1 Discussion of Background and Problem

1.1 Background

Bangalore, officially Bengaluru, is the capital of the Indian state of Karnataka. It has a population of over ten million, making it a megacity and the third-most populous city and fifth-most populous urban agglomeration in India. It is located in southern India, on the Deccan Plateau at an elevation of over 900 m (3,000 ft) above sea level. Its multi-ethnic, multi-religious,[promotional language] and cosmopolitan character[promotional language] is reflected by its more than 1000 Hindu temples, 400 mosques, 100 churches, 40 Jain Basadis, three Sikh gurdwaras, two Buddhist viharas and one Parsi fire temple located in an area of 741 km² of the metropolis.[citation needed] The religious places are further represented by the proposed Chabad of the Jewish community. The numerous Bahá'ís have a society called the Bahá'í Centre.

Bengaluru is sometimes referred to as the "Silicon Valley of India" (or "IT capital of India") because of its role as the nation's leading information technology (IT) exporter. Indian technological organisations ISRO, Infosys, Wipro and HAL are headquartered in the city. Bangalore is the second fastest-growing major metropolis in India. Recent estimates of the metro economy of its urban area have ranked Bangalore either the fourth or fifth-most productive metro area of India. It is home to many educational and research institutions. Mysore Sandal Soap is produced in this city. The city also houses the Kannada film industry also called Sandalwood.

Delhi, city and national capital territory, north-central India. The city of Delhi actually consists of two components: Old Delhi, in the north, the historic city; and New Delhi, in the south, since 1947 the capital

of India, built in the first part of the 20th century as the capital of British India.

One of the country's largest urban agglomerations, Delhi sits astride (but primarily on the west bank of) the Yamuna River, a tributary of the Ganges (Ganga) River, about 100 miles (160 km) south of the Himalayas. The national capital territory embraces Old and New Delhi and the surrounding metropolitan region, as well as adjacent rural areas. To the east the territory is bounded by the state of Uttar Pradesh, and to the north, west, and south it is bounded by the state of Haryana.

The pattern of land use in Delhi was influenced considerably by the implementation (albeit partial) of the Delhi Development Authority's 20-year (1962–81) master plan. Broadly, public and semipublic land use was concentrated in the Central Secretariat area of New Delhi and in the Old Secretariat area in the Civil Lines, with subsidiary centres developing in the Indraprastha Estate (an office complex) in the east and in Ramakrishnapuram (an office-cum-residence complex) in the south. A large number of small manufacturing establishments have entrenched themselves in almost every part of Old Delhi, but the main industrial areas have gravitated toward Najafgarh Road in the west and the large planned Okhla Industrial Estate in the south. Land for commercial use is found mainly in the Chandni Chowk and Khari Baoli areas, both in the north; in the Sadar Bazar of Old Delhi; in the Ajmal Khan Road area of Karol Bagh in western Delhi; around Connaught Place in New Delhi; and in the areas of Lajpat Nagar and Srojini Nagar in the south. A number of district and local shopping centres have developed in other localities

1.2 Problem

This projects aims to compare city of Delhi and Bangalore in India. It compares venues and location data of Delhi and Bangalore. With this we try to find how similar or dissimilar neighbourhoods of these cities are. We consider factors like types of venues like food, venues to visit. This project analyzes the neighborhoods in each of these two cities and tries to understand what is popular in them and what they have to offer to someone who is contemplating to make a choice on seeking a life in either of the metro cities.

1.3 Interest

People who would be interested in this study are those who would like to create a projection of potential life and activities in these metro city neighborhoods if the subject moves to live in one of them. Also people who would like to travel for tourism in these cities might also be interested in this project.

2) Data Acquision and Cleaning

2.1 Data Sources

For this study, we needed data about neighborhoods in each of these metro cities. We use data published by the government on postal codes for all India. We download the CSV from following site: https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude.

Here, we download the CSV, filter the csv and convert it into two parts to get csv data of Bangalore and Delhi. After this it is read it into a pandas dataframe.

2.2 Data Cleaning

We shall then clean up the unnecessary columns in the CSV, which is not relevant or useful for our current study. Post office names (***office name***) will be used as the neighborhood names. The imported dataset for delhi is shown as follows:

	Circle Name	Region Name	Division Name	Office Name	Pincode	OfficeType	Delivery	District	StateName
0	Delhi Circle	NaN	Delhi East Division	Anand Vihar SO	110092	SO	Non Delivery	EAST DELHI	Delhi
1	Delhi Circle	NaN	Delhi East Division	Azad Nagar SO East Delhi	110051	SO	Non Delivery	SHAHDARA	Delhi
2	Delhi Circle	NaN	Delhi East Division	Babarpur SO North East Delhi	110032	SO	Non Delivery	SHAHDARA	Delhi
3	Delhi Circle	NaN	Delhi East Division	Badarpur Khadar BO	110090	BO	Non Delivery	NORTH EAST DELHI	Delhi
4	Delhi Circle	NaN	Delhi East Division	Balbir Nagar SO	110032	SO	Non Delivery	SHAHDARA	Delhi

We now see that there are the same *Pincode* values for different neighborhoods. Neighborhood names with the same ***Pincode*** will be combined as a single row by pandas groupby() function.

The first few records of the dataset we now have after this cleanup and curation appear as below.

Out[100]:

	Pincode	Neighborhood	District
0	110001	Baroda House SO,Bengali Market SO,Bhagat Singh...	CENTRAL DELHI
1	110002	AGCR EXTENSION COUNTER,AGCR SO,Ajmeri Gate Ext...	CENTRAL DELHI
2	110003	Delhi High Court Extension Counter SO,Delhi Hi...	CENTRAL DELHI
3	110004	Rashtrapati Bhawan SO	CENTRAL DELHI
4	110005	Anand Parbat Indl Area SO,Anand Parbat SO,Bank...	CENTRAL DELHI

The next step is to add the the required information to dataset. We need the longitude and latitude values for the neighborhoods. We use the *Nominatim* library from *geocoders.geopy* package to find the longitude and latitude for each of the neighborhoods and then we add these to the original dataset.

Out[109]:

	Pincode	Neighborhood	District	Latitude	Longitude
0	110001	Baroda House SO,Bengali Market SO,Bhagat Singh...	CENTRAL DELHI	28.698548	77.219391
1	110002	AGCR EXTENSION COUNTER,AGCR SO,Ajmeri Gate Ext...	CENTRAL DELHI	28.698548	77.219391
2	110003	Delhi High Court Extension Counter SO,Delhi Hi...	CENTRAL DELHI	28.698548	77.219391
3	110004	Rashtrapati Bhawan SO	CENTRAL DELHI	28.698548	77.219391
4	110005	Anand Parbat Indl Area SO,Anand Parbat SO,Bank...	CENTRAL DELHI	28.698548	77.219391

Similarly we import and clean the data for bangalore.

	Pincode	Neighborhood	District	Latitude	Longitude
0	560001	Bengaluru G.P.O.,CMM Court Complex S.O,Dr. Amb...	BENGALURU	12.97912	77.5913
1	560002	Bengaluru Corporation Building S.O,Bengaluru C...	BENGALURU	12.97912	77.5913
2	560003	Malleswaram S.O,Palace Guttahalli S.O,Swimming...	BENGALURU	12.97912	77.5913
3	560004	Basavanagudi H.O,Gaviopuram Extension S.O,Mava...	BENGALURU	12.97912	77.5913
4	560005	Fraser Town S.O	BENGALURU	12.97912	77.5913

Thus at end we have two datasets ,i.e., for Bangalore and delhi. Thus now we have the required dataset to analyse neighborhoods and now we can start to analyse our data with Foursquare API.

2.3 Feature Selection and Extraction

We remove all the redundant neighbourhoods. The dataset consists of neighbourhood name, their latitudes and longitudes and their pin code and district of neighbourhoods.

We require only the Neighbourhood, Latitude and Longitude column from dataset and pin code and District column can be ignored.

Key Features:

- **Neighbourhood,**
- **Latitude**
- **Longitude**

Dropped features:

- **Pincode:** This is because Foursquare API does not consider pincodes for getting location.
- **District:** We drop this because this is same for all the values.

At end we find there are 96 neighbourhoods for Delhi dataset and 110 neighbourhoods for Bangalore dataset.

```
In [119]: delhi_df.shape
Out[119]: (96, 5)

In [120]: bangalore_df.shape
Out[120]: (110, 5)
```

We use the neighbourhood, latitude and longitude to get various features from Foursquare API. We can extract various types of venues like:

- Fun Venues
- Food Venues
- Trending Venues
- Top Venues
- Coffee venues
- Shopping Venues