BANGALORE NEW RESTAURANT SELECTION

IBM Coursera Applied Data Science Capstone Project Report-----
Rahul Muralidharan

Introduction

India is one of the most populous nations in the world (1.3B) just beginning to reap the benefits of the "demographic dividend" i.e. a large group of the population is in the working/employment bracket leading to more economic activity and growth. Bangalore often referred to as the 'Silicon Valley' of India is now a hub of startups and new employment opportunities. Many people from across India migrate to Bangalore in search of lucrative employment opportunities.

The client "JanakiRam (JR)" is a leading chain of restaurants based in North India. It has outlets mainly in Delhi and Mumbai where it has a loyal customer base. JR has enjoyed a fair amount of success which it wishes to replicate in the South. JR has chosen as Bangalore as its pilot location to launch its outlets in South India.

JR has approached us to determine an ideal location to start its foray. JR wants to ensure that the location chosen is economically feasible for setup as well as a key hub where it can start getting customers from the very beginning.

Business Problem: "Determine a location/set of locations for a restaurant in Bangalore to open a new restaurant"

The project is of utmost importance to the client JR and other restaurant players looking to expand their businesses in Bangalore. It can also be used for new business entrants to understand which localities have what types of entry opportunities to venture into.

Data

Data Source 1:

The city of Bangalore is divided into multiple areas or boroughs within which there are a set of neighbourhoods. Data on the boroughs and neighbourhoods was obtained from scraping Wikipedia.

Link : -

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Central [edit]

Name	Image	Summary
Cantonment area		The Cantonment area in Bangalore was used as a military cantonment during the British Raj in the 19th century. After the Indian independence, the area merged with rest of the city but still retains names for localities and streets used by the British as well as the colonial architecture in many buildings. Some localities which come under this area are Richmond Town, Austin Town, Murphy Town, Fraser Town, and Cox Town.
Domlur		Formerly part of the Cantonment area, Domlur has become a major hub for IT outsourcing firms and military establishments. The Embassy GolfLinks Business Park and HAL Airport are located in/near Domlur.

Central would be considered as Borough and Cantonment and Domlur would be Neighbourhoods

Data Source 2:

Geocoding: The latitude/longitude of the neighbourhoods was not easily available from the geocoder API. After multiple iterations, it was decided to populate this information manually by googling each neighbourhoods separately and finding its coordinates.

Demographics: The demographics of each locality were also determined through google search and reasonable approximations

Commercial Property Rates: Primary research with rental brokers for a few neighbourhoods was conducted to determine the average price per sqft. This was adjusted for error and approximated for other neighbourhoods

Data Source 3:

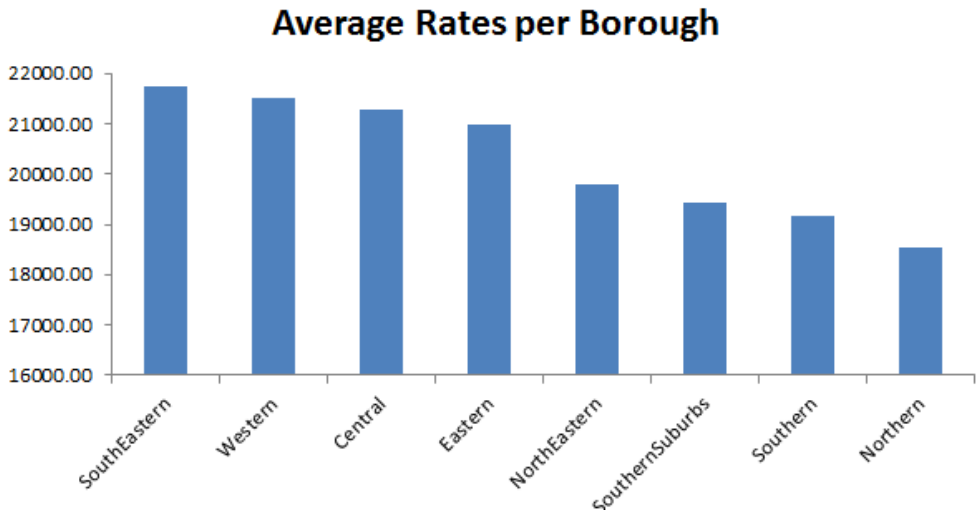
Data from Data Sources 1&2 were combined into an excel workbook '**Bangalore *Geocoding Information.xls*'** with all the information determined to be used for the actual data analysis. The excel workbook was uploaded on Dropbox. The workbook can be downloaded from the below link-

https://dl.dropboxusercontent.com/s/fan21pjc5qwjov5/Bangalore%20Geocoding%20Information.xlsx?dl=0

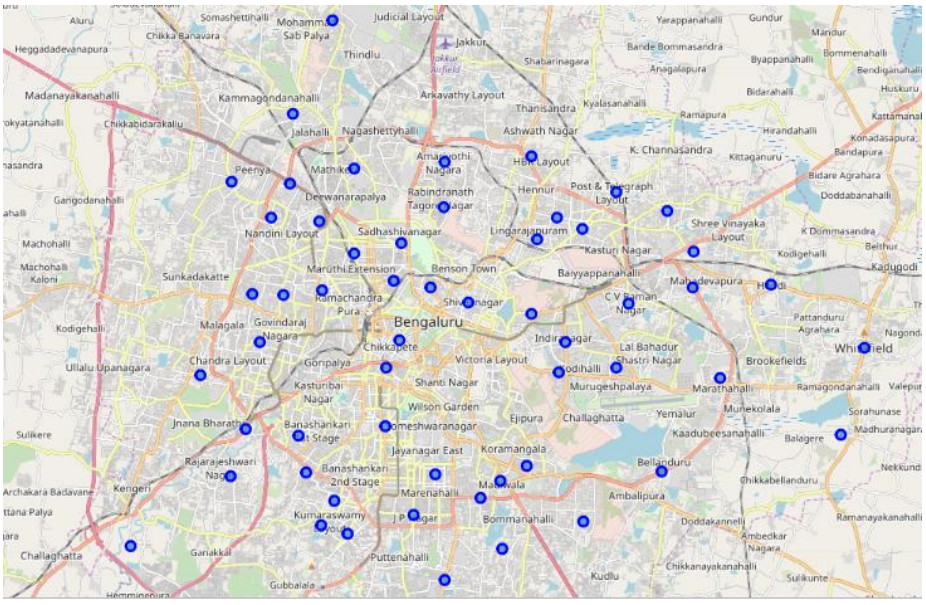
Methodology

The data was pretty clean since the dataset was prepared in-house and did not need any additional cleaning.

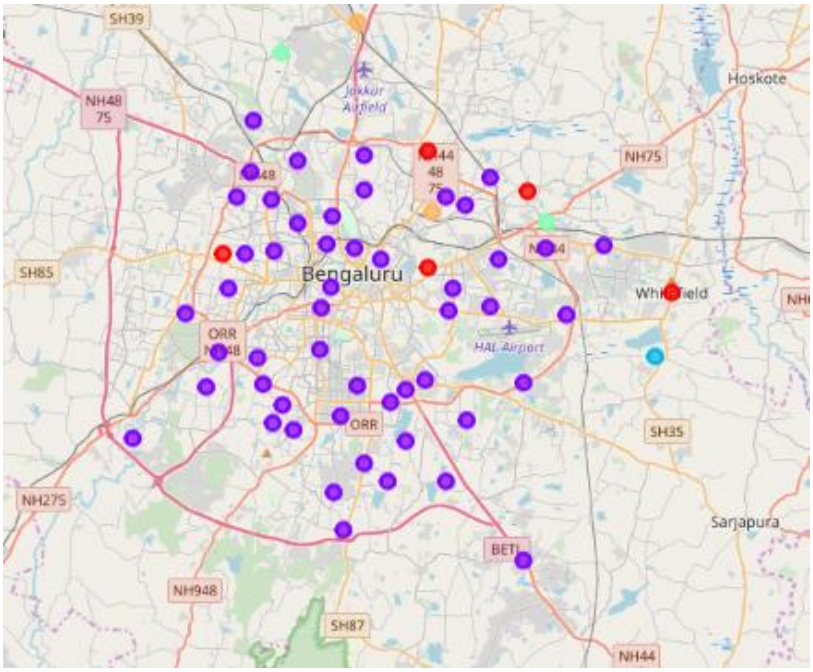
Initial exploratory analyses were conducted to understand the data better. An analysis revealed that North & South boroughs were comparatively economical than the West, East and Central boroughs



Foursquare API was used to obtain venue information for all the neighbourhoods of Bangalore. This was used to plot them on Folium



K-means classification was run on the dataset to classify neighbourhoods into clusters based on the frequency of commonly occurring venues at each neighbourhoods with k=5.



5 unique clusters were formed with Cluster 1 highlighted in purple showing restaurants as the most commonly occurring venue.

Results

The client plans to open a 2000 sqft facility and the rates were applied to the neighbourhoods falling under Cluster 1 to determine the set of locations preferred – Banashankari, Cantonment, JP Nagar, Basavangudi and Bommanahalli are the top 5 areas recommended.

Discussion

The model was built on the premise of the commercial rate viability to determine the recommended areas. This could be refined to add dimensions such as consumer preferences towards a certain cuisine, distance from schools/colleges, proximity to malls. The model employed k-means clustering. Other methods such as DBSCAN could be used to test and refine it further

Conclusion

The client JR was able to narrow down on a list of potential locations and started the process of buying discussions. This model using k-means clustering served as an effective tool to give a directional sense of the areas to be explored relying on data and not merely speculation