

The Battle of Neighborhood

Er. Babloo Sah

Data Scientist

Introduction/Business Problem:

It has always been difficult for person to accommodate in new society. People spent lot of time in selecting proper Neighborhood while relocating to new places irrespective of their period of stay or reason of relocation (new home, new jobs, better life, opportunities etc.). One thing that people cannot leave behind is the society’s values and their surrounding environment.

There was a period of time when infrastructure was limited but had plenty of time. But today’s scenario are quite different. We have sufficient infrastructure but we lack time. Customer are so calculative about their time and could hardly visit one neighborhood per week. It has been hard task for real estate agent to showcase them appropriate neighborhood. Our data show around 80% of our visiting customer had negative response toward neighborhood that was presented to them. They mostly compare neighborhoods with their current and expect similar or better than that.

We intend to build a model that will select the neighborhood which are similar to customer needs and present us with the list of all possible neighborhood that customer would be interested in. We want most out of our visiting customer to grow our real estate business.

Data:

Based in the nature of problem we will use segmentation and clustering algorithm to cluster similar kind of neighborhood. We will need few specific data from customer to have insight of their neighborhood.

1. Current Customer city
2. Relocation city name

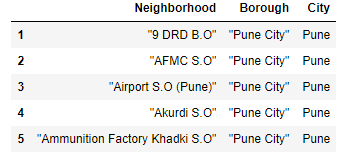
If customer is relocating within a city, we will not include that customer as they have better insight of their city and can suggest themselves better one.

These customer data will be collected from customer requirement form.

Later we will use government open data source to import Borough and neighborhood names of relocating city. We will import the data from csv file downloaded from below government database.

[India postal code](https://data.gov.in/resources/all-india-pincode-directory) : <https://data.gov.in/resources/all-india-pincode-directory>

A sample of data extracted from the government is as shown below:



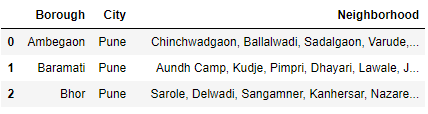
Based on the data obtained from government database, we will find out list of venues around the different neighborhood of both the cities and cluster them based on their similarity. We will use Foursquare API to obtain required venues details. Below are the list of venues that we will use to categories our neighborhood.

1. Building
2. Education
3. Nightlife
4. Food
5. Travel
6. Parks and Outdoors
7. Shops
8. Art and Entertainment

Methodology:

Our model is mainly focused to properly group similar localities of both city. In order to categorize locality, we will use unsupervised learning algorithm. In particular K-means clustering algorithm.

To implement this algorithm we will perform some data preprocessing step to clean the data. These extracted data like neighborhood, Borough name, etc. has some unwanted character which needs to be removed. We will merge both the city neighborhood and create single dataset. It will be easy to group them together.



Our model is planned to categorize the city’s localities based on trending venues of that localities. We will use Foursquare API to obtain venue details. Latitude and longitude of each borough will be extracted using Nominatim. We will feed these location information to Foursquare API to obtain venue details.

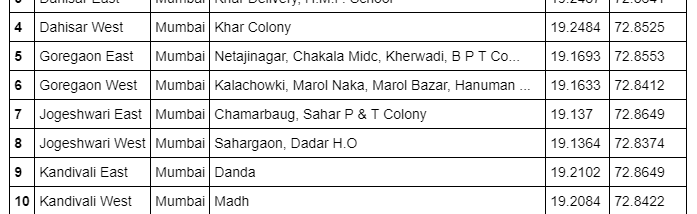


Figure 1: Location of each Borough



Figure 2: Venue count for each Borough

The Above information was feed to k-mean algorithm that clustered the Borough in five different categories based on their venue details of each neighborhood.

Result:

We cluster the similar neighborhood of both the relocating city and current city of customer. By referencing the information extracted from above, we will be able to showcase similar neighborhood to customer.



Figure 3: Neighborhood cluster table

Neighborhood belonging to each cluster are similar. The top venues of each customer will help in identifying the significant venue of each neighborhood.

Figure 4: Neighborhood top venues

The neighborhood of both the city are as shown belon on mas. Neighborhood with similar charcteristic are represented with similar color.

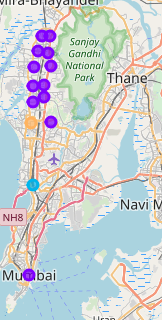
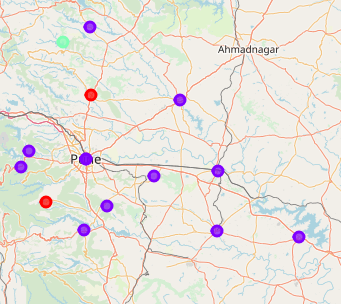


Figure 5: Mumbai Neighborhood Clustering

Figure 6: Pune neighborhood clustering

Discussion:

We used Foursquare API to extract venues details of each neighborhood. Based on venue details, the group of similar neighborhood was obtained. Apart from that we can group neighborhood based on trending attributes of that venues.

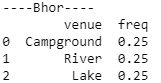
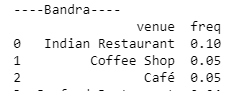
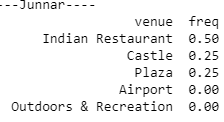


Figure 7: Junnar Venues

Figure 8:Bandra Venues

Figure 9: Bhor Venues

As shown in above fig, we can clearly distinguish the type of place and its significance. For example Bhor is a tourist place with landscape while Junnar has historical monuments.

Conclusion:

With this project, we were successfully able to compare neighborhood of two different cities and group them based on different venues.