IBM Applied Data Science Capstone Project

- By Nishant Agarwal (June 2020)

Title – *Opening a Hostel around Jaipur District, Rajasthan, India*

REPORT

Introduction:

In todays time education field is growing exponentially and in that too ther seems to be competition as to who is the best. In this competition the students/Professors finds it hard as in to which they should join. When this decision is over a big yet important factor arises that is where to live and how to manage for it within certain conditions. Many Institutes provides their facilities for it nowadays where most of the essentials are being done.

As for the other institutes, peoples have to search for accommodation within their taste and budget. This arises the need for Hostels which include accommodation to live with various facilities included as well as food mess in it which depends upon the owner owing the hostel.

Jaipur is the capital and the largest city of the Indian state of Rajasthan. As of 2011, the city had a population of 3.1 million, making it the tenth most populous city in the country. Jaipur is also known as the Pink City, due to the dominant colour scheme of its buildings. It is located 268 km (167 miles) from the national capital New Delhi, India. It is also a district of Rajasthan which covers various cities under it. In addition to its role as the provincial capital, educational, and administrative centre, the economy of Jaipur is fuelled by tourism, gemstone cutting, the manufacture of jewellery and luxury textiles, and information technology. Jaipur is nowadays being popular for various good institutes for study especially for the entrance exams like JEE & NEET and not to mention some good college/Universities of all sorts of departments.

In such a populous city/district/capital there occur need for hostels for students/teachers such that they can live like home, away from their homes

and most of their demand is to take the one with best possible location such that they have access to most daily essentials and entertainment places to visit.

So, our problem would be to search for the best place in Jaipur district (which includes many cities other than Jaipur) to build a Hostel which could cover maximum possible cities near it such that more people could access to it which would definitely benefit everyone. The **target audience** is all the people who are in search for Hostels for accommodation in best location, peoples who want to buy and open one at a best possible location and someone who wants to work or provide work to a hostel for income purpose.

Data:

Data containing Pin Codes, latitude, longitude and other stuff of India was being taken and used from Kaggle (https://www.kaggle.com/n1sarg/indian-postal-codes). This data is then modified manually and useful information related to only Jaipur district was being extracted to be used for the project. It would be then be cleared and a Data Frame which include all the useful information which we require will be made using pandas and other libraries.

The data which is finally used in the notebook from the start to be used in the project could also be accessed by the following below link –

https://drive.google.com/file/d/1yX-BqZ_Em6mBiqrJJrNW-z2gq61PgUv4/view?usp=sharing

Methodology:

Firstly, we need to get the list of neighbouring cities to be included in the project, which are to be near and under Jaipur district. We need their names, latitudes, longitudes and their area code i.e Pin Codes. As mentioned in the above section the following data was successfully retrieves to further process the project. We need to get the geographical coordinates of cities (Jaipur especially) in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using

Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the map of Jaipur district.

The data has to be shortened to reduce the time taken by api and the unnecessary space it takes. So, only data of unique pin codes was being kept and rest was to be eliminated.

Next, we will use Foursquare API to get the top 150 venues (although it gives maximum 100 in every case) that are within a radius of 10000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each city and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each city by grouping the rows by city and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Hostel" data, we will filter the "Hostel" as venue category for the neighbourhoods. This was done as to have minimum impact out of competition from other hostels.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Hostel". The results will allow us to identify which city have higher concentration of other Hostels while which city have fewer number of Hostels. Based on the occurrence of Hostels in different cities, it will help us to answer the question as to which cities/locations are most suitable to open new Hostels.

Results + Discussion + Conclusion:

According to this, it is analysed and concluded that:

All the clusters above are great locations for hostel to be built according to our requirements.

Most of the hostels are already available in cluster 0 so opening a new one their does not impact much and thus it will be a bad idea.

In cluster 2, there is less number of hostels so it could be a good choice to make a new one but before coming to conclusion if take a look at cluster - 1, it is surprisingly great for our conditions to meet as it contains little to NO hostels and has a great location too.

So, Cluster - 1 is definitely the best choice to be picked among the three for a hostel to be built.

The following is the resulted clustered graph containing all three different clusters.