

REAL ESTATE

Digital Squad

- Razan Alorefan
- Rasha Alerhili
- Seham Alotibi
- Maha Almutairi
- Atheer Alhomrani





TABLE OF CONTENTS

01

INTRODUCTION

02

2030 VISION

03

DATASET INFORMATION

04

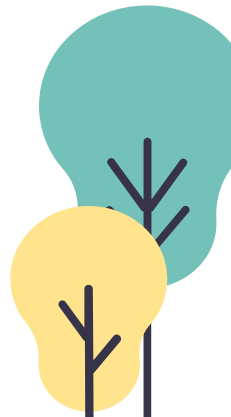
Data Exploration

05

MACHINE LEARNING

06

DASHBOARD



INTRODUCTION

Many people are interested in buying real estate, whether for use in housing or work.

Renting a property can be an ideal option for those wanting extra flexibility and less responsibility.



MYHOME
REAL ESTATE

2030 VISION

The Kingdom's Vision 2030

is one of the giant projects that will have a significant impact in the economic aspect in the Kingdom of Saudi Arabia.

The General Real Estate Authority

was established to be the central regulatory authority for the real estate sector.

Real Estate

Is one of the important economic areas

Ejar platform

is one of the the General Authority for Real Estate documentation service.

PROJECT OBJECTIVE

The objective of this analysis is to help us understand the reason in increasing in real estate in Saudi Arabia and how these factors are used to predict the housing price.



BUSINESS PROBLEM

The rise in rent prices increases the suffering of the citizen and prolongs the period of owning a property.



BUSINESS PROBLEM

Ministry of Housing Provide different solution:

- **Lone and Land** : Help people buy and build their own house.
- **Lone**: If you have land you can start to build you own house.
- One solution to buy a house or real estate far away from downtown like growing districts.

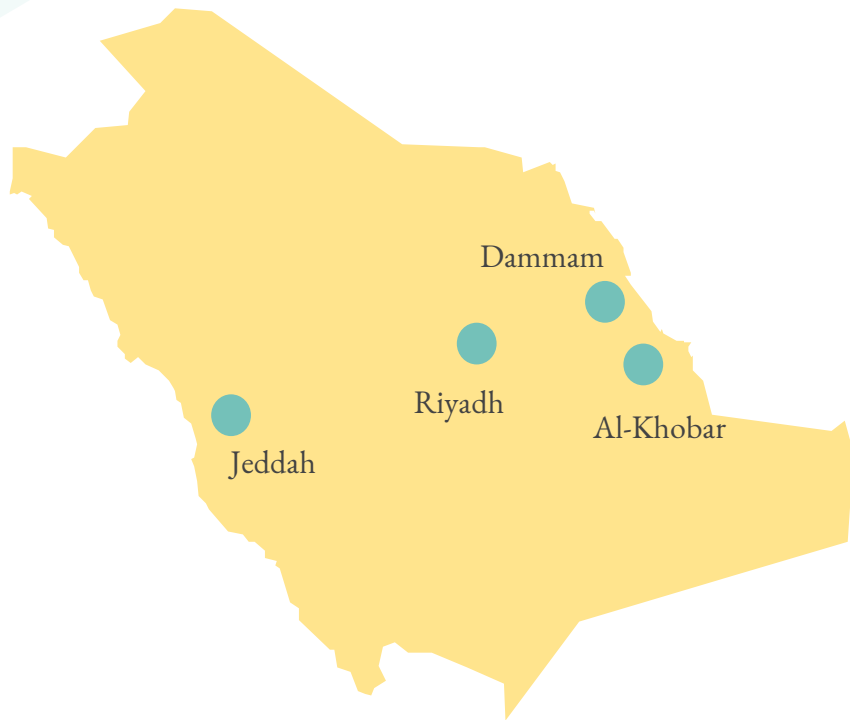


**And still
going
strong!**

DATASET INFORMATION

Saudi Arabia Real Estate (AQAR) is the dataset in this project has taken from Kaggle. The dataset is collected and scrapped from the “Aqar website”.

The dataset contains 3718 Records and 24 columns about the price of rental and details houses in Riyadh, Jeddah, Dammam and Al-Khobar.



DATASET INFORMATION

Variable	DataTye
city	Object
district	Object
front	Object
size	Int
propertyage	Int
bedrooms	Int
bathrooms	Int
livingrooms	Int
kitchen	Int
garage	Int
driverroom	Int
maidroom	Int
furnished	Int
ac	Int
roof	Int
pool	Int
frontyard	Int
basement	Int
duplex	Int
stairs	Int
elevator	Int
fireplace	Int
price	Int
details	Object



DATASET INFORMATION

	size	property_age	bedrooms	bathrooms	livingrooms	kitchen	garage	driver_room	maid_room	furnished	ac
count	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000	3718.000000
mean	390.968531	5.064820	5.083916	4.606509	2.243948	0.909360	0.802044	0.495697	0.795320	0.123453	0.560785
std	1565.056135	7.590427	1.230040	0.703449	0.916436	0.287135	0.398512	0.500049	0.403522	0.329001	0.496358
min	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	280.000000	0.000000	4.000000	4.000000	2.000000	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000
50%	330.000000	2.000000	5.000000	5.000000	2.000000	1.000000	1.000000	0.000000	1.000000	0.000000	1.000000
75%	400.000000	7.000000	6.000000	5.000000	3.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000
max	95000.000000	36.000000	7.000000	5.000000	5.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

	city	district	front
count	3718	3718	3718
unique	4	192	10
top	Khobar	Al Yaqoot	North
freq	976	177	917

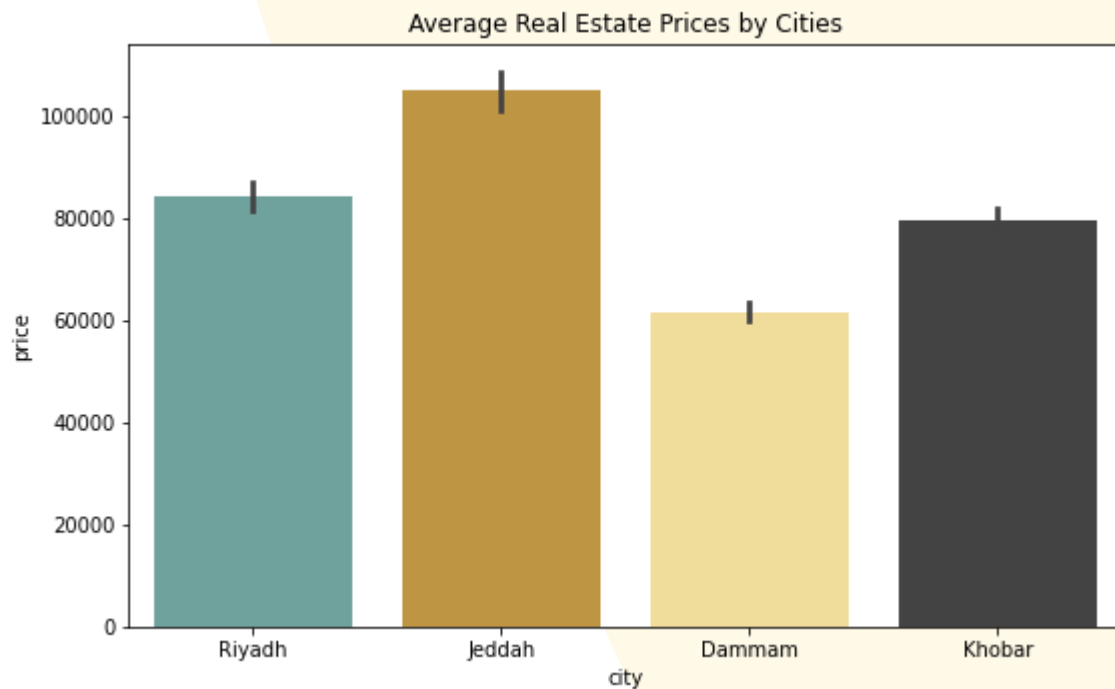
```
df.isnull().sum()
```

```
city          0
district      0
front         0
size          0
property_age  0
bedrooms      0
bathrooms     0
livingrooms   0
kitchen       0
garage        0
driver_room   0
maid_room     0
furnished     0
ac            0
roof          0
pool          0
frontyard     0
basement      0
duplex        0
stairs        0
elevator      0
fireplace     0
price         0
dtype: int64
```

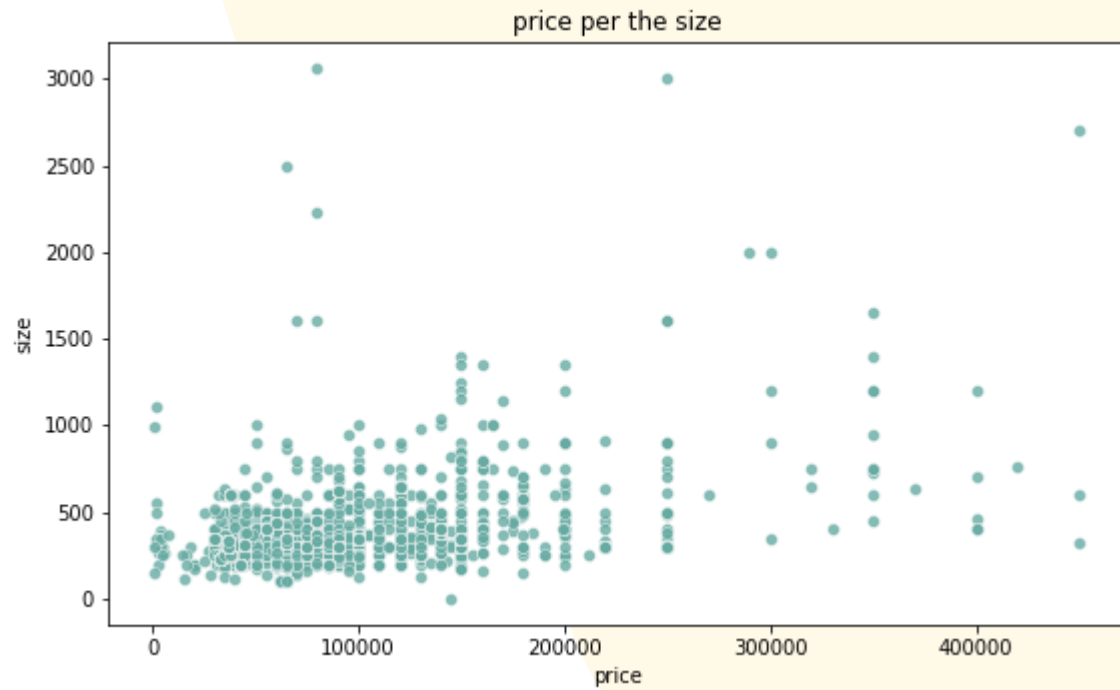
Data Exploration



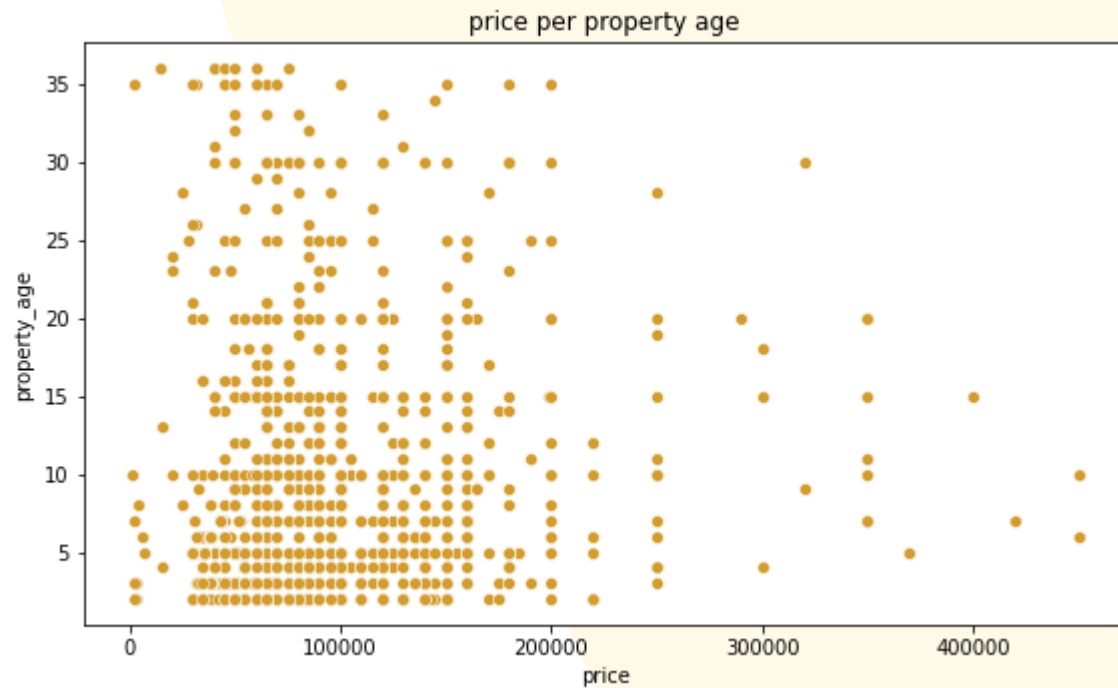
Average Real Estate Prices by Cities



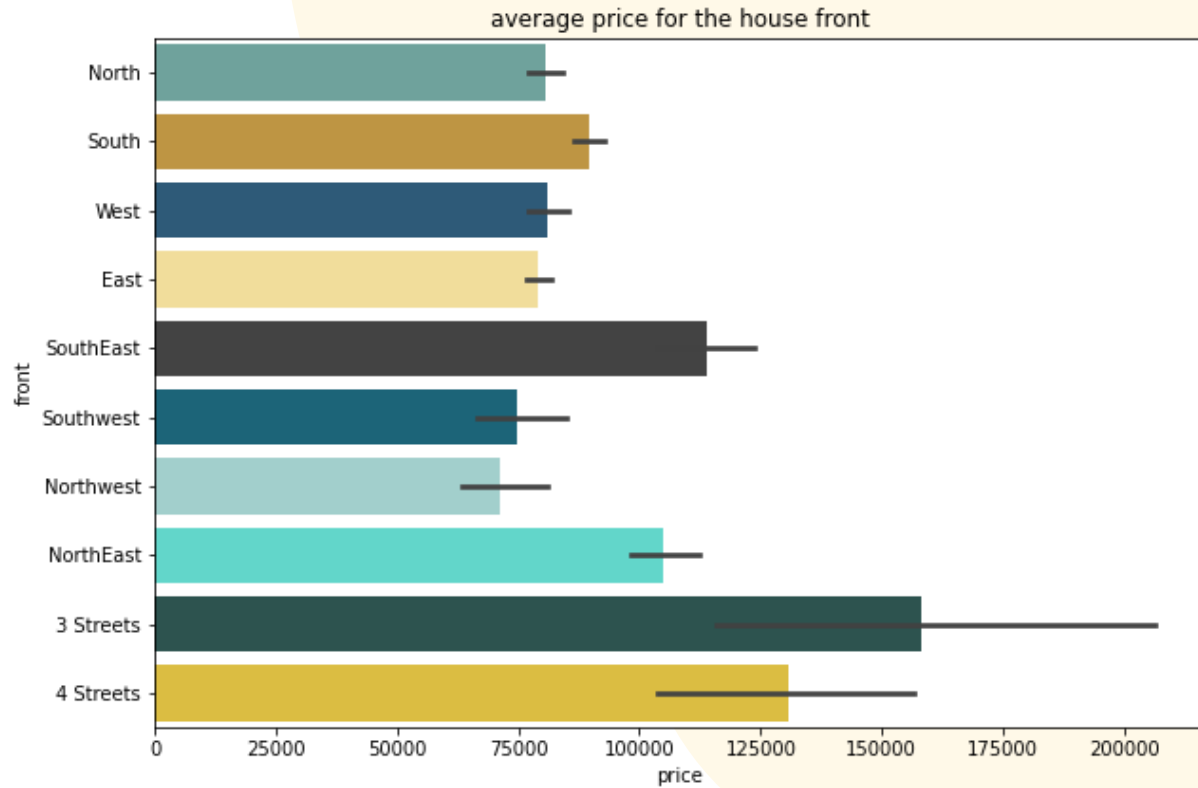
Price for the size of the property



Price for the age of the property

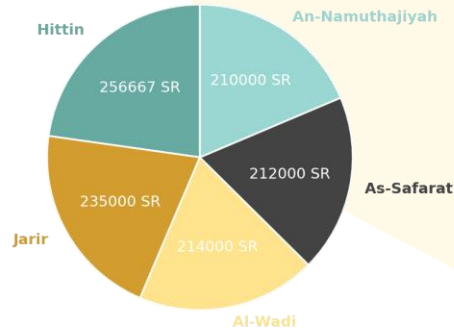


Average price of houses in relation to the front of the house

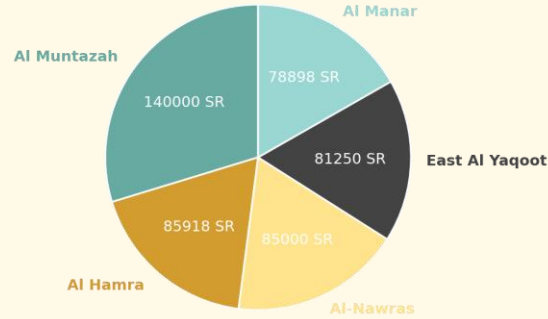


The five most expensive neighborhoods in the main cities in Saudi Arabia

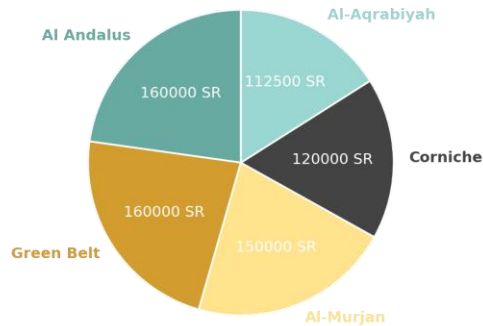
Riyadh



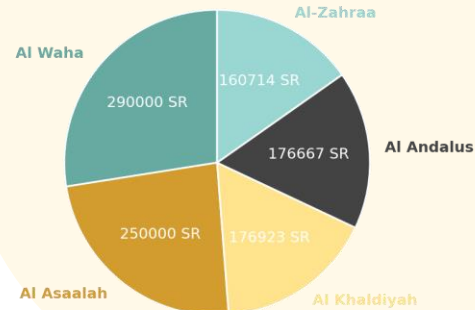
Dammam



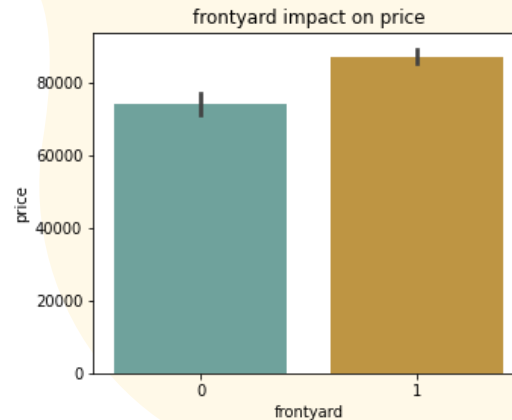
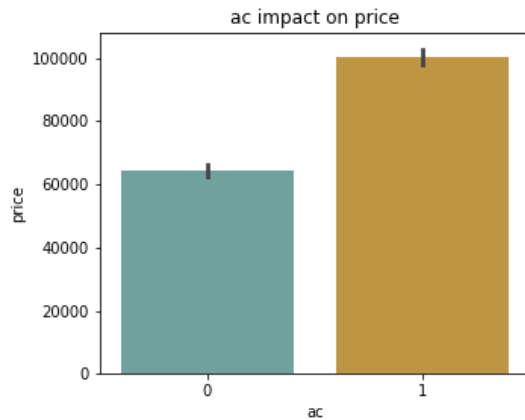
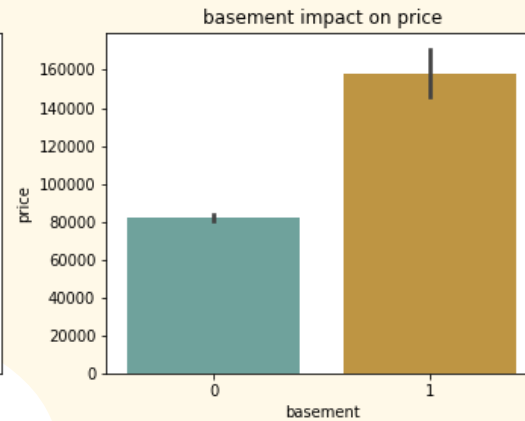
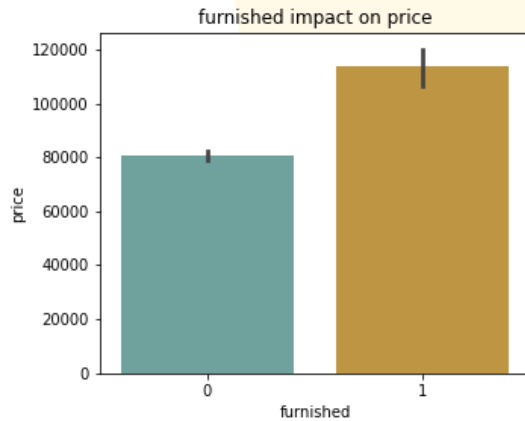
Khobar



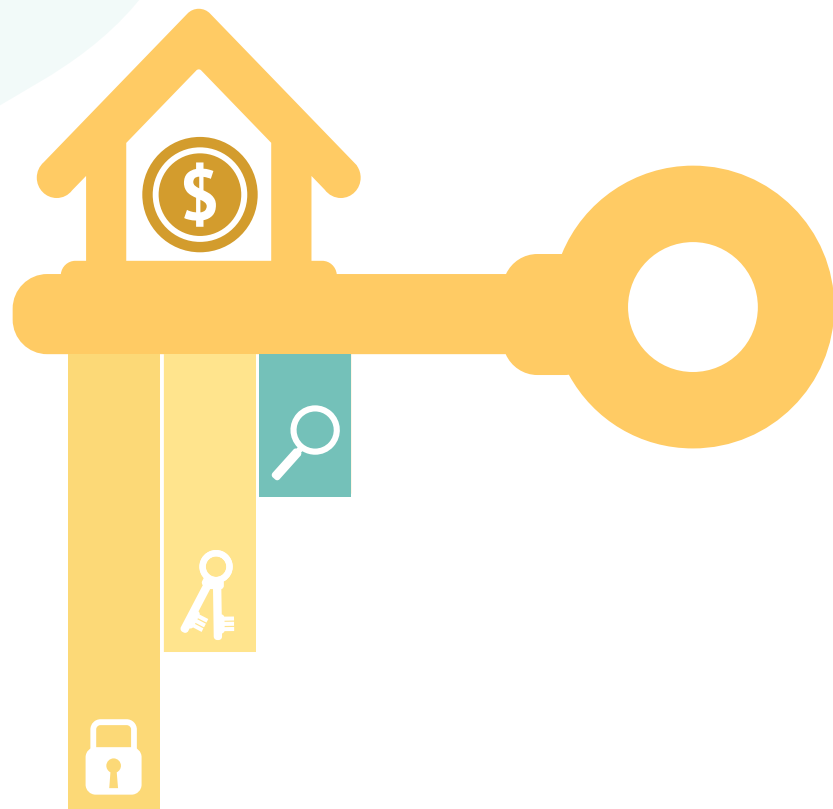
Jeddah



Features that impact the price of housing



MACHINE LEARNING



First: ML Model Preprocessing.

Second: Developing the models

- Testing their performance.

Third: Model Optimization:

- Optimize the performance by removing the outlier.
- Optimize the performance by Hyperparameter tuning.

Fourth: Building a Machine Learning Pipeline.



Features Engineering:

Convert categorical variable into dummy variables → city.

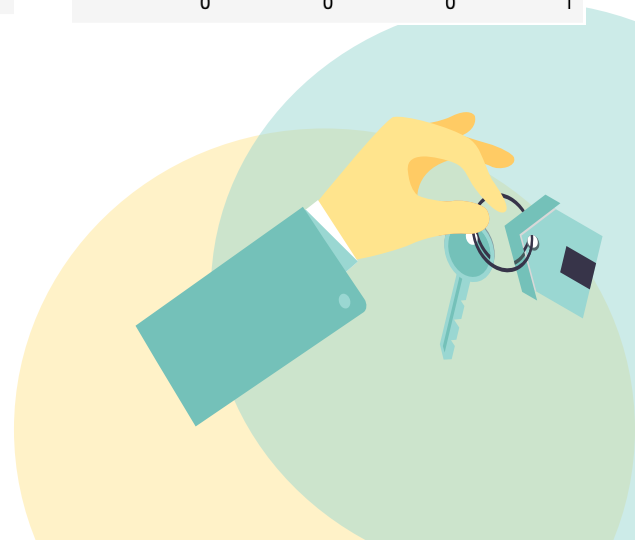
Set the target and Features:

- Target column → price.
- Features → all the columns except price, district, and front.

Split the Dataset:

- 20% Test.
- 80% Train.

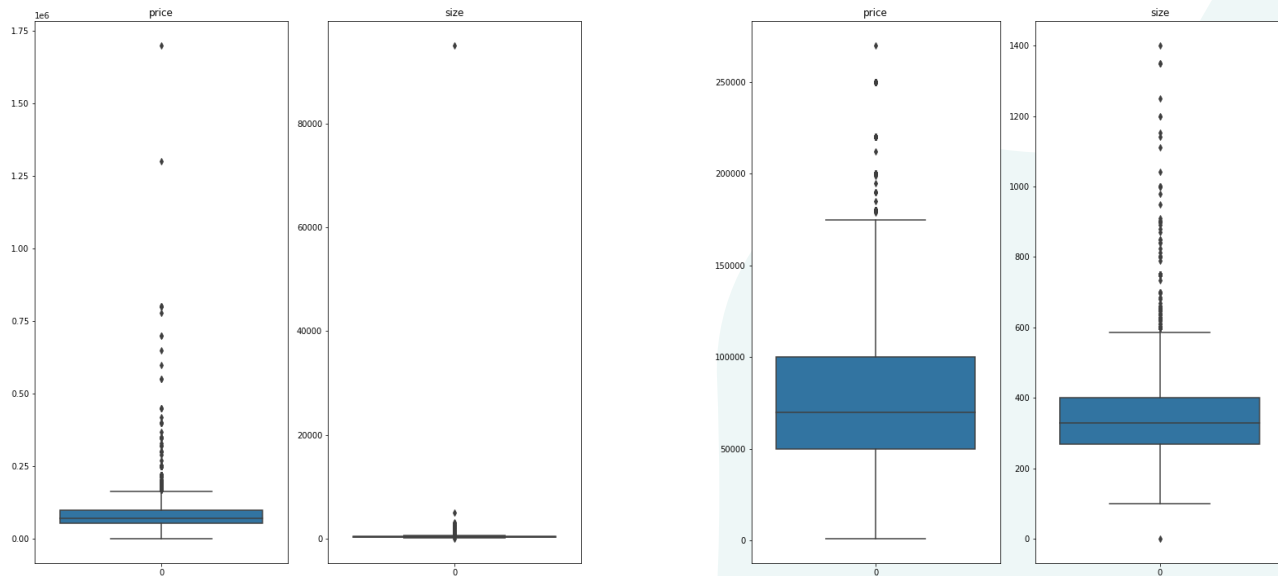
	city	city_Dammam	city_Jeddah	city_Khobar	city_Riyadh
0	Riyadh	0	0	0	1
1	Riyadh	0	0	0	1
2	Riyadh	0	0	0	1
3	Riyadh	0	0	0	1
4	Riyadh	0	0	0	1



Model Evaluation : Test the performance

	Linear Regression	Decision Tree regression	Random Forest regression	Support Vector regression	XGBoost regression
Mean Absolute Error	30062.88	31660.46	18239.24	37403.03	31275.2
Mean Square Error	2748414999.47	3563501985.16	3243962486.32	4118390438.9	2795255218.6

Model Optimization : Optimize the performance by removing the outlier



Remove price > 300000 and size > 1500.

Model Optimization : Optimize the performance by removing the outlier

	Decision Tree regression	Random Forest regression	Support Vector regression	XGBoost regression
Mean Absolute Error	20745.56	10328.29	32634.69	28050.14
Mean Square Error	937800126.14	574407681.31	2248397932.23	1422117744.08



Model Optimization : Optimize the performance by Hyperparameter tuning.

Used GridSearchCV to Hyperparameter tuning.

	Decision Tree regression	Random Forest regression	Support Vector regression
Mean Absolute Error	12710.95	9891.37	20506.07
Mean Square Error	685797509.41	543868427.89	1480126024.55

Building a ML Pipeline:

The Selected Model → Random Forest regression.

Building a ML Pipeline for Random Forest regression:

- Numerical Pipeline.
- Categorical Pipeline.
- Merging the Pipelines.
- Full Pipeline.

	Mean Square Error	Model Score	Mean Absolute Error
Random Forest regression	543868427.89	0.74	9891.37





Percentage Prediction Error:

Calculating the percentage prediction error between the actual price and predict price to check the model accuracy by count the 20%, 40% and 80% error difference of the actual price.

Percentage Prediction Error	20% Difference	40% Difference	80% Difference
Count	150	81	25

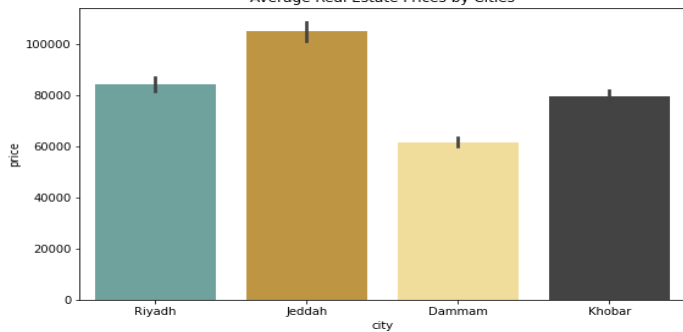
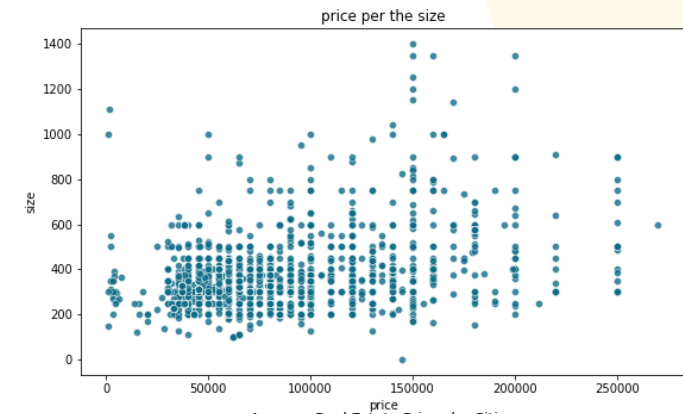
Prediction Performance

The model prediction performance by count how many time the predicted price was more than the actual price and how many time it's less than the actual price or it's equal the actual price

Prediction Performance	Predicted > Actual	Predicted < Actual	Predicted = Actual
Count	162	132	440

MACHINE LEARNING

Actual Price



Predicted Price



Dashboard





Saudi Arabia Real Estate

Rent Houses

Clear Filters



Bath rooms

All

Living rooms

All

Bed rooms

All

87.39K

Average of price

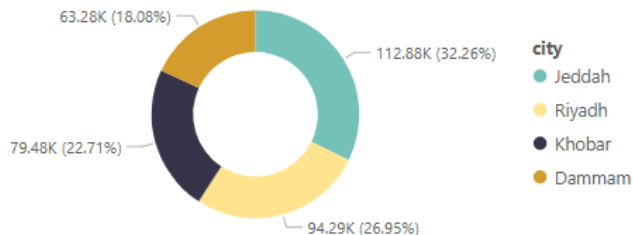
3718

Count of houses

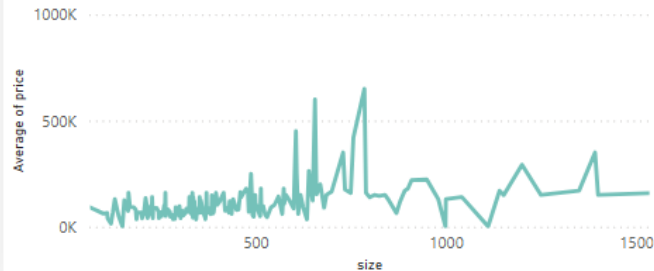
188

Count of district

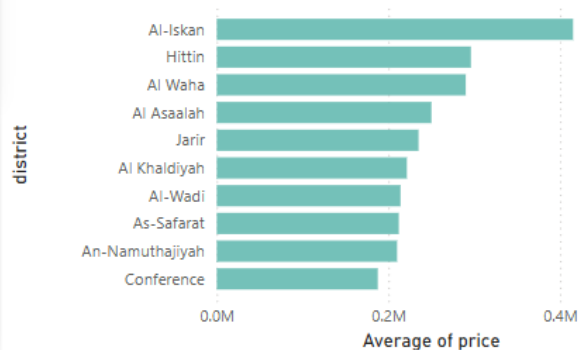
Average of price by city



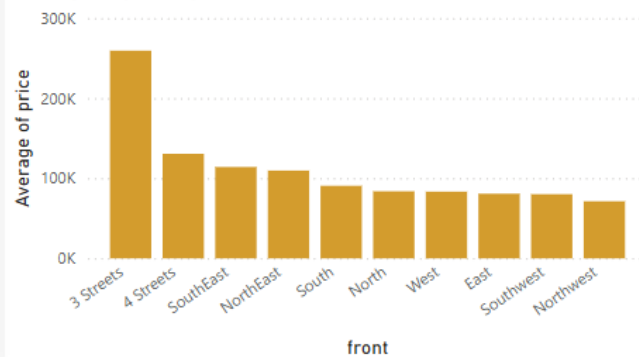
Average of price by size



Average of price by district



Average of price by front



THANKS

Any question?

Digital Squad

