IBM Data Science Capstone Project- The Battle of the Neighborhoods

Analysing Chicago Neighborhoods

By: Neelabh Sen

1.0 Introduction

Chicago, the largest city on the banks of Lake Michigan in the state of Illinois is the third most populous city in the United States and one of the fastest growing multicultural metropolitan centers in the world, with rapidly developing infrastructure, a beautiful waterfront, lively nightlife and liberal expression of diverse cultural identities and ideas attracting a large number of immigrants domestically and from around the world each year. The city is in fact, one of the most racially diverse metropolitan areas in the US, with a large chunk of the population being formed by African American and Asian minorities. This also promotes business and restaurants offering a large variety of cuisines ranging from Indian and Middle-Eastern dishes to Vietnamese and Japanese delicacies have flourished in the city which has become a *melting pot* of cultures from around the world. One of the major civic problems that the city faces however, is that the crime rate remains above the national average despite numerous measures that have been taken by state and federal law enforcement agencies over the last few years. In spite of these problems, Chicago remains one of the hottest destinations for setting up new business, as is evidenced by the fact that it was named among the top global cities in the US on the '*Ease of Doing Business Index*'. Furthermore, *KPMG* ranks Chicago as one of the most cost-effective cities in the world for doing business, ahead of New York and Los Angeles.

Problem:

The background of the business problem that we will be addressing in this project is as follows- a Japanese immigrant family is moving to Chicago and setting up a business which imports authentic ingredients (such as *Nori (Seaweed), Wasabi, Miso, Sushi Rice, Rice Vinegar* etc.) and other food stuffs from Japan and sells them to local businesses such as eateries and restaurants. Thus, the problem is two-fold:

- The business should be set-up in an area which is in close proximity to well-established Japanese restaurants.
- The neighborhood should be well-suited to a young immigrant family, i.e., it should have all the necessary amenities in the vicinity.
- Furthermore, the neighborhood should also be as crime-free as possible, i.e., with a low crime rate.

This business problem (although very specific for the purpose of this project) can be of broad interest to a 'large **target audience'** anyone setting up a business in the city of Chicago, or planning to move to the city for work and settling here with family. In the following sections, I will detail the data that will be used for this analysis, the methodology as well as a brief discussion of the results.

2.0 Data Required

In order to address our previously stated business problem, we would require various kinds of data from numerous sources which have been listed and briefly described below:

- Crime Data from the Chicago City Data Portal which maintains an extensive record of all crimes that have been reported in Chicago over the last two decades including- date, type of crime, exact location, district, police ward etc.
    - Data Source: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2
    - Description: As the original dataset is extremely large and cannot be handled properly for analysis, only the data from the year 2016 has been used as representative. This data has been used to build a Choropleth map of the city of Chicago showing the intensity of crime in various neighborhoods.
    - Example of the Data:

[2]:

	Case Number	Date	Primary Type	District	Ward_Number	Community Area	Year	Latitude	Longitude
0	HZ250496	05/03/16 23:40	BATTERY	10.0	24.0	29.0	2016	41.864073	-87.706819
1	HZ250409	05/03/16 21:40	BATTERY	3.0	20.0	42.0	2016	41.782922	-87.604363
2	HZ250503	05/03/16 23:31	PUBLIC PEACE VIOLATION	15.0	37.0	25.0	2016	41.894908	-87.758372
3	HZ250424	05/03/16 22:10	BATTERY	15.0	28.0	25.0	2016	41.885687	-87.749516
4	HZ250455	05/03/16 22:00	THEFT	15.0	28.0	25.0	2016	41.886297	-87.761751

- Chicago city data which includes the names and boundaries of all neighborhoods was obtained from the Chicago Data Portal in ‘csv’ format as well as Geospatial data was obtained in ‘json’ format.
    - Data Source: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9
    - Description: Using the above data in conjunction with the *GeoPy* geocoding libraries for Python, latitudes and longitudes for each neighborhood was obtained and organized into a *Pandas* dataframe.
    - Example of the Data:
      Dataframe from the imported CSV File:

[11]:
```
# Reading the data into a Pandas Dataframe
df = pd.read_csv("chicago_neighborhoods.csv")
df.head()
```

[11]:

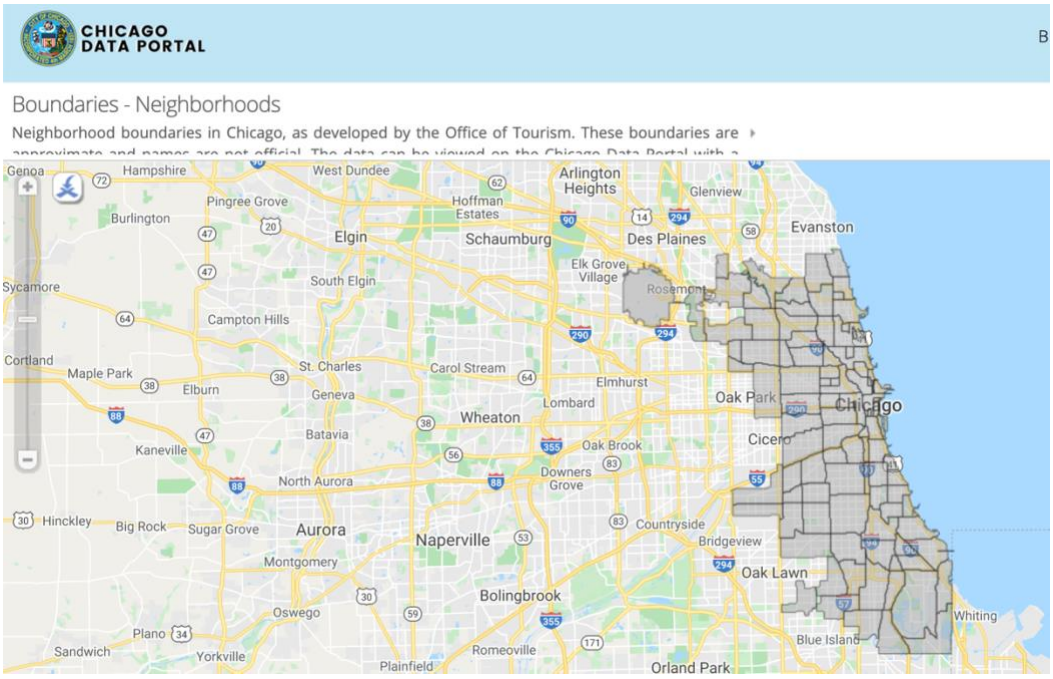
	PRI_NEIGH	the_geom	SEC_NEIGH	SHAPE_AREA	SHAPE_LEN
0	Grand Boulevard	MULTIPOLYGON (((-87.60670812560372 41.81681377...	BRONZEVILLE	4.849250e+07	28196.837157
1	Printers Row	MULTIPOLYGON (((-87.62760697485348 41.87437097...	PRINTERS ROW	2.162138e+06	6864.247156
2	United Center	MULTIPOLYGON (((-87.66706868914602 41.88885187...	UNITED CENTER	3.252051e+07	23101.363745
3	Sheffield & DePaul	MULTIPOLYGON (((-87.65833494805533 41.92166144...	SHEFFIELD & DEPAUL	1.048259e+07	13227.049745
4	Humboldt Park	MULTIPOLYGON (((-87.74059567509266 41.88782316...	HUMBOLDT PARK	1.250104e+08	46126.751351

After Cleaning the Data and Adding Latitude and Longitude:

[12]:

	Borough	Neighborhood	Latitude	Longitude
0	BRONZEVILLE	Grand Boulevard	41.813923	-87.617272
1	PRINTERS ROW	Printers Row	41.873787	-87.628900
2	UNITED CENTER	United Center	41.880683	-87.674185
3	HUMBOLDT PARK	Humboldt Park	41.905767	-87.704174
4	GARFIELD PARK	Garfield Park	41.882088	-87.715917

Geospatial Data:



- Data about the venues and amenities within a 1km radius of each neighborhood and the locations of established Japanese restaurants in Chicago was obtained using the *Foursquare API*.
    - Data Source: Foursquare credentials can be obtained by creating a developer account on https://developer.foursquare.com/

Defining Foursquare API Credentials

```
[22]: CLIENT_ID = '4W0W0WHMGIC31GMXQZWOWVXQQGTP5KCLWDBC33U5K0SDYKRY'
CLIENT_SECRET = 'XXHAOM32AVNZDYWS1EHRWIGLOKVCCOI3XZZSLL1KBASRDIPX'
VERSION = '20180605' #Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: 4W0W0WHMGIC31GMXQZWOWVXQQGTP5KCLWDBC33U5K0SDYKRY
CLIENT_SECRET:XXHAOM32AVNZDYWS1EHRWIGLOKVCCOI3XZZSLL1KBASRDIPX
```

- Description: Foursquare's API was used to obtain the names of venues as well as related information such as location, customer ratings, tips etc.
- Example:
  Defining Foursquare credentials:

Example of Venue Data obtained through the Foursquare API:

[59]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BRONZEVILLE	41.813923	-87.617272	Ain't She Sweet Cafe	41.816817	-87.613004	Coffee Shop
1	BRONZEVILLE	41.813923	-87.617272	Chicago Blues District	41.810071	-87.614105	Jazz Club
2	BRONZEVILLE	41.813923	-87.617272	Sip & Savor	41.816817	-87.612876	Coffee Shop
3	BRONZEVILLE	41.813923	-87.617272	Parkway Ballroom	41.813142	-87.616064	Food
4	BRONZEVILLE	41.813923	-87.617272	Blues Brothers Mural / Shelly's Loan & Jewelry...	41.809391	-87.619517	Plaza

3.0 Methodology

- After importing all the necessary python libraries to our Jupyter Lab environment, we initiate the process of importing and cleaning the data.
- First, we import crime data from the Chicago city data portal into a *Pandas* dataframe and display the data as follows:

	Case Number	Date	Primary Type	District	Ward_Number	Community Area	Year	Latitude	Longitude
0	HZ250496	05/03/16 23:40	BATTERY	10.0	24.0	29.0	2016	41.864073	-87.706819
1	HZ250409	05/03/16 21:40	BATTERY	3.0	20.0	42.0	2016	41.782922	-87.604363
2	HZ250503	05/03/16 23:31	PUBLIC PEACE VIOLATION	15.0	37.0	25.0	2016	41.894908	-87.758372
3	HZ250424	05/03/16 22:10	BATTERY	15.0	28.0	25.0	2016	41.885687	-87.749516
4	HZ250455	05/03/16 22:00	THEFT	15.0	28.0	25.0	2016	41.886297	-87.761751

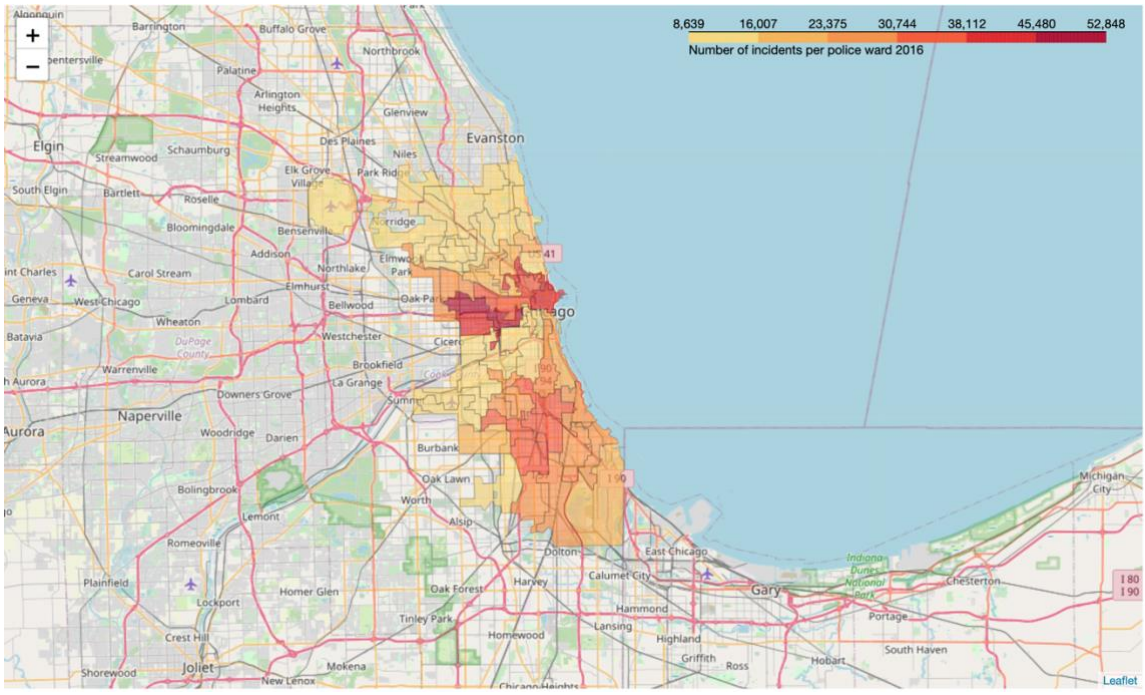
- Our purpose is to create a choropleth map showing the intensity of crime in different areas within the city. For this, we clean our dataframe and display only the two columns required- Police ward number and Number of crimes:

[5]:
```
# Converting the ward numbers into type string
df_crime_count['Ward'] = df_crime_count['Ward'].astype('int')
df_crime_count['Ward'] = df_crime_count['Ward'].astype('str')
df_crime_count.head()
```

[5]:

	Ward	Count
0	28	52415
1	42	45240
2	24	45172
3	2	39862
4	27	36655

- Following this, we use the ward boundary data from the Chicago Data Portal and create a Choropleth map using the library *Folium* as follows:



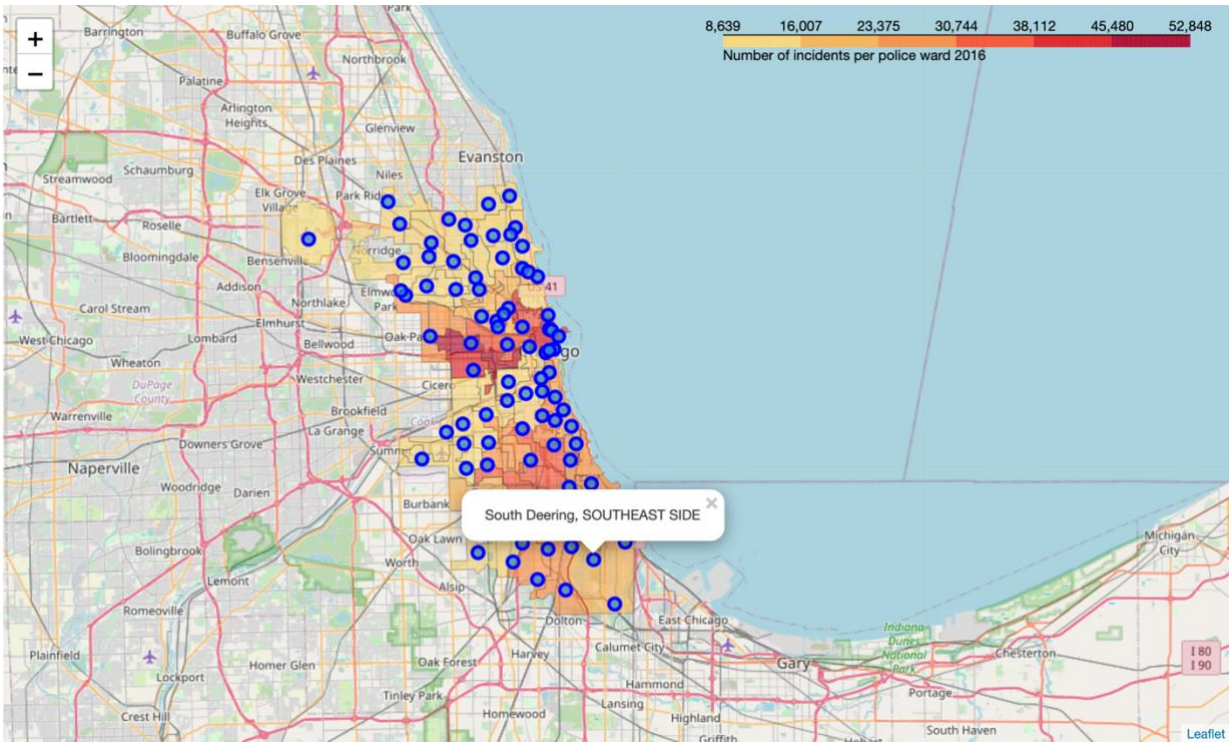
As can be seen in the choropleth map above, the intensity of crime is especially high in the inner cities and near the south central regions of the city, which makes these areas unsuitable for setting up a business or settling down.

- Following this, we Chicago neighborhood data and make a new dataframe which contains the name of each borough, neighborhood and its latitude and longitude using geopy's geocoder. We show the locations of these neighborhoods on our previously made map along with markers:

```
[12]: # Checking the resulting Dataframe
neighborhoods.head()
```

[12]:

	Borough	Neighborhood	Latitude	Longitude
0	BRONZEVILLE	Grand Boulevard	41.813923	-87.617272
1	PRINTERS ROW	Printers Row	41.873787	-87.628900
2	UNITED CENTER	United Center	41.880683	-87.674185
3	HUMBOLDT PARK	Humboldt Park	41.905767	-87.704174
4	GARFIELD PARK	Garfield Park	41.882088	-87.715917



- We then use the *Foursquare API* to find out the venues within a 1km radius of each neighborhood and create a new dataframe showing the most common venues for each neighborhood as follows:

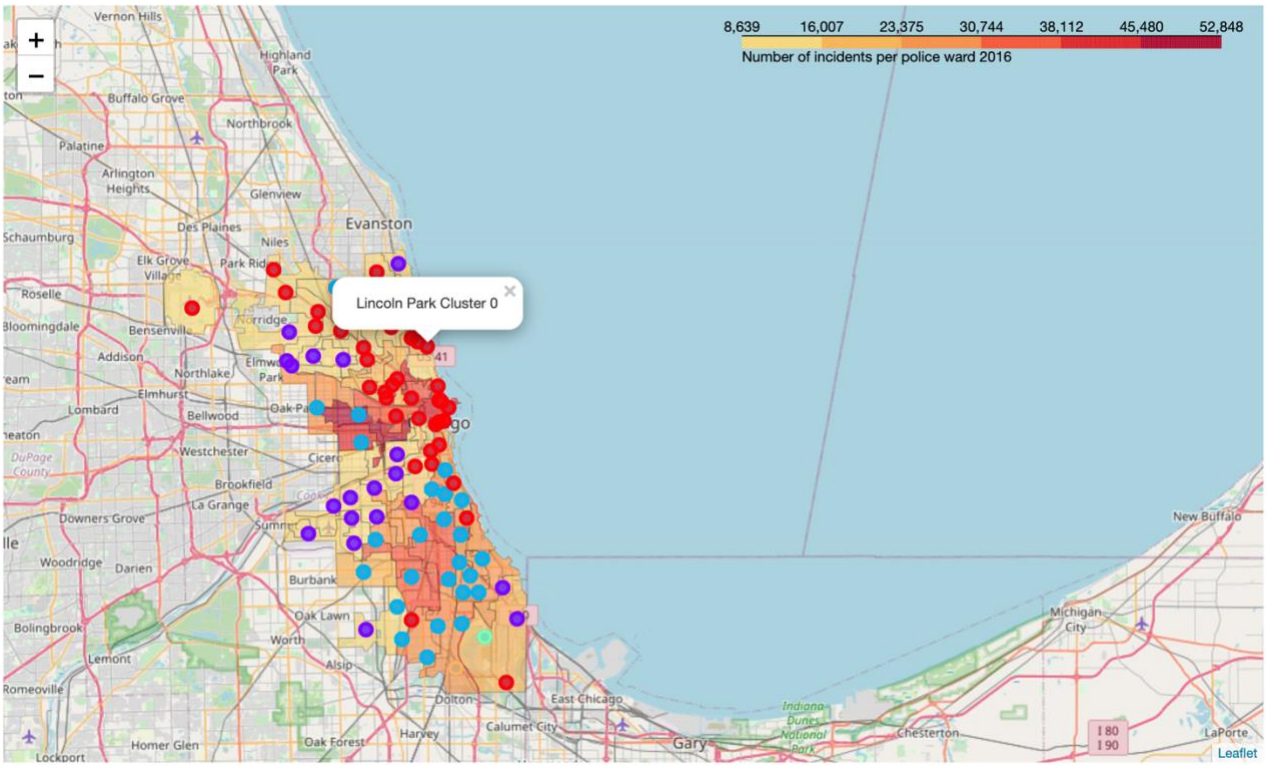
[25]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albany Park	Middle Eastern Restaurant	Ice Cream Shop	Bakery	Park	Mexican Restaurant	Sandwich Place	Asian Restaurant	Pizza Place	Grocery Store	Burger Joint
1	Andersonville	Grocery Store	Coffee Shop	Breakfast Spot	Pizza Place	Vietnamese Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Bakery	Thai Restaurant	Asian Restaurant
2	Archer Heights	Mexican Restaurant	Mobile Phone Shop	Bar	Clothing Store	Fast Food Restaurant	Grocery Store	Eastern European Restaurant	Taco Place	Big Box Store	Shoe Store
3	Armour Square	Chinese Restaurant	Pizza Place	Park	Bar	American Restaurant	Business Service	Bakery	Mexican Restaurant	Italian Restaurant	Asian Restaurant
4	Ashburn	Seafood Restaurant	Fast Food Restaurant	Park	Fried Chicken Joint	Hot Dog Joint	Clothing Store	Donut Shop	Discount Store	Sandwich Place	Furniture / Home Store

- Following these we use the '*K-means'* machine learning clustering algorithm in order to divide these nighborhoods into clusters based on the most common venues in their vicinity. The aim here is to understand the types of businesses and amenities which operate in these neighborhoods and therefore divide these neighborhoods based on their overall ambience and lifestyles.

- The K-means clustering algorithm was used as it easily scales to large datasets, is computationally faster than hierarchical clustering and generalizes clusters of different shapes and sizes.

- These clusters were then visualized by superimposing over our preciously made map of Chicago as follows:



- After the clustering was done, *Foursquare API* was used again to find out which neighborhoods had a Japanese restaurant as according to our business problem, the new business should be set up in close vicinity to Japanese restaurants. All the Japanese restaurants and their neighborhoods were put into a new *Pandas* dataframe:

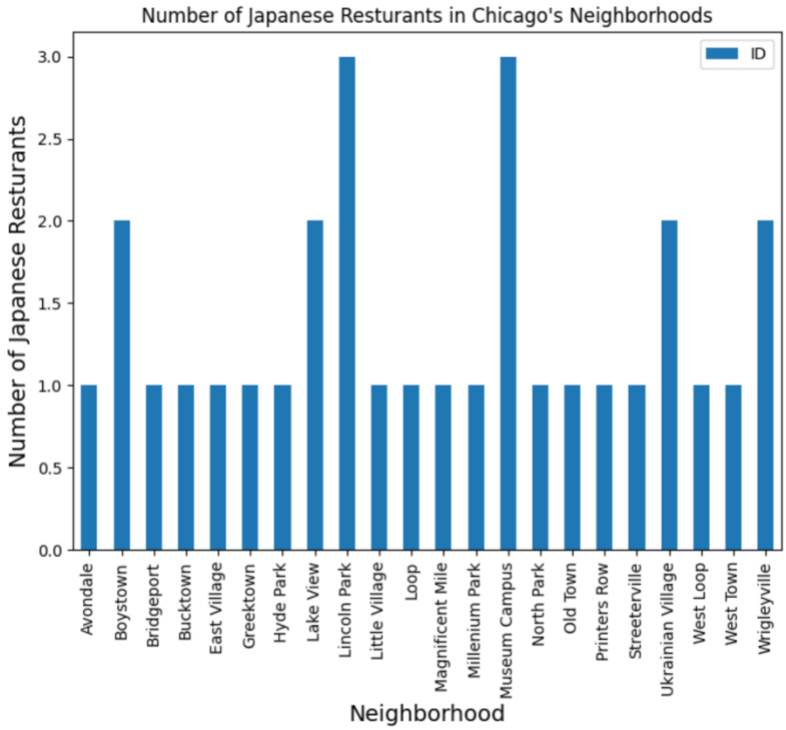
[39]:
```
# Checking the Dataframe
print(japanese_rest_chicago.shape)
japanese_rest_chicago
```

```
(30, 4)
```

[39]:

	Borough	Neighborhood	ID	Name
0	PRINTERS ROW	Printers Row	51018fa2e4b0ca484f8ac272	UMAI Japanese Kitchen & Sushi
1	LITTLE VILLAGE	Little Village	51018fa2e4b0ca484f8ac272	UMAI Japanese Kitchen & Sushi
2	IRVING PARK	Avondale	57e584e7498e0eda481b3d60	Ryuu Asian BBQ
3	LOOP	Loop	51018fa2e4b0ca484f8ac272	UMAI Japanese Kitchen & Sushi
4	GREEKTOWN	Greektown	54296aa7498e36575ee9a8b6	Momotaro
5	MUSEUM CAMPUS	Museum Campus	51343046e4b08f70ffe26b8f	SumoSam
6	MUSEUM CAMPUS	Museum Campus	4efd2d1a9adffb90ca7a8fdb	Tokyo Tokyo
7	MUSEUM CAMPUS	Museum Campus	5be575095c68380039ad6a91	Pepper Lunch
8	LAKE VIEW	Lake View	4b78c82af964a52024e22ee3	Yoshi's Cafe
9	LAKE VIEW	Lake View	4c5851eab7a31b8d243d52da	Ukai Japanese Restaurant
10	LINCOLN PARK	Lincoln Park	5591d9b5498e462cb30a5f4c	Glaze Teriyaki
11	LINCOLN PARK	Lincoln Park	4b78c82af964a52024e22ee3	Yoshi's Cafe
12	LINCOLN PARK	Lincoln Park	5b1f1c55b9b37b002cb719f8	Gyu-Kaku
13	STREETERVILLE	Magnificent Mile	4c3e115bdb3b1b8d09e66495	Gyu-Kaku Japanese BBQ
14	WEST LOOP	West Loop	51018fa2e4b0ca484f8ac272	UMAI Japanese Kitchen & Sushi
15	NORTH PARK	North Park	4b11d699f964a520be8523e3	Midori
16	STREETERVILLE	Streeterville	4c3e115bdb3b1b8d09e66495	Gyu-Kaku Japanese BBQ
17	MILLENIUM PARK	Millenium Park	4c3e115bdb3b1b8d09e66495	Gyu-Kaku Japanese BBQ
18	HYDE PARK	Hyde Park	4b5291f9f964a5209f8227e3	Kikuya Japanese Restaurant
19	BUCKTOWN	Bucktown	5483b98e498ed2ba51fe0caa	Izakaya Mita
20	WRIGLEYVILLE	Wrigleyville	4c5851eab7a31b8d243d52da	Ukai Japanese Restaurant
21	WRIGLEYVILLE	Wrigleyville	4b78c82af964a52024e22ee3	Yoshi's Cafe

- Using simple descriptive statistics, a bar chart showing the number of Japanese restaurants in each of these neighborhoods was plotted as follows:



As can be seen in the above graph, neighborhoods like *Museum Campus, Lincoln Park, Lakeview, Ukrainian Village, Wrigleyville and Boystown* have the highest number of established Japanese Restaurants in the city of Chicago. In addition to these, some other neighborhoods also have a well-established Japanese restaurant. This indicates that Japanese food is seen as a delicacy which is enjoyed on special occasions and not as regularly as fast-food variants such as hamburgers and pizzas.

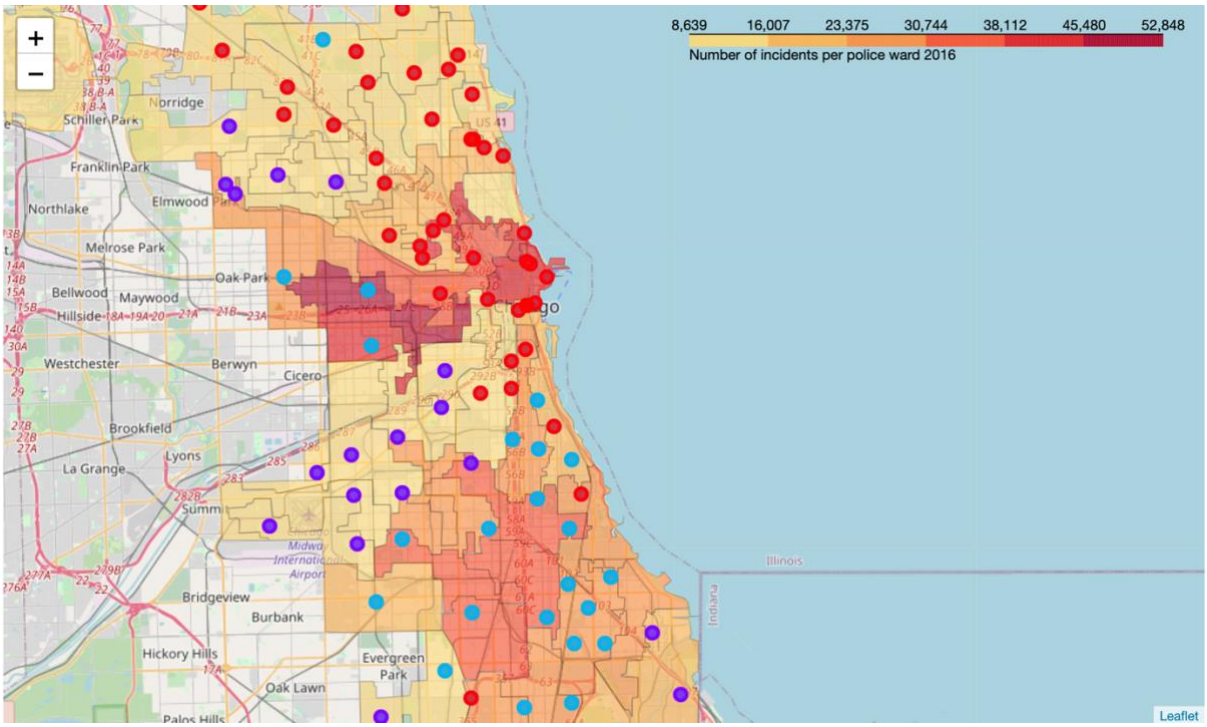
- Following this, customer ratings and tips were obtained for each Japanese restaurant using *Foursquare API* and the average rating for restaurants in each neighborhood was displayed in a *Pandas* dataframe as follows:

[48]:

	Neighborhood	Average Rating
20	West Town	9.000000
11	Magnificent Mile	8.900000
17	Streeterville	8.900000
12	Millenium Park	8.900000
5	Greektown	8.700000
19	West Loop	8.600000
4	East Village	8.600000
16	Printers Row	8.600000
9	Little Village	8.600000
10	Loop	8.600000
1	Boystown	8.400000
21	Wrigleyville	8.400000
7	Lake View	8.400000
14	North Park	8.300000
18	Ukrainian Village	8.050000
8	Lincoln Park	8.000000
3	Bucktown	7.900000
0	Avondale	7.800000
15	Old Town	7.700000
2	Bridgeport	7.600000
6	Hyde Park	7.400000
13	Museum Campus	7.266667

4.0 Results and Discussion

Visualizing the clusters:

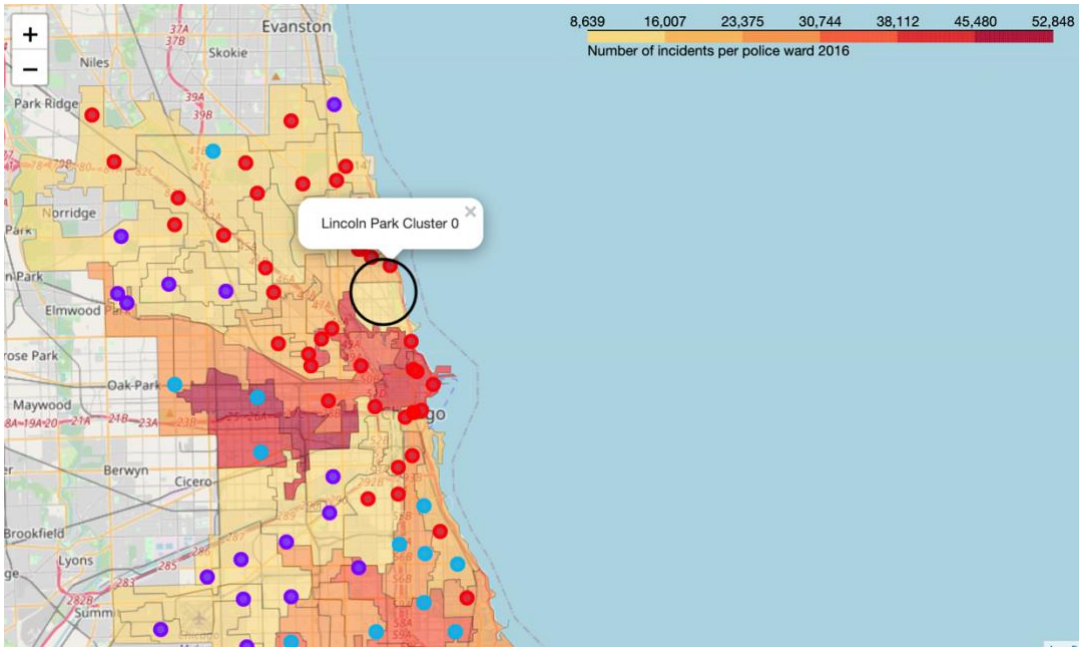


Looking at the average ratings table above, *West Town*, *Magnificent Mile* and *Streeterville* fare the best in terms of customer ratings, but these neighborhoods have just one established Japanese restaurant. Although *Museum Campus* has 3 Japanese Restaurants, the average rating indicates that the quality of food and customer satisfaction is not up to the mark.

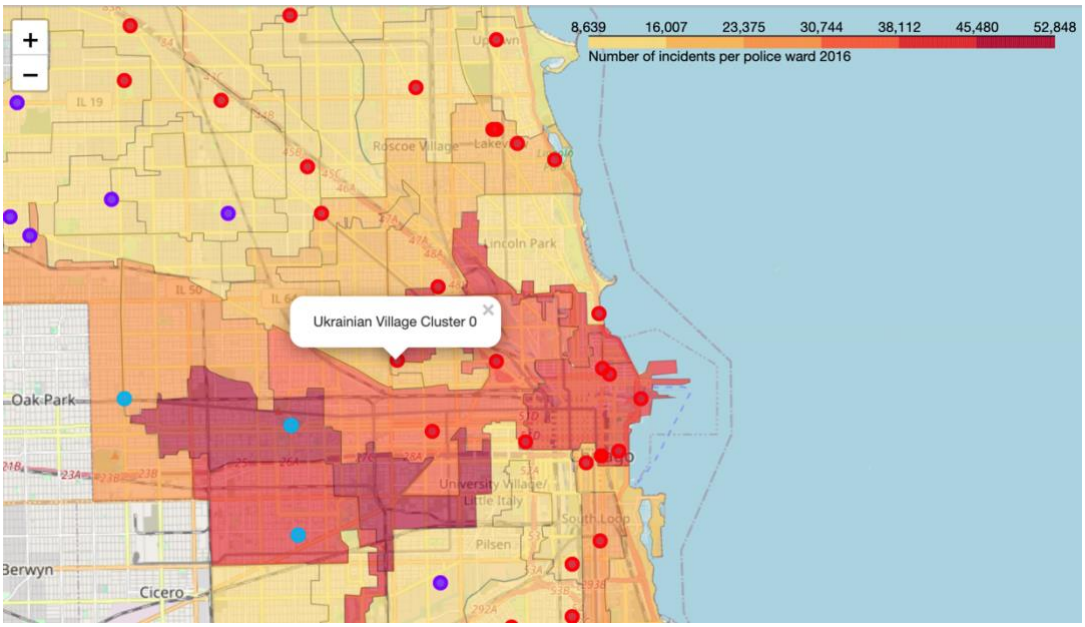
On the other hand, *Lincoln Park* which also has 3 Japanese restaurants has a decent average rating of 8.0. *Boystown*, *Lakeview*, *Wrigleyville* and *Ukrainian Village* which all have 2 well-established Japanese restaurants receive really good average ratings around 8.4 and recommendations from customers. Therefore, these neighborhoods as well as *Lincoln Park* are well suited for establishing a Japanese cuisine ingredients suppliers business.

After superimposing these results with our previous clustering and crime intensity map (see map), it can be seen that out of these shortlisted neighborhoods, *Ukrainian Village* can be rejected as it lies in an area where the crime intensity is very high (see image below).

The best option for setting up the business is around the *Lincoln Park* neighborhood (see image below), as the crime intensity is low and it lies in cluster 0 (red cluster on the map) with numerous recreational venues like cafes, parks, bookshops as well as **3 Japanese restaurants** situated in the vicinity. This makes it an ideal location for setting up a Japanese ingredients business as well as settling with a young family. The other nearby options include *Boystown, Wrigleyville and Lake View* which are neighborhoods situated in close proximity to the Lincoln Park area.



This image shows the Lincoln Park Neighborhood and surrounding area, which according to our analysis, is the ideal location for setting up the traditional Japanese ingredients business and also for raising a young family.



This image shows the high crime intensity neighborhoods of Chicago and the Ukrainian Village neighborhood in close proximity.

5.0 Conclusion

This project provides just a basic outline for setting up a new business in the city of Chicago and helps prospective businessmen get a better understanding of the localities in terms of the most common popular venues in and around these neighborhoods. It is a demonstration of the skills acquired and using the latest techniques in data science to solve a real-world problem as it is always a good idea to explore all possible options before setting up a new business or settling in a new city with a family. Safety should always be number one priority and we have inculcated that in our analysis by superimposing the best neighborhood locations onto a choropleth map of Chicago showing the intensity of crime in different areas. An extension to this project can be a detailed analysis of the budgetary constraints of a new business and finding neighborhoods which are not too expensive in terms of enjoying basic amenities.

After our analysis, we can recommend that the area surrounding the ***Lincoln Park*** neighborhood (*as shown in the above image*) is best suited to our new business and settling down with a young family. Some alternative options in the vicinity include- *Boystown*, *Wrigleyville* and *Lake View*.

Thanks for Reading!