Capstone Project - The Battle of Neighbourhoods (Part-1)

-SUHAS SOMASHEKAR

Strategic Location for Establishing a Restaurant

Finding the optimal location in opening an eatery in Bangalore, India using k-Means Clustering

BACKGROUND

Bangalore is the third-most populous city in India with a population of 10 million.It is referred 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 economy of Bangalore's metropolitan area have ranged from $45 to $83 billion.With its population growing rapidly, it has high requirements for quality Restaurants.

The success of establishing a new restaurant depends on several factors: demand, brand loyalty, quality of food, competition, and so on. In most cases, a restaurant’s location plays an essential determinant for its success. Hence, it is advantageous and of utmost importance to determine the most strategic location for establishment in order to maximize business profits.

DATA SECTION

Data Sources

1. The neighbourhoods of Bangalore alongside their respective postal codes and boroughs were scraped from Wikipedia.
   Geographical coordinates for each neighbourhood were extracted from https://commons.wikimedia.org/wiki/Category:Suburbs_of_Bangalore.
   We will find the latitude and longitude of each neighborhood and cluster them according to the restaurants present in each neighborhood fetched from foursquare location data. Then we will make a decision examining each cluster of neighborhoods.

Category:Suburbs of Bangalore

From Wikimedia Commons, the free media repository

Subcategories

This category has the following 59 subcategories, out of 59 total.

A

- Agara, Bangalore (2 C, 6 F)
- Arekere (5 F)

B

- Banashankari (1 C, 5 F)
- Banaswadi (2 F)
- Basavanagudi (5 C, 11 F)
- Begur, Bangalore (1 C, 6 F)
- Bellandur (1 C, 4 F)
- BEML (7 F)
- Bengaluru Pete (9 C, 4 F)
- Bidadi (2 C, 2 F)
- Bommasandra (33 F)
- Brigade Road, Bangalore (3 C, 8 F)

C

- Chandapura (4 F)

D

K

- Kettohalli (1 C)
- Kodihalli, Bangalore (1 C, 4 F)
- Konanakunte (1 F)
- Koramangala (1 C, 12 F)
- Krishnarajapura (3 C, 3 F)
- Kundalahalli (96 F)

M

- Madiwala (1 C, 6 F)
- Magadi (2 C, 10 F)
- Mahadevapura (2 C)
- Majestic (Bangalore) (1 C)
- Malleswaram (4 C, 2 F)
- Marathahalli (8 C, 1 P, 30 F)
- Mathikere (1 C)
- Murugeshpalya (4 C, 12 F)

N

- Nagarbhavi (1 C)

2.For identifying the number of restaurants in the vicinity of each neighbourhood, we will be utilizing Foursquare API, more specifically, its *explore* function. One has to register for a

Foursquare developer account to access their API credentials.

5. Use the Foursquare API to explore the neighborhoods

In [98]:
```
# define Foursquare Credentials and Version
CLIENT_ID = 'A2VC5XGM0A0G3EUICIQTHSN4KKKX14QLG2DQKSZ5AHBSIYVR' # your Foursquare ID
CLIENT_SECRET = 'X2IBNGDICER2NWFOGW0VGKD3DPWBWG3SQPWC5UG32OHGC1BQ' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: A2VC5XGM0A0G3EUICIQTHSN4KKKX14QLG2DQKSZ5AHBSIYVR
CLIENT_SECRET:X2IBNGDICER2NWFOGW0VGKD3DPWBWG3SQPWC5UG32OHGC1BQ
```

Data Clearing

Data downloaded or scraped from multiple sources will be combined into one table. If there are a lot of missing values for certain neighbourhoods, due to lack of record keeping,such values will be removed.

Few assumptions will be made to construct the dataframe:

- Only the cells that have an assigned borough will be processed; boroughs that were not assigned will be ignored.
- Neighbourhoods missing more than two census data value will be dropped.
- A column that features the group rows by neighborhood and by taking the mean of the frequency of occurrence of each category will be constructed.