Objectives

Objectives of the final assignments were to define a business problem, look for data in the web and, use Foursquare location data to compare different locations to figure out which neighborhood is suitable for starting a restaurant business. As prepared for the assignment, I go through the problem designing, data preparation and final analysis section step by step.

1. Introduction¶

1.1. Problem Statement

Chennai also known as Madras is the capital of the Indian state of Tamil Nadu. Located on the Coromandel Coast off the Bay of Bengal, it is the biggest cultural, economic and educational center of south India. According to the 2011 Indian census, it is the sixth-most populous city and fourth-most populous urban agglomeration in India. The city together with the adjoining regions constitutes the Chennai Metropolitan Area, which is the 36th-largest urban area by population in the world.

Chennai as the only South Asian city to feature in its 2015 "Top 10 food cities" list. Chennai was also named the ninth-best cosmopolitan city in the world by Lonely planet.

The south Indian city of Chennai is fast emerging as a destination for information technology outsourcing and has seen a growing number of IT parks being built here. Most of the upcoming complexes are being built along the **IT Corridor**.

- To find a suitable Area for starting a Restaurent near the IT HUBS of Chennai.
- This analysis will help to find suitable place for starting a new restaurent as the need for food is more among the office during lunch hours.
- This will help the new entrepreneurs who are in need of information regarding business ideas.

1.2. Import Necessary Libraries

We used following packages in this assignments. **Geospy, BeautifulSoup, Foursquare, Matplotlib and Folium**.

- **BeautifulSoup:** Beautiful Soup is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.
- **GeoSpy:** Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.
- **Foursquare:** The Foursquare is to help you discover and share information about businesses and attractions around you.
- **Folium:** Folium makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map.
- **Matplotlib:**Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc.,

2. Data Preparation

The south Indian city of *Chennai* is fast emerging as a destination for information technology outsourcing and has seen a growing number of IT parks being built here. List of IT parks details are getting from https://en.wikipedia.org/wiki/List_of_tech_parks_in_Chennai.

The BeautifulSoup is used for parsing and extract data from HTML. The result is uploaded into GitHub for later reference.

```
response = requests.get('https://en.wikipedia.org/wiki/List_of_tech_parks_in_Chennai')

soup = BeautifulSoup(response.text, 'lxml')

table = soup.find('table', {'class':'wikitable sortable'})
table_rows = table.find_all('tr')

chennai_list = []

for tr in table_rows:
    td = tr.find_all('td')
    row = [i.text for i in td]
    chennai_list.append(row)

chennai_df = pd.DataFrame(chennai_list)

chennai_df.columns = ['CompanyName', 'Area', 'Location', 'Investment', 'StartYear', 'CompaniesList']

chennai_df.dropna(axis=0, inplace=True)                              .

chennai_df = chennai_df.drop(chennai_df.columns[[1, 3, 4]], axis=1)

for column in chennai_df.columns:
    chennai_df[column] = chennai_df[column].str.strip()

chennai_df = chennai_df[pd.notnull(chennai_df['Location'])]
```

	CompanyName	Location	CompaniesList
0	Tidel Park	Taramani	Cisco Systems, Hexaware Technologies, Sify, Te...
1	Olympia Tech Park	Guindy	HP, Verizon, ABN Amro, Visteon, Hewitt, Merril...
2	SRM Tech Park	Potheri	SRM Institute of Science and Technology
3	Bahwan CyberTek IT Park	Thorappakkam	Zafin Labs, Beroe Consulting India, Identive T...
4	IITM Research Park	Taramani	Chakra Network Solutions, Saint-Gobain, Centre...

2.1 Getting Coordinates of Major Areas.

The OpenCage Geocoder (https://api.opencagedata.com/geocode/v1/json) is used to provide text to latitude and longitude via a RESTful API. The result is stored Github for further reference.

```
longitude = []
latitude = []
for index, row in chennai_df.iterrows():
    try:
        # api-endpoint
        URL = "https://api.opencagedata.com/geocode/v1/json"

        # defining a params dict for the parameters to be sent to the API
        PARAMS = {'pretty':'1', 'q': row['CompanyName'] + ', ' + row['Location'], 'key':'5238a8d178f0435cb5e6d6519281a1e0', 'language':'en'}

        # sending get request and saving the response as response object
        r = requests.get(url = URL, params = PARAMS)

        # extracting data in json format
        json_coordinates = r.json()

        if json_coordinates['results']:
            print(row['CompanyName'], json_coordinates['results'][0]['geometry'])
            longitude.append(json_coordinates['results'][0]['geometry']['lng'])
            latitude.append(json_coordinates['results'][0]['geometry']['lat'])
        else:
            print(row['CompanyName'], ' No data')
            longitude.append(0)
            latitude.append(0)
    except ValueError:
        print('Error in:' + row.Location)

chennai_df['longitude'] = longitude
chennai_df['latitude'] = latitude
# chennai_df.to_csv('chn_companies_coordinates.csv', index=False) ## uncomment if you want to export it in csv file
chennai_df.head()
```

	CompanyName	Location	CompaniesList	longitude	latitude
0	Tidel Park	Taramani	Cisco Systems, Hexaware Technologies, Sify, Te...	80.240407	12.984907
1	Olympia Tech Park	Guindy	HP, Verizon, ABN Amro, Visteon, Hewitt, Merril...	80.220365	13.008710
2	SRM Tech Park	Potheri	SRM Institute of Science and Technology	80.045089	12.825011
3	Bahwan CyberTek IT Park	Thorappakkam	Zafin Labs, Beroe Consulting India, Identive T...	80.278470	13.087840
4	IITM Research Park	Taramani	Chakra Network Solutions, Saint-Gobain, Centre...	80.240407	12.984907