CHICAGO JOB RELOCATION

Python Capstone Project

1. Introduction

In a recent bit of good news, a friend was promoted within her company as a director. The exciting new position requires her to relocate to Chicago, IL within the next few months. She is ecstatic at the opportunity since she has worked hard to move up in the company but also fortunate to be near family. As a single mother of a 4-year-old, her priority is to find a neighborhood that would be suitable for her and her daughter. She aims to find a great elementary school as she will start kindergarten within the next couple of years along with activities and events that they both can enjoy.

My friend is an outgoing individual that loves to dance and enjoys going out to cafes and bistros in her current home in Long Beach, CA. She is a very social person and a strong independent woman. Her daughter is a tiny little lady that may be shy when meeting new people but is a really fun and playful girl once she gets comfortable.

The goal of this project is to look for a neighborhood that would be a good fit for my friend and her daughter. We will break down the neighborhoods that make up the city and find the top schools that are available. Once we find those schools and where they are located, we will focus on the venues that may fit their lifestyle.

2. Data

2.1. Sources

The data that will be needed to answer her question will require us to map out the neighborhoods that make up the city. Once that has been established, we will find information on the top schools and the resources available in those neighborhoods. After scouring the web for information we have found the following to solve our question.

a. **Chicago Neighborhoods:** The JSON file will create polygons of each neighborhood on top of the city map. (https://github.com/jkgiesler/parse-chicago-neighborhoods)
b. **School Ranking:** A list of the top 50 schools in Chicago. The rankings are based on criteria such as Standard Test Score, Student-to-Teacher Ratio, etc. (https://www.schooldigger.com/go/IL/schoolrank.aspx)
c. **Resources:** An API provided by Foursquare will allow us to pull information for the various options that is made available for families such as parks, youth centers, etc. (https://foursquare.com/)

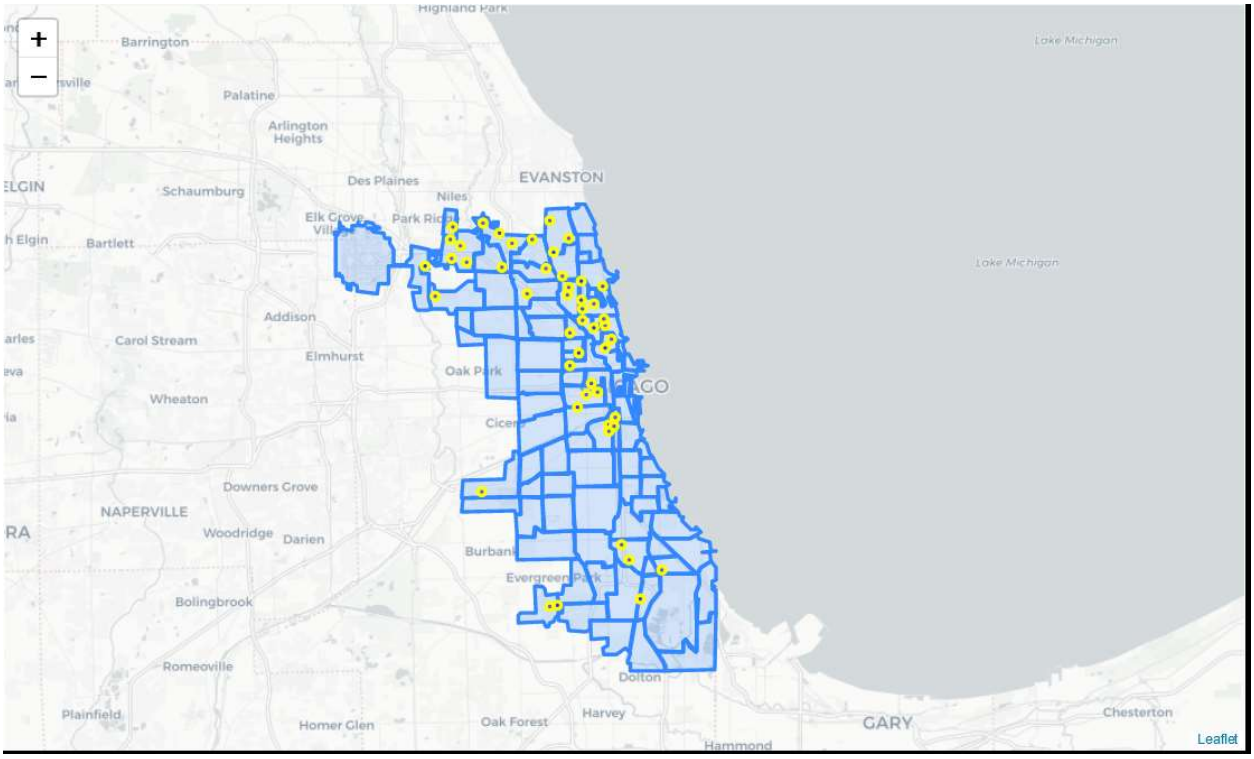
2.2 Data Manipulation

The first step in the project is to find out how the City of Chicago is broken down. Thanks to Jason Geisler, he was able to provide a GeoJSON file that created polygons of each neighborhood. Looking at

the file, it was found that he had broken down the city by Primary Neighborhoods and Secondary Neighborhoods. We focused on the primary neighborhoods in this instance. The polygons were layered out on top of the city and mapped out.

Now that we have segmented the neighborhoods, we would like to know what are the top elementary schools in the city and where they are located. After searching the web, I found the http://www.schooldigger.com that ranked schools in each state based on the Average Standard Score. I selected the State of Illinois and was able to filter by the Chicago Public School District. The goal for this part of the data selection was to use Web scraping but due to the website being an Active Server Page and its complexity, I was not able to use the BeautifulSoup package. However, I was fortunate to find that the website did allow for the download of the data into a CSV file. The file contained the top 50 elementary schools in the district, which was read into a dataframe. The data contained the physical addresses.

To get information on where these schools were located, Jason had also provided a module that converted the addresses into GPS coordinates. I ran each school into it and created new columns creating this new set of information. The module was also able to get the neighborhood that each school was located. The data was mapped out get a visual idea of schools' position.



We see that most of the top schools are found North of Downtown Chicago. The data will have us only focus on these neighborhoods. The dataframe was previously updated with where the schools are and, so, we summarize the data into a table listing those neighborhoods and the number of schools within them.

	Neighborhood
Lake View	4
Norwood Park	4
Lincoln Park	3
Sauganash,Forest Glen	3
West Ridge	3
Little Italy, UIC	2
Bridgeport	2
Chatham	2
Mount Greenwood	2
North Center	2
Uptown	1
River North	1
West Town	1
North Park	1
Armour Square	1
Chinatown	1

	Neighborhood
Calumet Heights	1
Edison Park	1
Irving Park	1
Dunning	1
Old Town	1
West Loop	1
East Village	1
Garfield Ridge	1
Jefferson Park	1
Sheffield & DePaul	1
Pullman	1
Lincoln Square	1
Albany Park	1
Lower West Side	1
Bucktown	1

```
In [44]: neigh_school_count.shape
Out[44]: (31, 1)
```

In order to move to the next step in the project, we would like to find the most popular venues within 500 feet of the neighborhood. However, the neighborhood information that we have only contains the shape of it. So, we will take the coordinates of the school in a particular neighborhood and, if multiple, will take the centroid of those schools. The points will be used to load into the Foursquare API and pull the venues near them.

3. Data Analysis

3.1 Methodology

The Foursquare data will now provide all of the available venues for each of the neighborhoods that we have filtered. The API will be set to search within 500 feet of the schools for any type of venue.

We plug in 1 neighborhood, in this case Albany Park, by using the GPS coordinates and radius into a URL. The requests module pulls at most 100 venues along with the category they belong to and their coordinates. The following table displays the first 5 venues for the neighborhood.

	name	categories	lat	lng
0	Jaafer Sweets	Bakery	41.969691	-87.708085
1	Great Sea Chinese Restaurant	Chinese Restaurant	41.968496	-87.710678
2	Brazilian Bowl	Brazilian Restaurant	41.968537	-87.708558
3	Lindo Michoacan	Grocery Store	41.968864	-87.708453
4	la Michoacana Premium	Ice Cream Shop	41.968559	-87.706510

Now that we get an idea of how the data is structured, a module is constructed to perform the same steps that was done for Albany Park. The module will find the venues for each neighborhood and get a summary of how many of the venues were pulled.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Albany Park	35	35	35	35	35	35
Armour Square	12	12	12	12	12	12
Bridgeport	13	13	13	13	13	13
Bucktown	58	58	58	58	58	58
Calumet Heights	7	7	7	7	7	7
Chatham	9	9	9	9	9	9
Chinatown	44	44	44	44	44	44
Dunning	4	4	4	4	4	4
East Village	72	72	72	72	72	72
Edison Park	12	12	12	12	12	12
Garfield Ridge	12	12	12	12	12	12
Irving Park	22	22	22	22	22	22
Jefferson Park	8	8	8	8	8	8
Lake View	85	85	85	85	85	85
Lincoln Park	42	42	42	42	42	42
Lincoln Square	73	73	73	73	73	73
Little Italy, UIC	55	55	55	55	55	55
Lower West Side	22	22	22	22	22	22
Mount Greenwood	4	4	4	4	4	4
North Center	16	16	16	16	16	16
North Park	10	10	10	10	10	10
Norwood Park	9	9	9	9	9	9
Old Town	69	69	69	69	69	69
Pullman	1	1	1	1	1	1
River North	52	52	52	52	52	52
Sauganash,Forest Glen	21	21	21	21	21	21
Sheffield & DePaul	33	33	33	33	33	33
Uptown	33	33	33	33	33	33
West Loop	35	35	35	35	35	35
West Ridge	43	43	43	43	43	43
West Town	27	27	27	27	27	27

We now want to categorize the venues and get the frequency of the type for each neighborhood. The data is laid out in a table with each category listed in its own column and showing the frequency of the type of venue for each neighborhood. The table below shows the first 5 neighborhoods and the frequency of the type of venue.

	Neighborhood	Adult Boutique	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	...	Video Game Store	Video Store	Vietnamese Restaurant	Weight Loss Center	Whisky Bar	W
0	Albany Park	0.0	0.000000	0.0	0.000000	0.0	0.0	0.028571	0.0	0.0	...	0.0	0.000000	0.028571	0.0	0.0	
1	Armour Square	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.000000	0.000000	0.0	0.0	
2	Bridgeport	0.0	0.076923	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.076923	0.000000	0.0	0.0	
3	Bucktown	0.0	0.000000	0.0	0.017241	0.0	0.0	0.017241	0.0	0.0	...	0.0	0.017241	0.000000	0.0	0.0	
4	Calumet Heights	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.000000	0.000000	0.0	0.0	

5 rows × 213 columns

Once we display the frequency of each neighborhood, we sort the venues for each neighborhood from most to least.

```
----Albany Park----
                venue  freq
0                Park  0.06
1  Mexican Restaurant  0.06
2          Donut Shop  0.06
3      Ice Cream Shop  0.06
4       Grocery Store  0.06


----Armour Square----
                venue  freq
0  Chinese Restaurant  0.25
1       Grocery Store  0.17
2    Business Service  0.08
3  Mexican Restaurant  0.08
4    Storage Facility  0.08
```

A module is created to now list the most common venues for each neighborhood.

	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	Albany Park	Park	Mexican Restaurant	Donut Shop	Ice Cream Shop	Grocery Store	Persian Restaurant	Supermarket	Coffee Shop	Cocktail Bar	Soccer Field
1	Armour Square	Chinese Restaurant	Grocery Store	Business Service	Mexican Restaurant	Storage Facility	Pizza Place	Mobile Phone Shop	Bakery	Cosmetics Shop	Poke Place
2	Bridgeport	Pizza Place	Chinese Restaurant	Bakery	Café	American Restaurant	Italian Restaurant	Dessert Shop	Video Store	Mexican Restaurant	Tanning Salon
3	Bucktown	Bar	Pizza Place	Coffee Shop	Smoke Shop	Hot Dog Joint	Park	Dive Bar	Convenience Store	Thai Restaurant	Tea Room
4	Calumet Heights	Rental Car Location	Liquor Store	Deli / Bodega	Supplement Shop	Wings Joint	Women's Store	Ramen Restaurant	Pet Service	Other Repair Shop	Outdoor Supply Store

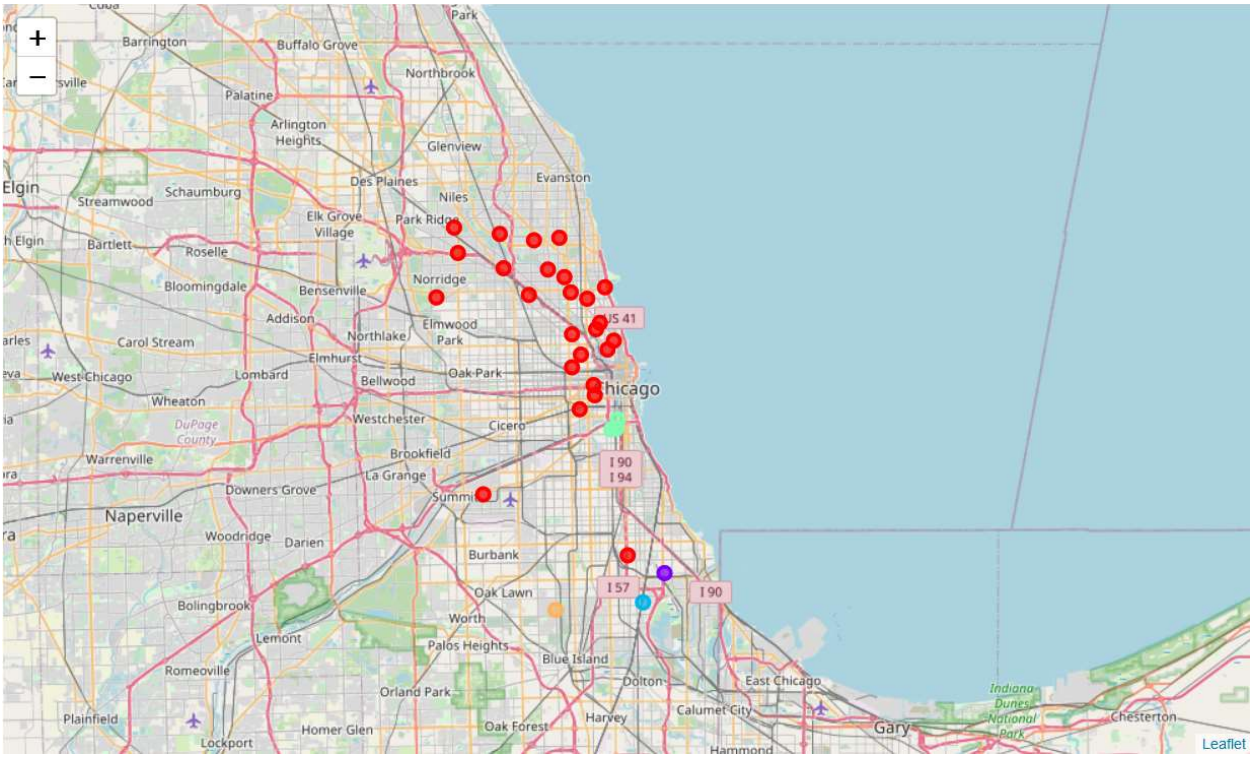
With the information now set, we will start with the cluster analysis to group the neighborhoods.

3.2 Cluster Analysis

In order to segment the neighborhoods based on the types of venues made available, we will perform a cluster analysis. We will set the number of clusters to 5 and run the analysis through the machine learning tool. The machine assigns each neighborhood a Cluster label based on the more common venues.

	Neighborhood	long	lat	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
0	Albany Park	-87.709385	41.971108	0	Grocery Store	Donut Shop	Ice Cream Shop	Mexican Restaurant	Coffee Shop	Sandwich Place	Thrift / Vintage Store	Hot Dog Joint	Cocktail Bar	Supermarket
1	Armour Square	-87.635136	41.843465	3	Chinese Restaurant	Grocery Store	Storage Facility	Asian Restaurant	Pizza Place	Mexican Restaurant	Mobile Phone Shop	Adult Boutique	Performing Arts Venue	Other Repair Shop
2	Bridgeport	-87.641031	41.841721	3	Pizza Place	Chinese Restaurant	Video Store	American Restaurant	Italian Restaurant	Café	Dessert Shop	Bakery	Tanning Salon	Ice Cream Shop
3	Bucktown	-87.683775	41.918823	0	Bar	Coffee Shop	Pizza Place	Hot Dog Joint	Dive Bar	Thai Restaurant	Park	Pool	Dance Studio	Ice Cream Shop
4	Calumet Heights	-87.583301	41.726501	1	Rental Car Location	Liquor Store	Women's Store	Wings Joint	Supplement Shop	Museum	Outdoors & Recreation	Park	Paper / Office Supplies Store	Pakistani Restaurant

We plot the neighborhoods on the map of Chicago and assign each cluster a unique color.



The majority of the neighborhoods fall under one cluster which is found mostly North of Downtown Chicago. The other clusters have at most 2 neighborhoods that are categorized. We can look at the data itself to see why they were broken up in such a way.

	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	Albany Park	Grocery Store	Donut Shop	Ice Cream Shop	Mexican Restaurant	Coffee Shop	Sandwich Place	Thrift / Vintage Store	Hot Dog Joint	Cocktail Bar	Supermarket
3	Bucktown	Bar	Coffee Shop	Pizza Place	Hot Dog Joint	Dive Bar	Thai Restaurant	Park	Pool	Dance Studio	Ice Cream Shop
5	Chatham	Gas Station	Chinese Restaurant	Sandwich Place	Gym / Fitness Center	Men's Store	Harbor / Marina	Fast Food Restaurant	Sporting Goods Shop	Café	Adult Boutique
7	Dunning	Laundry Service	Sports Bar	Playground	Park	Fast Food Restaurant	Convenience Store	Adult Boutique	Persian Restaurant	Other Repair Shop	Outdoor Supply Store
8	East Village	Bar	Café	Yoga Studio	Pizza Place	Sandwich Place	Sushi Restaurant	New American Restaurant	Mediterranean Restaurant	Italian Restaurant	Spa
9	Edison Park	Theater	Pizza Place	Salon / Barbershop	Seafood Restaurant	Hot Dog Joint	Bar	Mexican Restaurant	Breakfast Spot	Liquor Store	Italian Restaurant
10	Garfield Ridge	Bakery	Eastern European Restaurant	Gym / Fitness Center	Park	Construction & Landscaping	Pizza Place	Mexican Restaurant	Bar	Bank	Other Repair Shop
11	Irving Park	Bus Station	Coffee Shop	Bar	Train Station	Pharmacy	Brewery	Pizza Place	Tapas Restaurant	Gas Station	Bookstore
12	Jefferson Park	Train Station	Convenience Store	Donut Shop	Chinese Restaurant	Furniture / Home Store	Filipino Restaurant	Video Store	Ice Cream Shop	Concert Hall	Outdoors & Recreation
13	Lake View	Women's Store	Mexican Restaurant	Café	Bar	Cosmetics Shop	Coffee Shop	Clothing Store	Optical Shop	Boutique	Bus Station

In looking at the first cluster, we see more variety of venues. We see grocery stores, restaurants and cafes in this table. Of the 49 neighborhoods, 43 fall under this category. Let's take a look at the next cluster.

	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
4	Calumet Heights	Rental Car Location	Liquor Store	Women's Store	Wings Joint	Supplement Shop	Museum	Outdoors & Recreation	Park	Paper / Office Supplies Store	Pakistani Restaurant

The second cluster does show a neighborhood not very suitable for my friend and her daughter. The top venues are rental cars and liquor stores. Let's move to the next.

	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
23	Pullman	Cosmetics Shop	History Museum	Adult Boutique	Pet Service	Other Great Outdoors	Other Repair Shop	Outdoor Supply Store	Outdoors & Recreation	Pakistani Restaurant	Paper / Office Supplies Store

Again, this cluster does not seem suitable. It has more venues more suitable to adults. Not a good neighborhood for a child.

	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
1	Armour Square	Chinese Restaurant	Grocery Store	Storage Facility	Asian Restaurant	Pizza Place	Mexican Restaurant	Mobile Phone Shop	Adult Boutique	Performing Arts Venue	Other Repair Shop
2	Bridgeport	Pizza Place	Chinese Restaurant	Video Store	American Restaurant	Italian Restaurant	Café	Dessert Shop	Bakery	Tanning Salon	Ice Cream Shop
6	Chinatown	Chinese Restaurant	Asian Restaurant	Dim Sum Restaurant	Korean Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Tea Room	Taiwanese Restaurant	Pizza Place

The fourth cluster shows neighborhoods that are heavily focused on restaurants. They all happen to be South of Downtown Chicago.

	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
18	Mount Greenwood	Diner	Park	Fast Food Restaurant	Home Service	Pet Service	Other Great Outdoors	Other Repair Shop	Outdoor Supply Store	Outdoors & Recreation	Pakistani Restaurant

Finally, the last cluster seems to be very unique than the other neighborhoods. The venues are more focused on services. It is shown to not have regular restaurants until its tenth top venue and scarce on grocery stores.

4. Results and Conclusion

4.1 Summary

After reviewing all the neighborhoods in each of the clusters we have come to the conclusion on which ones would be favorable to my friend and daughter.

- Those neighborhoods found on the north side of Chicago are where she should focus on searching a place to live. Most of the top 50 schools in Chicago will be found in this area.
- The venues that may be of interest for my friend are found here. Most of the venues we found in the first cluster have a wide variety of services such as restaurants, grocery stores, and activities for her daughter to enjoy.
- The other clusters included venues that were not appropriate for kids to enjoy. There were rarely parks and businesses more dedicated to adults.

4.2 Discussion

Due to the large number of neighborhoods in the cluster that we selected; we could really dig into focusing our data finding venues catering children. We could filter the category in our Foursquare search to look for places such as parks, playgrounds, gyms and dance studios. Also, we can look at the specific restaurants that they both can enjoy in a more detailed analysis.

4.3 Conclusion

The project was very insightful in helping to find the neighborhoods that may be suitable for my friend and daughter. Cluster analysis and python is a very powerful tool that easily segmented the neighborhoods for us.

In addition to cluster analysis, python was very powerful in organizing the data. With so may options available to us, we were to use data from various sources to help with our analysis. It also helped that other people's modules assisted in the data wrangling as that was able to map out the City of Chicago and point out where the schools are located.