

# **Analyzing New Delhi for Opening a New Business**

Abhiroop Santra

July 8, 2020

## **1. Introduction**

### **1.1 Background**

For this project, we assume that our client Acme Corp. is a large conglomerate with successful businesses running across several sectors. They have hotel chains, restaurants, shopping malls and fitness centers. Acme Corp. intends to expand its business in New Delhi, India. As India is one of the most populous countries in the world, hence New Delhi, the capital city of India can be an attractive spot to open a new business. This will give Acme Corp. access to a large customer base. Also, this city can serve as a good foothold from where it can expand to the rest of the country.

### **1.2 Problem**

New Delhi is a vast city with several localities. It is crucial for Acme Corp. to choose the right locality to establish a particular business. If our client opens a certain business where the demand for such services will be low then it is unlikely the business venture will succeed. Hence we need to determine, a good location for our client to open their business. This project aims to answer this question for our client and help them decide which locations are more suitable for a certain type of business.

### **1.3 Interest**

This analysis is essential for our client as opening a business in the ideal location will not only increase profits but will also allow them to expand in the country in

the coming years. The report will be interesting to the board of directors and the senior management who is our target audience.

## **2. Data Acquisition and Cleaning**

### **2.1 Data Sources**

We will require the data on the various localities of New Delhi. We will need, at least, the names of all the localities that we are planning to analyze. Fortunately, there is a data set available from Kaggle which can be downloaded from [here](#). This data is from 2018, but as the locality details of New Delhi is unlikely to change in this period of time, so we can consider it accurate.

### **2.2 Data Cleaning**

The data that we have obtained contains information of several cities. So first, we need to filter the data only for New Delhi. Once this is done, we will end up with all the listed restaurants in New Delhi from 2018.

Next, we will be selecting the names of all the localities provided in the filtered data. As this data set is a bit old, so the restaurant details are probably not up-to-date. Hence we will ignore the restaurant data, which we will obtain later from Foursquare API. For now, we will simply select the various localities available.

After this step, we need to drop any row that does not have locality data as we will not be able to conclude anything from these rows.

The next step would be to remove any duplicate locality names from the data. This will also reduce the number of rows to work with. Once this is accomplished, we will have a complete list of all localities in New Delhi.

However, we are still missing an important information, that is the coordinates (latitude and longitude) of each locality. We will see how we can obtain this in the next section.

### **2.3 Fetching the Missing Details**

We will be using the Python Geocode library to get the coordinates of each locality. There is a small drawback to this data. As the locality information is not

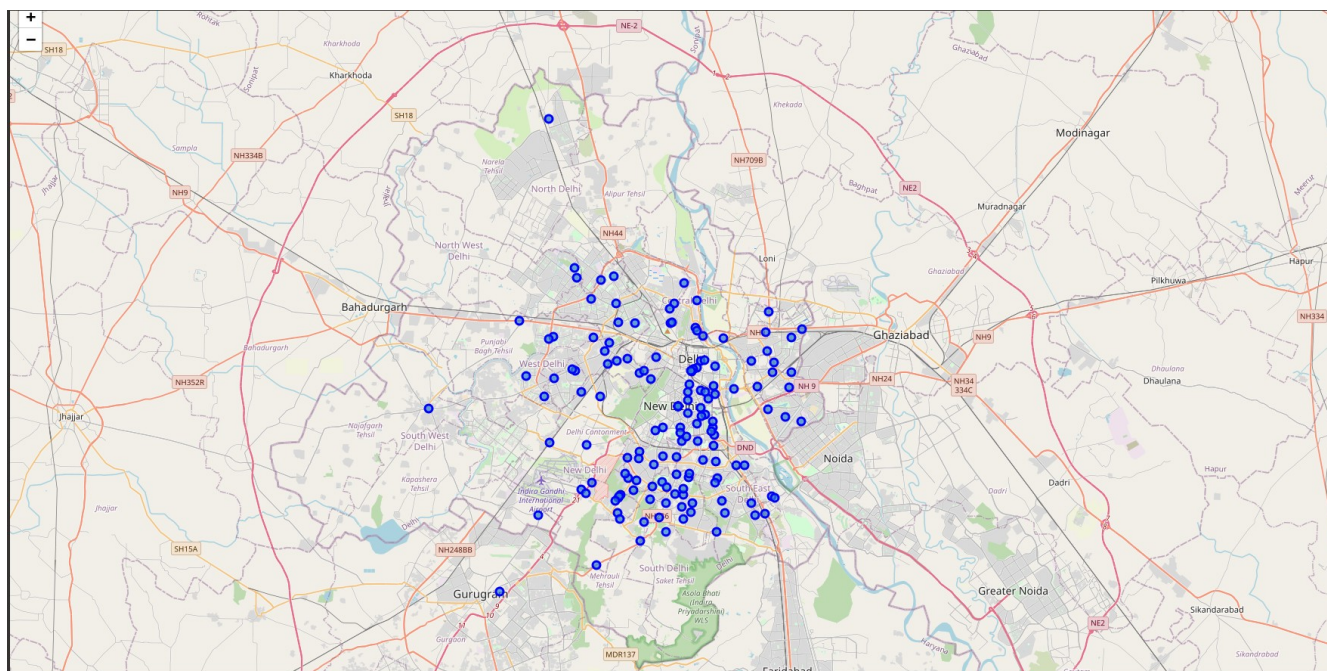
supplied officially by the New Delhi Municipality, it might have some incorrect locality names. In this case, we will be ignoring these incorrect localities as rectifying the locality names is beyond the scope of this project.

We will loop through our data and assign each locality its respective coordinates. Once this is done, we will have the data with which we can start the analysis.

## 3. Exploratory Data Analysis

### 3.1 Plotting the Localities

We will be using the Python Folium library to plot the localities using the coordinates received from the previous step, to the map of New Delhi. We will be able to view the locations of each locality on the map which will help us to analyze how each of them are connected and this will also influence the demand of a certain type of business.



### 3.2 Using Foursquare API for Popular Venues

We will be using the Foursquare Places API to get the popular venues for each locality. We do a trial run by selecting the first locality from our data and getting

the nearby venues within a radius of 500 meters. Also, we restrict the maximum number of venues to 100, so that we don't overload the API.

The API returns a JSON result. We will be extracting the venue name, the category of the venue and its coordinates.



	name	categories	lat	lng
0	100% Rock	Pub	28.535552	77.197523
1	Cafe Coffee Day	Café	28.538589	77.198683
2	Kuzart Lane	Café	28.537990	77.198368
3	Waves Restaurant	Indian Restaurant	28.538582	77.198771
4	Qutub Residency Hotel New Delhi	Hotel	28.535727	77.197161

We see that the API returns our required data successfully. So we will create a function that will loop through all the locality rows in our data and will fetch the venues in each locality along with their categories and coordinates.

Also, we group the new data, obtained from the API, based on the locality and check how many venues were fetched from Foursquare for each locality.

### 3.3 Analyzing the popularity of Venue Categories

Next, we will try to find out which types of venues are popular in each locality. This will help our stakeholders get an idea about the most lucrative location to open a certain type of business.

We see that there are 186 unique types of categories in the data.

To start this analysis, we will first do one-hot encoding on our data using the venue categories, to convert each category to a column. Next, we group the result based on the Locality name and find the mean of each category. This will show us which particular categories occur more frequently in a locality. A higher mean translates to a higher frequency of occurrence. This indicates which type of business is most likely to succeed in this locality.

	Locality	ATM	Accessories Store	Airport	Airport Food Court	Airport Terminal	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	Auto Dealership	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar	Bed & Breakfast	Beer Garden	Bike Shop	Bistro	Bookstore	Boutique
0	Adchini	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
1	Aerocity	0.000000	0.052632	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.000000	0.052632	0.000000	0.00	0.000000	0.000000	0.000000
2	Alaknanda	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.200000	0.00	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
3	Ambience Hall, Vasant Kunj	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.065217	0.000000	0.00	0.000000	0.00	0.000000	0.021739	0.000	0.021739	0.000000	0.021739	0.00	0.000000	0.021739	0.000000
4	Anand Lok	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
5	Anand Vihar	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
6	Asaf Ali Road	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.200000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
7	Barakhamba Road	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.062500	0.125000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.125000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
8	Bhikaji Cama Place	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.090909	0.000000	0.00	0.000000	0.00	0.000000	0.090909	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
9	Chanakyaपुरी	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000

Now, we loop through each row in our data and find out which are the top 5 venue categories in each locality. These categories have the highest frequency of occurrence and hence are most likely to succeed in their corresponding locality.

```
-- --Adchini-- --
      venue  freq
0         Pub  0.29
1        Café  0.29
2 Indian Restaurant  0.14
3  Parsi Restaurant  0.14
4         Hotel  0.14

-- --Aerocity-- --
      venue  freq
0        Hotel  0.32
1   Coffee Shop  0.11
2 Shopping Mall  0.05
3   Train Station  0.05
4  Thai Restaurant  0.05

-- --Alaknanda-- --
      venue  freq
0   BBQ Joint  0.2
1 Indian Restaurant  0.2
2   Steakhouse  0.1
3 Middle Eastern Restaurant  0.1
4 Food & Drink Shop  0.1
```

Our next task is to create a dataframe that will have all the locality names along with some columns containing the most frequent venue categories in each locality. We will be using this dataframe as the input for our clustering machine learning model.

To achieve this, we create a custom function to return the most frequent venue categories for a locality and then we loop through our locality data and retrieve the venue category data. Next, we append the most frequent venue categories as columns next to the locality names.

The result will look something like the following:

	Locality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adchini	Pub	Café	Indian Restaurant	Parsi Restaurant	Hotel	Yoga Studio	Farmers Market	French Restaurant	Food Truck	Food Court
1	Aerocity	Hotel	Coffee Shop	Indian Restaurant	Train Station	Buffet	Bed & Breakfast	Punjabi Restaurant	Plaza	Thai Restaurant	Shopping Mall
2	Alaknanda	Indian Restaurant	BBQ Joint	Pizza Place	Food & Drink Shop	Steakhouse	Middle Eastern Restaurant	Coffee Shop	Restaurant	Yoga Studio	Farmers Market
3	Ambience Mall, Vasant Kunj	Coffee Shop	Shopping Mall	Fast Food Restaurant	Café	Asian Restaurant	Clothing Store	Italian Restaurant	Chinese Restaurant	Indian Restaurant	Deli / Bodega
4	Anand Lok	Café	Dessert Shop	Hotel	Golf Course	Other Great Outdoors	Music Venue	Metro Station	Yoga Studio	Food Truck	Food Court

## 4. Clustering the Localities

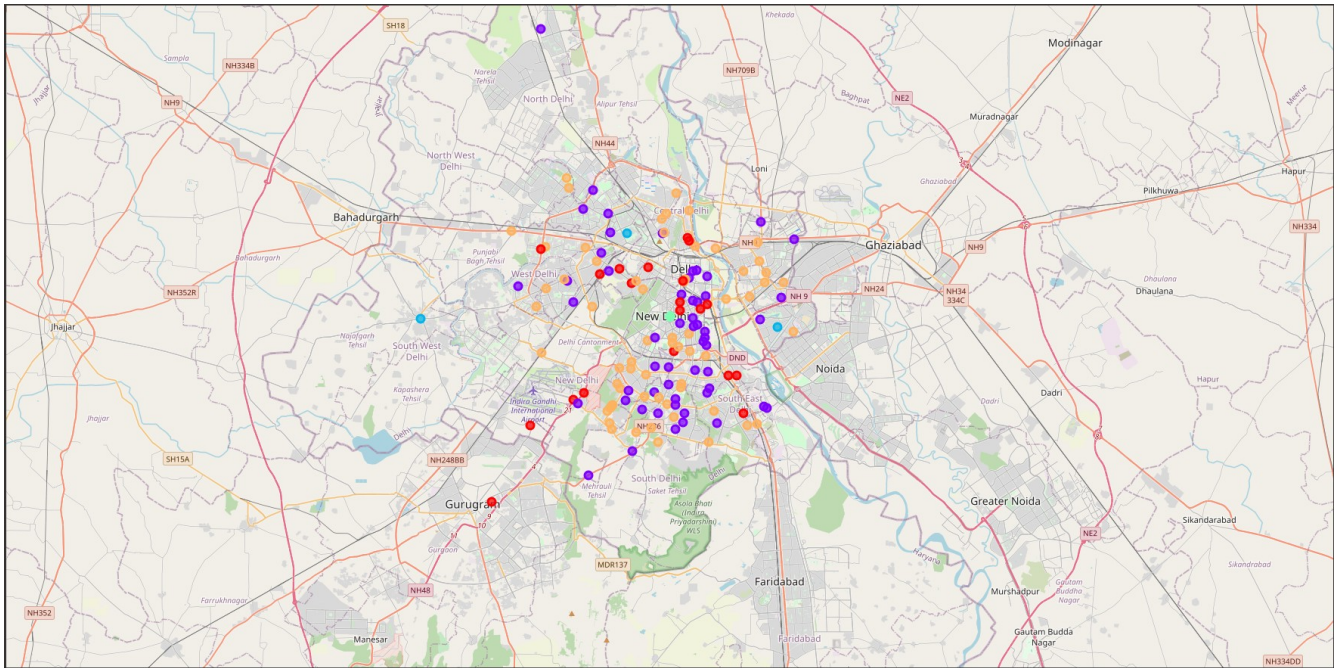
### 4.1 Generate the Clusters

We will be using K-Means algorithm provided by the Python scikit-learn library to cluster all the neighborhoods. We will start with an initial assumption of 5 clusters and will proceed with generating the model. Once the clustering is completed on the dataframe generated in the earlier step, each locality of the dataframe will be assigned a cluster.

### 4.2 Plot the clusters on the Map

Our next step would be to plot the generated clusters on the map of New Delhi, so that we can get a better idea how each cluster is geographically located. The generated clusters are shown in the map below:





## 4.3 Analyzing the clusters

We will analyze the clusters, one by one, to determine which cluster is more suitable to build which type of business.

We check the rows in cluster 1 and find that hotels are the 1<sup>st</sup> or 2<sup>nd</sup> most frequent venue category in most localities. This signifies that hotels are in high-demand in the localities grouped inside cluster 1. Hence, it would seem that opening a hotel in these localities might be profitable.

Next, we inspect cluster 2 and find that it contains several restaurants and places to dine, which means a restaurant business will be very successful if it is opened in one of the localities grouped in cluster 2.

After this, we check the 3<sup>rd</sup> cluster and find that ATMs are the most common occurrence in the localities of this cluster. Since Acme Corp. also is in Banking, they can easily set up an ATM machine or a bank branch in these localities.

Now, we analyze the 4<sup>th</sup> cluster and see that it is mostly having Music Venues, Yoga Studios and Rest Areas. So, people usually visit these localities to rest and relax after work and pursue recreational activities. Hence, if Acme Corp. plans to open a gym or a fitness center, the localities of cluster 4 will be the most suitable spot.

We inspect the last cluster and find that it has some shopping malls, some cafes, markets and other small shops (which might be a part of the shopping malls).

Hence, if a shopping mall is opened in one of these localities, it would ensure that there is a high number of footfall in the mall and will ensure its success.

## **5. Conclusion**

In this analysis, we scout the city of New Delhi to find which localities are suitable for which types of business. We used Foursquare API to get the most frequent venue categories in each locality and inferred which types of business establishments will most likely succeed in which localities. This report will help Acme Corp board of directors and Senior Management to decide which localities to aim for when opening a hotel, a restaurant or a shopping mall.

## **6. Future Directions**

Although this analysis predicted the likelihood of success of a business in each locality of New Delhi based on the popularity of the establishment type (venue category), several other factors can influence the success of a business, which were not present in the initial data. Some of these factors include the availability of infrastructure like roads, electricity, water, etc. Other factors include governmental factors like tax rates, real estate prices, etc.

With the availability of additional data such as these, we can create other machine learning models which will also help to determine a suitable locality based on operational cost and construction cost and this will further enhance the likelihood of the business venture to succeed.