# Chicago Community Areas Segmented by Income and Population Size: Selecting an Optimal Location for Pop Up Food Pantries

# REPORT: Final Capstone Project Jan 2021

**Introduction: Problem and Background**

Chicago is the 3rd largest city in the United States with a population of 2,693,976.[1] For statistical and planning purposes, Chicago’s 200+ neighborhoods are divided into 77 distinct community areas.[2] These areas were first defined by researchers at the University of Chicago in the 1920s and still are used for urban planning initiatives and to compare equity of access to government and public services, in addition to informing commercial decisions (e.g., where to build new housing developments, entertainment, etc.) and other public and private sector analyses.

During the Covid19 Pandemic, many businesses in Chicago were forced to close by gubernatorial order. Unemployment in the state sharply increased leaving many people without stable income. Additionally, riots over police brutality erupted in the summer of 2020. These riots resulted in extensive damage to small business throughout the city. Among the businesses permanently closed were grocery stores and affordable restaurants. Many lower income urban dwellers rely on low cost options from local restaurants, food trucks and outdoor vendors for their meals. Furthermore, Chicago is in the top ten cities in the US in terms of restaurants per capita[7].

Many parts of Chicago are considered food deserts for their lack of affordable healthy grocery shopping options. Food deserts were already a problem before the riots and pandemic lockdowns and have only gotten worse.[3,4] Currently, the demand for food banks or food pantries is higher than ever as many people cannot afford healthy food and rely on volunteer organizations for assistance.[4] The increased demand for food bank services has also coincided with a decrease in food bank resources as less people have been able to donate money and food.

Because of this increased strain on food banks, it is more important than ever to make wise use of existing resources. Stakeholder Traveling food banks, philanthropist organizations and government agencies that aim to reach the maximum number of needy citizens need to go directly to the areas that need it most. The purpose of this project will be to identify the lowest income and most populous communities within the city where temporary food banks can be established to reach the maximum number of neediest citizens.

This analysis will address two primary research questions:

1. How are the 77 community areas of Chicago similar with respect to income and population and most popular venues (specifically grocery/restaurant)?
2. Where should food non-profit orgs establish temporary food banks to serve the most needy people in the city?

**Data Description**

To answer the above two research questions, I will use two primary data sources to cluster and segment Chicago communities. The first is community data released in 2020 from the Chicago Metropolitan Agency for Planning (CMAP).[5] This data set contains many variables related to community population, average income, access to transportation, and more. The second data service that will be used here is the Foursquare API to explore how these communities differ in access to grocers and restaurants. The data sets will be merged for clustering and segmentation to develop recommendations to identify the best places for temporary, or new, food banks.

Geocordinates representing the center of each of the 77 Chicago areas were obtained via the Google Maps Platform Geocoding API. Chicago area names were manually entered into the geocoder at <https://developers-dot-devsite-v2-prod.appspot.com/maps/documentation/utils/geocoder> to visually confirm area location within Chicago as some Chicago areas have generic names that can be found in other cities, towns and even suburban Chicago areas. For more information on the tool, see the overview at <https://developers.google.com/maps/documentation/geocoding/overview>. The resultant latitude and longitudinal coordinates were saved into a csv file along with the area name.

Chicago Community Area (CCA) data was downloaded from the CMPA Data Hub at <https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data>. The CMAP Data Hub is operated by the Chicago Metropolitan Agency Data Hub and according to their about statement is “the source for data and information relevant to comprehensive planning in the seven county metropolitan Chicago region. As the official regional planning organization for northeastern Illinois, CMAP prepares data, analyses, and evaluations on land use, transportation and environmental topics. These activities are critical to providing objective assessments of current and future regional conditions”

The CCA Profile CSV file contained 77 records (corresponding to each Chicago community areas) and 221 columns related to demographic data compiled mostly from the 2010 US Census and July 2020 American Community Data Snapshots. These fields include median income, median education, age, race and ethnicity, access to transportation, employment and more. For the purposes of this analysis, the three column Chicago area geocoordinate file was merged with the select columns from the CCA dataset. The final data set contained: area name, latitude, longitude, total population, and median income.

**Methodology**

In order to segment the 77 Chicago Community areas, a k-means clustering model was applied. This model is widely used in a number of real world operations and was used in this exercise to segment the 77 areas along income and population size dimensions. The data cleaning, modeling and subsequent mapping of the areas were conducted in Python, documented in Jupyter notebooks via the IBM Developer Skills Network environment.

After initial dependencies were installed, and the dataset loaded, the data were normalized using the StandardScaler() function from sklearn. Normalization is a statistical technique that helps algorithms interpret features with different magnitudes and distributions equally. A total of nine clusters were selected for inclusion in the model. The choice of nine is driven by the assumption that high, middle and low income designations will be viewed across highly populated, medium populated and low populated areas. This would result in a 3 x 3 tiered stratification across population and income. A total of 12 iterations were run and centroid values were checked. The resulting model seemed satisfactory. Upon completion of the test, all 77 Chicago community areas were assigned to a cluster.

To analyze popular grocer and restaurant venues in each area, I utilized the Foursquare API to explore each neighborhood. First, I defined my credentials and the version. Then I wrote a GET request to obtain the top 50 venues in Albany Park within a radius of 500 meters. After that GET request was sent and the results were examined, I borrowed the function that extracts the category of the venue. Then the json file was cleaned and structured into a pandas dataframe. Once that was completed, I checked the number of venues returned by Foursquare. As a final step to that first phase, I created a function to repeat this process for all neighborhoods in Chicago. Next, I checked to see how many venues were returned for each neighborhood. I ran code to check how many unique categories were curated from all returned venues and then I was ready to analyze the neighborhood venues.

Additional documentation is provided in my Jupyter notebook on github at: <https://github.com/reynajac/Chicago-Community-Food-Insecurity-Analysis/blob/main/Chicago_Community_Segmentation.ipynb>

**Results**

According to the k-means clustering, 3 of the clusters represented particularly low income and low to high population figures. It was initially believed that the neediest citizens would be in the most populated areas. An examination of the median income from each cluster shows that the lowest income households are in the least populated areas. Three clusters emerged as the home to the most lower income households. Those are Clusters 0, 5 and 7. These 3 clusters account for just under half (49%) of all Chicago communities.

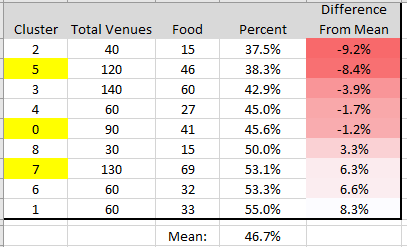
**Figure 1: Neighborhood Clustering**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Cluster*** | ***Total Population*** | ***Median  Income*** | ***Cluster Description*** |
| 0 | 43,314 | $32,933 | Auburn Gresham, Austin, Chatham, Chicago Lawn, Douglas, Grand Boulevard, New City, South Chicago, South Deering, South Lawndale, South Shore |
| 1 | 44,474 | $72,519 | Ashburn, Dunning, Garfield Ridge, Jefferson Park, Lincoln Square, Logan Square |
| 2 | 40,291 | $110,912 | Forest Glen, Lincoln Park, North Center, The Loop |
| 3 | 34,600 | $51,759 | Belmont Cragin, Bridgeport, Calumet Heights, Edgewater, Hegewisch, Hyde Park, Kenwood, North Park, O'Hare, Uptown, Washington Heights, West Elsdon, West Lawn, West Ridge |
| 4 | 41,657 | $99,167 | Beverly, Edison Park, Mount Greenwood, Near North Side, Near South Side, West Town |
| 5 | 16,944 | $24,771 | Armour Square, Burnside, East Garfield Park, Englewood, Fuller Park, Greater Grand Crossing, North Lawndale, Oakland, Riverdale, Washington Park, West Englewood, West Garfield Park, Woodlawn |
| 6 | 42,383 | $62,223 | Albany Park, Avondale, Clearing, Irving Park, Morgan Park, Portage Park |
| 7 | 28,844 | $42,660 | Archer Heights, Avalon Park, Brighton Park, East Side, Gage Park, Hermosa, Humboldt Park, Lower West Side, McKinley Park, Montclare, Pullman, Rogers Park, Roseland, West Pullman |
| 8 | 66,864 | $84,500 | Lake View, Near West Side, Norwood Park |

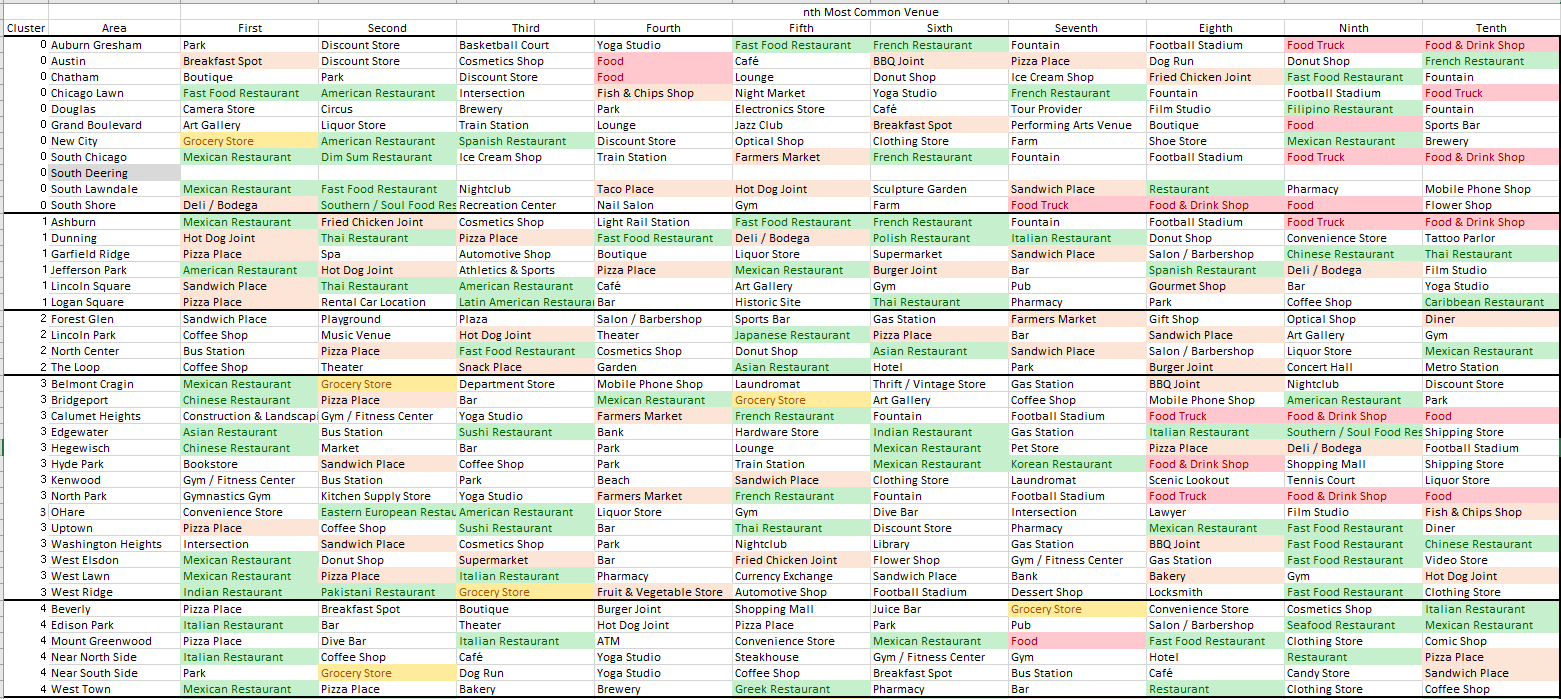
The Foursquare data yielded limited insights for the purposes of our stakeholders interested in locating an optimal location for a food bank. The ten most common venues for each neighborhood were collected and charted in a grid format (See Figure 2). No clear pattern emerged and there were some notable issues. First, 3 of Chicago neighborhoods-- Pullman, Riverdale and South Deering-- returned no results. These are among the most economically disadvantaged areas. Why this occurred requires further analysis beyond the scope of this report.

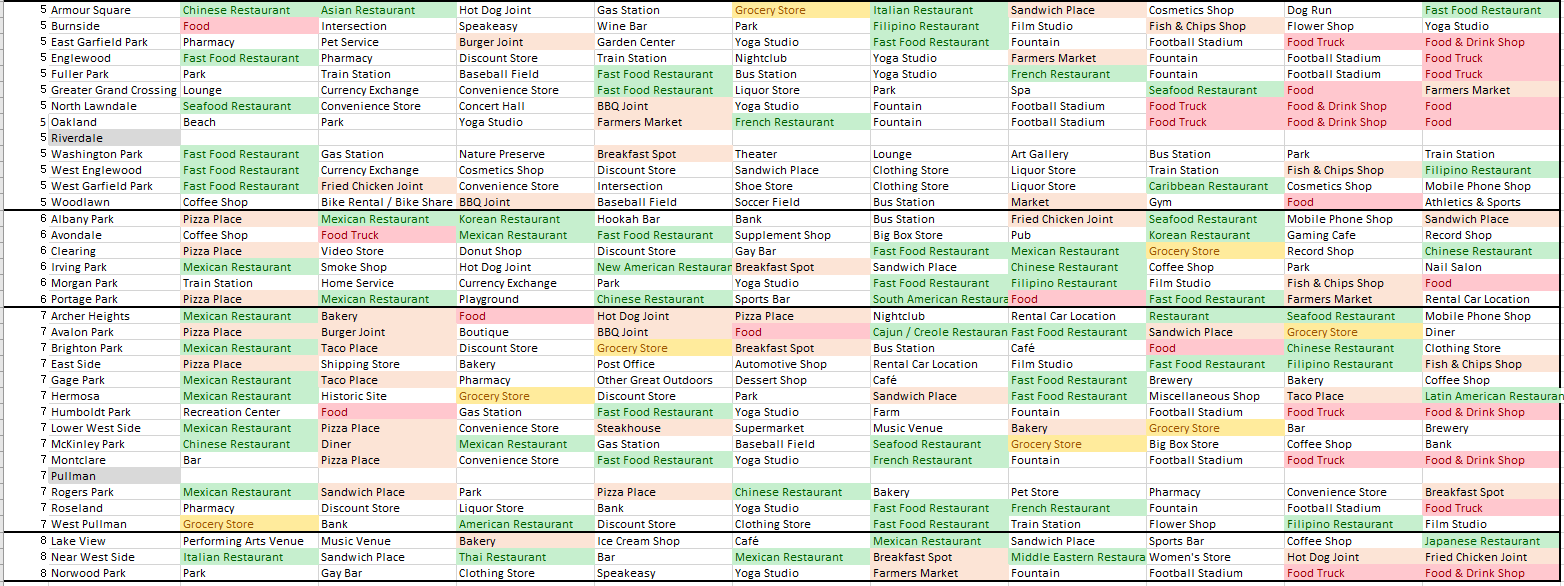
According to the frequency analysis performed on the venue returns, on average, 46.7% of all venues in each cluster were food related. These included restaurants, grocers/supermarkets and bakeries. See Figure 3 for additional detail.

**Figure 3: Frequency of Food Venues by Total Venues in Top Ten**



**Figure 2: First Through Tenth Most Common Venues in Each Chicago Neighborhood**





**Discussion and Conclusion**

The goal of this research was to determine the optimal place for a food bank in these difficult times where resources are scarce. It was initially assumed that the areas most economically disadvantaged and where the largest population is, would be ideal. However the cluster analysis indicated that the areas beset by the lowest median household income are also the least populated. This makes sense in retrospect as the most populated areas are the most sought after by home owners and renters so rents in these areas would likely be high. Therefore, lower income city dwellers tend to live in sparsely populated areas in the city. Stakeholders wishing to focus building new or temporary food pantries should focus on median income rather than population size of the area. Areas within clusters 0, 5 and 7 are the most ideal areas to target.

While income and population size might indicate areas where there are many disadvantaged households, other variables could also refine who is in the most need such as average household size and current employment data. Addition of these variables into a new model might reveal additional valuable insights.

A second assumption of this research was that lower income areas that are “food deserts”[8] would make good targets for organizations wishing to set up food banks or pantries. The issue we encountered in the course of this analysis when using Foursquare data is that a multitude of restaurants and food-related establishments are scattered about the city. One interesting finding is that the highest income cluster (Cluster 2, median income of $110K) had the least amount of food related venues listed among the top 10. This might be attributable to high rent and property taxes in those areas which might preclude all but the most profitable establishments from settling in those areas. The presence of food/restaurant establishments in an area may be a poor indicator of household need. An additional issue is that the quality and price of the food available through these restaurants, markets and delis is unknown. The quality of the products offered there is beyond the scope of this assessment. Because the nature of the restaurant and grocery venues varies widely, stakeholders should make their decisions on opening new food pantries based on the income and population clustering noted above and disregard the frequency of Foursquare venue type. This is not to say that this Foursquare data is not relevant, but simply more analysis may be required.

Additionally, when determining the optimal locations for a pop-up food pantry, volunteer organizations may need to weigh more practical considerations in their decision making, including:

* Access to public transportation for the target groups
* Availability of car parking for individuals traveling from other parts of the city
* Flexibility and cooperation of local alderman to select and protect sites
* Ease of arranging for street closures should sites be placed outdoors
* Cost of erecting distribution facilities for outdoor venues

**Footnotes**

1. US Census quick facts page: <https://www.census.gov/quickfacts/chicagocityillinois>
2. <https://en.wikipedia.org/wiki/Community_areas_in_Chicago>
3. <https://www.chicagoreporter.com/food-deserts-persist-in-chicago-despite-more-supermarkets/>
4. <https://www.chicagotribune.com/coronavirus/ct-life-coronavirus-food-insecurity-tt-20200403-20200403-ytanbm6j75e2fhjitctqhgljay-story.html>
5. <https://chicago.eater.com/2021/1/20/22231602/chicago-food-deserts-fresh-food-healthy-hood-we-go-us>
6. <https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data>

1. <https://thefoodoasis.com/the-number-of-people-per-restaurant-in-26-major-cities/>
2. <https://foodispower.org/access-health/food-deserts/>