**Analysing Various Neighbourhood in Bangalore to Establish a Restaurant**

**Introduction**

Bangalore 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. 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. Bengaluru has one of the most highly educated workforces in the world. It is home to many educational and research institutions. Hence to open a business in such a city, one must consider many parameters. Two of the most important and essential parameters are finding a right location and type of the establishment. Hence, the objective is to open a restaurant which is located in an optimal and feasible location and choose right kind or cuisine of restaurant which can make good profit and comparatively easy to operate. Best way to get the solution is by analysing the neighbourhood i.e analyse various services present across a chosen area and based on their popularity, ratings cluster them into groups and with the help of results, right decision can be made.

**Data acquisition and cleaning**

The dataset containing pin code along with respective neighbourhood, latitude and longitude was extracted from <https://data.gov.in/> website. This huge dataset consisted data of every state and city present in India. As our aim is to extract data on Bangalore urban only, first step was to copy complete data to excel file. Filter was applied on city column and resulting data was copied to another excel file. As majority of pin-codes consist of one or more neighbourhood name, there were multiple entries and this was solved by removing duplicates as we needed one pin-code along with its neighbourhood name and location details. Once this is done file was saved as a csv file. This is the cleaned data file which is used as main input file for the analysis.

**Methodology**

As a developer, it is important to know social, and specifically, location-based services that use social platforms. Most of these services use some kind of RESTful service you can use to request JavaScript Serialized Object Notation (JSON) or XML data. Some, such as Facebook, don't require much authentication, while others, like Foursquare, have some feeds that do require authentication and some that don't.

Foursquare is chosen particularly for two reasons. The first is because it's the simplest feed to access, but unwrapping it gives you a pretty good preview of what you can expect with other feeds. The second is that it doesn't require authentication - after all, we are not looking for details on who is actually checked in, just raw counts. The venues public feed provides a great deal of functionality for this project. Foursquare is a social location service that allows users to explore the world around them. One can download the Foursquare application to their iPhone, Blackberry, or Android phone and sign up for free, then connect their Foursquare accounts to their other social media accounts.

Foursquare launched its API in November 2009, allowing application developers to extend the platform in interesting ways. Developers can build location management tools, custom search engines, and even games and other tools that interact with the Foursquare API. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server. That being said, in this article, you use a set of PHP libraries to help simplify requests and responses, but this is an entirely optional step.

Outputs can be requested in XML or JSON format, making requests to URLs that look like this*: http://api.foursquare.com/v1/user.json*. If we don't use an extension on your request, then XML is served back . A request *to http://api.foursquare.com/v1/user* results in an XML output.

There are both GET and POST methods you can use, which means you aren't just limited to reading from feeds, and also do useful stuff like checking in and creating locations using the API. As for rate limits, the application is limited to 200 requests per hour per method,

For this project following steps are carried out

* Convert addresses into their equivalent latitude and longitude values.
* Use the Foursquare API to explore neighbourhoods in Bangalore City.
* Use the **explore** function to get the most common venue categories in each neighbourhood, and then use this feature to group the neighbourhoods into clusters.
* Use the *k*-means clustering algorithm to complete this task.
* Use the Folium library to visualize the neighbourhoods in New York City and their emerging clusters.

*Importing Libraries:*

First part of the code is to import necessary libraries for smooth execution. The python libraries used are

* Numpy - library to handle data in a vectorized manner
* Pandas - library for data analysis
* Json - library to handle JSON files
* Nominatim - convert an address into latitude and longitude values
* Requests - library to handle requests
* matplotlib.cm, matplotlib.colors - Matplotlib and associated plotting modules
* KMeans - for clustering
* Folium - map rendering library

*Importing Data:*

Data is imported in IBM Watson Studio and automatically code is generated to save the values into data frames using pandas library. Exploring the data it can be seen that columns are similar to excel file and an extra missing value column having values *NaN* was created and later dropped using *drop* command.

*Use geopy library to get the latitude and longitude values of Bangalore City:*

In order to define an instance of the geocoder, we need to define a user\_agent. We will name our agent nyblr\_explorer. Next a map of Bangalore with neighbourhoods superimposed on top was created. Folium is a great visualization library. We can zoom into the map, and click on each circle mark to reveal the name of the neighbourhood. However, for illustration purposes, we simplify the map by segmenting and clustering only the neighbourhoods in Yelahanka new town area. So we slice the original dataframe and create a new dataframe of the Yelahanka new town data. As we did with all of Bangalore City, let's visualize Yelahanka new town and neighbourhoods in it.

*Define Foursquare Credentials and Version:*

Next, we are going to start utilizing the Foursquare API to explore the neighbourhoods and segment them. First we get the neighbourhood's name and then the neighbourhood's latitude and longitude values. Then we get the top 100 venues that are in Yelahanka new town within a radius of 500 meters. Next we send the GET request and examine the results. We know that all the information is in the *items* key. Before we proceed, the get\_category\_type function from the Foursquare lab is borrowed. Now we are ready to clean the json and structure it into a *pandas* dataframe and check how many venues were returned by Foursquare?

*Explore Neighbourhoods in Yelahanka new town:*

We create a function to repeat the same process to all the neighbourhoods in Yelahanka new town. Then we write the code to run the above function on each neighbourhood and create a new dataframe called *yelahanka\_venues* and then check the size of the resulting dataframe. Next we check how many venues were returned for each neighbourhood and then find out how many unique categories can be curated from all the returned venues.

*Analyze Each Neighbourhood:*

First we start with one hot encoding. Next, we group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category. Then print each neighbourhood along with the top 5 most common venues and put that into a pandas dataframe. Function to sort the venues in descending order was written and new dataframe and display the top 10 venues for each neighbourhood was created.

*Cluster Neighbourhoods:*

Final step is to Run k-means to cluster the neighbourhood into 5 clusters. And then we create a new data frame that includes the cluster as well as the top 10 venues for each neighbourhood. Finally, the resulting clusters is visualized.

**Results**

After conducting the above steps, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster. Let us name them Cluster 1 to 5 for easy understanding.



**Fig 1. Cluster 1**

Above figure illustrates region of cluster 1 along with most 10 common areas in the neighbourhood. For this type of venue is fast food restaurant followed by Indian restaurant. And least common venue is dessert shop.

Similarly in case of cluster 2 which also contains few regions of Bangalore north and east, most common venue is farm, followed by wine bar and the least common spot visited by crowd are event spaces.

In case of cluster 3, which lies in the outskirts of Bangalore, most common venue visited are the Convenience stores and next most common visited place is flea market. The least favourite venue is the farm.

Cluster 4 is the most important one for us. This cluster covers the area where we wish to establish our restaurant i.e Yelahanka new town. The most favourite venue visited here are Indian restaurant and followed by pizza place and hotels. Least visited place is hockey arena.



**Fig 2. Cluster 2**

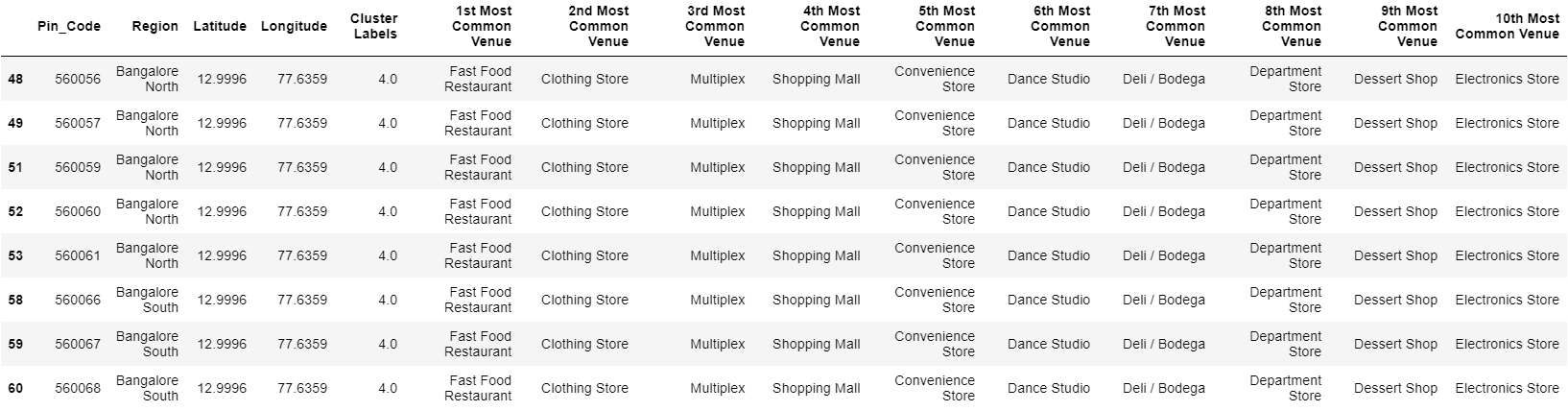


**Fig 3. Cluster 3**



**Fig 4. Cluster 4**

In case of cluster 5 , fast food restaurant is preferred the most. Followed by clothing store and multiplexes. Electronics store takes the last position.



**Fig 5. Cluster 5**

**Discussion**

Above discussed are the results of the k means clusters operated on the foursquare data. Next step is to draw a solution to our objective using above mentioned results. Yelahanka new town has pin code 560064 which lies in cluster 4. In that cluster the most favourable venues are Indian Restaurants, Pizza place and hotel. If we choose to establish one of the above mentioned venues, the competition would be high as there are many similar venues already established in the neighbourhood. On the other hand the least favourable venue is Hockey arena, if we choose to open this kind of venue the competition would be less but the probability of it becoming successful and profitable is low. Third option is to choose the one which is in between i.e 5th most visited venue is coffee shop and 4th one is a bakery. This option promises to be successful, not as good as an Indian restaurant but the competition is less. If we can establish an coffee shop along with a blend of bakery items this idea could be unique and in the initial stage the sales might be average but we can be sure that it wont be worse.

**Conclusion**

Opening a business in our case a restaurant needs a lot of planning and research work. The basic parameter is the location and deciding type of restaurant. This project helps us in analysing these issues and draw a conclusion based on machine learning method i.e k means clustering. K-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centres or cluster centroid), serving as a prototype of the cluster. The observations are imported from foursquare api which contains details of venues present in a neighbourhood. The neighbourhood is downloaded from government of India website which contains pin code along with latitude and longitude of areas present in Bangalore urban. 5 is value of k chosen for optimal solution and the venues are divided into 5 clusters. The visualization of clusters helps us decide whether the assumed location for establishment is feasible. Later the outcomes of clustering helps in drawing a conclusion regarding what type of restaurant would be successful. Based on these parameters we have decided to open a coffee shop along with a bakery in Yelahanka new town area of Bangalore