

Capstone Project-The Battle of Neighborhoods



CHICAGO ILLINOIS

Introduction



- **Background**

decision on store-locations is one of the most important strategic decisions the retailer has to make for its long term success. Finding the right location for a new store is a process that takes careful consideration. Population, neighborhood demographics, income distribution, crime rate, local competition are all factors taken into consideration when grocery chains look for a new store location.

- **Problem**

The aim of this project is to find location for opening of retail store, specifically grocery stores location across Chicago. This report will be targeted to stakeholder's interest in opening grocery store in community area with low crime rate and higher per capita income.

Methodology



Data Acquisition

The data acquired for this project is a combination of data from three sources.

- The first source of data is scraped from a [Wikipedia page](#) that contains the list of Chicago community area.
- The second data source of the project uses a [Chicago crime data](#) that shows the crime per community area in Chicago.
- Third data source is [Chicago Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012](#). This dataset contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” for each Chicago community area, for the years 2008 – 2012

Methodology



Data cleaning and processing

The data cleaning for each of the three sources of data is done separately.

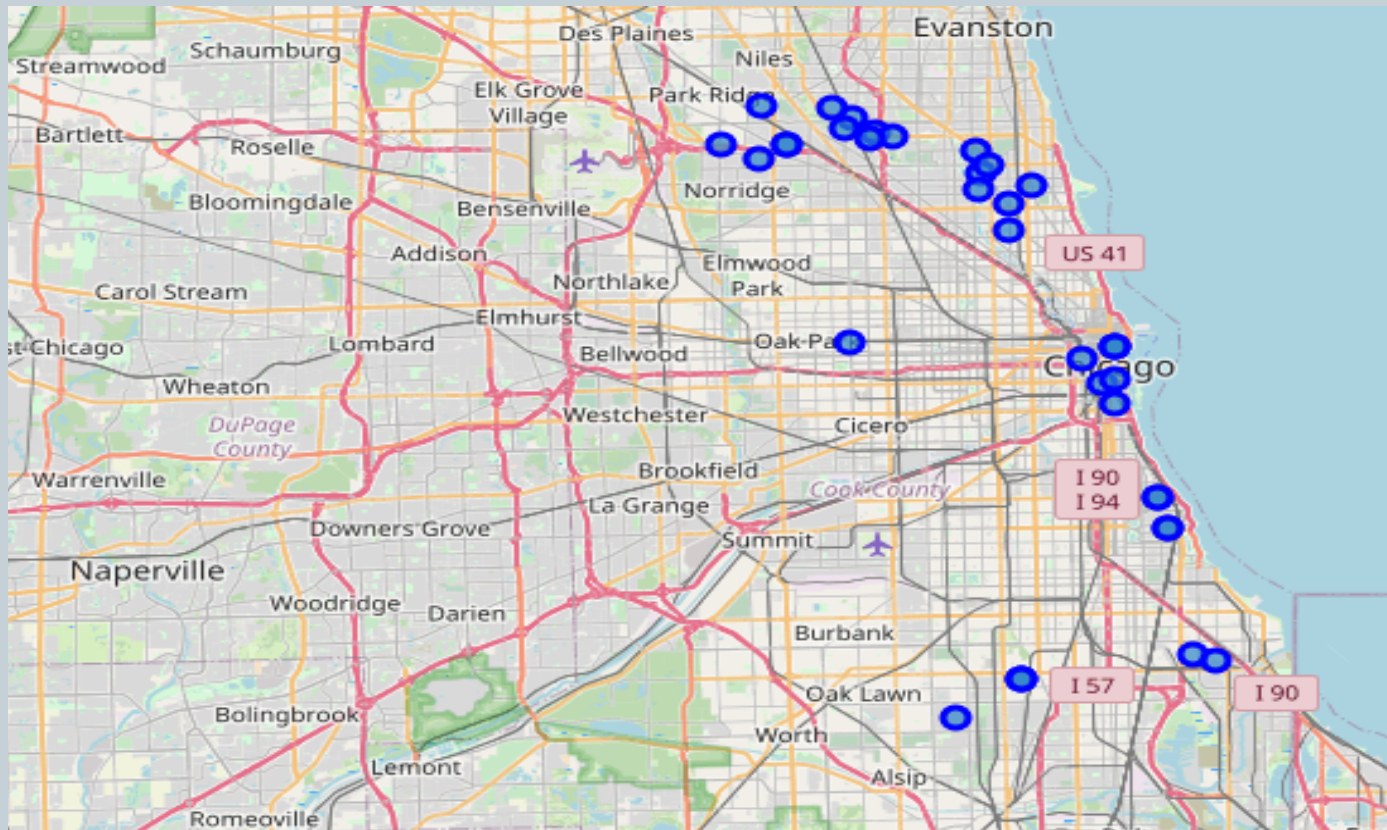
- The First data is scraped from a [Wikipedia page](#) using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website
- The second data from the [Chicago crime data](#), the crimes during the most recent year (2020) are only selected. The major categories of crime are segregated group by community area code to get the total crimes per the community area
- The third dataset [Chicago Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012](#)

Data cleaning and processing(Cont.)

- The two datasets are merged on the Community area names to form a new dataset that combines the necessary information in one dataset
- The coordinates of the neighborhoods is be fetched using Open Cage Geocoder to create a final consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighborhood's co-ordinates

	community area	Neighborhood	Latitude	Longitude	per_cap_income	Total_Case
0	Beverly	Beverly	41.718153	-87.671767	39523	236
1	Beverly	East Beverly	41.718153	-87.671767	39523	236
2	Beverly	West Beverly	41.718153	-87.671767	39523	236
3	Norwood Park	Big Oaks	41.885310	-87.622130	32875	221
4	Norwood Park	Norwood Park East	41.985590	-87.800582	32875	221

Neighborhood in Chicago



Neighborhoods with high per capita income and lowest crime rate were selected. There are 35 neighborhoods which has been selected with above criteria, they are visualized on a map using folium on python

Modeling

Using the final dataset containing the selected neighborhoods along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API

```
(1080, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beverly	41.718153	-87.671767	Ridge Park	41.718378	-87.667921	Park
1	Beverly	41.718153	-87.671767	Jimmy Jamm's Sweet Potato Pies	41.721181	-87.669373	Bakery
2	Beverly	41.718153	-87.671767	Top-Notch Beefburgers	41.721281	-87.675382	Burger Joint
3	Beverly	41.718153	-87.671767	Southtown Health Foods	41.721257	-87.674822	Grocery Store
4	Beverly	41.718153	-87.671767	GBN Nail Salon	41.721371	-87.668500	Cosmetics Shop

One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods

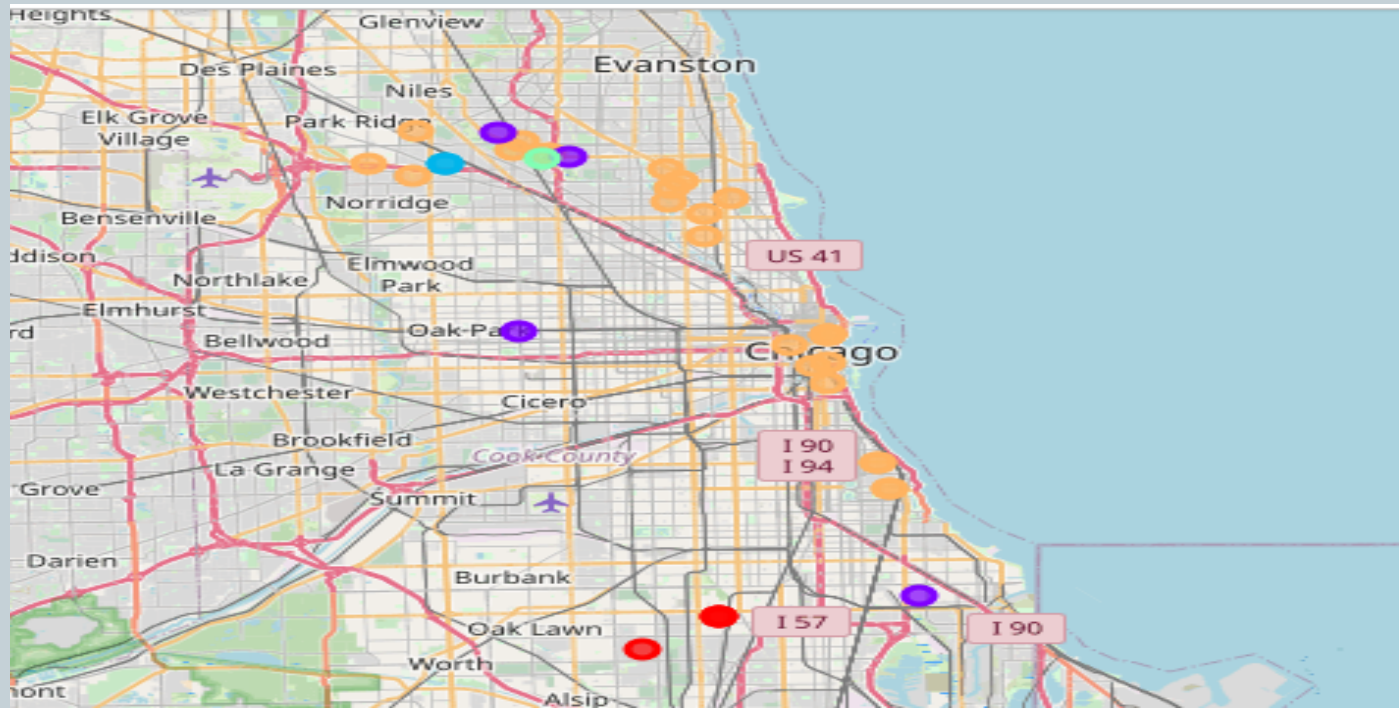
```
neighborhoods_venues_sorted.head()
```

	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	Beverly	Cosmetics Shop	Grocery Store	Park	Italian Restaurant	Farmers Market	Burger Joint	Boutique	Train Station	Bakery	Yoga Studio
1	Big Oaks	Hotel	Coffee Shop	Plaza	Steakhouse	Seafood Restaurant	Park	Sandwich Place	American Restaurant	Hotel Bar	Museum
2	Bowmanville	Sandwich Place	New American Restaurant	Ice Cream Shop	Bar	Filipino Restaurant	Garden	Dive Bar	Supermarket	Mobile Phone Shop	Coffee Shop
3	Budlong Woods	Bakery	Middle Eastern Restaurant	Karaoke Bar	Nightclub	Discount Store	Sushi Restaurant	Cajun / Creole Restaurant	Food & Drink Shop	Mexican Restaurant	Greek Restaurant
4	Calumet Heights	Bus Station	Gym / Fitness Center	Park	Yoga Studio	Garden	French Restaurant	Fountain	Football Stadium	Food Truck	Food Court

Results



After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library



Each cluster is color coded for the ease of presentation

Cluster 1: Looking into the neighborhoods in the first cluster. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Cosmetic Shop, Grocery Store, Parks, Bakery, Restaurant and stadium

	Neighborhood	per_cap_income	Total_Case	Cluster Labels	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	Common Venue
0	Beverly	39523	236	0	Cosmetics Shop	Grocery Store	Park	Italian Restaurant	Farmers Market	Burger Joint	Boutique	Train Station	Bakery	S
1	East Beverly	39523	236	0	Cosmetics Shop	Grocery Store	Park	Italian Restaurant	Farmers Market	Burger Joint	Boutique	Train Station	Bakery	S
2	West Beverly	39523	236	0	Cosmetics Shop	Grocery Store	Park	Italian Restaurant	Farmers Market	Burger Joint	Boutique	Train Station	Bakery	S
30	Mount Greenwood	34381	130	0	Cosmetics Shop	Vineyard	Home Service	Park	Yoga Studio	Falafel Restaurant	French Restaurant	Fountain	Football Stadium	-

Cluster 2 Looking into the neighborhoods in the second cluster clustered from, Forest glen, Calumet Heights, Near South Side which consist of venue such as Bus Station, Park, Nature Preserve

	Neighborhood	per_cap_income	Total_Case	Cluster Labels	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	
14	Calumet Heights	28887	358	1	Bus Station	Gym / Fitness Center	Park	Yoga Studio	Garden	French Restaurant	Fountain	Football Stadium	Food Truck	
15	Central Station	59077	492	1	Park	Bus Station	Intersection	Liquor Store	Gym	Train Station	Grocery Store	Shoe Repair	Donut Shop	
24	Sauganash	44164	153	1	Park	Indian Restaurant	Asian Restaurant	Fast Food Restaurant	Pharmacy	Yoga Studio	Falafel Restaurant	French Restaurant	Fountain	
26	Wildwood	44164	153	1	Nature Preserve	Trail	Theater	Park	French Restaurant	Fountain	Football Stadium	Food Truck	Food Court	

Cluster 3: Neighborhood in third cluster ,clustered from, Norwood Park having similar common venues, the most common venues in these neighborhoods are Park, Yoga studio, Farmers market

	Neighborhood	per_cap_income	Total_Case	Cluster Labels	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
4	Norwood Park East	32875	221	2	Park	Yoga Studio	Farmers Market	Furniture / Home Store	French Restaurant	Fountain	Football Stadium	Food Truck	Food Court	Food & Drink Shop
5	Norwood Park West	32875	221	2	Park	Yoga Studio	Farmers Market	Furniture / Home Store	French Restaurant	Fountain	Football Stadium	Food Truck	Food Court	Food & Drink Shop

Cluster 4: Looking into the neighborhood in fourth cluster, the most common venues in these neighborhoods are Golf Course, Yoga studio, Gas Station, Furniture Store , French Restaurant

	Neighborhood	per_cap_income	Total_Case	Cluster Labels	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
25	South Edgebrook	44164	153	3	Golf Course	Yoga Studio	Gas Station	Furniture / Home Store	French Restaurant	Fountain	Football Stadium	Food Truck	Food Court	Food & Drink Shop

Activate Windows
Go to PC settings to activate

Cluster 5: Looking into the neighborhood in fifth cluster. The cluster five is the biggest cluster with 24 of the 35 neighborhoods. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Hotel Cafe, Restaurants, Football Stadium, Coffee Shop, and Yoga Studio

	Neighborhood	per_cap_income	Total_Case	Cluster Labels	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	
3	Big Oaks	32875	221	4	Hotel	Coffee Shop	Plaza	Steakhouse	Seafood Restaurant	Park	Sandwich Place	American Restaurant	
6	Old Norwood	32875	221	4	Hotel	Pizza Place	Café	Gym / Fitness Center	Parking	Metro Station	Grocery Store	Sandwich Place	Co
7	Oriole Park	32875	221	4	Football Stadium	Video Store	Gym	Park	Yoga Studio	Falafel Restaurant	French Restaurant	Fountain	F
8	Union Ridge	32875	221	4	Coffee Shop	Food Truck	Sandwich Place	Gym	Mediterranean Restaurant	Mexican Restaurant	American Restaurant	Salad Place	P
9	Bowmanville	37524	561	4	Sandwich Place	New American Restaurant	Ice Cream Shop	Bar	Filipino Restaurant	Garden	Dive Bar	Supermarket	Pr

Results and Discussion



The objective of the business problem was to help stakeholders identify one of the safest Neighborhood with higher per capita income in Chicago Illinois, and an appropriate neighborhood within the community area to set up a commercial establishment especially a Grocery store. This has been achieved by first making use of Chicago crime data to identify a safe community area with considerably higher per capita income for any business to be viable. After selecting the community area it was imperative to choose the right neighborhood where grocery shops were not among top 10 most common venues in a close proximity to each other. We achieved this by grouping the neighborhoods into clusters to assist the stakeholders by providing them with relevant data about venues and safety and population with higher individual income of a given neighborhood.

Conclusion



We have explored the crime data to understand different types of crimes in all Community area of Chicago and later segregate them based on per capita income, this helped us for selecting area with high per capita income and lowest crime rate. Once we confirmed the community area the number of neighborhoods for consideration also comes down, we further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.