The Battle of Neighbourhoods-Week 4 Report

(Comparing the similarity of neighbourhoods in the cities of India)

Karan Thakur

IBM Data Science Professional Certificate

Contents

1. Introduction

1.1 Background

India is the seventh largest country in the world by geographical area and world's second-most populous country with more than 1.38 crore people. According to Census of India 2011, there are 7935 cities. The number of metropolitan cities having million plus population has also increased from 35 to 53 as per 2011 census. The name of such cities can be found here. In the current modern world citizens migrate from one place to another for a better opportunity. As per the article, Bangalore's population comprises of 42% migrants.

1.2 Problem

The world is facing an unprecedented threat from the COVID-19 pandemic. It is difficult for a citizen who wants to explore the whole new city and choose the locality in order to find a place of reside.

1.3 Interest

The aim of the project is to compare the cities of India to find the similarities based on the venues in the neighbourhood. This helps the working class people who may have to move to new city and are looking for the comfortable locality to stay. The locality they would want to choose is dependent on the lifestyle they carry that is in turn dependent majorly on the venues in the nearby locality. For example some people prefer to stay in a neighbourhood which has restaurants, shopping malls in the near vicinity whereas some people prefer staying near to parks, recreational zones etc. In this project, I will be making an attempt to compare the similarities of cities Pune and Bangalore City, and this methodology can be extended to compare any cities of India.

2. Data collection and cleaning

2.1 Data Sources

Since post office is a public facility that provides mail services, including accepting of letters and parcels, providing post office boxes to every nook and corner of the country, I will be using the all india pincode dataset from the Open Government Data(OGD) Platform India to derive the neighbourhoods of the city. The dataset has attributes like Office name, Pincode, Office type, Division, Region, Taluk, District, State, Latitude and Longitude.

```
1  all_india_postal_df = pd.read_csv("all_india_PO_list_without_APS_offices_ver2_lat_long.csv")
2  all_india_postal_df.head()
```

	officename	pincode	officeType	Deliverystatus	divisionname	regionname	circlename	Taluk	Districtname	statename	Telephone	Related Suboffice
0	Achalapur B.O	504273	B.O	Delivery	Adilabad	Hyderabad	Andhra Pradesh	Asifabad	Adilabad	TELANGANA	NaN	Rechin S.C
1	Ada B.O	504293	B.O	Delivery	Adilabad	Hyderabad	Andhra Pradesh	Asifabad	Adilabad	TELANGANA	NaN	Asifabac S.C
2	Adegaon B.O	504307	B.O	Delivery	Adilabad	Hyderabad	Andhra Pradesh	Boath	Adilabad	TELANGANA	NaN	Echoda S.C
3	Adilabad Collectorate S.O	504001	S.O	Non-Delivery	Adilabad	Hyderabad	Andhra Pradesh	Adilabad	Adilabad	TELANGANA	08732-226703	NaN
4	Adilabad H.O	504001	H.O	Delivery	Adilabad	Hyderabad	Andhra Pradesh	Adilabad	Adilabad	TELANGANA	08732-226738	NaN

```
1  all_india_postal_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 154797 entries, 0 to 154796
Data columns (total 15 columns):
officename          154797 non-null object
pincode             154797 non-null int64
officeType          154797 non-null object
Deliverystatus      154797 non-null object
divisionname        154797 non-null object
regionname          154797 non-null object
circlename          154797 non-null object
Taluk               146488 non-null object
Districtname        154761 non-null object
statename           154761 non-null object
Telephone            23418 non-null object
Related Suboffice   125269 non-null object
Related Headoffice  153987 non-null object
longitude              142 non-null float64
latitude               142 non-null float64
dtypes: float64(2), int64(1), object(12)
memory usage: 17.7+ MB
```

Figure 1: Dataset from OGD India platform with data for post offices.

On examining, we find that the dataset is not updated with the latitudes and the longitudes for all the records. To solve this, I will be using geopy, which is a Python client for several popular geocoding web services. The geopy makes it easy to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

We will also use the FourSquare API to retrieve the venues within the radius of 1.5 km of each neighbourhood. Foursquare is a location data provider with information of all type of venues and events within an area of interest. Such information includes venue names, locations, latitudes, longitudes, venue categories, etc.

2.2 Data Cleaning and feature selection

Since we are interested in comparing the neighbourhood of Pune and Bangalore City, we will filter out the data of other cities. And also we do not want the features like officeType, Deliverystatus, divisionname, regionname, circlename, Telephone, Related Suboffice, Related Headoffice for this analysis. We will also be filtering out the duplicate data from the dataset. We need latitudes and longitudes, but since the

dataset is yet to be updated with the details, we will be dropping these features from the dataset and will use the geopy client to retrieve the latitudes and longitudes.

```
1  pune_blore_df.head()
```

	officename	pincode	officeType	Deliverystatus	divisionname	regionname	circlename	Taluk	Districtname	statename	Telephone	
52175	A F Station Yelahanka S.O	560063	S.O	Delivery	Bangalore East	Bangalore HQ	Karnataka	Bangalore North	Bangalore	KARNATAKA	080-28478031	
52176	Agram S.O	560007	S.O	Delivery	Bangalore East	Bangalore HQ	Karnataka	Bangalore South	Bangalore	KARNATAKA	080-25300192	
52177	Air Force Hospital S.O	560007	S.O	Non-Delivery	Bangalore East	Bangalore HQ	Karnataka	Bangalore North	Bangalore	KARNATAKA	080-25302699	
52178	Amruthahalli B.O	560092	B.O	Delivery	Bangalore East	Bangalore HQ	Karnataka	Bangalore North	Bangalore	KARNATAKA	NaN	Sah
52179	Anandnagar S.O (Bangalore)	560024	S.O	Non-Delivery	Bangalore East	Bangalore HQ	Karnataka	Bangalore North	Bangalore	KARNATAKA	080-23431761	

Figure 2: Filtered data for Pune and Bangalore

To fetch the coordinates of the neighbourhood we need to pass the address, city to the API and hence I will be creating a derived feature Neighbourhood with the concatenation of officename, Districtname and statename. Dataset with the latitudes and longitudes for Bangalore is as below:

```
1  pune_blore_df.head()
```

	officename	pincode	Taluk	Districtname	statename	Neighbourhood	Latitude	Longitude
0	Air Force Hospital	560007	Bangalore North	Bangalore	KARNATAKA	Air Force Hospital, Bangalore, KARNATAKA	12.964	77.6275
1	Amruthahalli	560092	Bangalore North	Bangalore	KARNATAKA	Amruthahalli, Bangalore, KARNATAKA	13.0665	77.5966
2	Arabic College	560045	Bangalore North	Bangalore	KARNATAKA	Arabic College, Bangalore, KARNATAKA	13.03	77.6209
3	Attur	560064	Bangalore North	Bangalore	KARNATAKA	Attur, Bangalore, KARNATAKA	13.0995	77.568
4	Austin Town	560047	Bangalore North	Bangalore	KARNATAKA	Austin Town, Bangalore, KARNATAKA	12.9613	77.6153

Figure 3: Pune and Bangalore dataset with latitudes and longitudes

We have 101 and 142 neighbourhoods along with geographical coordinates for Pune and Bangalore respectively.

Visualising the neighbourhoods on the map for Pune and Bangalore using folium package
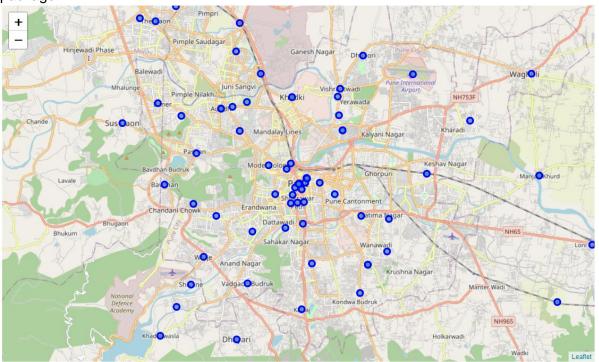

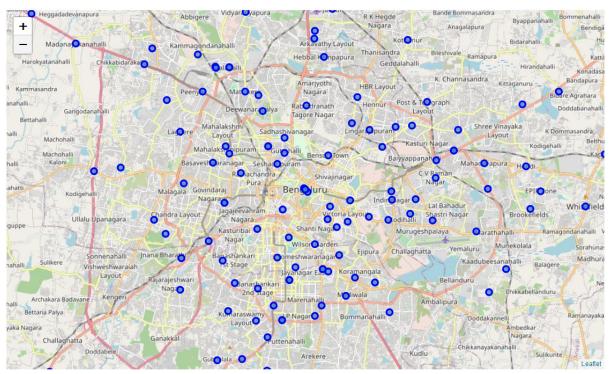Figure 4: Map of Pune along with the neighbourhoods

Figure 5: Map of Bangalore along with the neighbourhoods

The FourSquare's explore API will help us in providing with the list of the venues in the neighbourhood. For each of the neighbourhood area, we have chosen the limit to be 100, and the radius to be 1500 meters.

```
1  print(punebangalore_venues.shape)
2  punebangalore_venues.head()
```

(9558, 7)

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Air Force Hospital, Bangalore, KARNATAKA	12.964027	77.6275	League of Extraordinary Gamers	12.967099	77.636919	Gaming Cafe
1	Air Force Hospital, Bangalore, KARNATAKA	12.964027	77.6275	Corner House	12.969973	77.637453	Ice Cream Shop
2	Air Force Hospital, Bangalore, KARNATAKA	12.964027	77.6275	Vivanta by Taj	12.973365	77.619951	Hotel
3	Air Force Hospital, Bangalore, KARNATAKA	12.964027	77.6275	Retro Lounge Bar	12.972614	77.626700	Bar
4	Air Force Hospital, Bangalore, KARNATAKA	12.964027	77.6275	Foodhall	12.973486	77.620117	Department Store

Figure 6: Dataframe with the venues in a neighbourhood

These venues will be used for further analysis.

3. References

- http://mohua.gov.in/cms/number-of-cities--towns-by-city-size-class.php
- https://www.thehindu.com/news/cities/bangalore/migrants-constitute-42-of-bengalurus-population/article28734588.ece