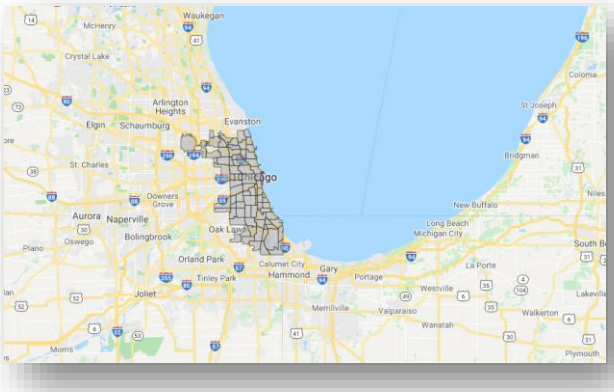
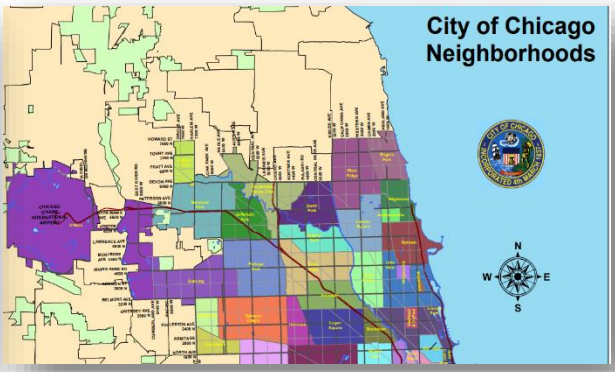
# The Battle of Neighborhoods – Chicago Market Entry

## 1. Introduction and Problem Statement

One of my clients mentioned to me that they are planning to setup an **Indo Chinese** restaurant in Chicago. However, since he is not from Chicago, he wanted me to do some analysis on the different neighborhoods in Chicago and recommend 2 neighborhoods where he will have greater chances of succeeding if he sets up his restaurant. He was open to suburbs and the downtown areas of Chicago. Based on the recommendation, he will perform a physical feasibility study and make his decision. These recommendations will speed his time to entry into the Chicago market.

This type of analysis will be useful to any business trying to expand into other markets and doesn't want to or have the necessary local presence or intelligence to make a **Go or No-Go** decision based on the local market data. A community area boundaries maps is shown below.




## 2. Data Description

The data for performing the explorations come from the following sources:

- **Chicago neighborhood dataset** - https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago . This data will help us group and analyze different neighborhoods based on the community areas.
- **Location Data from Foursquare** - for each neighborhood venues and venues categories were gathered for analysis, narrowing down of the neighborhoods to top 5 and clustering. In addition, we can get a granular view of the types of businesses in these neighborhoods.

The Chicago dataset sample is shown below:

List of neighborhoods by community area [ edit ]

| Neighborhood | Community area |
|---|---|
| Albany Park | Albany Park |
| Altgeld Gardens | Riverdale |
| Andersonville | Edgewater |
| Archer Heights | Archer Heights |
| Armour Square | Armour Square |
| Ashburn | Ashburn |
| Ashburn Estates | Ashburn |
| Auburn Gresham | Auburn Gresham |
| Avalon Park | Avalon Park |
| Avondale | Avondale |
| Avondale Gardens | Irving Park |
| Back of the Yards | New City |
| Belmont Central | Belmont Cragin |
| Belmont Gardens | Hermosa |
| Belmont Heights | Dunning |
| Belmont Terrace | Dunning |
| Beverly | Beverly |
| Beverly View | Ashburn |

As we can see from the dataset, neighborhoods are organized by community areas. This provides an indication that we narrow down to the best 2 neighborhoods, we need to group data by community areas and rank them. However, as seen this dataset doesn't have the Latitude and Longitude coordinates. So, we need to supplement this dataset with the location for each neighborhood obtained from using the GeoLocator service.

### 3. Methodology Overview

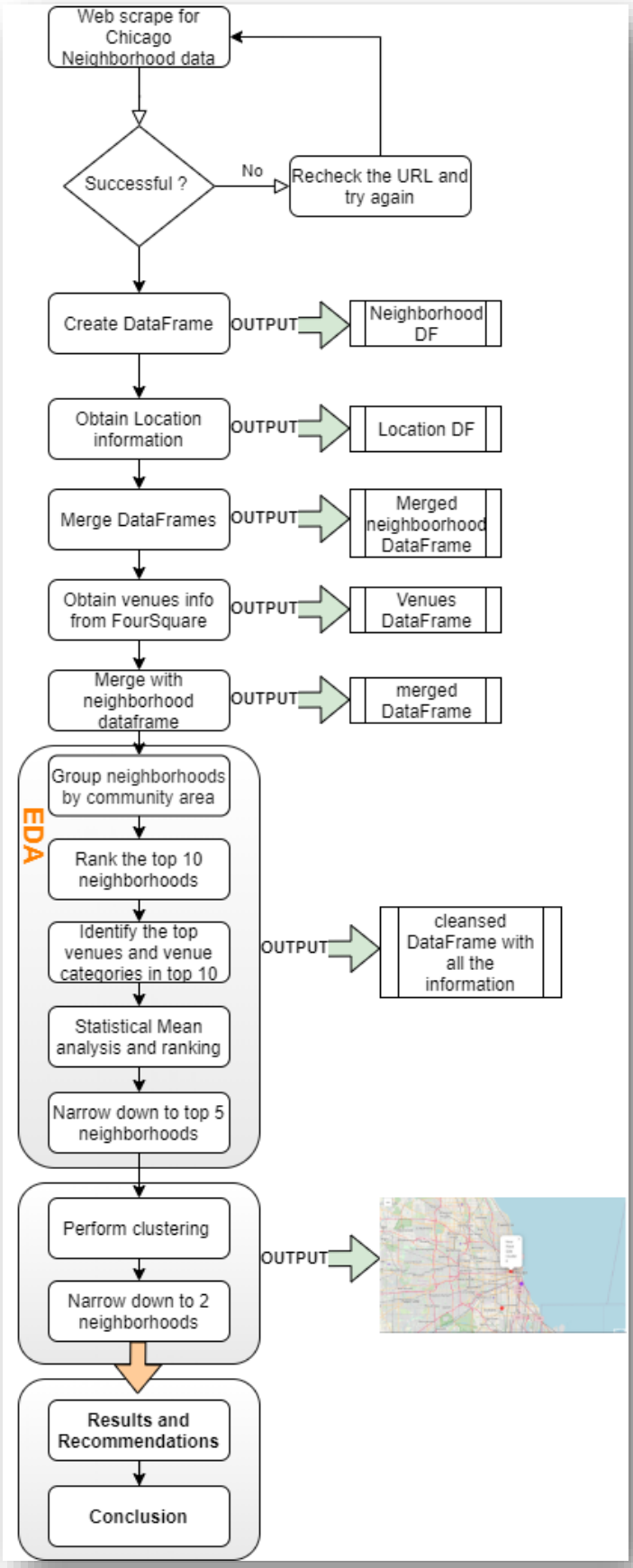
The key steps involved in this project include:

- Web scrape Community and Neighborhood data, process it and store it into Pandas DataFrame. I created few functions that will make getting the data

easier. These functions get data such as GeoLocator information, venues and venue details based on the neighborhood information.

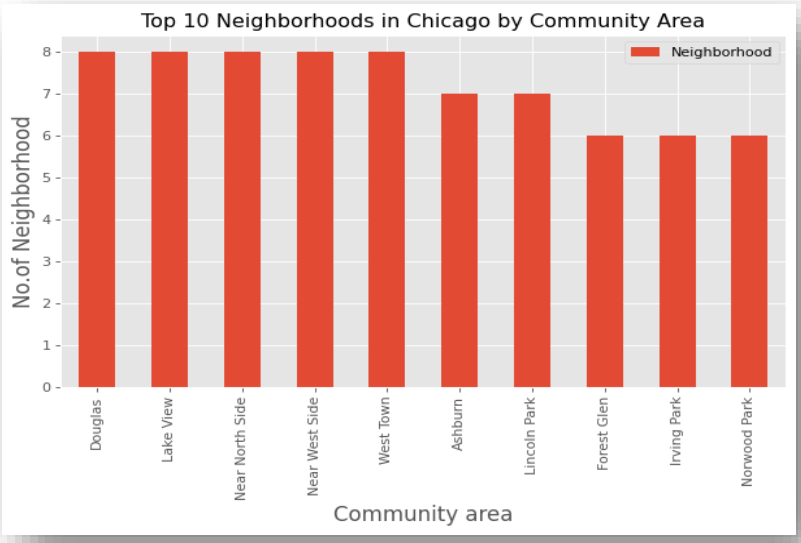
- Obtain data from Foursquare such as venues, venue categories etc. that would complement the neighborhood data.
- Determine the location coordinates and append it to the neighborhood data
- Perform exploratory data analysis (EDA) on the merged data (Foursquare and neighborhood) such as -
    - grouping of neighborhoods by community area,
    - listing the top 10 neighborhoods and identifying the top venues and venue subcategories in these neighborhoods
    - narrowing down to top 5 neighborhoods using statistical means comparison
- Perform unsupervised machine learning K-mean clustering algorithm and mapping to narrow down to top 2 neighborhoods
- Provide recommendations and conclusion

The flowchart below describes the steps and the outputs of these steps.


Web scrape for Chicago Neighborhood data
Successful ?
No
Recheck the URL and try again
Create DataFrame
OUTPUT
Neighborhood DF
Obtain Location information
OUTPUT
Location DF
Merge DataFrames
OUTPUT
Merged neighboorhood DataFrame
Obtain venues info from FourSquare
OUTPUT
Venues DataFrame
Merge with neighborhood dataframe
OUTPUT
merged DataFrame
EDA
Group neighborhoods by community area
Rank the top 10 neighborhoods
Identify the top venues and venue categories in top 10
Statistical Mean analysis and ranking
Narrow down to top 5 neighborhoods
OUTPUT
cleansed DataFrame with all the information
Perform clustering
Narrow down to 2 neighborhoods
OUTPUT
Results and Recommendations
Conclusion

## 4. Results

Based on the grouping of neighborhoods based on the community areas and ranking the top 10, we get the following bar chart.



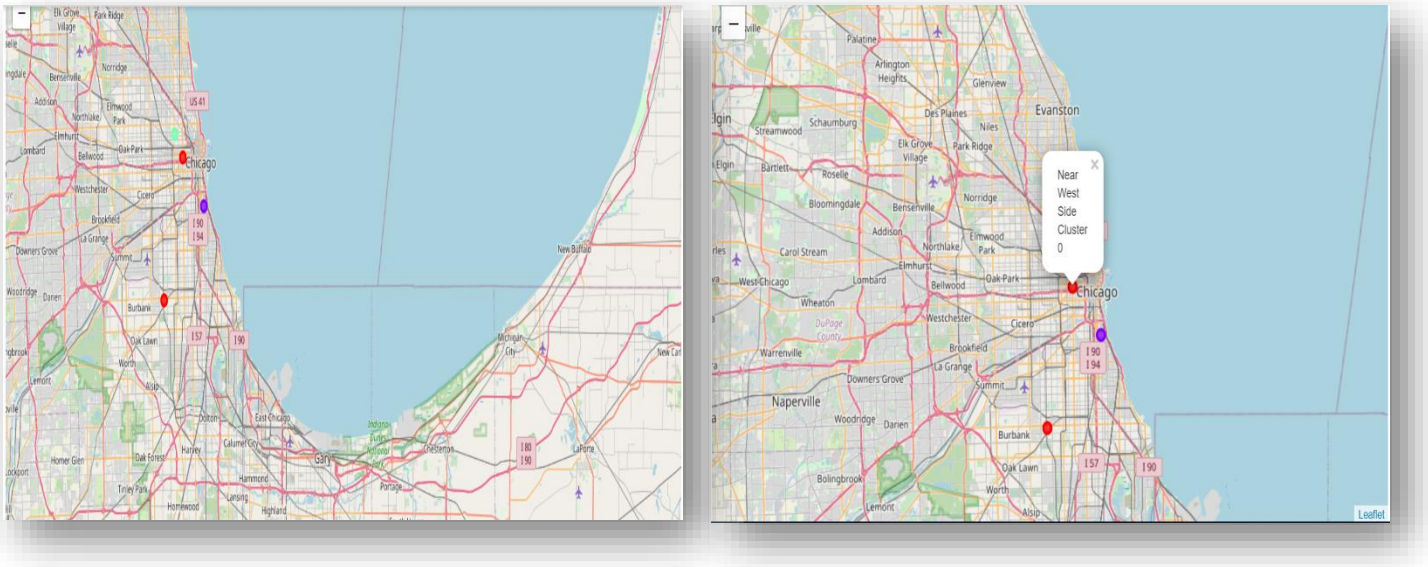

After gathering the venue categories, I ran the mean statistical analysis on the neighborhoods. This helped me to trim down the interested neighborhoods to 4 as seen in the DataFrame below.

| | Neighborhood | American Restaurant | Bakery | Bank | Bar | Big Box Store | Burger Joint | Bus Station | Chinese Restaurant | Cocktail Bar | Coffee Shop | Donut Shop | Fried Chicken Joint | Gas Station | Grocer Stor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Ashburn | 0.0 | 0.000000 | 0.071429 | 0.071429 | 0.071429 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.071429 | 0.000000 | 0.000000 | 0.071429 | 0.14285 |
| 1 | Douglas | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00000 |
| 2 | Near West Side | 0.0 | 0.071429 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.071429 | 0.071429 | 0.071429 | 0.000000 | 0.071429 | 0.071429 | 0.000000 | 0.14285 |
| 3 | West Town | 0.2 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.2 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00000 |

By obtaining the top 8 venues for the top 4 neighborhoods, it became clear that the all these 4 neighborhoods were good candidates for a feasibility study.

| | 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 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Ashburn | Mobile Phone Shop | Mexican Restaurant | Grocery Store | Wings Joint | Gym / Fitness Center | Bank | Bar | Big Box Store |
| 1 | Douglas | Park | Wings Joint | Fried Chicken Joint | Bakery | Bank | Bar | Big Box Store | Burger Joint |
| 2 | Near West Side | Grocery Store | Hookah Bar | Karaoke Bar | Bus Station | Donut Shop | Fried Chicken Joint | Sandwich Place | Chinese Restaurant |
| 3 | West Town | History Museum | American Restaurant | Post Office | Burger Joint | Fried Chicken Joint | Bakery | Bank | Bar |

However, after I ran the k-means clustering on the data and added the clusters to the data frame I wanted to map out all the different clusters to get a nice visualization of the similar neighborhoods. Based on the above map, it looks like the clusters are centered around the 2 neighborhoods - **Near West Side** and **Ashburn**. This giving us our 2 clusters that were the best for doing a feasibility study.




## 5. Recommendations

As part of improving the granularity of the clusters, I would recommend expanding the top "n" neighborhoods from 8 to 15 and increasing the venue from top 8 to top 12. This would bring in additional data points that will increase the granularity of the data. Another dimension that will help us further narrow down the 2 locations is the "sentiment" analysis data such as Tips and recommendation data from Foursquare. Also, if we can cross correlate with other specific Indo Chinese (or lack of) that will further strength our analysis.

## 6. Conclusions

Based on the number of neighborhoods, the number of venues within these neighborhoods and the number of variety of venues in the neighborhoods, I would recommend the 2 best places to start a new restaurant business is in the "**Near West Side**" and "**Ashburn**" neighborhoods. As additional support the venues details indicate that there are not that many restaurants these areas, which provides a good probability of success for my client succeed if they open a restaurant in these neighborhoods.