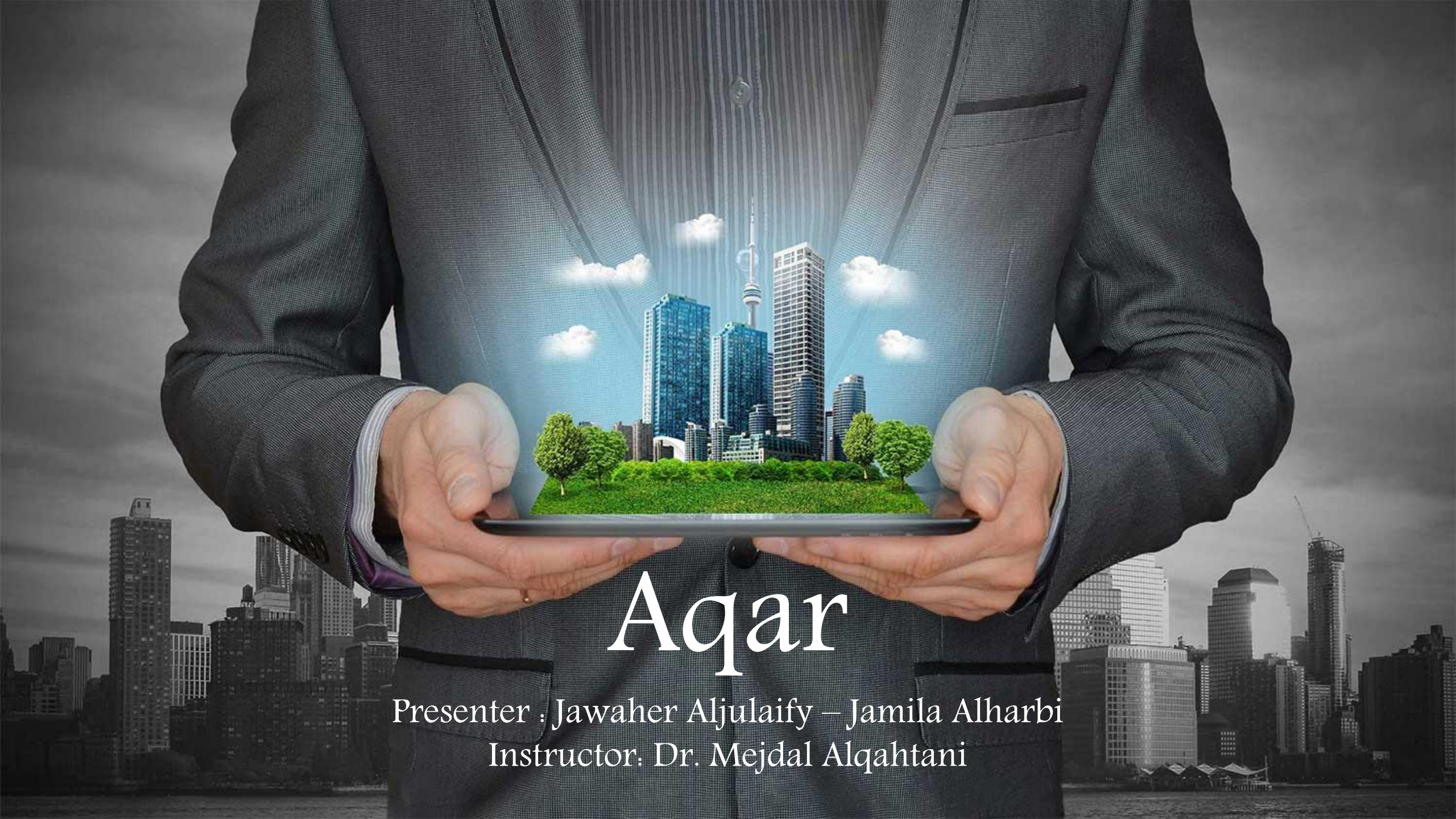
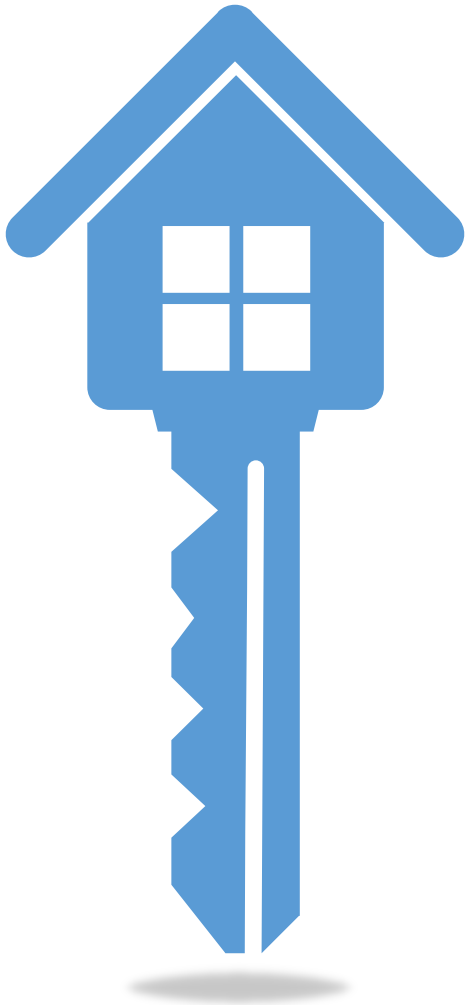


Aqar

Presenter : Jawaher Aljulaify – Jamila Alharbi
Instructor: Dr. Mejdal Alqahtani

Introduction



Due to the information gap about house prices and the difference in house prices from one neighborhood to another, the individual may face a problem estimating house prices and may be subject to deception by real estate owners.

Therefore, we will build a model that predicts house prices in various neighborhoods in Saudi Arabia's capital Riyadh based on several features.

Data Structure

	Price	Size	n_rooms	n_bathroom	Districts
0	750,000 ريال	240 م²	7	5	Aldar-Albaida
1	6,500,000 ريال	1237 م²	7	5	Alhamra
2	1,650,000 ريال	375 م²	5	5	street
3	6,000,000 ريال	625 م²	5	5	Almalga
4	1,100,000 ريال	325 م²	5	5	Okath
...	...	...	...	...	...
21973	1,400,000 ريال	510 م²	5	5	Alrawdah
21974	2,000,000 ريال	623 م²	7	4	Alrawdah
21975	1,110,000 ريال	330 م²	5	5	NaN
21976	800,000 ريال	200 م²	1	1	Aljaradeh
21977	1,300,000 ريال	235 م²	5	5	NaN

21978 rows × 5 columns

We have 21978 rows and 5 columns

	Price	Size	n_rooms	n_bathroom	Districts
5754	4000100.0	525.0	7	5	Al-Narjes
11379	8000000.0	1185.0	7	5	Alhamra
12981	1900000.0	355.0	3	3	Aleshah
21309	1800000.0	1050.0	7	5	Bader
6978	1400000.0	312.0	5	5	Alhazem
...	...	...	...	...	...
16850	800000.0	200.0	1	1	Aljaradeh
6265	1200000.0	250.0	5	5	Tuwaiq
11284	800000.0	200.0	1	1	Aljaradeh
5390	4700000.0	500.0	5	5	Almalga
15795	1400000.0	330.0	5	3	Al-Remal

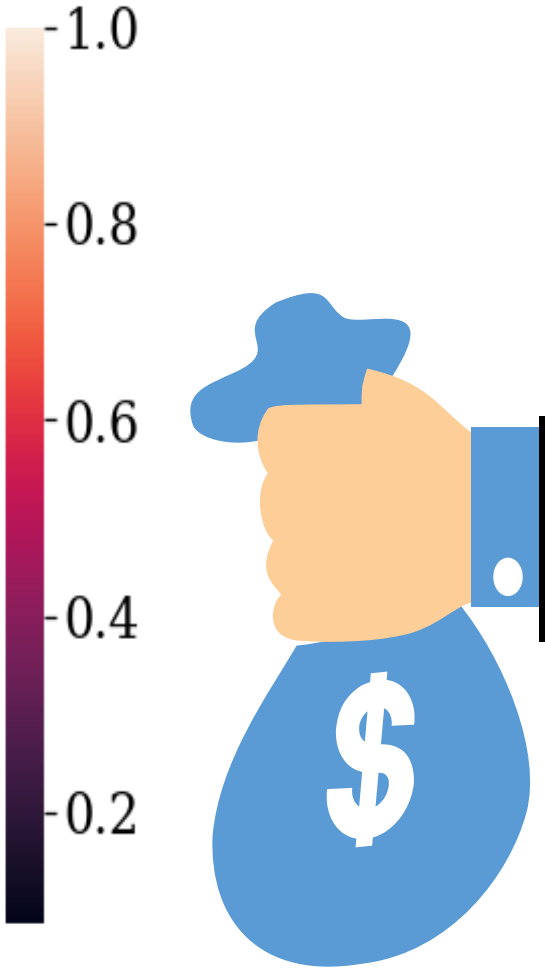
11073 rows × 5 columns

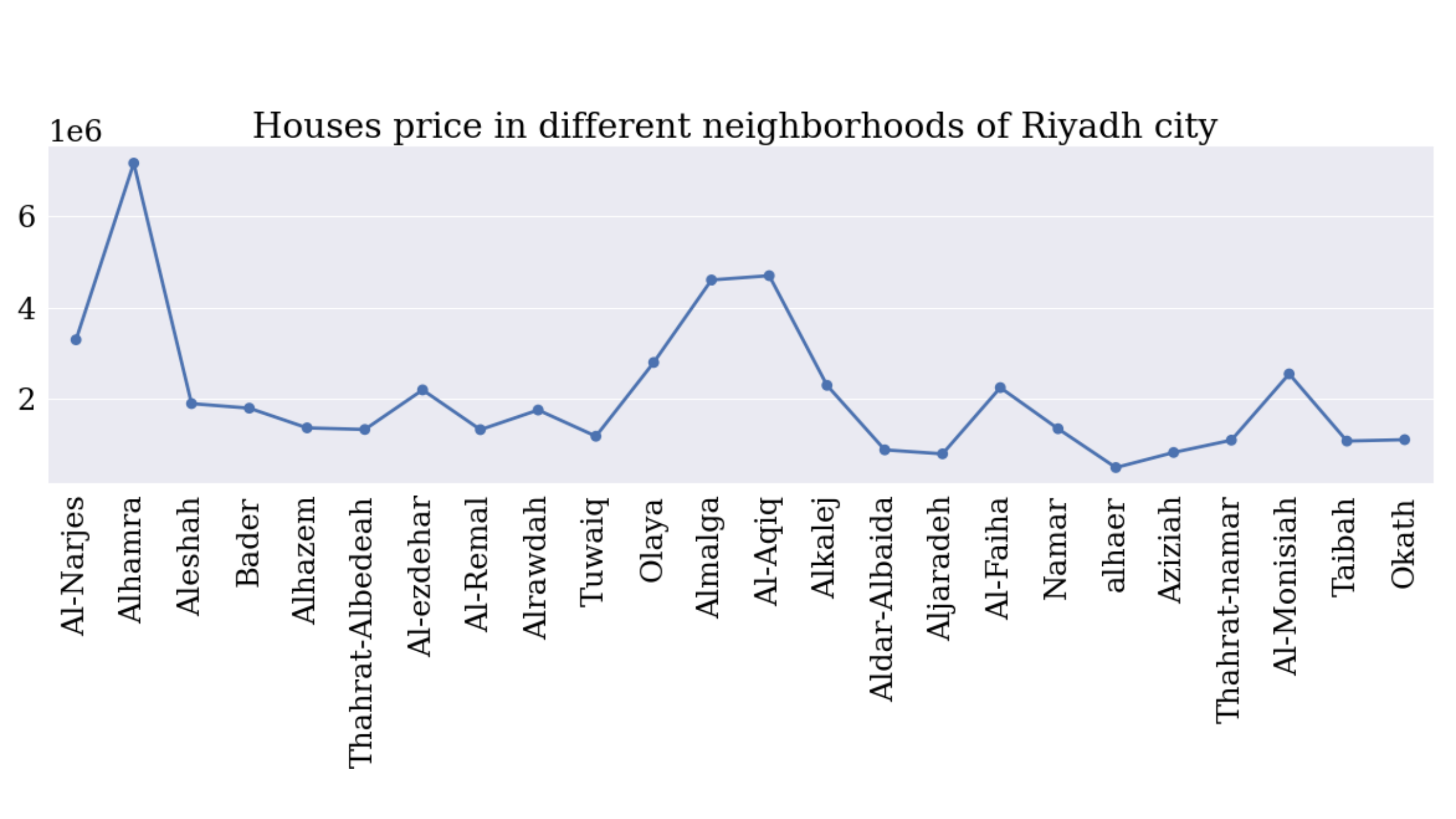
We have 11073 rows and 5 columns

After cleaning In Training set

Correlation

	Price	Size	n_rooms	n_bathroom
Price	1	0.74	0.086	0.21
Size	0.74	1	0.33	0.13
n_rooms	0.086	0.33	1	0.56
n_bathroom	0.21	0.13	0.56	1




Houses price in different neighborhoods of Riyadh city
1e6
6
4
2
Al-Narjes
Alhamra
Aleshah
Bader
Alhazem
Thahrat-Albedeah
Al-ezdehar
Al-Remal
Alrawdah
Tuwaiq
Olaya
Almalga
Al-Aqiq
Alkalej
Aldar-Albaida
Aljaradeh
Al-Faiha
Namar
alhaer
Aziziah
Thahrat-namar
Al-Monisiah
Taibah
Okath

Cross Validation

Before Converting categorical values

```
Linear mean cv r^2: 0.643114 +- 0.020
Ridge mean cv r^2: 0.643114 +- 0.020
poly mean cv r^2: 0.697513 +- 0.018
```

Cross Validation

After Converting categorical values

```
Linear mean cv r^2: 0.977385 +- 0.000
Ridge mean cv r^2: 0.977383 +- 0.000
poly mean cv r^2: 0.999262 +- 0.000
```

Then we have 11073 rows and
27 columns

Cross Validation

After Adding Log for Price

```
Linear mean cv r^2: 0.993010 +- 0.000
Ridge mean cv r^2: 0.993009 +- 0.000
poly mean cv r^2: 0.999976 +- 0.000
```

Cross Validation

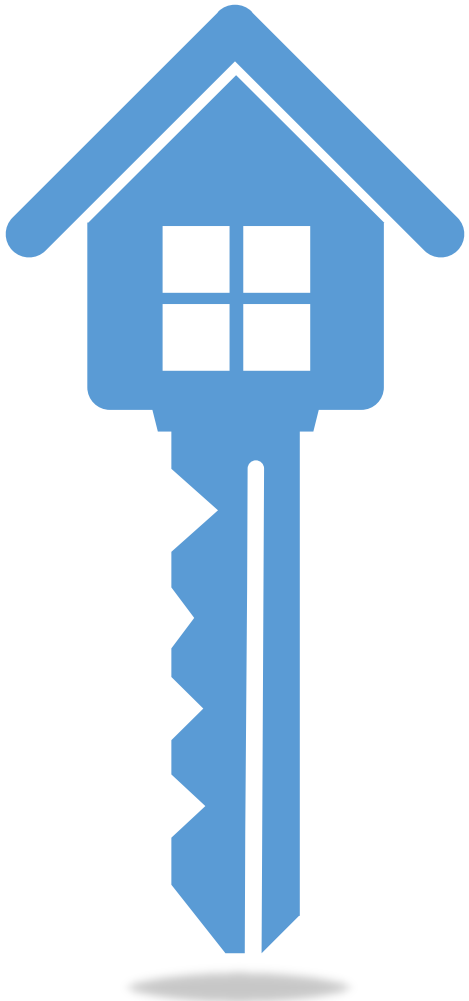
After Adding more feature in another data set

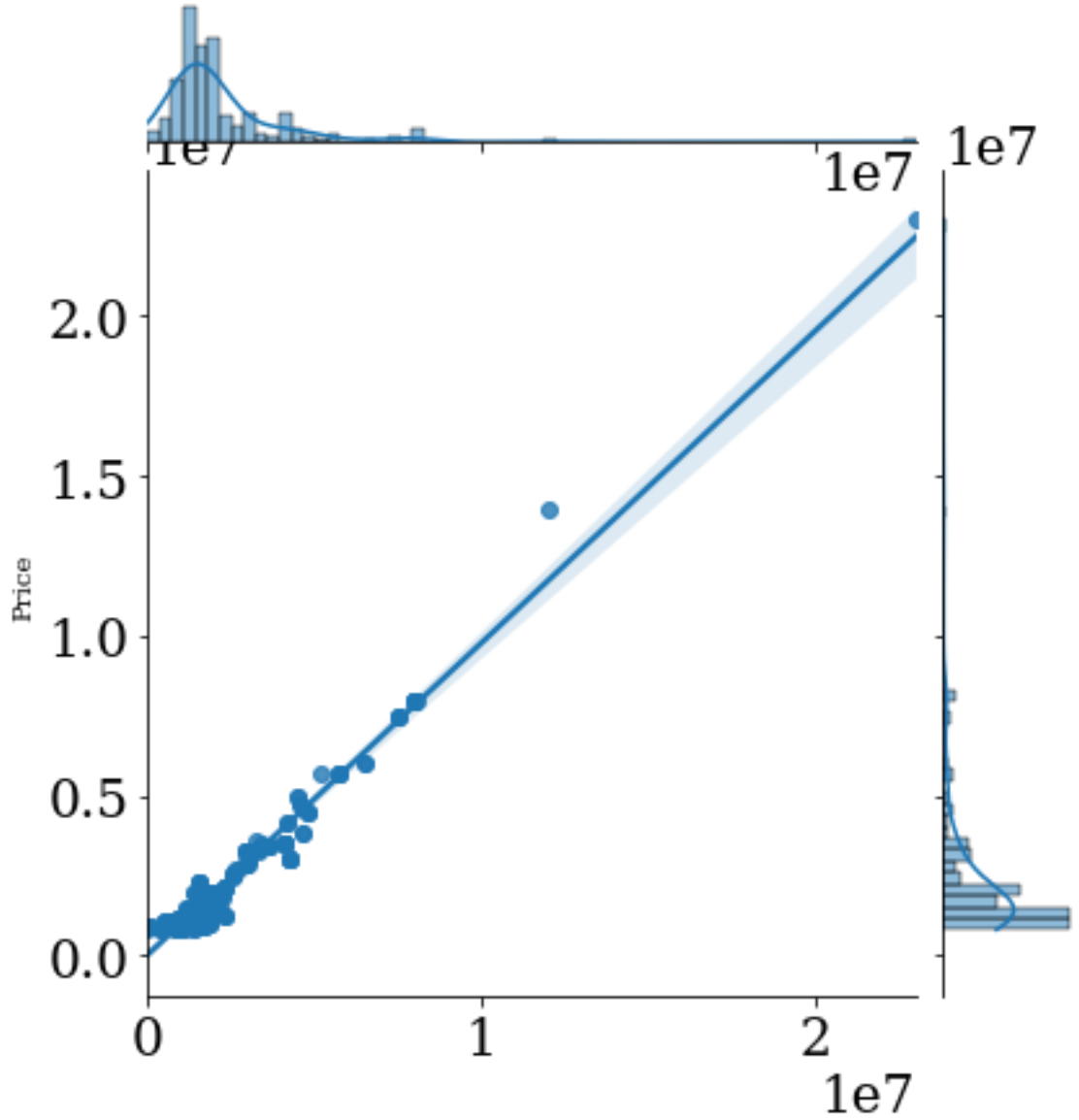
```
Linear mean cv r^2: 0.960574 +- 0.020
Ridge mean cv r^2: 0.960662 +- 0.020
poly mean cv r^2: 1.000000 +- 0.000
```

	Price	Districts	Area	Front	Bedrooms	Salon	Street_width	Age
0	1650000.0	AlArid	200.0	north	7	1	20.0	5.0
2	1270000.0	AlRemal	285.0	west	4	1	5.0	20.0
3	1250000.0	AlKaleej	560.0	north	5	3	3.0	18.0
6	8000000.0	Al-Hamra	1185.0	west	7	1	5.0	25.0
10	2000000.0	AlRemal	378.0	north	7	2	5.0	15.0
...	...	...	...	...	...	...	...	...
2077	1350000.0	Al-Naseem	312.0	east	7	3	5.0	10.0
2078	1300000.0	Al-Muraba	498.0	south-east	6	5	10.0	35.0
2080	1270000.0	AlRemal	285.0	west	4	1	5.0	20.0
2081	900000.0	AlDarAlbaida	230.0	south-west	7	3	20.0	0.0
2083	2000000.0	Al-Rawdah	623.0	north	7	2	5.0	36.0

1336 rows × 8 columns

2084 rows and 9 columns




1e7
1e7
1e7
2.0
1.5
1.0
0.5
0.0
Price
0
1
2
1e7

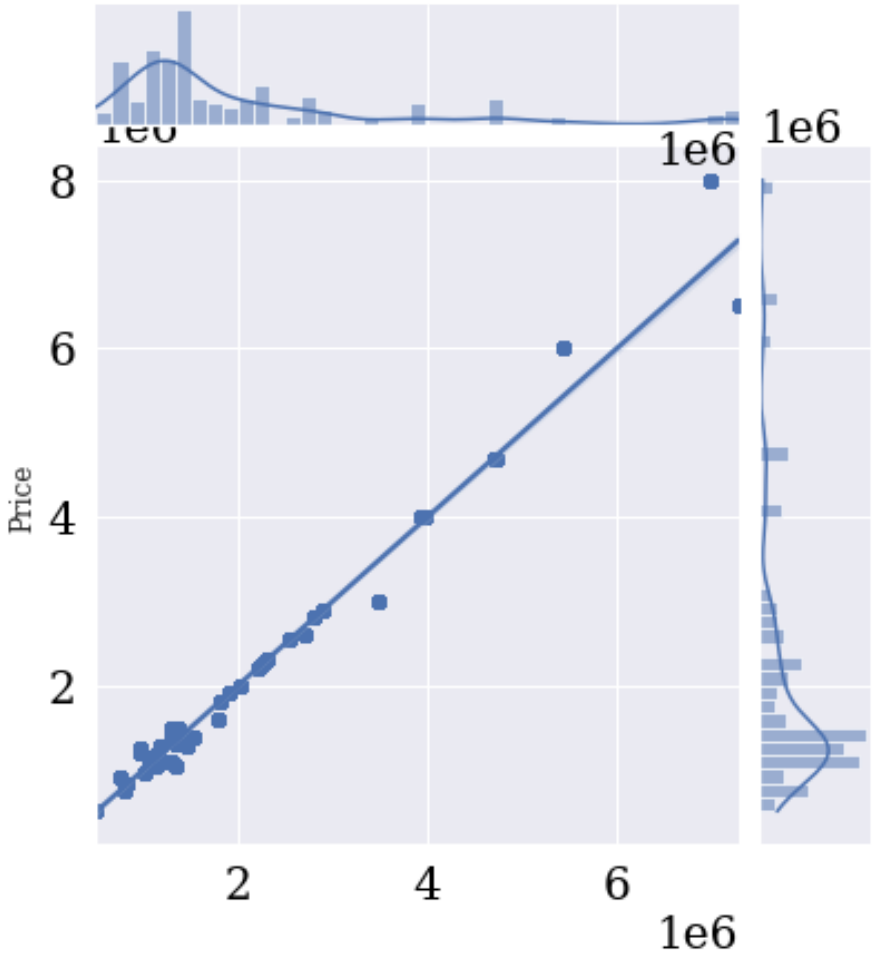
Lasso

Train set evaluation:0.8931568405641584

ElasticNet

Train set evaluation:0.9264798241758878

Select final model



Linear regression

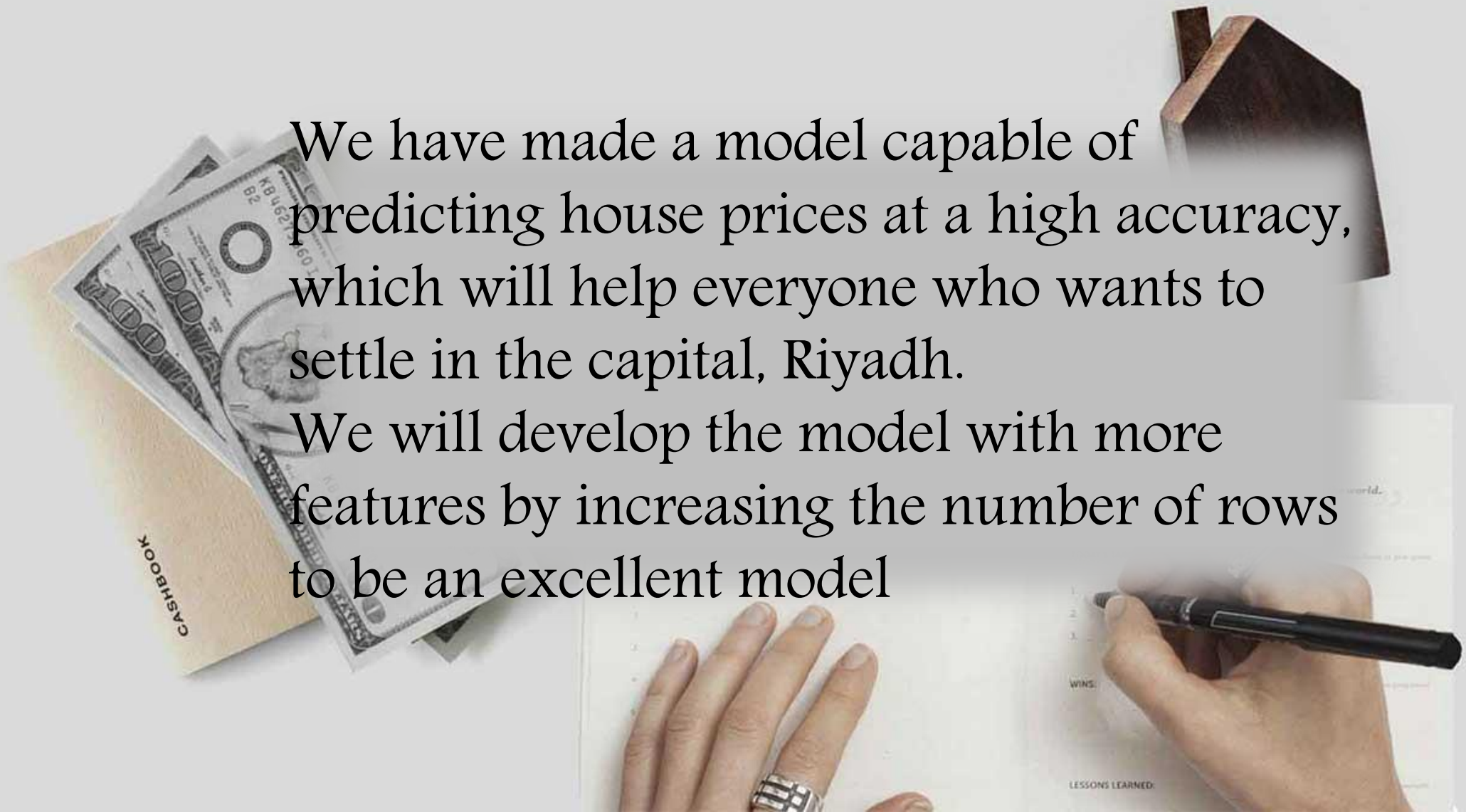
```
Linear Regression test R^2: 0.978
```

Select final model

Linear regression

$$y_{pred} = f(x) = -1460096.21 + 5514.50\,x_1 + 6483.28\,x_2 + 101744.14\,x_3 + 2682497.17\,x_4 + 1320239.07\,x_5 + 907483.44\,x_6 + 1581892.75\,x_7 + 303360.36\,x_8 - 470147.26\,x_9 - 25533.42\,x_{10} + 904065.02\,x_{11} + 969432.84\,x_{12} + 342552.06\,x_{13} + 991067.99\,x_{14} + 567318.40\,x_{15} + 2614178.97\,x_{16} - 813003.84\,x_{17} + 356558.27\,x_{18} - 3489536.64\,x_{19} + 196341.31\,x_{20} + 1903874.45\,x_{21} - 294238.84\,x_{22} + 344079.80\,x_{23} + 622834.53\,x_{24} + 205039.35\,x_{25} - 2329375.57\,x_{26}$$

Conclusion



We have made a model capable of predicting house prices at a high accuracy, which will help everyone who wants to settle in the capital, Riyadh.

We will develop the model with more features by increasing the number of rows to be an excellent model

THANK YOU