

MARKET SEGMENTATION ANALYSIS

Analysing the Real estate market for a Housing App startup

Sridhar Rambhatla
Somya R Patra

OVERVIEW:

Probe into the Real Estate market in India using Market Segmentation and analyse segments through which we can intrude into the market.

By 2040, it is estimated that the real estate market of India will grow by a huge margin and share a large part of the country's GDP. The inkling of the jump in numbers can be verified by the increase in the number of huge and luxurious investments into this sector. There is a surge in the number of people investing and due to SOP's introduced in the sector, there has been a huge jump in the number of housing sales and its other relevant sectors. This was also supplemented by the availability of completions of projects in many of the emerging cities like Bengaluru, Pune.

Historically, India's real estate sector has been unorganised, with factors such as the lack of a centralised title registry that provides title guarantee, lack of uniformity in local laws and their application, nonavailability of bank financing, high-interest rates, and transfer taxes, as well as a lack of transparency in transaction values preventing structured dealing.

Even though the rental market in many major IT hubs has fallen due to Covid-19, due to the major boom of the population in big cities the rental market is bound to surge, and with the new housing rules and Acts which have safeguarded the whole sector against vulnerabilities will pull in many investors into the market.

In this report we will segment the customer and real estate market to find insights and create a strategy to deploy an online real estate listing application.

As per the innovation Adoption life cycle, we have chosen Pune as the entry point for the case study. Also, because it is one of the major rising IT hubs in the country and is growing tremendously in the real estate industry.

Real Estate Market Overview:

India is set to double its urban population between 2018 and 2030, with estimates projecting that by 2028, New Delhi would become the most populous city on the planet. This large-scale migration to urban centers in India is bound to create pressure on housing markets.

Pune is among the few cities in India that have accelerated steadily and have made it among the 100 top cities in the country for its Real Estate sector. Over the past five years, Pune's real estate sector has evolved a lot and has been influenced by numerous factors. Among them, demonetization and GST implementation are the two major aspects that created a slight stir within the real estate sector across India. However, since Pune's real estate has always been known to be steady, it balanced it out and stood as a clear winner even under this critical situation. For such reasons, it is deemed that the next 10 years from now will be wonderful and lucrative for the real estate market of Pune.

Despite the ongoing crises and slowdown in the real estate and infrastructure sector, the residential property prices in Pune registered a marginal QoQ growth of 0.7% during the quarter ended June 2021. Further, with the lockdown being lifted partially since mid-June, the construction and allied activities have started to normalize. Subsequently, the supply and demand which was affected by the lockdown restrictions April-May is expected to grow. Additionally, citing affordability and proximity to major industrial and commercial hubs, homebuyers in Pune were seen shifting towards the peripheral localities, stagnating property prices in the city centre in the last 5 years.

Pune city has clocked about 53,000 sales of housing units in the period January-July 2021 as compared to about 49,000 units sold in January-July 2019 indicating the growth of almost 8%, in terms of value, Pune clocked Housing Sales of approx. Rs. 27,500 Cr.

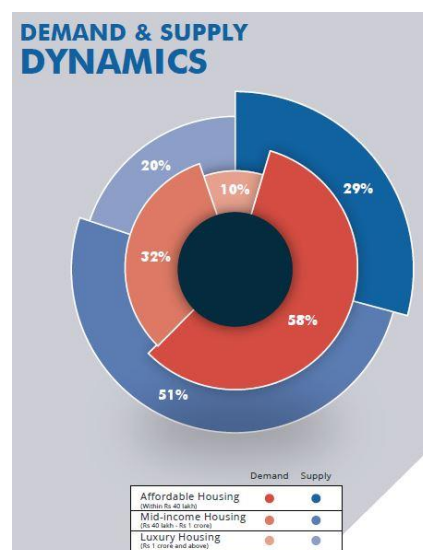
In 2021, the average rate of property in Pune currently stands at Rs 5,083 per sq. ft, after recording a 3% annual appreciation in the July

quarter of calendar year 2021. The average rental rate in the city is Rs 19,880 per month, as per listings on Housing.com.

Factors that make it an ideal time to invest in Pune realty		
Attractive property rates and home loan interest rates.	Gradual appreciation in property prices.	Developers focusing on quality and completion of existing projects.



This here demonstrates the rates at which the real estate sector has been burgeoning and how it will keep on increasing its pace in the coming years. It is divided into four parts according to India's regional area, and almost all quarters of it have prediction of equal growth in the upcoming years



FERMI ESTIMATION:

" It is often necessary to make quick estimates when neither time nor resources are available for making traditional assessments. This is particularly true at the ideation stage of product development when even a gross estimate could be useful for heading off ill-advised expenditures. "- Philip M. Anderson, Cherie Ann Sherman

A Fermi estimate is one done using back-of-the-envelope calculations and rough generalizations to estimate values which would require extensive analysis or experimentation to determine exactly. It is a helpful starting point for gaining insight into the order of magnitude estimation.

Problem Breakdown: With more than 70 per cent of home buyers now present on the internet and over 80 per cent of available properties getting posted online in recent times, the demand for online housing applications is rising. Housing applications can perform a variety of functions and serve many purposes like: Home valuation, Recommending real estate to the User etc.

The estimation of the housing/ real estate market infographics starts with understanding the factors like locality, and customer behaviour responsible for changes in real estate prices. In terms of a housing app, making profit and providing renting/housing services is the primary goal. In order to find/estimate the correct customer base we asked ourselves a few questions like: What kind of tenant are the renters or buyers looking for? What kind of locality do tenants of pune prefer? What kind of property is more popular based on the segments?

Breaking down the housing application market these questions can help us estimate the target customer and gauge the performance of the application in the market Fermi estimation can be used to calculate the potential customer base.

DATA SOURCES:

Pune House Listing Data -

<https://www.kaggle.com/saipavansaketh/pune-house-data?select=Pune+house+data.csv>

Rental Price Listing, Pune

<https://www.kaggle.com/anantsakhare/rental-price-of-indias-it-capital-pune-mh-ind>

Demographic data

<https://www.kaggle.com/frtgmn/dunnhumby-the-complete-journey>

DATA PREPROCESSING:

Handling Atypical Data and Removing Null Values:

Converting range of prices to float using the mean.

	area_type	availability	size	society	total_sqft	bath	balcony	price	site_location
30	Super built-up Area	19-Dec	4 BHK	LedorSa	2100 - 2850	4.0	0.0	186.000	Gultekdi
122	Super built-up Area	18-Mar	4 BHK	SNontle	3067 - 8156	4.0	0.0	477.000	Ganeshkhind
137	Super built-up Area	19-Mar	2 BHK	Vaarech	1042 - 1105	2.0	0.0	54.005	Khadaki
165	Super built-up Area	18-Dec	2 BHK	Kinurg	1145 - 1340	2.0	0.0	43.490	Prabhat Road
188	Super built-up Area	Ready To Move	2 BHK	MCvarar	1015 - 1540	2.0	0.0	56.800	Wakadewadi

```
def convert_sqft_to_num(x):
    tokens = x.split('-')
    if len(tokens) == 2:
        return (float(tokens[0])+float(tokens[1]))/2
    try:
        return float(x)
    except:
        return None

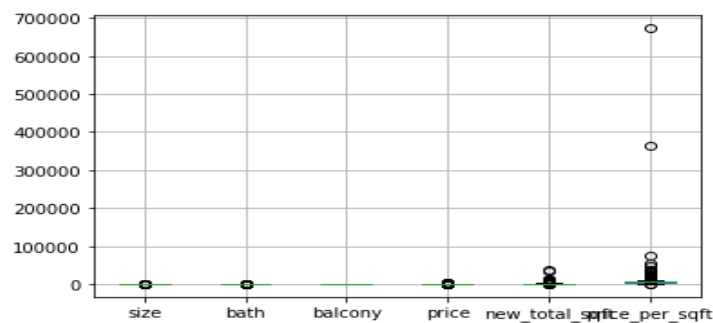
PHD['new_total_sqft'] = PHD.total_sqft.apply(convert_sqft_to_num)
PHD = PHD.drop('total_sqft', axis='columns')
PHD.head()
```

	area_type	availability	size	society	bath	balcony	price	site_location	new_total_sqft
0	Super built-up Area	19-Dec	2 BHK	Coomee	2.0	1.0	39.07	Alandi Road	1056.0
1	Plot Area	Ready To Move	4 Bedroom	Theanmp	5.0	3.0	120.00	Ambegaon Budruk	2600.0
3	Super built-up Area	Ready To Move	3 BHK	Soiewre	3.0	1.0	95.00	Aundh	1521.0
5	Super built-up Area	Ready To Move	2 BHK	DuenaTa	2.0	1.0	38.00	Balaji Nagar	1170.0
11	Plot Area	Ready To Move	4 Bedroom	Prtry M	5.0	3.0	295.00	Bopodi	2785.0

Removing Outliers:

A box plot tells us, more or less, about the distribution of the data. It gives a sense of how much the data is actually spread about, what's its range, and about its skewness. Q1 is the first quartile of the data, i.e., to say 25% of the data lies between *minimum* and Q1. Q3 is the third quartile of the data, i.e., to say 75% of the data lies between *minimum* and Q3. To detect the outliers using this method, we define a new range, let's call it the decision range, and any data point lying outside this range is considered an outlier and is accordingly dealt with.

<matplotlib.axes._subplots.AxesSubplot at 0x229b3e45988>



```
# Using
q1 = df.quantile(0.25)
q3 = df.quantile(0.75)
IQR = q3-q1
IQR

df = df[~((df) > (q3 + (1.5 * IQR))).any(axis=1)]
df.head()
```

Normalization:

Tools: Sklearn - StandardScaler()

Standardisation generally means changing the values so that the distribution's standard deviation equals one. Scaling is often implied.

```
ss=StandardScaler()
data_num_scaled=pd.DataFrame(ss.fit_transform(data_num),columns=data_num.columns)
data_num_scaled=data_num_scaled.reset_index()
data_num_scaled.head()
```

	index	size	bath	balcony	price	new_total_sqft	price_per_sqft
0	0	-0.645688	-0.432273	-0.792653	-0.980046	-0.726911	-1.059776
1	1	0.983871	1.186616	-0.792653	0.720798	0.578057	0.644689
2	2	-0.645688	-0.432273	-0.792653	-1.012585	-0.406983	-1.362330
3	3	-0.645688	-0.432273	-0.792653	-1.012585	-0.884068	-0.992704
4	4	0.983871	1.186616	0.531427	2.332541	2.623910	0.866870

Encoding:

Tool: Sklearn - Pandas - pd.get_dummies

```
df_encoded=pd.get_dummies(df,columns=['area_type','site_location'])
df_encoded.head()
```

	availability	size	bath	balcony	price	new_total_sqft	price_per_sqft	area_type_Built-up Area	area_type_Carpet Area	area_type_Plot Area	...	site_location_Vadgaon Budruk	site_lo
0	0	2.0	2.0	1.0	39.07	1056.0	3699.810606	0	0	0	...	0	
3	1	3.0	3.0	1.0	95.00	1521.0	6245.890861	0	0	0	...	0	
5	1	2.0	2.0	1.0	38.00	1170.0	3247.863248	0	0	0	...	0	
12	1	2.0	2.0	1.0	38.00	1000.0	3800.000000	0	0	0	...	0	
14	1	3.0	3.0	2.0	148.00	2250.0	6577.777778	0	0	1	...	0	

SEGMENT EXTRACTION:

The approach used for Segmentation:

Clustering:

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.

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.

K Means:

The k-means clustering algorithm is an iterative process of moving the centers 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 or a number of iterations reached.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Elbow Method: The Elbow method is one of the most popular ways to find the optimal number of 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. The formula to calculate the value of WCSS (for 3 clusters) is given below:

$$WCSS = \sum_{P_i \text{ in Cluster1}} \text{distance}(P_i, C1)^2 + \sum_{P_i \text{ in Cluster2}} \text{distance}(P_i, C2)^2 + \sum_{P_i \text{ in Cluster3}} \text{distance}(P_i, C3)^2$$

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. To find the optimal value of clusters, the elbow method follows the below steps:

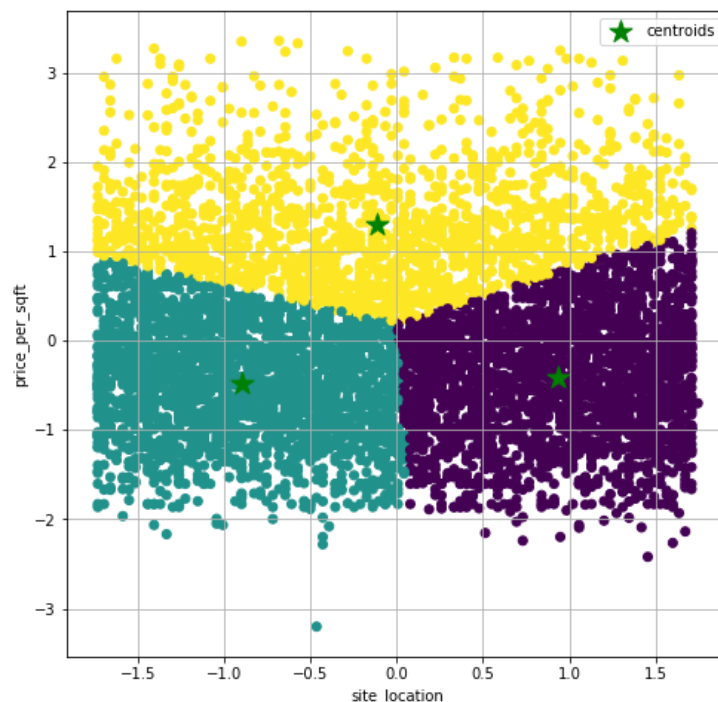
- It executes the K-means clustering on a given dataset for different K values (ranges from 1-10).
- For each value of K, calculate the WCSS value.

- Plots a curve between calculated WCSS values and the number of clusters K.
- The sharp point of bend or a point of the plot looks like an arm, then that point is considered as the best value of K.

Note - Online housing applications maintain their own consumer data. Hence the data is not readily found on the internet as it is a privacy breach. Since we do not have the relevant consumer data, generalization and assumptions shall be made.

Geographic Segmentation:

Geographic segmentation is an effective methodology used by organizations with large national or international markets to better understand the location-based attributes that comprise a specific target market. Consumers that live in different geographic regions typically display varying needs, wants, and cultural characteristics that can be specifically targeted for more efficient and better marketing



The above cluster shows how sellers in a certain locality sell their homes at a higher price than other localities, this shows that the geographic

attributes have high significance in determining the customer base and house prices.

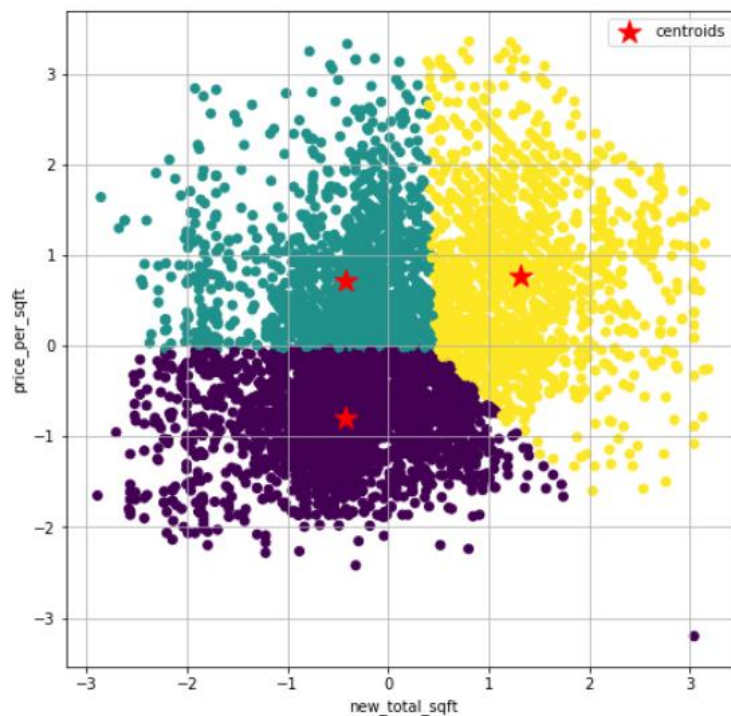


Fig-2

This cluster shows the localities/data points with size and price. Yellow cluster shows the localities with bigger sized and expensive houses. Purple shows small houses which are moderately priced and blue shows small houses with high prices. These fluctuations happen due to the psychological differences of people in different locality

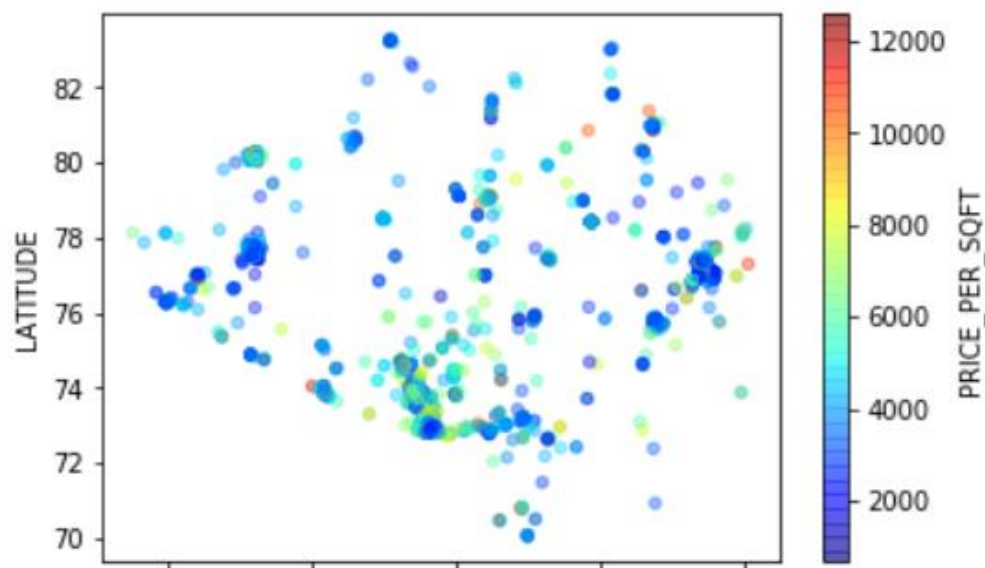


Fig-3

Fig3 shows the average prices of properties in different localities. These clusters are made from a real estate listing dataset

Demographic Segmentation:

Demographics are the data that describes the composition of a population, such as age, race, gender, income, migration patterns, and population growth. It helps in blending a particular potential customer into a well-defined cluster, where every representative in there has at least a similar base and this not only helps by increasing the revenue to bring those potential customers in but also gentrifying the agglomerate and its vicinity.

Data like age, income, spending score, family size, ethnicity, job type, Marital size and the likes help in facilitating this segment. It also helps in painting a definitive character of the customer and making ends meet for them.

The fig below also helps in substantiating the viewpoint of how important the demographic segmentation is to this sector and it is also evident that Employment, Marital status and Gender play the biggest role in it.

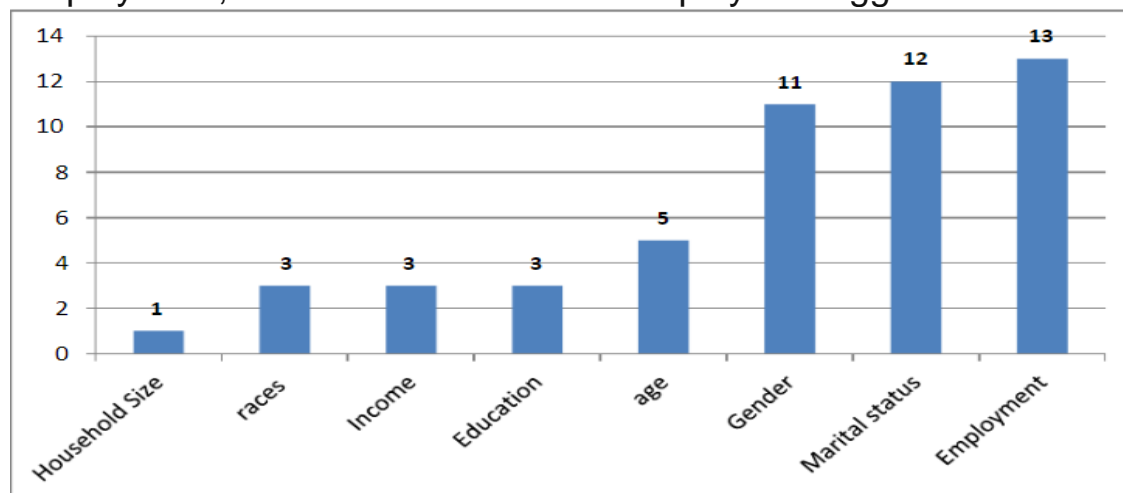


Fig - The ranking of frequency for the total number of significances between demographic criteria and property criteria

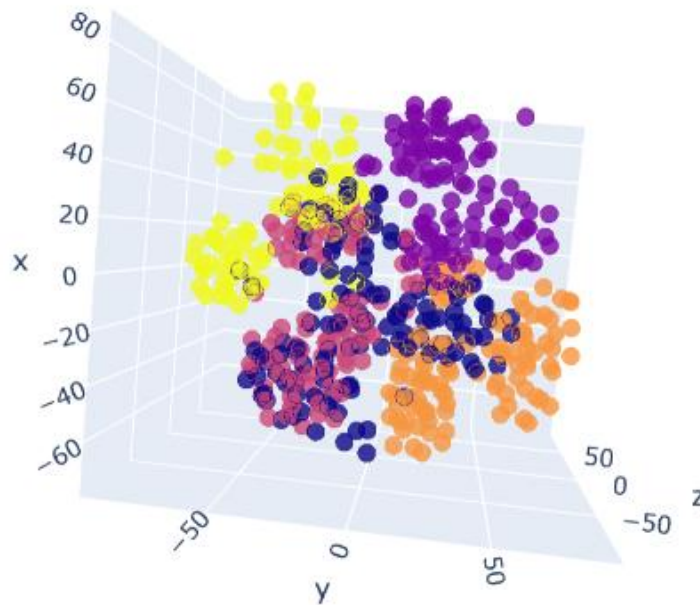


Fig a

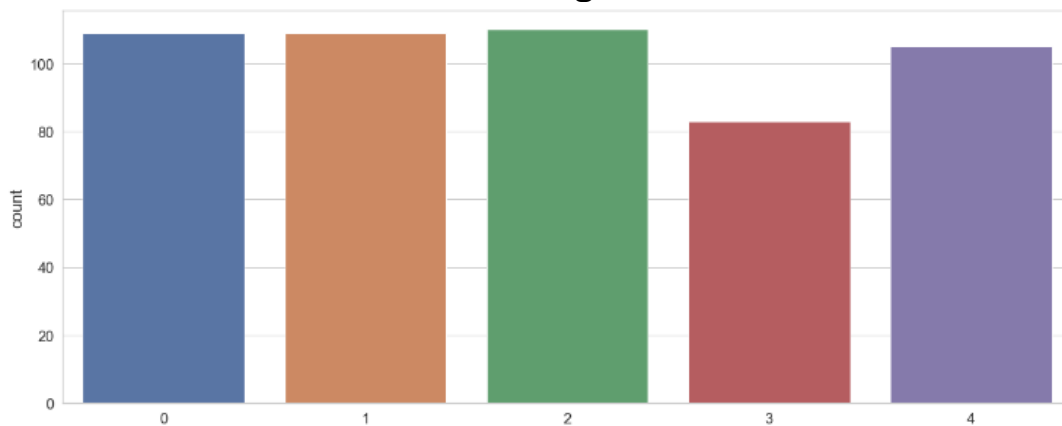


Fig b

In the above cluster at fig a is where x represents the age, y the salary and z whether the person has ownership of the house or is a renter. This cluster gives us a clear description of which age group has what kind of financial status and whether they own a home or rent it. Based on this the potential customers can be selected. Fig b shows us about the count of data in each cluster

Psychographic Segmentation:

In real estate, psychographics refers to the study of consumers based on their psychological and cognitive attributes, such as their beliefs, values, thoughts, hopes, and goals. Psychographics go beyond demographics (e.g., age, gender, race, and location) to understand consumers' purchasing behaviours. Psychographic deals with the question of why a customer would want to buy something and this always lies within their beliefs, values, goals, and attitudes. In this sector, psychographics entirely depends upon the customer who plays a very big role in this segmentation, and depending upon their perspective it either makes the deal or offsets the whole process. It all depends on the buyers/ tenant's lifestyle, attitudes, and interests. It matters the most in the luxurious section of the market where top-notch properties and the amenities available with them matter the most to the customers.

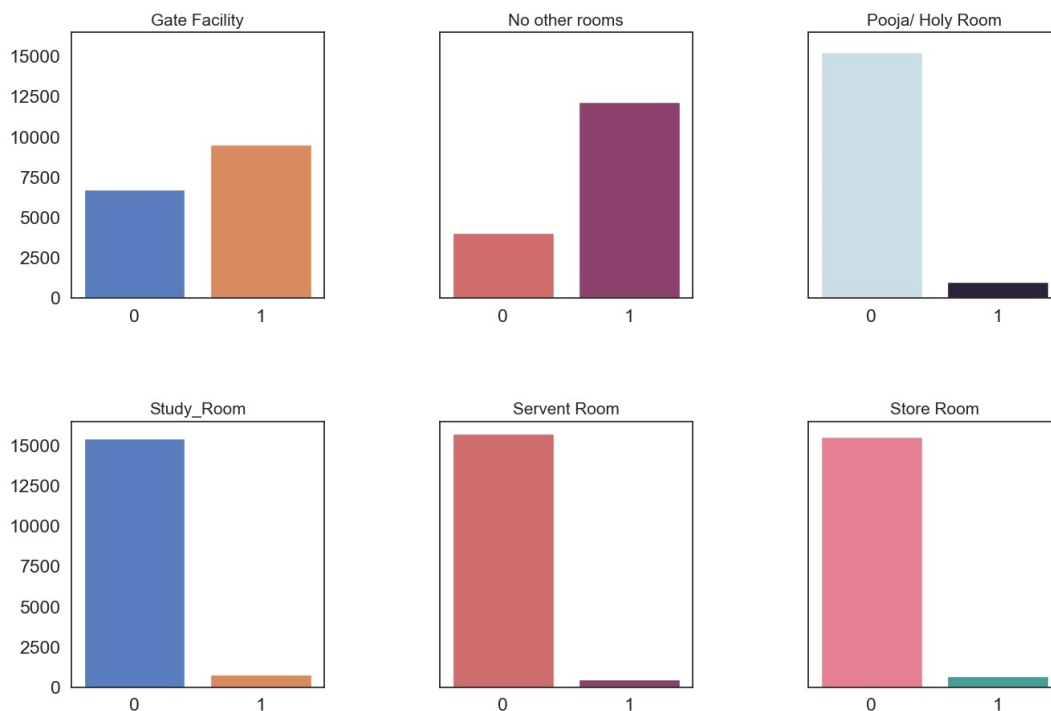


Fig 1. Facilities available in the houses

Fig 1 describes all the details that a house or apartment in Pune may or may not contain. This brings us across many of the aesthetics and interests that people might have while looking for a house like the puja/holy room or the availability of study room or servant room. While Fig 2 enlightens about the disposition of the house like the direction in

which the apartment is facing, whether the house is furnished or not, the floor on which the house is situated, the age of the property and so on.

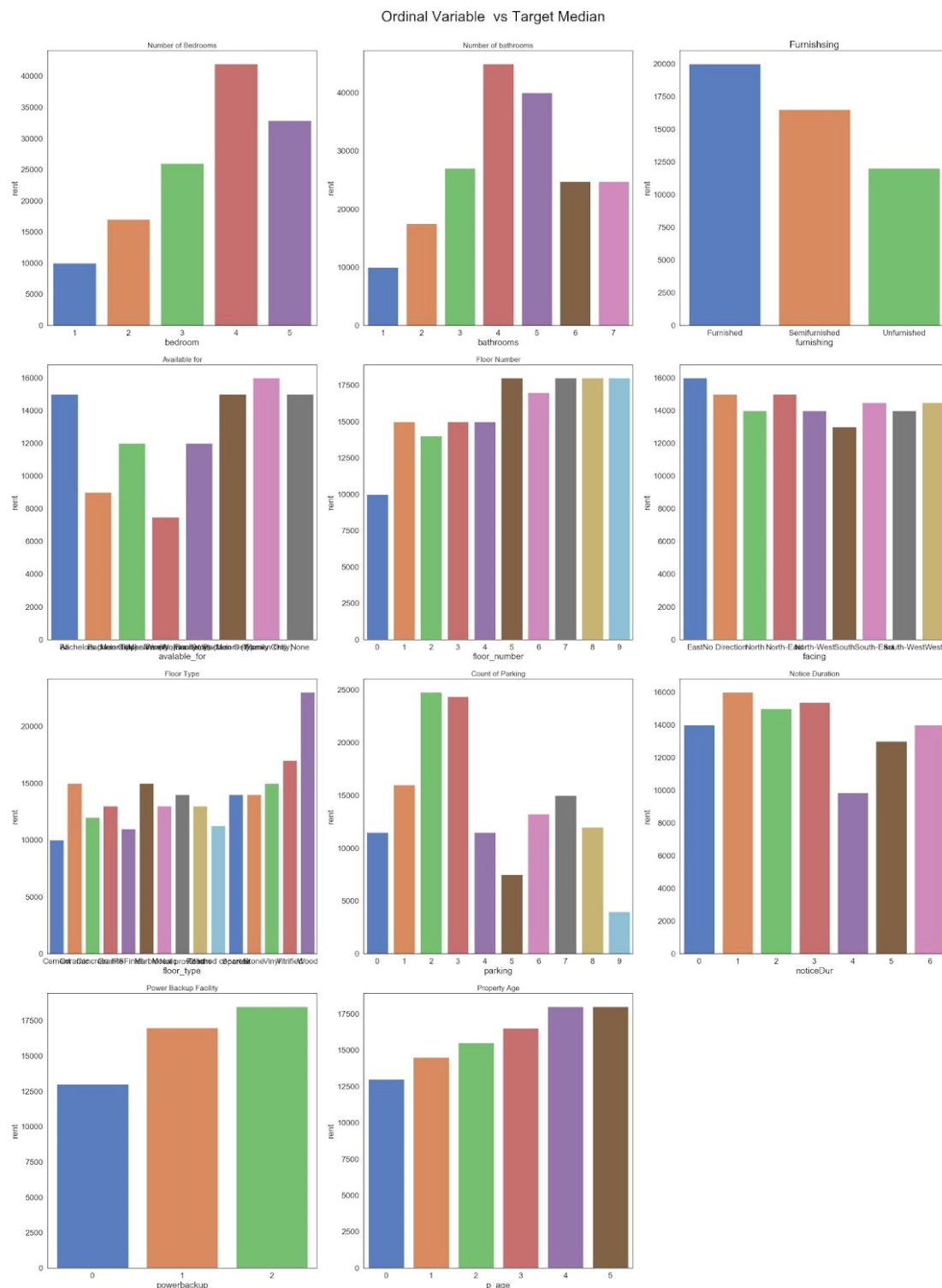


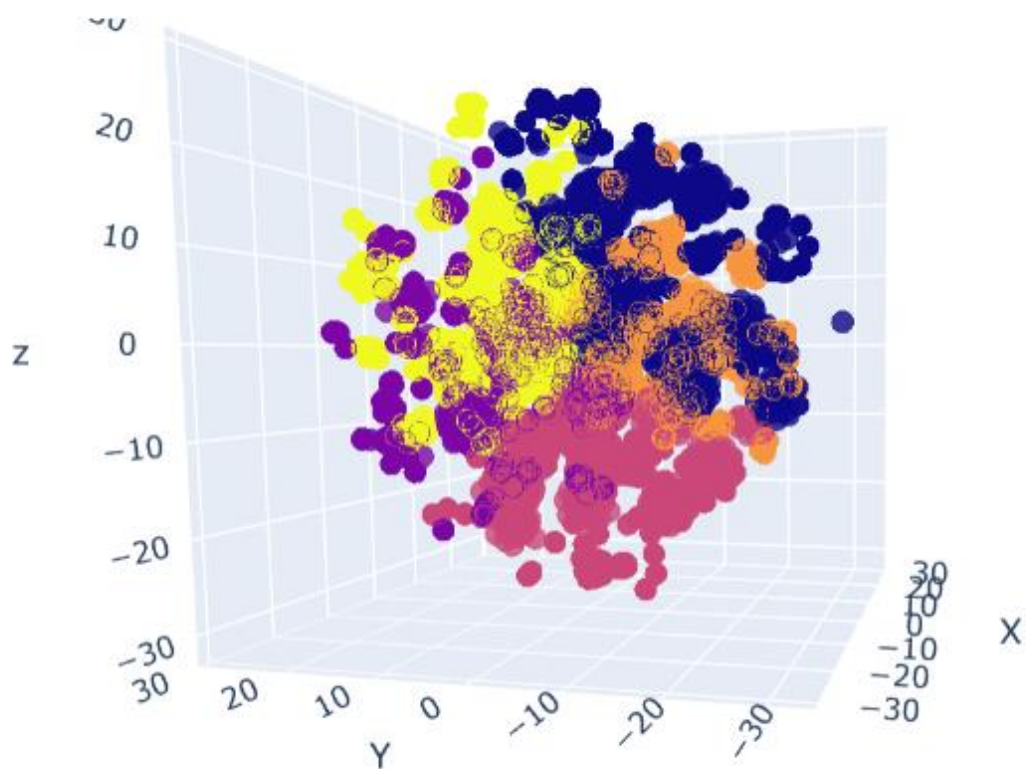
Fig-2. Disposition of the Houses on the market

Behavioural Segmentation:

Understanding factors that affect the real estate market are found to be very important and crucial due to the inability of the Utility Theories which have been used to study consumer behaviour to provide convincing explanations to understand the market. From Behavioural perspectives, insights on consumer purchase behaviour and explanations were given on what drives consumers to purchase housing. Including consumer behaviour concepts from sociology, psychology and marketing can therefore be beneficial to the study of real estate.

	N	Minimum	Maximum	Mean	Std. Deviation
Price range	220	2	5	4.25	0.768
Availability of loan	220	2	5	4.2	0.75
Fire resistant	220	2	5	3.8	0.87
Maintenance cost	220	2	5	3.75	0.77
Well built	220	2	5	3.7	0.901
Location	220	2	5	3.68	0.843
Ventilation	220	2	5	3.65	0.734
Developer's reputation	220	2	5	3.65	0.727
Proximity to amenities	220	2	5	3.63	0.638
Accessible garage	220	1	5	3.48	0.743
Locality	220	2	5	3.44	1.043
View from house	220	2	5	3.24	0.783
Power supply	220	1	5	3.16	1.012
Water supply	220	1	5	3.12	1.042
Appearance	220	2	5	3.02	0.941
Vastu compliance	220	2	5	2.95	0.875
House no.	220	1	5	2.95	1.247
Interior	220	2	5	2.9	0.835

This data was from a survey collected to analyse their behaviour by giving a numerical value to what they think is important and whether they will buy or rent the house based on these features. Referencing these we scoured the data and made clusters which somewhat shows a similar kind of inference. The financial factors – price range and availability of home loan – were found to be most important by home buyers. The financial or economic factors were followed by construction quality, location, developer's reputation, ventilation and proximity to amenities on the basis of importance for the home buyers.



In this cluster x represents furnishing of the house, y represents bedroom and z the rent of the house. Even though it's heavily populated, it had to be divided into 5 clusters, each cluster having its own features corresponding with the other values in the other two columns. It holds valuable information like the unfurnished houses with more no. of rooms still have a higher price than the rest, while houses that are semi-furnished or furnished but with fewer bedrooms have a decent price range but that heavily depends upon the locality.

TARGET SEGMENT:

The segments that we are going to invest and target are the Geographic segment, Demographic segment and Behavioural segment. These segments help us understand the market insights of a city accurately and can be used to target a particular customer segment. We will use these segments to develop a feasible strategy to launch a housing application into the market.

MARKET STRATEGY:

A marketing mix includes multiple areas of focus as part of a comprehensive marketing plan. The term often refers to a common classification that began as the four Ps: product, price, placement, and promotion.

Marketing mix- 4P's



Product - Generating a user profile on the application would give us the necessary insight on the behaviour based on the searches and interactions. Using that we can customize and streamline the target audience as the market grows. Property Recommendations based on the user interests can be made for faster sale of the property and providing the user their dream property. Our product would offer all kinds of real estate listings.

Price - It would be a free to use website and an application but the primary source of income would be advertising and promotion of properties

Place- Initially the application would be deployed in Pune but as the market and customers grow it can be used all over India

Promotion - Promotion can be done by collecting the users interest and in which demographic and psychographic clusters they would lie in. This would add more value to the company

Potential customer Base – Since 70-80 of home buyers now present on the internet, we can assume that a good start-up can get 10-15% of the market share. 60000 new houses are added daily on the internet for sales according to TOI. So, assuming we get a market share of 6000-8000 houses every day, and for each house at least 5 people visit the application. That gives us 30000 views on the application every day.

Target price – Promotions and advertisers pay huge amount of money for sales. Google AdSense pays \$10 per 1000 views. Private advertisers pay higher price but for estimation we consider Google AdSense.

Potential Profit – $(30000)/1000 * \$10 = \300

\$300 profit is generated everyday.

REFERENCES:

1] The Rental Housing Market

[Institutionalising The Rental Housing Market In India -2019](#)

2] Rental Housing India

<https://pdf.savills.asia/asia-pacific-research/india-research/rental-housing-india.pdf>

3)Fermi Estimation

<http://t.www.na-businesspress.com/JABE/Jabe105/AndersonWeb.pdf>

<https://indianexpress.com/article/opinion/columns/model-tenancy-act-homeowners-tenants-7427337/>

<https://www.techsciresearch.com/report/india-rental-housing-market/4811.html>

5) Behaviour in the purchasing of a residential property

[https://www.researchgate.net/publication/337049277 Factors Affecting Buying Behaviour in the Purchase of Residential Property A Factor Analysis Approach](https://www.researchgate.net/publication/337049277_Factors_Affecting_Buying_Behaviour_in_the_Purchase_of_Residential_Property_A_Factor_Analysis_Approach)

6) Magicbricks propindex

<http://property.magicbricks.com/microsite/buy/propindex/index.html>

GitHub:

<https://github.com/SridharRambhatla/Market-Segmentation-Pune-Real-Estate>