Segmenting the Community Areas of Chicago for offsite ATM installations

Shagul Hameed Meeran

05 September 2019

1. Introduction

1.1 Background

Banks predominantly start with locating ATM in their branches. That establishes the primary ATM (Core) network. Off-site ATMs are installed at places where foot-traffic is high.

Typically, a Retail Bank's marketing department would research areas of interest where the bank should have its off-site ATM installed. Such research is grouped by various parameters, of which the following are key parameters:

- Estimated traffic per day
- Location popularity
- Possible venues that generate cash transactions (e.g. Restaurants, Groceries)
- Service Routes (for Security company, is the ATM within a route or away from it)
- Other ATMs in the vicinity
- Leasing space issues / costs
- Marketing potential
- Communication & Power infrastructure
- Threat scale for vandalism and/or crime
- Competition Factor (if there are multiple ATMs in the same location)
- Vestibule Preferences

All these parameters are taken into account and prioritized and then the banks then take the next step in trying to find out the operating costs of such an ATM

1.2 Problem

Though there are some mechanisms to find out most of the above said parameters, it requires field work predominantly for the factors such as location popularity, other ATMs and available venues mix that promote cash transactions. This field work is meant to be time consuming and costly and also different banks would conduct the same exercise for the same locations.
If this field work is accompanied by some data from machine learning that provides the nature of the neighborhood, its popularity and the available venues mix, then the field work can be reduced a lot and focused only on the key possible areas and neighborhoods.

This machine learning should be able to segment the neighborhoods into different clusters based on the types of venues and popularity and other interested factors.

1.3 Interest

All commercial banks that needs to plan to open its offsite ATM locations or extending its existing ones would be able to use this segmentation to take some informed decisions. Also it gives back the information about the possible competition from other banks in every locations based on their existing ATM networks.

2. Data acquisition and cleaning

2.1 **Data Sources**

The primary list of community areas of Chicago can be web scrapped from the Wikipedia
https://en.wikipedia.org/wiki/Community_areas_in_Chicago

This forms the base data for our dataset. All other information will be added to these base data as additional columns or features.

There are the following additional information that would be obtained from the city of Chicago Portal

1. Grocery Stores
2. Sidewalk Café Permits

Along with the latitude and longitude of each community areas of Chicago would be provided with an external CSV file.

All the venues details for each community areas would be extracted from Foursquare API passing the latitude and longitude of each community area and created as features.

Also, not all venues are considered. The following filters will be applied to the venue list. The idea to prioritize the venues that promotes cash transactions.

1. Specific venues on arts & entertainment (Arcade, Art Gallery, Bowling Alley, Casino, Concert Hall, Historic Site, Memorial Site, Movie Theater, Museum – all types, Stadiums – all types, Theme Park, Zoo)
2. All type of restaurants, bakery, bistro, cafeteria, cafes, coffee shop, diner, etc.
3. All type of bars and nightclubs
4. All grocery and small stores

All these venues count will be merged for each community areas and finally the data will be grouped together into the following columns.

1. Entertainment
2. Food
3. Lifestyle

4. Retail
5. Utilities

Community Area	Latitude	Longitude	Entertainment	Tourism	Food	Night Life	Health	Retail

2.2 Data Cleaning

a) Extract the base information (community areas)

The base information of all community areas (community area number, community area name) would be extracted from a Wikipedia page.

Link to Wikipedia page: [Community areas in Chicago]

We shall read the html source with requests and extract the table data from it

	comareano	CommunityArea	Neighborhoods
0	8	Near North Side	Cabrini–Green,The Gold Coast,Goose Island,Magn...
1	32	Loop	Loop,New Eastside,South Loop,West Loop Gate
2	33	Near South Side	Dearborn Park,Printer's Row,South Loop,Prairie...
3	5	North Center	Horner Park,Roscoe Village
4	6	Lake View	Boystown,Lake View East,Graceland West,South E...

Figure 1: Base data extracted from Wikipedia

b) Get the community areas and zip codes mapping

To add additional information to the above dataset, it is required to have zip codes as all these datasets are having zip codes only. So the mapping has been loaded from an external CSV.

	comareano	zip_code
0	1	60626
1	1	60645
2	1	60660
3	2	60626
4	2	60645

Figure 2: Community Area No/Zipcode Mapping

c) Load sidewalk cafe data from Chicago data portal

Chicago Data portal provides API for many different information about Chicago neighbourhoods. We shall use the sidewalk cafe permits data so as to get the count of sidewalk cafes at each neighbourhood

zip_code	sidewalkcafes
60601	525
60602	356
60603	264
60604	194
60605	551

Figure 3: Number of sidewalk cafe permits for each zipcode

d) **Load groceries data from Chicago data portal**

We shall use the groceries data so as to get the count of groceries at each neighbourhood. Again, this has been extracted from same data portal

zip_code	groceries
60601	3
60605	2
60607	4
60608	18
60609	20

Figure 4: Number of grocery stores

e) **Merge the sidewalk cafe and groceries data and aggregate at community area level**

We have the datasets for sidewalk cafes and groceries extracted from Chicago data portal. Also, we have dataset having zip code and community area number mapping. Let's merge all these datasets together using zip code column from each of these dataset.

After the dataset is merged, it should be aggregated at the community area level as our input data is required to be built at community area level
Finally merge the dataset with the base data extracted from Wikipedia site. And finally the dataset looks like this.

	comareano	CommunityArea	Neighborhoods	sidewalkcafes	groceries
0	8	Near North Side	Cabrini–Green,The Gold Coast,Goose Island,Magn...	3995.0	17.0
1	32	Loop	Loop,New Eastside,South Loop,West Loop Gate	3665.0	9.0
2	33	Near South Side	Dearborn Park,Printer's Row,South Loop,Prairie...	756.0	10.0
3	5	North Center	Horner Park,Roscoe Village	4919.0	42.0
4	6	Lake View	Boystown,Lake View East,Graceland West,South E...	4217.0	25.0

Figure 5: Merged data of community areas with total sidewalk cafes and groceries

f) Geocoding for community areas of Chicago

The geocoding for each community areas is loaded from an external CSV file. The same file will be loaded as a data frame and then merged with the final input created in the previous step.

	comareano	CommunityArea	Neighborhoods	sidewalkcafes	groceries	Latitude	Longitude
0	8	Near North Side	Cabrini–Green,The Gold Coast,Goose Island,Magn...	3995.0	17.0	41.9039	-87.6315
1	32	Loop	Loop,New Eastside,South Loop,West Loop Gate	3665.0	9.0	41.8786	-87.6251
2	33	Near South Side	Dearborn Park,Printer's Row,South Loop,Prairie...	756.0	10.0	41.8608	-87.6257
3	5	North Center	Horner Park,Roscoe Village	4919.0	42.0	41.9467	-87.6883
4	6	Lake View	Boystown,Lake View East,Graceland West,South E...	4217.0	25.0	41.9398	-87.6589

Figure 6: Community areas with geo coordinates

This completes first part of our data preparation process. Also this forms our base data. Before proceeding further, lets try to plot these community areas in a map.

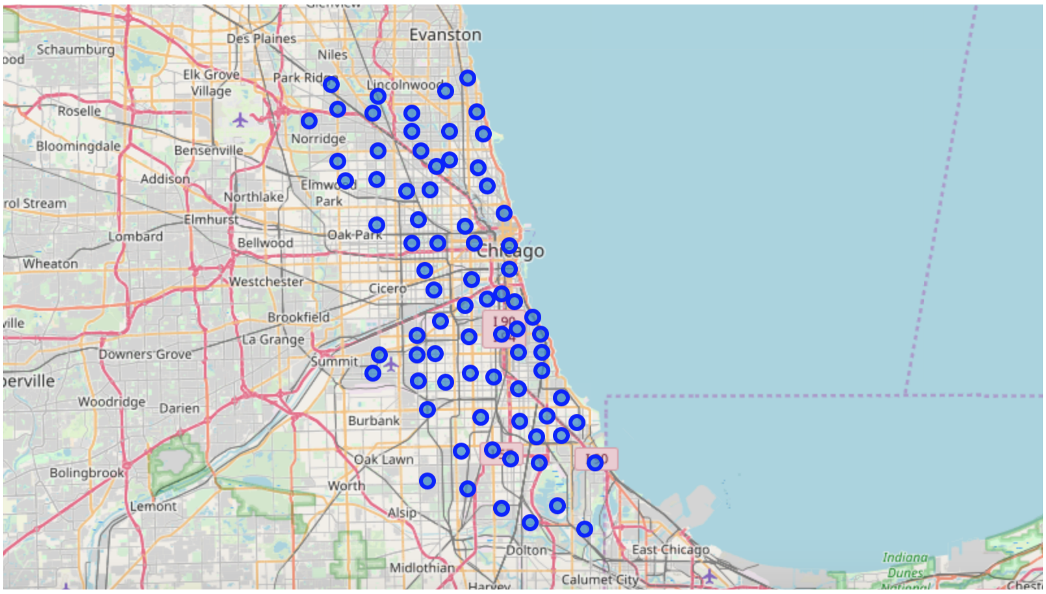

Figure 7: All Chicago community areas plotted (to be segmented later)

3. Exploratory Data Analysis

3.1. Build Venues Information

Only the Only the base information we collected so far might not be sufficient as features to cluster the neighbourhoods.
In addition to the base information we collected above for all the community areas, we should collect more information about the other venues and business outlets present in those areas. For this, we shall try to use the Foursquare Places API that provides venue details around any specific geographical location with a geo coordinate.

As we have already collected geo coordinates for each community area, let's use this to feed Foursquare API to scan all the venues in that neighbourhood (1500 metres radius) and extract the list for every community area and build the venues dataset out of it.

For this we have to write a function with Python and Pandas that would make a request to Foursquare API with latitude and longitude values of each community area and get back a JSON, which would be parsed to extract only the required information and then build a dataset out of it. Foursquare places API will produce the list of venues with details like Venue category, Venue name, Venue location and so on.

This will bring a new dataset like this.

	CommunityArea	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Near North Side	41.9039	-87.6315	Chicago Q	41.903599	-87.630277	BBQ Joint
1	Near North Side	41.9039	-87.6315	Sparrow	41.903277	-87.628972	Cocktail Bar
2	Near North Side	41.9039	-87.6315	Blue Door Kitchen & Garden	41.903171	-87.630508	Restaurant
3	Near North Side	41.9039	-87.6315	3 Arts Club Cafe	41.905800	-87.630470	Café
4	Near North Side	41.9039	-87.6315	Division Street Farmers Market	41.903908	-87.630015	Farmers Market

Figure 8: Venues extracted using Foursquare Places API

3.2 Venues Insights

We are only interested in the venue category, rather than other details. We have got 240 unique categories extracted. Let's take a quick look on how much data is extracted overall as venues information for each community area

CommunityArea	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Albany Park	34	34	34	34	34	34
Archer Heights	25	25	25	25	25	25
Armour Square	13	13	13	13	13	13
Ashburn	5	5	5	5	5	5
Auburn Gresham	4	4	4	4	4	4
Austin	11	11	11	11	11	11
Avalon Park	9	9	9	9	9	9
Avondale	34	34	34	34	34	34
Belmont Cragin	15	15	15	15	15	15
Beverly	17	17	17	17	17	17

Figure 9: Venues count for first few community areas

Not all community has a lot of venues, some of them have very least and some community areas like Pullman are not at all available. Let's take a quick visualisation of this distribution of venues below

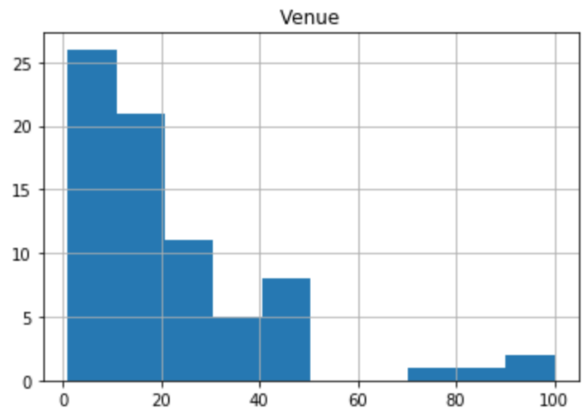

Figure 10: Distribution of Venues across community areas

As we see, the venues are mainly concentrated on a specific group of community areas, and for the rest of it, it's vert sparse. Our area of interest are those community areas having more venues, which means those areas churn out more transactions which in turn needs more ATM installations.

Also, not all the venues are our area of interest. We are particularly looking for venues which promotes cash transactions more than card transactions.
For example, a Coffee shop attracts more cash transactions than a French restaurant, same way, a Flower Shop would have more cash transactions than a Gym.

So, we have to categorize these venues into different buckets so as to tell which categories promotes cash transactions more

3.3 **Categorizing the Venues**

As pointed above, it is required to categorize the venues so as to separate the venue categories that promotes cash transactions.

First, all the venues will be grouped under larger buckets as:

- Entertainment
- Food
- LifeStyle
- Retail
- Utilities

Also, these buckets are further sub divided into type1 and type2, in which type1 would attract more cash transactions than type2.

Let's build these venue types with predefined list and then apply that to the venues we have extracted above by adding that category as an additional column.

	CommunityArea	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	venue_type
0	Near North Side	41.9039	-87.6315	Chicago Q	41.903599	-87.630277	BBQ Joint	Food Type1
1	Near North Side	41.9039	-87.6315	Sparrow	41.903277	-87.628972	Cocktail Bar	Entertainment Type1
2	Near North Side	41.9039	-87.6315	Blue Door Kitchen & Garden	41.903171	-87.630508	Restaurant	Food Type2
3	Near North Side	41.9039	-87.6315	3 Arts Club Cafe	41.905800	-87.630470	Café	Food Type1
4	Near North Side	41.9039	-87.6315	Division Street Farmers Market	41.903908	-87.630015	Farmers Market	Retail Type1

Figure 11: Dataset having the Venue Types (Larger buckets than Venue category)

3.4 **Build the feature set for clustering**

Let's do one hot encoding on this new column 'venue_type' and convert these built categories as features against the community area. This will bring all the venue types as columns against each venue row. Now every venue type becomes a feature.

	CommunityArea	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
0	Near North Side	0	0	1	0	0	0	0	0	0	0
1	Near North Side	1	0	0	0	0	0	0	0	0	0
2	Near North Side	0	0	0	1	0	0	0	0	0	0
3	Near North Side	0	0	1	0	0	0	0	0	0	0
4	Near North Side	0	0	0	0	0	0	1	0	0	0

Figure 12: Venue Types as features

Let's group these data by community area and count the number of venues found against each of these categories for every community area. These builds the feature set for the community areas which will be used in clustering of these community areas later

	CommunityArea	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
0	Albany Park	4	0	11	8	1	0	4	1	5	0
1	Archer Heights	3	0	4	7	0	0	5	1	1	4
2	Armour Square	1	0	4	5	1	0	1	0	0	1
3	Ashburn	0	0	1	1	1	0	0	0	1	1
4	Auburn Gresham	0	0	0	0	0	0	1	0	2	1

Figure 13: Venue Type Features grouped at community area level

With this dataset having feature set, it's time to merge the original dataset which contains all the information extracted so far from Chicago data portal and geo coordinates.

Also, let's create a dataset from this merged dataset to be used for clustering, by dropping all the unnecessary columns from the data frame. This will be final feature set of Chicago community areas

	CommunityArea	sidewalkcafes	groceries	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
8	Rogers Park	424.0	24.0	5.0	1.0	12.0	14.0	1.0	1.0	6.0	0.0	2.0	0.0
9	West Ridge	595.0	38.0	1.0	0.0	3.0	19.0	0.0	0.0	4.0	1.0	1.0	0.0
10	Uptown	1624.0	25.0	6.0	1.0	15.0	12.0	1.0	2.0	5.0	2.0	4.0	0.0
11	Lincoln Square	2336.0	68.0	12.0	4.0	18.0	18.0	0.0	11.0	9.0	6.0	2.0	2.0
3	North Center	4919.0	42.0	7.0	0.0	12.0	6.0	2.0	2.0	9.0	3.0	4.0	2.0

Figure 14: Final features set for clustering

3.5 **Convert the feature set to categorical values**

As all the columns in the feature set contains continuous numeric values, it should be converted to categorical values so as to defined as the features for clustering.

For this, let's identify the mean values for every column, and then categorize all the numerical values to categorical codes: High (1) and Low(0) based on whether the count value is less than mean value or not.

	CommunityArea	sidewalkcafes	groceries	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
8	Rogers Park	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0
9	West Ridge	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
10	Uptown	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
11	Lincoln Square	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0
3	North Center	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 15: Final features set with categorical values — all set and ready for clustering

4. Clustering of Chicago Community Areas

4.1 K-Means Clustering

K-Means clustering would be used to cluster the community areas with the features prepared above. We shall keep the cluster numbers as 3, to cluster all community areas into three clusters.

Once the cluster labels are identified for all community areas with K-Means clustering, let's merge the cluster labels computed above for each community area to the original data frame that contains all the venues information and other details about the community area including geo coordinates

The following provides the dataset with cluster labels resulted from K-Means clustering

	Cluster Labels	comareano	CommunityArea	Neighborhoods	sidewalkcafes	groceries	Latitude	Longitude	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2
8	0	1	Rogers Park	East Rogers Park	424.0	24.0	42.0106	-87.6696	5.0	1.0	12.0	14.0	1.0	1.0
9	1	2	West Ridge	Arcadia Terrace,Peterson Park,West Rogers Park	595.0	38.0	42.0006	-87.6926	1.0	0.0	3.0	19.0	0.0	0.0
10	0	3	Uptown	Buena Park,Argyle Street,Margate Park,Sheridan...	1624.0	25.0	41.9665	-87.6533	6.0	1.0	15.0	12.0	1.0	2.0
11	0	4	Lincoln Square	Ravenswood,Ravenswood Gardens,Rockwell Crossing	2336.0	68.0	41.9687	-87.6890	12.0	4.0	18.0	18.0	0.0	11.0
3	0	5	North Center	Horner Park,Roscoe Village	4919.0	42.0	41.9467	-87.6883	7.0	0.0	12.0	6.0	2.0	2.0

Figure 16: Cluster labels define which cluster that community areas belongs to

Let's plot the final data frame that contains the cluster labels into the Chicago map, which would represent the clusters easily. Each cluster is defined by a different colour marker.

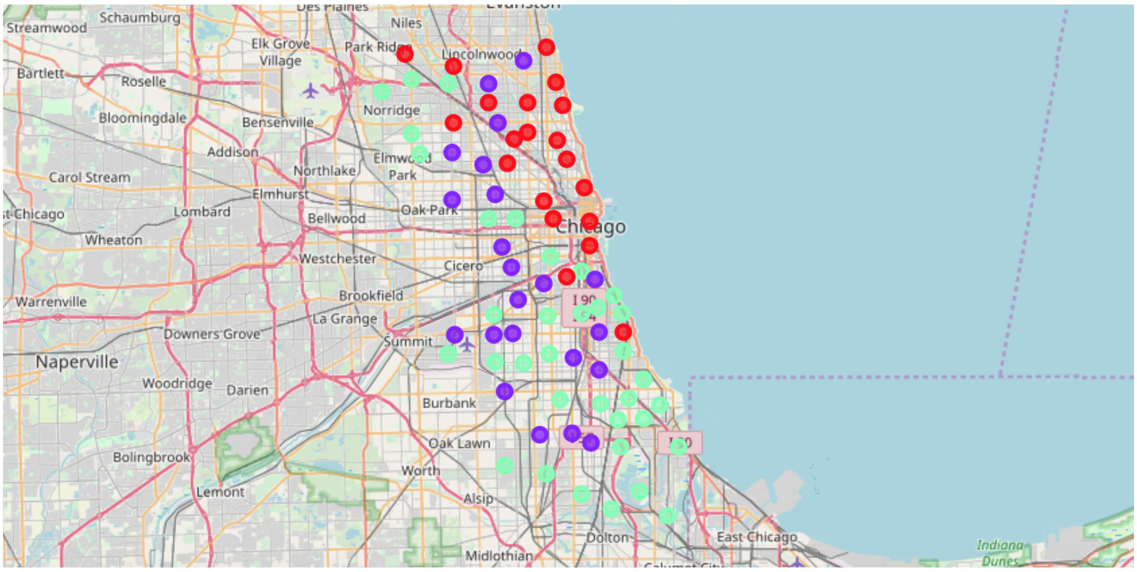

Figure 17: Chicago Community Area Clusters

Here cluster 1 are red markers, cluster 2 are blue markers and cluster 3 are green markers.

4.2 Cluster Insights

From the clusters built above, we can try to get some insights about the nature of each cluster so as to pick the cluster which would be good for new ATM installations

Let's try to load the community areas of each cluster to understand the type of area.

Cluster 1 community areas:

CommunityArea	sidewalkcafes	groceries	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
Rogers Park	424.0	24.0	5.0	1.0	12.0	14.0	1.0	1.0	6.0	0.0	2.0	0.0
Uptown	1624.0	25.0	6.0	1.0	15.0	12.0	1.0	2.0	5.0	2.0	4.0	0.0
Lincoln Square	2336.0	68.0	12.0	4.0	18.0	18.0	0.0	11.0	9.0	6.0	2.0	2.0
North Center	4919.0	42.0	7.0	0.0	12.0	6.0	2.0	2.0	9.0	3.0	4.0	2.0
Lake View	4217.0	25.0	7.0	4.0	11.0	4.0	0.0	4.0	1.0	0.0	1.0	0.0
Lincoln Park	5804.0	46.0	21.0	9.0	25.0	18.0	2.0	7.0	3.0	5.0	3.0	2.0
Near North Side	3995.0	17.0	7.0	2.0	20.0	15.0	1.0	15.0	4.0	7.0	3.0	4.0
Edison Park	52.0	2.0	8.0	0.0	8.0	8.0	1.0	1.0	1.0	0.0	1.0	1.0
Forest Glen	265.0	27.0	2.0	1.0	6.0	6.0	0.0	2.0	3.0	1.0	5.0	1.0
Albany Park	1438.0	47.0	4.0	0.0	11.0	8.0	1.0	0.0	4.0	1.0	5.0	0.0
Portage Park	332.0	31.0	3.0	0.0	8.0	6.0	0.0	3.0	3.0	0.0	3.0	5.0
Avondale	1551.0	77.0	8.0	0.0	9.0	8.0	1.0	1.0	4.0	2.0	1.0	0.0
Logan Square	4471.0	93.0	4.0	0.0	5.0	5.0	0.0	2.0	1.0	2.0	1.0	2.0
West Town	4874.0	62.0	6.0	2.0	12.0	19.0	2.0	2.0	3.0	2.0	2.0	0.0
Near West Side	2490.0	37.0	7.0	3.0	10.0	8.0	0.0	4.0	1.0	3.0	3.0	4.0
Loop	3665.0	9.0	17.0	7.0	32.0	20.0	1.0	2.0	7.0	4.0	3.0	7.0
Near South Side	756.0	10.0	4.0	0.0	10.0	6.0	0.0	1.0	3.0	2.0	6.0	3.0
Hyde Park	150.0	15.0	2.0	0.0	13.0	9.0	0.0	1.0	4.0	6.0	5.0	4.0
Bridgeport	355.0	46.0	5.0	3.0	12.0	10.0	0.0	0.0	10.0	4.0	3.0	0.0
Edgewater	931.0	25.0	3.0	0.0	9.0	14.0	1.0	3.0	7.0	3.0	3.0	1.0

Figure 18: Cluster 1 neighbourhoods (high potential areas)

From the mix of venues and other business establishments from this cluster, this cluster seems to be appropriate for new ATM installations for any banks. Most of these venues promote cash transactions than card transactions, which would make this cluster favourite for ATMs

Cluster 2 Community Areas:

CommunityArea	sidewalkcafes	groceries	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
West Ridge	595.0	38.0	1.0	0.0	3.0	19.0	0.0	0.0	4.0	1.0	1.0	0.0
North Park	797.0	39.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
Irving Park	906.0	35.0	0.0	0.0	5.0	9.0	0.0	1.0	3.0	1.0	1.0	0.0
Belmont Cragin	361.0	58.0	1.0	0.0	2.0	6.0	0.0	0.0	4.0	1.0	1.0	0.0
Hermosa	849.0	60.0	2.0	0.0	2.0	7.0	0.0	0.0	1.0	0.0	2.0	0.0
Humboldt Park	2104.0	87.0	0.0	0.0	4.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
Austin	83.0	61.0	0.0	0.0	7.0	0.0	1.0	0.0	2.0	1.0	0.0	0.0
North Lawndale	164.0	52.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
South Lawndale	149.0	60.0	2.0	0.0	8.0	11.0	0.0	0.0	3.0	2.0	1.0	0.0
Douglas	243.0	34.0	2.0	0.0	12.0	4.0	1.0	2.0	3.0	1.0	5.0	0.0
Washington Park	171.0	42.0	1.0	1.0	3.0	0.0	0.0	2.0	2.0	1.0	3.0	0.0
Roseland	4.0	40.0	1.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	1.0	0.0
Garfield Ridge	38.0	40.0	1.0	0.0	6.0	2.0	0.0	3.0	3.0	1.0	1.0	0.0
Brighton Park	22.0	40.0	0.0	0.0	7.0	3.0	0.0	0.0	2.0	1.0	3.0	0.0
McKinley Park	152.0	58.0	2.0	0.0	3.0	3.0	0.0	0.0	5.0	1.0	1.0	1.0
West Elsdon	36.0	35.0	1.0	0.0	6.0	4.0	0.0	1.0	3.0	0.0	0.0	0.0
Gage Park	56.0	70.0	0.0	0.0	6.0	2.0	0.0	1.0	2.0	0.0	0.0	1.0
Englewood	3.0	39.0	1.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	2.0	0.0
Greater Grand Crossing	35.0	45.0	0.0	0.0	2.0	3.0	0.0	3.0	1.0	0.0	3.0	1.0
Ashburn	36.0	34.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0
Beverly	25.0	29.0	0.0	0.0	8.0	2.0	1.0	1.0	3.0	2.0	0.0	0.0
Washington Heights	24.0	39.0	1.0	0.0	5.0	2.0	2.0	1.0	1.0	0.0	0.0	3.0

Figure 19: Cluster 2 neighbourhoods (moderate potential areas)

From the mix of venues and other business establishments from this cluster, this cluster seems to be good for new ATM installations, but not having greater potential like cluster 1. Most of these venues promote both cash transactions and card transactions, which would make this cluster suitable for banks to run secondary ATMs or lease/tie-ups rather than running on its own

Cluster 3 Community Areas:

CommunityArea	sidewalkcafes	groceries	Entertainment Type1	Entertainment Type2	Food Type1	Food Type2	Lifestyle Type1	Lifestyle Type2	Retail Type1	Retail Type2	Utilities Type1	Utilities Type2
Norwood Park	161.0	17.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	3.0	1.0
Jefferson Park	94.0	13.0	1.0	0.0	4.0	4.0	0.0	0.0	0.0	1.0	0.0	1.0
Dunning	175.0	16.0	0.0	0.0	9.0	7.0	1.0	3.0	0.0	0.0	0.0	0.0
Montclare	175.0	16.0	0.0	0.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
West Garfield Park	0.0	8.0	1.0	0.0	8.0	1.0	1.0	0.0	1.0	8.0	1.0	0.0
East Garfield Park	17.0	12.0	0.0	1.0	1.0	1.0	0.0	0.0	3.0	0.0	0.0	2.0
Lower West Side	335.0	26.0	2.0	1.0	3.0	6.0	0.0	0.0	3.0	0.0	1.0	0.0
Armour Square	225.0	28.0	1.0	0.0	4.0	5.0	1.0	0.0	1.0	0.0	0.0	1.0
Oakland	18.0	6.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	7.0	0.0
Fuller Park	20.0	20.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	6.0	0.0
Grand Boulevard	157.0	30.0	4.0	6.0	4.0	2.0	0.0	3.0	1.0	0.0	2.0	1.0
Kenwood	137.0	10.0	2.0	1.0	6.0	3.0	0.0	3.0	1.0	1.0	5.0	0.0
Woodlawn	32.0	19.0	1.0	0.0	5.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0
South Shore	33.0	29.0	0.0	0.0	1.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0
Chatham	3.0	27.0	0.0	0.0	3.0	0.0	0.0	2.0	1.0	0.0	2.0	0.0
Avalon Park	14.0	24.0	0.0	0.0	6.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0
South Chicago	14.0	22.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0
Burnside	1.0	10.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	2.0	2.0
Calumet Heights	14.0	24.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
Pullman	2.0	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
South Deering	13.0	16.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
East Side	13.0	14.0	0.0	0.0	3.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0

Figure 20: Cluster 3 neighbourhoods (low potential areas)

From the mix of venues and other business establishments from this cluster, this cluster seems to be not good for new ATM installations. The number of venues in all categories itself is very less in these areas compared to cluster 1 and 2.

5. Conclusion

From this study, we would like to recommend the community areas included in **cluster 1** to be more potential areas for new ATM installations for any banks which looks for offsite ATM installations.

To see this geographically, most of the **north-eastern neighbourhoods** of Chicago is areas of interest for banks. Same way, most mid neighbourhoods have mediocre interest, which still has some potential. Likewise western and southern neighbourhoods least attracts bank interests with their lack of business venues.

6. Future directions

The aim of this study is to provide some meaningful insights of the community areas to be considered from its cash transaction potential which is one of the key factors to run ATMs in that neighbourhood. This study satisfies that requirement and identifies the favourable set of community areas having that potential.

Although this segmentation of community areas recommend a cluster of community areas for ATM installations, this should be considered as a study and additional research should be done on these community areas further with other parameters like foot traffic, competitions of other banks, existing ATM installations and so on.

This concludes this study of community areas of Chicago.