

Capstone Project – The Battle of Neighborhoods (Week 1)

Note on

“A description of the DATA and how it will be used”

Table of Contents

A. Project:.....2

B. DATA required:.....2

C. Source of Data:.....3

D. Description and using the Data:.....4

E. Error/ Exception handling for data:7

A. Project:

The client is an EVENT MANAGEMENT entity, and specializes in conducting corporate seminars/ conferences.

The project deals with the problem of selecting a suitable area within a host of cities to set up an office. The client wants to buy/ take rent a commercial office at a suitable location in any of the four metro cities in India, i.e. in any of the following cities:

1. Mumbai,
2. Kolkata,
3. New Delhi, or
4. Chennai.

It plans to open up an office which will meet the following parameters/criteria:

1. The office has to be located in an area where commercial property rates are low.
2. The office has to be located in an area where there are a good number of restaurants so that its customers may have variety of choices for their snacks/ food in case hotel menu is not to their liking.
3. The office has to be located in an area where there are a good number of hotels so that it can arrange for both conferences and lodging for the customers.
4. The office has to be located within the vicinity of the city airport so that its customers may save time in transit.

B. DATA required:

The user will need the following data:

Sl. No.	DATA Required
1	Commercial property rates of different areas in a city.
2	Number of restaurants available in a locality/ area in the city
3	Number of hotels available in a locality/ area in the city.
4	Distance of a location/ area from the city airport.

C. Source of Data:

It is to be noted that the project is **USER INTERACTIVE** and the user will have the option of choosing preferences, **including the source of data.**

The following table illustrates the source of data sourced for this project:

DATA Required	DATA Source
Commercial property rates of different areas in a city.	Option I: Use Existing database already loaded in the notebook. Option II: Top 25 commercial zones for the selected city from website (www.magicbricks.com). The Notebook program uses the "Beautiful Soup" package for web-scrapping.
Number of restaurants available in a locality/area in the city	For meeting both the criteria, the following are used: 1. The Geolocator library. 2. The Foursquare Location data.
Number of hotels available in a locality/area in the city.	
Distance of a location/ area from the city airport.	The Geolocator library.

D. Description and using the Data:

This section deals with how the data arrived from the above mentioned sources have been utilized for the project.

(i) Commercial property rates of different areas in a city.

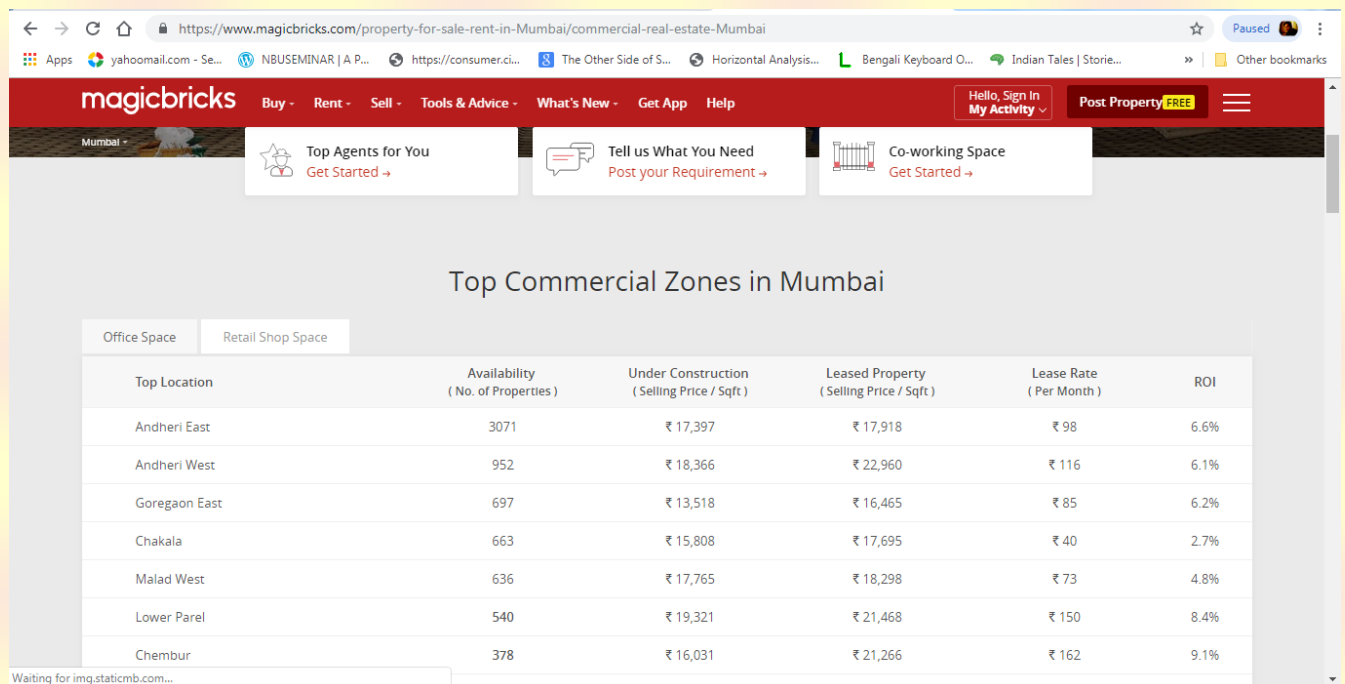
(a) **Option I:** Use **Existing database** already loaded in the notebook.

The client has the option to use the existing databases for the 4 cities, which shows the top 25 commercial zones for each city.

The source of the existing database is the website: www.magicbricks.com.

(b) **Option II:** Top 25 commercial zones for the selected city from website (www.magicbricks.com)

The client will have the option of using data directly from this website. A screen-shot of this website is given below:



The screenshot shows the Magicbricks website interface. The header includes navigation links for Buy, Rent, Sell, Tools & Advice, What's New, Get App, and Help. A red banner at the top right says 'Hello, Sign In My Activity' and 'Post Property FREE'. Below the header, there are three main sections: 'Top Agents for You', 'Tell us What You Need', and 'Co-working Space'. The main content area is titled 'Top Commercial Zones in Mumbai' and features a table with columns for Office Space, Retail Shop Space, Top Location, Availability, Under Construction, Leased Property, Lease Rate, and ROI. The table lists seven locations: Andheri East, Andheri West, Goregaon East, Chakala, Malad West, Lower Parel, and Chembur.

Office Space	Retail Shop Space	Top Location	Availability (No. of Properties)	Under Construction (Selling Price / Sqft)	Leased Property (Selling Price / Sqft)	Lease Rate (Per Month)	ROI
		Andheri East	3071	₹ 17,397	₹ 17,918	₹ 98	6.6%
		Andheri West	952	₹ 18,366	₹ 22,960	₹ 116	6.1%
		Goregaon East	697	₹ 13,518	₹ 16,465	₹ 85	6.2%
		Chakala	663	₹ 15,808	₹ 17,695	₹ 40	2.7%
		Malad West	636	₹ 17,765	₹ 18,298	₹ 73	4.8%
		Lower Parel	540	₹ 19,321	₹ 21,468	₹ 150	8.4%
		Chembur	378	₹ 16,031	₹ 21,266	₹ 162	9.1%

The Notebook program uses the “**Beautiful Soup**” package for web-scraping.

- Either of the options will result in creating a data-frame providing details of the top 25 commercial zones in the city.
- In case the website refuses to connect for web-scraping, the programme automatically redirects the user to the existing database for the chosen city.
- **Difference between the 2 options:**

While the option of using web-scraping through Beautiful Soup will result in updated data available on the website, the data available by using the existing database option is as on May 2019 and cannot be changed, hence not dynamic. The program has provided both the options as web-scraping may always not be possible as the website sometimes blocks scrapping for security reasons.

- **A sample of the dataframe created (for city: Kolkata) is as below:**

```
For running on existing database, enter 'A',
For scrapping through website, enter 'B'
```

```
Enter your choice B
```

```
You have decided to use the scrape through the website
https://www.magicbricks.com/property-for-sale-rent-in-Kolkata/commercial-real-estate-Kolkata
```

[7]:

	Location	No_of_prop	SP_under_cons	SP_leased_prop	Rentals_per_month	ROI
1	Salt Lake City	1489	7460	8109	49	7.3%
2	Rajarhat	329	5353	5956	25	5%
3	New Town	168	4616	5905	44	8.9%
4	Park Street	151	22004	22664	85	4.5%
5	Kalighat	130	7650	8138	52	7.7%
6	A J C Bose Road	106	15500	15965	80	6%
7	Ballygunge	103	10031	11146	39	4.2%
8	Kasba	101	7403	8226	38	5.6%
9	Gariahat	95	20240	22000	41	2.2%
10	Camac St-Park Street area	85	19500	20085	98	5.8%
11	Salt Lake City Sector 2	83	7651	8502	18	2.6%

Columns description of the above table:

1. **Location** : Top Commercial Areas/ locations within the city.
2. **No_of_prop** : Number of commercial properties available.
3. **SP_under_cons** : Selling price of properties under construction.
4. **SP_leased_prop** : Selling price of readily available properties.
5. **Rentals_per_month** : Existing rates if properties are taken on rent.
6. **ROI** : Return on investment if the property is held for investment purpose.

PS : All the prices/ rates are in *Rs.per sq.ft*

USAGE of the above data:

The above data-frame thus arrived will be used to rank/ score the locations based upon the type of property (*to be chosen by the user*).

- (ii) Number of restaurants & hotels available in a locality/area in the city.

For meeting both the criteria, the following are used:

1. The **Geolocator** library.
2. The **Foursquare Location** data.

USAGE of the above data:

The Geolocator library.	The Foursquare Location data.
<p>Get the coordinates of different commercial locations within the city, helping in:</p> <ol style="list-style-type: none">a. Providing input for Folium Map.b. Providing input for Fousquare location data to get venues.c. Providing input to calculate distance of each location from the city airport, thus enabling to rank/ score each location.	<p>Get the list of venues and venue categories for each location within a city. Input from the Geolocator library is used for the same.</p> <p>This will be used to determine the number of hotels and restaurants for each location, thus enabling to rank/ score each location.</p>

E. Error/ Exception handling for data:

Since the program is **User Interactive** and is dependent upon the inputs from the user, quality/ quantity of data is prone to be affected due to **invalid entries** by the user.

However, the program is designed to handle errors/ exceptions arising out of invalid entries.

The programme will also provide an explanation/ message why the input entered by the user is invalid. The examples below substantiate the above:

1. Invalid entry for choice of City:

```
print("\n\nYou have chosen :\n\n",city)
```

Available Cities

```
['kolkata', 'chennai', 'new delhi', 'mumbai']
```

Enter Your City: new york

INVALID ENTRY TRY AGAIN

Enter Your City: kolkata

You have chosen :

kolkata

2. Invalid entry for choice of dataset:

For running on existing database, enter 'A',
For scrapping through website, enter 'B'

Enter your choice c

INVALID ENTRY TRY AGAIN

For running on existing database, enter 'A',
For scrapping through website, enter 'B'

Enter your choice B

You have decided to use the scrape through the website

<https://www.magicbricks.com/property-for-sale-rent-in-Kolkata/commercial-real-estate-Kolkata>

[7]:

	Location	No_of_prop	SP_under_cons	SP_leased_prop	Rentals_per_month	ROI
1	Salt Lake City	1489	7460	8109	49	7.3%

3. Invalid entry for choice of property:

```
df_final

Enter Your Choice
r
INVALID ENTRY... PLEASE TRY AGAIN
Enter Your Choice
pr
You have chosen to take property on rent..... getting database.
```

[8]:

	Location	No_of_prop	Price/rentals
1	Salt Lake City	1489	49
2	Rajarhat	329	25
3	New Town	168	44
4	Park Street	151	85

-----END OF NOTE-----