Data requirements :

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyze such large dataset and produce analytics to find a solutions.

Let's consider the base scenario :

Suppose i want to find a restaurant, then logically, i need 3 things :

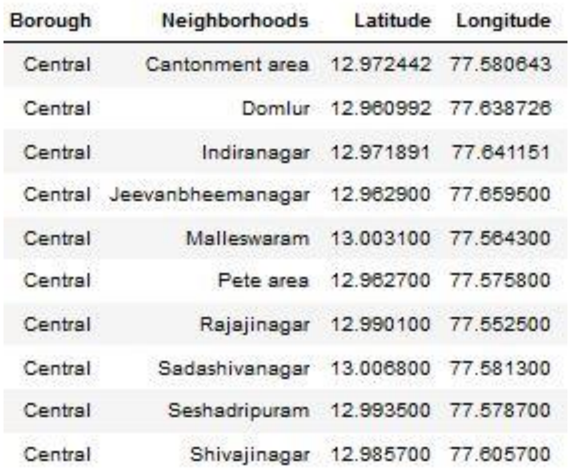
1. Its geographical coordinates(latitude and longitude) to find our where exactly it is located.
2. Population of the neighborhood where the restaurant is located.
3. Average income of neighborhood to know how much is the restaurant worth.

Let’s take a closer look at each of these :

1. To access location of a restaurant, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively.
2. Population of a neighborhood is very important factor in determining a restaurant's growth and amount of customers who turn up to eat. Logically, the more the population of a neighborhood, the more people will be interested to walk openly into a restaurant and less the population, less number of people frequently visit a restaurant. Also if more people visit, better the restaurant is rated because it is accessed by different people with different taste. Hence is is very important factor.
3. Income of a neighborhood is also very important factor as population was. Income is directly proportional to richness of a neighborhood. If people in a neighborhood earns more than an average income, then it is very much possible that they will spend more however not always true with very less probability. So an restaurant assessment is proportional to income of a neighborhood.

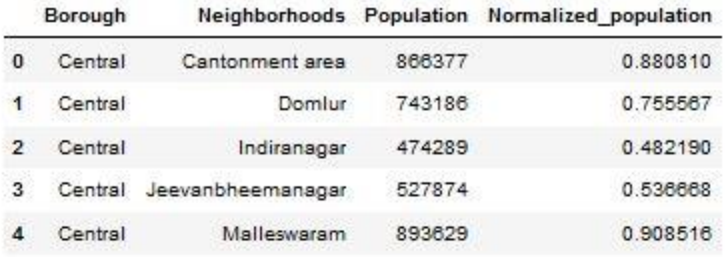
Data collection :

1. Collecting geographical coordinates is not difficult but after searching on Google for more than 2 days, it was not available on open source data websites such as Wikipedia, India gov website, census report websites etc. So i decided to use Google maps API to fetch latitude and longitude but Google API has limited number of calls that i could make with my free account. So it would take around 15 - 20 days to fetch location of all the neighborhoods in Bangalore. Initially i scrapped list of neighbor's using beautifulSoup4 from Wikipedia. The table headings becoming the boroughs and data becoming the neighborhoods. Bangalore has 8 boroughs and 64 neighborhoods. So i manually googled each neighborhood to find its corresponding latitude and longitude. After doing so, i produced the following dataframe.



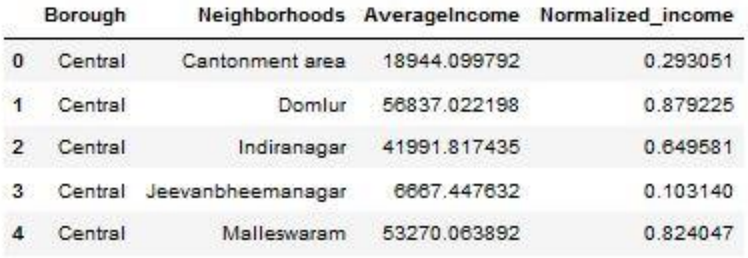
Borough	Neighborhoods	Latitude	Longitude
Central	Cantonment area	12.972442	77.580643
Central	Domlur	12.960992	77.638726
Central	Indiranagar	12.971891	77.641151
Central	Jeevanbheemanagar	12.962900	77.659500
Central	Malleswaram	13.003100	77.564300
Central	Pete area	12.962700	77.575800
Central	Rajajinagar	12.990100	77.552500
Central	Sadashivanagar	13.006800	77.581300
Central	Seshadripuram	12.993500	77.578700
Central	Shivajinagar	12.985700	77.605700

- Population by neighborhood is again easy to find out given that its readily available. But in case of Bangalore, it is again not the case. i was able to find population data for few cities. Here is the link. Rest other neighborhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The dataframe for Bangalore neighborhood population looks like :



	Borough	Neighborhoods	Population	Normalized_population
0	Central	Cantonment area	866377	0.880810
1	Central	Domlur	743186	0.755567
2	Central	Indiranagar	474289	0.482190
3	Central	Jeevanbheemanagar	527874	0.536668
4	Central	Malleswaram	893629	0.908516

- Income by neighborhood is again easy to find out given that its readily available. But incase of bangalore, it is again not the case. i was able to find Income data for main city. Here is the link. Neighborhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The dataframe for Bangalore neighborhood population looks like :

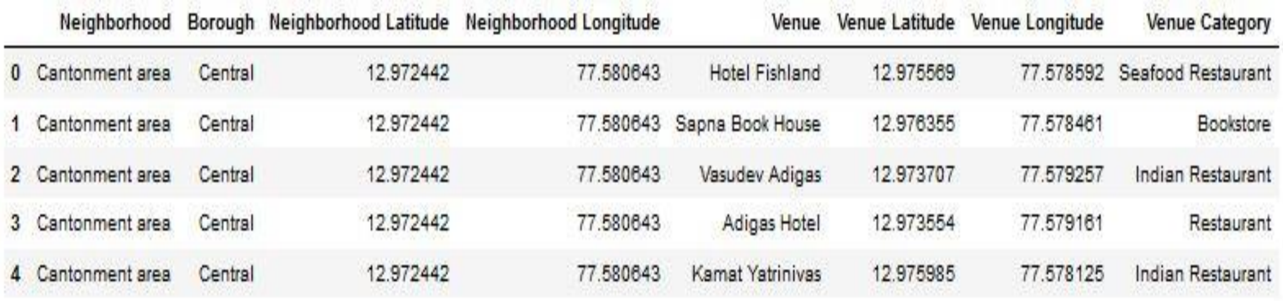


	Borough	Neighborhoods	AverageIncome	Normalized_income
0	Central	Cantonment area	18944.099792	0.293051
1	Central	Domlur	56837.022198	0.879225
2	Central	Indiranagar	41991.817435	0.649581
3	Central	Jeevanbheemanagar	6667.447632	0.103140
4	Central	Malleswaram	53270.063892	0.824047

- FourSquare API :

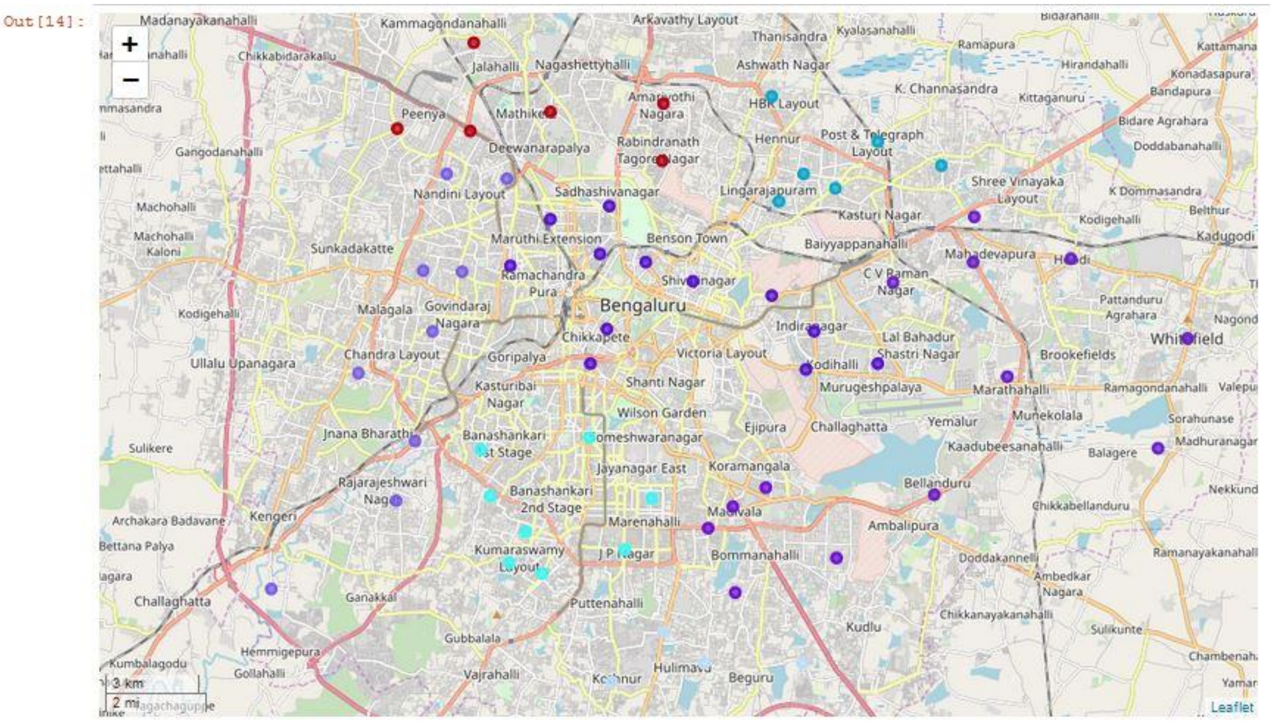
Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius(in my case : 500mts) and also corresponding coordinates, venue location and names. After calling, the following dataframe is created

:



	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cantonment area	Central	12.972442	77.580643	Hotel Fishland	12.975569	77.578592	Seafood Restaurant
1	Cantonment area	Central	12.972442	77.580643	Sapna Book House	12.976355	77.578461	Bookstore
2	Cantonment area	Central	12.972442	77.580643	Vasudev Adigas	12.973707	77.579257	Indian Restaurant
3	Cantonment area	Central	12.972442	77.580643	Adigas Hotel	12.973554	77.579161	Restaurant
4	Cantonment area	Central	12.972442	77.580643	Kamat Yatrinivas	12.975985	77.578125	Indian Restaurant

The following map is produced by marking all the neighborhoods in Bangalore city.


Out[14]:
+
−
Madanayakanahalli
Kammagondanahalli
Arkavathy Layout
Thanisandra
Kyalasanahalli
Ramapura
Kattamana
Chikkabidarakallu
Jalahalli
Nagashettyhalli
Ashwath Nagar
Hirandahalli
Konadasapura
K. Channasandra
Bandapura
Kittaganuru
Amaravothi Nagara
HBR Layout
Peenya
Mathikere
Bidare Agrahara
Gangodanahalli
Hennur
Post & Telegraph Layout
Doddabanahalli
Deewanarapalya
Rabindranath Tagore Nagar
Shree Vinayaka Layout
Nandini Layout
Sadhashivanagar
Lingarajapuram
K Dommasandra
Machohalli
Kasturi Nagar
Belthur
Kodigehalli
Machohalli Kaloni
Maruthi Extension
Benson Town
Kadugodi
Sunkadakatte
Baiyyappanahalli
Mahadevapura
Hoodi
Ramachandra Pura
Shivajinagar
C V Raman Nagar
Pattanduru Agrahara
Malagala
Govindaraj Nagara
Bengaluru
Nagond
Kodigehalli
Indiranagar
Whitefield
Lal Bahadur Shastri Nagar
Chikkapete
Chandra Layout
Goripalya
Victoria Layout
Brookefields
Ullalu Upanagara
Jodihalli
Kasturibai Nagar
Shanti Nagar
Murugeshpalaya
Marathahalli
Ramagondanahalli
Valepu
Wilson Garden
Munekolala
Sorahunase
Ejipura
Challaghatta
Yemalur
Jnana Bharathi
Banashankari 1st Stage
Someshwaranagar
Kaadubeesanahalli
Madhuranagar
Sulikere
Balagere
Jayanagar East
Koramangala
Rajarajeshwari Nagar
Banashankari 2nd Stage
Bellanduru
Nekkund
Madivala
Chikkabellanduru
Archakara Badavane
Kengeri
Marenahalli
Ambalipura
Bettana Palya
Kumaraswamy Layout
J P Nagar
Bommanahalli
Doddakannelli
Ramanayakanahalli
Ambedkar Nagara
Challaghatta
Ganakkal
Puttenahalli
Chikkanayakanahalli
Kudlu
Sulikunte
Gubbalala
Kumbalagodu
Hemmigepura
Chambenah
Gollahalli
Vajrahalli
Hulimavu
Kothnur
Beguru
Yamar
3 km
2 mi
Leaflet