The Battle of Neighborhoods – week 1

30st April 2020

1.Introduction

1.1 BACKGROUND

According to a survey conducted by Indian government 23% of India's population runs their own business. When a person wants to choose a city of india to start his own food business, he needs to go through various factors like city population, most visited place in the city, etc.,

By analysing various cities with respect to the population, most visited places, transportation facilities, types of restaurants/ food chains, a person can finalise his business location and the cuisine.

1.2 Problem Statement

The dataset consists of the latest population of various cities in India as of 2019. The problem statement is to find out the best cities in India when a person desires to migrate to a city where there is a moderate population and to start his own food business.

This project aims to help the person to choose a city with moderate population and to go through the most common visited venues in that city to choose the best city and choose a best suited cuisine accordingly.

2.Data Acquisition and Pre-Processing

2.1 Data Collection and description

The dataset used in this project is downloaded from the https://simplemaps.com/data/in-cities which consists of city wise India's population data along with the data of location coordinates and migrated population.

The data set consists of 7 columns

1. **City:** Name of the city
2. **Lat:** The city's Latitude
3. **Lng:** The city's Longitude
4. **Country** : India
5. **Iso2** : India's location code
6. **Admin** : state name
7. **Capital**: category of city
8. **Population** : population of city
9. **Migrated population**: the population migrated to that city

2.2 Data Cleaning

The data is downloaded from the url of given website as a dataframe - the raw data before data preprocessing looks like in the image below

	city	lat	lng	country	iso2	admin	capital	population	population_proper
0	Mumbai	18.987807	72.836447	India	IN	Mahārāshtra	admin	18978000.0	12691836.0
1	Delhi	28.651952	77.231495	India	IN	Delhi	admin	15926000.0	7633213.0
2	Kolkata	22.562627	88.363044	India	IN	West Bengal	admin	14787000.0	4631392.0
3	Chennai	13.084622	80.248357	India	IN	Tamil Nādu	admin	7163000.0	4328063.0
4	Bengalūru	12.977063	77.587106	India	IN	Karnātaka	admin	6787000.0	5104047.0

- After the **preprocessing**
    1. Removing NA values
    2. Removing unnecessary columns
    3. Creating new columns for migrated population and location(combination of city, state and country)
    4. Label encoding the categorical value in capital column

        1. Delhi - Capital of India - 3 (category after encoding)

        2.Ahmedabad - minor - 2

        3. Hyderabad- Capital of state - 1

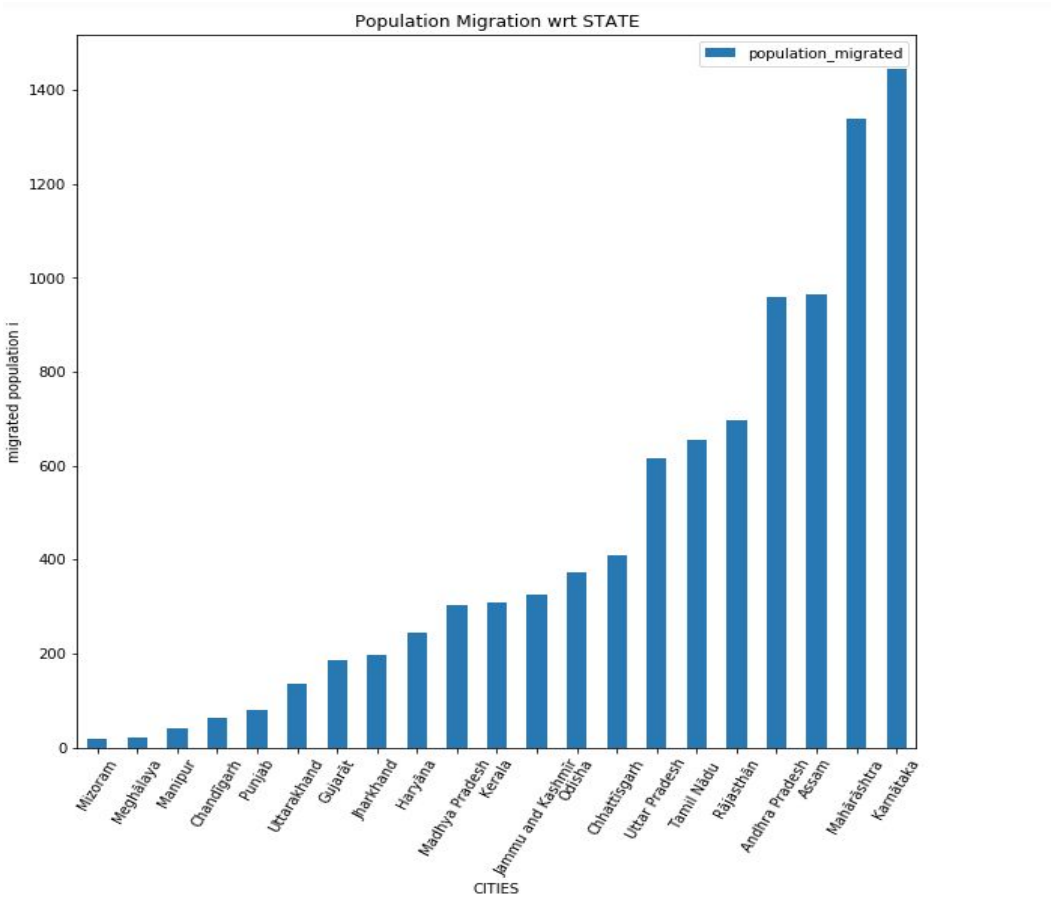
        4. Hauora - non capital city - 0

3.Exploratory Data Analysis :

3.1 filtering the data according to the requirements :

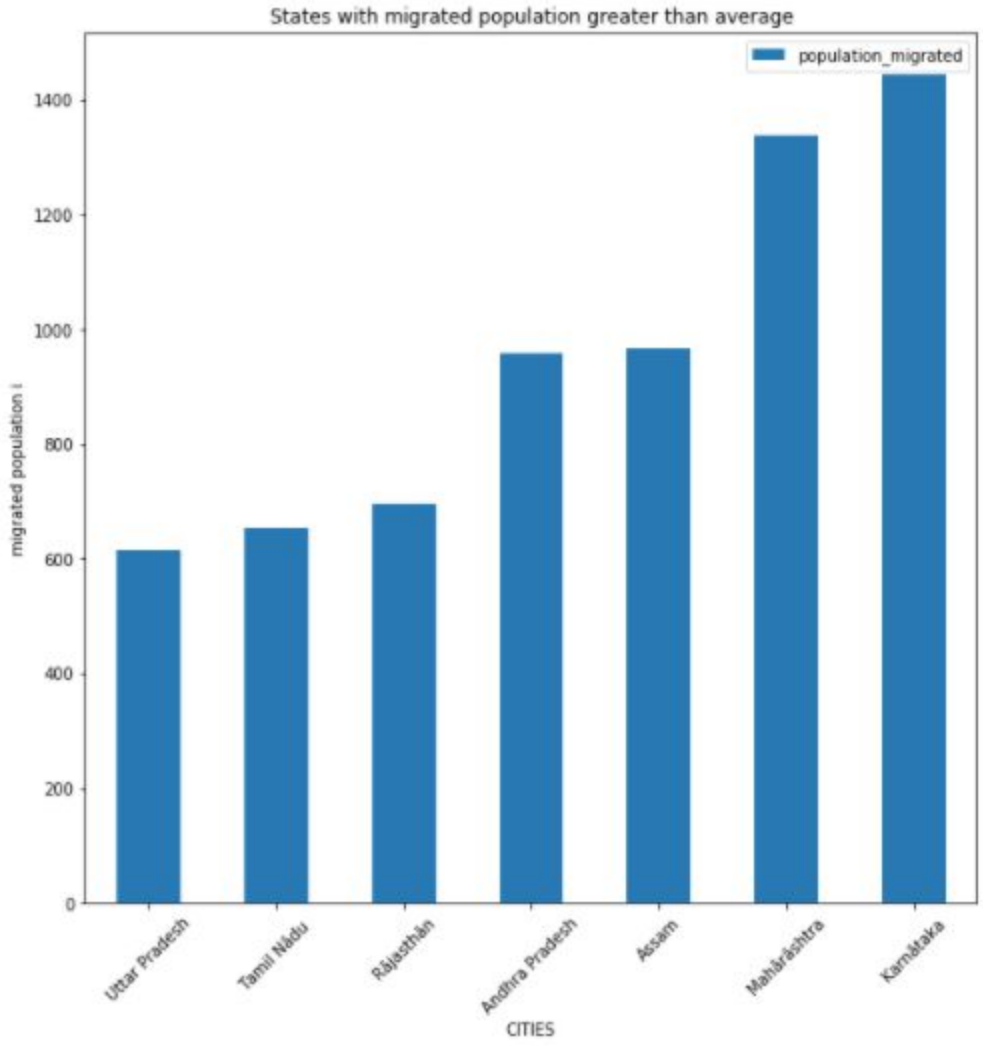
3.1 .1 filtering Moderately populated areas :

The cities are filtered by using the function which generates the quartile data - Here the cities with top and bottom 20% population are eliminated by using the function np.percentile. The 60% of cities are held for our further assessment



3.1 .2 filtering highly migrated areas :

The cities which are highly migrated are filtered for our further analysis which can be achieved by only considering the cities with higher than average population migration. Finally there are 7 states where the migration is more.



3.1.3 Final data :

The final data after all the filtering is used in foursquare API to get the venue details.

	Location	capital	city	country	lattitude	longitude	population	population_migrated	population_proper	state
136	Rāichūr,Karnātaka,India	0	Rāichūr	India	16.205459	77.355670	255240.0	29278.0	225962.0	Karnātaka
133	Bijāpur,Karnātaka,India	0	Bijāpur	India	16.827715	75.718988	271064.0	194.0	270870.0	Karnātaka
130	Chikka Mandya,Karnātaka,India	0	Chikka Mandya	India	12.545602	76.895078	285034.0	150189.0	134845.0	Karnātaka
129	Hospet,Karnātaka,India	0	Hospet	India	15.269537	76.387103	286007.0	88161.0	197846.0	Karnātaka
121	Bīdar,Karnātaka,India	0	Bīdar	India	17.913309	77.530105	300136.0	96065.0	204071.0	Karnātaka

3.2 Generating the venues using Foursquare API :

The foursquare api is used to get the venue details of the given location coordinates within the radius of 1000 meters.The API generated 3800 + venues within the given locations. Those locations are further analysed to get the most commonly visited venues.

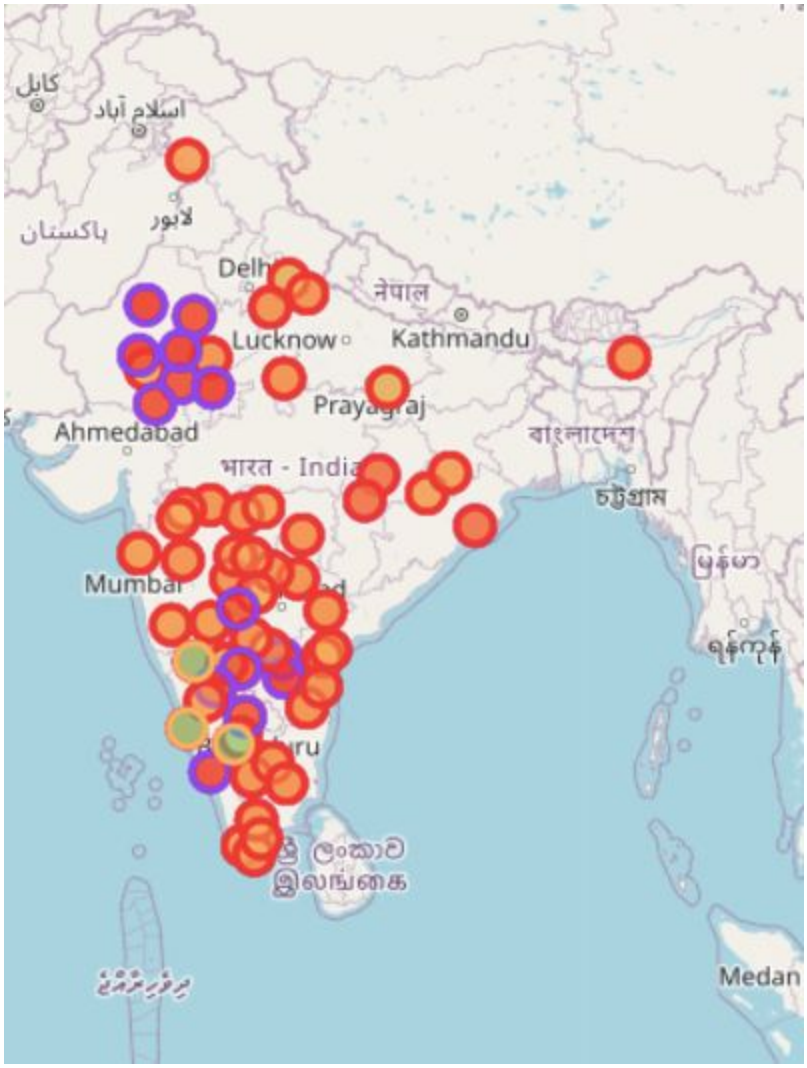
```
: (3848, 7)
```

	city	city Latitude	city Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rāichūr	16.205459	77.35567	Baskin-Robbins	16.197120	77.343124	Ice Cream Shop
1	Rāichūr	16.205459	77.35567	Aahar	16.206770	77.353980	Indian Restaurant
2	Rāichūr	16.205459	77.35567	Liberty Exclusive Showroom	16.207966	77.351776	Shoe Store
3	Rāichūr	16.205459	77.35567	Raichur Railway Station	16.193046	77.339428	Train Station
4	Rāichūr	16.205459	77.35567	Hill Fort, Raichur	16.203551	77.349017	Mountain

4.Result

The foursquare API generated data gives us the top 10 most visited venues along with their coordinates which is merged with the final_data in step 3.1.3 to apply KMeans algorithm to get the similar data. The similarity can be found by using “cluster labels” column. The similar places are assigned with the same cluster. Let us use the data to generate Map which locates the clusters.

population_proper	state	...	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	Cluster Labels	9th Most Common Venue	10th Most Common Venue
225962.0	Karnātaka	...	Ice Cream Shop	Shoe Store	Train Station	Mountain	Zoo	Food & Drink Shop	French Restaurant	2	Food Truck	Food Stand
270870.0	Karnātaka	...	Café	Indian Restaurant	Train Station	Hotel	Historic Site	French Restaurant	Food Truck	4	Food Stand	Food Court
134845.0	Karnātaka	...	Convenience Store	Bus Station	Ice Cream Shop	Train Station	Smoke Shop	Fried Chicken Joint	French Restaurant	2	Food Truck	Food Stand
197846.0	Karnātaka	...	Diner	Vegetarian / Vegan Restaurant	Hotel Bar	Historic Site	Zoo	Food	French Restaurant	0	Food Truck	Food Stand
204071.0	Karnātaka	...	Train Station	Historic Site	Food & Drink Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	4	Food Court	Food



```
merged_city_df[merged_city_df['Cluster Labels'] == 0]

#vegetarian restauranta +hotels + historic sites + zoo+ mountain +train and bus
```

```
merged_city_df[merged_city_df['Cluster Labels'] == 1]
# concentrated on shopping malls + zoo+ chicken/French
```

```
merged_city_df[merged_city_df['Cluster Labels'] == 2]
#icecream shops+ bus + train + food shops
```

```
merged_city_df[merged_city_df['Cluster Labels'] == 3]

# concentrated on cafe + park +food courts +lake +hotels
```

```
merged_city_df[merged_city_df['Cluster Labels'] == 4]

# concentrated on food cafes and indian & french restaurants and have train and bus facilities with historic places
```

Here you can see the most common visited places in that particular cluster. You can choose the city as per your requirement for example : If a person wants to stay in a city where there are more shopping malls which gives more business if that person takes a franchise in a shopping mall, he gets more profits and the % of people who knows about that shop increases.