The Battle of Neighbourhoods

Introduction

People often move from country to country or between cities for work meetings, where they don't know much about the neighbourhoods of that particular city. So, they may struggle to find the perfect hotel near their work place. Hence this problem is solved by gathering the data of city and make the clusters of all the hotels. So that when user will put the location of his work place he will get recommendation of all the hotels near that location which are present in the same cluster.

Data

We will use Geopy to get the Latitude and Longitude values of the cities and user workplace. Then to get the data about hotels and its name and distance all the other things we will be using Foursquare. We will request foursquare for hotel data which will be returned in Json file. Then information will be extracted from that json file to DataFrame which will be further be used in clustering process.

Above map show us the data Hotel data of New York city obtained from Foursquare. above data is obtained in the form of json file. json file is shown below

```
In [112]: data

Out[112]: {'meta': {'code': 200, 'requestId': '5d3351ac787dba0038452462'},
           'response': {'venues': [{'id': '49efcc88f964a52006691fe3',
             'name': 'Smyth Hotel',
             'contact': {},
             'location': {'address': '85 W Broadway',
              'crossStreet': 'Chambers St',
              'lat': 40.7151439,
              'lng': -74.0091826,
              'labeledLatLngs': [{'label': 'display',
                'lat': 40.7151439,
                'lng': -74.0091826}],
              'distance': 379,
              'postalCode': '10007',
              'cc': 'US',
              'city': 'New York',
              'state': 'NY',
              'country': 'United States',
              'formattedAddress': ['85 W Broadway (Chambers St)',
               'New York, NY 10007',
               'United States']}
```

Now to process this data we have to convert this json file into more structured form. So we converted it to DataFrame as shown below

```
df.head()
```

	Name	Distance	Lat	Lng	City	Address
0	Smyth Hotel	379	40.715144	-74.009183	New York	85 W Broadway
1	The Roxy Hotel	740	40.719341	-74.005044	New York	2 Avenue of the Americas
2	Four Seasons Hotel New York Downtown	284	40.712612	-74.009380	New York	27 Barclay St
3	Soho Grand Hotel	1036	40.721942	-74.004217	New York	310 W Broadway
4	Hotel 50 Bowery NYC	856	40.715936	-73.996789	New York	50 Bowery

Above is the DataFrame obtained from json file. it is cleaned and columns are named accordingly for further processing.

Methodology

First Lat and Lng values are stored in X. which will be used in clustering. Now the important thing about clustering is we have to guess the number of clusters in the given dataset. For that we will be using a method named elbow method which will help us get the amount of cluster to be used in the given dataset.
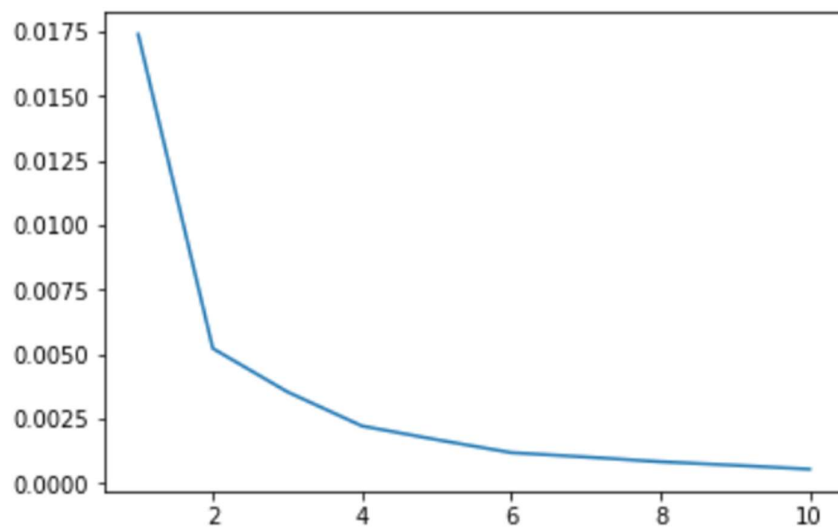
above is the graph obtained now we can see that at value 4 is perfect number. So we will initiate 4 centroids to **create 4 clusters** in our data.

Now we will import K-Means library from sklearn which a is machine learning library. The we will create a object of KMean with cluster value of 4. And with **init=k-means++**

Then we will get cluster values which we will add into our previously created DataFrame. Now this DataFrame will be used to make the recommendations to users based on the location user enters.

Results

User will be asked to put the location of his workplace where he is going to work. Using Geopy we will get the coordinates for entered address and then those coordinates will be provided to our k-mean model. Then k-means model will predict the cluster for that specific location and all the values of that cluster will presented to user to decide which hotel he want to get into.

Enter the neighbourhood name where you want to stay = Chinatown,New York

List of hotels will be recommended to user based on entered location: -

Recommended Hotels Name

```
In [114]: df.loc[df['Cluster'] == ans[0]]
```

Out[114]:

	Name	Distance	Lat	Lng	City	Address	Cluster
4	Hotel 50 Bowery NYC	856	40.715936	-73.996789	New York	50 Bowery	2
5	Hotel Indigo	1816	40.721762	-73.988092	New York	171 Ludlow St	2
8	The Leon Hotel	984	40.715994	-73.995171	New York	125 Canal St	2
10	Mercer Hotel	1486	40.724828	-73.998553	New York	147 Mercer St	2
14	SIXTY SoHo Hotel	1276	40.723990	-74.003160	New York	60 Thompson St	2
15	Hotel Chantelle	1567	40.718483	-73.989056	New York	92 Ludlow St	2
17	The Bowery Hotel	1924	40.726145	-73.991627	New York	335 Bowery	2
19	SIXTY LES Hotel	1807	40.722090	-73.988510	New York	190 Allen St	2
20	The Ludlow Hotel	1884	40.721857	-73.987204	New York	188 Ludlow St	2
25	Best Western Bowery Hanbee Hotel	1112	40.718446	-73.995198	New York	231 Grand St	2
30	Crosby Street Hotel	1353	40.723011	-73.997454	New York	79 Crosby St	2
35	The Nolitan Hotel	1254	40.720577	-73.995339	New York	30 Kenmare St	2

As shown in above image all the names of the hotels in **cluster 2** are close to **Chinatown, New York**. Hence these names are presented to user.

Discussion section

Accuracy of the above model is good but it is limited to New York city. We will need to create model for every city so it will be used globally. This same model I think can be used to predict the location for anyone who is interested in opening Hotel in this city by going through all the clusters and showing user the minimum and maximum number of hotels present in corresponding cluster or that location. So, the owner can use this information to make wise decision to choose location for his hotel.

Conclusion section

It was a nice learning experience. This was a very simple and interesting project. I learned how to use libraries and how data can used to for multiple things. as above model itself can be used to predict the locations of hotels for workers who are temporary coming to unknows city for working and also for recommending people who are looking to open hotel. I conclude this project by saying I learned and enjoyed while learning a lot and a lot more learning and fun is yet to come.