Naasir Haleem

11/29/20

Battle of The Neighborhoods Capstone

**Introduction:**

Chicago is one of the most segregated cities in the United States. Neighborhoods have strong ethnic and economic divisions. This is also reflected in the different level of access that neighborhoods have to grocery stores, banks, and other structures in communities. The starkest divisions are between neighborhoods in the North side of the city and those in the South side. By analyzing a neighborhood map of Chicago, I would like to segment the different parts of the city. I believe that this analysis would be the most useful to both government and non-government organizations that are trying to determine areas of the city to assist and how similar or dissimilar their approaches should be this assistance should be.

**Data that will be used:**

I'm going to read in the household income of neighborhoods throughout the city. This data can be found in a dataset from this link:

Link: <https://data.cityofchicago.org/Health-Human-Services/Per-Capita-Income/r6ad-wvtk>

Description of the data: This data was collected via census between 2008-2012. There are 6 categories that I will be using from this dataset. I decided not to use Percent of Housing Crowded, which is defined as percent of households that have more than 1 person per room, since I wanted to focus more on universal measures of community health.

1. Percent households below poverty (converted to Poverty Rate)

2. Percent aged 16+ unemployed (converted to Unofficial Unemployment Rate)

3. Percent aged 25+ without high school diploma (converted to Non-GED)

4. Percent aged under 18 or over 65 (converted to Percent Children or Retirement Age)

5. Per Capita Income (left as is)

6. Hardship Index (left as is)

Note: I want to describe two of these variables since they are not intuitive.

Per Capita Income is defined as the aggregate of tract level income divided by the total population. Thus, it can be pulled down if there are large families with children.

Hardship Index is the score that incorporates each of the six (I only use 5) socioeconomic indicators.

I'm going to read in the demographic makeup of the neighborhoods in Chicago. This data was collected through the 2010 Census so, while it may not be completely accurate today, it should still be very relevant to the task at hand.

Linked: <https://datahub.cmap.illinois.gov/dataset/2010-census-data-summarized-to-chicago-community-areas>

Description of the Data: The data here gives the percent makeup of the demographics of each neighborhood.

I'm going to first make a dataframe of the neighborhoods in Chicago. I will do this by reading in a csv of Chicago neighborhoods. I had to this indirectly by running this individuals code that gave a .txt file with locations throughout the neighborhoods. I then averaged the longitudes and latitudes of the neighborhood locations to give a central point in each neighborhood. The link can be found below:

Link: <https://github.com/jkgiesler/parse-chicago-neighborhoods>

Description of the Data: The data here gave a list of the neighborhoods in Chicago with several longitude and latitude points. As I said, I averaged out those points to give a central location within the neighborhood.

Lastly, I will use the Foursquare API to determine the venues that are most present in each of these neighborhoods.

**Methodology:**

I chose to use KMeans Clusters in order to segment the neighborhoods of Chicago because the process will help me to spot any substantial differences between neighborhoods. This process works by creating centroid points that are determined by the data that most closely surrounds that point. Since I am using 5 different socioeconomic factors, along with longitude and latitude, the centroid points will help to bin the different neighborhoods. I have chosen to use 7 of these points.

# Using FourSquare API for Clustering

Explanation: While looking at demographics and socioeconomic factors did help to segment the Chicago neighborhoods; I also want to understand if the venues present in each neighborhood differ from one another. Thus, I am going to merge this venue data with the dataframe that is in use and see if the clusters change substantially. This should help me to determine how disadvantaged communities are affected in relation to the venues that are most present within them. For a policy maker, or a non-profit organization, this could help determine what resources should be leveraged in these communities.

# Analysis Based on Socioeconomic/Demographic Information

**Analysis:** I can see from this map that for the most part there are clearly identifiable clusters of neighborhoods in Chicago. These are the following observations I made after briefly analyzing the map.

1: Cluster 5 is the most concentrated cluster. It is located almost entirely in the South-Western part of Chicago and, besides two neighborhoods, is located entirely south of the Eisenhower Expressway.

2: Cluster 1 and Cluster 0 both represent the greatest number of neighborhoods. They are both somewhat evenly distributed throughout the city, although Cluster 0 has a distinct northern cluster, and Cluster 1 has a defined southern cluster. The per capita income of Cluster 0 does appear to be greater than Cluster 1.

3: Cluster 4 and Cluster 2 could almost be lumped in together. They are located in the same relative part of the city and have a significantly above normal per capita income.

# For future reference:

Cluster 0 = Red

Cluster 1 = Purple

Cluster 2 = Blue

Cluster 3 = Light Blue

Cluster 4 = Pale Green

Cluster 5 = Olive Green

Cluster 6 = Orange

# Cluster 0 Analysis

**Analysis:**On a high level it is clear to see that Cluster 0 (represented by the red) is dispersed widely throughout the city. The Per Capita Income, which again is the aggregate income of the neighborhood divided by the total population, is around $25000. The household poverty rate is around 16%. In totality I would peg Cluster 0 as being a Middle-Class neighborhood.

**Additional Observations**

1. Cluster 0 is certainly extremely diverse. The percent of Blacks, Whites, and Latinos is all between 24% and 38%. Asians make up the final ~8% of the population. I think that this is supported by observing the sorted venues in the neighborhoods. Many different ethnic specific restaurants are represented.
2. It is important to note that the percentage of Children or Retirement Age Adults is fairly high. More than a third of the population falls into this category. This probably does mean that the average family here is young with children.

**Cluster Definition**

Cluster 0 can be defined as being Diverse, Middle Class, and Young

# Cluster 1 Analysis

**Analysis:** There are similarities between Cluster 1 and Cluster 0. Both clusters have Per Capita Income that is within $8000 of each other. The Non-GED numbers and Unemployment Rate also seem to be in line with one another. Additionally, both clusters can be found throughout the city and aren't necessarily geographically constrained. However, it is clear that by several metrics, Cluster 1 has more clear instances of greater community hardship than Cluster 0. For this reason, I would label Cluster 1 as being Lower/Middle Class.

**Additional Observations:**

1. Cluster 1 is predominately Black, with Hispanics or Latinos making up the other significant ethnicity in the neighborhood.
2. Similarly to Cluster 0, the Percent of Children or Retired Adults is fairly high as compared to other neighborhoods.

**Cluster Definition:**

I would define Cluster 1 as being Black/Hispanic/Latino, Lower/Middle Class, and Young

# Cluster 2 Analysis

**Analysis:**It is immediately apparent that Cluster 2 is substantially different than Cluster 0 and Cluster 1. The Per Capita Income is around $62000, and the Unemployment Rate is extremely low at around 5 percent. There is also significant geographic restraint in play, as neighborhoods that fall into Cluster 2 are close to the lake and are for the most part in the northern side of the city. Cluster 2 can be characterized as Upper Class.

**Additional Analysis:**

1. The demographics of Cluster 2 swing heavily toward Whites, who make up 72 percent of the Cluster. Blacks, Asians, Hispanics or Latinos, encompass the other 28 percent at a relatively similar rate. This is backed up by a noted lack of ethnic restaurants making up the venues.
2. Cluster 2 has the highest rate of working age adults at almost 80 percent of the population. These are also educated individuals where 95 percent have at least a GED.

**Cluster Definition:**

I would define Cluster 2 as North Side, Professional, and White

# Cluster 3 Analysis

**Analysis:**Cluster 3 is primarily grouped on the North side with 2 exceptions. Kenwood and Mount Greenwood are not located near the other neighborhoods in the cluster but are similar in all other ways. While Per Capita Income of Cluster 3 is not as high as Cluster 2, it is firmly within the Middle-Class tier.

**Additional Analysis:**

1. Cluster 3 is not as diverse as Cluster 0 yet shares some similarity in terms of demographics. While 55 percent of the Cluster is White, another 34 percent consists of Blacks and Hispanics or Latinos.
2. Cluster 3 is fairly educated, with 89 percent of the population holding a GED, and also seems to skew towards young Professionals.

**Cluster Definition:**

I would define Cluster 3 as North Side, Middle Class, and Diverse

# Cluster 4 Analysis

**Analysis:**Cluster 4 consists of a singular neighborhood and that is Near North Side. It is heavily White, and it is wealthy with a Per Capita Income of around $89000. It can immediately be classified as Upper Class.

**Additional Analysis:**

1. Interestingly, there seems to be a similar percent of working age adults that reside in Near North Side as Cluster 2. This suggests that most individuals are professionals.
2. The 2nd Most Common venue in Near North Side are Hotels. This suggests that this is an area of the city well trafficked by tourists.

**Cluster Definition:**

I would define Cluster 4 as Upper Class, Professional, and Highly Educated

# Cluster 5 Analysis

**Analysis:**Cluster 5 is probably the most interesting of any of these clusters. Cluster 5 is at a significant lack of resources. The Per Capita Income is $6000 less than the next lowest Per Capita Income of any neighborhood. The cluster is also extremely geographically constrained. In fact, it seems that there are pockets of the city where these neighborhoods lie next to one another. It would be to the interest of the city to direct substantial resources to neighborhoods within this cluster. Cluster 5 are most definitely in Critical Need.

**Additional Analysis:**

1. Blacks largely comprise this cluster, as do Hispanics or Latinos. Additionally, it is concerning that Fast Food Restaurants are so present within these communities.
2. Not only is the Unemployment Rate extremely high, the Percent of Working Age Adults is relatively low. This probably does indicate that even if there are many children within this Cluster, the families are plagued by unemployment.

**Cluster Definition:**

I would define this cluster as Black, In Critical Need, and Underserved.

**Cluster 6 Analysis**

**Analysis:**There does not appear to be any geographic constraint on Cluster 6. The Per Capita Income is Upper Middle Class.

**Additional Analysis:**

It appears that families are very present within this Cluster. The most common venues seem to bear this out, along with the relatively high Percent of Children or Retired individuals.

**Cluster Definition:**

I would define this Cluster as being White, Family Based, and Upper Middle Class

# Conclusion

I believe that it is clear to see from analysis of these different clusters that Chicago is indeed highly segregated, which can particularly be seen in the difference between North Chicago and South Chicago. Specific parts of North Chicago are White, Upper Class, and Professional. This does not necessarily remain true for South Chicago. It should also be noted that the neighborhoods of mainly Black and Hispanic or Latino individuals are significantly more underserved than their counterparts. This is true in both terms of the unemployment rate within these neighborhoods, as well as the venues that are present. I think with further analysis of these neighborhoods public officials will be able to more effectively serve the communities under their jurisdiction.