**IBM Data Science Capstone Project**

**Battle of the Neighbourhoods**

Shaju Unnikrishnan Nair

**Table of Contents**

Introduction 3

Business Problem description 3

Approach Taken 3

Step 1: Data identification and acquisition 4

Step 2: Data wrangling (preparing data for further use) 4

Step 3: Data analysis and visualization 5

Cluster 1: 5

Cluster 2: 6

Cluster 3: 6

Cluster 4: 7

Word Cloud - 1 7

Word Cloud - 2 8

Observations 8

Recommendations 9

# Introduction

Bangalore has one of the most vibrant startup eco-systems in the world. A lot of activity is happening with respect to new ideas, funding, startups, eco-systems and the like.

Food based startups are getting the maximum traction – be it restaurants or food delivery solutions. Venture capital funding is also generous on food based startups – perhaps the population density of Bangalore (and India at large), the demographic dividend that exists (with a lot of millennials entering the work force who are not interested in cooking their own food) and the increasing amount of disposable incomes, makes it a big market that is available to tap into by these food based startups.

# Business Problem description

A new startup in the "pocket-friendly menus" restaurants is wanting to identify the best places to open their outlets around the city of Bangalore. The menu will be multi-cuisine in nature and so will appeal to a wider audience. This approach will also give the startup the opportunity to tweak their menu based on seasonality, local preferences and changing customer tastes.

They have given me a mandate to identify the best places to open these outlets. Their mandate is to identify locations based on other competing outlets / venues and customer feedback for these outlets. Their desire is to open stores where there is less competition and people are not happy with the current options available.

Customer has also asked us to advise which type of cuisine and what type of restaurant should be opened.

# Approach Taken

To solve this, it’s imperative to collect the right kind of data and segment the market into clusters based on availability of eating outlets, the feedback of their customers. More importantly, this has to be done using a secondary research approach by using information freely available over the internet as the customer does not have the budget to do a primary research that includes field study and meeting with actual customers.

Broadly, there are 4 steps that will be taken:

**Step 1:** Data identification and acquisition

**Step 2:** Data wrangling (preparing data for further use)

**Step 3:** Data analysis and visualization

**Step 4:** Drafting Recommendations

## Step 1: Data identification and acquisition

Clearly, one needs the following data:

* Geographical data about Bangalore (Lat/Long information)
* Name of competing providers
* Location of their operations
* Type of services they provide
* Customer responses to their services / products

This can be acquired from the following sources:

* Get Geo data for Bangalore from geolocator
* Using Foursqure APIs, get the following about the current providers
  + Venue Name
  + Venue ID
  + Venue Location
  + Venue Category
  + Number of Likes received

**Note:**

“Likes” were taken as a basis for decision making. This is because if an outlet gets a higher number of “Likes”, then clearly it stands out in terms of customer satisfaction, quality and other parameters that are important for the customers.

It also means that if the company were to set up a restaurant in a location where there are other restaurants with a lot of “Likes” then the chances of the new restaurant succeeding is very less. Therefore it makes sense to open new restaurants where there is less competition and other restaurants have lower number of “Likes”.

## Step 2: Data wrangling (preparing data for further use)

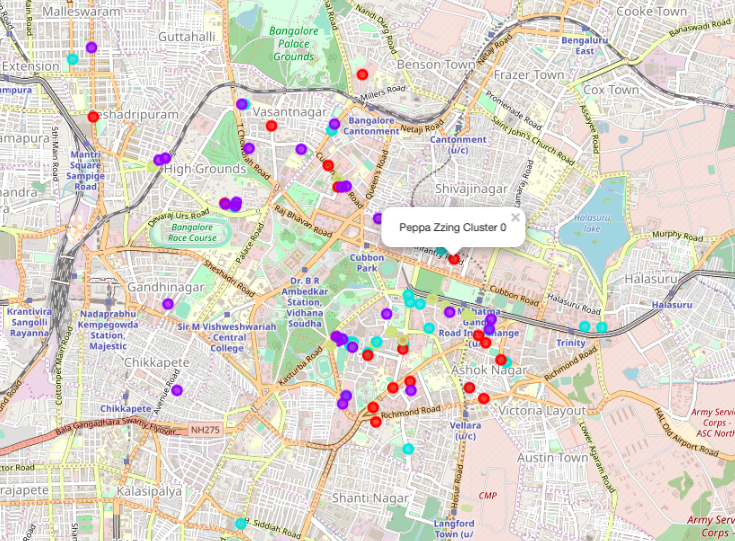
The following steps were taken to sanitize and clean up the data for further use during the analysis phase:

1. Collect the venues data into a Dataframe
2. Collect the Likes data into a different Dataframe
3. Combine dataframes to create a new one that has data about venues as well as the Likes
4. Collect the Remove data that is not relevant to our research
5. Merge categories to eliminate overlapping categories so that the data becomes more consistent
6. Convert categorical data into numeric binaries

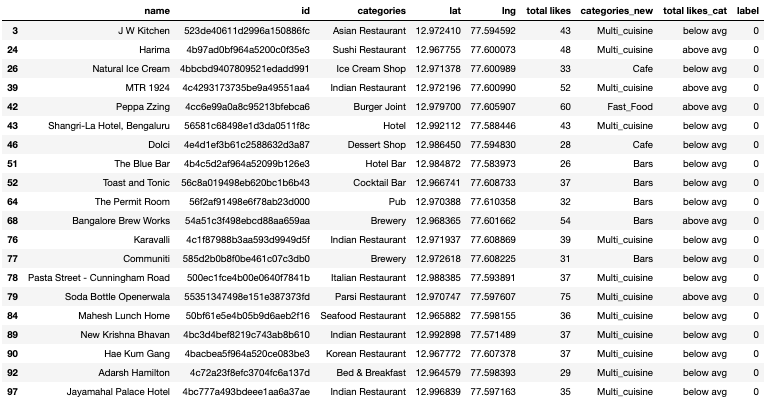
## Step 3: Data analysis and visualization

KMeans clustering was used to cluster the venues based on the likes that they’ve received.

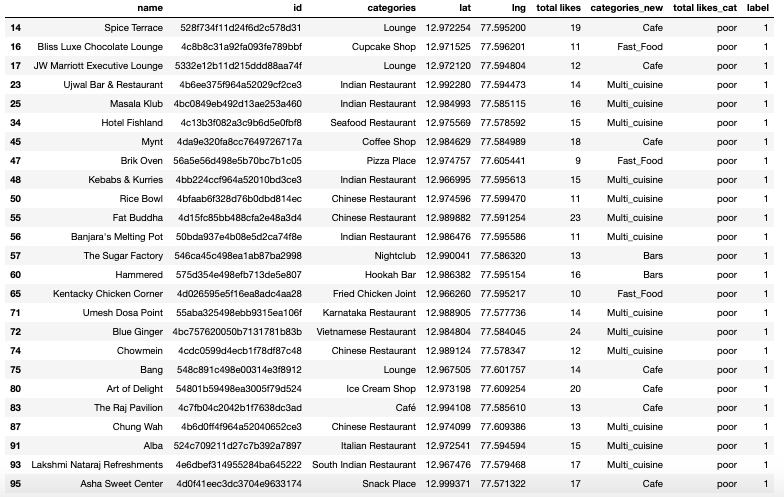
I created 4 clusters. The diagram below shows the 4 clusters superimposed on the map of Bangalore.



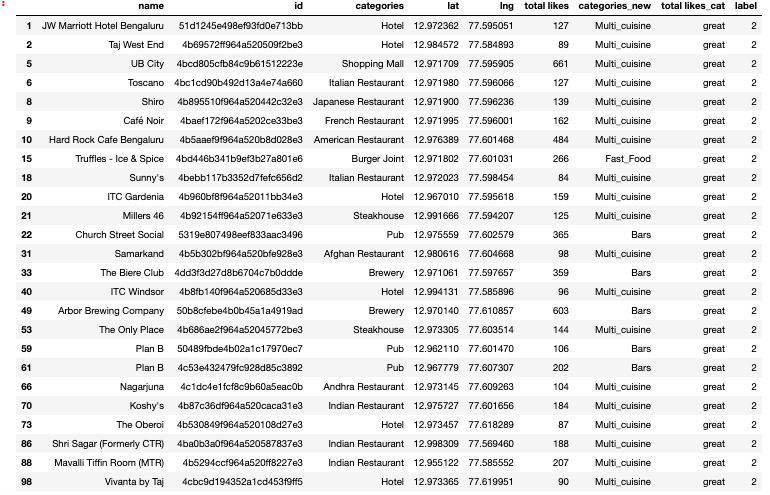
### Cluster 1:



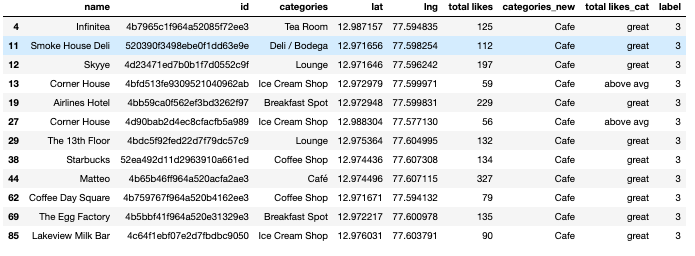
### Cluster 2:



### Cluster 3:



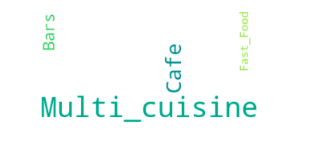
### Cluster 4:



### Word Cloud - 1

To identify the type of cuisine and the type of restaurant that they should start, we’ve also done a word cloud implementation.

The word cloud below clearly indicates that “Multi-cuisine” restaurants are more preferred by the potential customers in Bangalore.



From the analysis above, the following recommendations are being made :

1. It would make sense to open restaurants in and around restaurants in Cluster 2 to maximise on the lack of quality of competing eatries

2. Restaurant must cater to multi-cuisines

3. Focus should be on Indian cuisine

### Word Cloud - 2

The word cloud below clearly indicates that restaurants serving “Indian” food is much in demand. This was already known. However, given the cosmopolitan nature of the population of Bangalore, the customer wanted to be sure whether having yet another restaurant serving Indian food would be a good idea or not. Clearly, the data indicates that it would be.



# Observations

From the analysis above, it is evident that :

* + Not all 5 and 7 star hotels have good quality of service (based on the less number of likes they've received)
  + Cluster 1 has the best mix of eating joints – but the ratings are not encouraging
  + Cluster 2 has the maximum number of joints with a rating of “poor”
  + Cluster 3 has the best restaurants in the city
  + Cluster 4 is largely composed of Cafes and they provide decent service
  + Multi-cuisine restaurants are preferred
  + Indian food still remains the hot favourite among the population

# Recommendations

From the analysis above, the following recommendations are being made :

1. It would make sense to open restaurants in and around restaurants in Cluster 2 to maximise on the lack of quality of competing eatries
2. Restaurant must cater to multi-cuisines
3. Focus should be on Indian cuisine