

# THE BATTLE OF “NEIGHBORHOOD”

---

Thinam Tamang

March 31, 2020

## INTRODUCTION & BUSINESS PROBLEMS

The city of **Toronto** is the most populated city in **Canada**. It is the capital of the province of **Ontario** and home for more than **2.7 million people** which makes it the fourth most populated city in **North America**. In the **20th century** the city experienced significant industrial development and subsequently became the **English Speaking Banking, Financial and Commercial heart of Canada**. It was named as the most important financial centres in the world and all the credit goes to the **Medical Research, Film Production, Technology and Computer Science, The Arts, Education** and many more.

With its diverse cultural institutions, including **Museums and Art galleries, Festivals and Public events** and many more **Toronto becomes America's most visited city**.

These tourists promote the **Tourism Industries** which generate a very competitive market. As it is a very developed city, it basically deals with business cases. That's why any new investment or company that wants to move to **Toronto** needs to use market based insights that will help them to understand the **Business Environment** and allowing them for a strategy to reduce the risk. And increase the profit return on investment.

## DATA DESCRIPTIONS

A **Restaurant** is a business establishment that serves prepared meals and beverages on the premises in exchange of the payment. The food is prepared by the Chef. The term covers multiplicity of the places and a great diversity of the types of cuisine, both local and foreign. The restaurants are sometimes the device reserved for serving the meals within the large entity. When we speak about collective restaurants as opposed to the kitchen site, they can also be associated with a catering or grocery business. The restaurants offers more or less

comfort conditions and the restaurants is also called Fast because when the customer place the order they can get and eat in minutes or tens of minutes possibly standing.

While searching on the Internet, I found the Toronto has several Restaurants. Specific theme restaurants such as: **Greek Restaurants, Hasal Restaurants, Jewish Restaurants** and many more. There are 2 Restaurants in big Hotel like **Drake Hotel**. There are pizzaries like **Pizza Nova or Pizza Yolo**. 4 Bistros like **Bistro 990** and 5 Global Restaurants like **Big Smoke Burger or Fan's Restaurants**. 6 Restoring bars such as **Rivoli**.

All this leads to a competitive market that requires distinguishing such as Which criteria to choose to optimize the income? Presence of the supplier nearby? The demography of Toronto, Number of Competitors Nearby. And the demography of the Neighborhood. The contribution of the people of Toronto. The list can continue.... This analysis will allow an investor to be able to choose the optimal location for business and can use this analysis to move or open an extension for the business.

In this analysis, I will work out with the csv file of **Toronto Neighborhood**. I also used the **Four Square API** to know details about the **Neighborhood of Toronto** and it deals with 1.6 km around the city.

## PROBLEMS & TARGET

My analysis basically deals with finding the best Neighborhood in Toronto to establish a Restaurant and helping the investors for making profitable investment to earn the profit.

## DATA CLEANING & FEATURE SELECTION

At first I defined the libraries required for the analysis such as pandas, numpy, matplotlib, seaborn, KMeans and more. Some of the basic libraries are listed above. At part from this I used the FourSquare API to find the locations of different Neighborhoods which will be discussed more later. I downloaded the csv file of Toronto Neighborhood Profiles from Internet and saved it in my device. Then I read the Data using pandas. At initial phase the downloaded Data was not systematic to proceed my Analysis so I have to undergo the different cleaning technique such as removing the unnecessary rows and columns and I even undergo removing unnecessary columns.

I selected the simple features to carry my analysis such as Demographic Characteristics and Neighborhoods Index which was a bit challenging because a lot of Data were duplicated and some even don't have values.

The cleaned and ready to use Data Frame looks as shown in the figure below:

	Neighbourhood Number Neighbourhood Information	TSNS2020 Designation Neighbourhood Information	Population, 2016 Population and dwellings	Population, 2011 Population and dwellings	Population Change 2011-2016 Population and dwellings	Total private dwellings Population and dwellings	Private dwellings occupied by usual residents Population and dwellings	Population density per square kilometre Population and dwellings	Land area in square kilometres Population and dwellings	Children (0-14 years) Age characteristics	Youth (15-24 years) Age characteristics	
	Agincourt North	129.0	No Designation	29113.0	30279.0	-3.9	9371.0	9120.0	3929.0	7.41	3840.0	3705.0
	Agincourt South- Malvern West	128.0	No Designation	23757.0	21988.0	8.0	8535.0	8136.0	3034.0	7.83	3075.0	3360.0
	Alderwood	20.0	No Designation	12054.0	11904.0	1.3	4732.0	4616.0	2435.0	4.95	1760.0	1235.0
	Annex	95.0	No Designation	30526.0	29177.0	4.6	18109.0	15934.0	10863.0	2.81	2360.0	3750.0
	Banbury- Don Mills	42.0	No Designation	27695.0	26918.0	2.9	12473.0	12124.0	2775.0	9.98	3605.0	2730.0

## EXPLORATORY DATA ANALYSIS

Here at first I used the FOUR SQUARE API to get the coordinates of each Neighborhood. But unfortunately I was not able to get the coordinates for each Neighborhood so I had to assign coordinates for that Neighborhood myself by referring the Wikipedia and searching through the Internet. Then By the help of FOURSQUARE API, I was able to look at the sites that are about 1610 m which means (1 mile) around the Toronto. Then I listed the collection of Neighborhood obtained from FOURSQUARE API to the New Data Frame.

Examples Of Connecting To The FOURSQUARE API is represented below:

```

CLIENT_ID = 'W4RF2PNNJMJBMCW4D5MD2SODDFBKSCRQ51PPAFV5XKKEQF'
CLIENT_SECRET = 'S1LH2T5CYL1MU3LMXF0I1DXTSMLZI4B3CQIQ4JHGETVIP3NR'
VERSION = '20200330'
LIMIT = 100

i = 0
search_query = ''
radius = 1610

for index, row in df_final.iterrows():

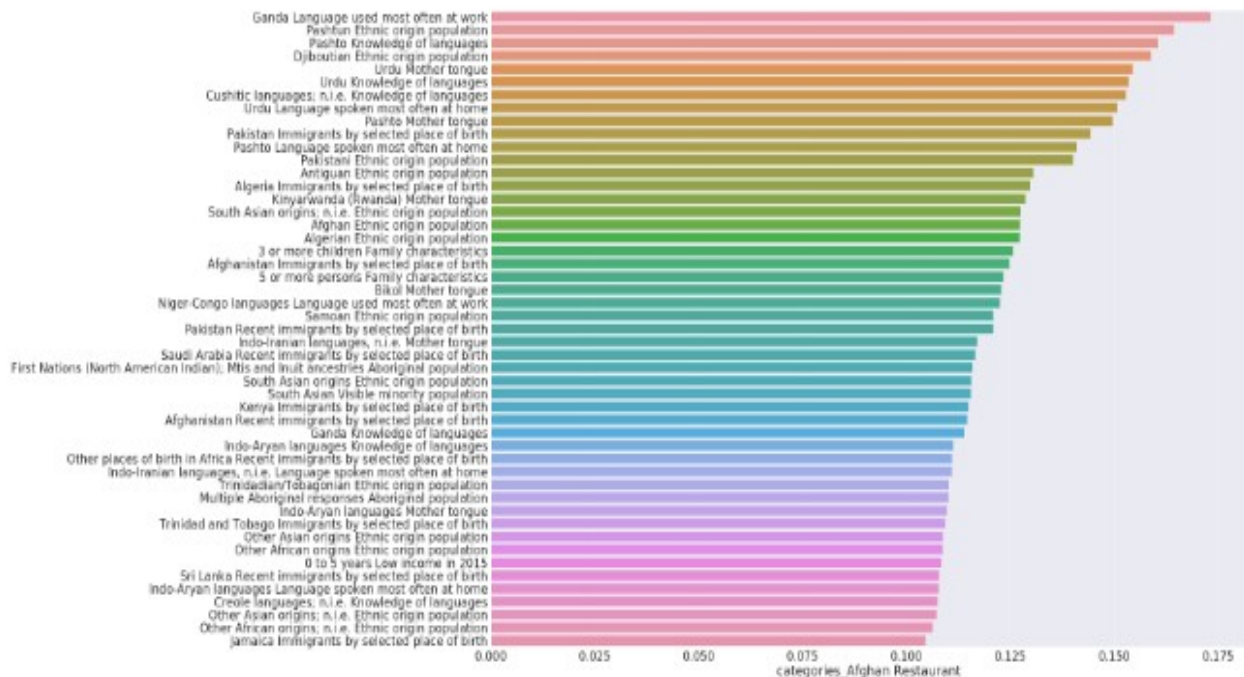
    url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{ }&v={}&query={}&radius={}&limit={}'.format(
        CLIENT_ID, CLIENT_SECRET, row['lat'], row['lon'], VERSION, search_query, radius, LIMIT)
    results = requests.get(url).json()
    results = json_normalize(results['response']['venues'])

```

## DATA VISUALISATION

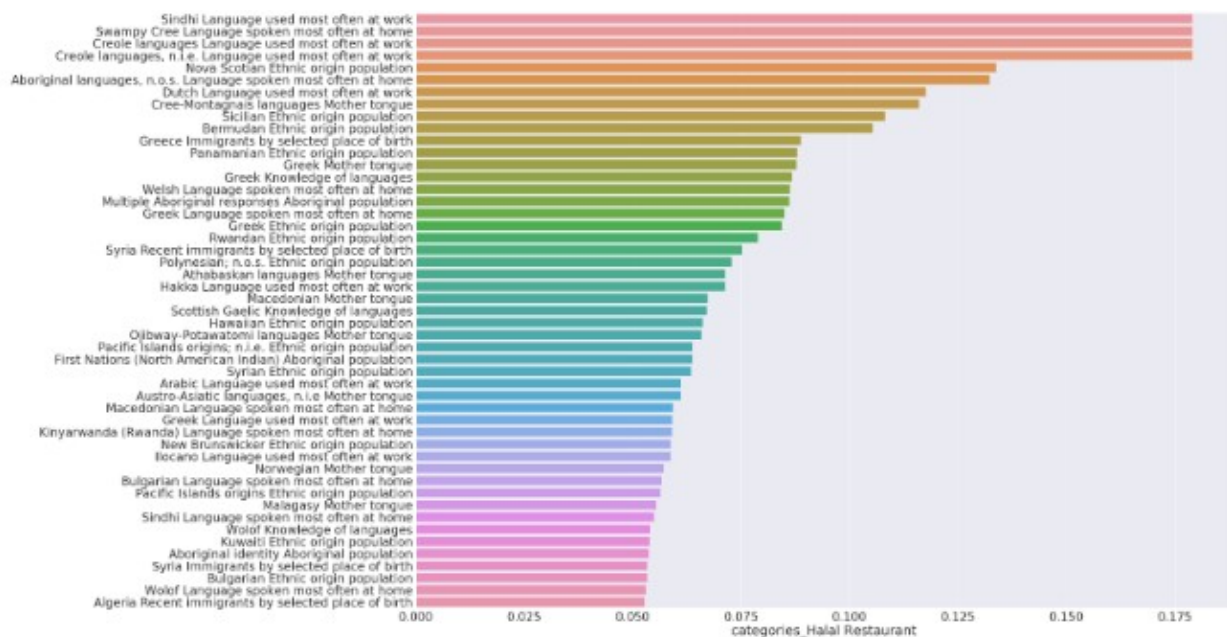
In this Analysis Process I applied Informative Visualization Techniques such as Folium, Bar Plot and Horizontal Plot. I applied Circle Marker to make the Visualization more Informative. I also used One Hot Encoding Technique. I tried to find out the correlation between different features such as correlation between Restaurants Popularity among different Neighborhoods.

### Correlation Of Afghan Restaurants



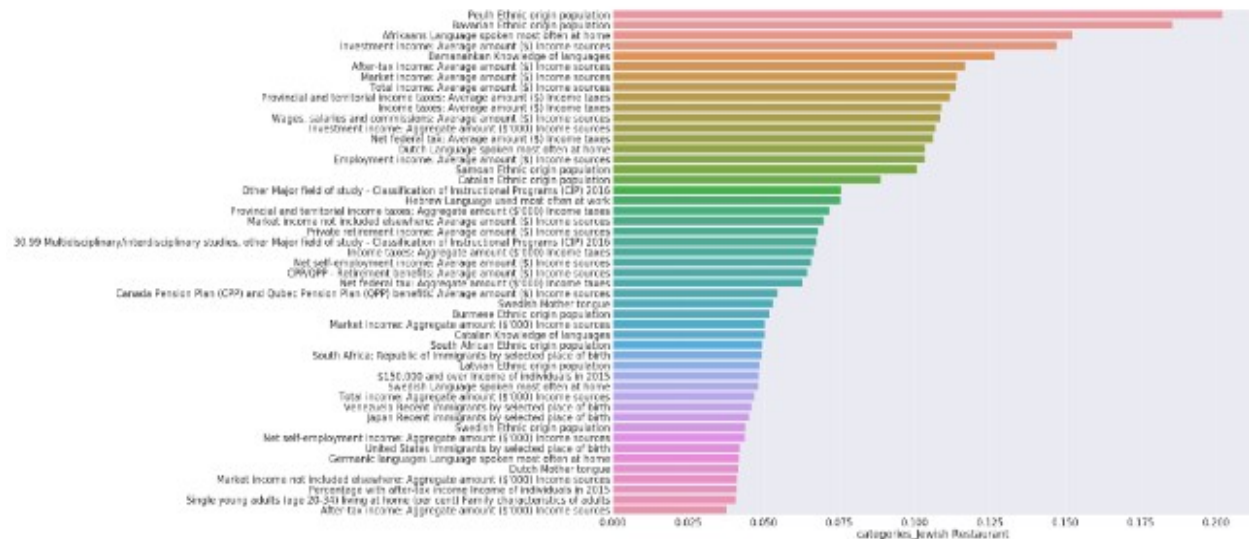
In the above plot we can see Ganda, Ush Tic, & Urdu languages are the most spoken languages in Afghan Restaurants. The above Bar Plot clearly shows that people who speak Ganda, Ush Tic, and Urdu Languages visit Afghan Restaurants because these languages are the most spoken languages in Africa, more in South East Africa & East Africa. This Bar Plot presents the Correlation between Afghan Restaurants and People of specific characteristics. We can also see that Afghan Restaurants are famous among the Muslim Community more.

## Correlation Of Halal Restaurants



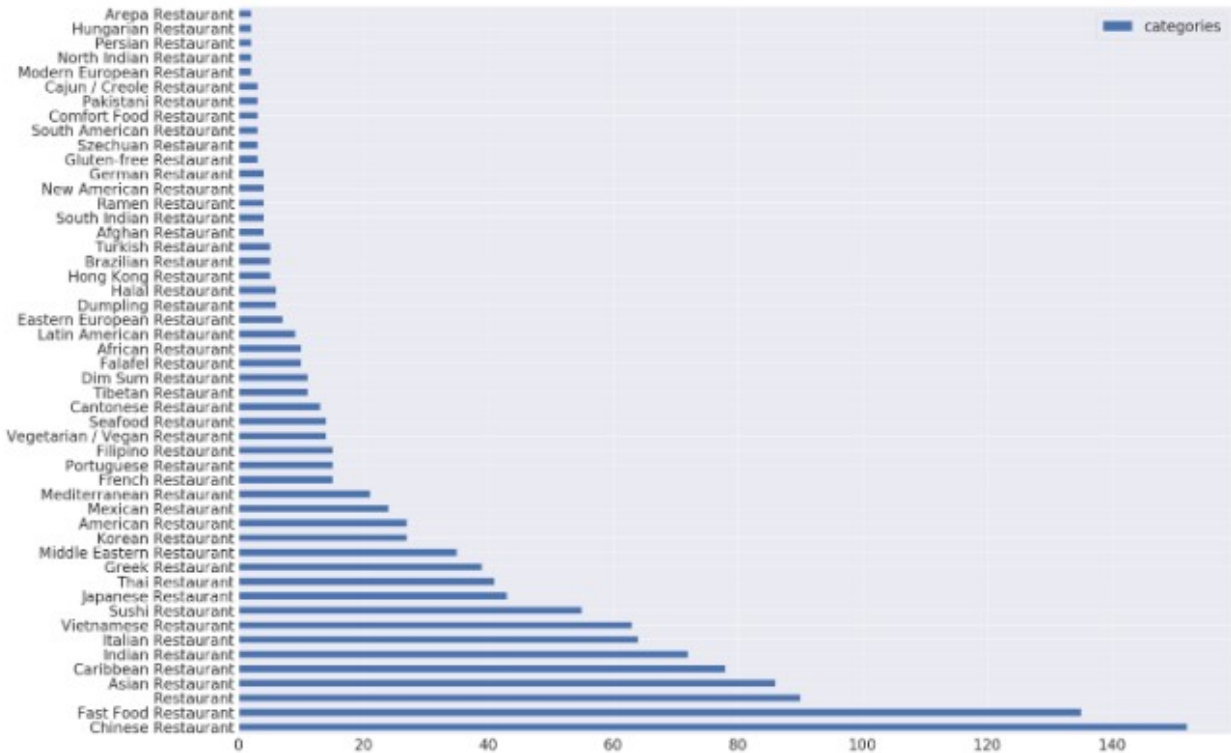
In the above Bar Plot we can see that Sindhi, Creol, & Swampy Creed are the most languages that are spoken in Halal Restaurants. The above mentioned Bar Plot makes crystal clear that people who speak Sindhi, Creol & Swampy Creed visit the most to the Halal Restaurants. The second is the street language which is little distorted and trixieme is a language which is spoken most in Northern Canada.

## Correlation of Jewish Restaurants



In the above mentioned Bar Plot is the Correlation of Jewish Restaurants among different Neighborhoods in Toronto. The Bar Plot presents that Peul and Bavarian ethnicities are the most common ethnic groups who visits the Jewish Restaurants the most. And In the second line of the Bar Plot represents the average income of People and it also indicates that the People visiting the Jewish Restaurants tends to have high average income.

## Restaurants In Canada



The above mentioned Horizontal Bar Plot represents the Number of Restaurants spread over the great Canada. As shown in the Horizontal Bar Plot it clarifies that the Chinese Restaurants and Fast Food Restaurants are the Restaurants which are most spread in Canada and obviously in the Toronto.

## CLUSTERING

My target in this Analysis is to find out the Neighborhoods which the most favorable for establishing or starting the Restaurants. In order to perform Clustering and to find out which Neighborhoods are the most favorable to open the Restaurants I had to pick out some features that can foster the Income in that area.

The features are listed as follows:

- Population
- Population Density
- Percentages Of Person Living Alone
- Total Income: Average
- Non Permanent Residents Immigrant

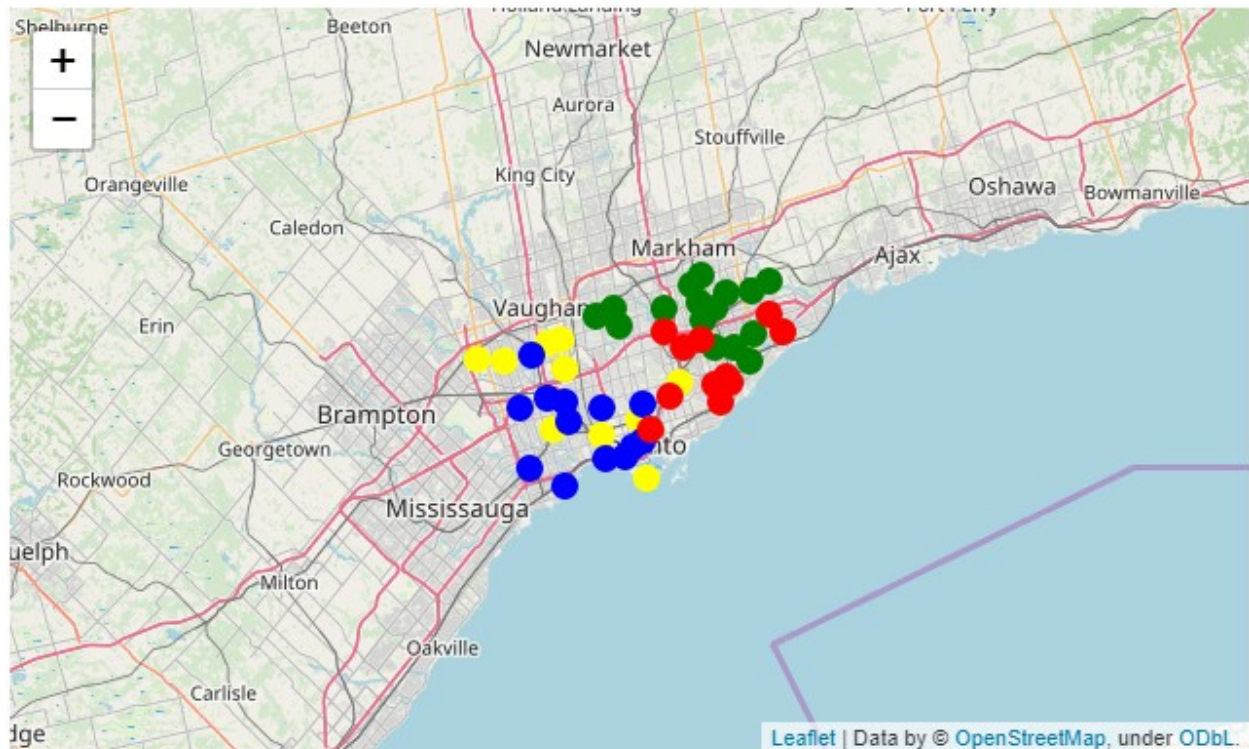
- Youth(15-25 years)
- Working Age(25-54 years)
- Females
- Males
- After Tax Income
- Latitude(To save the location)
- Longitude(To save the location)

Then I created the New Data Frame containing the above mentioned features inorder to carry out the Analysis.Then I normalized the newly created Data Frame using Pandas.And I calculated the average score of each Neighborhoods.

Then I arranged the Neighborhoods in the descending order and choosed the 50 best Neighborhoods according to the score.And join the above mentioned features with the Restaurants to carry out the Analysis process.Then I counted the Neighborhoods and added the new features to make the Data set ready for Training.

Then I fitted the KMeans on learning Data and assign the assign the label to each Neighborhood.And I created the Folium Map based on the Data Frame created after undergoing the above mentioned processs.The Folium Map looks like as follows:





Then I calculated the Clustering in each Neighborhoods. The score of each Clustering is mentioned as follows:

```
print(df_Pop[df_Pop.labels == 0 ].shape)
print('average Number_of_Restaurant cluster 1 : ' + str(df_Pop[df_Pop.labels
== 0 ][ 'Number_of_Restaurant' ].mean()))
print('average score cluster 1 : ' + str(df_Pop[df_Pop.labels == 0 ][ 'score'
].mean()))
df_Pop[df_Pop.labels == 0 ].head()
```

```
(11, 15)
average Number_of_Restaurant cluster 1 : 13.909090909090908
average score cluster 1 : 2.2980868355062793
```

I calculated the Clustering process for 4 times by undergoing the same process as mentioned as above.

## DISCUSSION OF THE RESULTS

Kmeans returned 4 clusters and each cluster has its own characteristics: • Cluster 1 in red: groups together neighborhoods with a relatively low score and with a relatively low restaurant count in its perimeter average Number Of Restaurant cluster 1 : 12.0 average score cluster 1 : 2.27 • Cluster 2 in blue: groups neighborhoods with a relatively high score and a relatively high number of restaurants also in its perimeter average Number Of Restaurant cluster 2 : 18.8 average score cluster 2 : 2.39

• Cluster 3 in green: groups neighborhoods with a relatively low score and a relatively high number of restaurants also in its perimeter average Number Of Restaurant cluster 3: 19.64 average score cluster 3: 2.30 • Cluster 4 in yellow: groups neighborhoods with a relatively high score and a relatively low number of restaurants also in its perimeter average Number Of Restaurant cluster 4: 13.692307692307692 average score cluster 4: 2.4528401540266525

## RECOMMENDATIONS OF THE RESULTS

The recommendations I bring for a restaurant opening after all this project:

- The criteria for opening a restaurant depend partly on the category of restaurant, if you want to open a Chinese restaurant, you target neighborhoods with a Chinese population quite present, as well as for an Italian restaurant or other.
- The Chinese restaurants and the fast foods are the most present by conceiving the competition is rough so it is preferable to choose another category.
- I recommend the East Willowdale, Mount Olive-Silverstone-Jamestown, Waterfront Communities-The Island and Dovercourt-Wallace Emerson-Junction neighborhoods to open a restaurant because they are part of a cluster that has the fewest restaurants so less competition and they have a high demographic score so it is favorable to a high turnover.

## CONCLUSIONS

We can rely on the results quoted before even if it remains imprecise and that because of the lack of data provided by the foursquare API, a premium account will give us the possibility of seeing the note of the people and it will facilitate us the work better, by using a far-off regression we will get the right selection of a neighborhood and the result will be more accurate. Also later we will consider more data to reinforce our choice like, the trade surrounding the restaurants, the crime scene of neighborhoods and others.