

REAL ESTATES IN RIYADH



Dimah Albunayyih
Tarfah Alabbad



CONTENTS

- INTRODUCTION
- AIMS AND OBJECTIVES
- TOOLS
- WORK FLOW
- RESULTS
- CONCLUSION



INTRODUCTION



Buy ▾ Rent ▾ **NEW** Add property Blog Find an Agency **NEW** New Projects

PURPOSE Buy ▾	LOCATION	PROPERTY TYPE Residential ▾	PRICE (SAR) 0 to Any ▾
BEDS All ▾	AREA (SQ. M.) 0 to Any ▾	BATHS All ▾	KEYWORD
AGENCY			ADVANCED FILTERS Any ▾

Properties for Sale in Saudi Arabia



Riyadh Region (3,524)	Western Region (1,224)	Eastern Region (1,006)	Hail Region (344)
VIEW ALL LOCATIONS ▾			

Furnish Type: **All** Furnished Unfurnished

Popular ▾ ☐ Active listings  

Be the first to hear about new properties

 **ALERT ME OF NEW PROPERTIES**



SAR 33,000,000
Al Ghadir, North Riyadh, Riyadh, Riyadh Region
Residential Building
Residential Units Building For Sale In Al Ghadir...
1,225 Sq. M.

 **CALL**  **EMAIL**



SAR 5 200 000



AIMS AND OBJECTIVES

- To perform web scraping on buyot website
- To create a dataset from the scraped data and perform EDA on it
- To choose the best regression model for our data by comparing scores to avoid over/under fitting
- To see if our train score close to the test and reach the goal of generalization



TOOLS



WORKFLOW

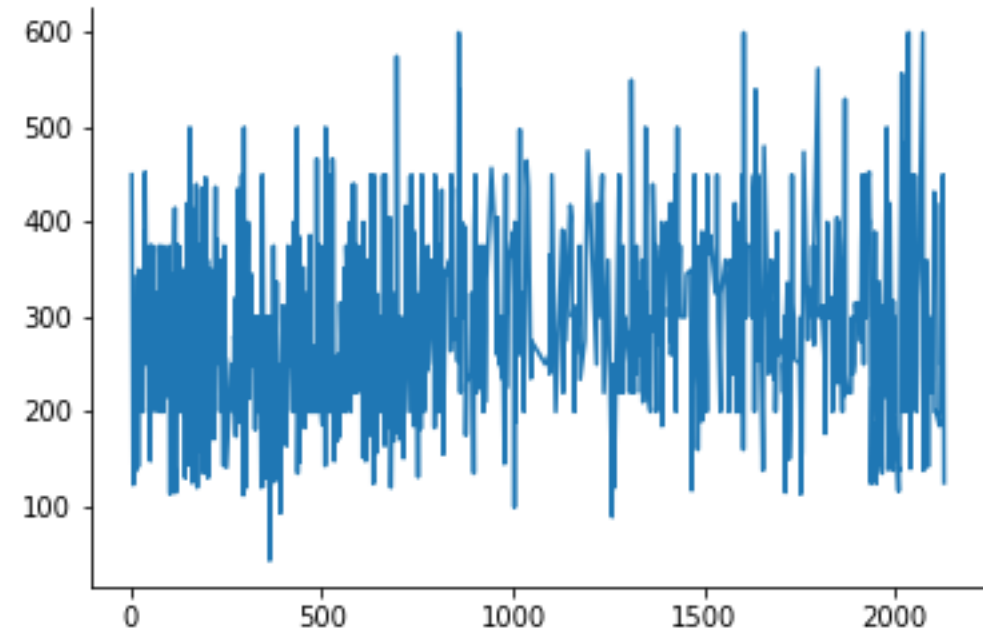
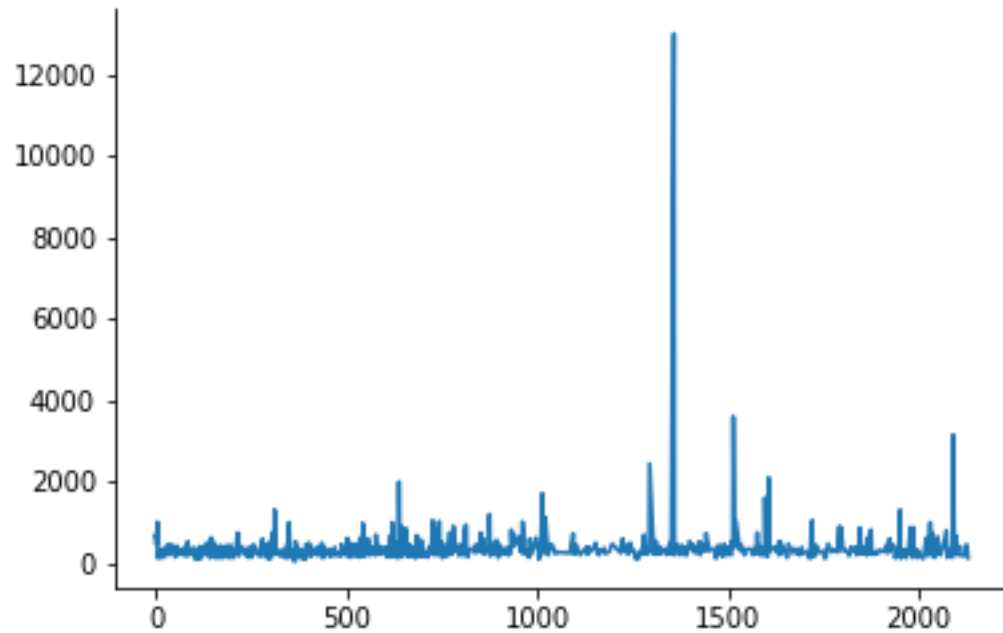


REALESTATEDATASET

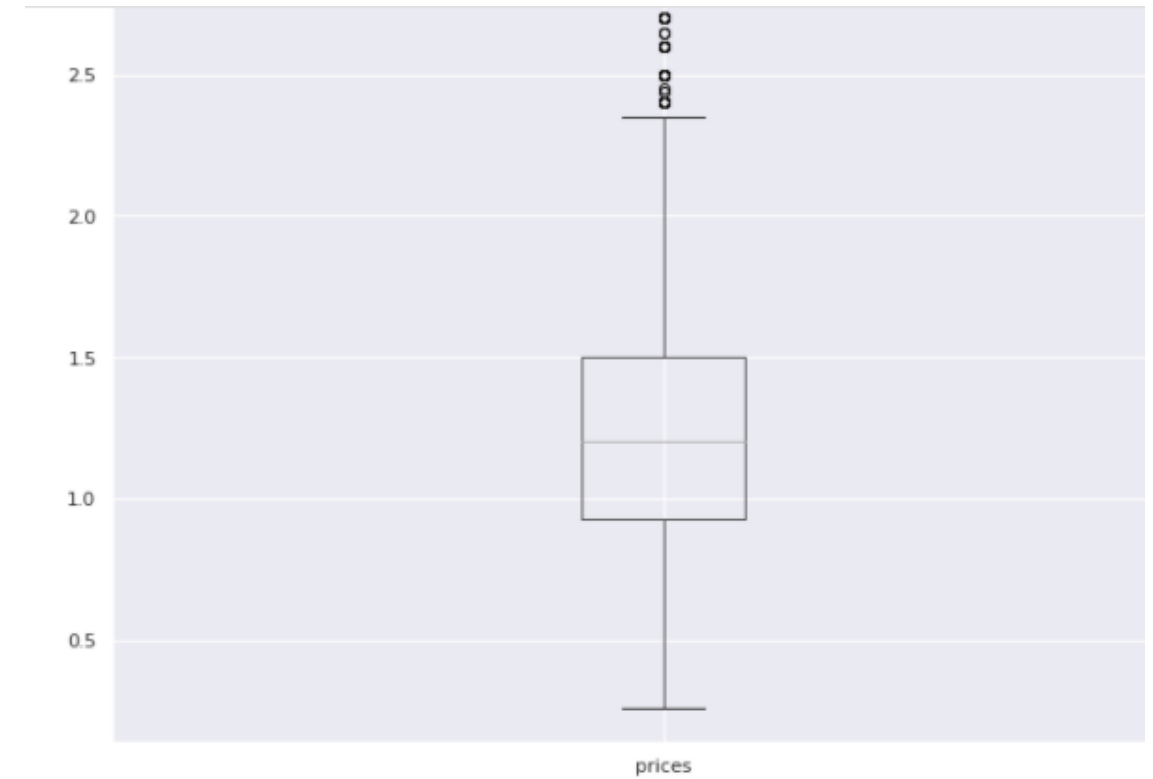
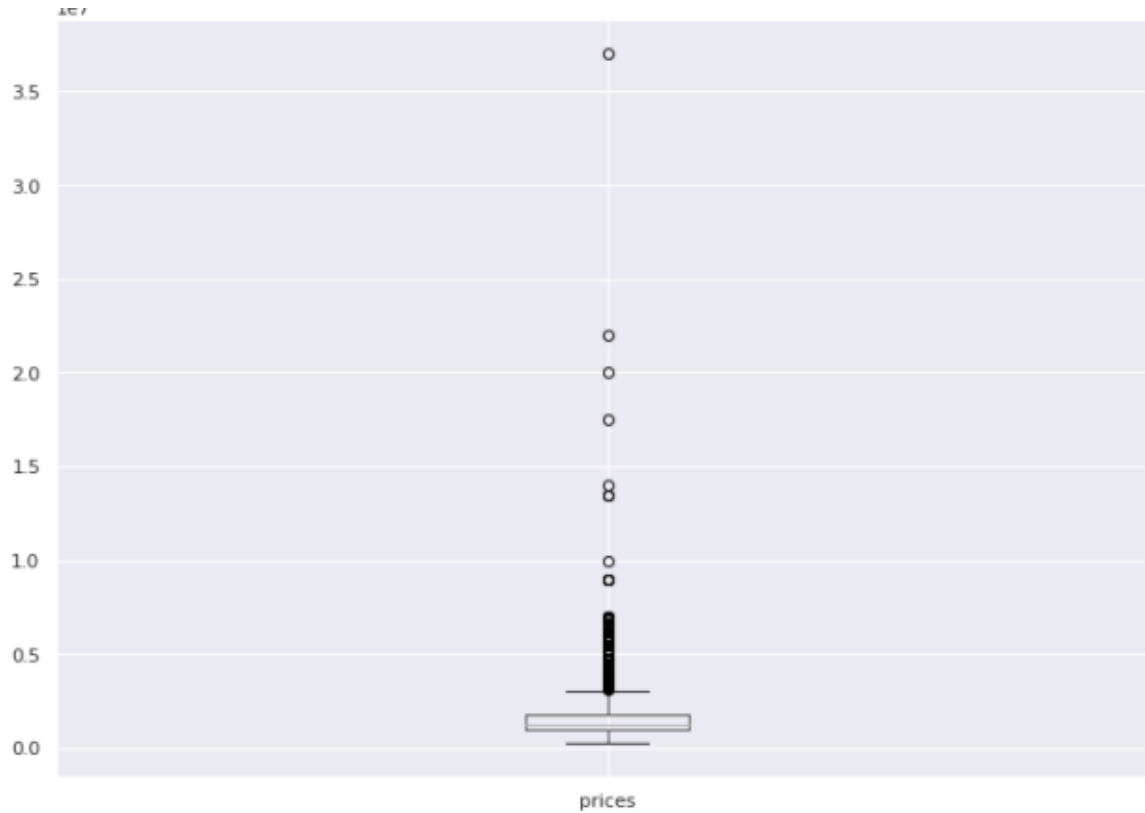
	neighborhood	regions	RealEstateType	square_meters	rooms_number	baths_number	prices	RealEstateTypeCode	RegionsCode
3	Al Nuzhah	North Riyadh	Villa	450	4	5	2500000.0	0	0
7	Al Sahafah	North Riyadh	Apartment	123	2	4	1750000.0	1	0
8	Al Sahafah	North Riyadh	Apartment	177	3	5	2092454.0	1	0
9	Al Munsiyah	East Riyadh	Villa	261	3	3	1900050.0	0	1
11	Al Munsiyah	East Riyadh	Villa	254	3	5	2004300.0	0	1
...
2121	Al Rimal	South Riyadh	Villa	250	4	3	1040000.0	0	2
2122	Al Dar Al Baida	North Riyadh	Villa	250	6	4	800000.0	0	0
2124	Al Mahdiyah	West Riyadh	Villa	200	6	6	1150000.0	0	3
2125	Laban	North Riyadh	Villa	450	4	6	1700000.0	0	0
2128	Dhahrat Laban	North Riyadh	Apartment	125	3	3	395000.0	1	0

1041 rows × 9 columns

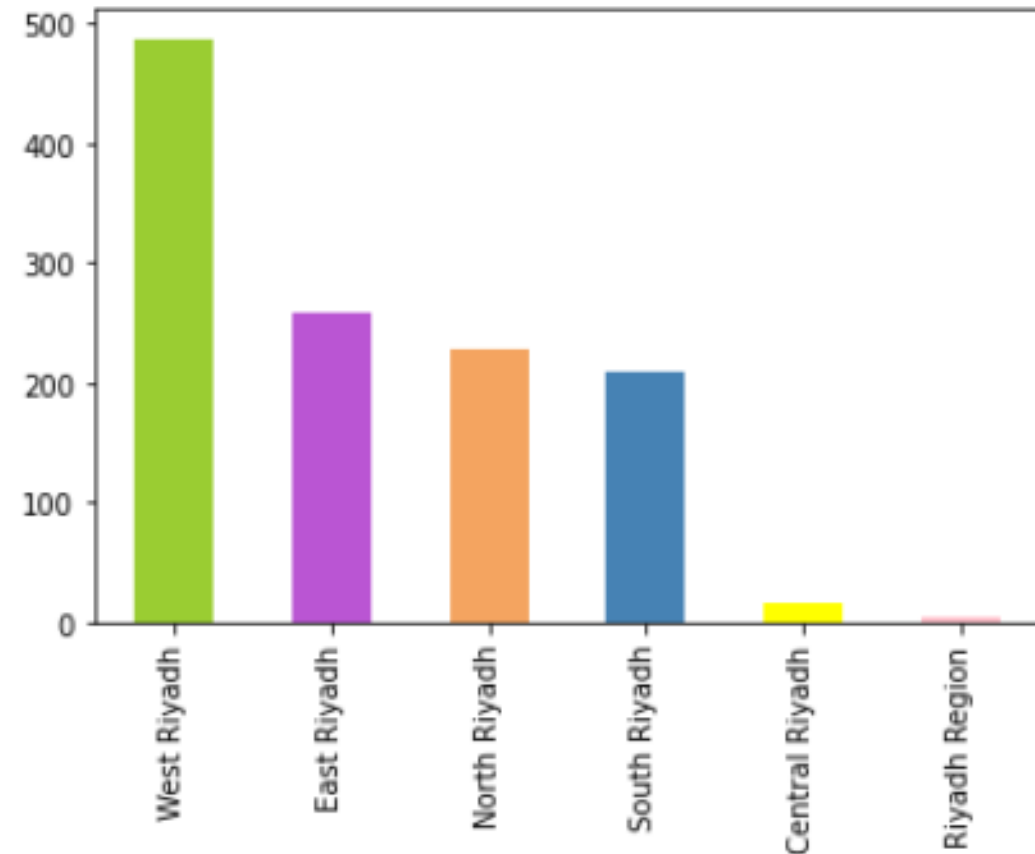
Square meters column before and after removing outliers



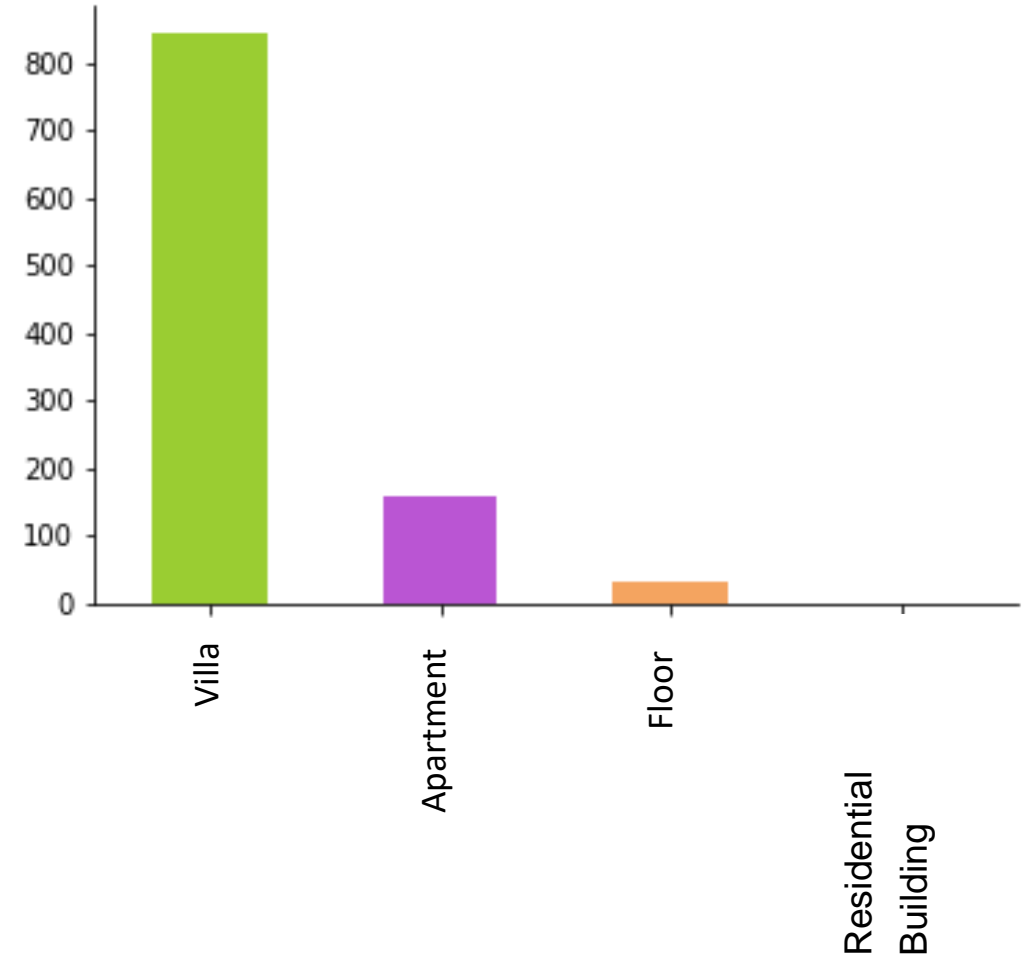
Prices column before and after removing outliers



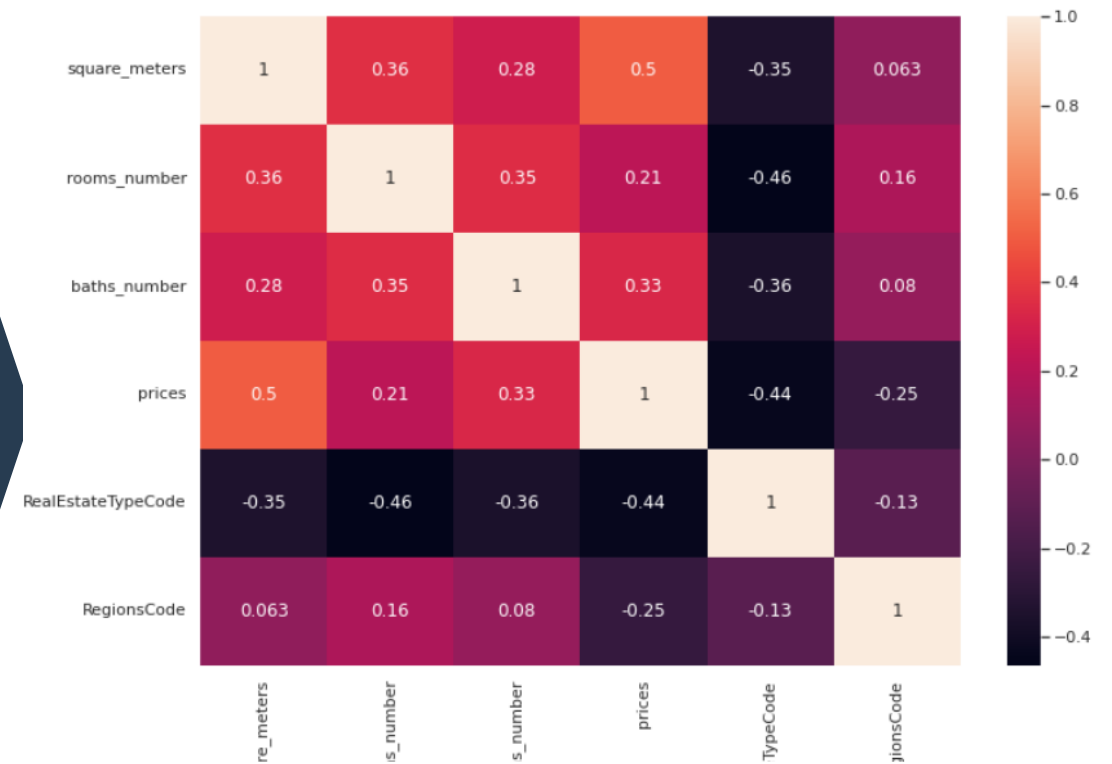
Counts of real estates sales for each region



Counts of real estates sales for each real estate type

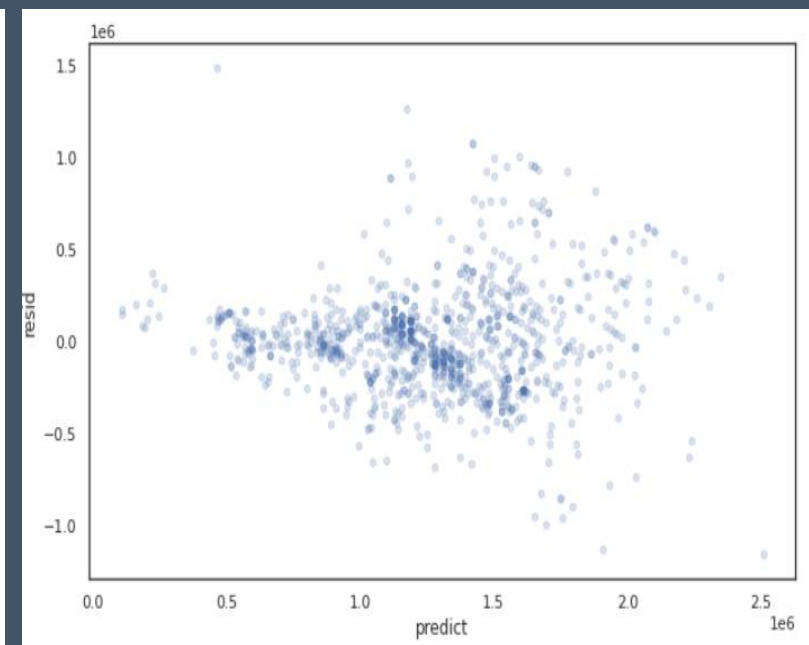
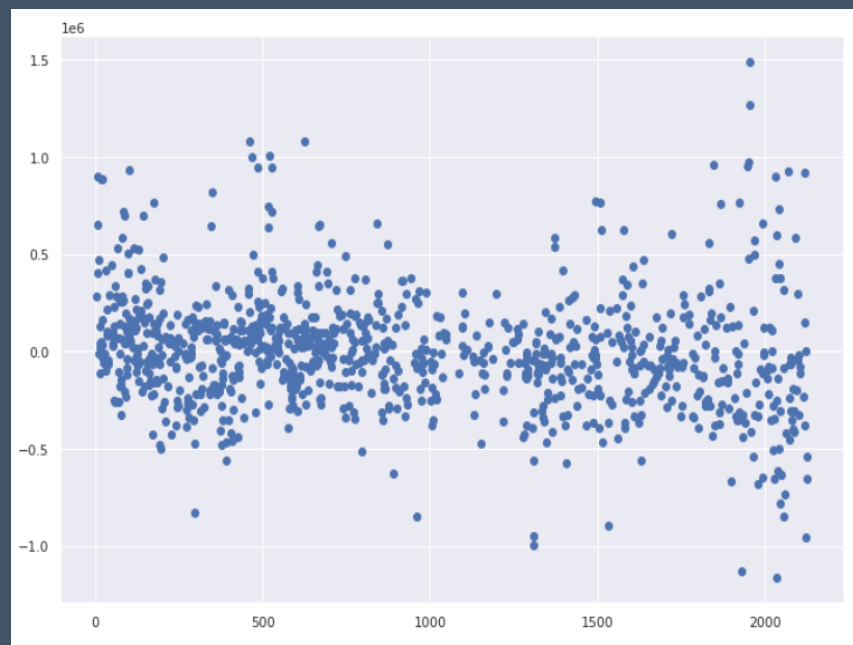
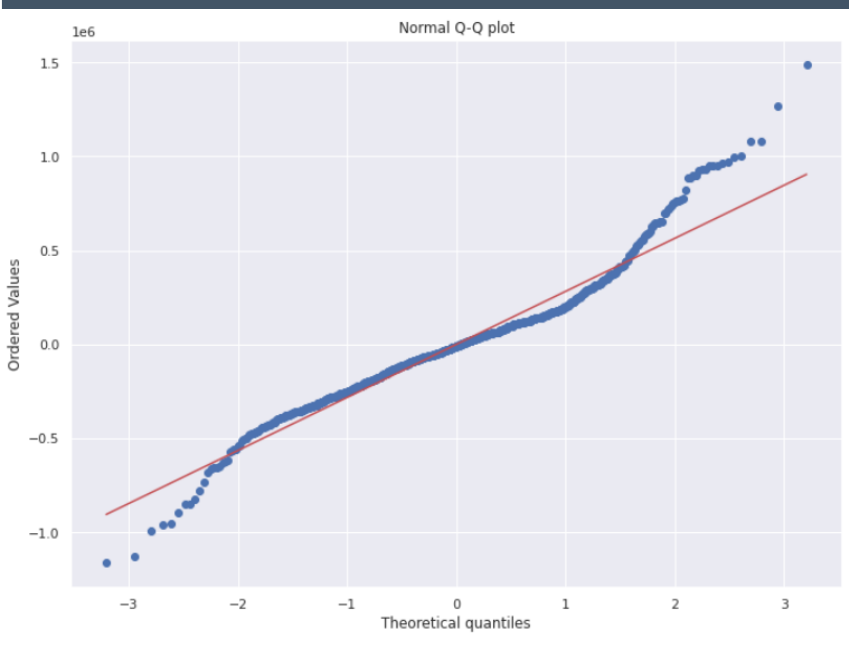


After cleaning the data we transform the categorical features into dummies then we use the heat map to see the correlation between the target and the features



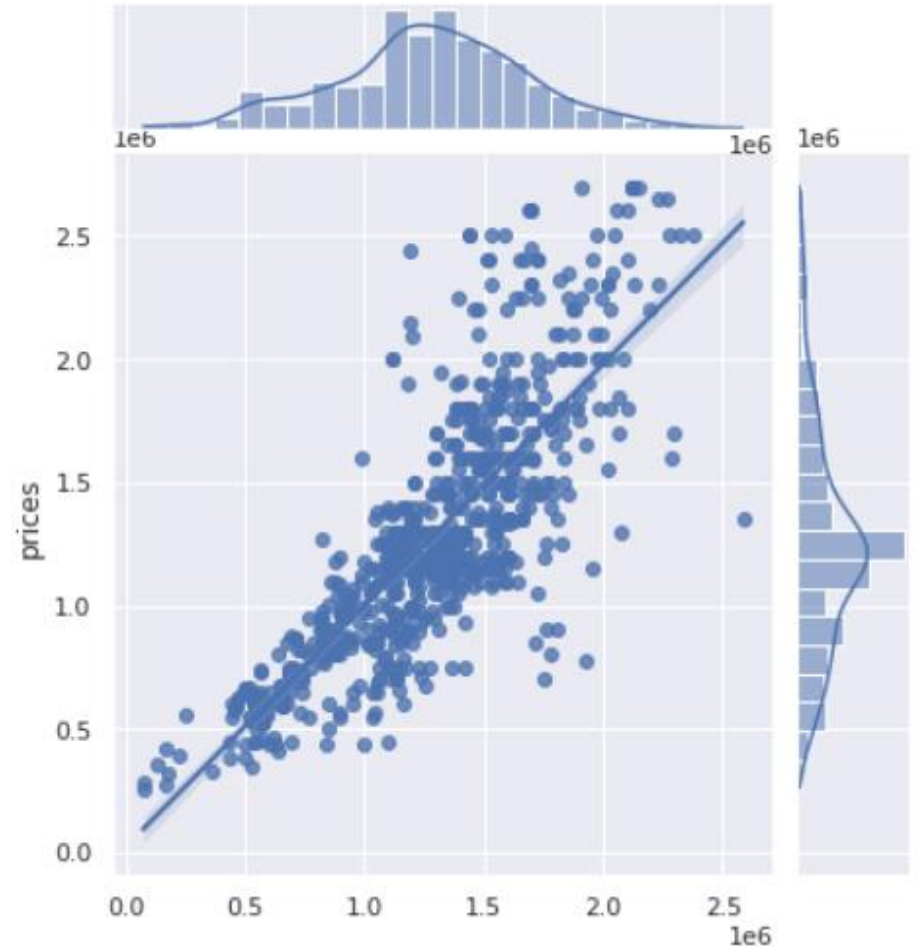


Linear regression assumptions



Baseline model

We split our data into train validate and test we choose to preform the cross validation on our data to make sure to have the best result and the score was R^2 0.638

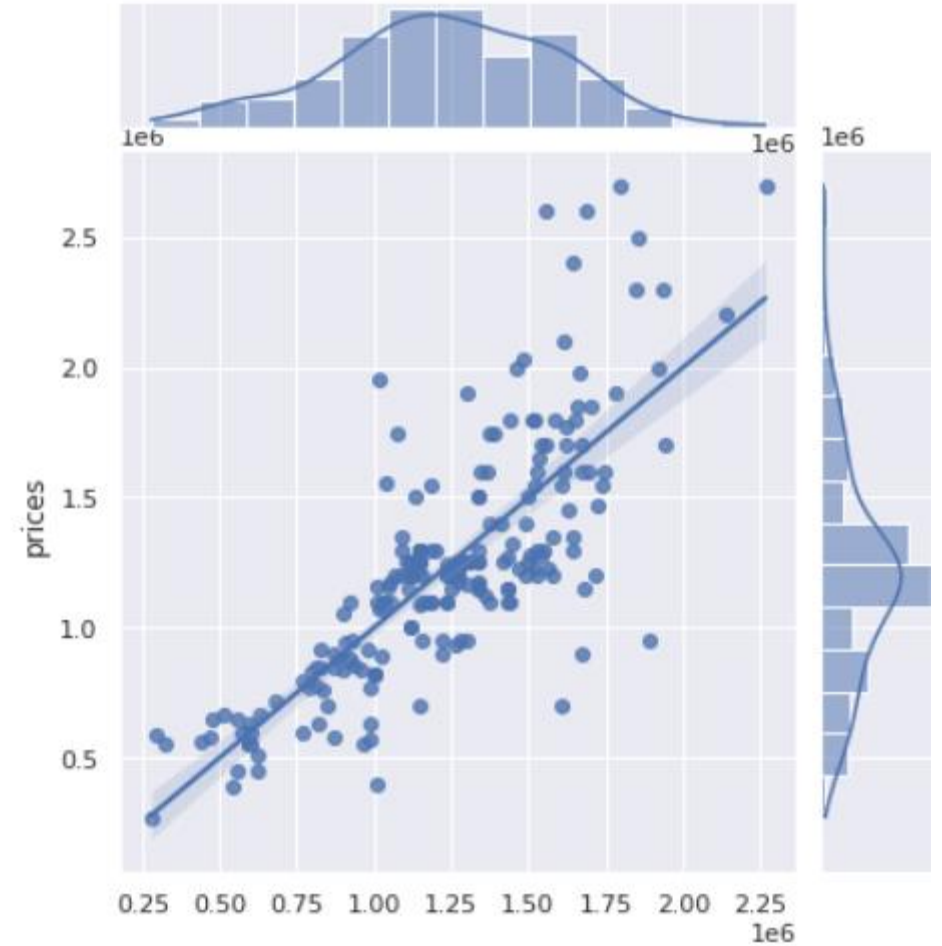


Now we want to improve our model so we are going to perform feature engineering on the data we exclude some of the features also we are going to use lasso for feature selection since the coefficients we have is high

Feature engineering R^2 score: 0.648

LASSO R^2 score: 0.362

Test to see if our model is generalized or not so the result is on the test set is R^2 score=0.590



CONCLUSION

- **The linear assumptions is valid on our data**
- **Linear model is the best model to represent our data**
- **The base line R^2 score is 0.638**
- **The LASSO R^2 score is 0.362**
- **The feature engineering R^2 score is 0.648**
- **The R^2 score of the test set is 0.590**

A close-up photograph of a person's hand, wearing a dark suit jacket and a white shirt, reaching out towards the viewer. The hand is positioned on the left side of the frame. On the right side, there is a dark blue rectangular overlay with a white triangle pointing downwards at the top center. Inside this overlay, the word "THANKS" is written in large, bold, white capital letters.

THANKS