REAL ESTATE MARKET SEGMENTATION

Analyzing the Real Estate Market in India

for HousingApp Tech Startup

Team Lead: Prashant Srivastava

Prashant Srivastava	*https://github.com/prashhhant213/Real-Estate-Market-Segmentation*
V Harsha Vardhan	*https://github.com/H234Arsh/Feynn-Labs-Task-3*
Abhishikta Dhar	*https://github.com/mistiiberry-exe/market_segment_analysis_real_estate*
Poshak Prajeet	*https://github.com/P0shak/Real-Estate-MS/blob/main/README.md*

ABSTRACT

Real estate sector is one of the most globally recognized sectors. It comprises of 4 sub sectors. Housing, Retail, Hospitality, & Commercial. The construction industry ranks third among the 14 major sectors in termsof direct, indirect and induced effects in all sectors of the economy.
In India, the real estate sector is the second-highest employment generator, after the agriculture sector. It is also expected that this sector will incur more non-resident Indian (NRI) investment, both in short termand the long term.

By 2040, the real estate market in India will grow to a peak. That is, it willreach to Rs. 65000 Crore from Rs. 12000 Crore. It will contribute to the 13% to the country's GDP. The sub sectors like Retail, Hospitality and commercial real estate are also growing significantly to provide high infrastructure.

The market segmentation is the process of dividing a broad group of consumer or target market into sub groups based on demographics, needs, common interests, geographic locations, psychographic behavior, etc. Market segmentation is one of the key building blocks of strategic marketing. Market segmentation is essential for marketing success

Now, we analyze the real estate market in India using Segmentation analysis and come up with a feasible strategy to enter the market in India.So that we offer this market demand information to the Housing App which provides online property rental service and Real Estate Listing Service.

Overview of Real Estate Market in India

Indian firms are expected to raise more than thirty-five trillion rupees INRthrough infrastructure and real estate investment in 2022. As the demandfor cloud resources is increasing, the space required for the cloud setup became unavoidable.

Tech giants like Google, Amazon, Facebook are insearch of land for configuring their data centers. According to the recent report, real estate demand for data centers is expected to increase by 15- 18 Million sq. ft by 2025. Manufacturing industry accounted for 24% increase in their office space. This is due to the enlargement of the E- commerce industry.

People are more likely to buy electronic gadgets andso the electronic manufacturing industry have grown to multifold environment. Bengaluru is expected to be the most favored property investment destination for NRIs, followed by Ahmedabad, Pune, Chennai, Goa, Delhi and Dehradun.

Indian real estate sector has witnessed high growth in recent times with rise in demand for office as well as residential spaces. Top three cities inIndia recorded a huge investment in total real-estate sector in 2021.

Mumbai - 39%

NCR Delhi - 19%

Bengaluru - 19%

To increase this market in India, Government of India along with the governments of respective states has taken several initiatives toencourage development.

The Smart City Project, with a plan to build 100 smart cities, is a prime opportunity for real estate companies. If they establish and succeed in this project, then the growth will be exponential. Government has created an AffordableHousing Fund (AHF) in the National Housing Bank (NHB) with an initial corpus of Rs. 10,000 crores.

The low home loan interest rates regime is expected to drive the housing demand and increase sales by 35-40% in the festive season in 2022.

DATA SOURCES

We have gathered some datasets which are somehow related to the case.Multiple datasets are combined together to gain good insights.

https://prsindia.org/policy/analytical-reports/land-records-and-titles-india

Property Rates Data:

https://www.magicbricks.com/Property-Rates-Trends/ALL-COMMERCIAL-rates-in-New-Delhi

https://rbidocs.rbi.org.in/rdocs/Publications/PDFs/RAPMSA07052015FL6A5F839BAB9B4D0190CE127D04DEF74C.PDF

Real Estate Industry Growth – Infographic: https://www.ibef.org/industry/real-estate-india/infographic

Bangalore Real Estate Dataset: https://www.kaggle.com/datasets/amitabhajoy/bengaluru-house-price-data

Packages and Tools used

Numpy
Pandas
Matplotlib
Seaborn
Plotly
Sklearn

PROJECT SYNOPSIS

Importing Packages and tools

Importing Libraries

```
[ ]  import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import plotly as py
     import plotly.graph_objs as go
     import plotly.figure_factory as ff
     import plotly.graph_objects as go
     import plotly.express as px
     from plotly.subplots import make_subplots
```

Uploading Dataset

```
# Importing consumer buying behavior study dataset
df = pd.read_csv("C:\\Users\\SUSNATA BISWAS\\OneDrive\\Desktop\_All_Cities_Cleaned.csv")
df.head()
```

	seller_type	bedroom	layout_type	property_type	locality	price	area	furnish_type	bathroom	city
0	OWNER	2.0	BHK	Apartment	Bodakdev	20000.0	1450.0	Furnished	2.0	Ahmedabad
1	OWNER	1.0	RK	Studio Apartment	CG Road	7350.0	210.0	Semi-Furnished	1.0	Ahmedabad
2	OWNER	3.0	BHK	Apartment	Jodhpur	22000.0	1900.0	Unfurnished	3.0	Ahmedabad
3	OWNER	2.0	BHK	Independent House	Sanand	13000.0	1285.0	Semi-Furnished	2.0	Ahmedabad
4	OWNER	2.0	BHK	Independent House	Navrangpura	18000.0	1600.0	Furnished	2.0	Ahmedabad

Data Pre-processing

Checking are there any null values in the data setThere are no null values in the data. We have 193000+ rows of data with 10 features. By applying df.info (), we can see that we have no null values.

```
[ ] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 193011 entries, 0 to 193010
Data columns (total 10 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   seller_type    193011 non-null  object
 1   bedroom        193011 non-null  float64
 2   layout_type    193011 non-null  object
 3   property_type  193011 non-null  object
 4   locality       193011 non-null  object
 5   price          193011 non-null  float64
 6   area           193011 non-null  float64
 7   furnish_type   193011 non-null  object
 8   bathroom       193011 non-null  float64
 9   city           193011 non-null  object
dtypes: float64(4), object(6)
memory usage: 14.7+ MB
```

Observing Unique values in each features

```
# Observing unique value for object dtype'property_type','House Loan','Wife Working','Make']:city
for col in ['seller_type','bedroom','layout_type','property_type','locality','price','area','furnish_type', 'bathroom' , 'city']:
  print(col,':',df[col].unique())
```

```
seller_type : ['OWNER' 'AGENT' 'BUILDER']
bedroom : [ 2.  1.  3.  4.  5.  6. 10.  7.  8. 15.  9. 12. 14. 11.]
layout_type : ['BHK' 'RK']
property_type : ['Apartment' 'Studio Apartment' 'Independent House' 'Independent Floor'
 'Villa' 'Penthouse']
locality : ['Bodakdev' 'CG Road' 'Jodhpur' ... 'Renuka Nagar' 'Gananjay Society'
 'Aundh Gaon']
price : [20000.  7350. 22000. ... 15120.  8640.  5850.]
area : [1450.  210. 1900. ... 3575. 1721.  258.]
furnish_type : ['Furnished' 'Semi-Furnished' 'Unfurnished']
bathroom : [ 2.  1.  3.  4.  5.  6. 12.  7.  8. 15.  9. 18. 19. 10. 16. 14.]
city : ['Ahmedabad' 'Bangalore' 'Chennai' 'Delhi' 'Hyderabad' 'Kolkata' 'Mumbai'
 'Pune']
```

ANALYSIS

Behavioral and psychographic analysis

Behavioral

Behavior segment extraction the similarities in behavior or reported behavior.A wide range of possible behaviors can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. This type of Survey data is cheap and easy to collect. But survey data can becontaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. Carefully selecting the variables that are included as segmentation variable in common sense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution. Developing a good questionnaire typically requires conducting exploratory or qualitative research. Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Survey data is prone to capturing biases. A response bias is a systematic tendency torespond to a range of questionnaire items on some basis other than the specific itemcontent.

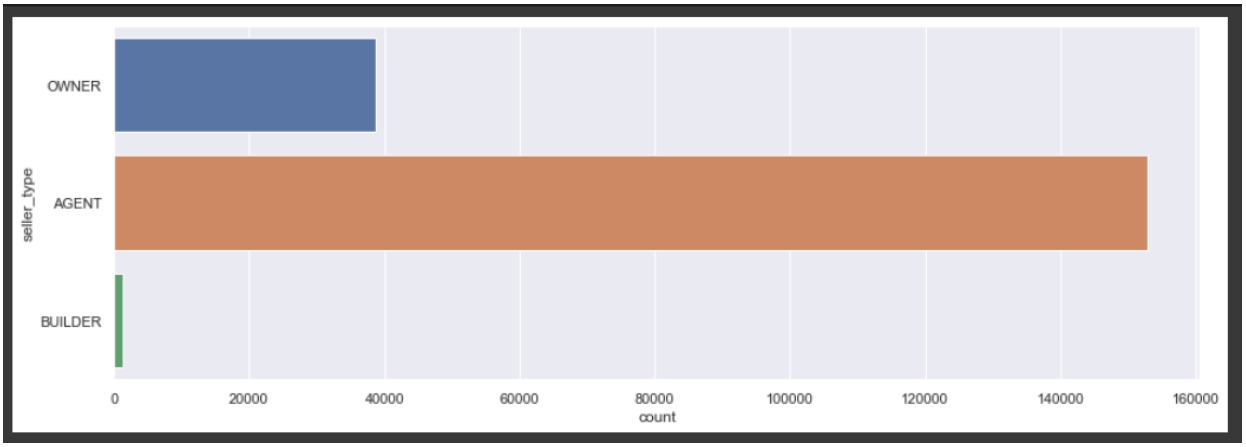
Psychographic Analysis

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term

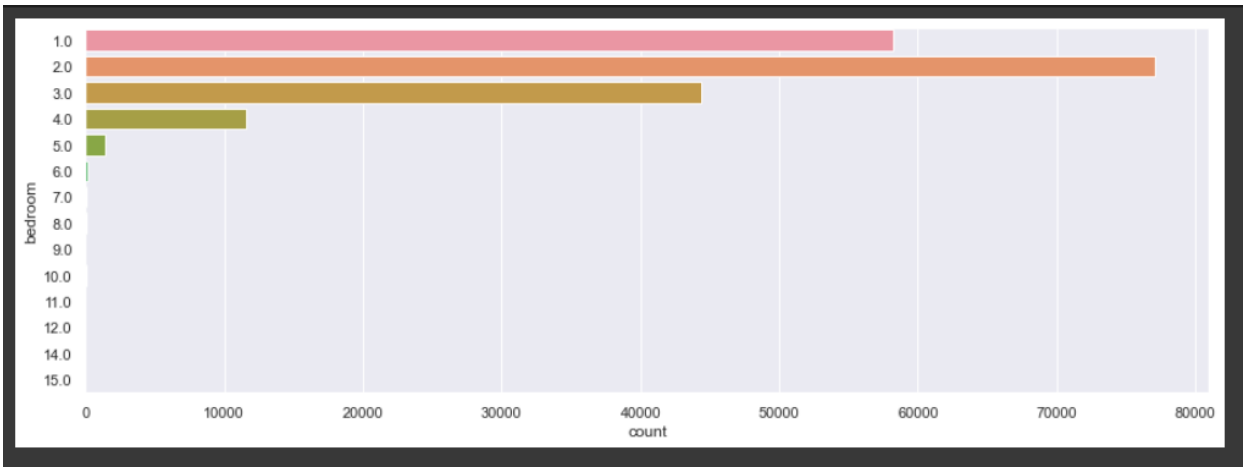
psychographic segmentation is used.

Psychographic criteria are, by nature, more complex than geographic or socio-demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest.

Seller type and its count

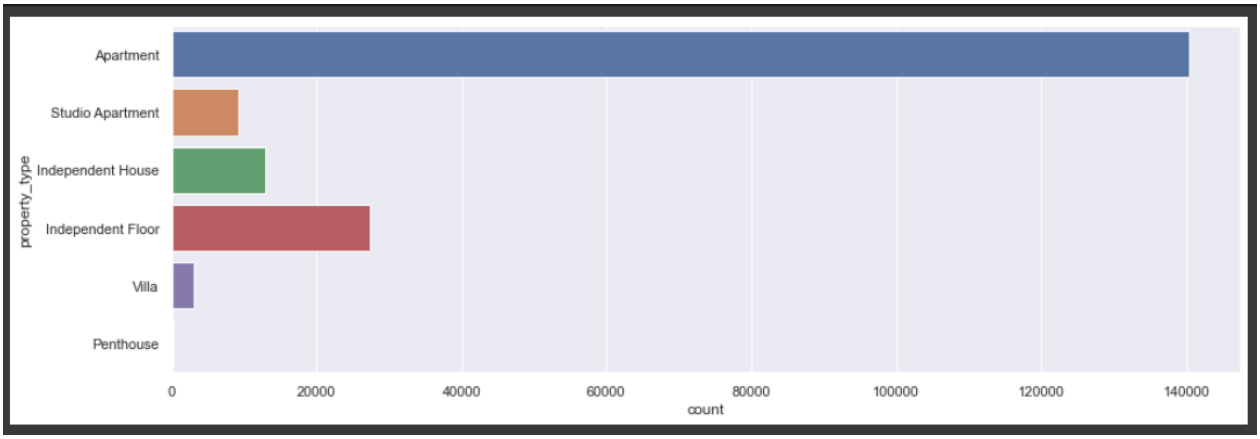


Number of Bathrooms in Houses

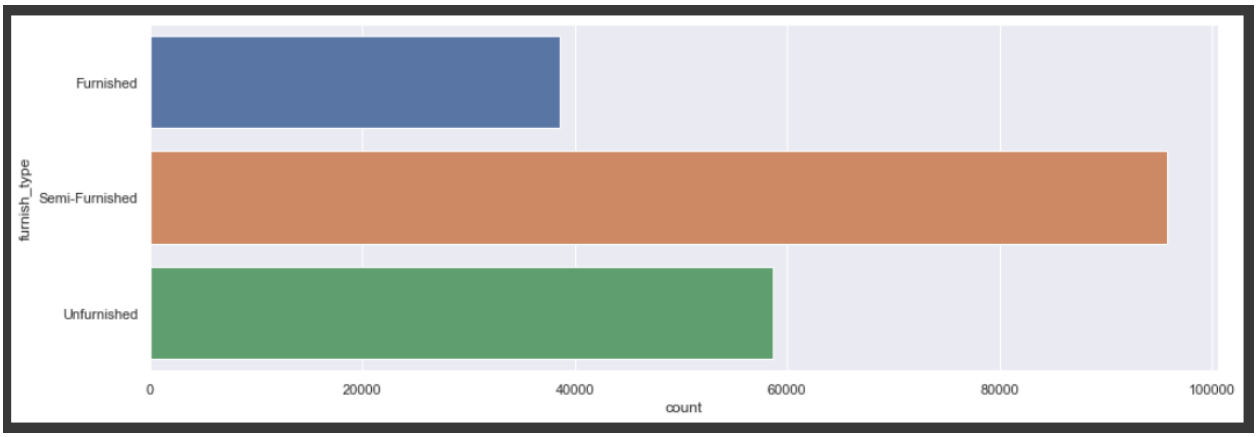


Property Type in India

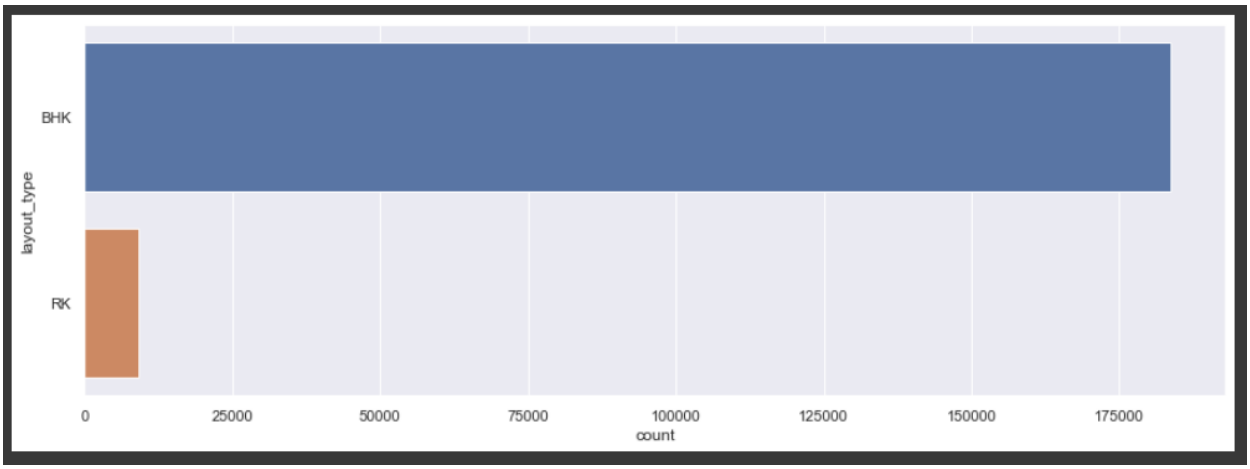
This denotes the behavioral attributes usually followed by the people. From this type of analysis, we could be able to find the most preferred type of property by the people.



Furnish Type



Layout Type Preference



Observations from the Analysis

Seller type: Agents sell the greater number of flats.

Bedrooms: Attachment with 2 attached bedrooms is preferred mostly.

Layout type: BHK is preferred by most of the sellers.

Property type: Apartment is the most selling property.

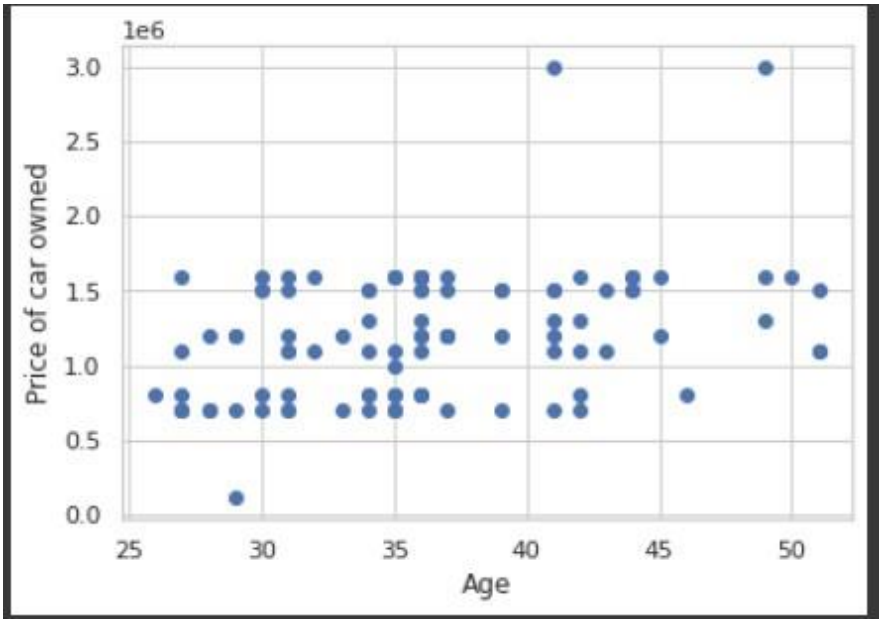
Furnished type: Semi-furnished is the most furnished property.

Bathrooms: Attachment with 2 bathrooms is preferred mostly.

Demographic Analysis

It describes the composition of a population, such as age, race, gender,income, migration patterns, and population growth.

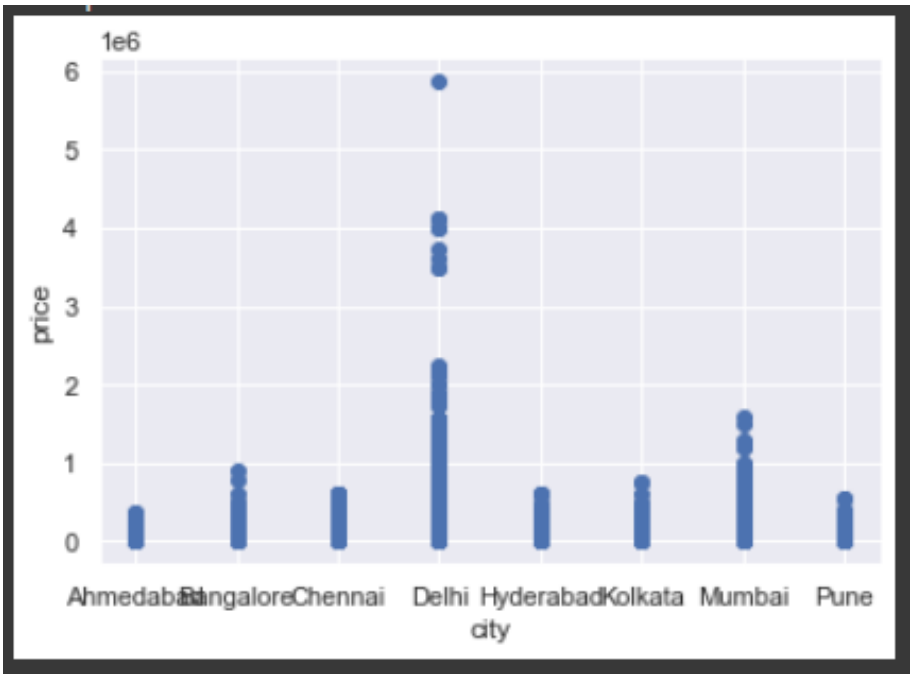
Data like age, income, spending score, family size, ethnicity, job type,Marital size and the likes help in facilitating this segment.

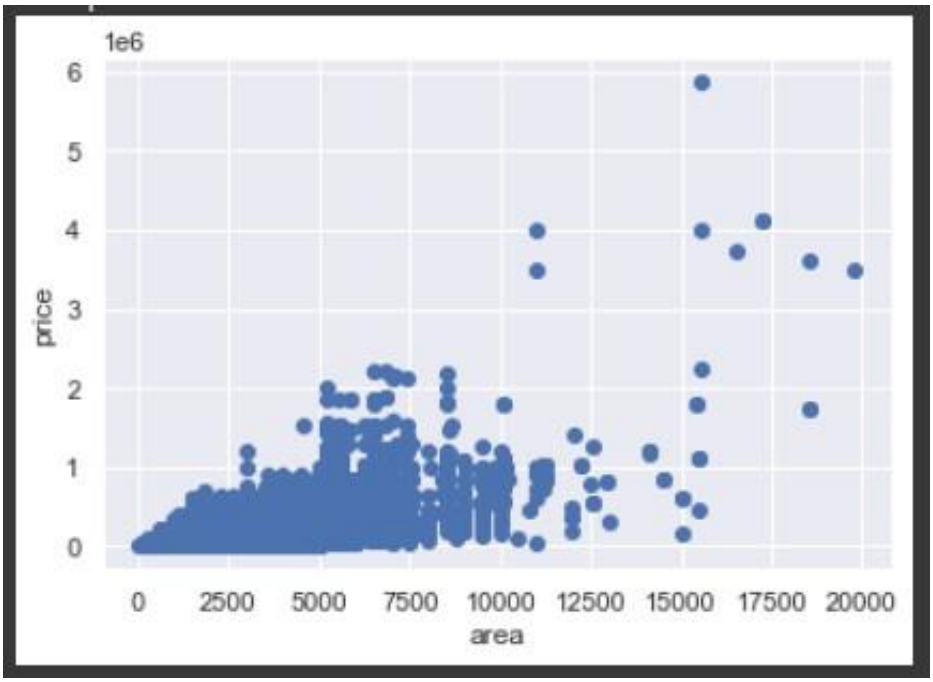


Geographic analysis

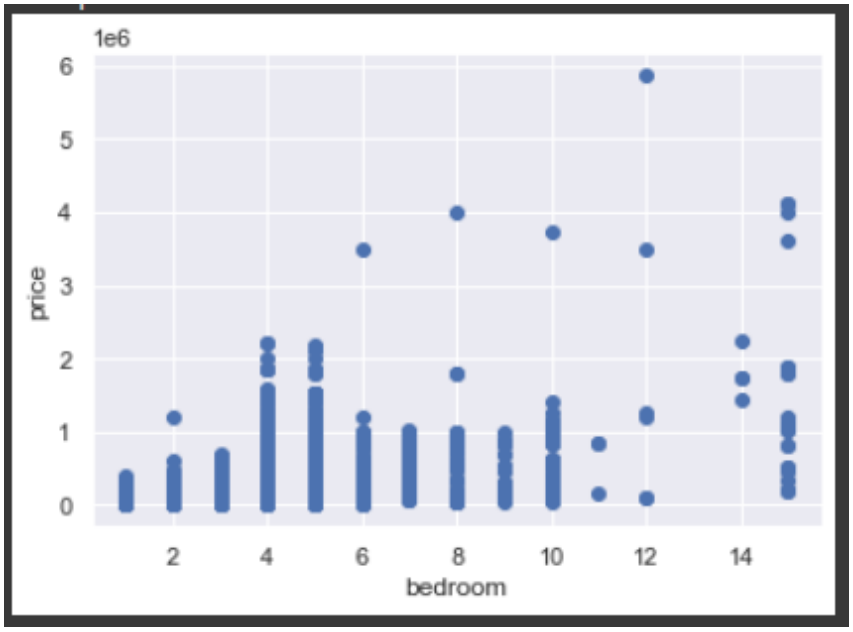
Geographic information is segmentation criterion used for the purpose ofmarket segmentation. When geographic segmentation is used – the consumer's location is considered.

The price of houses in Delhi seems to be high compared to otherplaces.

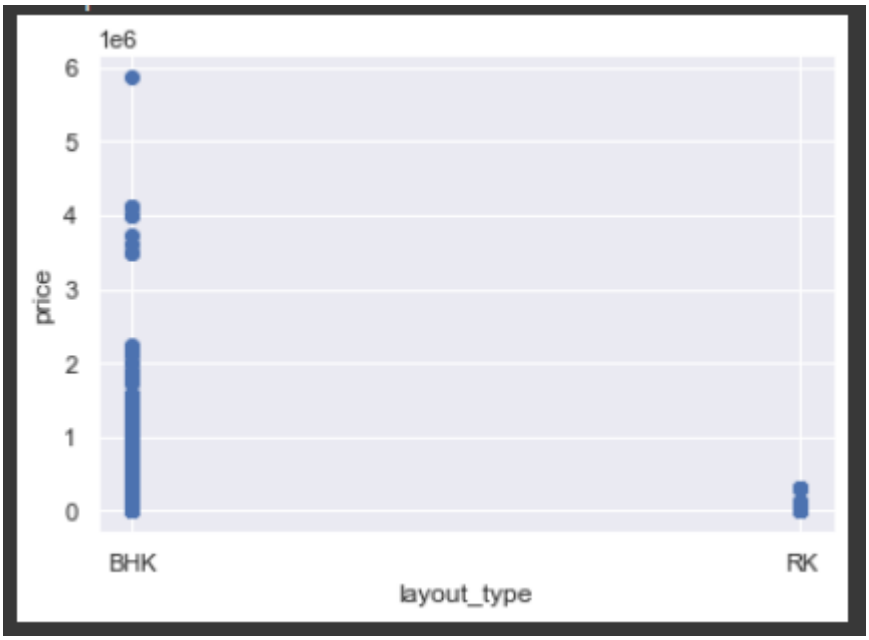




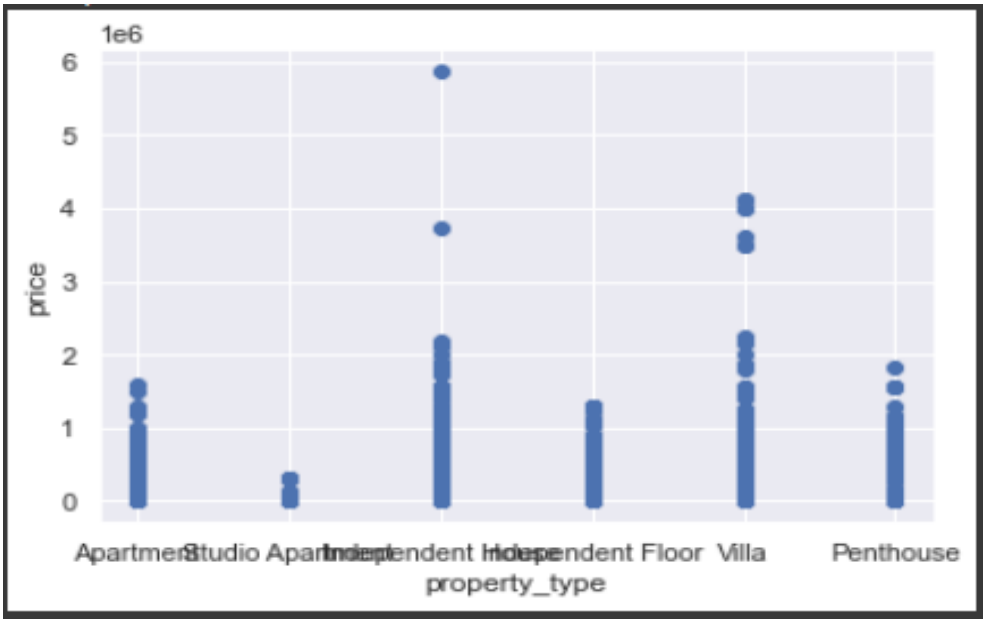
Number of bedrooms impacting price of houses



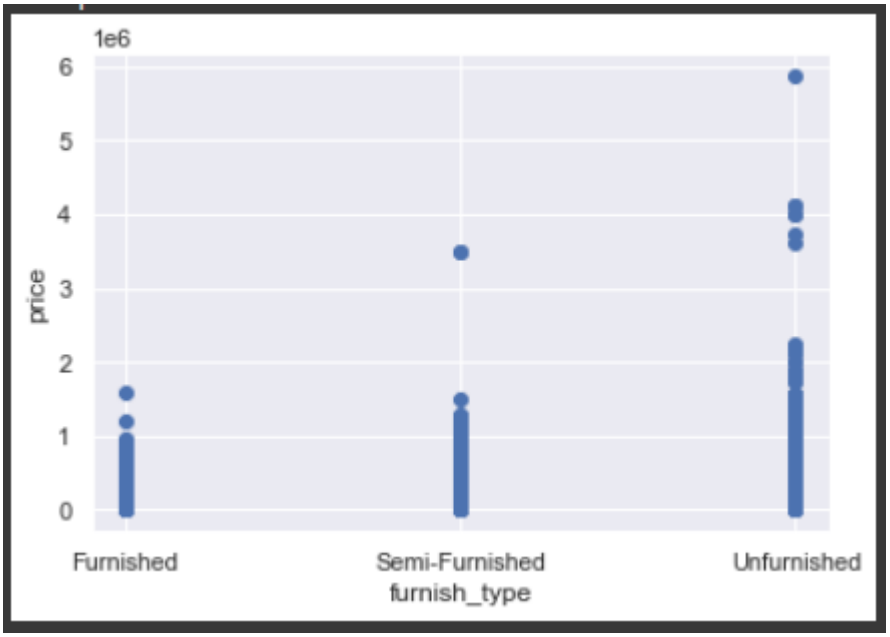
Layout type impacting Price



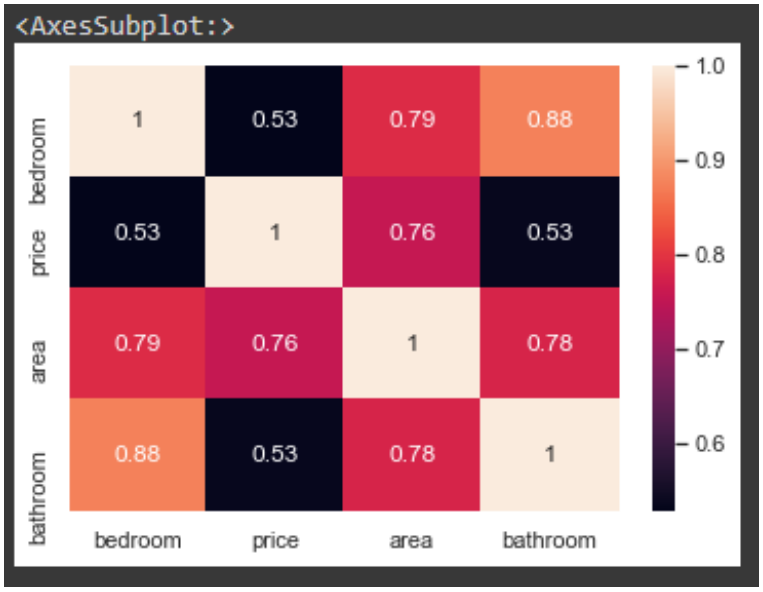
Property Type and its Price range



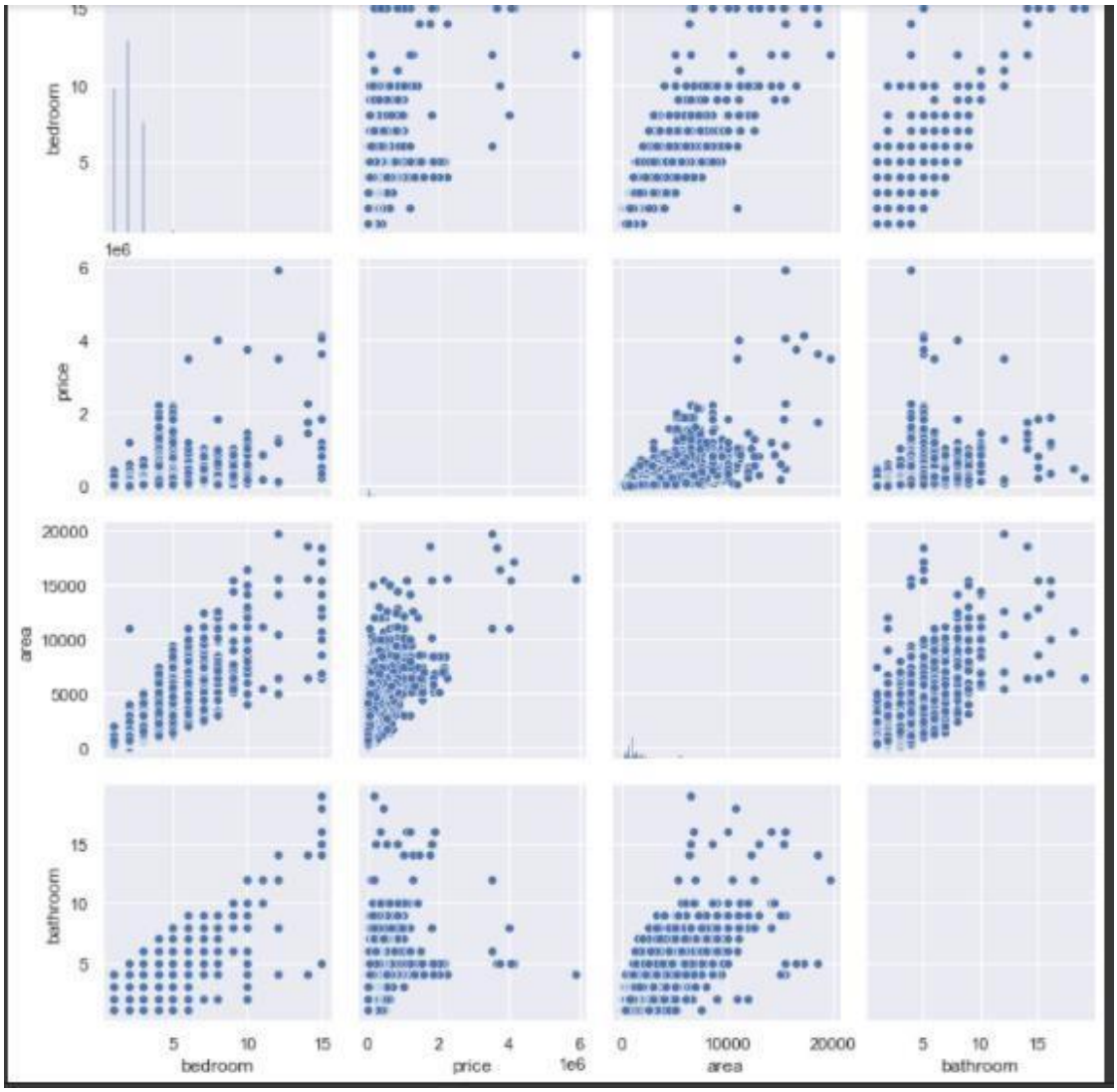
Furnishment Impacting price



Correlation Analysis



Pair Plot describing multiple correlation between variables


bedroom
price
area
bathroom
1e6
bedroom
price
area
bathroom

SEGMENTATION

We use clustering for the segmentation.

Clustering is the task of grouping observations in such a way thatmembers of the same cluster are more similar to each other and membersof different clusters are very different from each other. Clustering is the task of grouping observations in such a way that members of the same cluster are more similar to each other and members of different clusters are very different from each other.

We use K Means clustering for the segmentation.

k-means clustering algorithm is an iterative process of moving thecenters of clusters or centroids to the mean position of their constituent points, and reassigning instances to their closest clusters iteratively until there is no significant change in the number of cluster centers possible ora number of iterations reached.

Using K means clustering

```
[ ]  wcss = []

     for i in range(1, 11):
         kmeans = KMeans(n_clusters = i, init = 'k-means++',
                         max_iter = 300, n_init = 10,random_state = 0)
         kmeans.fit(X_scaled)
         wcss.append(kmeans.inertia_)
```

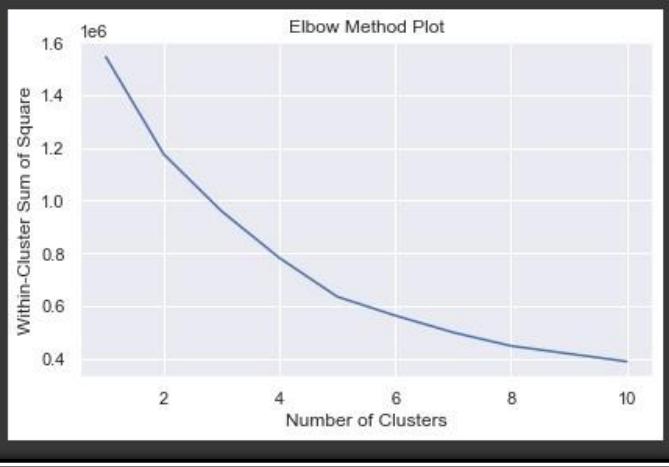
Elbow method for optimal K

Elbow method is one of the most popular ways to find the optimal numberof clusters. This method uses the concept of WCSS value. WCSS stands for Within Cluster Sum of Squares, which defines the total variations within a cluster.

It is the sum of the squares of the distances between each data point and its centroid within a cluster1 and the same for the other two terms. To measure the distance between data points and centroid, we can use any method such as Euclidean distance or Manhattan distance.

Number of cluster vs WCSS

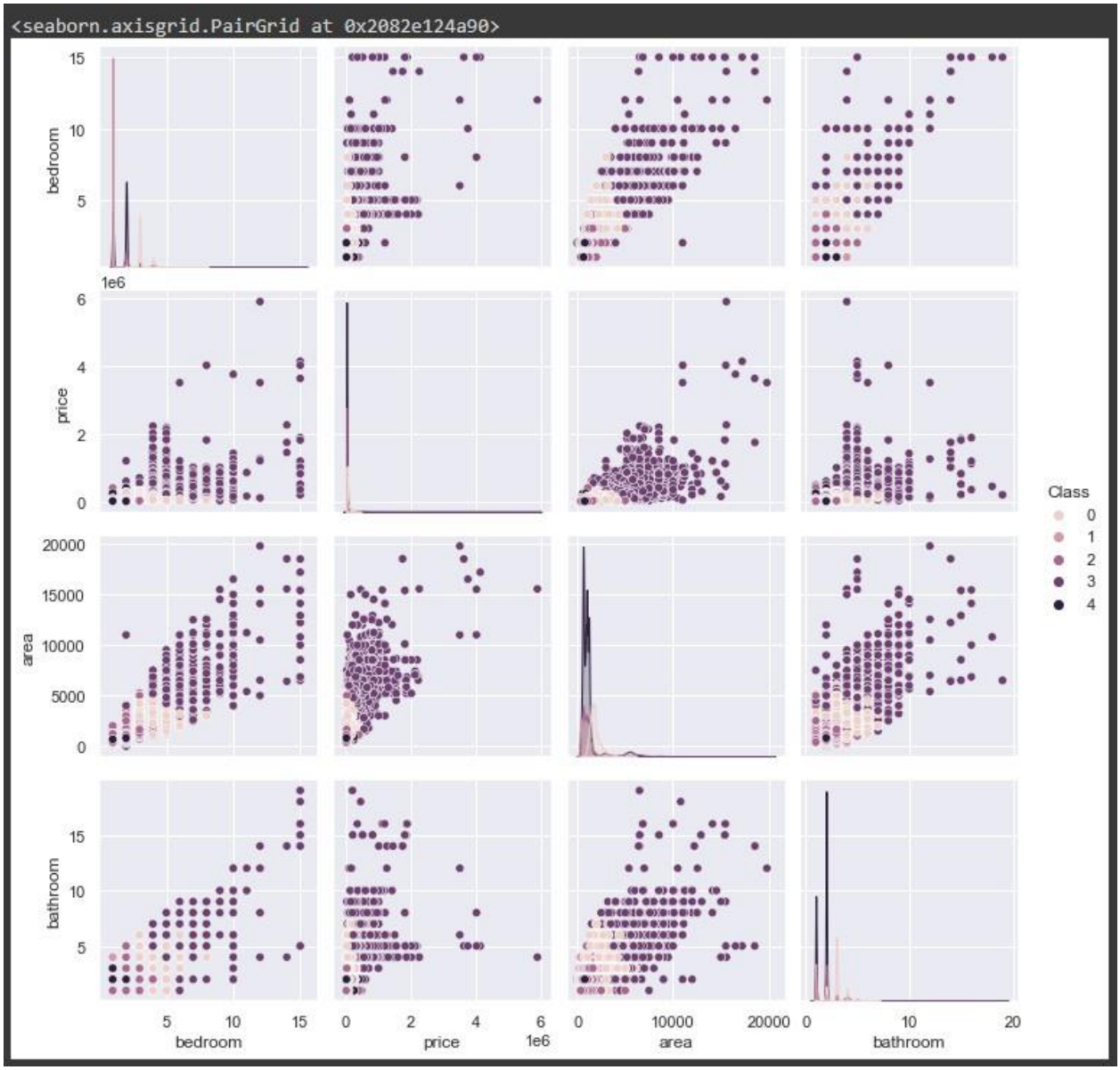
```
plt.plot(range(1, 11), wcss)
plt.title('Elbow Method Plot')
plt.xlabel('Number of Clusters')
plt.ylabel('Within-Cluster Sum of Square') # Within cluster sum of squares
plt.tight_layout()
plt.show()
```



Fit KMeans into data

```
[ ] kmeans = KMeans(n_clusters = 5, init = 'k-means++',
                    max_iter = 300, n_init = 10,random_state = 42)
    kmeans.fit(X_scaled)
```

```
KMeans(n_clusters=5, random_state=42)
```



Here we can see 5 clusters based on bedroom, price, area and bathroom.

Target Segment

We are going to target only few of the segments which will impact the market. Those segments are behavioral, demographic, geographic segments. Behavioral and psychographic segment helps in understanding the behavior and characteristics of the consumer/customer, so that we cansuggest them according to their interests.

Demographic segment helps toidentify the age group of people, income of the different levels of people.So, it will definitely help us to target the specific range of people and improve the business.

Geographic segments mainly focus on the location.Targeting customers where the demand is high will be a useful strategy toincrease the business. Now we will target the above said segments and develop a strategy to launch our Housing services App.

Market Strategy

Marketing mix is key to success in any business. The marketing mix is acombination of factors(strategies) that can be combined by a company toinfluence consumers to purchase a product. There are 4 Ps of marketing.

Product, Price, Promotions, Place

1. **Product**

A product is a commodity produced or built to satisfy the need of anindividual or a group. Here the property is the product. It is important to do extensive research before buying/selling a propertyas it has a fluctuating life cycle. So, our system will recommend andlist out the property based on the behavioral and psychographic segmentation.

2. **Price**

Price is a very important component of the marketing mix. The price of the product is basically the amount that a customer pays for it to enjoy the product. It is the most essential and crucial element of a marketing plan because it paves way for the profit and loss. Adjusting the price of the product, even a little bit has a big impact on the entire marketing strategy as well as greatly affecting the salesand demand of the product in the market. Things to keep on mind while determining the cost of the product are, the competitor's price, list price, customer location, discount, terms of sale, etc., The price of the property can be

determined by the geographic segmentation.

3. Place

Place is a very important part of marketing mix strategy. We shouldposition and distribute our product in a place that is easily accessible. Our application can be accessible throughout India. So customers from all over the places can interact with the applicationand know the pricing.

4. Promotion

Promotion is a marketing communication process that helps topublicize the product and its feature to the public. It is the most expensive and essential components of the marketing mix, that helps to grab the attention of the customers and influence them to buy theproduct. Promotions can be done using segmentation results. Specific type of properties can be suggested to the specific set of people based on their behavioral, demographic segments.

PROJECT DEPLOYMENT IN GRADIO

Interface: The user interface is created using Gradio, which allows users to enter the details of a house and get the predicted price.

Data Loading and Preprocessing

- The necessary data columns and the model are loaded from saved artifacts.
- The columns.json file contains the data columns, and the model is stored in the BREM.pickle file.
- Locations are extracted from the data columns.

Price Prediction Function

- get_estimated_price(location, sqft, bhk, bath) is the core function for predicting the price.
- It processes the inputs, handles the location indexing, and uses the loaded model to predict the price.

Gradio Interface

- Inputs for the interface include total square feet, location, number of bedrooms, and number of bathrooms.
- The interface is designed to be user-friendly, allowing users to input details and receive a predicted price instantly.

Example Explanation

Consider two cases to understand how the number of bathrooms affects the price for the same square feet:

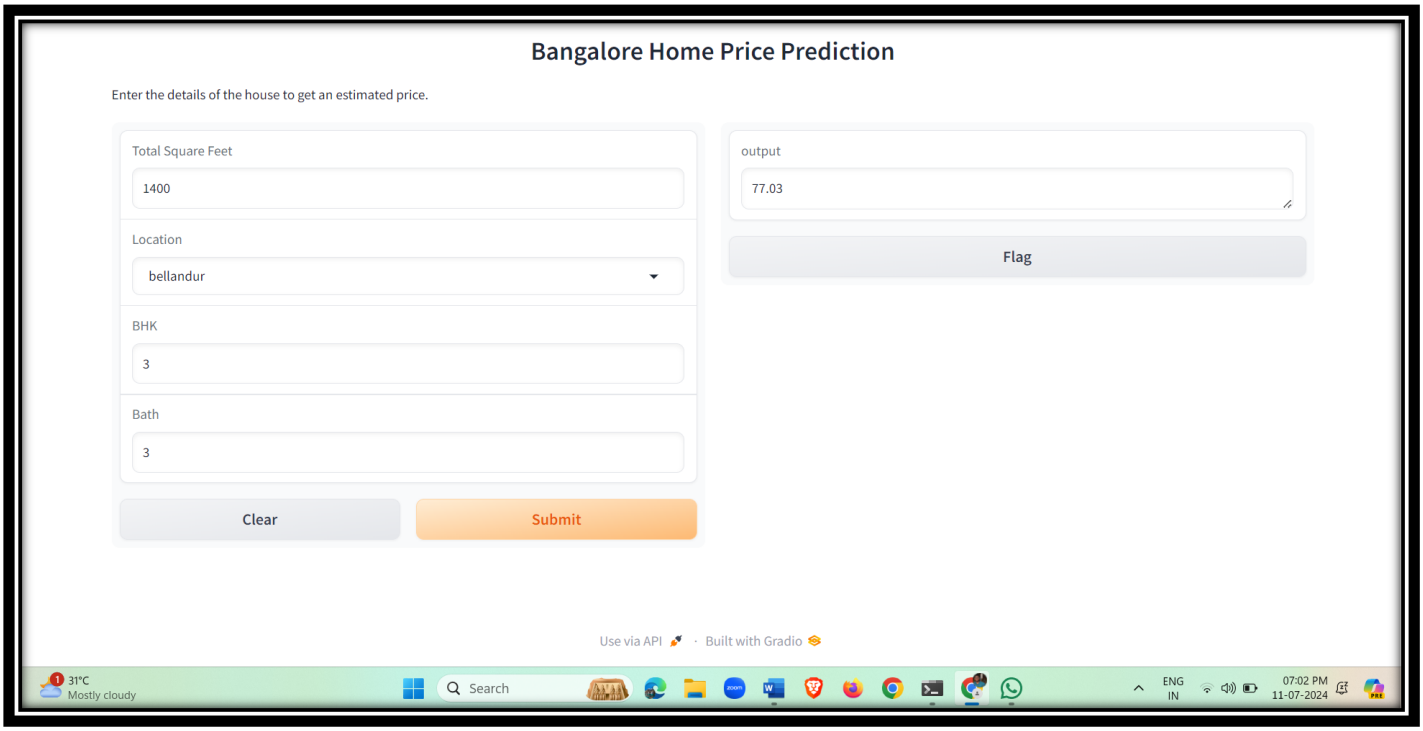
a. **Case 1**:

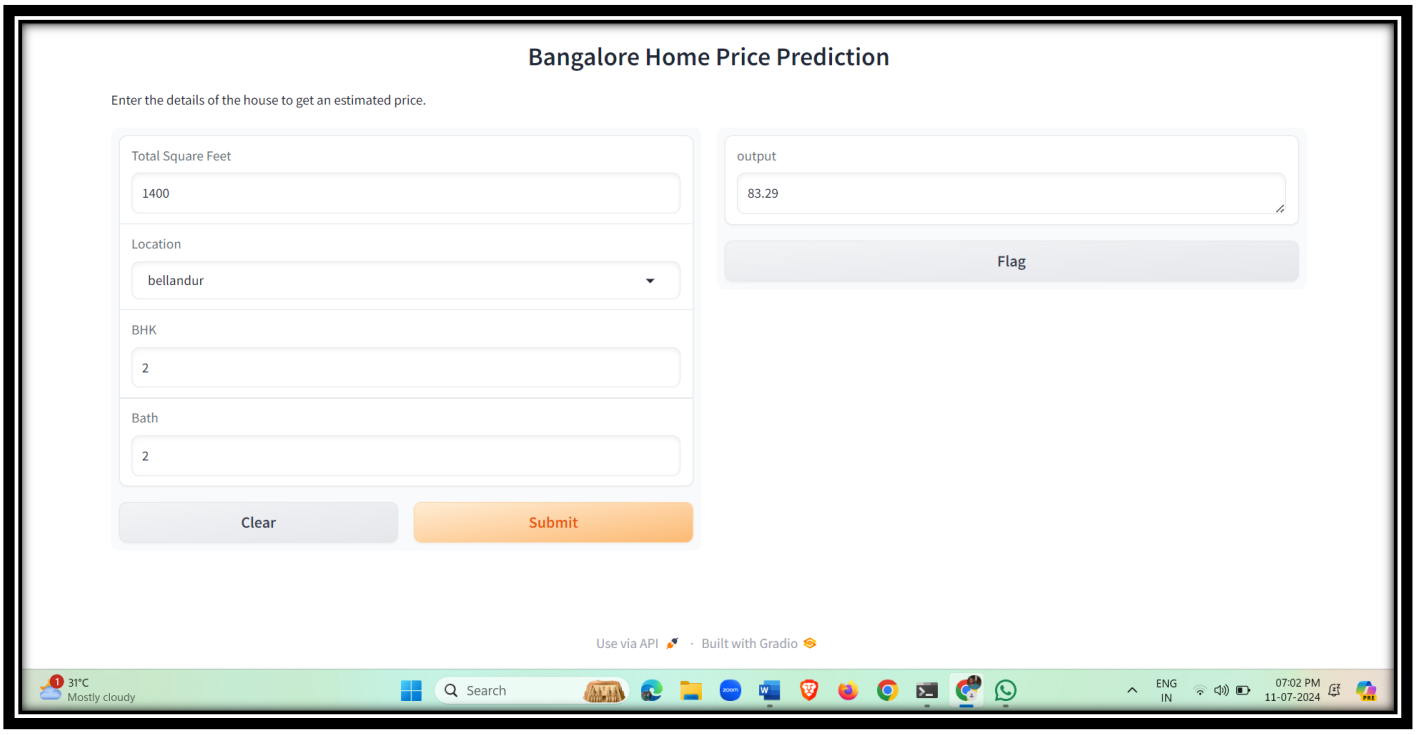
- **Details**: 2 bathrooms, 2 BHK, 1400 sq ft.
- **Predicted Price**: 83 lakhs.
- **Explanation**: More bathrooms may imply better amenities or a more luxurious setup, justifying a higher price.

b. **Case 2**:

- **Details**: 3 bathrooms, 3 BHK, 1400 sq ft.
- **Predicted Price**: 77.03 lakhs.

<u>Explanation</u>: Despite having an extra bathroom, the overall space allocated to each room might be reduced, resulting in a lower perceived value and thus a lower price.





These examples illustrate that the distribution of space within the same total area can significantly affect the price. A more balanced distribution (adequate space per room) can sometimes be valued higher than a layout with more amenities but reduced room space.

CONCLUSION

The Bangalore Home Price Prediction project provides an efficient and user-friendly way to estimate home prices based on key factors. By understanding how different features influence the price, users can make more informed decisions. The integration with Gradio makes it accessible and easy to use for a wide audience.

POTENTIAL CUSTOMER BASE

In India, there are 600 million smart phone users in 2021. So, we can obviously say that this Tech Startup for Housing App is a good initiativeand we can achieve a good profit margin from this business.

Normal mobile users are also switching to the smart phone. So, there will be further increase in the income and share for the upcoming years. Recent Reports said that there will be investment of around 35 Trillion rupees inReal estate sector.

So as our app acts as a bridge between buyer and seller,it will definitely produce a huge profit. Even if 1% of the mobile users visit our app, then the count will be around 60000. If 1% of that 60K people show some interest to buy/sell property, then 600 people will get benefited from our application.

TARGET PRICE

Advertisers pay a wholesome of money for the sales of their products. Some private advertisements can be displayed on the applications. Also google provides AdSense service. From these services we can gain someamount.

Google pays $1 for 1000 views and private advertisers can be charged $2 for 1000 views.

So totally we gain $3 for 1000 views.

POTENTIAL PROFIT

Potential Profit = Potential Customer Base * Target Price Range

= (60000/1000) * $3

= $180

$180 is the potential profit generated every day.