Coursera Capstone
IBM Applied Data Science Capstone

Opening a New Hotel for Tourism in the most populous cities of India

By: Nakul Lakhotia

July 2020

Business Problem

- About 10.89 million foreign tourists visited India in 2019, an increase of 3.1% from the year before. Forex earnings from inbound tourists rose 8.2% to Rs 2.2 lakh crore in the past year.
- Since the domestic and international tourist inflow has been increasing every year, there is a bright future for the hospitality industry in India.
- Hotels are definitely one of the fastest-growing sectors in the tourism sector and it is truly justified as accommodation is the key part in the development of any country or region's tourism

Data

To solve the problem, we will need the following data:

- List of India neighbourhoods/cities which are highly populated according to Census 2011. This defines the scope of this project which is confined to the country of India.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to Hotels. We will use this data to perform clustering on the neighbourhoods
- Source- https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population

UA[a]	State/Territory	Population (2011)[4]
Mumbai	Maharashtra	18,394,912
Delhi	Delhi	16,349,831
Kolkata	West Bengal	14,112,536
Chennai	Tamil Nadu	8,696,010
Bangalore	Karnataka	8,520,435
Hyderabad	Telangana	7,749,334
Ahmedabad	Gujarat	6,361,084
Pune	Maharashtra	5,057,709
Surat	Gujarat	4,591,246
Jaipur	Rajasthan	3,073,350
Kanpur	Uttar Pradesh	2,920,496

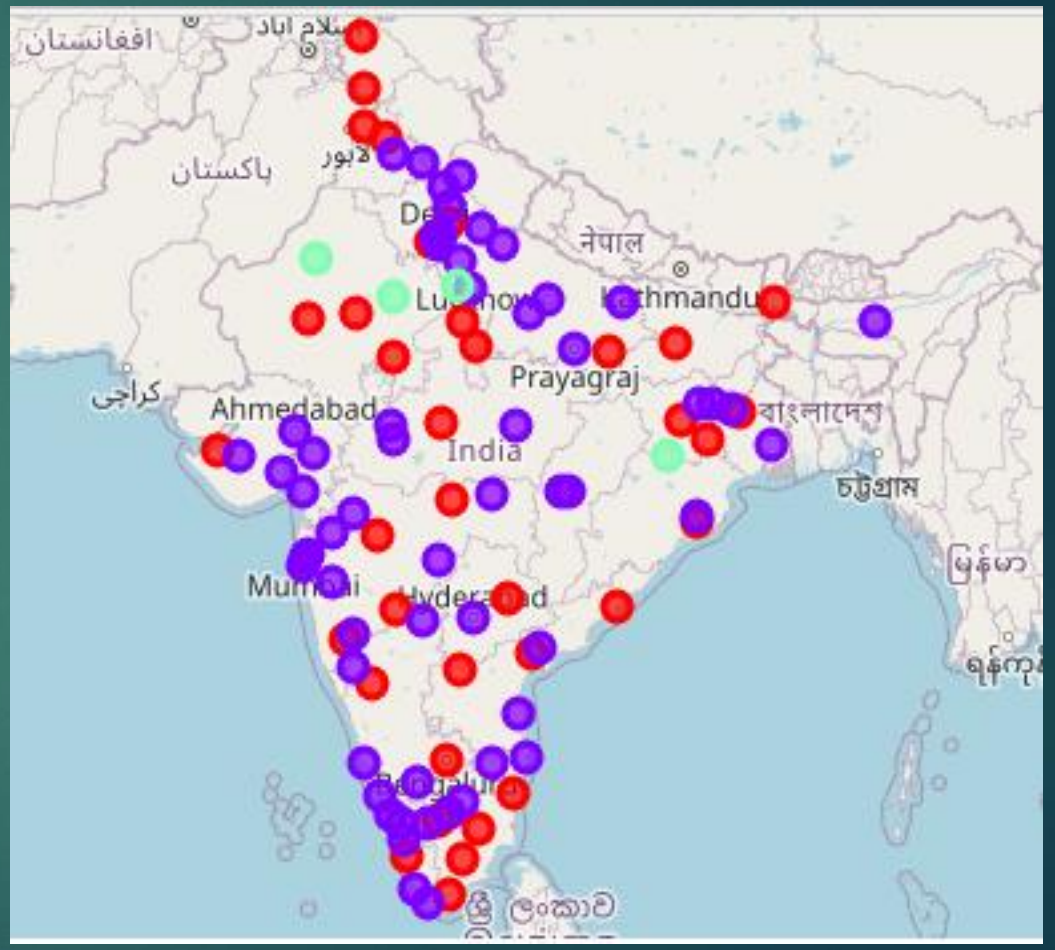
Methodology

- Web scraping Wikipedia page for neighbourhoods list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by neighbourhood and taking the mean of the frequency of occurrence of each venue category
- Filter venue category by Hotels
- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

	Neighborhood	Population	Latitude	Longitude
0	Mumbai	18394912	18.94017	72.83483
1	Delhi	16349831	28.63410	77.21689
2	Kolkata	14112536	22.57053	88.37124
3	Chennai	8696010	13.08362	80.28252
4	Bangalore	8520435	12.96618	77.58690

Results

- Cluster 0: Neighbourhoods with moderate number of hotels
- Cluster 1: Neighbourhoods with low number to no existence of hotels
- Cluster 2: Neighbourhoods with high concentration of hotels
- The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.



Discussion

- Highest number in cluster 2 and moderate number in cluster
- Cluster 1 has very low number to no hotels in the neighbourhoods
- Oversupply of hotels mostly happened in the north western part of the country, with the extreme north and extreme southern regions still having very few hotels.
- This project recommends hotel builder and chains to capitalize on these findings to open new hotels in neighbourhoods in cluster 1 with little to no competition.
- Existing hotel chains are advised to avoid neighbourhoods in cluster 2 which already have high concentration of hotels and suffering from intense competition.

Limitations and Future Scope

- In this project, we only consider one factor i.e. frequency of occurrence of hotels, there are other factors such as population, tourism places, market places, and hospitals that could influence the location decision of a new hotel.
- In this research such data are not available in the foursquare api to the neighbourhood level required by this project.
- This project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain better results.

Conclusion

- **Answer to business question:** The neighbourhoods in cluster 1 are the most preferred locations to open a new hotel
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open new hotels