

REAL ESTATE MARKET VALUATION

By Group 4 Hammad Arif (21100305)
Grey Hats Shahzeb Naveed (21110316)
 Mustafa Hassan(21100298)
 Danyal Haroon (21100041)
 Arham Saeed (21100062)



PROBLEM STATEMENT

Current situation of Pakistan's real estate market

BY AIYZA JAVAID , (LAST UPDATED NOVEMBER 20, 2016)

And its long-term economic forecast

A recent research survey has indicated that following the amendments in the Income Tax Ordinance-2001, about 16,000 out of 18,000 real estate developers have finished their operations in the local real estate market

Innovation Is The Way To Go For Real Estate Sector In Pakistan

February 25, 2020 Guest Desk Estate, Innovation, Pakistan, Real, sector

The real estate sector in Pakistan has been going through good and bad times due to the country's economic situation.

Unfair games not a level playing field

Mansoor Ahmad December 2, 2020

Iqra Aziz Shares Loved-up Photo With Yasir Hussain, Says She 'misses Him'

Minal Khan Posts Touching Snap With Mother During

Even though, Real Estate Market is one of the biggest sectors in Pakistan, there is a wide variation in value of properties (with the same general amenities) across the country, which makes it difficult for buyers and sellers to evaluate true market value of a property?



ROAD MAP OF DATA SCIENCES PROJECT

PHASE 1

Data Collection

PHASE 2

Data Cleaning

PHASE 3

Exploratory Data Analysis

PHASE 4

Building a Machine Learning Model

PHASE 5

Evaluation of the model and Conclusion





Data Overview

	property_id	location_id	page_url	property_type	price	location	city	province_name	latitude	longitude	baths	area	purpose	bedrooms	date_added
0	237062	3325	https://www.zameen.com/Property/g_10_g_10_2_gr...	Flat	10000000.0	G-10	Islamabad	Islamabad Capital	33.679890	73.012640	2.0	4 Marla	For Sale	2.0	2/4/2019
1	346905	3236	https://www.zameen.com/Property/e_11_2_service...	Flat	6900000.0	E-11	Islamabad	Islamabad Capital	33.700993	72.971492	3.0	5.6 Marla	For Sale	3.0	5/4/2019
2	386513	764	https://www.zameen.com/Property/islamabad_g_15...	House	16500000.0	G-15	Islamabad	Islamabad Capital	33.631486	72.926559	6.0	8 Marla	For Sale	5.0	7/17/2019
3	656161	340	https://www.zameen.com/Property/islamabad_bani...	House	43500000.0	Bani Gala	Islamabad	Islamabad Capital	33.707573	73.151199	4.0	2 Kanal	For Sale	4.0	4/5/2019
4	841645	3226	https://www.zameen.com/Property/dha_valley_dha...	House	7000000.0	DHA Defence	Islamabad	Islamabad Capital	33.492591	73.301339	3.0	8 Marla	For Sale	3.0	7/10/2019
...	
179833	17355248	6754	https://www.zameen.com/Property/gulshan_e_maym...	House	26500000.0	Gadap Town	Karachi	Sindh	25.029909	67.137192	0.0	9.6 Marla	For Sale	6.0	7/18/2019
179834	17355249	680	https://www.zameen.com/Property/gadap_town_gul...	House	12500000.0	Gadap Town	Karachi	Sindh	25.017951	67.136393	0.0	8 Marla	For Sale	3.0	7/18/2019
179835	17355250	6757	https://www.zameen.com/Property/gulshan_e_maym...	House	27000000.0	Gadap	Karachi	Sindh	25.015384	67.116330	0.0	9.6 Marla	For Sale	6.0	7/18/2019

Overview of the Raw Dataset

WHY ZAMEEN.COM?

- Reliable Dataset
- True (Closest) Representative of Population
- No other official extensive data available on property

**168,446 ROWS
19 COLUMNS**

Understanding The Variables

Property_id, Location_id

Property_type: States whether the property is a flat or house.



Price: Mentions the selling price of the mentioned property in Rupees.



Location: the area within the city ie G-10

City, Province_name

Latitude, Longitude



Bath: The total number of bathrooms available in the property.



Area: mentions the area of the property in Kanal or Marla.

Purpose: States whether the property is for sale or rent.



Bedrooms: The total number of bedrooms available.

Date_added: The date property was listed on Zameen.com

Description: Additional Information.



Agency: Whether the property is listed directly by the seller or by an Agency Amenities

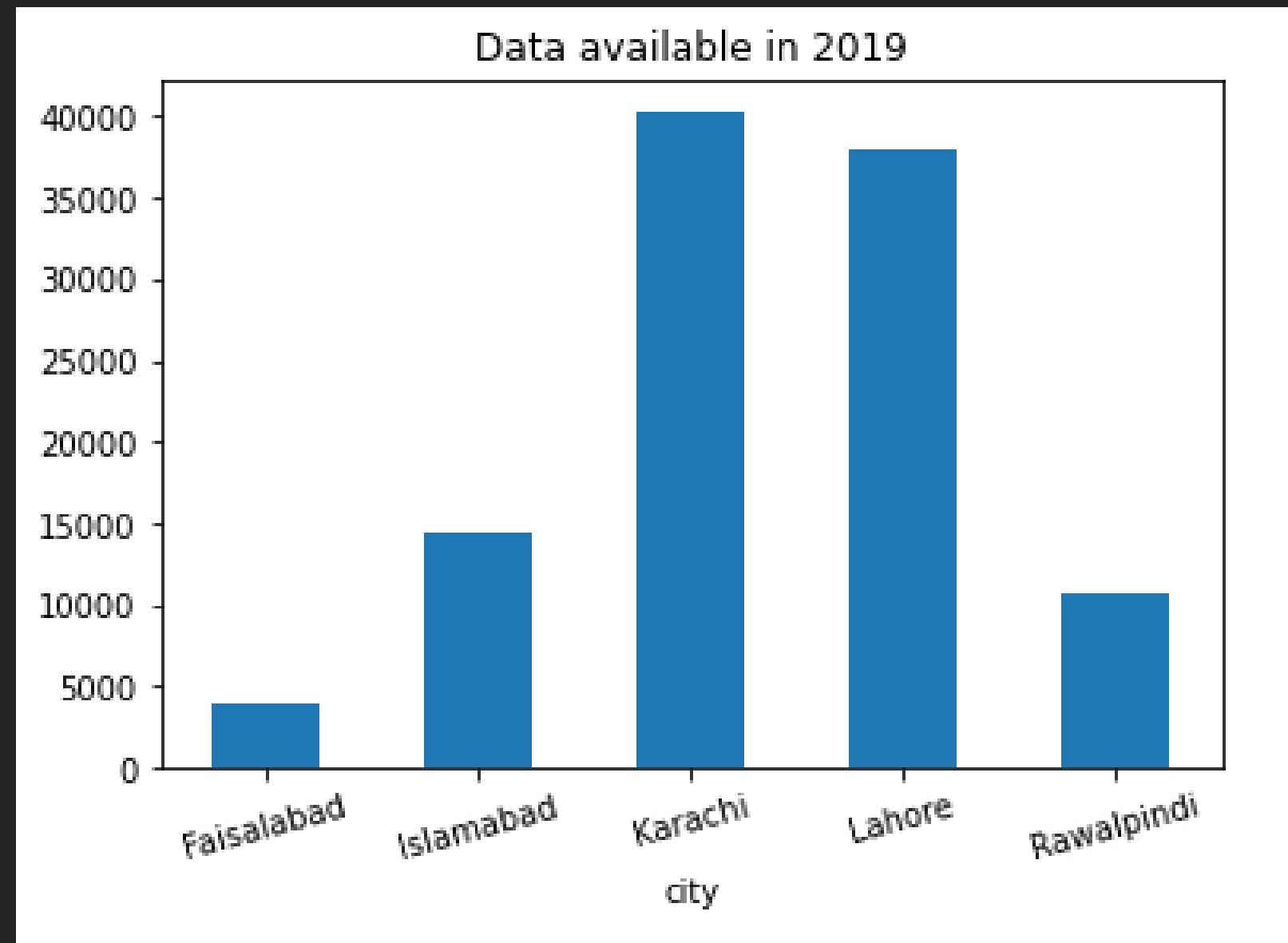
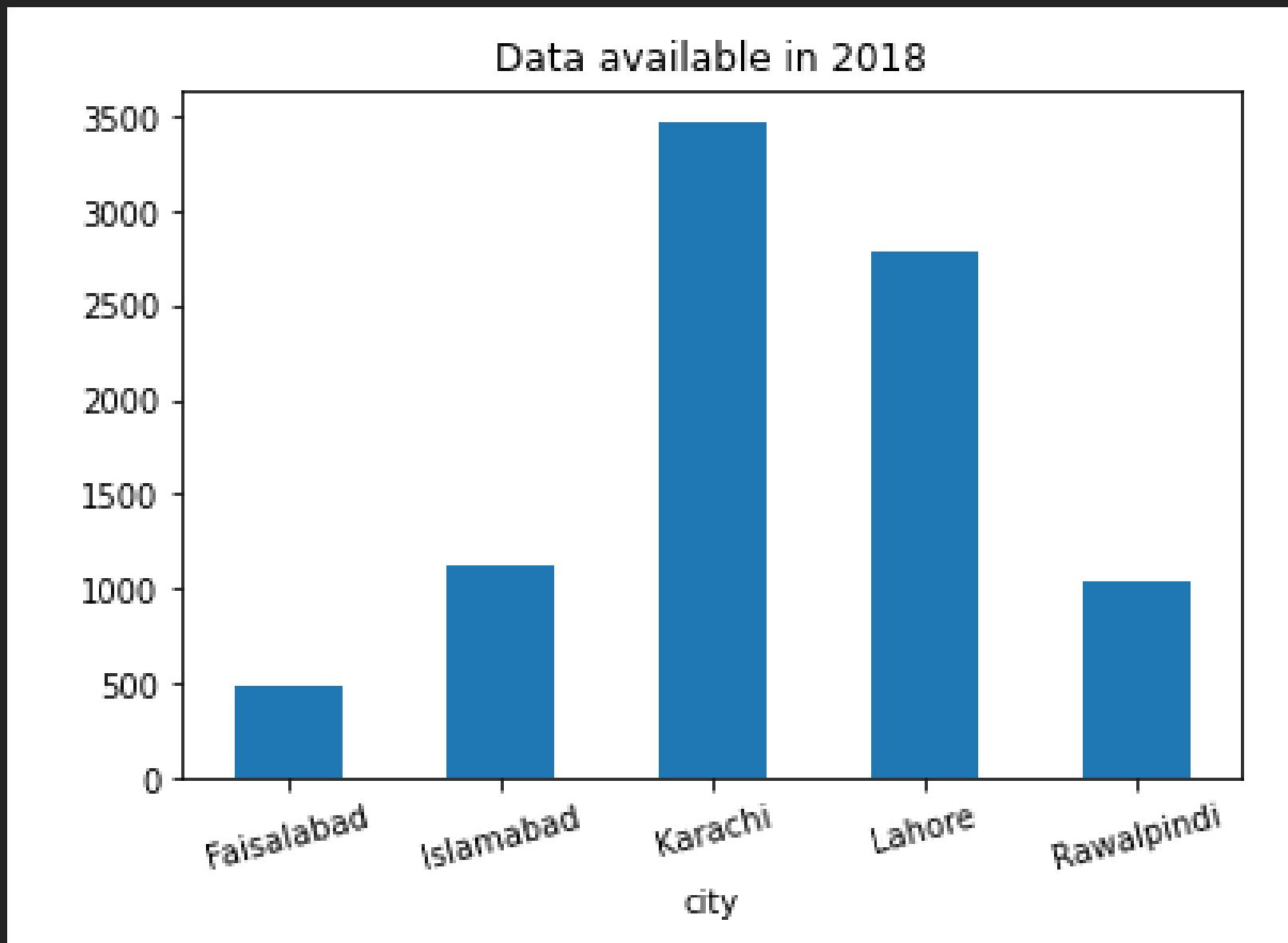


Data Visualisation



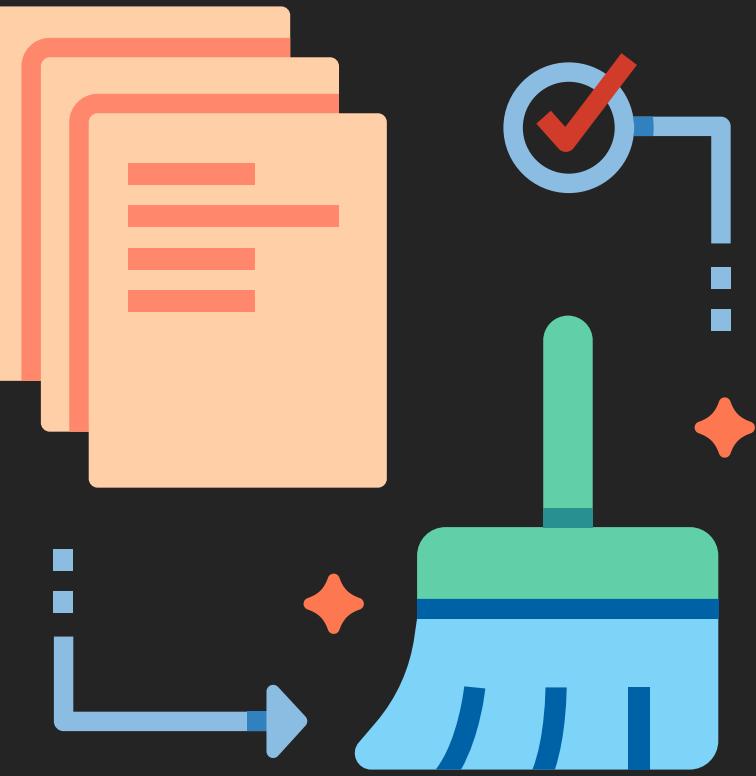
Comparison of Data available on Yearly basis

The dataset encompasses property listings on zameen.com from the last 2 quarters of 2018 and whole of 2019, with the majority of data coming from 2019 for the cities of Karachi and Lahore.



Data Cleaning

After understanding our data, we further refined our data by undertaking the following steps:

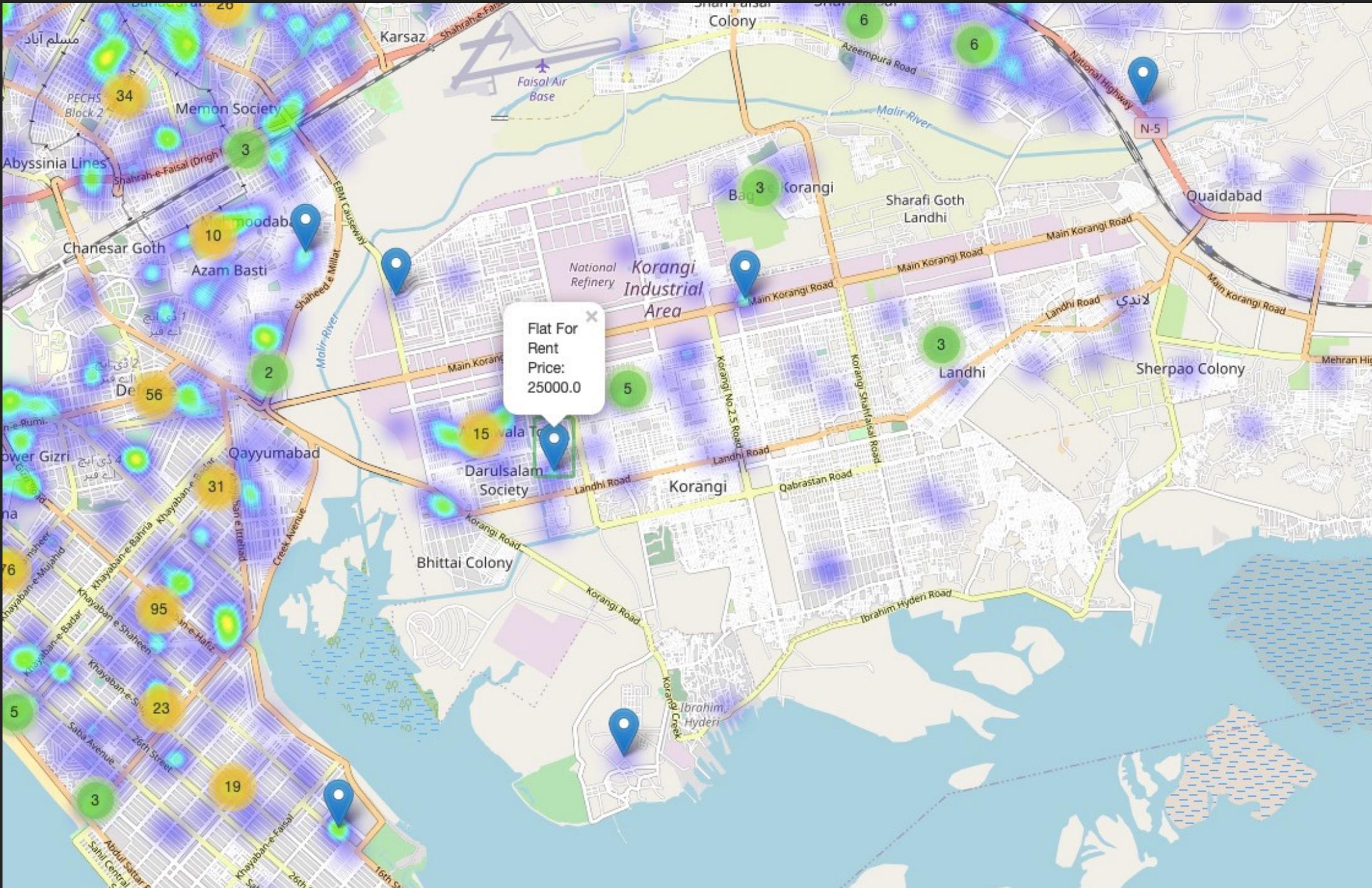


- The units of the values under the area column were converted from Kanal and Marla to a square feet.
- The data was further divided into yearly and quarterly data.
- The 'NaN' entries under the Agency columns were converted to a 'No.'
- We also removed outliers which were not in line with the market value from the price column. This was done to improve data accuracy.

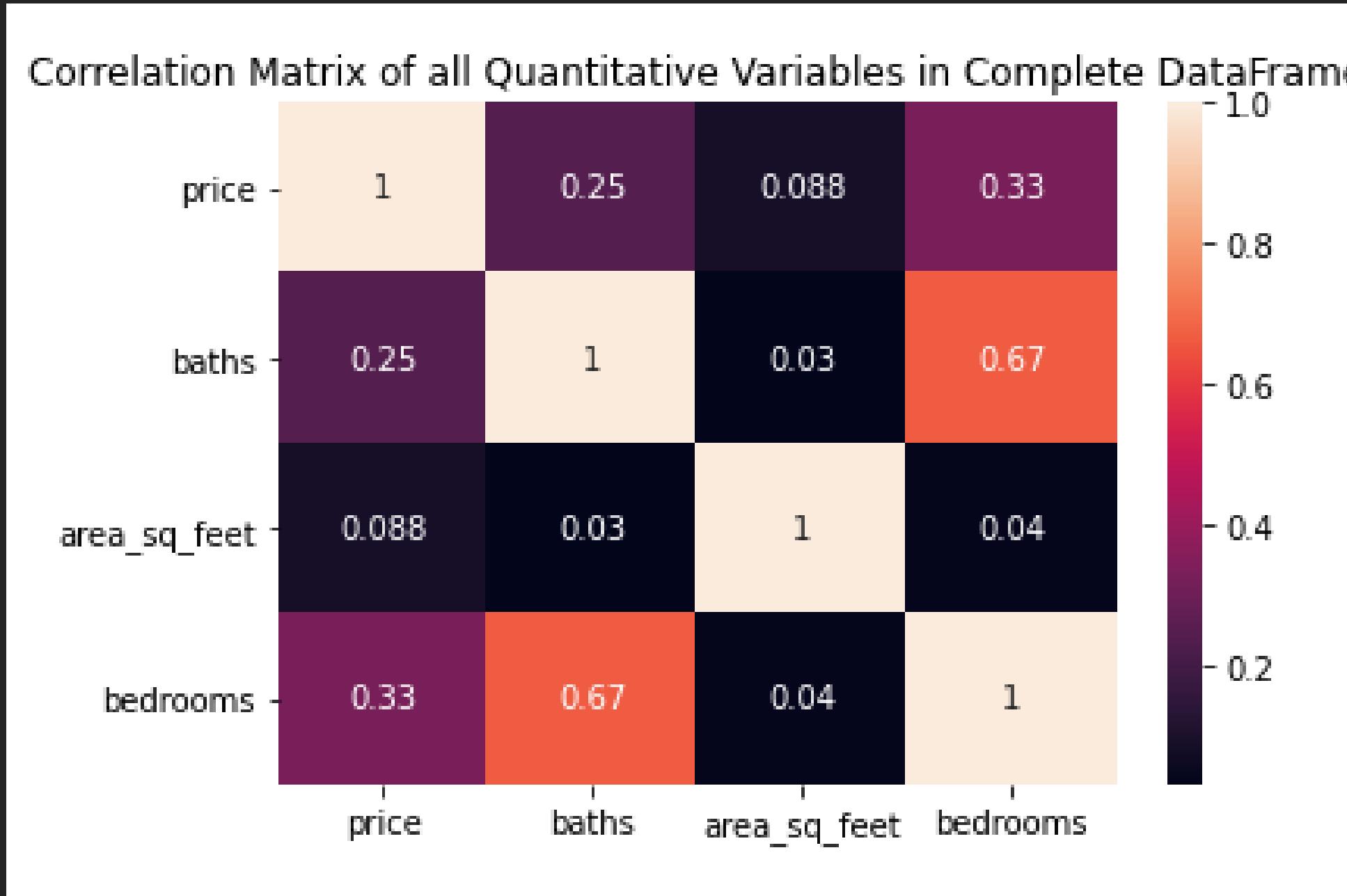
CITY-WISE GEOSPACIAL ANALYSIS

Insights:

Our map shows the spread of properties across Karachi, Lahore, Islamabad and Faisalabad. It reveals which areas are popular for sale and for rent. This will be a good basis on where one may be able to find variations in price between areas within cities.



Analyzing Correlations



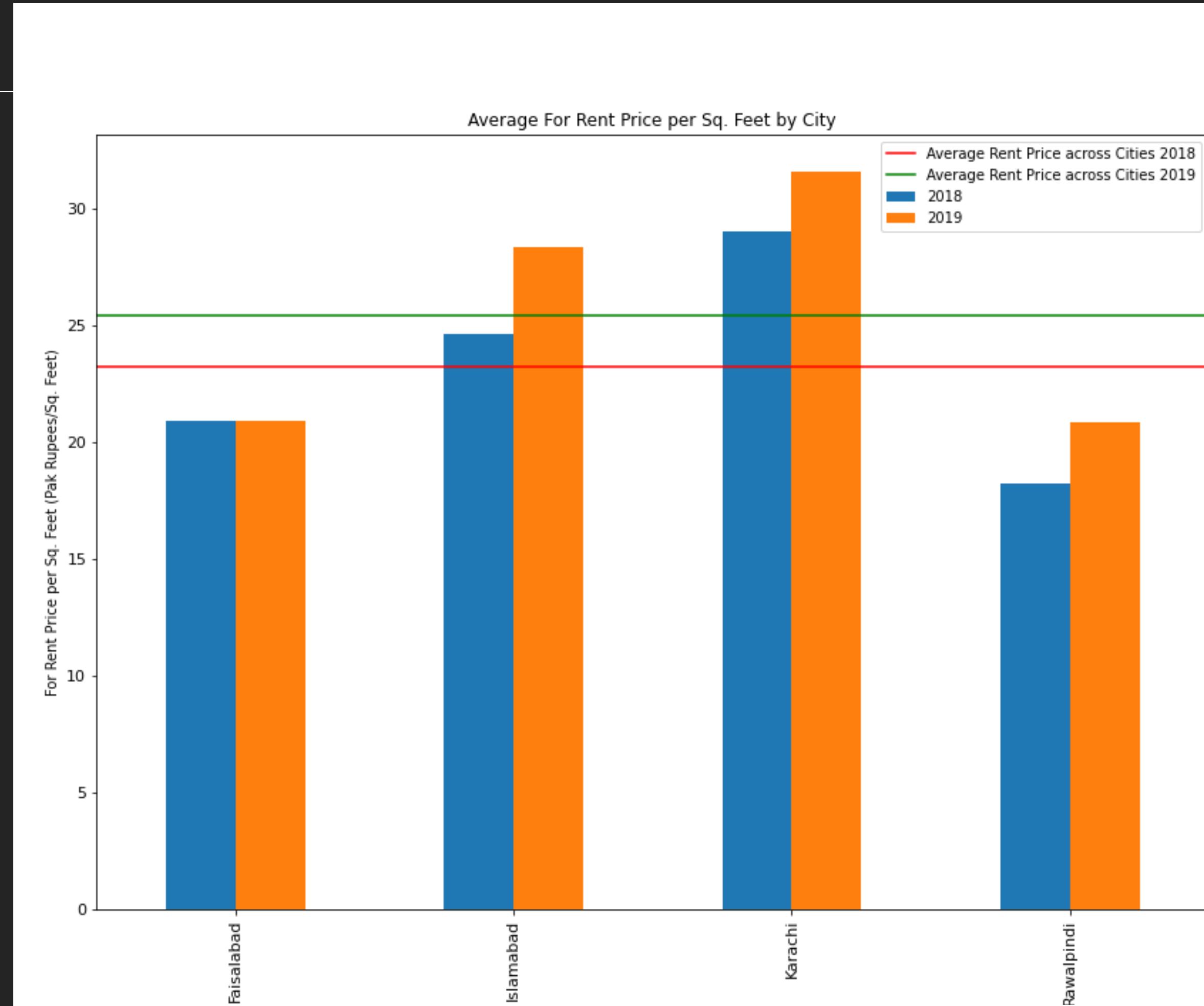
CORRELATION MATRIX

This is a correlation matrix for the price, area per sq feet, bedroom and bathrooms for the properties.

● ●

Price Per Sq. Ft

The above analysis shows that hiring an agency will definitely help increase the average selling price by 6.8% and the average rental value by 17.4%. We, therefore, recommend hiring an agency to sell or rent a property across the country.

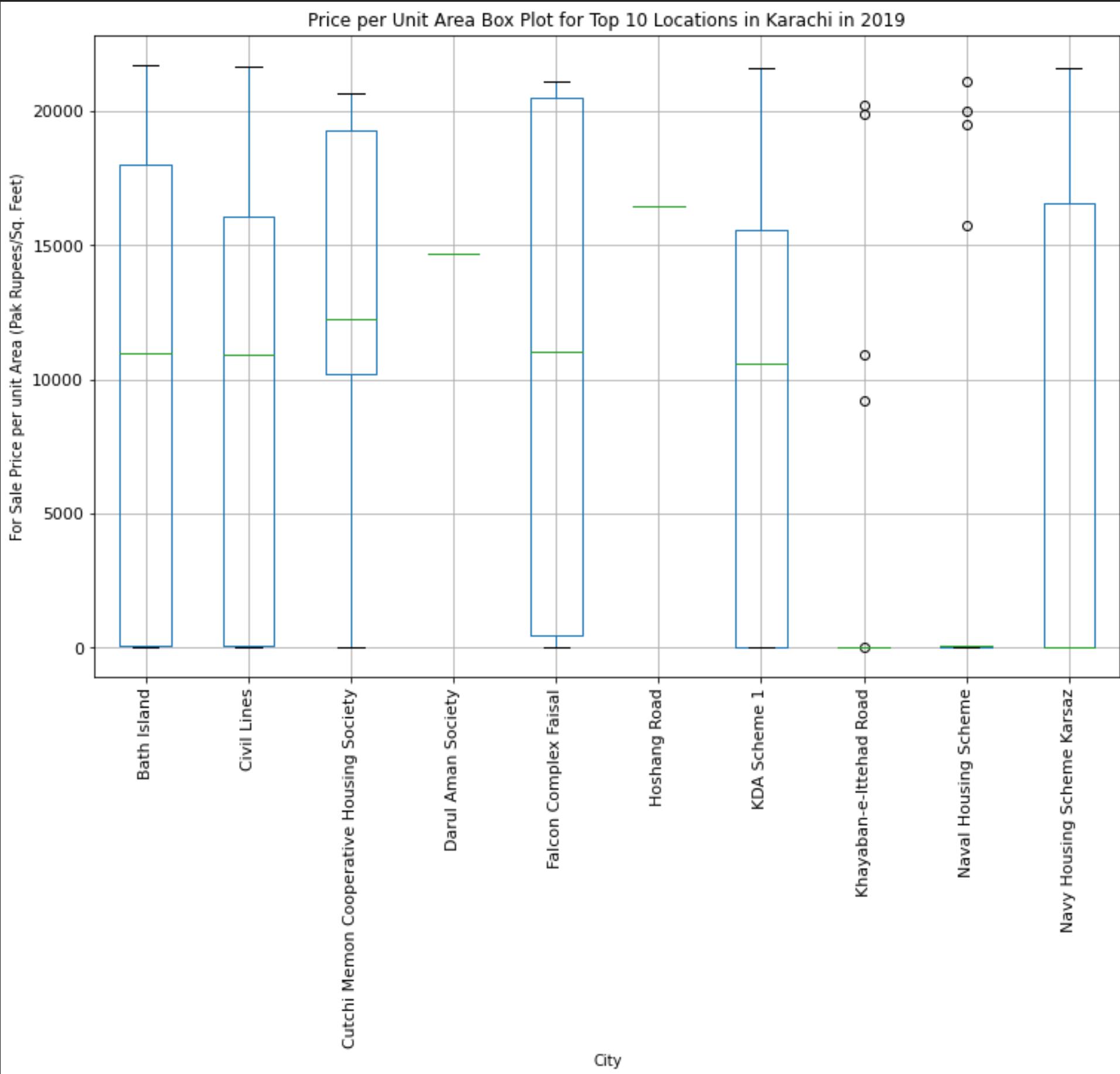


Price Per Unit Area

2018 AND 2019

Compared the selling price of different cities between 2018 and 2019

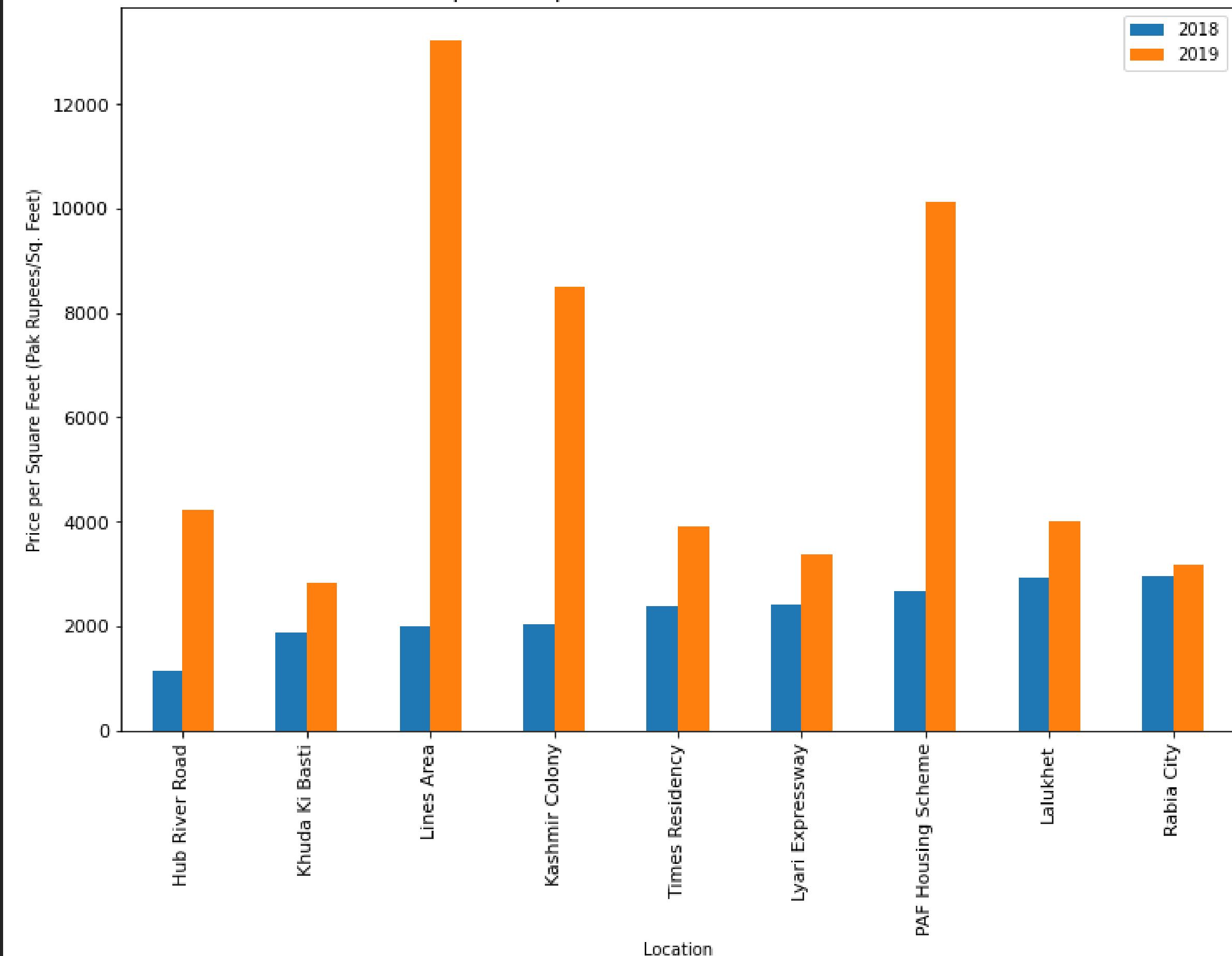
The average selling price has increased from 7000 Rupees per square feet to 7500 Rupees per square feet, an increase of 7.14%



Cheapest Areas

We have seen that Karachi has been the hotspot for the real estate sector. Therefore, in the figure below we have attempted to identify the ten cheapest areas in Karachi in 2019 and 2018.

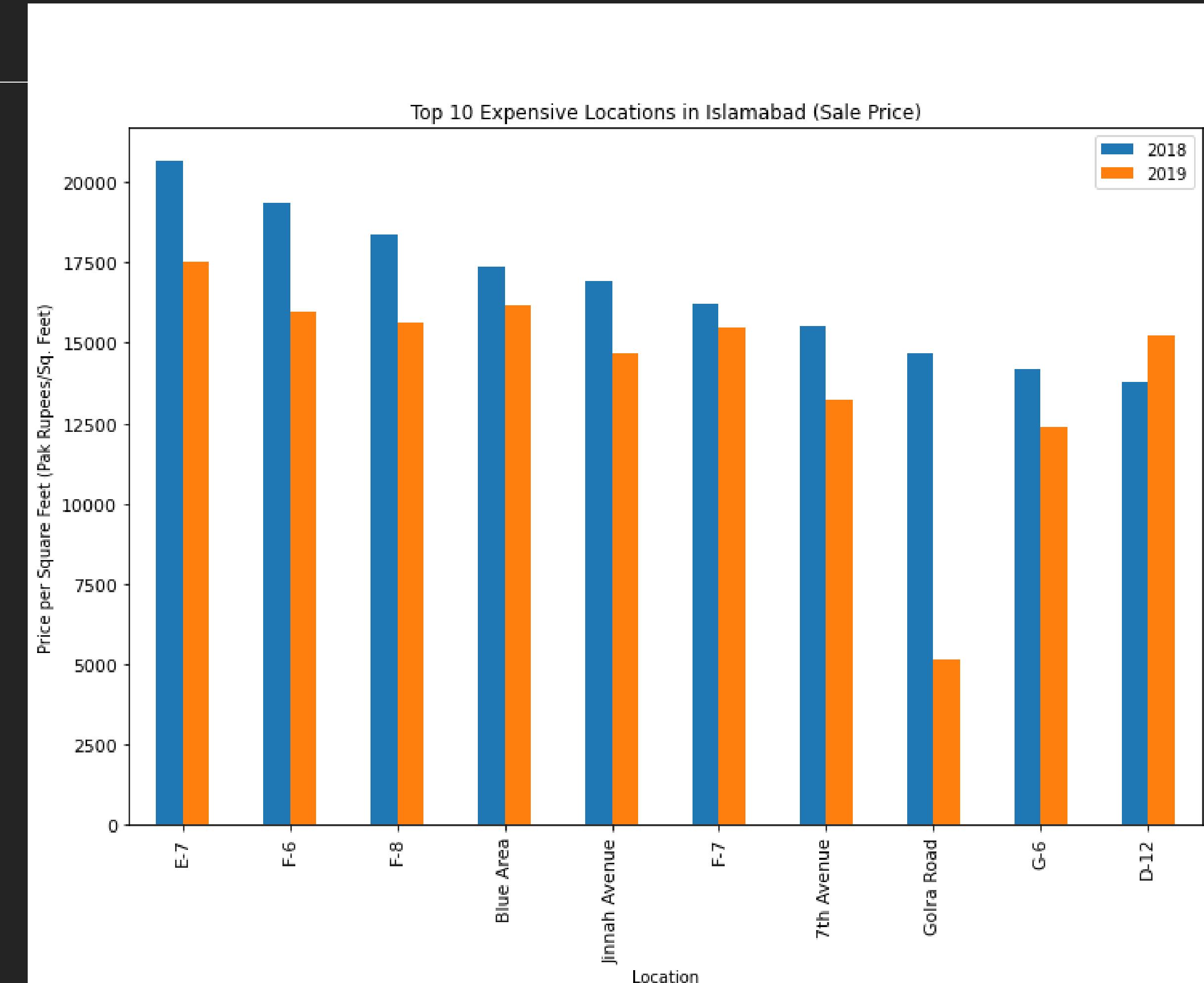
Top 10 Cheapest Locations in Karachi (Sale Price)



••

Most Expensive Areas

Islamabad has some of the most expensive properties in the country. In the figure below we have attempted to identify the ten most expensive areas in Islamabad in 2019 and 2018.



Machine Learning

1) COLLECTING DATA

2) INITIAL DATA CLEANING

3) CLEANING MISSING DATA FOR MACHINE LEARNING

- Handling Categorical like & continuous Variables like longitude and latitudes
- Outlier Clipping (per-value, not via SQL)
- Data Normalization / Feature Scaling (Price per sq rt)

4) TRAIN & EVALUATE

- Linear Regression with L2 Regularization/Ridge Regression
- Cross Validation
- Trained the model on log(Price)
- Feature Engineering - hot encoding (encoding categorical and text features)
- Best alpha estimated



..

Machine Learning

- **3 QUANTITATIVE FEATURES: NO OF BATHROOMS, BEDROOMS, AREA (SQFT)**
- **4 QUALITATIVE FEATURES: LOCATION, PROPERTY TYPE, CITY, PURPOSE (RENT/BUY)**
- Cross validation:
- 30% of Training Data in Cross Validation Training Data
- 10% of remaining Training Data to check Cross Validation
- Alpha -> Checked from 0.1 to 0.0001
- Best Alpha: 0.01

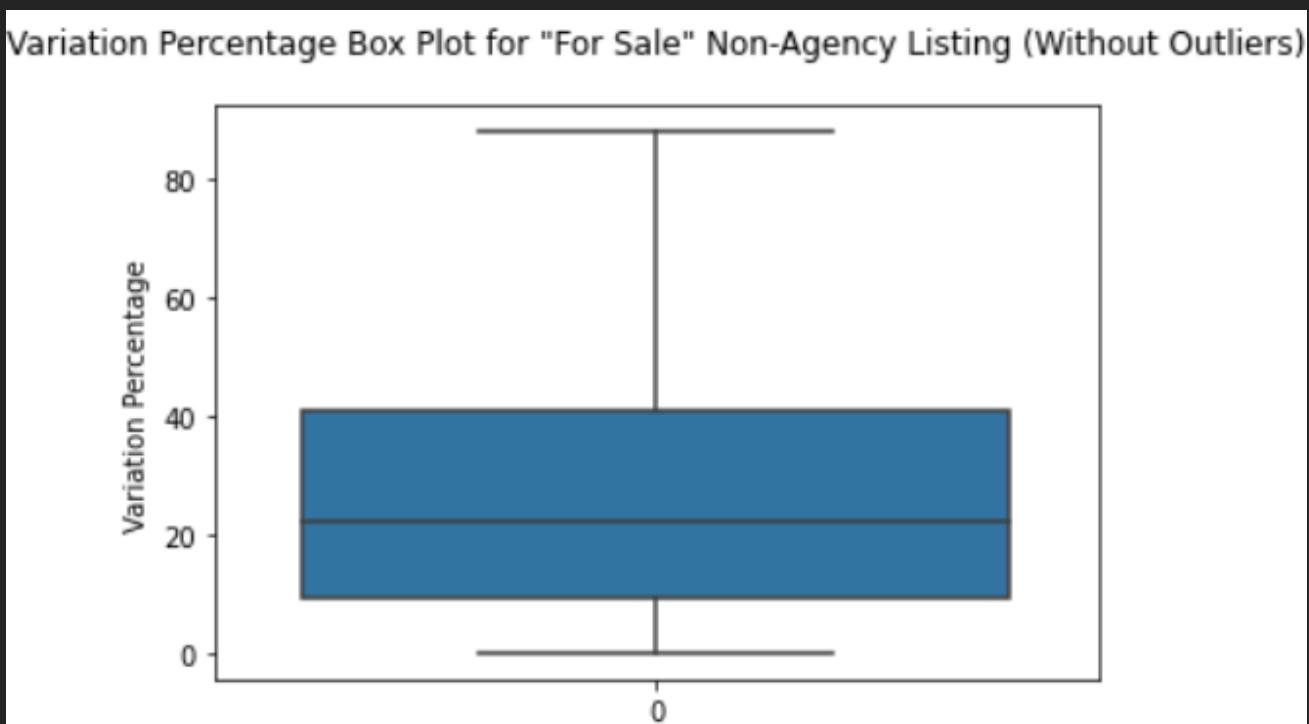
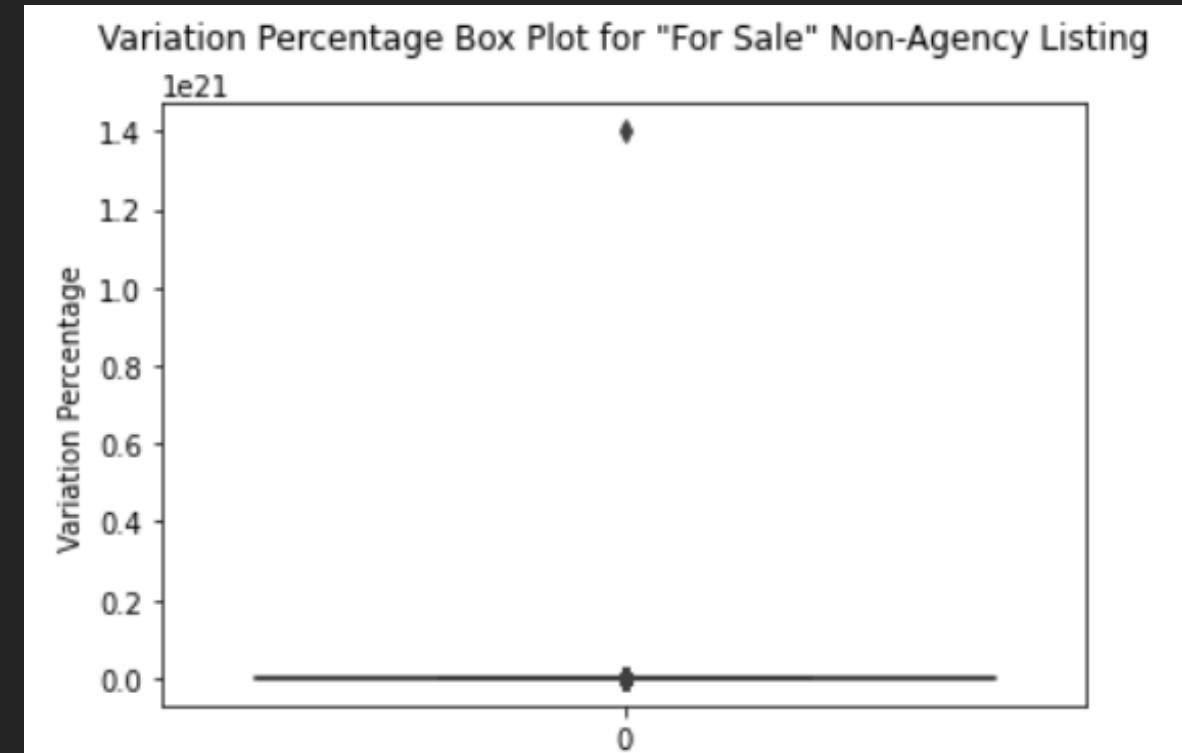
Results

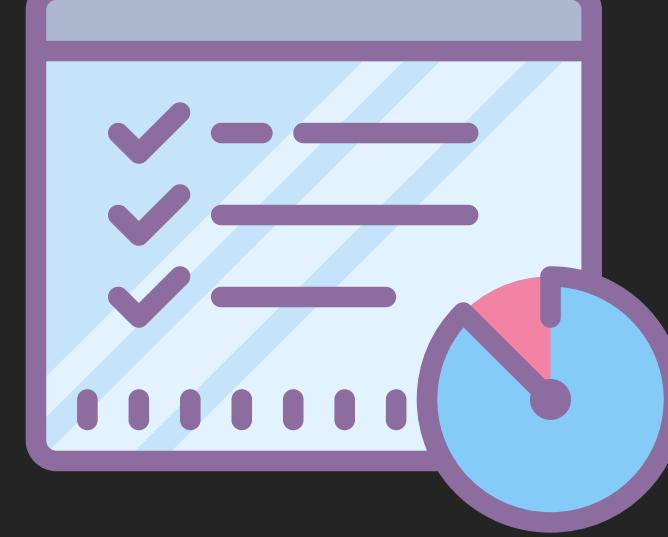
	<i>With Agency</i>	<i>Without Agency</i>
<i>Mean Square Error of Linear Regression for Test:</i>	0.23416	0.28486
<i>Root Mean Square Error of Linear Regression for Test:</i>	0.48390	0.55337
<i>Mean Square Error of Linear Regression for Train:</i>	0.17004	0.18127
<i>Root Mean Square Error of Linear Regression</i>	0.41236	0.42576
<i>Mean Log Square Error of Linear Regression</i>	0.00106	0.00094
<i>R Square of Linear Regression</i>	0.97081	0.96076

Outliers

Outliers in Predicted Prices

- Non-Agency Listings resulted in a few (0.5-1.5%) extreme predictions.
- Agency Listings resulted in no extreme predictions.
- This gives birth to the hypothesis that Ordinary Users compared to Real Estate Agencies, have more chances of listing inconsistent prices.





Limitations

- Due to unavailability of data across different years, analysis and machine learning for across time was not possible.
- Other factors also affect the valuations of real estate, e.g. availability of amenities nearby property.
- Only the bid price of the properties are available, not the transaction price.

⋮

CONCLUSION

••

Our analysis of Zameen.com's property data revealed several interesting trends of online property listings in Pakistan:

- 1) Properties listed through property agents are much more accurate in terms of price.
- 2) The model can be used to [predict a range of prices](#) (Fair Market Range) for a given property on a property listing website e.g Zameen and shown along with the listed price. This is similar to how the Kelly Blue Book Range is shown on price listing sites in the US.
- 3) There exists a variance in prices of similar properties. One of the reasons is that there is a large margin between transaction and demand prices of properties.

