Identifying market opportunity for restaurant business based in Bangalore

Usha Priyadharshini

31st August 2019

1 Background

Bangalore is a cosmopolitan city in India with population of over 12.5 million and roughly 700 km^2 and is the capital of Indian state, Karnataka. Being a resident of this city, I decided to choose Bangalore to do my analysis project. As can be seen from the figures mentioned above, Bangalore is a city with high population and population density. Moreover, Bangalore being IT hub of India, the residents of Bangalore come from different parts of the country. In fact, residents of the city are not only from India – There is a sizeable population from all over the world. It is often said that the migrants from outside of Karnataka outweighs the locals from the state.

1.1 Problem and Interest

With such an ethnically diverse population and most of them office goers, restaurant business has become a thriving business – albeit a difficult one. One of the difficulties arise from catering to the multitudes of taste that cuisines from different parts of India has to offer. Also, with globalization and establishment of multinational food chains, dwellers of Bangalore have found their taste in cuisine from all over the world. Since it is difficult for a restaurant business to offer all possible cuisines, it is valuable to identify if there is a demand supply shortage for a particular cuisine category and then cater to that market

Bangalore is one of the largest cities in India and is known for its notorious traffic jams. Given this, it is less likely for customers living far from the restaurant location to visit, leave alone repeated or frequent visits. Therefore, it is important to find a location that has lesser number of restaurants addressing the target market that we identify so that the competition is reasonable to cater to the population living in and around that location

In this project, we attempt to address these key questions for aspiring restaurant owners in the city of Bangalore:

- **Is there an opportunity in restaurant business in Bangalore that is not adequately addressed?**
- **If such an opportunity does exist, what would be the ideal location in the city to start our venture so that competition is minimized?**
- **From available data, can we identify the potential customers for our new venture?**

2 Data Acquisition

In order to identify the target market, we need two data sets – one to analyze the suppy and the other to understand the demand. To analyse supply, we need information on restaurants serving different cuisines in Bangalore. This information can be obtained from foursquare using the search

API to search for categories under 'Food'. From the response we extract the venue ID, name, location and the type of restaurant that indicates what type of cuisine is served in the restaurant. Similarly, to understand demand, we need information on preference of residents for different type of cuisines. This information can also be obtained from foursquare likes API that gives the like count for a venue. Aggregate of likes for venues under a specific category is a good indicator of customer preference for that cuisine type. Though may not exactly indicate the number of visitors, aggregate like count for a category has high correlation to the number of visitors preferring a particular cuisine category

In order to identify the target location, we need locality information of Bangalore and the number of restaurants of target category in that location. For locality information, we use Bangalore municipality geo division data from:

- https://raw.githubusercontent.com/datameet/Municipal_Spatial_Data/master/Bangalore/BBMP.GeoJSON

To find the number of restaurants of target category in each municipal location, we use the foursquare search API to find restaurants of target category that are located within a specified radius of each municipal division headquarter location

In order to identify potential customer, we use likes API from foursquare. From the likes API, we can extract information on different cuisine categories that a particular user liked and use that to segment users based on their preference for cuisine types and identify potential customer for our new venture

3 Methodology

In the following sections, we detail the methodology employed to address the key questions listed in section 'Problems and Interest'

3.1 Supply analysis: Understanding existing restaurants serving different cuisines

In order to establish a new venture in restaurant business we need to have a grasp on the existing restaurants in terms of the cuisine it serves and for each cuisine type, the number of restaurants serving that particular cuisine type. Using foursquare search API, we find this information on existing restaurants in and around Bangalore, within a radius of 50 km. The data collected has information on venue ID, venue name, location details and the category of cuisine it serves:

	Venue ID	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	4ba0b3a0f964a520587837e3	Shri Sagar (Formerly CTR)	12.998309	77.569460	Indian Restaurant
1	4b92154ff964a52071e633e3	Millers 46	12.991666	77.594207	Steakhouse
2	4f0b2674e4b0ca3a5efdaeaf	Ice N Spice by truffles	13.028079	77.573116	Burger Joint
3	4b5aaef9f964a520b8d028e3	Hard Rock Cafe Bengaluru	12.976389	77.601468	American Restaurant
4	4bd446b341b9ef3b27a801e6	Truffles - Ice & Spice	12.971802	77.601031	Burger Joint

Then we group the restaurants based on venue category. Since the venue ID is unique, grouping by venue category and getting unique value will list the number of available restaurants for that category:

Venue Category	Venue Count
Indian Restaurant	16
Café	10
Italian Restaurant	7
Bakery	5
Burger Joint	5

Then we sort the restaurant category based on the number of available restaurants and plot a bar chart diagram to visualize the number of available restaurants per category

It is possible that we have concentrated population of certain ethnicity in particular localities so that restaurants of particular category are more in number in those localities as the customers in those locations have higher preference for certain category of restaurants. For example, if we have concentrated population of people from say state of Gujarat in particular location, then there is a high likelihood of more restaurants serving Gujarati cuisines in that locality. For our analysis we would like to weed out any such outliers.

The data we have is limited for such an analysis and the limitations are discussed in 'Discussion and Observation Section'. However, for the sake of completeness, we analyse the available data by considering only restaurant categories where enough representation exists in the data set. We use python folium library to visualize the location details of each restaurant category – Each restaurant category is represented with different colour and we superimpose the venue locations to analyse if a restaurant category is spread across the city of Bangalore.

3.2 Demand analysis: Understanding user preference for different cuisines

Next, we perform demand analysis by understanding the user preference for different cuisines and identify the high demand category. To achieve this, we use four square APIs, and find the number of likes for the restaurants and group them over categories to find the number of likes for a particular type of category:

Venue Category	Like Count
Indian Restaurant	1002
Burger Joint	873
Breakfast Spot	737
Café	618
American Restaurant	483

We merge the data on venue count per restaurant category (described in earlier section 'Supply analysis') with the like count per restaurant category:

Venue Category	Venue Count	Like Count
Indian Restaurant	16	1002
Café	10	618
Italian Restaurant	7	407
Bakery	5	207
Burger Joint	5	873

For each category, we plot the number of restaurants and the number of likes to understand if the demand for a particular category outweighs the current available restaurant count. As the two features differ by an order of magnitude in our data set (restaurant counts are of the order of tens, while the like counts are of the order of hundreds), we do linear normalization of both features by considering minimum value as zero. Linear normalization ensures features in the graph remains comparable.

3.3 Competition analysis for our new venture

Once we identify the target market in terms of the cuisine type(s) we would like to serve in our new venture, the next logical step is to identify location where the competition is not intense for the target category of restaurants. In order to identify the location, we segment the Bangalore city into its municipal division and find the number of restaurants available for the target restaurant category in each municipal division.

The geoJSON file we used, segments the city of Bangalore into 198 municipal divisions, and for each division lists the co-ordinates of division headquarter location. We use the foursquare search API to find the number of restaurants of target category that lies within the radius of 1km from the headquarter location. For the purpose of illustration, we have used the target category of 'American restaurants and Burger Joints'. Once we obtain the data on number of restaurant of target category of interest per municipal division, we merge that information with the municipal division data:

	Name	Latitude	Longitude	Venue Count
0	Shantala Nagar	12.969867	77.606464	18
1	HSR Layout	12.913718	77.646426	17
2	Hoysala Nagar	12.978456	77.638783	16
3	Jogupalya	12.973725	77.632594	16
4	Domlur	12.963361	77.638040	15

We can now use this data to get a choropleth map to visualize the competition intensity in different municipal location in Bangalore for restaurants serving American cuisine and/or burgers

3.4 Data analysis for potential customer identification

Once we identify the market opportunity and the competitive advantage on location, we prepare to launch our new venture. For our new venture to be successful, we need to reach potential customers as quickly as possible. In identifying potential customers, our aim is to maximize the likelihood of turnouts for the invitation we send. If we can get information on users who would be interested in the cuisine(s) our restaurant offers or cuisines similar to what our restaurants offer, that would be helpful in spreading news about our new venture.

To identify the users who has interests in cuisines similar to what we offer, we perform customer segmentation based on the cuisines they have liked in the past. Using four square APIs, we find users and the category of cuisine they have liked. Since the cuisine type is a categorical variable, we convert it into one-hot variable across the available categories:

	User ID	Afghan Restaurant	American Restaurant	Andhra Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bistro	Breakfast Spot	Burger Joint	...	Multicuisine Indian Restaurant	Pakistani Restaurant	Pizza Place	Resta
0	101642791	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
1	103103043	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
2	104972480	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
3	106466411	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
4	107105268	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	

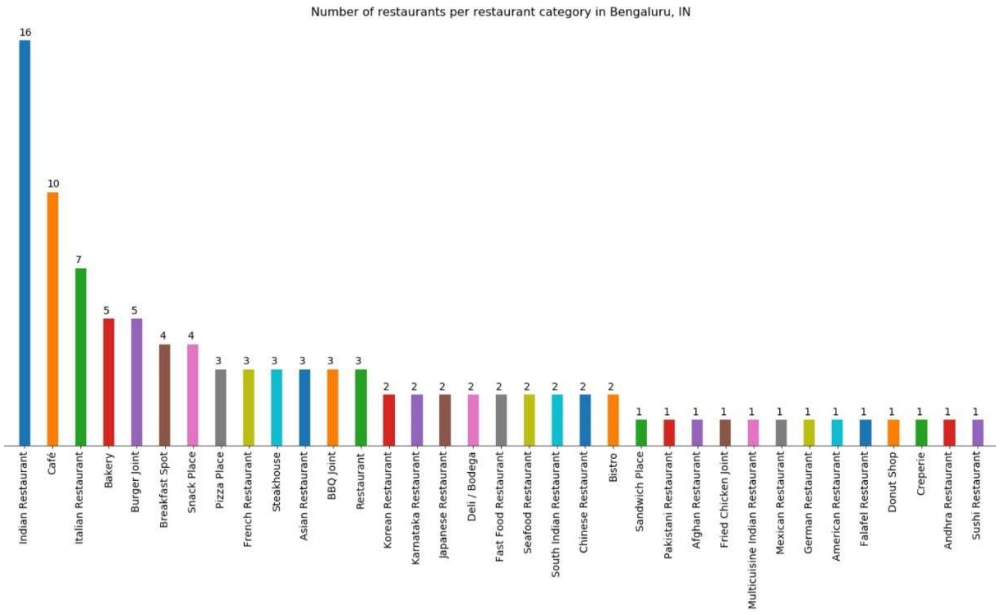
We perform K-Means clustering to segment the users into groups and analyse the groups to identify the potential customers

4 Results

In the following sections, we detail the results from the various analysis we described in the methodology section

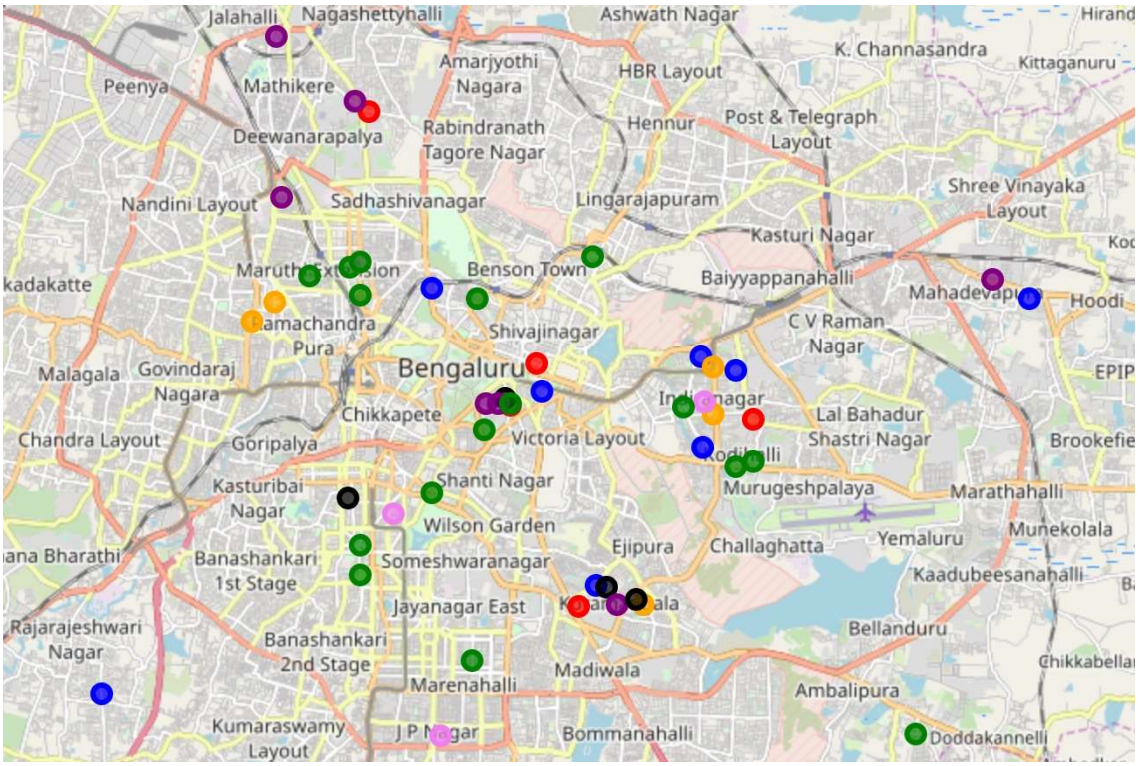
4.1 Results of supply analysis

To understand the market dominance in terms of the cuisine being served, we plot a bar chart diagram of number of available restaurants per category:



From the plot, it is clear that market is dominated with coffee shops and restaurants serving Indian and Italian cuisines. Unless there is an overwhelming demand, it might be difficult to penetrate this market.

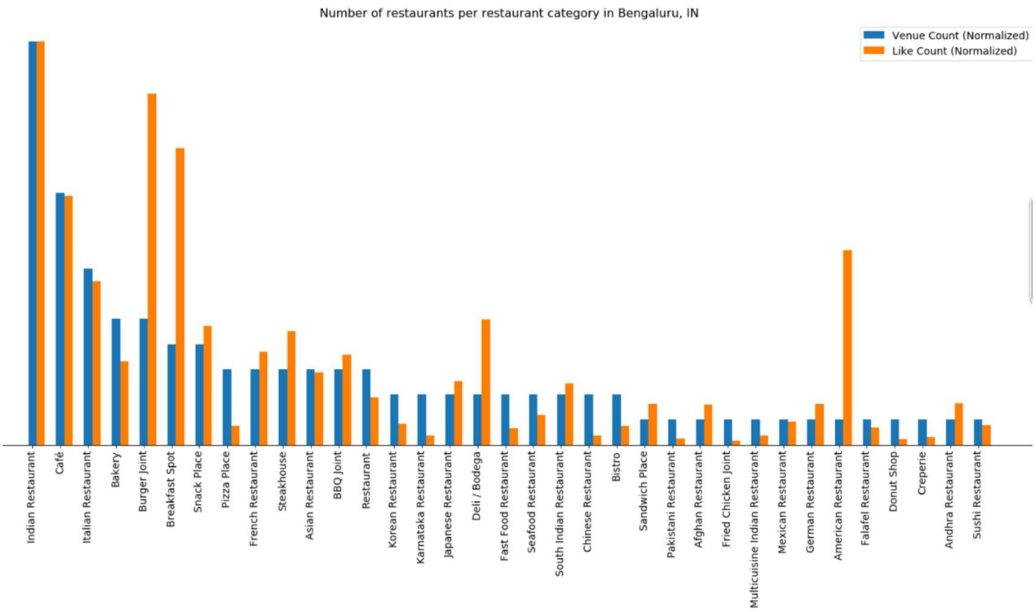
Further, we create a map of Bangalore, with the venue of restaurants of top categories (in terms of number of available restaurants for that category) superimposed on the map. Each category is represented with different colour to differentiate one category from the other:



From the map, we could observe that each category has moderate spread all over Bangalore. Though the analysis is performed on limited data, it is reasonable to conclude no particular restaurant category is concentrated in specific locality. Bangalore is indeed a true cosmopolitan city as far as its taste buds are concerned!

4.2 Results of demand analysis

To understand the supply-demand shortfall for a particular type of cuisine, we plot, for each restaurant category, the number of available restaurants and the number of likes for that restaurant category. Both features are linearly normalized.



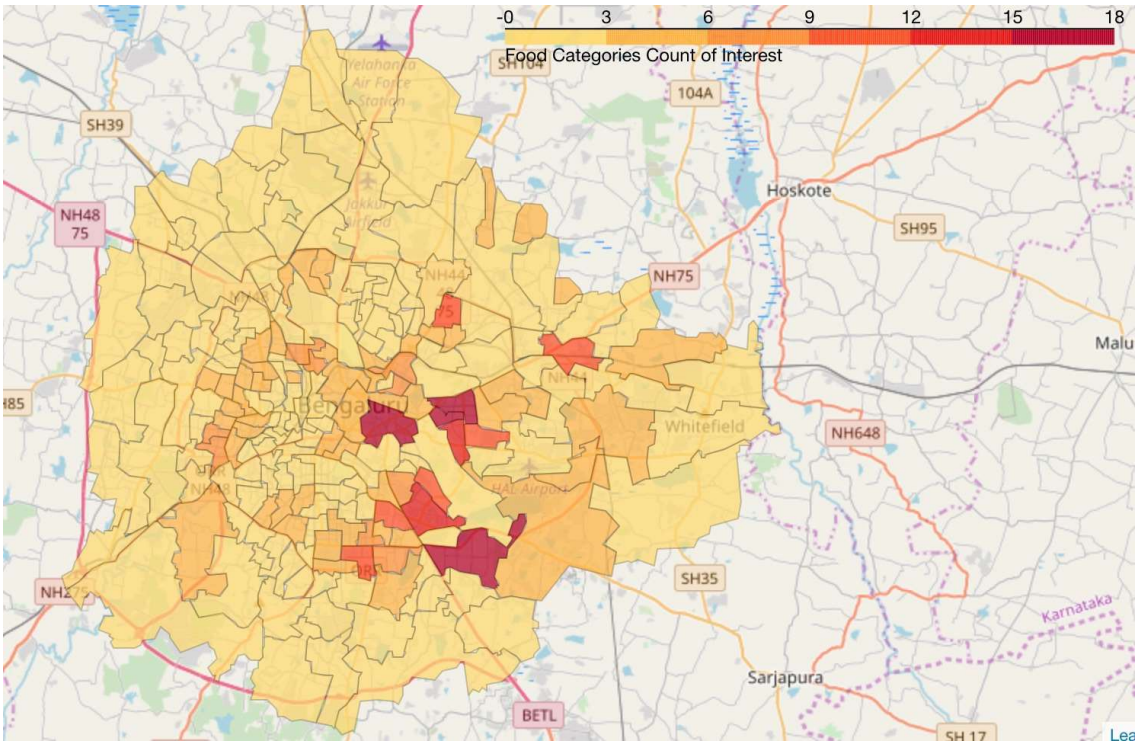
From the plot, we observe that there is a clear shortfall of restaurants that serve the following categories of cuisine as the normalized count of number of likes heavily outweighs the number of available restaurants in that category:

- Restaurants serving American cuisine and Burgers
- Restaurants serving breakfast

For the categories mentioned above, demand far exceeds the supply and hence offer a good market opportunity for aspiring restaurant owners.

4.3 Results of competition intensity analysis

To analyze the market competition for the opportunity we identified in earlier section, we use a choropleth map that shades each municipal division in the city of Bangalore based on the number of target restaurant category in that municipal division. Of the two high demand category we have identified earlier, we choose the 'American cuisine and Burgers' category as our point of interest for further analysis. Given below is the choropleth map for our chosen high demand category:



From the map, we observe that the restaurants serving American cuisine and burgers are more prominent in central, south and south western parts of Bangalore. As such northern Bangalore (extending from north east to north west) is a probable location candidate.

4.4 Results of data analysis for potential customer identification

K-Means clustering we performed on users, segmented the users into 5 groups based on their past interest for a particular type of cuisine. We analyzed the clusters by noting the most common

interest of cuisine categories for the users of the cluster. From the analysis, we infer that the users are segmented along the following line of interest on cuisine categories:

- Cluster zero users have interest in different types of cuisine
- Cluster one users have a special taste bud for coffees
- Cluster two users like to eat out for breakfast
- Cluster three users like Indian food
- Cluster four users who predominantly like Italian food, but also have an inclination for other foreign cuisines

Based on the inference we made on cluster data, we conclude that users of cluster zero and cluster four have high likelihood of turnout.

5 Discussion and Observations

For the initial analysis, we have limited data. Four square API's return only limited number of results for search APIs. While the current analysis does give us some insights, for a more meaningful analysis, we need to get search results with close-by locations (say a search per municipal division) and then merge those results weeding out the duplicates. An analysis based on such consolidated data can throw more insights. With the limited queries per day for foursquare APIs, we found this difficult to implement – but future work could consider this enhancement

To understand if population with specific taste buds are concentrated in a particular locality, we have marked the specific restaurant category in the map and have looked for spread. But for such an analysis, the data we had is insufficient as many restaurant categories had minimal sample points. Enhancing the data collection through mechanism mentioned above may provide better analysis

To understand demand, we count the likes per category. But since the counts are based on venues if our search returned more 'good' restaurants for a category and more 'bad' restaurants for a different category, then our demand inference may be biased. For the analysis presented in this report, we pretend such a bias does not exist. This problem will also be alleviated to a great extend with the data collection enhancement discussed earlier

We have based our analysis on the existing data available to us from four square and as such no inference can be made on data that is not available from this data set. For example, how do residents of Bangalore like 'Hawaiian food' for which our search did not return a result, possibly because one does not exist in Bangalore? Data need to be augmented from reviews of food festival etc. rather than basing it only on foursquare location data to alleviate this short coming

6 Conclusions

From the analysis we performed, we identify two market opportunities for opening a restaurant in Bangalore

- Restaurants that serves American food and Burgers
- Breakfast spots

We also observe that the northern (extending from northeast to northwest) Bangalore is less competitively intense for restaurants serving American cuisine and/or burgers and therefore an ideal candidate for opening our new venture.

Based on the cluster analysis of segmentation of users, we find that users who has a liking for multi-cuisine and Italian cuisine has higher chances of visiting our restaurant and generating business for our new venture. Therefore, we must target and lure these customers with discounts and rewards.