



01

Estates regression project

MOHAMMAD ALMALKI
MOHAMMAD ALAJMI

INSTRUCTOR:
DR.MEJDAL ALQAHTANI



02

Introduction and main goal

The main goal of this project is to study the Estates market and to train a model to predict the prices of the estates of and testing it using regression models.

WEB SCRAPING

The used DataSet in this project is derived from “Aqar” website, the data collected using web scraping library called beautiful soap

ABOUT THE DATASET

it is about the prices of estates in Riyadh city, the data has about 40 thousand row before the cleaning and about 7 features.

THE FEATURES OF DATASET

The columns describe a name, area, neighborhood within Riyadh, the number of bedrooms, bathrooms and the price of the Estate.

The DataSet

03



The columns describe a name, area in Square and the neighborhood within Riyadh. The columns also describe the number of bedrooms, the number of bathrooms and the price of the Estate.

04

The Dataset before the cleaning.

	Estate	Area	Roams	Price	Baths	city	Type
0	...فيلات للبيع في شارع البخموري ، حي عطيشة ، الريا	355 م²	3	ريال 1,900,000	3	riyad	villa
1	...فيلات للبيع في شارع رقم 357 ، حي الملقا ، الريا	500 م²	5	ريال 4,700,000	5	riyad	villa
2	...فيلات للبيع في شارع الهنوف ، حي الصحافة ، الريا	500 م²	2	ريال 3,300,000	4	riyad	villa
3	...فيلات للبيع في شارع جعفر بن احمد ، حي الدار الب	230 م²	5	ريال 880,000	5	riyad	villa
4	...فيلات للبيع في حي عكاظ ، الرياض ، الرياض	325 م²	5	ريال 1,100,000	5	riyad	villa
...
39573	...فيلات للبيع في شارع وادي نوار ، حي الملقا ، الر	875 م²	5	ريال 8,000,000	5	riyad	villa
39574	...فيلات للبيع في طريق العزيزية ، حي الدار البيضاء	280 م²	4	ريال 1,100,000	5	riyad	villa
39575	...فيلات للبيع في شارع محمد بن حمد بن حامد ، حي اش	243 م²	5	ريال 1,700,000	5	riyad	villa
39576	...فيلات للبيع في شارع بحر العرب ، حي اشيلية ، ال	324 م²	4	ريال 1,950,000	4	riyad	villa
39577	...فيلات للبيع في شارع زين الدين الدابلسي ، حي الر	250 م²	4	ريال 1,250,000	4	riyad	villa

	Estate	Area	Roams	Price	Baths	city	Type	street	Neighborhood
0	فيلا للبيع في شارع البغموري حي طليشة	355	3	1,900,000	3	riyad	villa	البغموري حي طليشة	طليشة
1	فيلا للبيع في شارع رقم 357 حي الملقا	500	5	4,700,000	5	riyad	villa	رقم 357 حي الملقا	الملقا
2	فيلا للبيع في شارع الهفوف حي الصحافة	500	2	3,300,000	4	riyad	villa	الهفوف حي الصحافة	الصحافة
3	فيلا للبيع في شارع جعفر بن احمد حي الدار البيضاء	230	5	880,000	5	riyad	villa	جعفر بن احمد حي الدار البيضاء	الدار البيضاء
5	فيلا للبيع في شارع رقم 254 حي النرجس	450	5	3,400,000	5	riyad	villa	رقم 254 حي النرجس	النرجس
...
39572	فيلا للبيع في شارع يونس الوفراوندي حي طويرق	300	5	1,200,000	5	riyad	villa	يونس الوفراوندي حي طويرق	طويرق
39573	فيلا للبيع في شارع وادي نوار حي الملقا	875	5	8,000,000	5	rivad	villa	وادي نوار حي الملقا	الملقا

After cleaning

ADDED FEATURES:

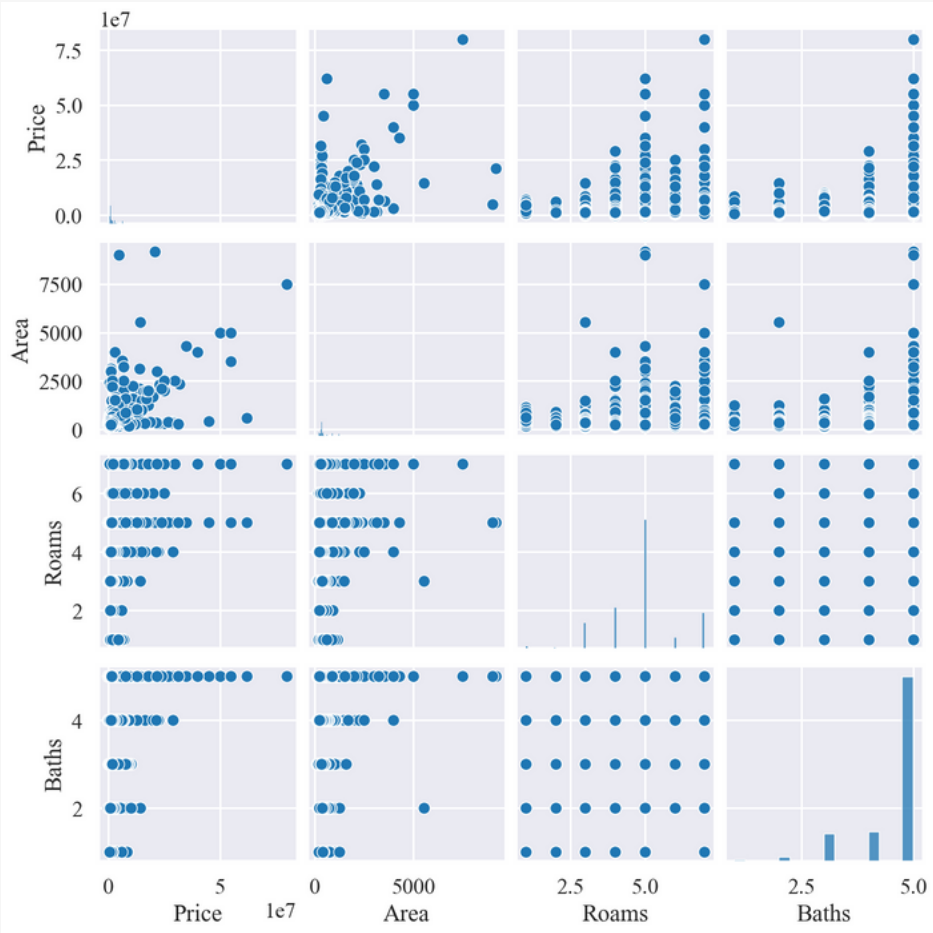
THE ESTATE NAME BROKE INTO TWO FEATURES

1-STREET OF THE ESTATE
2-THE NEIGHBORHOOD OF THE ESTATE

THE NAN'S VALUES OF NEIGHBOHOOD WERE DROPE,ALSO THE OBJECT VALUES TURNED INTO INT VALUES

the relation between features

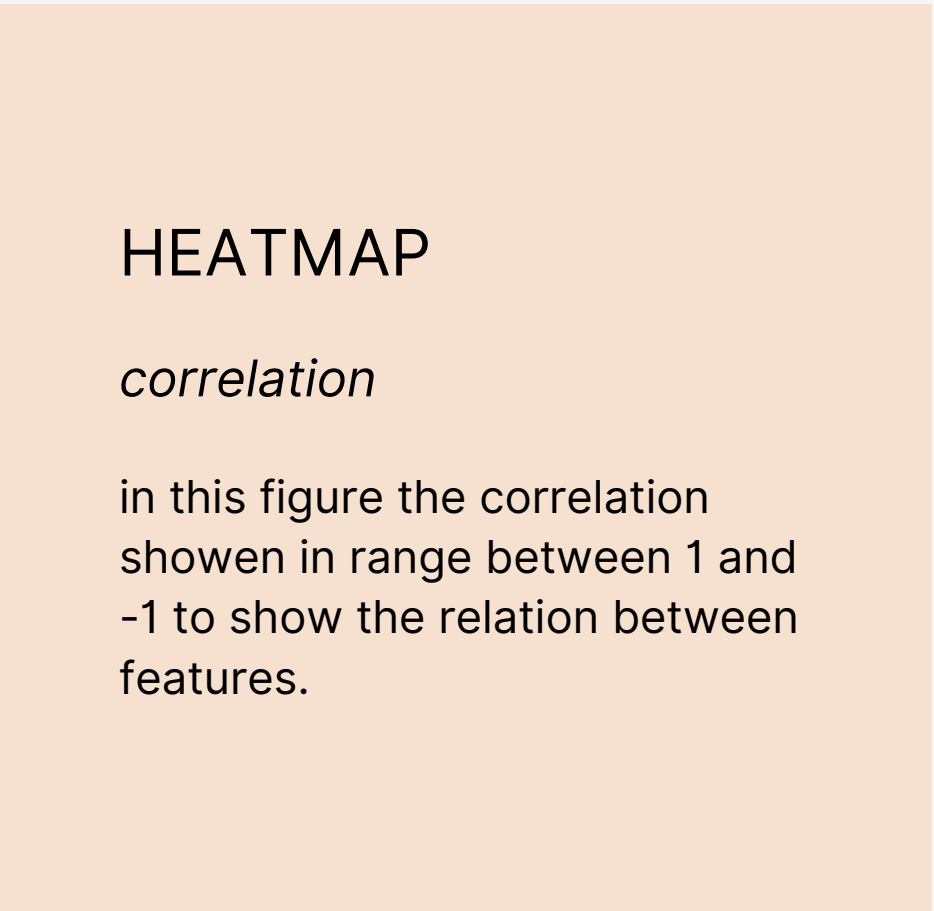
THE RELATION BETWEEN FEATURES



PAIRPLOT

increases and decreases?

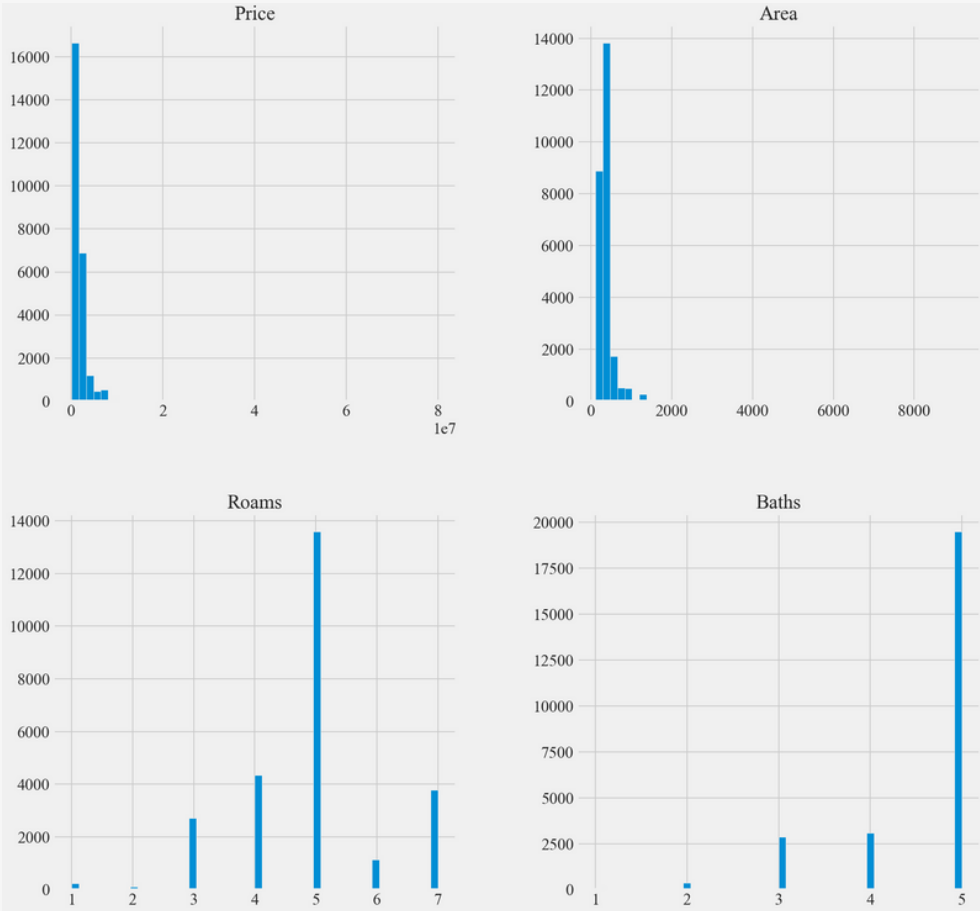
this figure shows there is an increase between two feature, the area and price has a relation.



HEATMAP

correlation

in this figure the correlation shown in range between 1 and -1 to show the relation between features.

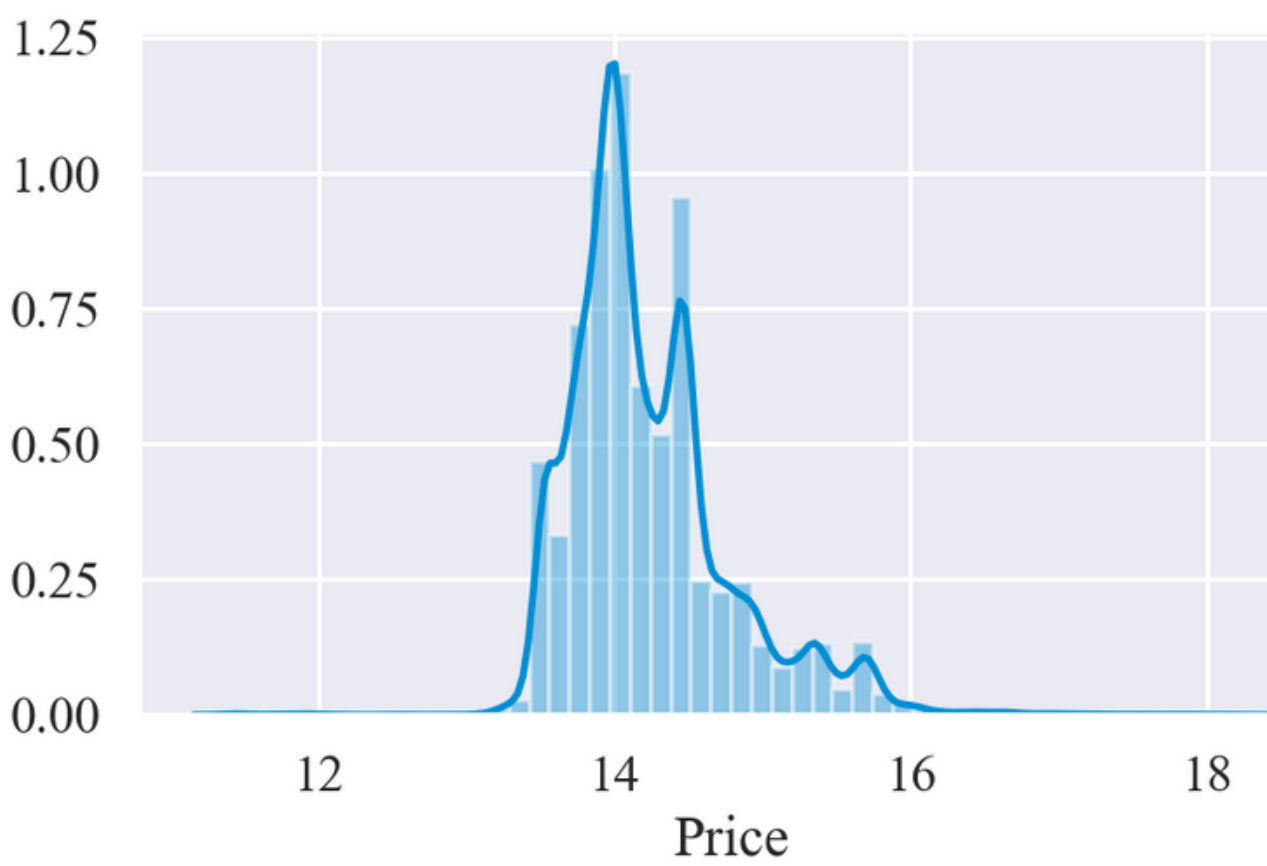


HIST PLOT

the distribution of features

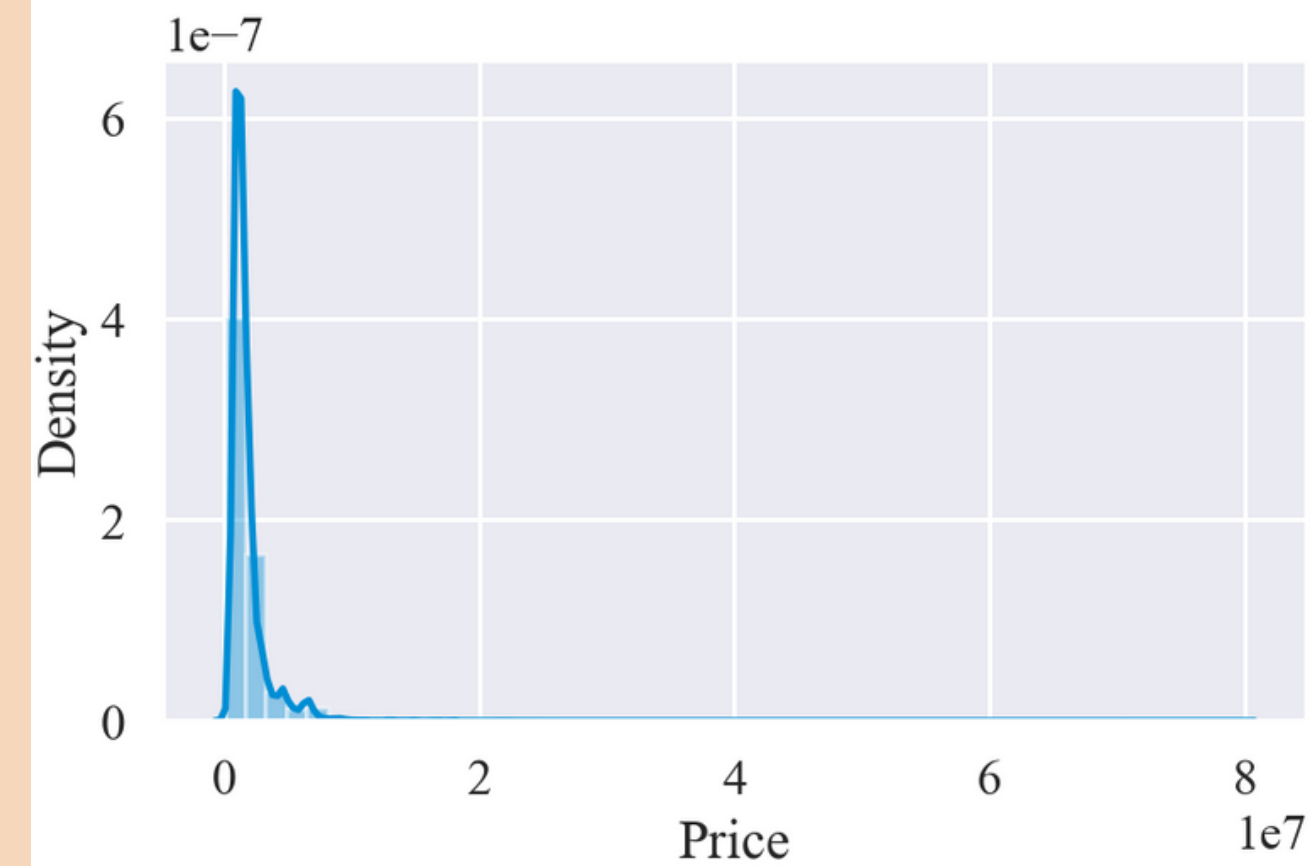
this figure shows the distribution of every feature values.

log transform



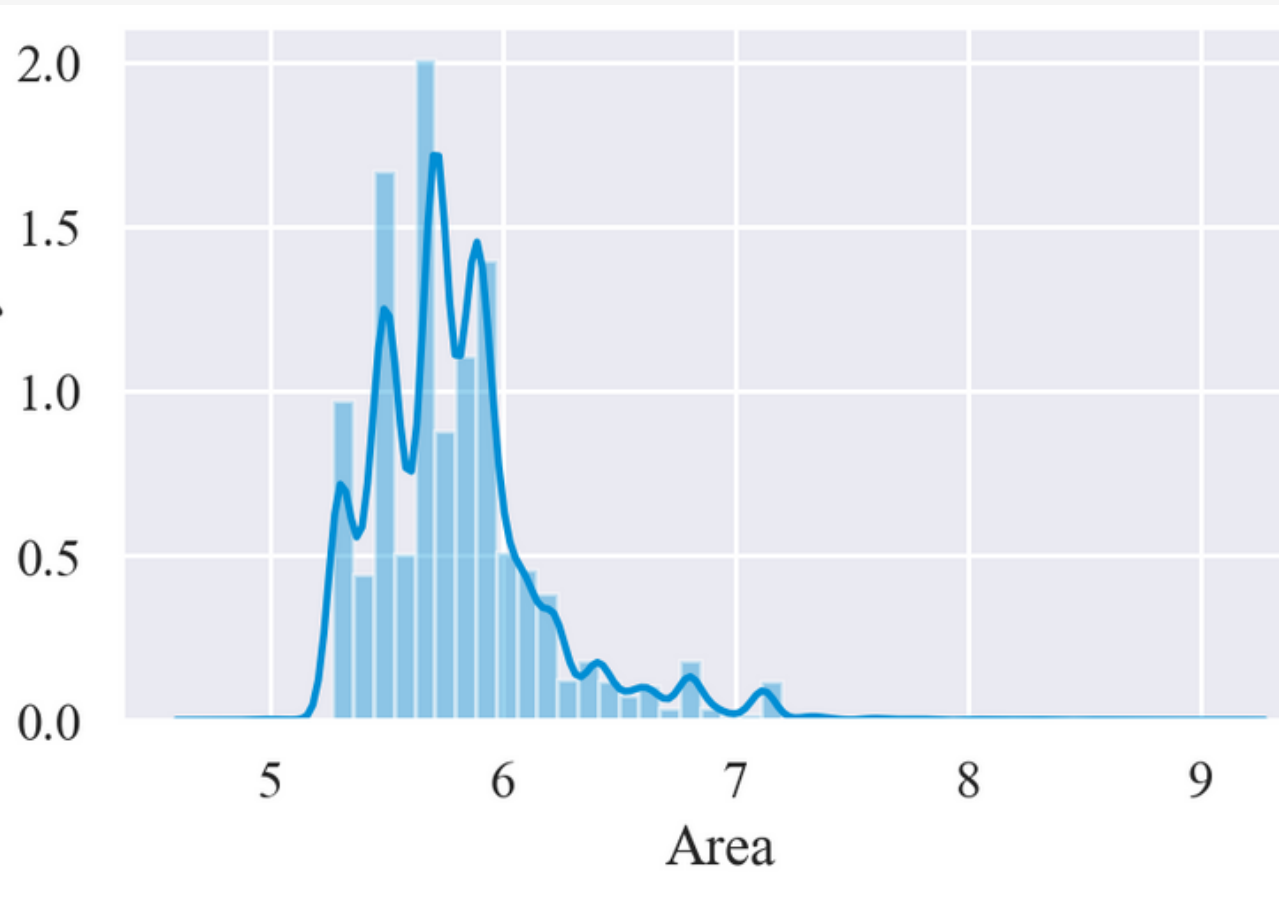
After

Price distribution
before and after
log transform



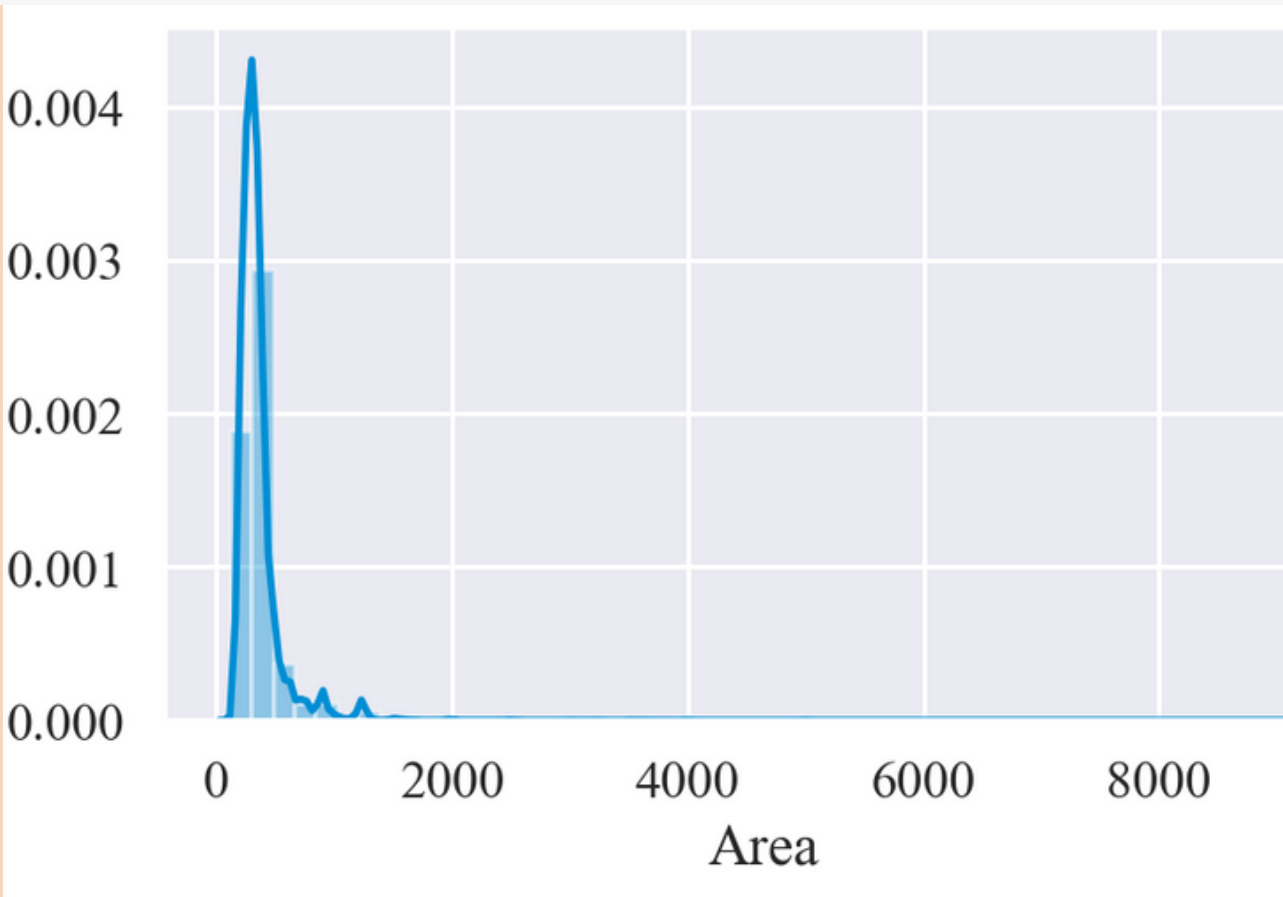
Before

log transform



After

Area distribution
before and after
log transform

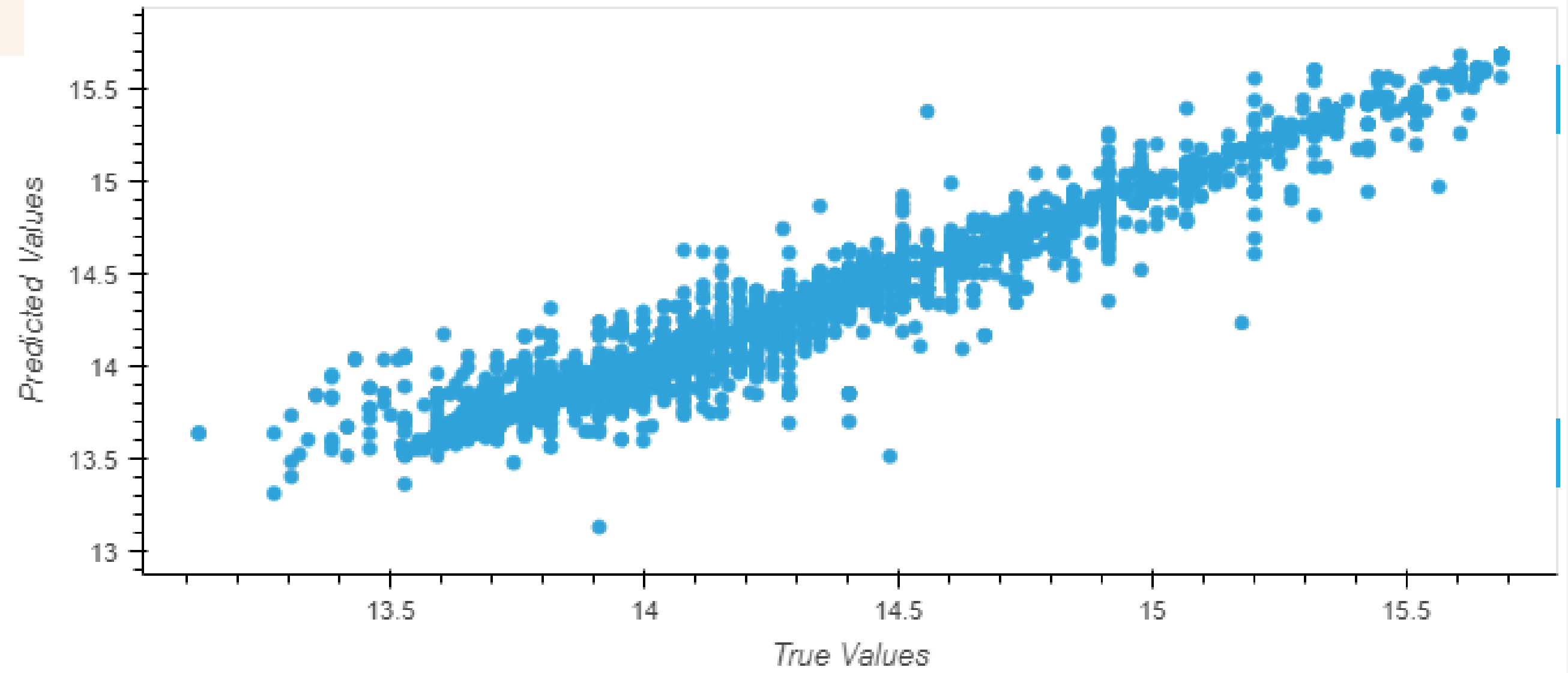


Before

The results of the prediction

This figure represented
by scatter plot it shows
the predicted values and
the true values

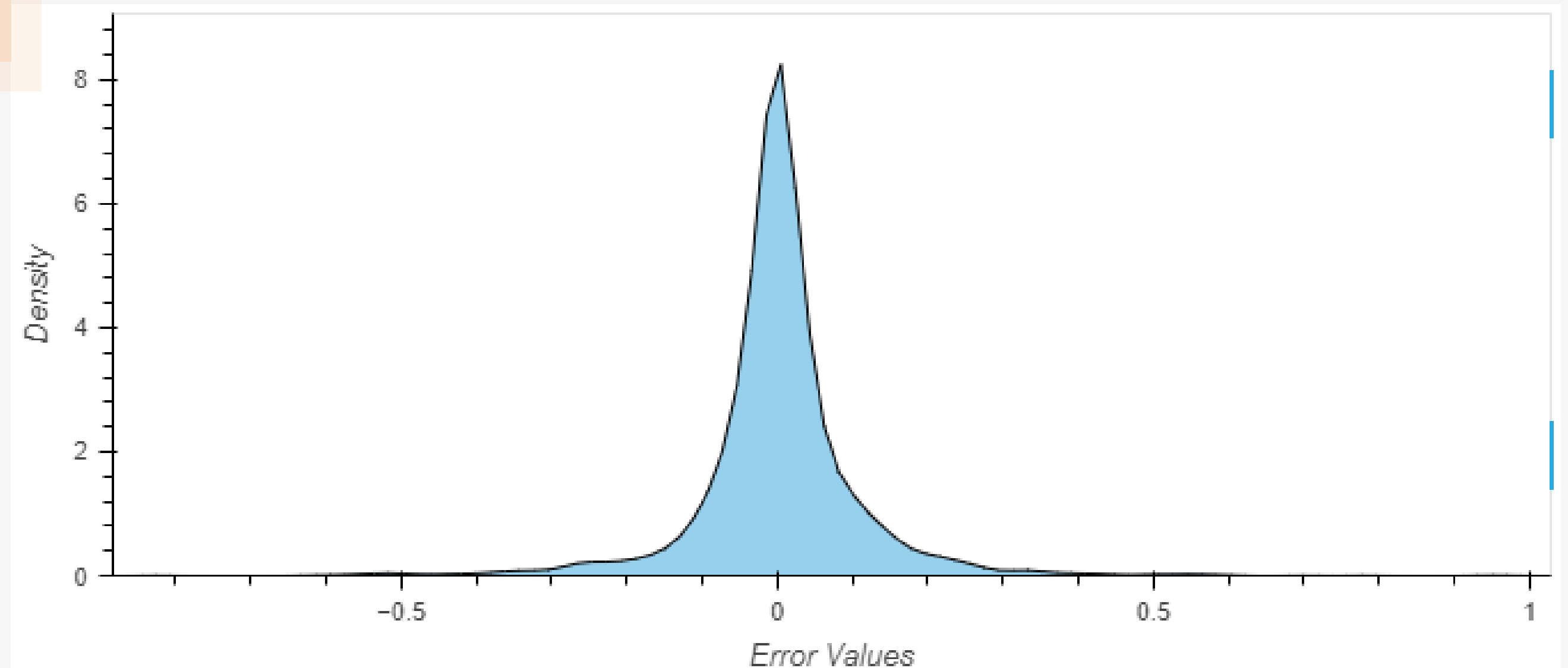
The results of the prediction



The density of the error values

This figure represented
by hv plot it shows the
density of the error
values

The density of the error values



comparison between models

the DataSet applied
on five models

11

	Model	MAE	MSE	RMSE	R2 Square	Cross Validation
0	Linear Regression	0.098681	0.020143	0.141925	0.918249	0.912443
1	Ridge Regression	0.098787	0.020157	0.141975	0.918192	0.912469
2	Lasso Regression	0.261338	0.110437	0.332320	0.551785	-0.023750
3	Polynomail Regression	0.096210	0.019465	0.139515	0.921002	0.000000
4	Random Forest	0.061577	0.011478	0.107138	0.953414	0.952802

ERROR VALIDATION TECHNIQUES

the error values measured in five methods
1-MAE
2-MSE
3-RMSE
4-R2 square
5-Cross Validation

WHAT IS THE BEST MODEL
THAT PREDICT THE ESTATE
MARKET PRICES?

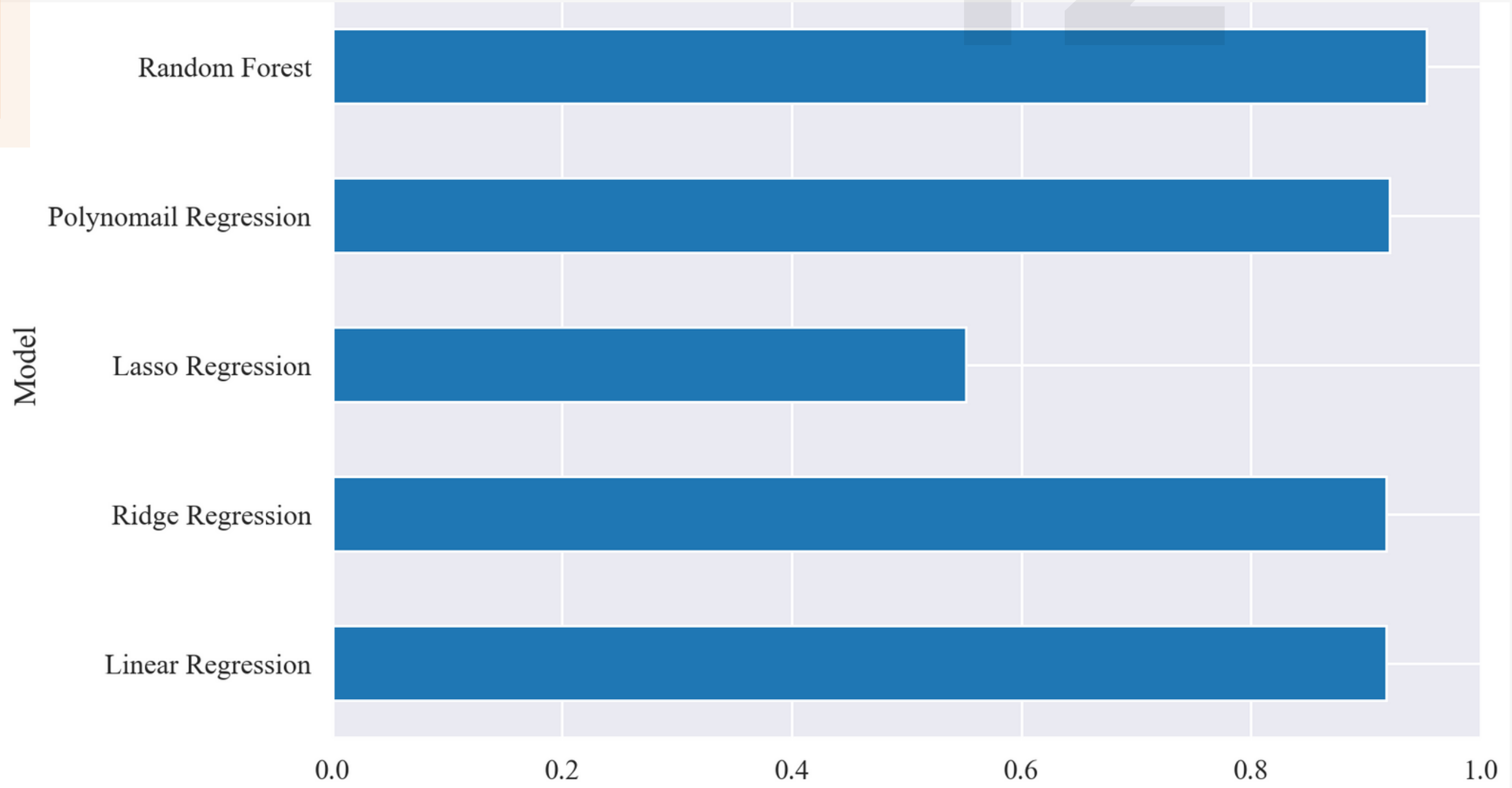
it appears that random forest model
had the best result with the highest
accuracy with 0.95 in R2 square.

comparison between
regression models

compare between the different models results

This figure represented
by Barh it shows the
comparasion between the
different models results

compare between the different models results



CONCLUSION

13

WHAT IS THIS PROJECT ABOUT?

the project aims to build a model that predict the prices of Riyadh estates using regression models.

THE DATASET

the DataSet derived from Aqar website, the data focuses on the features of Riyadh estates

THE BEST MODEL TO ACHIEVE THIS GOAL

it appears that random forest model had the best result with the highest accuracy with 0.95 in R2 square.



thank you

14

Thank you for
listening
<3

feel free to ask qeustions

