

# **The Battle of Neighborhoods- Coursera Capstone Project: Finding a place to open a profitable Afghani restaurant**

## **Table of Content**

1. Introduction: Business Problem
2. Data
3. Methodology
4. Analysis
5. Discussion
6. Results / Conclusion

## **1. Introduction: Business Problem**

The final project for the Coursera course: Applied Data Science Capstone of IBM Data Science Professional Certificate is to apply the learning from all the courses of that certificate. In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. In Module 4, we have to come up with an idea of real data science problem and find its solution.

### **Background information:**

New York is a major metropolitan area with more than 8.4 million (Quick Facts, 2018) people living within city limits. New York City is the largest city in the United States with a long history of international immigration. People came from many parts of the world. According to the 2007 American Community Survey estimates, New York City is home to approximately 50,000 people from Afghanistan. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, French, etc. I know that New Yorkers are open to new ideas and food and they would try and taste food from other countries.

### **Problem statement:**

Long before starting university, I always had an idea in my mind to open an authentic Afghani restaurant, especially the afghani bread and kababs are best. So in this final project, I will work on that idea and try to find the best place in New York City to open an authentic Afghani restaurant.

### **Target audience for this report**

- Business personnel who wants to invest or open a restaurant.
- Finding the best location for opening a profitable restaurant.
- Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it.

## **2. Data Section**

For this project we need the following data:

1. New York City data that contains Borough, Neighborhoods along with their latitudes and longitudes
- Data Source: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

- Description: This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.
- 1. Afghani restaurants in New York City
- Data Source: Foursquare API
- Description: By using this API we will get all the venues in the New York City. We can filter these venues to get only Afghani restaurants.

### 3. Methodology

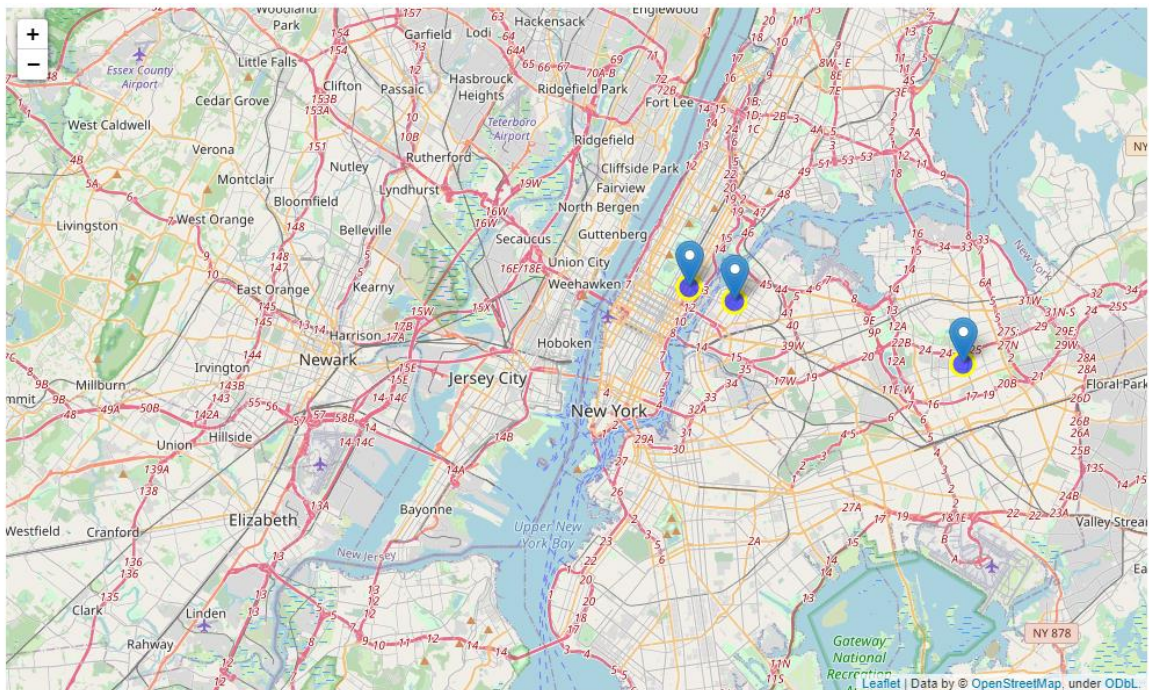
- Collect the new york city data from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset).
- Using Foursquare API we will get all venues for each neighborhood.
- Filter out all venues which are Afghani Restaurants.
- Data Visualization and some statistical analysis.
- Analyzing using Clustering (Specially K-Means):
  1. Find the best value of K
  2. Visualize the neighborhood with a number of Afghani Restaurants.
- Compare the Neighborhoods to Find the Best Place for Starting up a Restaurant.
- Inference From these Results and related Conclusions

## 4. Analysis

We got 3 Afghan Restaurants across the New York City.

|   | Borough   | Neighborhood | Latitude  | Longitude  | Average Rating |
|---|-----------|--------------|-----------|------------|----------------|
| 0 | Manhattan | Lenox Hill   | 40.768113 | -73.958860 | 7.8            |
| 1 | Queens    | Ravenswood   | 40.761705 | -73.931575 | 8.1            |
| 2 | Queens    | Utopia       | 40.733500 | -73.796717 | 6.8            |

Out[35]:



As we continue our analysis, we see below that Manhattan has one restaurant and Queens has two restaurants. From the data of the Foursquare API, we can see that there are only three afghani restaurants in New York City, which is very unusual. I think the data is not that much updated or the afghani restaurants are not group accordingly. The afghani restaurants might be mistaken with Irani/Persian restaurant as these two cousins are much similar. These could be another project to include the Afghani and Irani restaurant together and see their ratings and likes. For this project, we will just consider the afghani restaurant for further data analysis and result/conclusion.

Out[30]:

|   | Neighborhood | Average Rating |
|---|--------------|----------------|
| 1 | Ravenswood   | 8.1            |
| 0 | Lenox Hill   | 7.8            |
| 2 | Utopia       | 6.8            |

## 5.Results / Conclusion

From the figure above, we can see that the Manhattan has one afghani restaurant and Queens have two afghani restaurants. So there is very good opportunity to open an afghani restaurant in neighborhood other than these two. From these information, we cannot conclude much where to open a new afghani restaurants which will give us the most profits. Based on the rating, likes and tips, I would say that the afghani restaurant in Queens neighborhood and in Utopia borough is the best one, that I would visit. I would not open a new restaurant near that one. For further evaluation of opening a afghani restaurant one should take consideration of the population data of the afghani origin people living in New York City. If we can combine that data with the data above then we can also make a very good decision to open a afghani restaurant where the population of the Afghani people are more and there might not be a afghani restaurant near to them or may have not good rating and likes. As a final note, all of the above analyses is depended on the adequacy and accuracy of Four Square data. A more comprehensive analysis and future work would need to incorporate data from other external databases.

Finally, to conclude this project, we have got a small glimpse of how a real-life Data science project looks like. I have used some frequently used python libraries to handle JSON file, plotting graphs, and other exploratory data analysis. Use Foursquare API to major boroughs of New York City and their neighborhoods. The potential for this kind of analysis in a real-life business problem is discussed in great detail. I would like to thank you Coursera and IBM for a such a great course and specialization certificate.