

**NAME: Sidharth Kumar Mohanty**

**DATE: 29<sup>th</sup> May 2020**

**TOPIC: RESTAURANT RECOMMENDER SYSTEM Blog**

### **post link**

I have written a blog post about this project. Here is the link :

<https://medium.com/@sidharth.ku178/restaurant-recommender-system-80c5ee2b9c35>

### **REVIEW LINK**

This is the review link to the caption proposal project

<https://review.udacity.com/#!/reviews/2319744>

### **Domain Background**

Bangalore is the third leargest city in India. The diversity of the cuisine available is reflective of the social and economic diversity of Bangalore. Roadside vendors, tea stalls, South Indian,

Muslim food, Chinese and Western fast food are all very popular in the city. Udupi restaurants, are very popular and serve predominantly vegetarian cuisine. The Chinese food and the Thai food served in most of the restaurants are can be customized to cater to the tastes of the Indian population. Bangalore can also be called a foodie's paradise because of its vast variety of foods and edibles with a touch of Bangalore's uniqueness and tradition.

### **DATASETS**

Initially I scrapped list of neighbor's using BeautifulSoup4 from this given link [wikipedia]([https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Bangalore](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)).

T he table headings becoming the boroughs and data becoming the neighborhoods. Bangalore has 8 boroughs and 64 neighborhoods. So i manually googled each neighborhood to find its corresponding latitude and longitude. After doing so, I produced the following data frame.

Borough	Neighborhoods	Latitude	Longitude
Central	Cantonment area	12.972442	77.580643
Central	Domlur	12.960992	77.638728
Central	Indiranagar	12.971891	77.641151
Central	Jeevanbheemanagar	12.962900	77.659500
Central	Malleswaram	13.003100	77.564300
Central	Pete area	12.962700	77.575800
Central	Rajajinagar	12.990100	77.552500
Central	Sadashivanagar	13.006800	77.581300
Central	Seshadripuram	12.993500	77.578700
Central	Shivajinagar	12.985700	77.605700

Income by neighborhood is again easy to find out given that it's readily available. But in case of Bangalore, it is again not the case. i was able to find Income data for main city. This is the link .

[\]\(https://en.wikipedia.org/wiki/List\\_of\\_Indian\\_cities\\_by\\_GDP\\_per\\_capita\)](https://en.wikipedia.org/wiki/List_of_Indian_cities_by_GDP_per_capita).

Neighborhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The data frame for Bangalore neighborhood population looks like:

	Borough	Neighborhoods	AverageIncome	Normalized_income
0	Central	Cantonment area	18944.099792	0.293051
1	Central	Domlur	56837.022198	0.879225
2	Central	Indiranagar	41991.817435	0.649581
3	Central	Jeevanbheemanagar	6667.447632	0.103140
4	Central	Malleswaram	53270.063892	0.824047

Population by neighborhood is again easy to find out given that it's readily available. But in case of Bangalore, it is again not the case. i was able to find population data for few cities. [Here is the link](<https://indikosh.com/dist/655489/bangalore>). Rest other neighborhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The data frame for Bangalore neighborhood population looks like:

	Borough	Neighborhoods	Population	Normalized_population
0	Central	Cantonment area	866377	0.880810
1	Central	Domlur	743186	0.755567
2	Central	Indiranagar	474289	0.482190
3	Central	Jeevanbheemanagar	527874	0.536668
4	Central	Malleswaram	893629	0.908516

## **Target Audience:**

Target audiences for this project does not limit to a person who keeps travelling but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food. People who rarely use restaurants would prefer to have the most rated restaurants nearby them and all this could be easily handled by our recommender system. So target for this project is basically everyone who is exploring different places or similar places

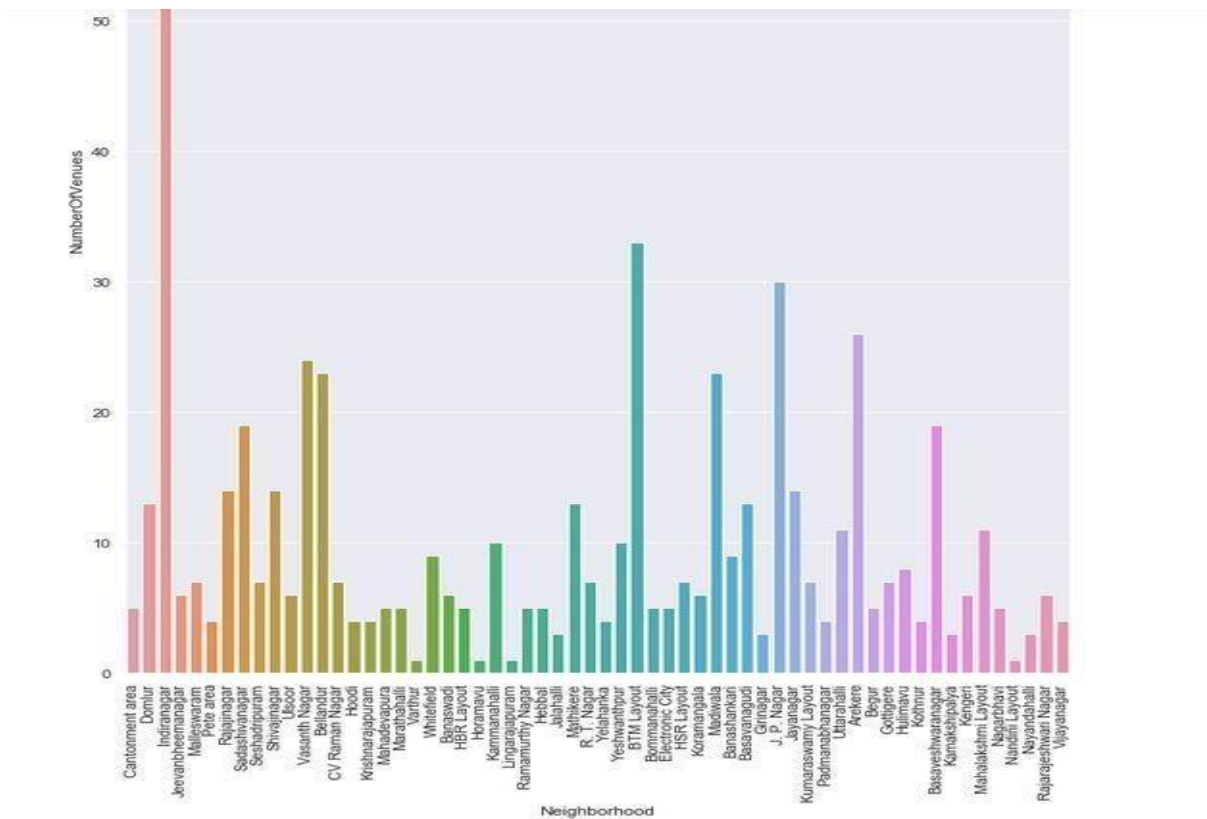
### **Success rate:**

With restaurants evolving, new food categories emerge, hybrid food starts to be more popular, we need a system that could help us access vast number of food varieties. It is impossible for a person to ask each and every one about their visit to a particular place and also not everyone remembers everything. On the other hand, Computers are good at remembering things, and with Machine learning to its peak, it high time technology will by our personal guidance and help us personally based on our likes and dislikes. So people would care about this project as their personal assistance and success rate could certainly increase with time.

### **Exploratory analysis:**

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

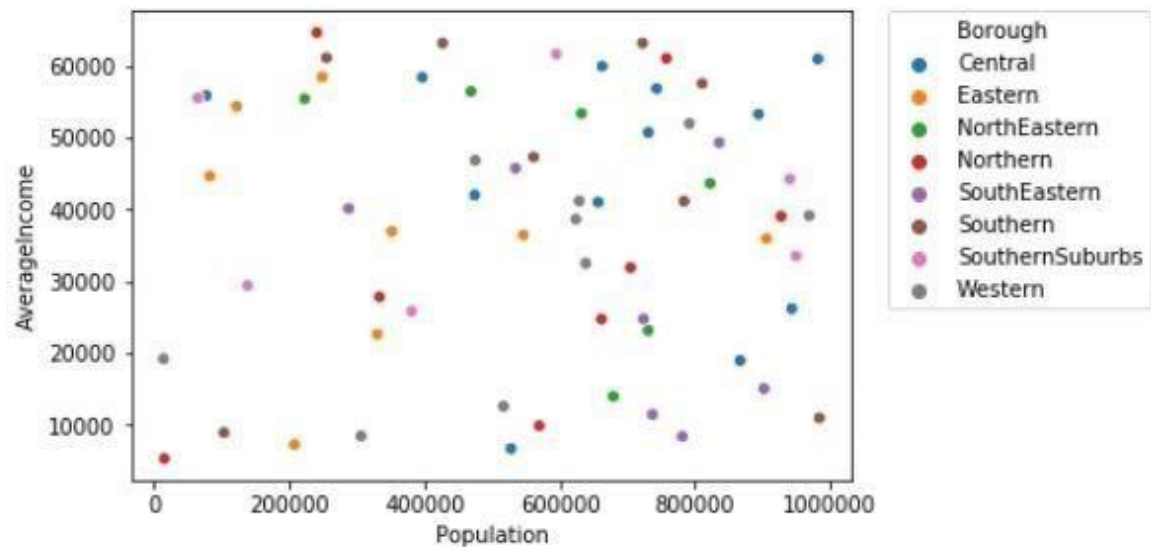
Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data. While exploring the dataset, I found out that Inderanagar has most number of venues while Varthur has the least.



## Inferential analysis:

Most important factors while building the recommender system were population and income. They are the most important factor because they have a nonlinear relationship according to our dataset.

It needed to make some inferential analysis to understand this nonlinear relationship. As the amount of population increases, it does not necessarily mean that average income of a neighborhood will also increase. It is true to most of the case but also many cases differ to follow this trend. Similarly, a neighborhood with less number of people may not necessarily have less average income. It is possible to have less number of people and more income and vice versa. This can be inferred from the following graph:



## RESULT

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking ‘Whitefield’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘Whitefield’.

The following image shows the result:

Out[211]:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Basaveshwaranagar	Venue Category_Ice Cream Shop	Venue Category_Indian Restaurant	Venue Category_Fast Food Restaurant	[0.6426377807870477]
1	Begur	Venue Category_Indian Restaurant	Venue Category_Indian Sweet Shop	Venue Category_Food Court	[0.7381321887351776]
2	Electronic City	Venue Category_Outlet Store	Venue Category_Furniture / Home Store	Venue Category_Bus Stop	[0.5423513638809381]

## DISCUSSION

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also during clustering, similar neighborhoods must be dumped into the right cluster. Another observation that we can make is that choosing number of clustering could produce very diverse results. Some may be over fitted or some may be under fitted. Hence analysis of number of clusters must be done. Ref elbow\_graph in the Methodology section.

## **CONCLUSION**

The recommender system is a system that considers factors such as population, income and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfil all their needs.

