

# 1. 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](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