Airbnb Prices Prediction

A study of Airbnb prices in Saudi Arabia

Presented by:
Elaf Alsaedi and Muntaha Jaber

Table of Contents

01

Introduction
Airbnb and project overview

02

Problem Statement
What is the problem we are trying to solve?

03

Workflow and Tools
Project outline and used tools

04

Dataset
How did we collect and clean our data?

05

Model and results
Experiments on different models and scoring

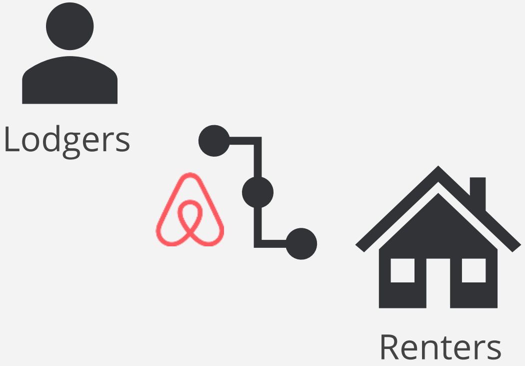
06

Streamlit Tool
Web app for model deployment

07

Conclusion
Final score and recommendations

01 Introduction


Lodgers
Renters


What is airbnb ?

02

Problem Statement

Problem Statement

Build a regression model to predict Airbnb room prices in Saudi Arabia

03 Workflow and Tools

Workflow

01
Data Collection
via Web Scraping

02
Data Cleaning
and Feature
selection

03
Building and
Evaluting the
model

04
Web application
and Dashboard

Tools

Web Scraping & data cleaning

- Pandas
- Beautiful soup
- Selenium

Creat ML Model

- linear regression

Experiments

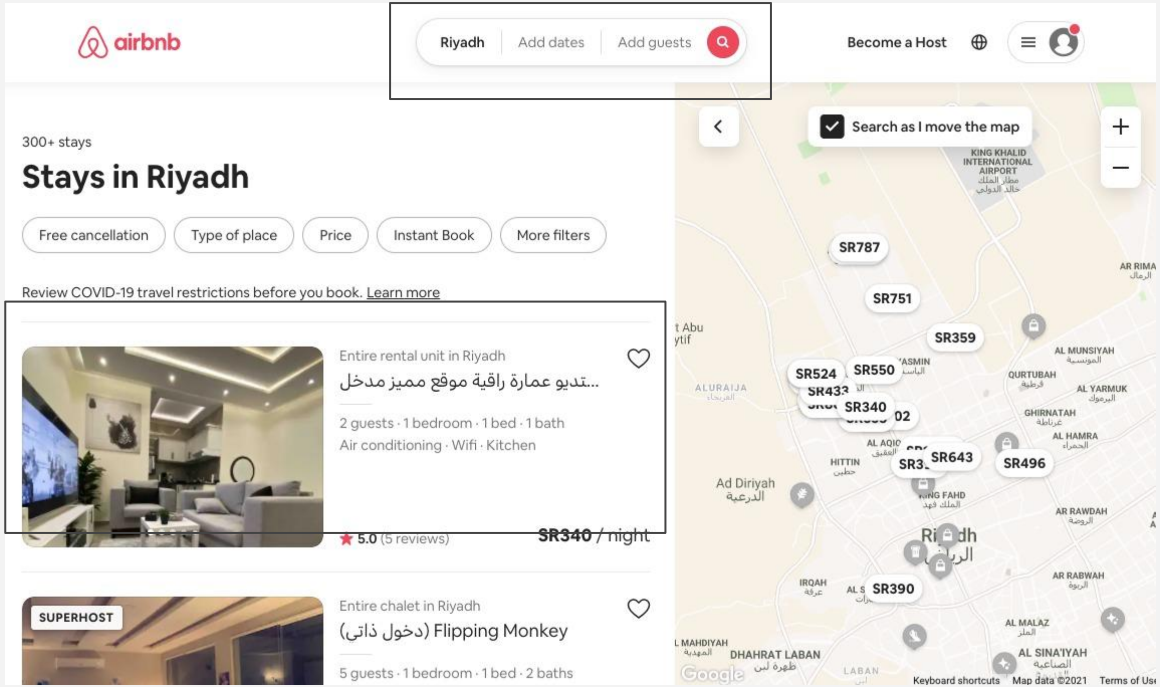
- Feature Engineering
- Scale
- Polynomials
- Selectkbest
- ElasticNet, Lasso, Ridge
- GridSearch

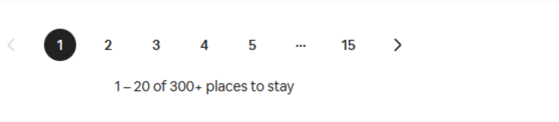
Web Application & Dashboard

- Streamlit
- Tableau

04 Dataset

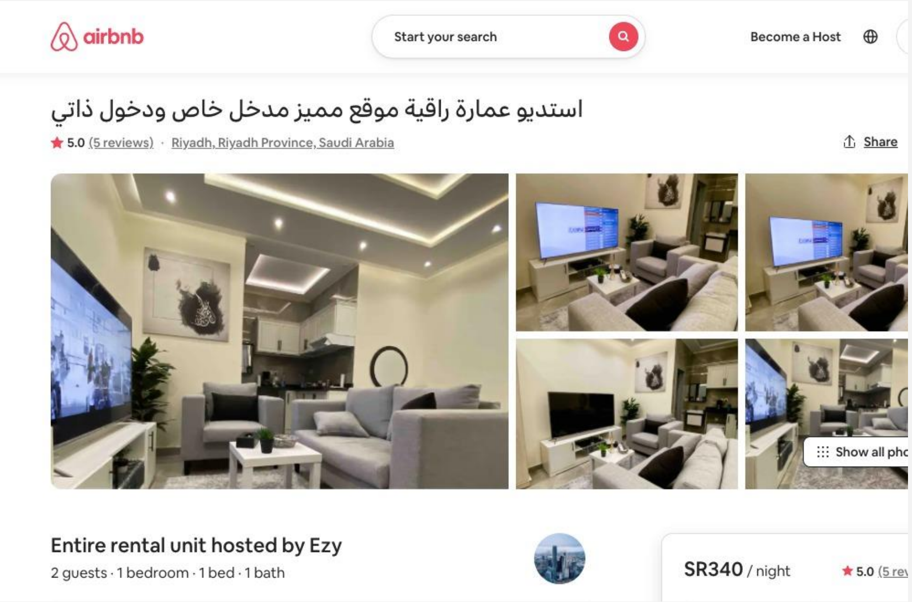
Airbnb Search



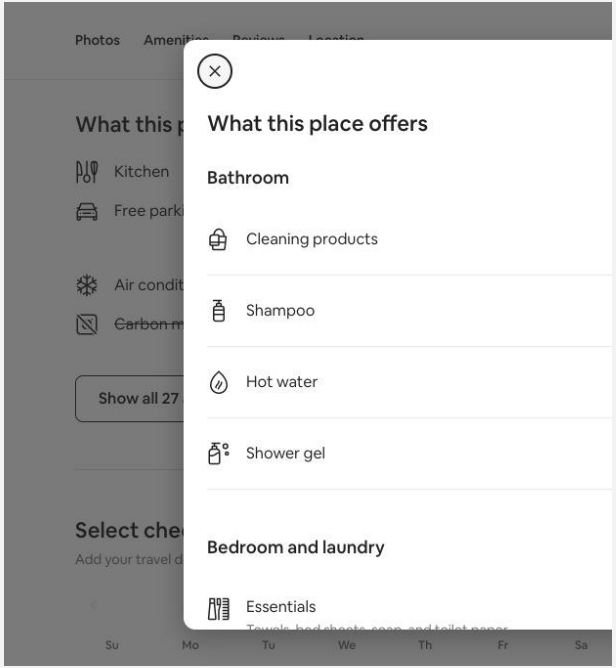


Airbnb room

What One Page Looks Like



Amenities List



DATASET

Data collection

Scraped pages of Saudi cites in Airbnb

Type of data collected

Different forms of data were collected, such as provinces, cities, pricing, the number of rooms, guests, and so on.

The number of data

the total amount of data collected was roughly 5K. After cleaning it became around 2K.

Time

Spent three days scraping the web
And Two Days Cleaning the data

Web scraping difficulties:

Inconsistency in city names, duplicates, and inconsistent values (no city or province)

Data Sample

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
1		rating	city	province	superhost	price	house_typ	link	Room-dar	Suitable fo	Window g	Carbon m	Cooking b	Private en	Pets allow	Smoking a	Long term	Breakfast	Essentials	Lock on be	First aid ki	Hot tub	Private liv	Host greet	Fire pit	Outdoor d	Self check	Lockbox	Outdoor s	Ou
2	0	0	buraydah	6	0	1100	Earth hous	https://w	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	0	buraydah	6	0	2200	Entire cha	https://w	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
4	2	0	buraydah	6	0	400	Shared ro	https://w	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
5	3	0	buraydah	6	0	550	Private ro	https://w	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	
6	4	0	buraydah	6	0	2500	Entire vill	https://w	0	0	0	0	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	
7	5	0	buraydah	6	0	500	Earth hous	https://w	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	
8	6	0	buraydah	6	0	714	Entire cha	https://w	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
9	7	0	buraydah	6	0	2700	Farm stay	https://w	0	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
10	8	0	buraydah	6	0	700	Entire cha	https://w	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	9	0	buraydah	6	0	900	Entire cha	https://w	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	
12	10	0	buraydah	6	0	1000	Entire cha	https://w	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	
13	11	4.75	buraydah	6	0	470	Entire cha	https://w	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
14	12	0	mulayda	6	0	543	Entire ren	https://w	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1	0	
15	13	3.75	buraydah	6	0	150	Entire loft	https://w	0	1	0	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	
16	14	0	buraydah	6	0	1200	Entire vill	https://w	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	
17	15	0	buraydah	6	0	600	Tiny hous	https://w	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
18	16	0	buraydah	6	0	259	Entire cha	https://w	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
19	17	5	buraydah	6	0	380	Entire resi	https://w	0	1	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	
20	18	0	buraydah	6	0	300	Entire cha	https://w	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
21	19	0	buraydah	6	0	307	Private ro	https://w	0	1	0	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	
22	20	5	buraydah	6	0	600	Entire cha	https://w	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
23	21	0	buraydah	6	0	77	Entire ren	https://w	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	
24	22	0	buraydah	6	0	180	Entire ren	https://w	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	
25	23	0	buraydah	6	0	222	Entire con	https://w	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	
26	24	0	buraydah	6	0	300	Private ro	https://w	0	1	0	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	
27	25	0	buraydah	6	0	120	Private ro	https://w	0	0	0	1	1	0	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	
28	26	0	buraydah	6	0	250	Farm stay	https://w	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
29	27	0	buraydah	6	0	250	Entire loft	https://w	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
30	28	0	buraydah	6	0	600	Entire cha	https://w	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
31	29	4.67	buraydah	6	0	750	Entire cha	https://w	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
32	30	0	mulayda	6	0	113	Entire ren	https://w	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
33	31	0	buraydah	6	0	850	Entire resi	https://w	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	
34	32	0	buraydah	6	0	550	Entire loft	https://w	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
35	33	0	mulayda	6	0	400	Entire cha	https://w	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
36	34	0	buraydah	6	0	560	Entire ren	https://w	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
37	35	0	buraydah	6	0	1100	Entire cha	https://w	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	

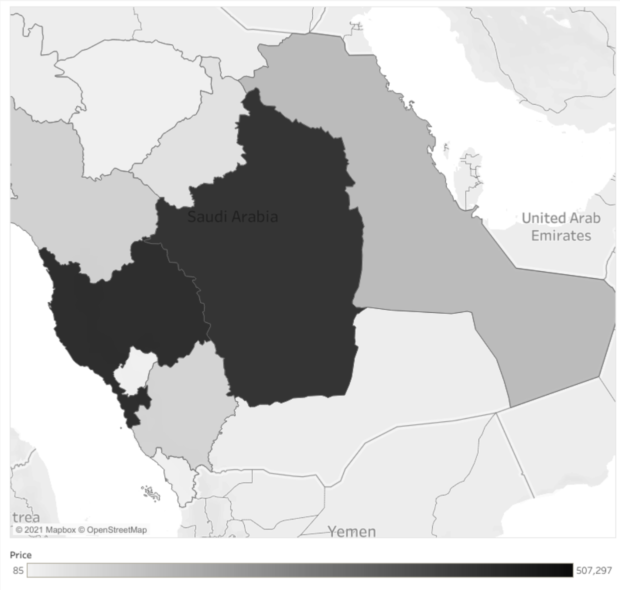
Columns

```
['rating',
 'city',
 'province',
 'superhost',
 'price',
 'house_type',
 'link',
 'rating_count',
 'mean_price_per_prov',
 'Room-darkening shades',
 'Suitable for events',
 'Window guards',
 'bbq',
 'hot_water',
 'child_friendly',
 'heating',
 'home_utensils',
 'appliance',
 'sauna',
 'entertainment',
 'safety',
 'boat',
 'internet']
```

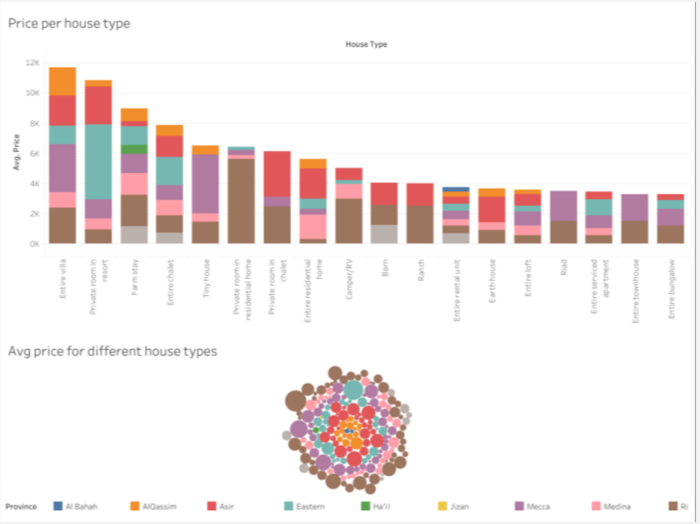
EDA DashBoard

[Link]

House Price Per Province



House Price Per Province and Type



Data Cleaning

Categorical columns

Incorrect information

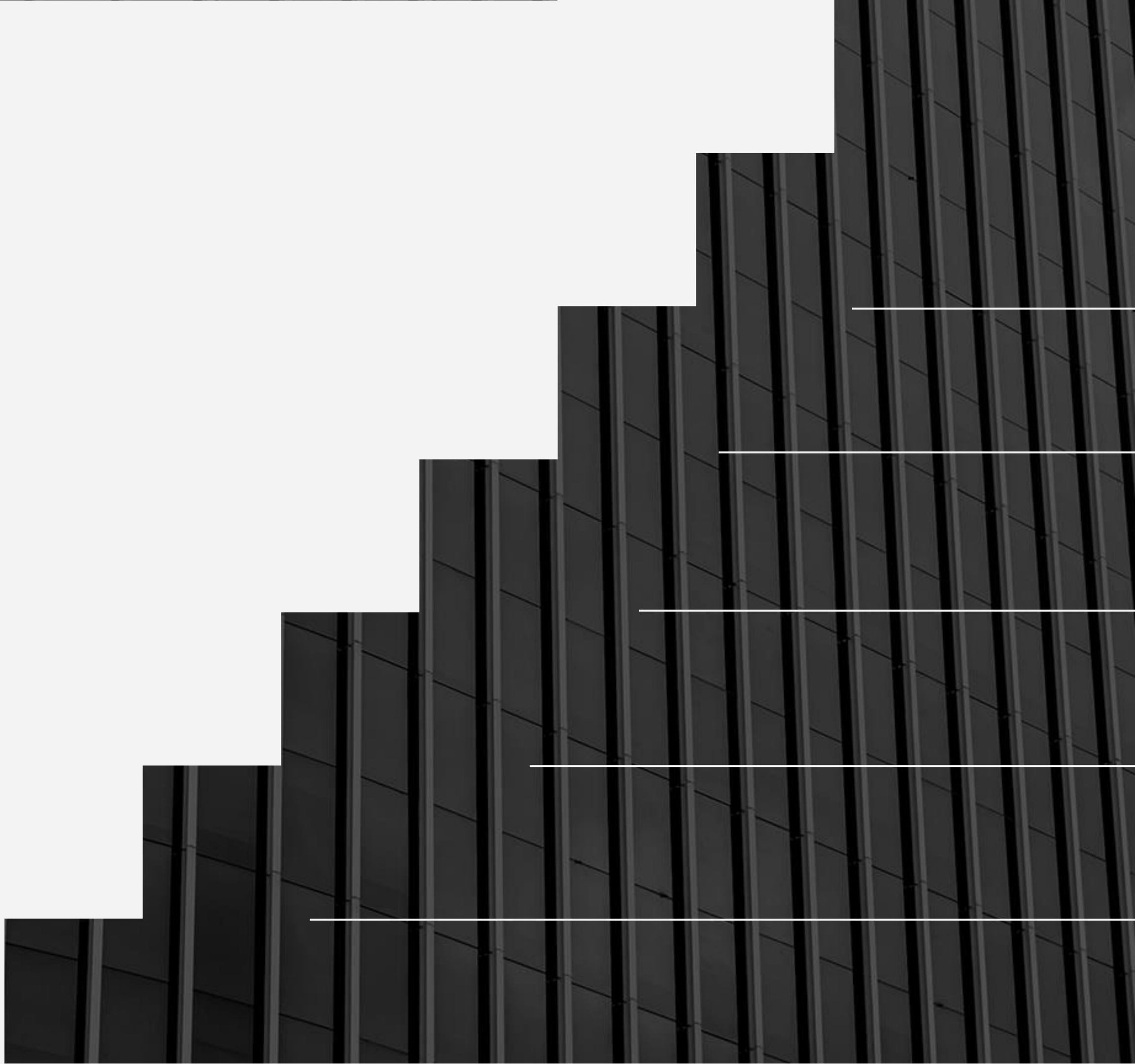
Duplicates and null values

Large Features and outliers

05 Model and experiments

Experiments

- Scale
- Polynomial features
- Lasso
- Ridge
- Elastic Net
- Feature selection
- Remove outliers



Experiments

R2: 33.849609115994596
MAE 380.841963077402

K-best with 25 features

R2: 36.50024047047358
MAE: 383.2017860980756

Lasso with gridsearch

Vanilla Linear Regression

R2: 34.566135965863275
MAE 402.0275488129502

GridSearch k best and linear regression

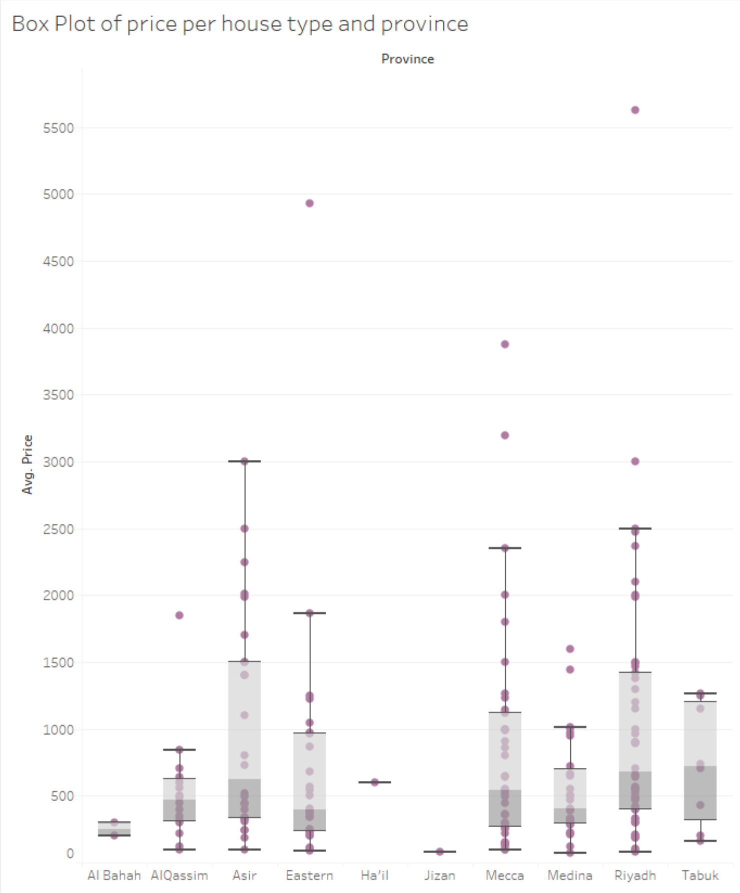
R2: 36.32524852538283
MAE 394.4275967064814

Best Model

1- Remove Outliers
2- Gridsearch with ElasticNet
3- Select kbest features (62 features)
4- Apply Random Forest them linear regression

R2: 47.12%
MAE 273.85

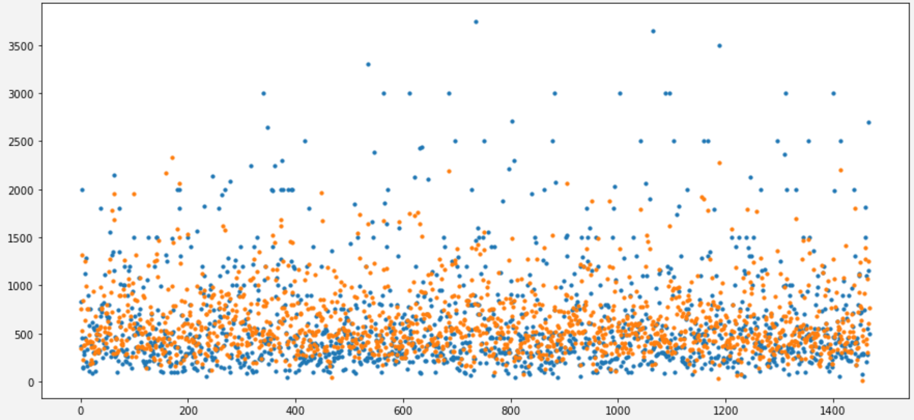
Outliers



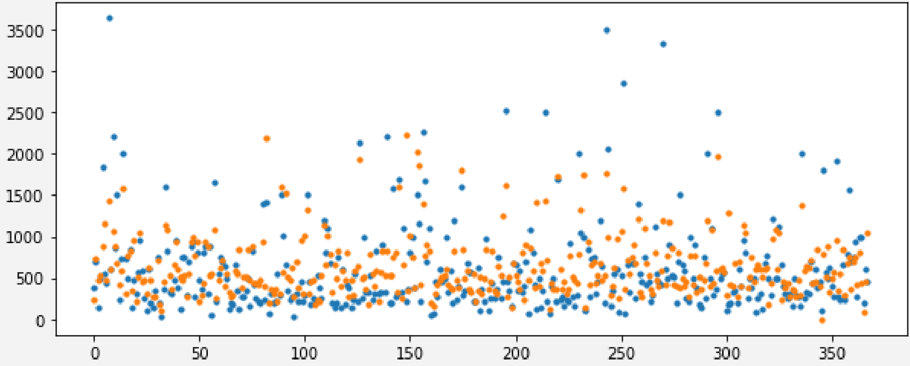
Box Plot of price per house type and province
Province
Avg. Price
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Al Bahah
AlQassim
Asir
Eastern
Ha'il
Jizan
Mecca
Medina
Riyadh
Tabuk

Price Plot

Train Data vs Predicted Data



Test Data vs Predicted Data



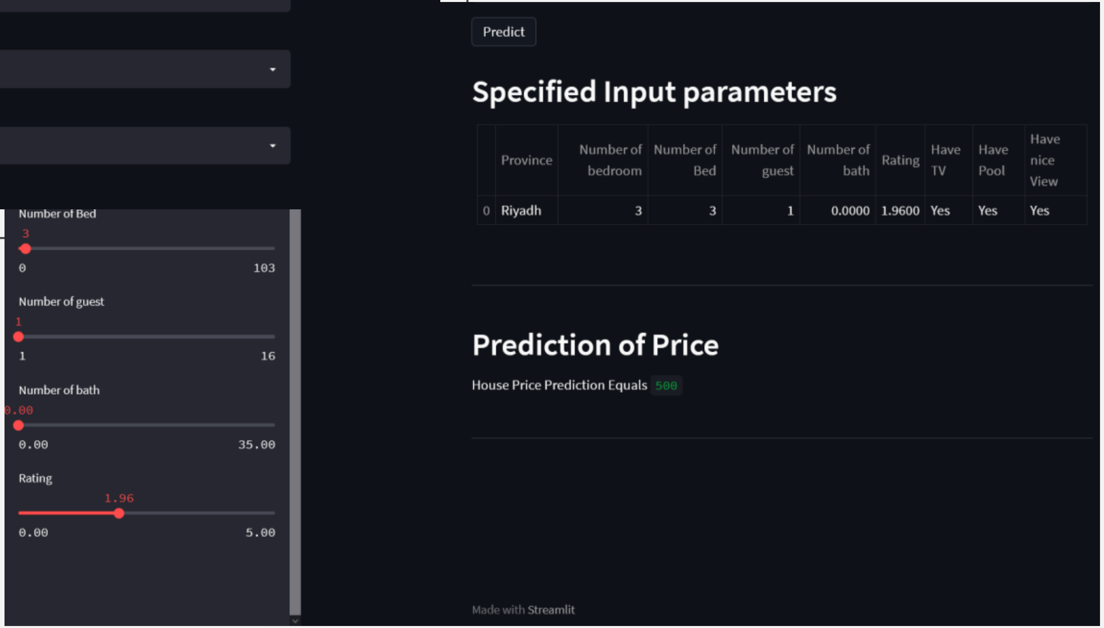
06

StreamLit Tool

Streamlit Tool

- Streamlit is an open-source app framework for Machine Learning and Data Science
- Develop Airbnb price prediction web app in Saudi Arabia


Specify Input Parameters
Number of bedroom
3
0
44
Number of Bed
3
0
103
Number of guest
1
1
16
Number of bath
0.00
0.00
35.00
Rating
1.96
0.00
5.00
Saudi Arabia House Price Prediction App
This app predicts the Saudi Arabia House Price for the Airbnb Website!
Province
Riyadh
Have Pool
Yes
Have TV
Yes
Have nice View
Yes
Predict


Number of Bed
3
0
103
Number of guest
1
1
16
Number of bath
0.00
0.00
35.00
Rating
1.96
0.00
5.00
Predict
Specified Input parameters
Province
Number of bedroom
Number of Bed
Number of guest
Number of bath
Rating
Have TV
Have Pool
Have nice View
0
Riyadh
3
3
1
0.0000
1.9600
Yes
Yes
Yes
Prediction of Price
House Price Prediction Equals 500
Made with Streamlit

Website App

07 Conclusion

Results and Recommendations

Best r2 score of 47.12%

- Collect more meaningful features with more data.
- Try a different machine learning model.


Thank You