

# **IBM Applied Data Science Capstone**

## **Opening new Indian restaurant in Atlanta: An exploration of neighborhoods in Atlanta, Georgia, USA**

By,

Kanchan Basnet

June 2019



## **Introduction**

The population of Indian American in Atlanta is going up. Many people who come from India to Atlanta work in IT sectors. The preferred location of their job in southeastern region of USA is Atlanta. As Atlanta is becoming a hub city for the IT and health sector companies, it is attracting many people of South Asian origin, most of whom work in the IT sector. In addition to those companies, Atlanta has some well-known big universities such as Georgia Institute of Technology, Georgia State University, Emory University. These universities attract a lot of students who like Indian foods. Because Indian cuisines are distinctive from other foods primarily because of their spicy taste, Indian foods are becoming popular for the people other than Indian origin also. Entrepreneur who is interested in opening a new Indian restaurant might want to do research on the distribution of Indian restaurant, quality and types of the cuisine they serve, price of the foods, neighborhoods where Indian people prefer to live, their outing and eating habit, companies where Indian people work, and number of Indian students who are attending in the universities, etc. The development of the business plan as well as the success of the business depends on these factors. In this project, the analysis of the distribution of the Indian restaurant in neighborhoods of Atlanta will be explored with the goal of providing a suggestion of a location for the new Indian restaurant.

## **Business problem**

Opening a new restaurant involves the research of a location in a city. In this project, we are trying to explore Atlanta neighborhoods to provide a recommendation of a suitable region for business success. Planning to open an Indian restaurant demands the data analysis of the distribution of Indian restaurants in the Atlanta area. In addition to

that, it requires an understanding of various aspects of targeted customers. This includes neighborhoods where most of the Indian communities live, the number of Indian American in a particular area, the number of Indian students studying in the nearby universities, etc. For an investor or an entrepreneur who is planning to open a new restaurant, these business challenges need to be tackled based on facts and figures before he or she starts making a business plan.

### **Target Audience**

The targeted audience of this project is business entrepreneurs or investors who are planning to invest in restaurant business. Data analysis from this project might be very useful for the Indian business investors who are investing in other sectors that target Indian populations since this project provides the indirect information about the neighborhoods where most of the Indian people live. Moreover, the information generated on Indian communities will be helpful for the Indian organizations who are working to build their communities better.

### **Data source**

The Wikipedia page [https://en.wikipedia.org/wiki/Category:Neighborhoods in Atlanta](https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Atlanta) was used to find and extract the name of the neighborhoods in Atlanta, Georgia. Once the neighborhoods' list was extracted by using web scraping technique, geographical coordinates for those neighborhoods are required in order to plot them in a folium map and also to get the name list of the venues in the particular region.

## **Data extraction methodology**

A list of 123 neighborhoods was extracted from Wikipedia page using web-scraping packages BeautifulSoup4 and python-requests. The list was then converted to pandas data frame and cleaned so that it was ready for the geographical coordinates to append. Python geopy.geocoder package was used to search the longitude and latitude of the neighborhoods in Atlanta. The obtained geographical data was then merged with the names of the neighborhoods in the pandas data frame.

Foursquare API was used to extract the venue data of a particular neighborhood in Atlanta. According to Foursquare, more than 150,000 partners and developers are using the webpage. Once the venues data was obtained, data cleaning and data wrangling was done, which makes the data ready for visualization and clustering. Folium package was used for the map visualization. Towards the end, machine learning technique k-means clustering was used to cluster the extracted list of Indian restaurants into four different clusters. Each cluster was then studied, and the decision for the appropriate spot where chances of business success would be maximized was recommended.

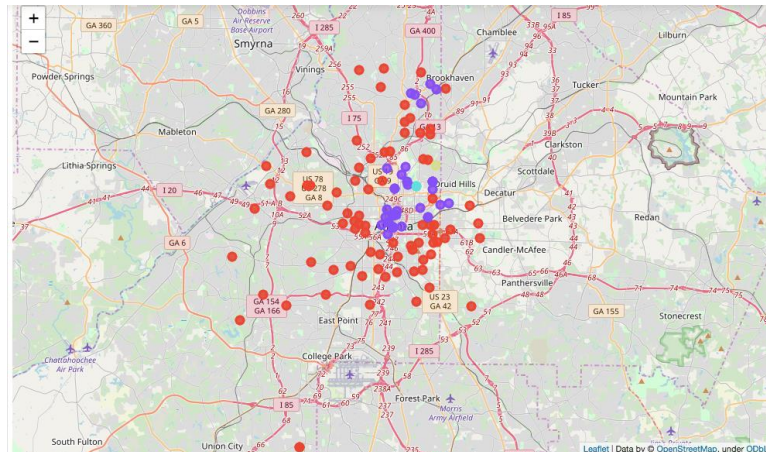
Foursquare API was used to extract the number of top 100 venues within 1500 meters radius of the city. This process required client\_id, client\_secret, and Foursquare API version. The obtained venues were then appended into the Neighborhoods data frame. Venues in each neighborhood were then grouped and counted using groupby and count function. Then, get-dummy function was used to convert venues categories into numerical values in order to use machine learning algorithm.

The obtained data frame that contains categorical value was then filtered based on vaenuecategory and grouped based on neighborhood. Once the data was preprocessed,

the number of Indian restaurants in each neighborhood was filtered out. Now, to explore the distribution of Indian restaurants around the city, the k-means clustering algorithm was used. This divides the Indian restaurants into 4 clusters. The cluster was visualized using folium map visualization technique. This gives a clear understanding of the distribution of restaurants in each cluster.

## Results

Out of 123 neighborhoods in Atlanta, only 41 neighborhoods have Indian restaurants. Most of the Indian restaurants are concentrated in cluster 0 and cluster 1. Those areas are located in downtown, east, and north side of the city center. Clusters 2 and 3 have only a few Indian restaurants. Based on the observations, it seems opening a new Indian restaurant in cluster 2 and 3 would face less competition as compared to cluster 0 and 1.



In the picture above, red color represents cluster 0, purple color belongs to cluster 1, light yellow cluster 2, and cyan color represents cluster 3.

## **Discussion**

Because clusters 0 and 1 have most of the Indian restaurants, one might think that these areas have high competition as compared to clusters 2 and 3. This simply might not be true. There are several factors that influence while deciding on opening an Indian restaurant in Atlanta.

The distribution of the population is one major factor. The majority of the Indian communities might be in cluster 0 and 1. If most of them are living in cluster 0 and 1, it makes sense to have the majority of restaurants in those areas. Opening of a restaurant in cluster 2 and 3 might be a blunder mistake if Indian populations are not living in those neighborhoods.

Another factor involves the location of companies where the majority of the Indians work. It is better to open a restaurant in nearby neighborhoods where Indian peoples work. If they work in cluster 0 and 1, but you opened a restaurant in clusters 2 or 3, then chances are high that your business will fail.

The crime rate in the neighborhoods also plays a vital role when you think of starting a new business. If you open a restaurant in cluster 2 or 3 simply because there is no competition, you might end up being a victim of burglary or other crimes. An investor does not want to put their money where there is a high crime rate. The reason that cluster 2 and 3 do not have that many Indian restaurants might be because of the higher crime rate in those neighborhoods.

Nearby universities also play an important role while determining a suitable spot for a successful restaurant business. If a restaurant is in a neighborhood where there is a large flow of Indian students, chances are high that the business will be successful.

In addition to these discussed aspects, other factors that affect the business also need to be considered before making any decision on opening a new restaurant.

### **Limitations and suggestions for future research**

This project does not explore the other factors that we discussed in previous section. Before making any decision, investor or entrepreneur needs to do further research on other factors that affect the chances of failure. K-means cluster analysis can also be done on Indian population distribution in the Atlanta area. Moreover, crime data can be searched and analyzed to divide the neighborhoods based on the number of crimes. This could give valuable information about the safest neighborhoods in Atlanta. In addition to these analyses, a simple search of the location of universities in Atlanta city and the tentative number of Indian students on those universities can be explored.

### **Conclusion**

In this project, we used a web-scraping algorithm BeautifulSoup4 to extract the list of the neighborhoods. Then python geocoder package was employed to extract the geographical coordinates. Our goal was to find a suitable cluster where the opening of a new Indian restaurant would be the right decision. Foursquare API was used to search the top 100 venues in the Atlanta neighborhood. Once the data was obtained, we cleaned and preprocessed the data to filter out the neighborhoods that have Indian restaurants. Out of 123 total neighborhoods, only 41 had Indian restaurants. We divided those neighborhoods into four clusters by using a machine learning algorithm K-means clustering. Out of those four clusters, cluster 0 and 1 have the highest number of neighborhoods that contains Indian restaurants. Clusters 2 and 3 have only a few Indian restaurants. If the Investor only makes a decision based on competition faced by the

restaurant, then probably opening a new Indian restaurant in cluster 2 or 3 is a good idea. But there are other factors that play a vital role. Population distribution, the crime rate in neighborhoods, number of students in universities, location of the universities, etc. also play a major role in the determination of suitable spot in Atlanta, neighborhoods. Although this project provides valuable information about the clustering of neighborhoods based on the distribution of Indian restaurant, it would be better for an investor to look into other factors and research other aspects also before making any decision about their future investment.

## References

1. List of Atlanta neighborhoods  
[https://en.wikipedia.org/wiki/Category:Neighborhoods\\_in\\_Atlanta](https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Atlanta).
2. Foursquare developer credentials  
<https://foursquare.com/developers/apps>
3. Atlanta city Wikipedia  
<https://en.wikipedia.org/wiki/Atlanta>
4. List of Indian dishes  
[https://en.wikipedia.org/wiki/List\\_of\\_Indian\\_dishes](https://en.wikipedia.org/wiki/List_of_Indian_dishes)
5. Indian-Asian population in Atlanta, Georgia, USA  
[http://documents.atlantaregional.com/gawsnapshots/ASIAN\\_INDIAN.pdf](http://documents.atlantaregional.com/gawsnapshots/ASIAN_INDIAN.pdf)