Trip Advisor – Explore New Places

Telson Thomas
August 18, 2019

1. Introduction:

1.1Background:

XYZ is a travel agency company. The Company would like to predict/advice their customers to plan their trip to explore new places the most efficient way.

There are many places in the World which people would like to explore. One of the most important factors that needs to be considered while planning to explore a new city/place is time. We would be happier or satisfied in our trip, if we could explore more places within the available time.

Therefore, it is advantageous for the company to predict the possible ways, by which their customers could make use of the time they are spending in the most efficient way. For example, the locations that would be covered in an area, what would be the weather condition, where do they need to plan the stay etc…

1.2Problem:

Based on the area a customer would like to explore, with the help of location data of that area and the categories of interest of the customer, a suggestion needs to be provided to the customer.

This project aims to group the areas in a location as different clusters and make decisions based on it.

1.3Interest:

The XYZ Company would be able to provide helpful information for their customers and they would be able to explore the place accordingly. The customers would be highly satisfied with the service the Company is providing and this would help to increase the reputation of the company and gain more customers.

2. Data acquisition and cleaning:

2.1Data:

One city will be analysed in this project. Based on the city, the latitude and longitude values will be obtained by using geopy module.

The data used in this project is provided by Foursquare location data. The city's latitude and longitude values will be used to extract the venues data from Foursquare location data using explore endpoint.
From the obtained results which is a JSON file, we will filter the columns and create a table/dataframe.

2.2Datacleaning:

From the dataframe we can get the list of unique categories in the data and decide what all categories needs to be considered based on the interest. For example, 'Movie Theatre' might not be of interest during exploring the places. The data can be filtered based on categories.

2.3 Feature:

After cleaning the data, we would get the list of venues which would be appropriate for the project.
The features of the data should be analysed, and the unnecessary features could be removed. For our project the main feature would be the latitude and longitude values.
Using the data the venues would be grouped into different clusters. Based on the clusters the decisions can be taken accordingly.

For this project we have chosen to explore a city in India. One of the best places to visit in India is Udaipur, which is in the state of Rajasthan in India. Below is the image of the Foursquare API data of Udaipur City, in the form of dataframe after cleaning.

	name	categories	lat	lng
0	Jagmandir Island Palace Hotel	Hotel	24.567959	73.678853
1	The Oberoi Udaivilas	Resort	24.577199	73.672243
2	Lalit Laxmi Vilas Palace Hotel	Hotel	24.593846	73.682412
3	Fatehsagar Lake	Lake	24.599325	73.680491
4	Maharana Pratap Smarak	History Museum	24.598036	73.680831
5	Radisson Blu	Resort	24.589976	73.667807
6	Brewmen Café	Café	24.601438	73.687107
7	Taj Lake Palace	Resort	24.575505	73.680170
8	City Palace Museum	Museum	24.576911	73.683544
9	Trident	Hotel	24.577066	73.669200
10	Natraj Dining	Indian Restaurant	24.572356	73.699663
11	Jharokha	Restaurant	24.575729	73.680244
12	Raaj Bagh Restaurant	Indian Restaurant	24.591376	73.677984
13	Jaiwana Haveli	Hotel	24.579221	73.682553

2. Methodology:

2.1Business Understanding:

Our main goal is to suggest the best possible way to a customer who would like to explore Udaipur City.

2.2Analytic Approach:

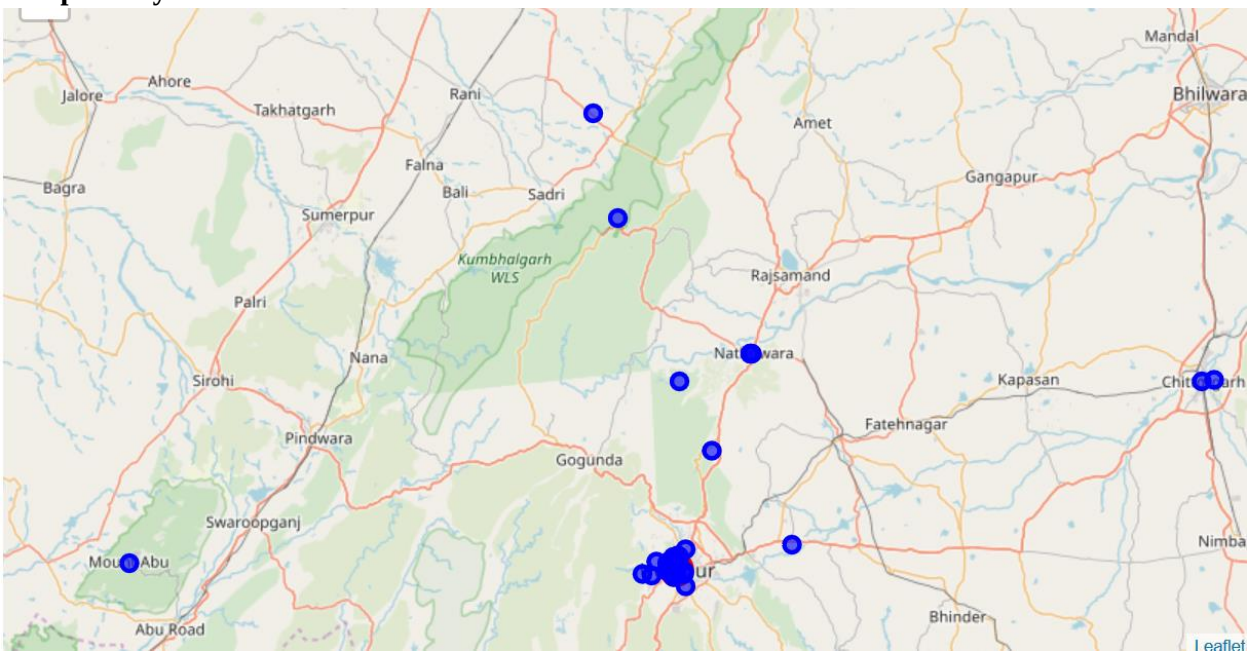
From the obtained clean data of the list of venues in Udaipur City, there are around 54 venues.

Clustering Algorithm should be used to group the data in the dataframes to different clusters and these clusters should be analysed and determined if further processing of the cluster is required to obtain the goal.

3. Exploratory Data Analysis:

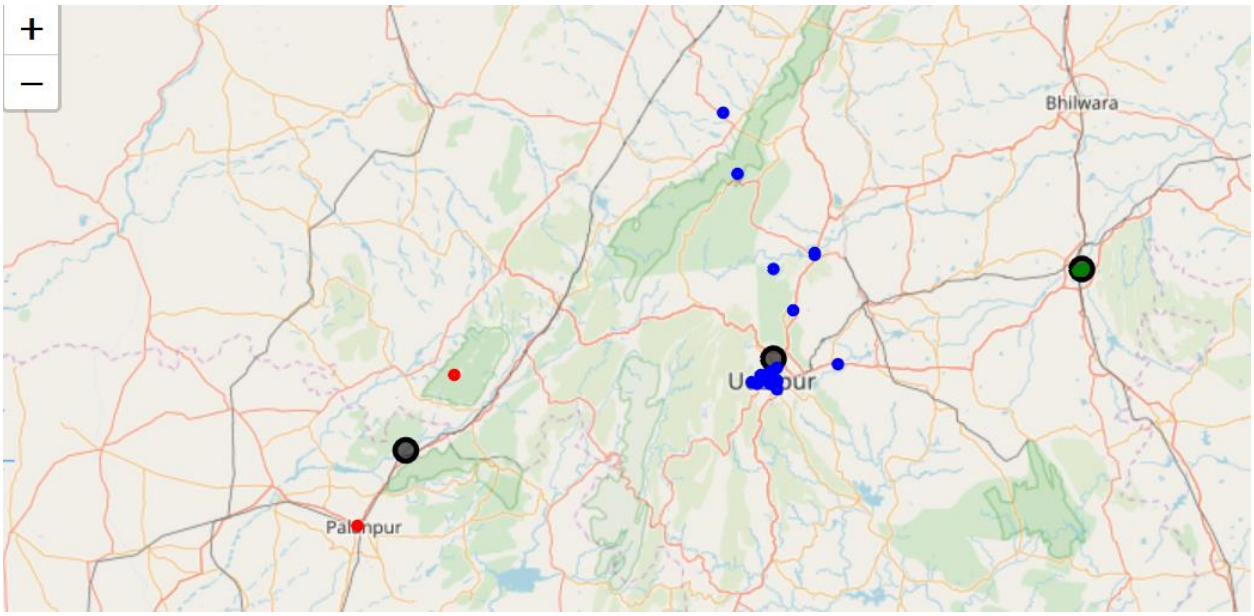
1. We use geopy and get the latitude and longitude values of the city we would like to explore.
2. The list of venues around the city can be obtained from Foursquare by using the Explore endpoint.
3. The data obtained from Foursquare should be transformed to a pandas dataframe.
4. We will use folium libraries to create map of the city with venues superimposed on top.

Udaipur City visualization



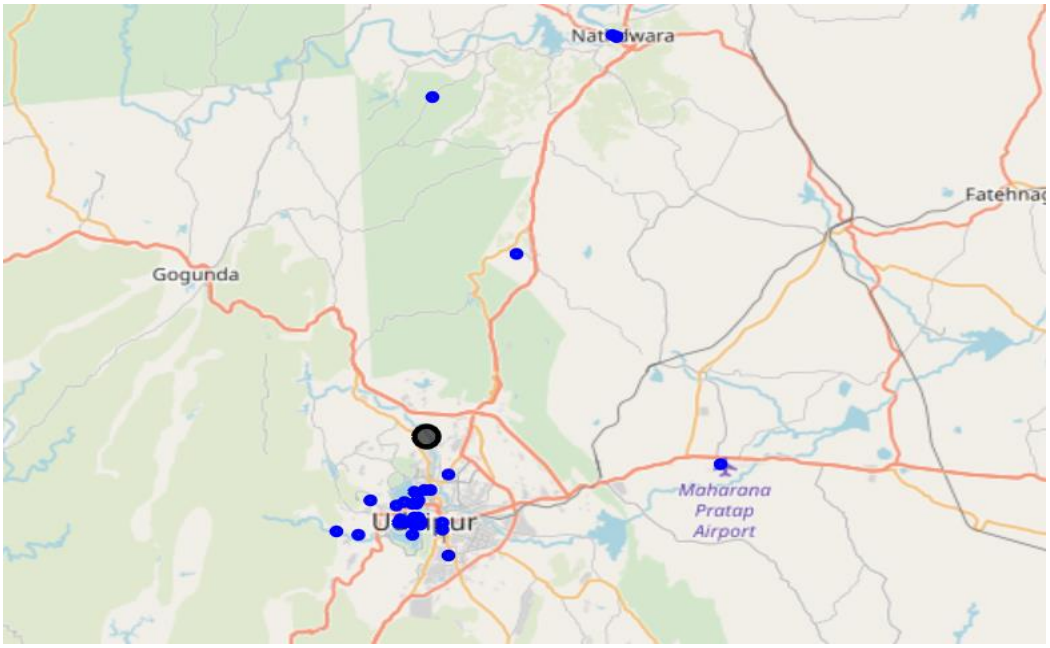
5. Apply K-Means Clustering based on the latitude and longitude values

Udaipur City visualization in clusters:



Udaipur City – Cluster0 Categories and Map:

```
array(['Hotel', 'Resort', 'Lake', 'History Museum', 'Café', 'Museum',
       'Indian Restaurant', 'Restaurant', 'Historic Site', 'Dessert Shop',
       'Hotel Bar', 'Mediterranean Restaurant', 'Rajasthani Restaurant',
       'Roof Deck', 'General Entertainment',
       'Vegetarian / Vegan Restaurant', 'Hookah Bar', 'Food', 'Castle',
       'Monument / Landmark', 'Bistro', 'Airport Lounge',
       'Fast Food Restaurant'], dtype=object)
```

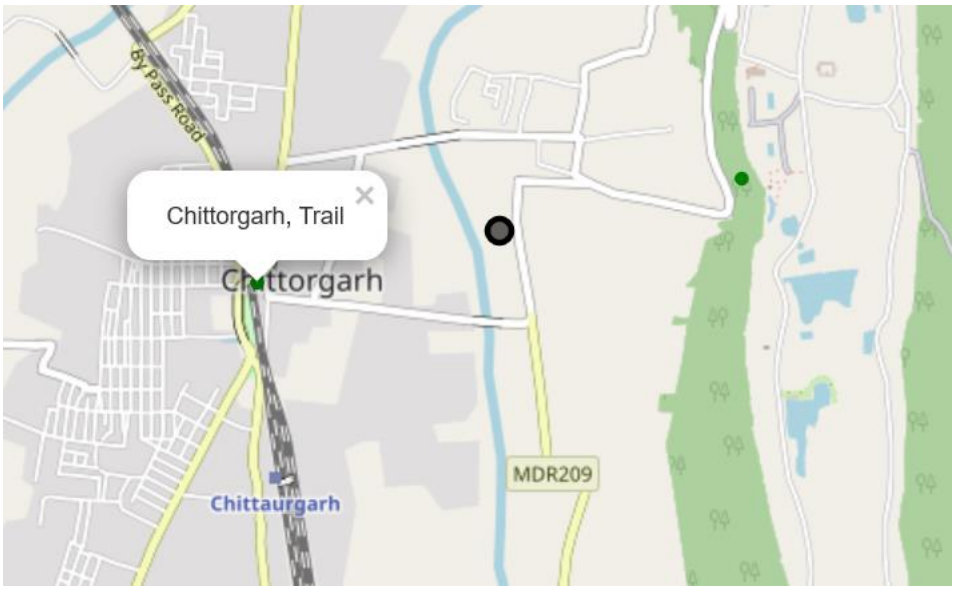


Udaipur City – Cluster1 Categories and Map:

```
array(['Indian Restaurant', 'Coffee Shop'], dtype=object)
```

Udaipur City – Cluster2 Categories and Map:

```
array(['Trail', 'Castle'], dtype=object)
```



Cluster1 covers around 50 venues, Cluster1 has 2 venues and Cluster2 has 2 venues.

4. Results:

From the map and dataframe we can understand that the maximum number of places can be explored in cluster0.
Cluster1 can be ignored as there are only two venues and they are Restaurant and Coffee Shop.
Cluster2 also has only venues but it is worth exploring the places in Cluster2 as there is a Trail and Castle.

5. Discussion:

We have obtained only two main venues in Cluster2 when the map was centred around Udaipur, let us explore Cluster2 to check if we could get more venues around it since it is worth exploring Cluster2 in detail.
Using the latitude and longitude values of the centroid of Cluster2 explore the area and get the different venues around it.

Below is the image of the Foursquare data venues around Cluster2.

	name	categories	lat	lng	labeledLatLngs	distance
0	Chittorgarh	Trail	24.883558	74.622478	[{'label': 'display', 'lat': 24.88355767698358...	1073
1	Chittorgarh Fort	Castle	24.887582	74.643277	[{'label': 'display', 'lat': 24.88758224620540...	1073
2	Vijay Stambh	Castle	24.887778	74.646342	[{'label': 'display', 'lat': 24.88777764298154...	1381
3	Chittorgarh Railway Station	Train Station	24.875425	74.623451	[{'label': 'display', 'lat': 24.87542529002411...	1477
4	Victory Tower	History Museum	24.874298	74.623502	[{'label': 'display', 'lat': 24.87429806175967...	1571

From the above observations of Cluster2, two extra spots (Castle and Museum) has been identified.
Also, as per the venue list there are no Hotels around Cluster2. But there is a Train Station which is an useful information.

6. Conclusion:

In this study, I analysed that to explore Udaipur it would be better to allocate more time to explore the places in Cluster0 and to prefer stay in one of the Hotels which is closer to the centroid of Cluster0 and then probably to explore the places in Cluster2. Rail Transport can be preferred to explore Cluster2 and the places in Cluster2 can be covered in a single day as the distance to different venues are within 3km. The western side of Udaipur can be given the least importance as there are not much interesting places to explore.
This project is performed on limited data. If a good amount of data is available which has the interesting and useful venues, weather condition of the areas etc., the decisions could be made in a much better way.