

**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**



## **Introduction**

Historically, the concept of hospitality is about receiving guests in a spirit of goodwill—especially strangers from other lands. Hospitality implies warmth, respect and even protection; it builds understanding and appreciation among cultures. The Latin root *hospes* is formed from *hostis*, which means “stranger” or “enemy.” Related words are *host*, *hospital*, *hostel* and *hotel*.

Today, hospitality also refers to a segment of the service industry that includes hotels, restaurants, entertainment, sporting events, cruises and other tourism-related services. As such, the hospitality industry is important not only to societies—but to economies, customers and employees.

## **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. The growth was 9.6% in 2018 and 15% in 2017, according to figures from the ministry and Reserve Bank of India.

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. Tourism and Hotel Industry always go hand in hand and the presence of enough hotels also adds value and quite a lot of factors and punches it within the region’s economy. The Existence of a Hotel isn’t enough to single-handedly boost a region’s tourism but they also give out a symptom of health 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

## Sources of Data and Extraction Methods

This page ([https://en.wikipedia.org/wiki/List\\_of\\_cities\\_in\\_India\\_by\\_population](https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population)) contains a list of most populous cities in India, with a total of 100 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

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