T5 Bootcamp

REAL ESTATE

PRICES PREDICTION

Presented by: Saleh , Samer

TABLE OF CONTENTS

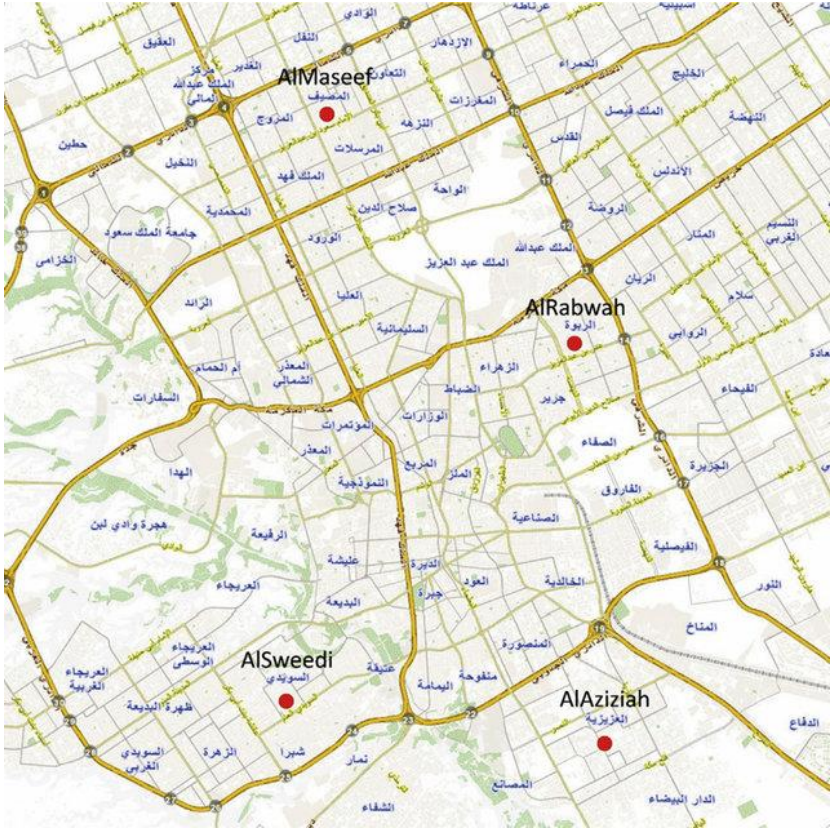
Introduction

Problem

In recents days, the price of houses in Riyadh city is a hot topic, the prices changes monthly and yearly because of the high demand. Most of the houses owners trying to set a selling price for their property does not know what is the best price which make them set a high price or to low.

Our solution

In this project, Our goal is to develop a linear regression model to predict the property prices according to area, rooms, bathrooms, and location. Which will help the owners to set the best price.



Approach and Methodology

1- We took the data from Aqar website.

2- We did a data cleaning , removing the extra space of the heading column and removing duplicate, null values and outliers.

3- We use these libraries:

- Pandas
- Numpy
- Patsy
- Scipy.stats
- Statsmodels
- LinearRegression, Ridge
- train_test_split, cross_val_score

4- Feature Enginerring

5- Feature Selection

Data

Data Source

We have used Web Scrapping to collect our data and the source was Aqar website.

Data columns

- AREA
- ROOMS
- APARTMENTS
- BEDROOMS
- BATHROOMS
- WIDTH OF STREET
- LIVING ROOMS
- PRICE

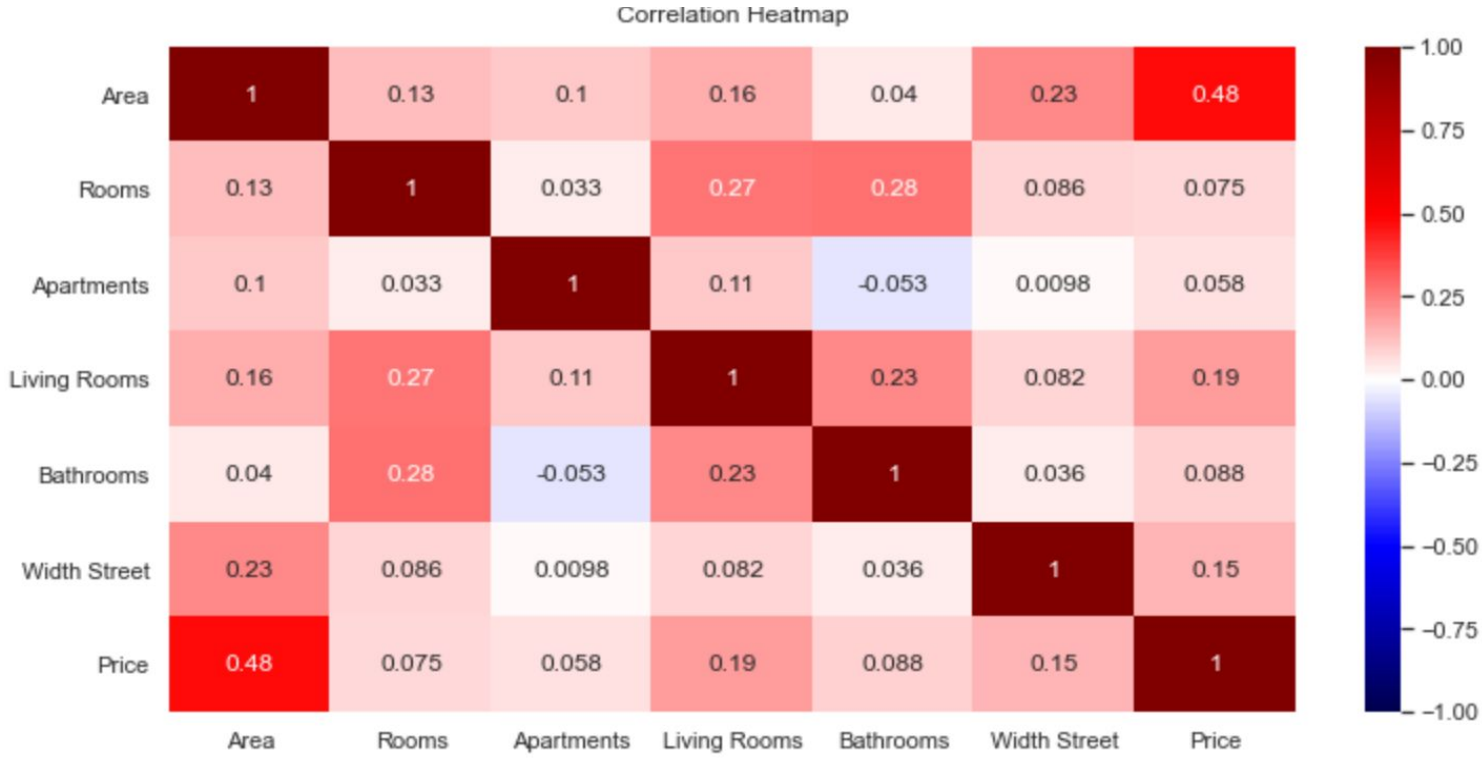
Data

	Area	Front	Rooms	Apartments	Living Rooms	Bathrooms	Width Street	Age	Price
0	360	شمال	5	2	3	5	15	new	1200000
1	404	غرب	7	1	1	4	16	new	1200000
2	375	شمال	7	2	2	5	15	new	1900000
3	300	شمال	7	1	3	5	18	new	1350000
4	300	شرق	7	1	3	5	15	new	1300000

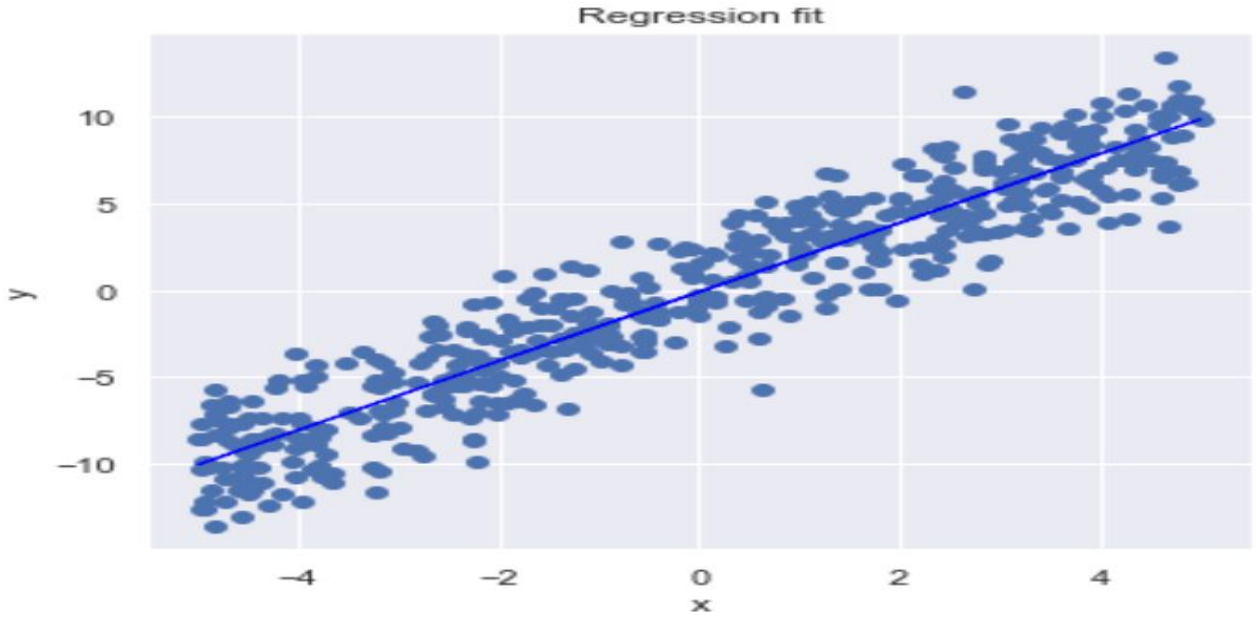
Data

	Area	Front	Rooms	Apartments	Living Rooms	Bathrooms	Width Street	Age	Price
0	360	4	5	2	3	5	15	2	1200000
1	404	3	7	1	1	4	16	2	1200000
2	375	4	7	2	2	5	15	2	1900000
3	300	4	7	1	3	5	18	2	1350000
4	300	2	7	1	3	5	15	2	1300000

Relations


Correlation Heatmap
Area 1 0.13 0.1 0.16 0.04 0.23 0.48
Rooms 0.13 1 0.033 0.27 0.28 0.086 0.075
Apartments 0.1 0.033 1 0.11 -0.053 0.0098 0.058
Living Rooms 0.16 0.27 0.11 1 0.23 0.082 0.19
Bathrooms 0.04 0.28 -0.053 0.23 1 0.036 0.088
Width Street 0.23 0.086 0.0098 0.082 0.036 1 0.15
Price 0.48 0.075 0.058 0.19 0.088 0.15 1
Area Rooms Apartments Living Rooms Bathrooms Width Street Price
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00

Linear Regression Model



Results

	Linear Regression	Ridge Regression
X (ALL)	R^2 = 0.67 MAE= 0.328 MSE= 0.442 RMSE= 0.779	R^2 = 0.77 MAE= 0.229 MSE=0.362 RMSE= 0.563
Age, Front, and Apartments.	R2 = 0.7 MAE= 0.549 MSE= 0.462 RMSE= 0.672	R^2 = 0.8 MAE= 0.432 MSE= 0.729 RMSE= 0.332
Age, Front, and Apartments, Rooms, and Bathrooms.	R^2 = 0.8 MAE= 0.629 MSE= 0.482 RMSE= 0.662	R^2 = 0.83 MAE= 0.542 MSE= 0.442 RMSE= 0.771

Conclusion

We have received a good results model and from using updated and real data. we develop two models linear and ridge. Both of these model has a very good results as a start

THANK YOU