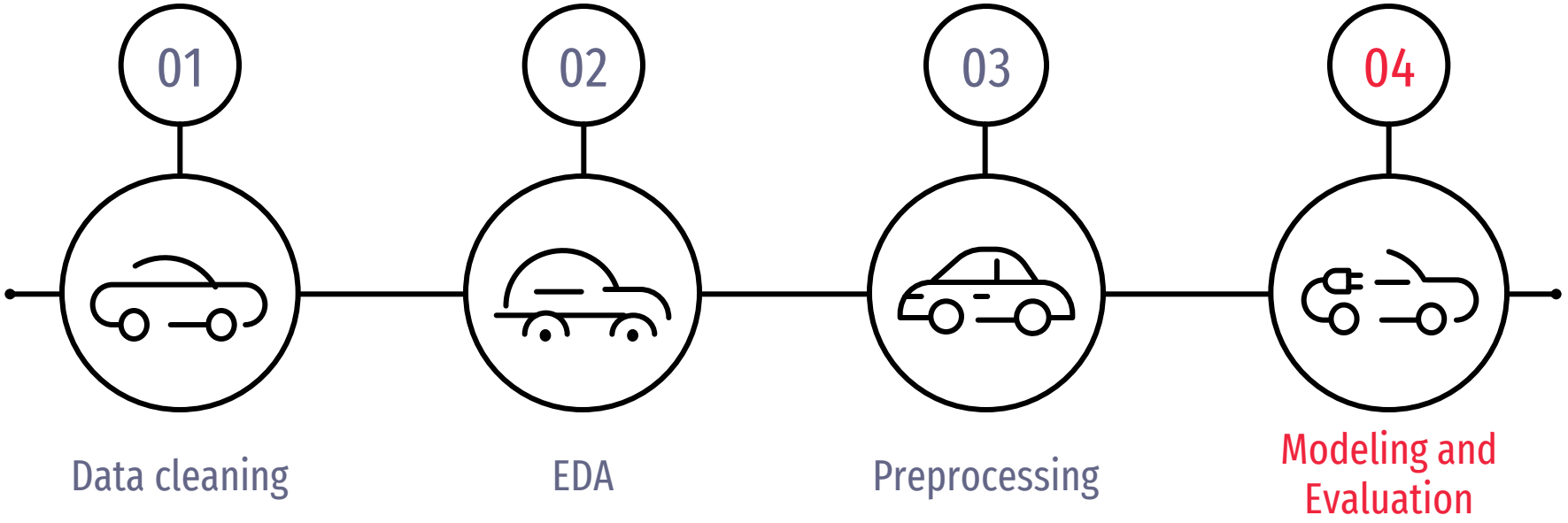

01
02
03
04
Data cleaning
EDA
Preprocessing
Modeling and Evaluation

Feature engineering and data scaling

Data for modeling

	Selling_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission_Type	Owner_Number	Mileage	Engine	Max_Power	Seats_Number	Car_Age	Car_Brand
0	13.017003	145500	Diesel	Individual	Manual	First Owner	23.40	1248.0	4.304065	5.0	1.945910	Maruti
1	12.821258	120000	Diesel	Individual	Manual	Second Owner	21.14	1498.0	4.639765	5.0	1.945910	Skoda

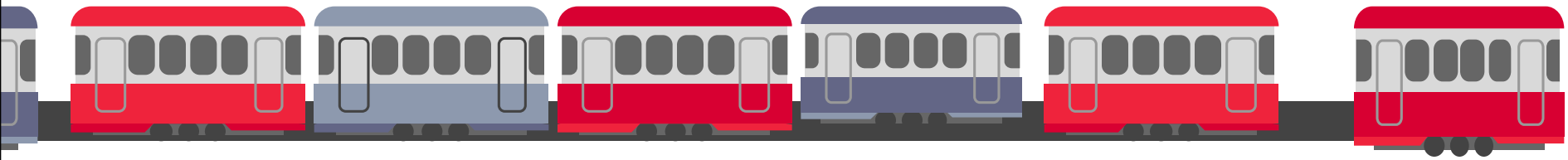
Stages


01
02
03
04
Data cleaning
EDA
Preprocessing
Modeling and Evaluation

Models

Lasso

01



02

XGBoost

03

Ridge

04

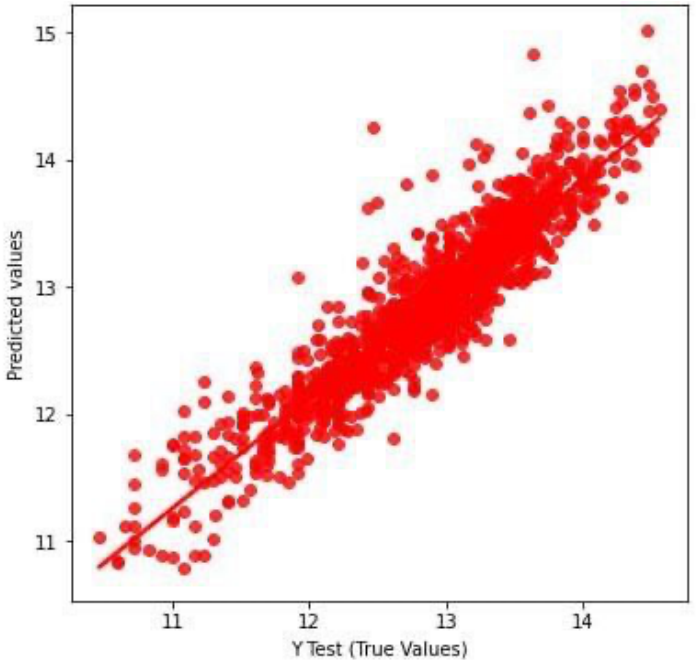
Ordinary Least Square

05

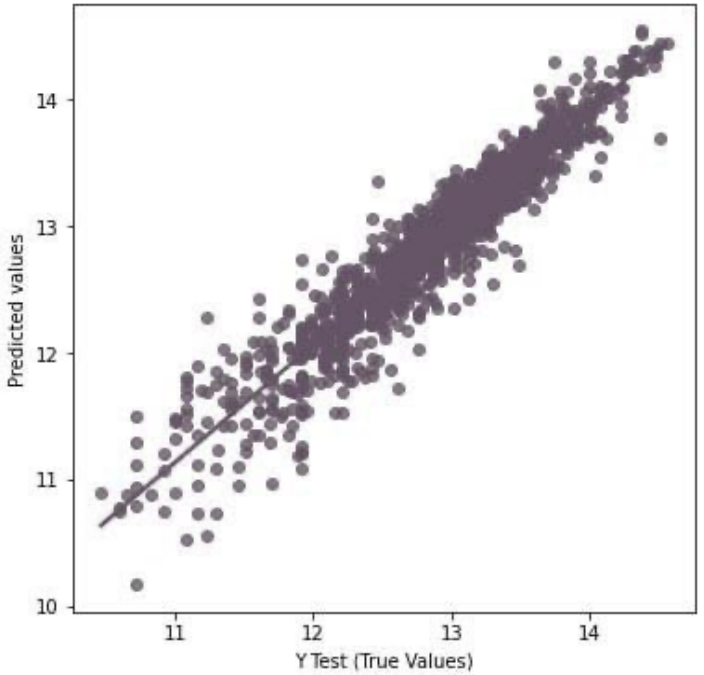
Random Forest

Comparing

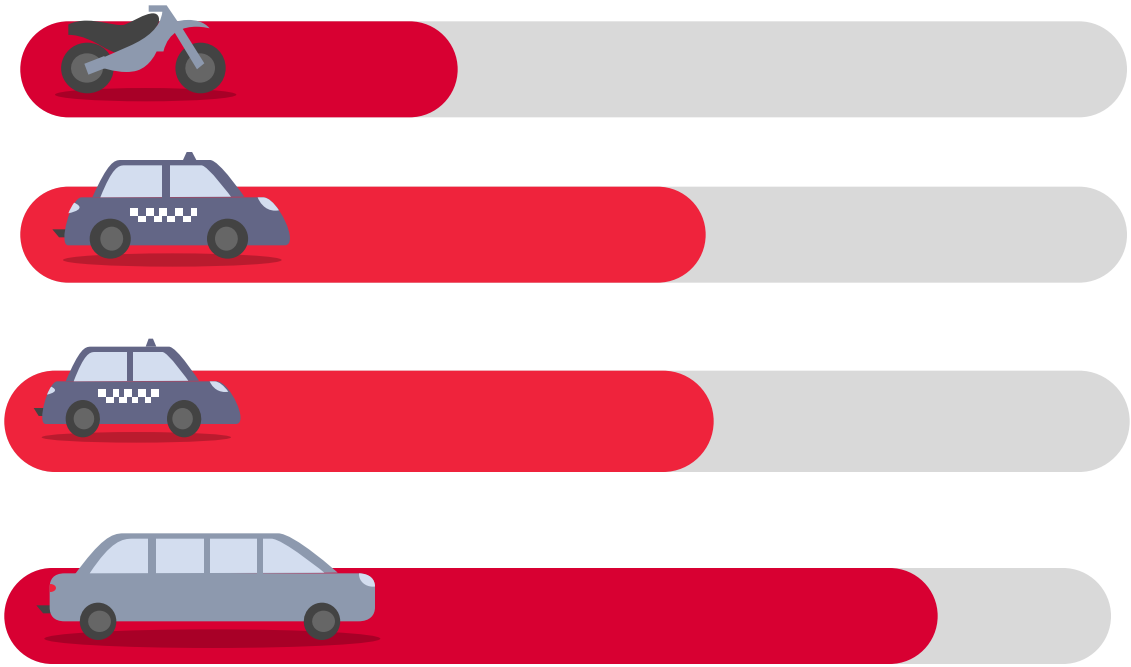
Ordinary Least Square



XGBoost



Accuracy



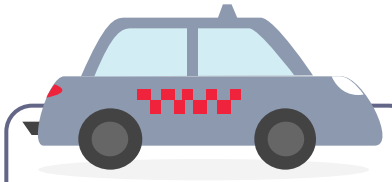
82% Linear regression (Lasso) and (Ordinary Least Square)

85% Linear regression (Radge)

91% Random Forest regression

99% XGBoost regression

Evaluation



Mean Absolute Error

16%

Root Mean Squared Error

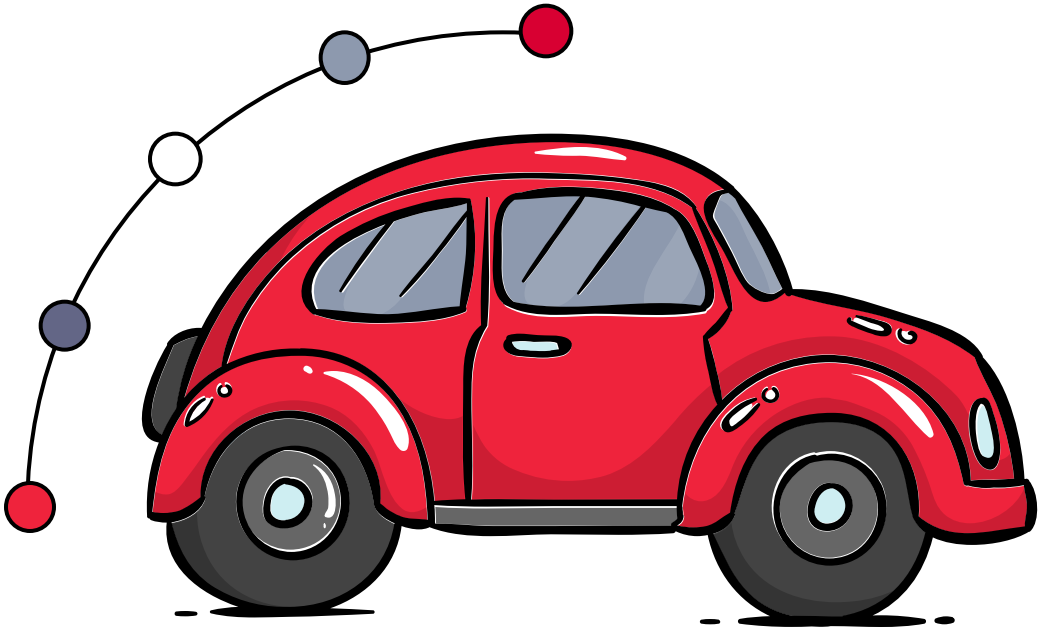
22%

Mean Squared Error

5%

R Squared Error

90%

Thanks!

Any questions?