Bengaluru House Price Prediction

A Data Analytics Project

By- Harshvardhan Bahukhandi, Swayam Kaushal , Vaibhav Arya

Contents

01

Introduction

To the dataset.

Introduction

Welcome to the Bangalore House Price Prediction project presentation. In this project, we delve into the dynamic realm of real estate in Bangalore, leveraging data analytics to discern the factors that significantly influence housing prices in the city.

Dataset Features:
Area Type
Availability
Location
Size
Society
Total Square Footage (Total_sqft)
Number of Bathrooms (Bath)
Number of Balconies (Balcony)
Price

Dataset overview

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Data Cleaning

Handling missing values

Null values in the dataset were begin identified and dropped from further analysis.

Standardize Data Types

Location, size, total_sqft ,bath and price is being further selected.

Handling outliers

Outliers in terms of house size is been handled.

Dataset cleaning

Shape of data after dropping na values and selection of specific dimensions.

```
df3.shape
```

```
(13246, 5)
```

	location	size	total_sqft	bath	price
0	Electronic City Phase II	2 BHK	1056	2.0	39.07
1	Chikka Tirupathi	4 Bedroom	2600	5.0	120.00
2	Uttarahalli	3 BHK	1440	2.0	62.00
3	Lingadheeranahalli	3 BHK	1521	3.0	95.00
4	Kothanur	2 BHK	1200	2.0	51.00

Converting total_sqft into numeric data by taking average in case of ranged values.

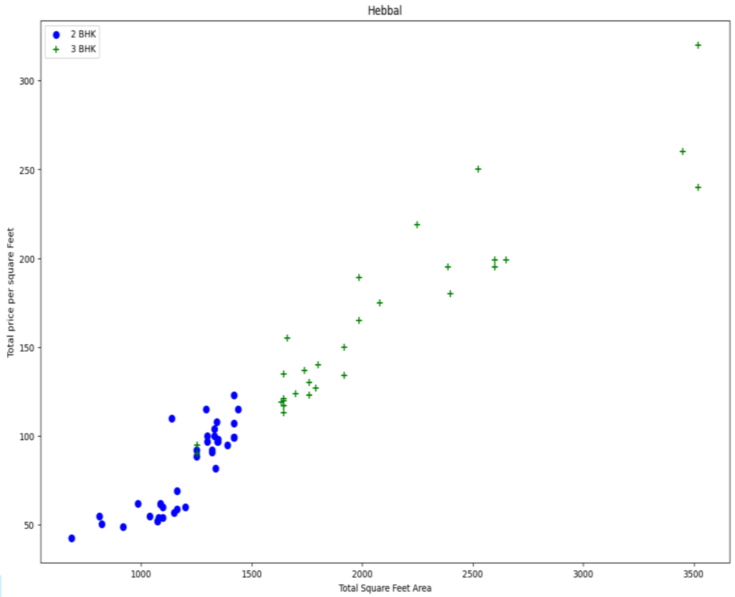
	location	size	total_sqft	bath	price	bhk
30	Yelahanka	4 BHK	2100 - 2850	4.0	186.000	4
122	Hebbal	4 BHK	3067 - 8156	4.0	477.000	4
137	8th Phase JP Nagar	2 BHK	1042 - 1105	2.0	54.005	2
165	Sarjapur	2 BHK	1145 - 1340	2.0	43.490	2
188	KR Puram	2 BHK	1015 - 1540	2.0	56.800	2
410	Kengeri	1 BHK	34.46Sq. Meter	1.0	18.500	1
549	Hennur Road	2 BHK	1195 - 1440	2.0	63.770	2
648	Arekere	9 Bedroom	4125Perch	9.0	265.000	9
661	Yelahanka	2 BHK	1120 - 1145	2.0	48.130	2
672	Bettahalsoor	4 Bedroom	3090 - 5002	4.0	445.000	4

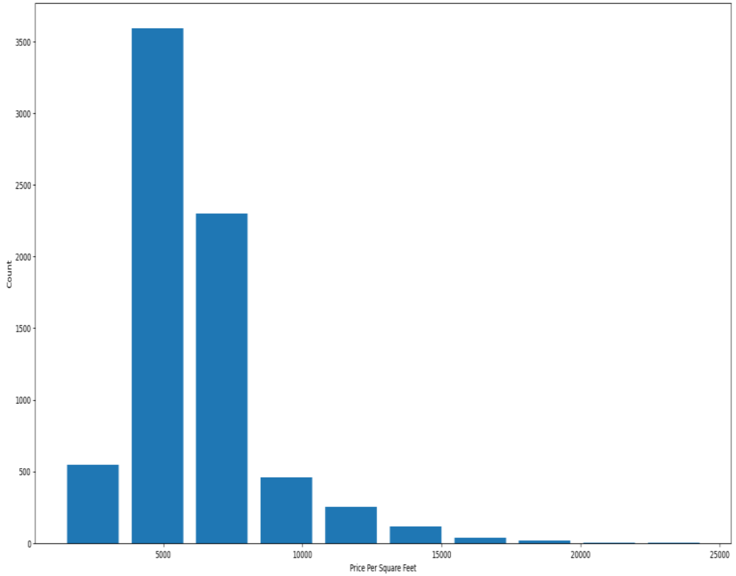
Removed ranged data from the dataset's column ['total_sqft'].

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699.810606
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615.384615
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305.555556
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245.890861
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250.000000

More data cleaning :

- Removed the data outliers which had the ratio of total_sqft to bhk less than 300.
- Plotted a scatter plot and histogram graph for the data using location as Hebbal.





Data cleaning continued:

- Removed the anomaly of no. of bathrooms more than the no. of rooms in a house.

```
df8[df8.bath>df8.bhk+2]
```

✓ 0.0s

	location	size	total_sqft	bath	price	bhk	price_per_sqft
1626	Chikkabanavar	4 Bedroom	2460.0	7.0	80.0	4	3252.032520
5238	Nagasandra	4 Bedroom	7000.0	8.0	450.0	4	6428.571429
6711	Thanisandra	3 BHK	1806.0	6.0	116.0	3	6423.034330
8411	other	6 BHK	11338.0	9.0	1000.0	6	8819.897689

Preparing the dataset for model:

- Removed size and price per sqft form the dataset.

	location	total_sqft	bath	price	bhk
0	1st Block Jayanagar	2850.0	4.0	428.0	4
1	1st Block Jayanagar	1630.0	3.0	194.0	3
2	1st Block Jayanagar	1875.0	2.0	235.0	3
3	1st Block Jayanagar	1200.0	2.0	130.0	3
4	1st Block Jayanagar	1235.0	2.0	148.0	2

- Preparing dummies for better prediction using the column locations.

```
dummies = pd.get_dummies(df10.location)
dummies.head(10)
```

✓ 0.0s Python

	1st Block Jayanagar	1st Phase JP Nagar	2nd Phase Judicial Layout	2nd Stage Nagarbhavi	5th Block Hbr Layout	5th Phase JP Nagar	6th Phase JP Nagar	7th Phase JP Nagar	8th Phase JP Nagar	9th Phase JP Nagar	...	Vishveshwarya Layout	Vis
0	True	False	False	False	False	False	False	False	False	False	...	False	
1	True	False	False	False	False	False	False	False	False	False	...	False	
2	True	False	False	False	False	False	False	False	False	False	...	False	
3	True	False	False	False	False	False	False	False	False	False	...	False	
4	True	False	False	False	False	False	False	False	False	False	...	False	
5	True	False	False	False	False	False	False	False	False	False	...	False	
6	True	False	False	False	False	False	False	False	False	False	...	False	
8	False	True	False	False	False	False	False	False	False	False	...	False	
9	False	True	False	False	False	False	False	False	False	False	...	False	
10	False	True	False	False	False	False	False	False	False	False	...	False	

Final Dataset:

	total_sqft	bath	bhk	1st Block Jayanagar	1st Phase JP Nagar	2nd Phase Judicial Layout	2nd Stage Nagarbhavi	5th Block Hbr Layout	5th Phase JP Nagar	6th Phase JP Nagar	...	Vijayanagar	Vishveshwarya Layout	Vishwapriya Layout	Vittasandra	Whitefield	Yelachenahalli	Yelahanka	Yelahanka New Town	Yelenahalli	Yeshwanthpur
0	2850.0	4.0	4	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1	1630.0	3.0	3	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
2	1875.0	2.0	3	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	1200.0	2.0	3	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	1235.0	2.0	2	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
5	2750.0	4.0	4	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
6	2450.0	4.0	4	True	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
8	1875.0	3.0	3	False	True	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
9	1500.0	5.0	5	False	True	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
10	2065.0	4.0	3	False	True	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False

Model Selection:

- Used Train test split function of sklearn to split our data into training and test dataset.
- Imported 3 model for regressor type of problems:\

```
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.tree import DecisionTreeRegressor

las = Lasso()
tree = DecisionTreeRegressor()
lr_reg = LinearRegression()
```

Comparing the results:

- Lasso model:
  score: 72%

```
las.fit(X_train,y_train)
las.score(X_test,y_test)
```
✓ 0.0s

```
0.7237775279429011
```

- Tree Model:
  score: 71%

```
tree.fit(X_train,y_train)
tree.score(X_test,y_test)
```
✓ 0.1s

```
0.7130411347889515
```

- Linear Regression model:
  score: 84%

```
lr_reg.fit(X_train,y_train)
lr_reg.score(X_test,y_test)
```
✓ 0.1s

```
0.8452277697874349
```

Final model Selection and Prediction Function:

- Final model : Linear Regression.
- Prediction Function :

```
def predict_price(location,sqft,bath,bhk):
    loc_index = np.where(X.columns==location)[0][0]

    x = np.zeros(len(X.columns))
    x[0] = sqft
    x[1] = bath
    x[2] = bhk
    if loc_index >= 0:
        x[loc_index] = 1

    return lr_reg.predict([x])[0]
```

Some Predicitons:

```
predict_price('1st Phase JP Nagar',1000, 2, 2)
✓ 0.0s
C:\Users\Vaibhav\AppData\Local\Packages\PythonSoftw
  warnings.warn(

83.49904677194546
```

```
predict_price('Indira Nagar',1000, 2, 2)
✓ 0.0s
C:\Users\Vaibhav\AppData\Local\Packages\Python
  warnings.warn(

181.27815484006592
```

```
predict_price('1st Phase JP Nagar',1000, 3, 3)
✓ 0.0s
C:\Users\Vaibhav\AppData\Local\Packages\PythonSoftwar
  warnings.warn(

86.80519395221248
```

```
predict_price('Indira Nagar',1000, 3, 3)
✓ 0.0s
C:\Users\Vaibhav\AppData\Local\Packages\Python
  warnings.warn(

184.58430202033293
```

Conclusion

The real estate price prediction model we've developed is a game-changer in the industry, built through careful planning, data analysis, and algorithmic innovation. By using advanced techniques and algorithms like linear regression, decision trees, and lasso regression, we've created a versatile tool that offers precise price estimates and insightful market analysis. This model boosts profitability, minimizes risks, and empowers stakeholders with the knowledge to navigate the dynamic real estate market effectively.