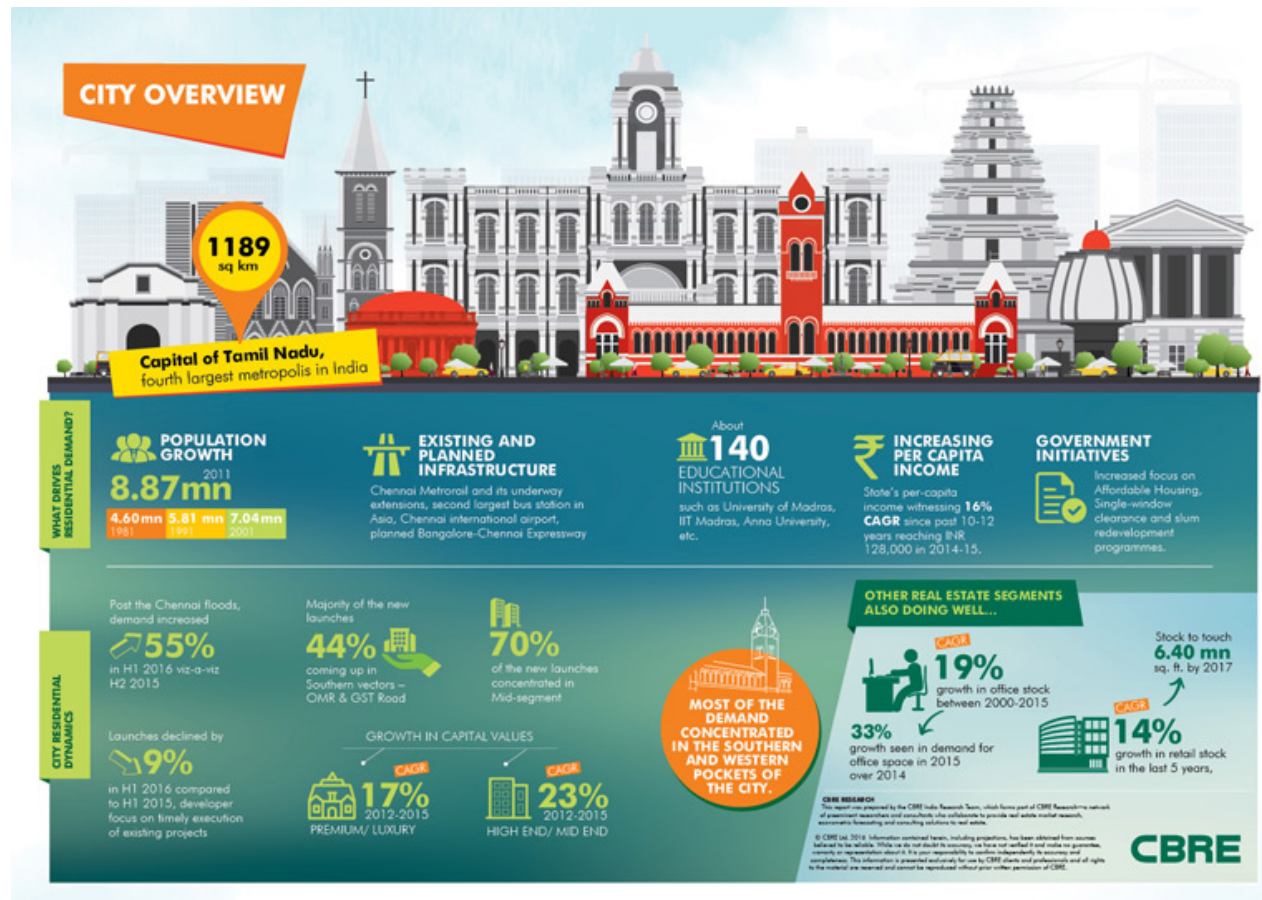CLIENT: VOQUE CONSTRUCTIONS

Site Selection Study for Residential Development Project in Chennai, India

Author: Amal Putti

Date: March 2020

(Picture Courtesy: CBRE)

1. INTRODUCTION

VOQUE CONSTRUCTIONS is one of Asia's largest real estate companies. Headquartered and listed in Singapore, it is an owner and manager of a global portfolio comprising integrated developments, shopping malls, lodging, offices, homes, real estate investment trusts (REITs) and funds. Present across more than 60 cities in over 10 countries, the Group focuses on Singapore and China as core markets while it continues to expand in markets such as Indonesia and Vietnam.

Voque is determined to foray into the Indian market and are considering in developing a multistory apartment project in Chennai. Chennai is the fourth largest metropolitan city of India. It is both the commercial and cultural capital city of the south-eastern state of Tamil Nadu. The city is developing rapidly and has become a fast-growing city in India. Currently there are many major industries in the field like automobile, Hardware manufacture, technology, health and education. It is ranked as a second major exporter of IT and is also well-known as the automobile capital of India.

Since this is Voque's first project in the country, and being unfamiliar with the local market of Chennai city, they have reached out to me to perform a study to:

1. *Identify neighborhoods in the city which have a low density of residences, as they consider these areas as ideal for their projects. Further classify these neighborhoods on the*

2. *Classify these neighborhoods based on the average price per square feet (Price PSQFT) into:*
   - *Low Budget Market*
   - *Premium Market*
   - *Luxury Market*

3. *Provide a list of the inventory of residential apartments currently available for sale in these neighborhoods and include their features such as:*

   - *Price*
   - *Floor Size*
   - *No. of Bedrooms*
   - *No. of Bathrooms*
   - *Furnishing Status (Unfurnished/Semi Furnished/Furnished)*
   - *Property Condition (Resale or New Property)*

2. DATA

Data used in this report was collected primarily from two sources – a leading Indian online real estate property listing website (Magicbricks) and a location data platform (Foursquare)

2.1. Inventory of Apartment Listings from https://www.magicbricks.com

The Magicbricks website gave us all the relevant information on the apartments that are currently for sale in Chennai. Data was scraped using Python libraries to extract the following information:

- *Neighbourhood* – This included all the localities in the Chennai Metropolitan area broadly classified into Chennai North, Chennai South, Chennai West and Chennai Central.

- *Neighbourhood Coordinates* – Latitude and Longitude of each neighborhood
- *Listing Information* – Price, Floor Size, No of Bedrooms and Bathrooms, Furnishing Status, Property Condition and Developer Information (where applicable)
- *Price per square feet (Price PSQFT)* – This Price per square feet was not directly available and was indirectly computed using the Price and Floor Size details from the Listing Information.

2.2. Residential Density of Neighborhoods from https://foursquare.com

Foursquare Places API's were leveraged through their Developer Platform to download Neighbourhood Attributes.

3. METHODOLOGY

The primary data for this analysis was scraped from the Magicbricks website. Since there is no accurate sales data on residential apartments for Chennai, an alternate approach was using data from a real estate listing website. The one limitation of using this data is the price of the listings is based on what the seller is expecting for the property, and not necessarily the final price which the unit would sell for. Pricing fluctuates based on market conditions and so it's not easy to predict what the sold price is. However, for the analysis needed to perform our client's study this data provides the most recent data available and all provides all the attributes of the available apartments in the market including Listing Price, Floor Size, # of Bedrooms, # of Bathrooms, and Geocoordinates. The data further allows us to classify the data based on listings for new developments vs. resale properties. This information is especially critical as it provides a valuable insight into how other developers are pricing their inventory.

3.1 Cleaning the Data

The data scraped from the magicbricks website had a few duplicates and missing information that was removed. Also, the Neighborhood (Locality) names were cleaned to remove the sub locality names as that is not necessary for our analysis. We also had to convert the data into appropriate Data types.

We obtained data for 108 Neighborhoods in the city.

3.2 Computing Price per sq. ft (Price PSQFT)

Price per sq. ft was not an attribute available in the data and had to be calculated. Since we had the listing price of the property (Price) and the size of the property (Floor Size), the Price per sq. Ft calculation was straightforward and was computed using the formula:

$$Price\ PSQFT = Price\ /\ Floor\ Size$$

Table 1: List of attributes extracted from the magicbricks webpages

ListingID	Locality	SizeType	Furnishing	Society	Price PSQFT
Property Type	Price	Bedrooms	Property Condition	Longitude	Listing URL
Region	FloorSize	Bathrooms	Developer	Latitude	

Figure 1: Data subset extracted and cleaned from the Magicbricks webpages

	ListingID	Property_Type	Region	Locality	Price	FloorSize	SizeType	Bedrooms	Bathrooms	Furnishing	Property_Condition
0	1532	Apartment	Chennai South	Sir Thyagaraya Road	103600000	3810.0	super area	4.0	5.0	Semi-Furnished	Resale
1	460	Apartment	Chennai South	Madras Boat Club Road	70000000	2200.0	carpet area	3.0	4.0	Semi-Furnished	New Property
2	1991	Builder Floor Apartment	Chennai South	Adyar	61000000	2786.0	super area	4.0	3.0	Furnished	Resale
3	1147	Apartment	Chennai South	Besant Nagar	29000000	1675.0	super area	3.0	3.0	Unfurnished	New Property
4	388	Builder Floor Apartment	Chennai South	Besant Nagar	24900000	1450.0	super area	3.0	3.0	Unfurnished	Resale

3.3 Exploring and Analyzing the neighborhoods using Foursquare API

Foursquare Places API's are leveraged through their Developer Platform to get the Neighbourhood Attributes for the localities. Fortunately, one of the attributes we get from this data is the 'Residence' category which allows us to identify residential neighborhoods which is one of the key data points required for our analysis.

Table 2: List of categories obtained using Foursquare API

Arts & Entertainment	College & University	Event	Food	Nightlife Spot
Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport

Figure 2: Data subset extracted from Foursquare API and appended to our data from magicbricks

Locality	Price_PSQFT	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
Abiramapuram	9282.000000	13.031990	80.258896	3	2	0	65	4	6	19	3	29	9
Adambakkam	6412.162261	12.989216	80.202051	1	2	0	7	3	6	7	3	4	3
Adyar	13258.000000	13.003388	80.257289	4	7	0	68	9	6	14	2	35	4
Alandur	5265.483234	13.000546	80.199760	4	5	1	15	3	5	6	2	3	7
Alwarpet	13649.000000	13.035668	80.250031	5	7	0	69	10	6	15	1	31	10

With data extracted using Foursquare, we not only have the Neighbourhood attributes (categories) for each Neighbourhood but also the count of the number of venues that fall in those categories. The count of venues can also be interpreted as the density of occurrence of a given category in a particular Neighbourhood. For ex, if we have a Neighborhood with a higher count of Residence categories, we can infer that the neighborhood is a residential neighborhood.

3.4 Clustering and Classifying Neighborhoods using k-means clustering algorithm

In this study, we use K-means algorithm for clustering neighborhoods. K-means is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. In our study, the neighborhoods will need to be grouped into clusters to identify neighborhoods with low residential density, and compare the other attributes of these neighborhoods with neighborhoods in other clusters.

3.4.1 Scaling the Data

Before we apply the k-means algorithm, we would need to scale the data. Scaling of data is necessary to normalize the data so is can be scaled to a fixed range between 0 and 1.

Figure 3: Data subset transformed to a scale between 0 to 1 after scaling

	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	0.157895	0.083333	0.000000	0.928571	0.307692	0.428571	0.904762	0.500000	0.617021	0.5625
1	0.052632	0.083333	0.000000	0.100000	0.230769	0.428571	0.333333	0.500000	0.085106	0.1875
2	0.210526	0.291667	0.000000	0.971429	0.692308	0.428571	0.666667	0.333333	0.744681	0.2500
3	0.210526	0.208333	0.333333	0.214286	0.230769	0.357143	0.285714	0.333333	0.063830	0.4375
4	0.263158	0.291667	0.000000	0.985714	0.769231	0.428571	0.714286	0.166667	0.659574	0.6250

### 3.4.2 Computing the optimum number of clusters

We used the 'Elbow' method to compute the optimum number of clusters for our dataset. From the "Elbow method" we see the optimum value of k as 3. This clusters the neighborhoods in the city into: *1) City Center 2) City Suburbs 3) City Outskirts.*
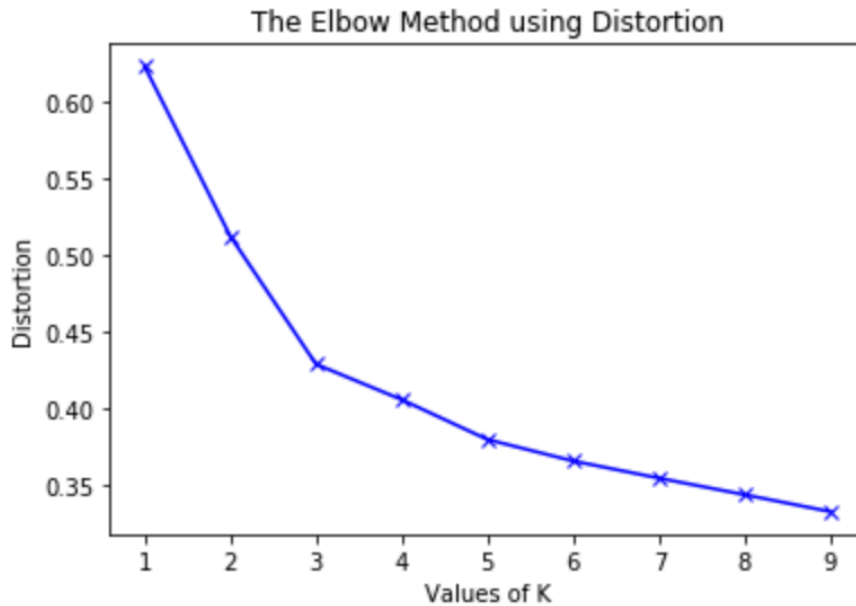
However, we found that increasing the value of k to 4, further splits the City Suburbs into Neighborhoods with established infrastructure and new neighborhoods with developing infrastructure. This is ideal for our study as typically new neighborhoods is where people prefer to invest in as the costs are lower, but they still have advantage of the new infrastructure being developed.
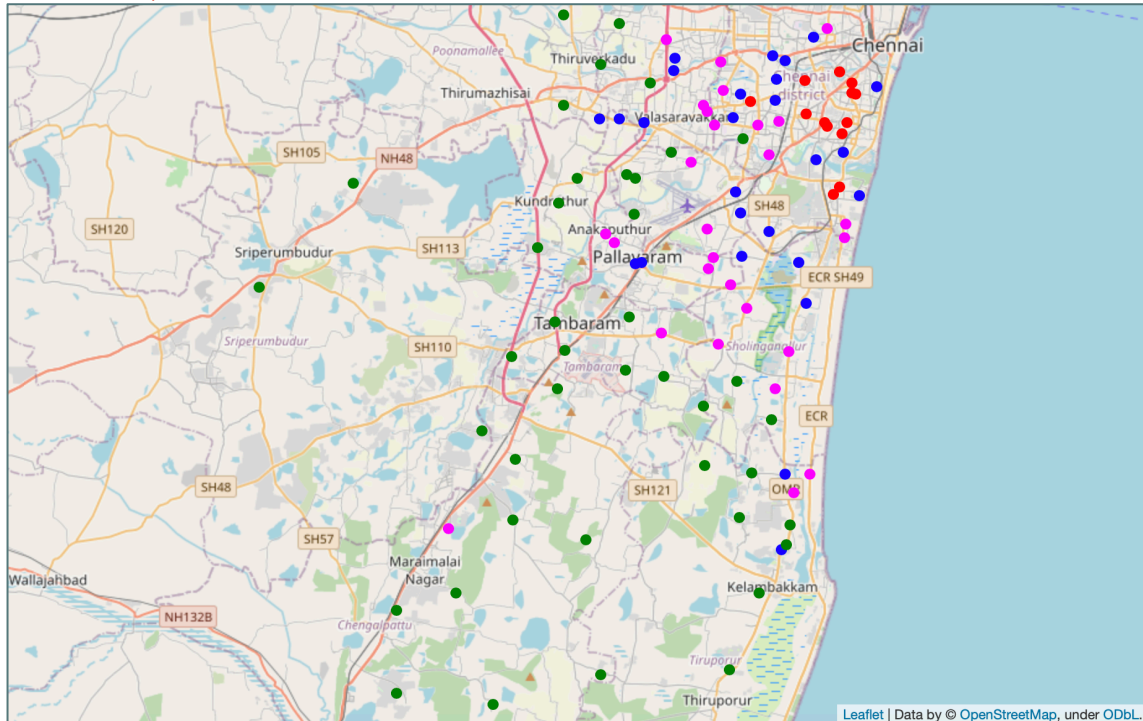
For our study further, we use the value k=4 and group our data into 4 clusters.

Figure 4: Plot of density of neighborhoods in a cluster by Category
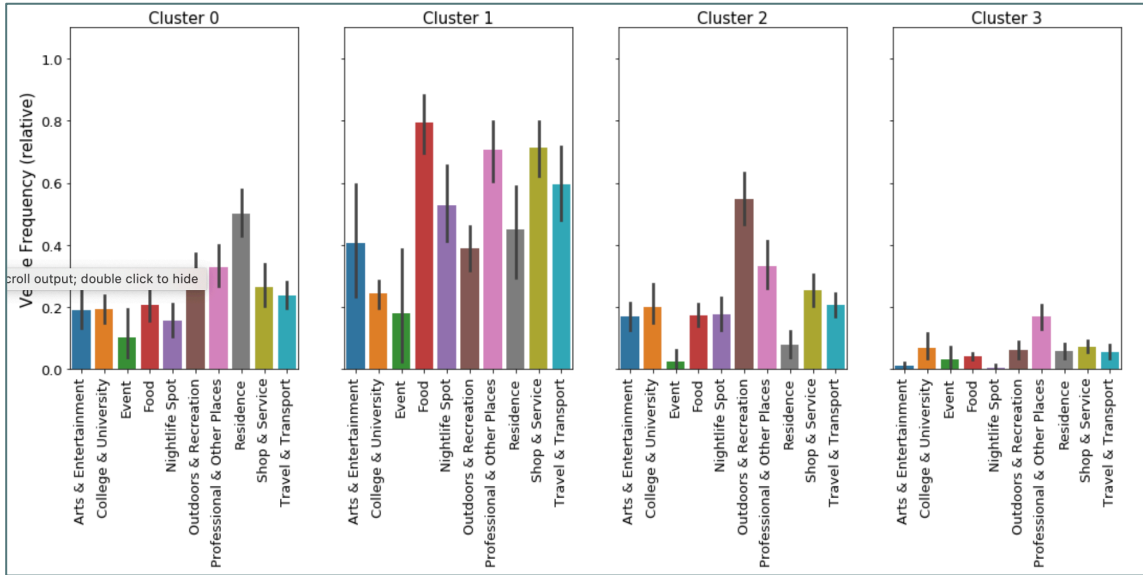
We will further map these neighborhoods on a map and visualize the neighborhoods clusters to be able to make observations.

Figure 5: Map of neighborhoods visualized by cluster

From the visualization and mapping of the clusters, we can make the following observations:
- **Cluster 0** are well established residential neighborhoods in the City Suburbs
- **Cluster 1** are neighborhoods with high density of both residential and commercial venues. These, as expected, are located in the center of the city.

- **Cluster 2** are neighborhoods with a low density of residences but have a similar profile as that of some of the more well-established residential neighborhoods.
- **Cluster 3** are neighborhoods in the outskirts of the city with a low density of both commercial and residential areas.

From these observations, neighborhoods in Cluster 2 and 3 have low density of residences. Given the proximity of Neighborhoods in Cluster 2 to the center of the city, which is advantageous, we will study them for further classification and analysis

3.5 Classifying Neighborhoods in Cluster 2 based on Average Price per sq. ft (Price PSQFT)

We further classify the neighbourhoods by the average price per sq. ft (Price PSQFT). The Price PSQFT allows us to understand how expensive a particular neighborhood is and determined as an average price per sq. ft. of all the individual active property listings in that neighborhood.

In Chennai, the well-established classifications of the markets based on Average Prices per sq. ft are:
- Low Budget Market - INR 2,500 to INR 6,000 per sq. ft
- Premium Market - INR 6,000 to INR 10,600 per sq. ft
- Luxury Market - INR 10,600 per sq. ft and above

Figure 6: Data showing Neighborhoods classified by Premium and Luxury Market

Market	Locality	Price_PSQFT	Latitude	Longitude
Luxury	Kottivakkam	12657.828280	12.975598	80.260239
	Thiruvanmiyur	10674.157300	12.983258	80.260666
Premium	Alwarthirunagar	6557.296534	13.043975	80.183716
	Ashok Nagar	10560.146920	13.036874	80.211647
	East Coast Road	7241.532977	12.848025	80.240552
	Manapakkam	6284.799150	13.016341	80.174619
	Purasawalkam	6487.000000	13.088560	80.250076
	Saidapet	9130.500000	13.020789	80.218315
	Uthandi	7055.181106	12.838152	80.231951
	Virugambakkam	7360.000000	13.055141	80.192688
	West Mambalam	8089.000000	13.038436	80.223373

Figure 7: Map of neighborhoods visualized by Market
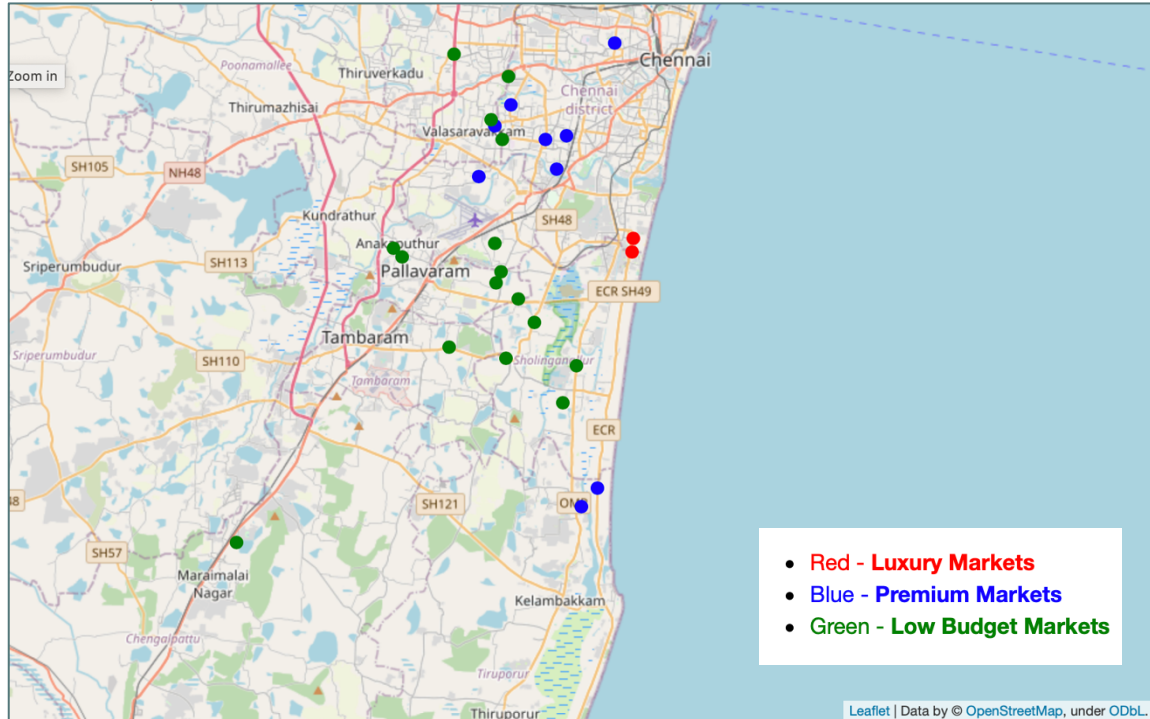
4. RESULTS

- We started by analyzing 108 Neighborhoods in Chennai City, and we were able to select 27 Neighborhoods that had a low density of residences and also had similar profile as that of some of the more well-established residential neighborhoods.

- Of these 27 Neighborhoods, we classified them further and identified 9 Premium Markets and 2 Luxury Markets which are potential neighborhoods to consider for a multi-story real-estate development.

- We are also able to provide a list of 893 active listings of apartments currently for sale along with their attributes for each of these listings in 9 Premium Markets and 2 Luxury Markets. This can be found here.

5. DISCUSSION

To further analyze the feasibility and cost benefit analysis of the shortlisted locations, one key data point required is the cost of land. If the cost of land is high, then the developer needs to sell the apartments at a higher price per sq. ft. If this price is significantly higher than the average selling price of the apartments in that neighborhood, the developer may find it difficult to offload his inventory. This study can be further augmented by factoring in this datapoint. The land price data can easily be obtained from magicbricks website.

Another datapoint that would be valuable for this study is the trend of the unsold inventory in these neighborhoods. If the trend shows that unsold inventory has been increasing, it indicates that the supply is exceeding demand and the developer would be competing with other developments. If such is the case, it wouldn't be advisable to launch a project in these neighborhoods.

6. CONCLUSION

In this report, we were able to provide recommendations of potential neighborhoods that meet the client's requirements. We were also able to give the client an understanding of the pricing in these neighborhoods, how the neighborhood profiles compare with each other, as well as provide an insight to compare other developments that are launched/to be launched in these neighborhoods.

We also recommend expanding this study to include analysis factoring in the cost of land prior to making a decision on the project.