Project – Exploring Delhi

1. Introduction: -

1.1 Background:

Delhi, India's capital territory, is a massive metropolitan area in the country's north. It is the largest commercial center in northern India. As of 2016, recent estimates of the economy of the Delhi urban area have ranged from $167 to $370 billion (PPP metro GDP), ranking it either the most or second-most productive metro area of India. The nominal GSDP of the NCT of Delhi for 2016–17 was estimated at ₹6,224 billion (US$90 billion), 13% higher than in 2015–16. The NCT covers an area of 1,484 square kilometers (573 sq. mi). According to the 2011 census, Delhi's city proper population was over 11 million, the second-highest in India after Mumbai, while the whole NCT's population was about 16.8 million. A large number of people from all over the world come to Delhi for different kinds of purposes such as study, career, business, etc. Hence this city hosts all kinds of people, be it unskilled or skilled like Software professionals, Government officials, Diplomats, Administrators, Students, and so on. Through this project, I will explore this beautiful city so that the newcomers can decide which area/neighborhood they would want to stay in or figure out the best places to eat/shop.

1.2 Problem:

Finding the best place for living is very difficult in the area like Delhi because of such a large population and variety. Everyone's lifestyle is different from each other and their preference of locality, as well as food, might also different so user wants to explore the places where they can live their lifestyle comfortably, and food is also available of their preferences. A lot of time is going to use for this purpose, and you have to travel a lot for it. As time is very precious in today's busy world, so finding the best place to stay is very close to impossible, i.e., you have to compromise with your lifestyle, but through this project, this big problem is easily solved.

2. Data Section:

- I made use of a python library called ***pgeocode,*** which is high-performance off-line querying of GPS coordinates, region name, and municipality name from postal codes. Distances between postal codes as well as general distance queries are also supported. The used GeoNames database includes postal codes for 83 countries.

  I used ' *index* postal_codes' function which creates a data frame of unique postal codes of a given country. The data frame consists of following columns:
  - country code: iso country code, 2 characters
  - postal code: postal code
  - place name: place name (e.g. town, city etc)

- state_name: 1. order subdivision (state)
- state_code: 1. order subdivision (state)
- county_name: 2. order subdivision (county/province)
- county_code: 2. order subdivision (county/province)
- community_name: 3. order subdivision (community)
- community code: 3. order subdivision (community)
- latitude: estimated latitude (wgs84)
- longitude: estimated longitude (wgs84)
- accuracy: accuracy of lat/lng from 1=estimated to 6=centroid

Using this function, I first obtained the postal code details for the entire country of India and then queried the data to retain the postal codes of Delhi.

- I used **Foursquare API** to get the most common venues of the given community of Bengaluru.

3. Target Audience: -

The target audience of this project is Software professionals, Government officials, Diplomats, Administrators, Students, and so on who come to Delhi for different kinds of purposes such as study, career, business, etc.

4. Methodology: -

- We use pgeocode package to create a dataset that contains the main places of Delhi where a client wants to live, to explore them, and find the nearby shops, restaurants, etc. for their comfortability as shown below:

```
In [8]: #print 5 rows
        Delhi.head()
```

Out[8]:

	country code	postal_code	place_name	state_name	state_code	county_name	county_code	community_name	community
0	IN	110001	New Delhi G.P.O., Parliament House, Connaught ...	Delhi	7	New Delhi	94.0	New Delhi	NaN
1	IN	110002	Civic Centre, Darya Ganj, A.G.C.R., I.P.Estate...	Delhi	7	New Delhi	94.0	New Delhi Central	NaN
2	IN	110003	Pandara Road, Delhi High Court, Delhi High Cou...	Delhi	7	Central Delhi	95.0	New Delhi	NaN
3	IN	110004	Rashtrapati Bhawan	Delhi	7	Central Delhi	95.0	New Delhi	NaN
4	IN	110005	Bank Street (Central Delhi), Karol Bagh, Anand...	Delhi	7	Central Delhi	95.0	New Delhi	NaN

- But this data frame has several columns which are of no use so we will drop them and the new data frame is:

```
In [12]: Delhi = Delhi[['postal_code','place_name', 'community_name', 'latitude', 'longitude']]
```

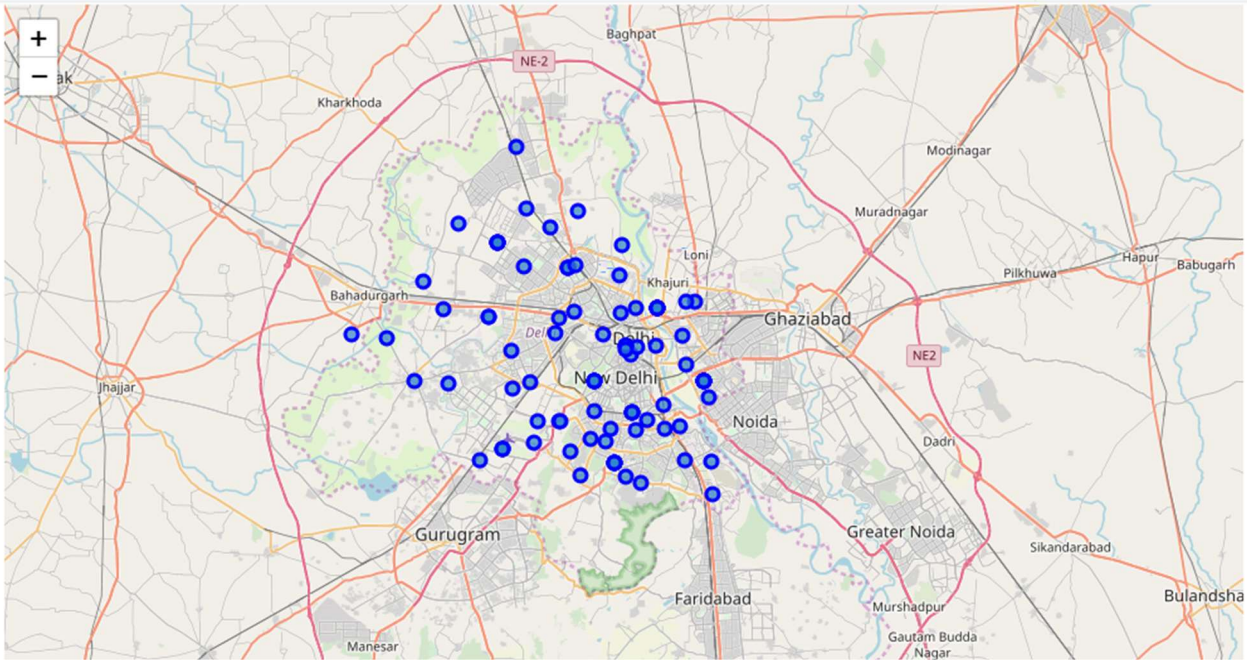
```
In [13]: Delhi.head(10)
```

Out[13]:

	postal_code	place_name	community_name	latitude	longitude
0	110001	New Delhi G.P.O., Parliament House, Connaught ...	New Delhi	28.6369	77.218229
1	110002	Civic Centre, Darya Ganj, A.G.C.R., I.P.Estate...	New Delhi Central	28.6453	77.245600
2	110003	Pandara Road, Delhi High Court, Delhi High Cou...	New Delhi	28.6431	77.225270
3	110004	Rashtrapati Bhawan	New Delhi	28.6453	77.212800
4	110005	Bank Street (Central Delhi), Karol Bagh, Anand...	New Delhi	28.6551	77.188775
5	110006	Baratooti, Dareeba, Delhi Sadar Bazar, S.T. Ro...	Delhi	28.6453	77.212800
6	110007	Delhi University, Malka Ganj, Rana Pratap Bagh...	Delhi	28.6764	77.207400
7	110008	Dada Ghosh Bhawan, Patel Nagar West, Patel Nag...	New Delhi	28.6453	77.212800
8	110009	Nirankari Colony, Dr.Mukerjee Nagar, Model Tow...	Delhi	28.7121	77.206000
9	110010	Dhaula Kuan, Subroto Park, R R Hospital, Kirby...	NaN	28.5724	77.139850

- Now create a map of Delhi using folium library and plot each postal code or place of the above data frame in that map for understanding the area of Delhi.

Out[17]:



- Check the community where a maximum no of locations exists because of it higher the probability of living in that place.

In [18]:
```
#group and count the areas according to their community name
Delhi.groupby("community_name").count()
```

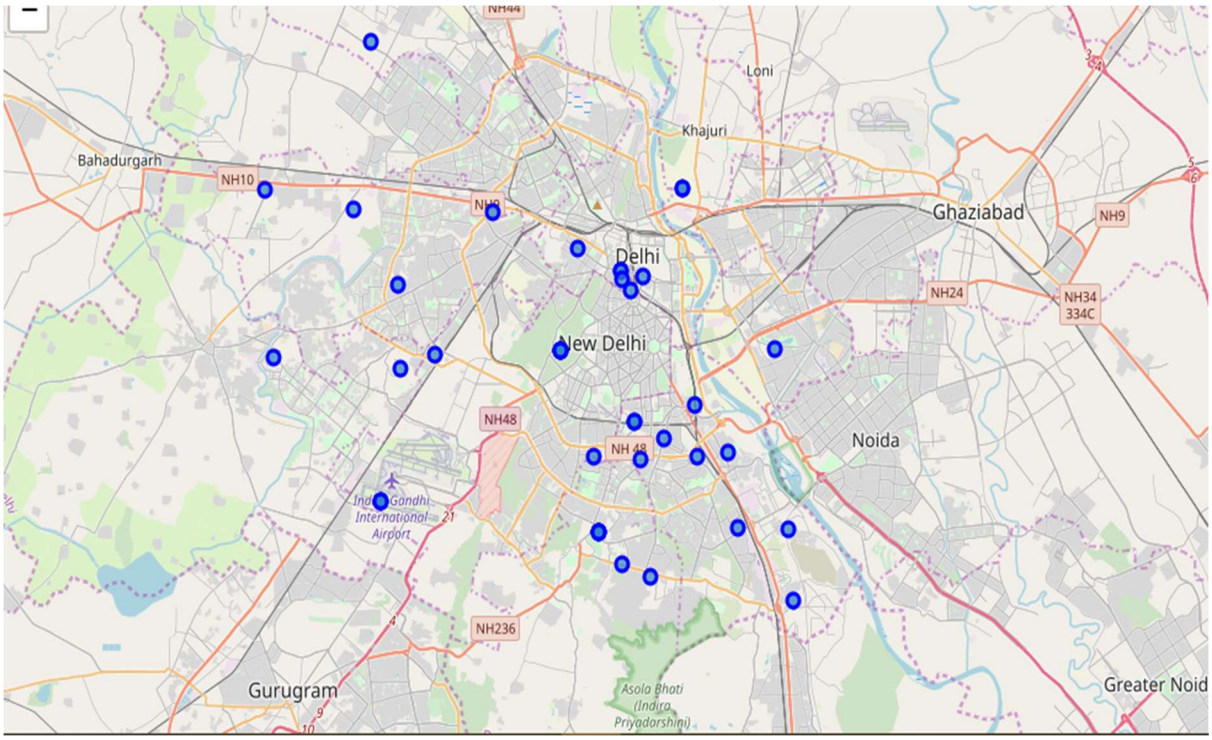
Out[18]:

	postal_code	place_name	latitude	longitude
community_name				
Delhi	25	25	25	25
Delhi North East	1	1	1	1
East Delhi	1	1	1	1
New Delhi	42	42	42	42
New Delhi Central	1	1	1	1
South Delhi	2	2	2	2
South West Delhi	1	1	1	1

As New Delhi has highest count, will be using for further clustering

In [19]:
```
Delhi_new = Delhi[Delhi['community_name'] == 'New Delhi'].reset_index(drop=True)
```

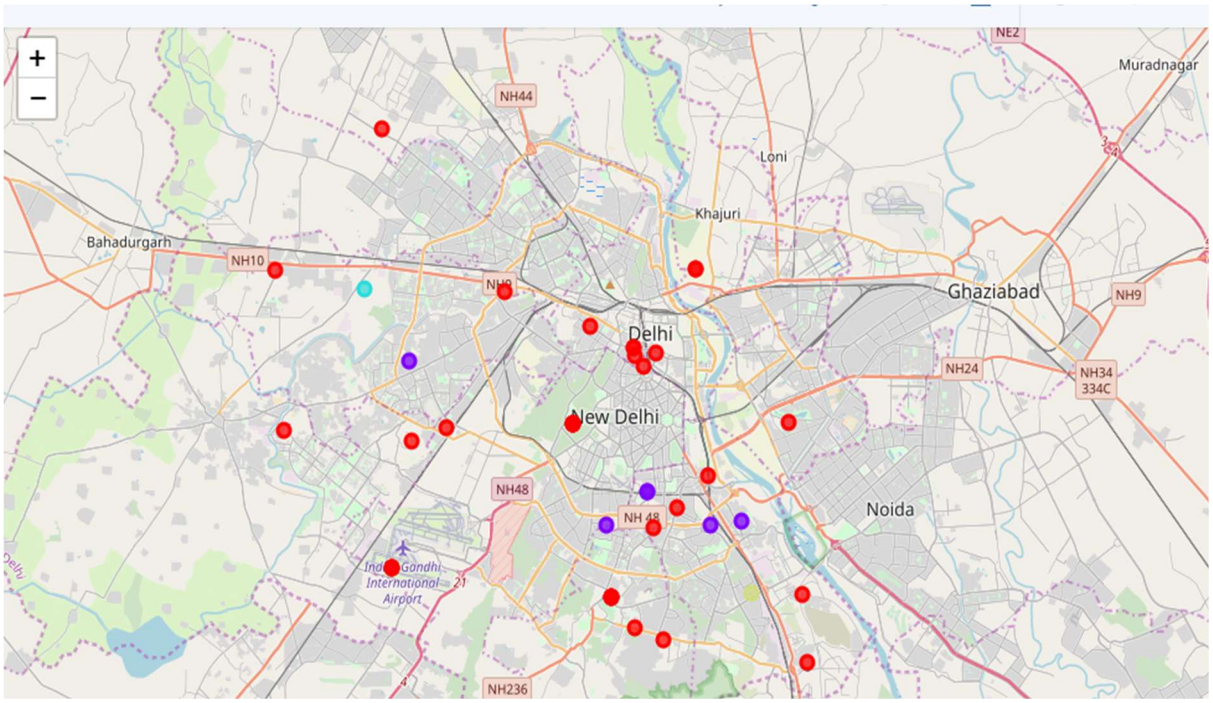
- As shown, the maximum no of postal areas are in New Delhi location, so we explore this particular community.



- Let us use Foursquare API now and get the venuses name and their type in all locations which results in a data frame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Sagar Ratna	28.635487	77.220650	Indian Restaurant
1	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Nizam's Kathi Kabab | निज़ाम काठी कबाब	28.634858	77.219462	Indian Restaurant
2	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Wenger's	28.633412	77.218292	Bakery
3	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Route 04	28.634890	77.220225	Bar
4	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Veda Restaurant	28.635210	77.218079	Indian Restaurant
5	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Naturals Ice Cream	28.634455	77.222139	Ice Cream Shop
6	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Khan Chacha | खान चाचा | خان چاچا	28.634202	77.220780	Indian Restaurant
7	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Odeon Social	28.634414	77.220936	Café
8	New Delhi G.P.O., Parliament House, Connaught ...	28.6369	77.218229	Connaught Place | कनॉट प्लेस (Connaught Place)	28.632731	77.220018	Plaza
9	New Delhi G.P.O., Parliament	28.6369	77.218229	Immigrants Project - A Cafe in	28.634055	77.218867	Café

- After that, we make clusters using K- means clustering and plot them in a new map with different colors as shown below:



- Each group has something different from each other, and we will check the cluster according to our requirements for living a comfortable life.

5. Result: -

We get three clusters from applying machine learning on the above data frame, and each cluster has venues according to their commonness.

- **Cluster-1: -**

	place_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mc Comm Ven
0	New Delhi G.P.O., Parliament House, Connaught ...	0	Indian Restaurant	Café	Chinese Restaurant	Hotel	Lounge	Deli / Bodega	Bar	Pub	Dessert Shop	Multiplex
1	Pandara Road, Delhi High Court, Delhi High Cou...	0	Train Station	Fast Food Restaurant	Metro Station	Food & Drink Shop	Gym / Fitness Center	Gym	Food Truck	Food Court	Ice Cream Shop	Food
2	Rashtrapati Bhawan	0	Hotel	Restaurant	Hostel	Pizza Place	Café	Indian Restaurant	Korean Restaurant	Food	Fast Food Restaurant	Motel
3	Bank Street (Central Delhi), Karol Bagh, Anand...	0	Ice Cream Shop	Hotel	Pharmacy	Snack Place	Smoke Shop	Food	Dessert Shop	Diner	Donut Shop	Falafel Restaura
4	Dada Ghosh Bhawan, Patel Nagar West, Patel Nag...	0	Hotel	Restaurant	Hostel	Pizza Place	Café	Indian Restaurant	Korean Restaurant	Food	Fast Food Restaurant	Motel
5	South Avenue, Nirman Bhawan,	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

- **Cluster- 2: -**

	place_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Tilak Nagar (West Delhi), Chaukhandi, Ashok Na...	1	Indian Restaurant	Print Shop	Train Station	Food	Dessert Shop	Diner	Donut Shop	Falafel Restaurant	Fast Food Restaurant	Food & Drink Shop
15	New Friends Colony, Zakir Nagar, Abul Fazal En...	1	Indian Restaurant	Office	Train Station	Food	Dessert Shop	Diner	Donut Shop	Falafel Restaurant	Fast Food Restaurant	Food Court
17	Tagore Garden, Subhash Nagar West, Subhash Nag...	1	Indian Restaurant	Bistro	Athletics & Sports	Modern European Restaurant	Deli / Bodega	Tea Room	History Museum	Gym / Fitness Center	Gym	Food Truck
18	Naraina Industrial Estate H.O, Naraina Village	1	Indian Restaurant	Bistro	Athletics & Sports	Modern European Restaurant	Deli / Bodega	Tea Room	History Museum	Gym / Fitness Center	Gym	Food Truck
19	Nauroji Nagar, Safdarjung Enclave, Ansari Nagar	1	Indian Restaurant	Ice Cream Shop	Nightclub	Department Store	Food	Dessert Shop	Diner	Donut Shop	Falafel Restaurant	Fast Food Restaurant
33	Sriniwaspuri, Sant Nagar (South Delhi), East O...	1	Wine Bar	Indian Restaurant	Pizza Place	Food	Dessert Shop	Diner	Donut Shop	Falafel Restaurant	Fast Food Restaurant	Food & Drink Shop

- **Cluster- 3: -**

Cluster 3 list shown below:

In [37]:
```
Delhi_new_merged.loc[Delhi_new_merged['Cluster Labels'] == 2, Delhi_new_merged.columns[[1] + list(range(5, Delhi_new_merged.shape[1]))]]
```

Out[37]:

	place_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
25	Andrewsganj, Gautam Nagar, Gulmohar Park, Sadi...	2	ATM	Train Station	Food Court	Diner	Donut Shop	Falafel Restaurant	Fast Food Restaurant	Food	Food & Drink Shop	Food Truck

As each cluster is shown above and now, a user has to decide which cluster is best according to common venues in them and after he/she will have to choose the place from that cluster so that he/she can live comfortably.

6. Discussion and Future Scope: -

As already said and we also know that Delhi is a very important city in India, and a huge number of people come here for their purpose. In this project, I use k means clustering algorithm which is most famous one and all the results based on this algorithm. Also, I use the data from pgeocode package which give offline dataset that maybe not entirely accurate. So someone can use the datasets from other places and use different machine learning algorithm with better optimization, this may enhance the results.

I ended the study by visualizing the data and clustering information on the Delhi map. In future studies, the web application can be carried out to direct users.

7. Conclusion: -

Not only users who come to Delhi for his/her purpose and wants to live there, but this project also helps the investors as well as the organisation that manages the city to check the variety of resources in this city and where they should mainly focus on maintaining the balance.

References:

1. **Delhi,** (https://en.wikipedia.org/wiki/Delhi)
2. ***pgeocode*** **(https://github.com/symerio/pgeocode)**
3. **Forsquare API https://developer.foursquare.com/**