

Hotels in INDIA

Marina Sruthi M

Predicting where to set hotels in INDIA

- Hotels are always a bigger opportunity in india.
- It might present a great opportunity for an entrepreneur.
- Find the most suitable location for the entrepreneur to open a new hotel in india.
- By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question

Data Acquisition and Cleaning

- List of cities in India.
- Latitude and Longitude of these cities.
- Venue data related to Hotels. This will help us find the cities that are most suitable to open a hotel
- Getting Latitude and Longitude data of these cities via Geocoder package
- Using Foursquare API to get venue data related to these cities

Data Cleaning

- Remove null values
- The data had postal column which is a string of city,state, country. We split them and store it in separate columns.

Methodology

The list of cities in India is gathered as said in data section.

In order to achieve it a web scraping is done using request and beautiful soup api can prettify it and also it is easier and more convenient to pull tabular data directly from a web page into dataframe.

We can get the latitude and longitude from the geocoders api or the above web scraping provides all the details.

Once the latitude and longitude is obtained we perform data cleaning. Then we can visualize the map of India using Folium package to verify whether these are correct coordinates.

Next, we use the Foursquare API to pull the list of top 100 venues within 500 meters radius. In order to obtain account ID and API key to pull the data from foursquare a developer account is created.

From Foursquare, we will be able to pull the names, categories, latitude and longitude of the venues.

With this data, I can also check how many unique categories that I can get from these venues.

Methodology

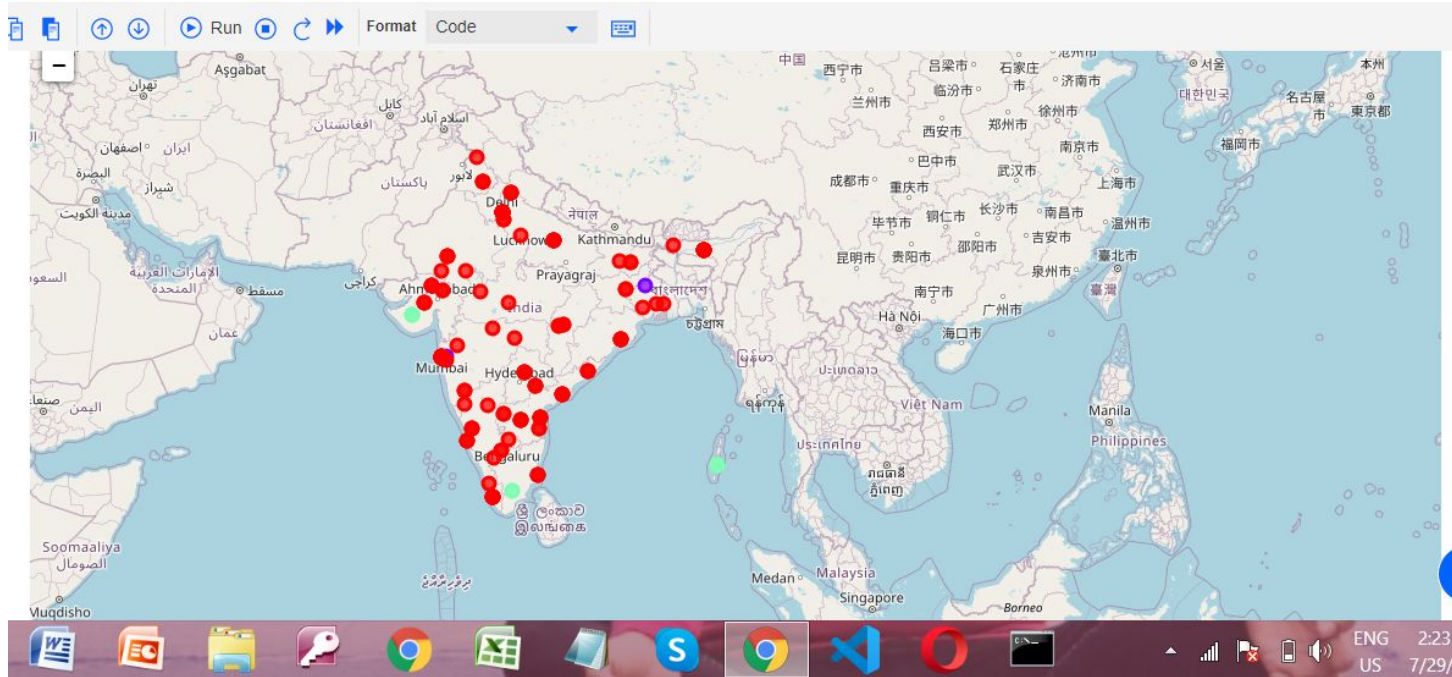
Then, we analyze each cities by grouping the rows by cities and taking the mean on the frequency of occurrence of each venue category. This is to prepare for the clustering to be done later.

Lastly, we perform the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, 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 it is highly suited for this project as well.

We have clustered the cities in india into 3 clusters based on their frequency of occurrence for “Hotel”. Based on the results (the concentration of clusters), we will be able to recommend the ideal location to open the hotel.

Visualisation



- Cluster 0: Cities with high number of Hotels
- Cluster 1: Cities with no hotels
- Cluster 2: Cities with no or less number of hotels

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.