

The Battle of Neighbourhoods

Data Science Capstone





Introduction





Background

Moving to a new place which is close to your heart with nice venues around starting own business requires deep venue wise analysis of a city.

Problem Statement

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.



Data Acquisition & cleaning

Data Collection and description:


Data is downloaded from the url : <https://simplemaps.com/data/in-cities>

Description :

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. **Proper population:** the population to that is native to the city

	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	Karnāṭaka	admin	6787000.0	5104047.0

Data Cleaning and preprocessing:

- 
1. Removing NA values
 2. Removing unnecessary columns(iso2)
 3. Creating new columns for migrated population and location(combination of city, state and country)
 4. Label encoding the categorical value in capital column
 - Delhi - Capital of India - 3 (category after encoding)
 - Ahmedabad - minor - 2
 - Hyderabad- Capital of state - 1
 - Hauora - non capital city - 0

	city	latitude	longitude	country	state	capital	population	population_proper	population_migrated	Location
4	Bengalūru	12.977063	77.587106	India	Karnāṭaka	1	6787000.0	5104047.0	1682953.0	Bengalūru,Karnāṭaka,India
3	Chennai	13.084622	80.248357	India	Tamil Nādu	1	7163000.0	4328063.0	2834937.0	Chennai,Tamil Nādu ,India
2	Kolkata	22.562627	88.363044	India	West Bengal	1	14787000.0	4631392.0	10155608.0	Kolkata,West Bengal,India
1	Delhi	28.651952	77.231495	India	Delhi	1	15926000.0	7633213.0	8292787.0	Delhi,Delhi,India
0	Mumbai	18.987807	72.836447	India	Mahārāshṭra	1	18978000.0	12691836.0	6286164.0	Mumbai,Mahārāshṭra,India

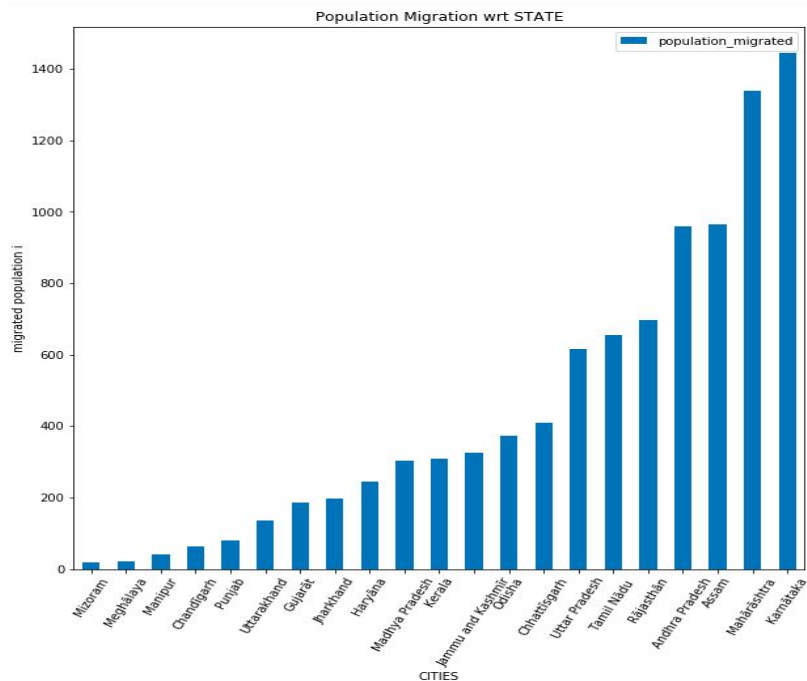


Exploratory data analysis

Filtering the data according to requirements

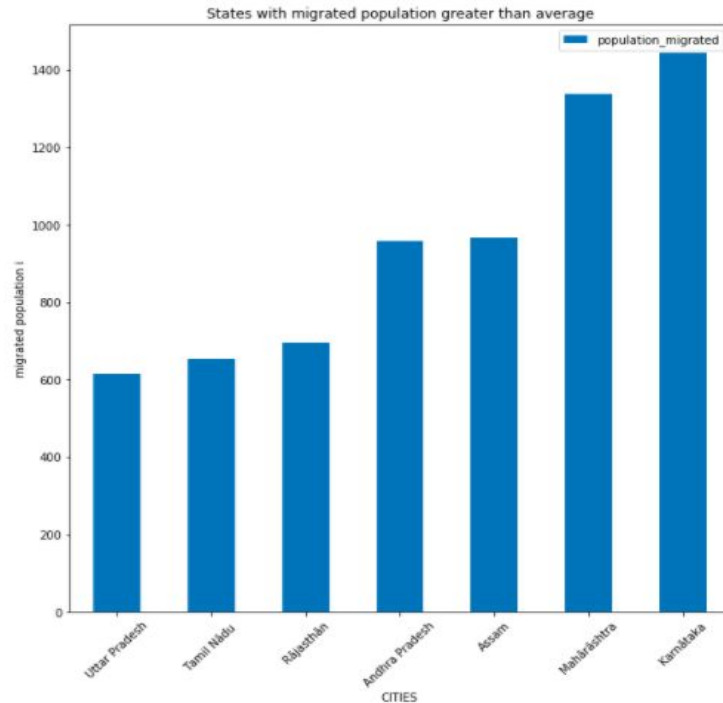
1. filtering Moderately populated areas :

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



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.





Modeling



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

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 K-Means 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.

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	Karnataka	...	Ice Cream Shop	Shoe Store	Train Station	Mountain	Zoo	Food & Drink Shop	French Restaurant	2	Food Truck	Food Stand
270870.0	Karnataka	...	Café	Indian Restaurant	Train Station	Hotel	Historic Site	French Restaurant	Food Truck	4	Food Stand	Food Court
134845.0	Karnataka	...	Convenience Store	Bus Station	Ice Cream Shop	Train Station	Smoke Shop	Fried Chicken Joint	French Restaurant	2	Food Truck	Food Stand
197846.0	Karnataka	...	Diner	Vegetarian / Vegan Restaurant	Hotel Bar	Historic Site	Zoo	Food	French Restaurant	0	Food Truck	Food Stand
204071.0	Karnataka	...	Train Station	Historic Site	Food & Drink Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	4	Food Court	Food



Results



The clusters generated by using the K Means algorithm can be seen below

```
: 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
```

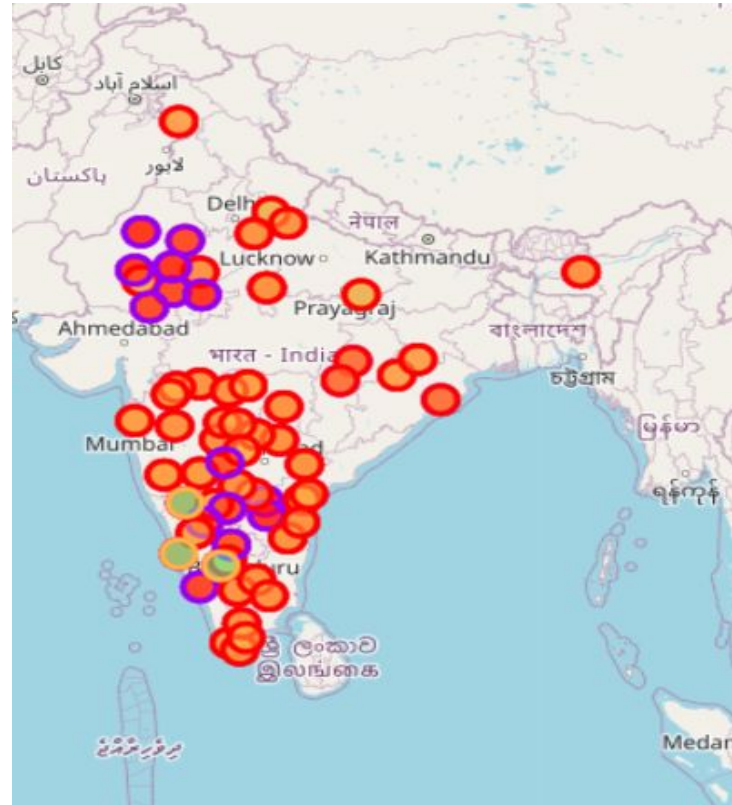
```
: merged_city_df[merged_city_df['Cluster Labels'] == 3]
```

```
# 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
```

The Map of the clustered data







Conclusion





1. This project helps a person to choose a location which is near to the desire to stay in a moderately populated area and also helps in fulfilling the aim of starting own business
2. This project helps 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.