

Predicting Used Car Prices Project

Raghad Alarifi
Reef Alturki



“

The idea behind our project is to predict used cars' market value by applying Linear regression model.



Data Sources

Cars.com

Around 10K car data
were scraped

CarGurus.com

Around 5K car data
were scraped

Additional dataset

Obtained from
Kaggle.com, contains
420K row, 22 columns

Final dataset consists of 435K rows and 11 columns

“

Difficulties



“

EDA

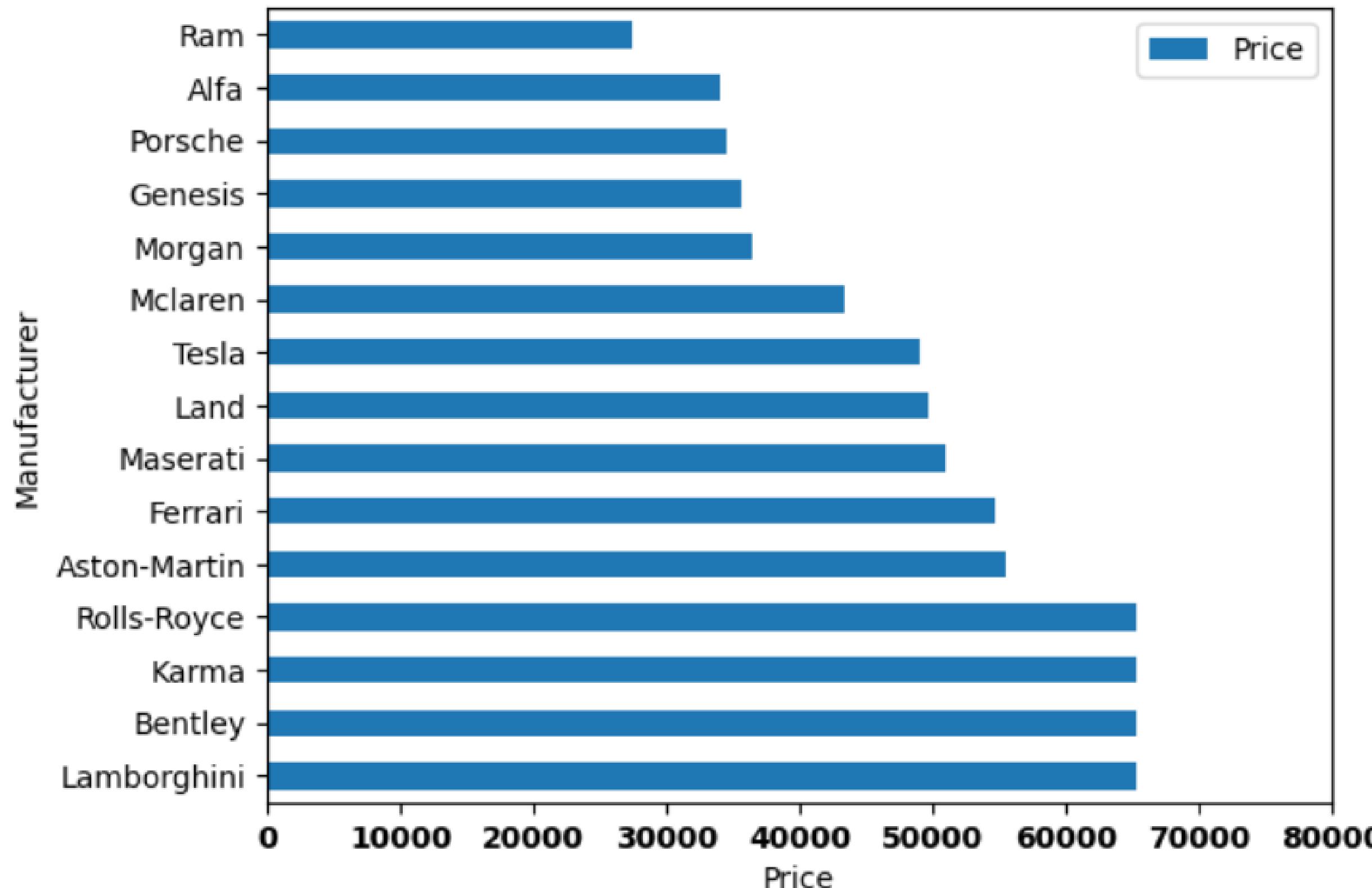
Data Cleaning

- Matching the format of the 3 datasets.
- Dropping unnecessary columns for the analysis.
- Dealing with null values.
- Detecting and removing outliers.

“

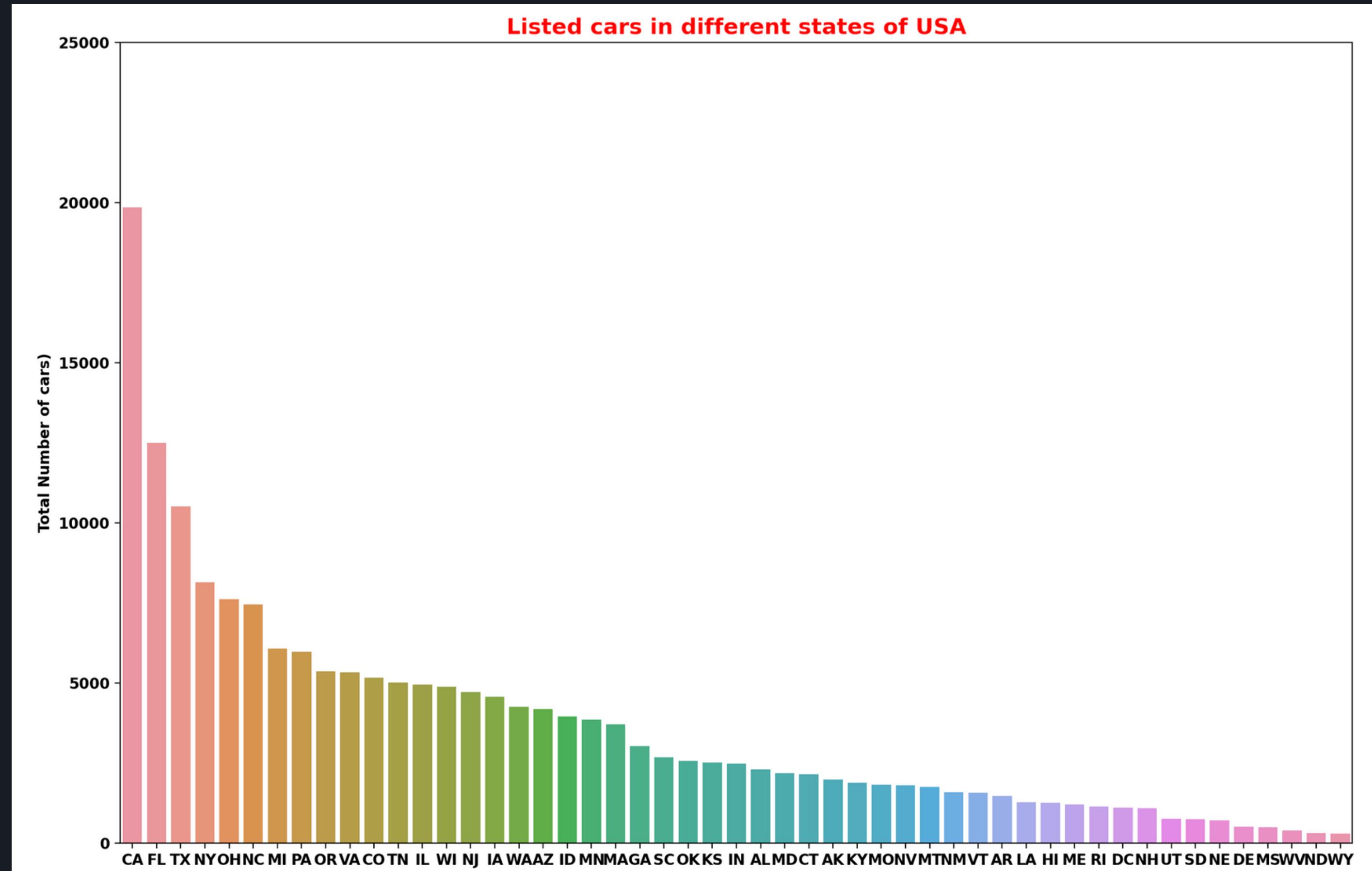
EDA

Average Price per Manufacturer



“

EDA



Final look of the data

Name	Year	Manufacturer	State	Color	Fuel type	Transmission	Engine	VIN	Mileage	Price
AMG C 43 Base 4MATIC	2019	Mercedes-Benz	CA	Black	Gasoline	Automatic	6 Cylinders	55SWF6EB8KU291212	24076.0	54392
Terrain SLT	2018	GMC	TN	Graphite Gray Metallic	Gasoline	Automatic	4 Cylinders	3GKALVEX1JL336991	45278.0	26659
Romeo Stelvio Ti	2018	Alfa	PA	Vulcano Black Metallic	Gasoline	Automatic	4 Cylinders	ZASFAKNN5J7B93222	13690.0	34998
Aviator Reserve AWD	2020	Lincoln	WI	Blue Diamond	Gasoline	Automatic	6 Cylinders	5LM5J7XC1LGL24283	3104.0	55995
XC60 T6 Momentum	2021	Volvo	WA	Fusion Red Metallic	Gasoline	Automatic	4 Cylinders	YV4A22RK1M1777654	18378.0	46880
Wrangler Unlimited Sahara	2020	Jeep	SC	Billet Silver Metallic Clearcoat	Gasoline	Automatic	4 Cylinders	1C4HJXEN7LW204339	48942.0	41889
CX-5 Touring	2020	Mazda	FL	Snowflake White Pearl	Gasoline	Automatic	4 Cylinders	JM3KFBCM3L0777706	49937.0	22939
GLC 300 Base 4MATIC	2020	Mercedes-Benz	CA	Iridium Silver Metallic	Gasoline	Automatic	4 Cylinders	WDC0G8EB5LF699139	14317.0	45991
LC 500 Base	2018	Lexus	MD	Caviar	Gasoline	Automatic	8 Cylinders	JTHHP5AY5JA003020	65360.0	67390
Fusion SEL	2020	Ford	UT	Oxford White	Gasoline	Automatic	4 Cylinders	3FA6P0CD9LR206514	48712.0	20489

“

Linear Regression

Our model achieved a score of 0.51088
and mean absolute error of 6995.07

“

Experiments

LASSO and Ridge regression

LASSO regression achieved mean absolute error of 6988.53

Ridge regression achieved mean absolute error of 6988.91

“

Thank you