TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

Moving is painful even though it is for the career or job of your dreams. Job relocation or temporary housing in a new location is a huge topic of discussion which is mostly commonly occurring scenario for an individual if not so often. There are so many articles and literatures about this topic where one have to make a life decision or choose for the best option possible. Multiple commercial players are available in market some of them include WHRG Group, SHRM, CapRelo Hellolanding etc. to aid and support individual going through relocation specific phase of their life. There are several factors to consider during relocation which include the new cost of living, weather, company's performance, individual's relationships, a new home or temporary housing and so on. In case of searching of a new home factors like proximity to essential services, convenience of transport, flexibility of agreements, a new lifestyle etc. have to be considered. Apart from personal relocation there are other specific use cases like traveling nurse housing, corporate housing, inter housing etc.

1.2. PROJECT DESCRIPTION

In this project let us focus on a particular case of relocation where an individual would want to relocate from hometown for a new job. Main focus of the current project is restricted to identifying locations near job place that are similar to places in hometown without loss in generality.

CHAPTER 2. DATA DESCRIPTION

2.1. NEIGHBORHOOD

Understanding of the neighborhood of the hometown and job location is essential for solving this problem. First hometown is chosen to be Chennai city, India, where Adyar and Kodambakkam are chosen to be zones of importance. Second the job location is chosen to be Bangalore, India, where zones of interest will be obtained in such a way that they are within 10 Km range from the office location of Whitefield area. Once the zone information of the cities are identified it is equally important to obtain the latitude and longitude geographical coordinates of the zones. After obtaining all relevant information, neighborhood of each zone is explored using Foursquare API. Later the neighborhood information is analyzed for similarities and differences. Finally an appropriate neighborhood in the job location similar to hometown is chosen based on the analysis results.

2.2. DATA SOURCE

Zone information is obtained from Wikipedia pages –

- https://en.wikipedia.org/wiki/List_of_Chennai_Corporation_zones
- https://en.wikipedia.org/wiki/List_of_wards_in_Bangalore

Geographical coordinate information is obtained from –

- https://indiamapia.com/Chennai.html
- https://indiamapia.com/Bangalore.html

Zonal and geographical information for important regions mentioned in this project are not readily available for direct usage. Hence, the date from the above mentioned pages are gathered and combined appropriately to be used in this project.

CHAPTER 3. METHODOLOGY

3.1. DATA EXTRACTION AND CLUSTERING

In this section, first collect the zonal information data from the mentioned website in the data section. Beautiful soup is used to extract data from HTML, which is useful for web scraping. First, inspect the webpage to identify the class attribute for the table of interest. Raise a get request to fetch the raw HTML content and parse the content. Read the table of interest with the identified class attribute and extract all table information.

After obtaining the zonal information, latitude and longitude location data for each zone is obtained using Geopy, Nominatim package which is commonly used for converting address to location information. The location data is visually rendered on a map using Folium.

Each neighborhood location is explored using foursquare API requests and the JSON response from the API request is processed to obtain only relevant information. The Obtained venue data is explored for total number of venues and unique categories.

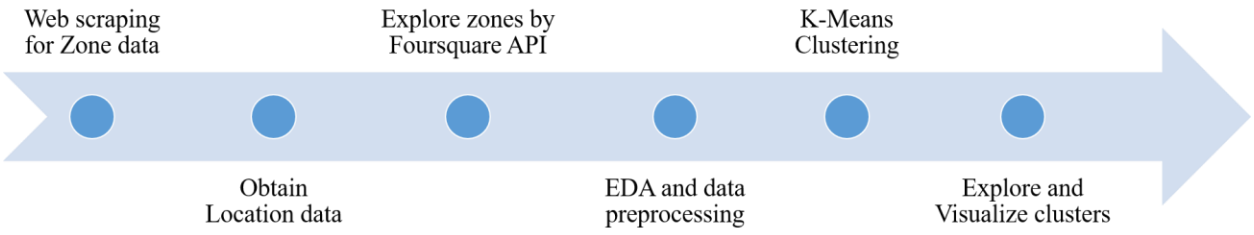

Figure 3.1 Data Extraction and clustering process

One hot coding is done to convert the unique categories to venues to numerical data and the average occurrence of a particular category of venue is obtained. Most

commonly occurring venues in a given neighborhood are identified and a data frame consisting of the neighborhood and its corresponding 10 most commonly occurring venues is prepared.

K-Means clustering algorithms is used to group the venues based on the similarities of the occurrence of venues. In this project, a cluster value of five is chosen to group the neighborhoods into five different categories. The Clustered information is then visually rendered using folium. Individual cluster groups are explored for their similarities and their descriptions. The complete data extraction and clustering process is shown in Figure 3.1.

3.2. IDENTIFYING SIMILAR NEIGHBORHOODS

The above extraction and clustering process is repeated for two main locations i.e. Chennai or hometown location and Bangalore or job location. Neighborhoods of both job and home location are explored first for similarities. Home location neighborhoods are filtered for regions of interest for comparison with job location neighborhoods. Neighborhoods in job location are filtered based on distance from office area which is chosen to be less than 10 Km.

The radial distance between two latitude and longitude location information is calculated using Geopy geodesic distance calculator module.

One important fact to observe is that the distance obtained is radial or perpendicular distance between two latitude and longitude location data and not road wise distance that is noticed commonly on any navigator like google maps.

The filtered location information from hometown and job location are combined in single data frame. The data is then visualized using folium renderer and explore of sufficient information. Already preprocessed data for previous clustering is used

which includes the average occurrence of a particular venue category for each neighborhood.

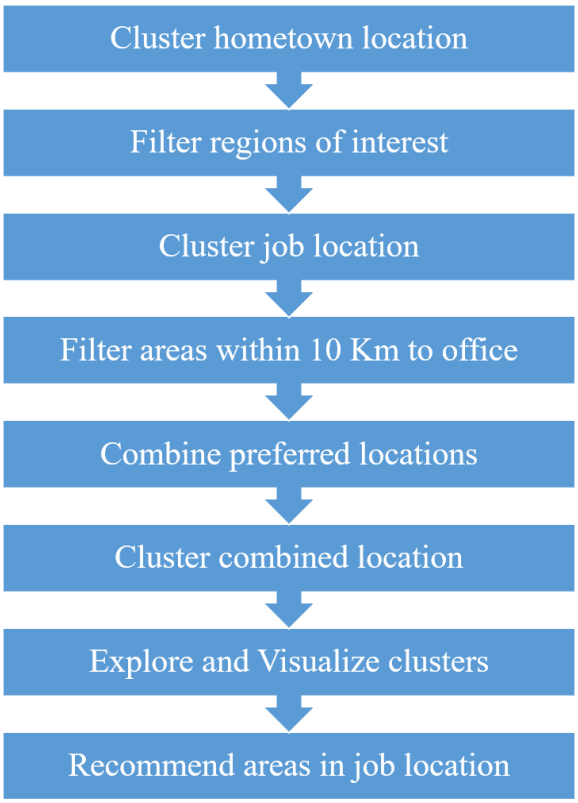

Figure 3.2 Identifying similar neighborhood

The combined filtered data is then clustered using K-Means into five groups. The groups are then visualized and explored individually for similarities and descriptions.

The area in job location, which falls in the same group of the hometown, would become the first choice for selection. In case of multiple areas and groups identified, the most similar hometown location is chosen and the job area in the same group, which is closest to office, are recommended. The process for identifying similar neighborhood is shown in Figure 3.2.

3.3. EXPLORING NEIGHBORHOODS OF CHENNAI AND BANGALORE

3.3.1 Extracting Zonal information

Zones of Chennai and Bangalore are obtained from the website mentioned in data section 2.2 using Beautiful soup. Obtained area and pincode information for Chennai and Bangalore is shown in Figure 3.3 and Figure 3.4 respectively. Total number of Location information include 299 and 295 for Chennai and Bangalore respectively.

	S.No.	Location	Pincode
0	1	Abiramapuram	600018
1	2	Adambakkam	600088
2	3	Adyar	600020
3	4	Agaram	600082
4	5	Alandur	600016

Figure 3.3 Chennai pincode

	S.No.	Location	Pincode
0	0	A F Station Yelahanka	560063
1	1	Adugodi	560030
2	2	Agara	560034
3	3	Agram	560007
4	4	Air Force Hospital	560007

Figure 3.4 Bangalore pincode

3.3.2 Obtain Location data

Latitude and longitude location data for Chennai and Bangalore can be obtained either by scraping through the webpages mentioned in section 2.2 or the zone information can be converted to latitude and longitude using the python module Nominatim from Geopy package. In this project the latter approach is used to obtain location information.

	S.No.	Location	Latitude	Longitude
0	2	Adambakkam	12.982221	80.209121
1	3	Adyar	13.006450	80.257779
2	4	Agaram	13.115860	80.230041
3	5	Alandur	12.994373	80.194284
4	6	Alandur North	12.994373	80.194284

Figure 3.5 Chennai location dataframe

	S.No.	Location	Latitude	Longitude
0	1	Adugodi	12.942847	77.610416
1	2	Agara	12.620112	77.479307
2	5	Amruthahalli	13.066513	77.596624
3	6	Anandnagar	13.033377	77.589523
4	7	Anekal	12.799122	77.680604

Figure 3.6 Bangalore location dataframe

The obtained location information is stored in pandas dataframe as shown in Figure 3.5 and Figure 3.6. A Snap of output from the Geopy Nominatim package during latitude and longitude extraction is shown in Figure 3.7 and Figure 3.8.

```
The geograpical coordinate of Adambakkam, Chennai are 12.9822215, 80.209121.
The geograpical coordinate of Adyar, Chennai are 13.00645, 80.2577791.
The geograpical coordinate of Agaram, Chennai are 13.1158603, 80.2300407.
The geograpical coordinate of Alandur, Chennai are 12.9943729, 80.1942837.
The geograpical coordinate of Alandur North, Chennai are 12.9943729, 80.1942837.
The geograpical coordinate of Ambattur H O, Chennai are 13.1139774, 80.1529202.
The geograpical coordinate of Ambattur West, Chennai are 13.1193746, 80.1507648.
The geograpical coordinate of Anakaputhur, Chennai are 12.9838095, 80.1688264.
The geograpical coordinate of Anna Road H O, Chennai are 13.1207974, 80.0948305.
The geograpical coordinate of Arcot Road, Chennai are 13.0499473, 80.21188.
The geograpical coordinate of Arumbakkam, Chennai are 13.074371, 80.2081312.
The geograpical coordinate of Ashok Nagar, Chennai are 13.0400731, 80.2159247.
The geograpical coordinate of Avadi, Chennai are 13.1254758, 80.09409.
The geograpical coordinate of Ayanavaram, Chennai are 13.0946157, 80.23541.
The geograpical coordinate of Besant Nagar, Chennai are 12.9996907, 80.2682259.
The geograpical coordinate of Broadway, Chennai are 13.0962367, 80.2864133.
The geograpical coordinate of Cathedral, Chennai are 13.0518669, 80.25307675773865.
The geograpical coordinate of Cemetry Road, Chennai are 13.1091962, 80.2861039.
```

Figure 3.7 Obtaining Chennai coordinates

```
The geograpical coordinate of Adugodi, Bangalore are 12.9428472, 77.6104155.
The geograpical coordinate of Agara, Bangalore are 12.6201123, 77.4793074.
The geograpical coordinate of Amruthahalli, Bangalore are 13.0665131, 77.5966238.
The geograpical coordinate of Anandnagar, Bangalore are 13.033377, 77.5895227.
The geograpical coordinate of Anekal, Bangalore are 12.799122, 77.68060437921741.
The geograpical coordinate of Arabic College, Bangalore are 13.0300088, 77.6208657.
The geograpical coordinate of Aranya Bhavan, Bangalore are 13.01061485, 77.57081951019086.
The geograpical coordinate of Ashoknagar, Bangalore are 12.9791198, 77.5912997.
The geograpical coordinate of Attibele, Bangalore are 12.7782588, 77.7712834.
The geograpical coordinate of Attur, Bangalore are 13.0994536, 77.5680336.
The geograpical coordinate of Austin Town, Bangalore are 12.9612736, 77.6152935.
The geograpical coordinate of Avalahalli, Bangalore are 13.0352777, 77.7365655.
The geograpical coordinate of Avenue Road, Bangalore are 12.9684724, 77.5789181.
The geograpical coordinate of Bagalgunte, Bangalore are 13.0566493, 77.5048222.
The geograpical coordinate of Bagalur, Bangalore are 13.1331868, 77.6687093.
```

Figure 3.8 Obtaining Bangalore coordinates

Important point to note is Nominatim package does not provide latitude and longitude coordinates for all the available areas. Python package could provide coordinates for 145 out of 299 areas in Chennai and 184 out of 295 areas in Bangalore. Hence only are areas that have valid coordinates are included for further processing.

3.3.3 Rendering and visualizing location data on folium map

Obtained zonal and coordinate information for areas in Chennai and Bangalore are visualized by rendering it in folium map. Maps of Chennai and Bangalore with neighborhood information is shown in Figure 3.9 and Figure 3.10.

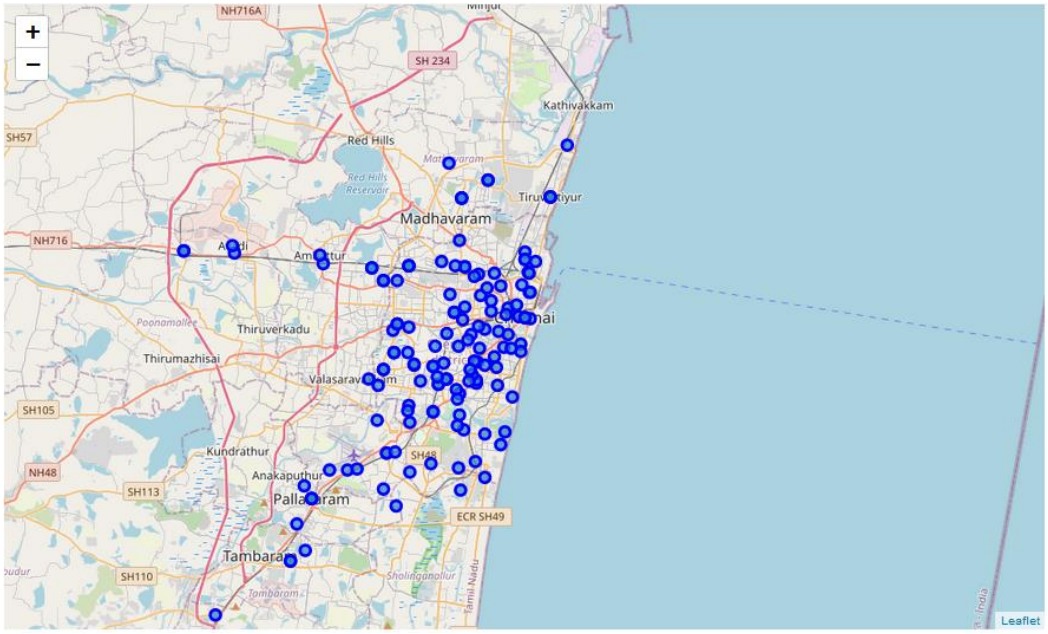

Figure 3.9 Chennai map with neighborhoods

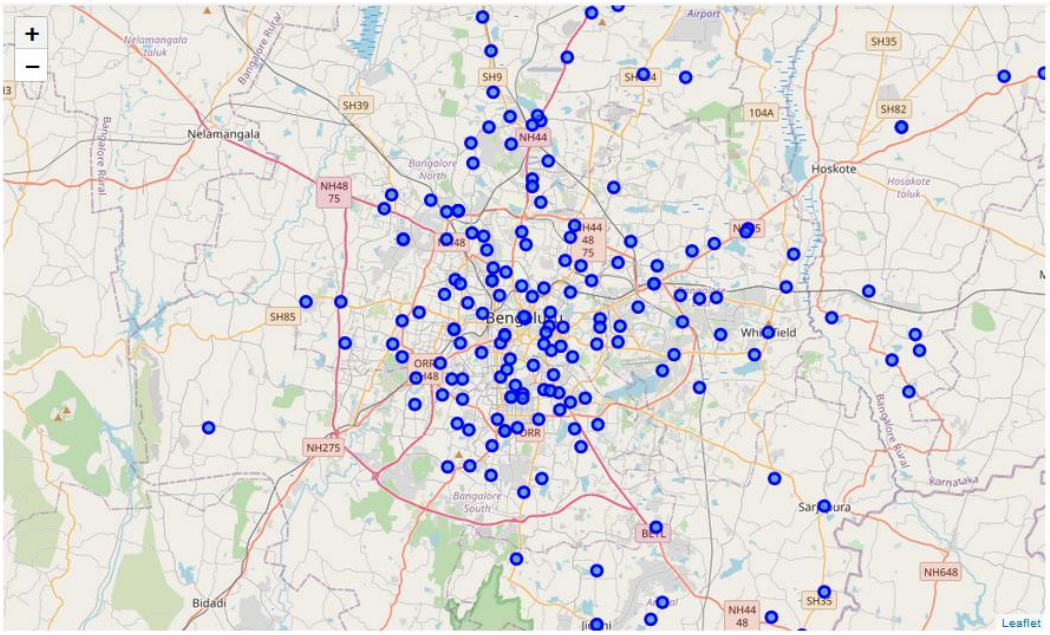

Figure 3.10 Bangalore map with neighborhoods

3.3.4 Explore zones by Foursquare API

Venues near each neighborhood in Chennai and Bangalore are explored using foursquare API through requests module. The obtained JSON file is processed to keep only relevant information and the venue information is stored in pandas dataframe.

3.3.5 Exploratory data analysis (EDA)

Obtained venue information from foursquare is then explored for –

- Total number of venues
- Number of venues near each neighborhood
- Unique categories of venues

Total number of venues

Total number of venues near Chennai and Bangalore are 3368 and 4600 venues respectively as shown in Figure 3.11 and Figure 3.12.

```
print(chennai_venues.shape)
chennai_venues.head(10)
```

```
(3368, 7)
```

Figure 3.11 Number of venues near Chennai

```
print(bangalore_venues.shape)
bangalore_venues.head(10)
```

```
(4600, 7)
```

Figure 3.12 Number of venues near Bangalore