Analysis of business opportunities for new childcare facilities in the city of Chicago, Illinois

ELINA VITOL

IBM DATA SCIENCE PROGRAM

CAPSTONE PROJECT, JULY 2020

Presentation outline

Introduction/Business problem

Data Sources/References

Methodology and Results

◦ PART 1. EXTRACTING DATA ABOUT DAYCARES IN CHICAGO USING FOURSQUARE

◦ PART 2. INTRODUCE THE DAYCARE DATA PROVIDED BY THE CITY OF CHICAGO

◦ PART 3. MERGE THE DAYCARE DATA FROM THE FOURSQUARE AND THE CITY OF CHICAGO. ANALYSIS OF THE MERGED DATASET IN COMBINATION WITH THE SOCIOECONOMIC DATA ON CHICAGO NEIGHBORHOODS.

Conclusions and future direction

Introduction/ Business problem

- Opening a new daycare business requires a thorough analysis of the existing competition as well as the analysis of socioeconomic situation in the given neighborhood.

- The city of Chicago, being the largest city in the Midwest United States, with the estimated population of 2.7 million, serves as a home to many different companies within the city limits. As a result, there is a large number of working parents looking for childcare arrangements in the Chicago area. According to the census data, there are approximately 175,000 children under the age of 5 in Chicago.

- The goal of this project is to analyze which neighborhoods in the city of Chicago would be attractive for starting a new daycare business, depending on the interests of new busines owners.

- The target audience for this project is people looking to start a new daycare business in Chicago.

Data sources

- Foursquare API will be used for analyzing the neighborhood data in Chicago. There is a specific venue category for daycares. Specifying the venue type will narrow down the results.

- We will explore the number of daycare facilities in the city of Chicago per neighborhood as well as the type of the facility. In addition to the Foursquare data, we will analyze the data provided by the city of Chicago about the existing early education programs.

- Information about specific neighborhoods in the city, including neighborhood names and corresponding zipcodes is provided by the city of Chicago. This data will be used for creating a map of existing childcare facilities and linking it to the socioeconomic data. The latter will be used as an indicator for the daycare price point opportunities in a given area, i.e. low income vs high income neighborhoods.

- The socioeconomic information, including per capita income per neighborhood is available from the Chicago census data.

References

1. https://kiddieacademy.com/franchising/

2. https://www.goddardschoolfranchise.com/franchise-cost.html

3. https://smallbiztrends.com/2017/11/child-care-franchise.html

4. https://www.census.gov/quickfacts/fact/table/chicagocityillinois,US/PST045219

5. https://developer.foursquare.com/

6. https://data.cityofchicago.org/Education/Chicago-Early-Learning-Programs-Map/2kih-a5ex

7. https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Chicago-Zip-Code-and-Neighborhood-Map/mapn-ahfc

8. https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

Venue category had to specified to find daycares on Foursquare. 39 results were returned, 3 had to be excluded since they were erroneously categorized as daycares. Excerpt from resulting dataframe is shown below.

	COMMUNITY AREA NAME	Neighborhood Latitude	Neighborhood Longitude	Site Name	Latitude	Longitude	Venue Category
0	Printers Row	41.870981	-87.629035	It Takes A Village	41.872199	-87.633867	School
1	Sheffield & DePaul	41.927188	-87.653670	Tiny Tots Incorporated	41.929380	-87.648550	Daycare
2	Hermosa	41.924348	-87.734740	Kidslife Daycare Center	41.922340	-87.737204	School
3	Avondale	41.938666	-87.711211	A-Karrasel Child Care	41.936617	-87.707797	Daycare
4	Avondale	41.938666	-87.711211	Kat Slawson	41.935631	-87.715263	Daycare
5	Logan Square	41.923193	-87.707389	Christopher House	41.926334	-87.709773	Daycare
6	Logan Square	41.923193	-87.707389	A-Karrasel Child Care	41.924895	-87.712903	Daycare
7	East Side	41.707314	-87.534902	Shinning Star	41.703346	-87.535548	Daycare
8	Grand Crossing	41.763247	-87.616134	Jellybean Learning Center II	41.765854	-87.615811	Daycare
9	Grand Crossing	41.763247	-87.616134	Allison's Infant &Toddler Center	41.765966	-87.613692	Daycare
10	Loop	41.880052	-87.626993	Bright Horizons at Cook County/City of Chicago…	41.882722	-87.629767	Daycare
11	Magnificent Mile	41.894784	-87.624188	Butler Children's Prep	41.897270	-87.627411	Daycare
12	Magnificent Mile	41.894784	-87.624188	catherineschildren	41.891188	-87.624625	Daycare
13	West Loop	41.877690	-87.648721	Lily Pad Nursery + Preschool	41.876259	-87.652870	Daycare
14	Andersonville	41.979854	-87.667865	Mi Casita Chicago	41.979643	-87.664005	Daycare
15	Woodlawn	41.778787	-87.601686	Busy bumble bee daycare Academy	41.777096	-87.606196	Nursery School

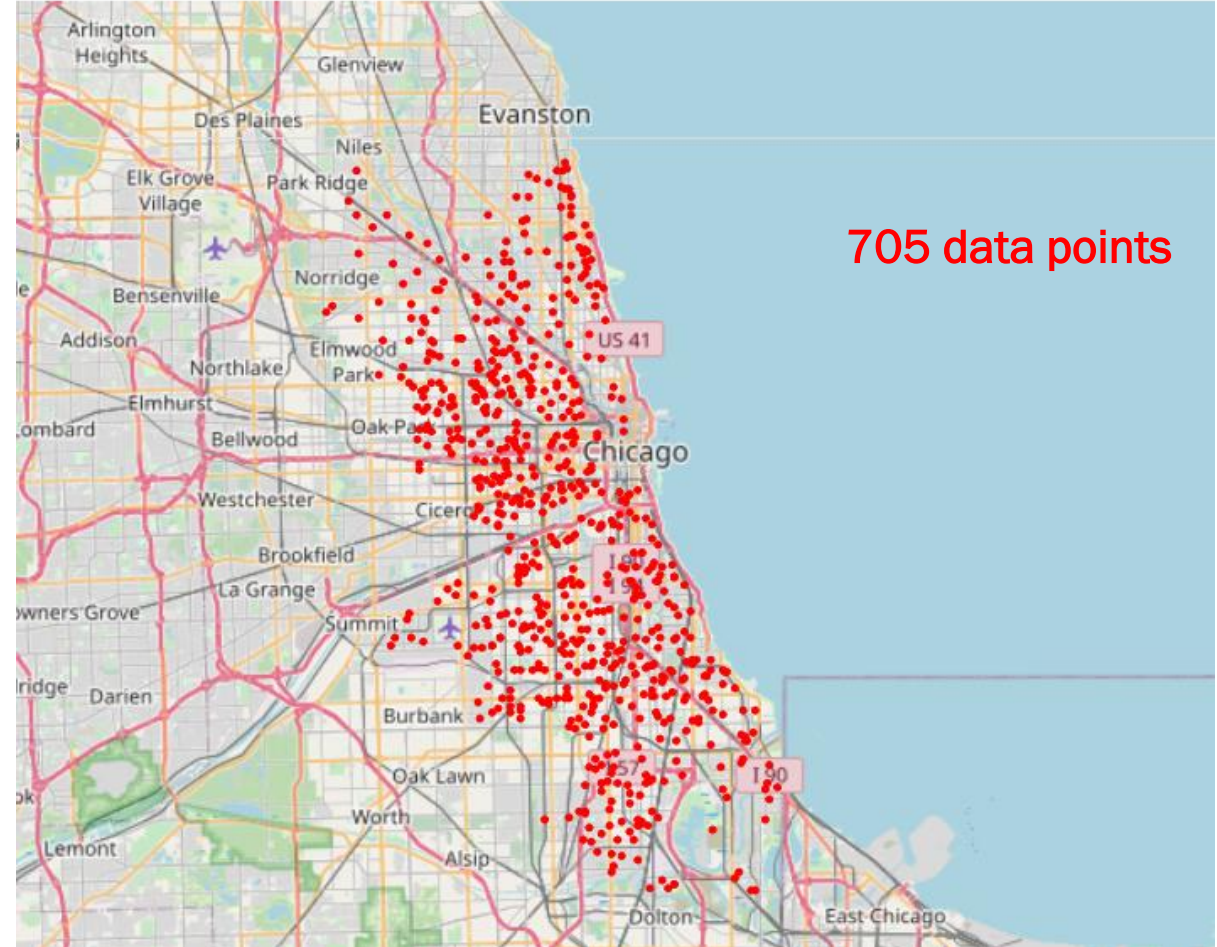# Data on daycares available on the Foursquare is very limited. We supplement it with the data from the city of Chicago.
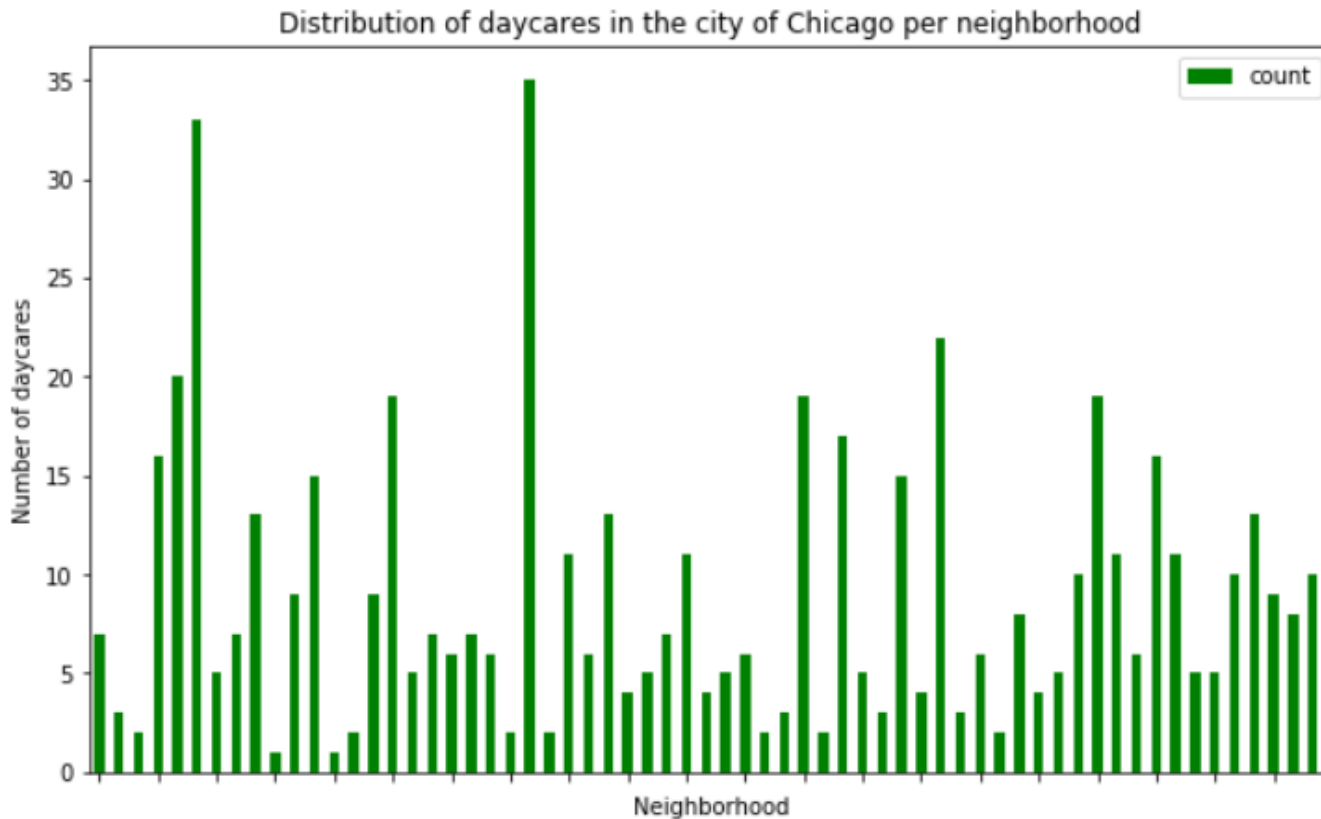

Daycare centers in Chicago, data from Foursquare — 36 data points
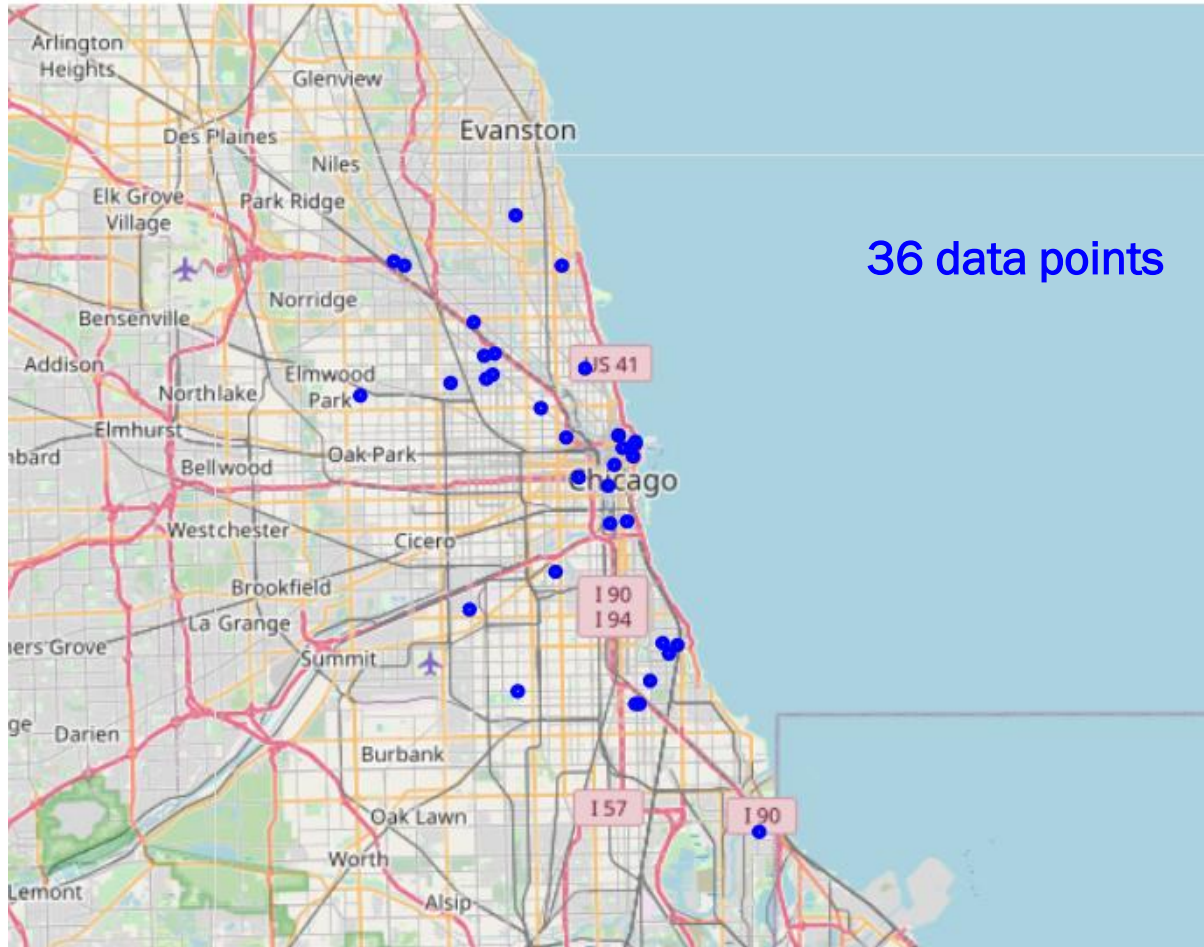
Daycare centers in Chicago, data from the city of Chicago — 705 data points

Exploratory analysis of daycare distributions per neighborhood

Distribution of daycares in the city of Chicago per neighborhood

```
count      63.000000
mean        8.841270
std         7.025774
min         1.000000
25%         4.000000
50%         7.000000
75%        11.000000
max        35.000000
Name: count, dtype: float64
```

- The number of daycares per neighborhood ranges from 1 to a maximum of 36.

- The graph shows the daycare density per neighborhood varies.

- Specific neighborhood names are shown on the next slide.

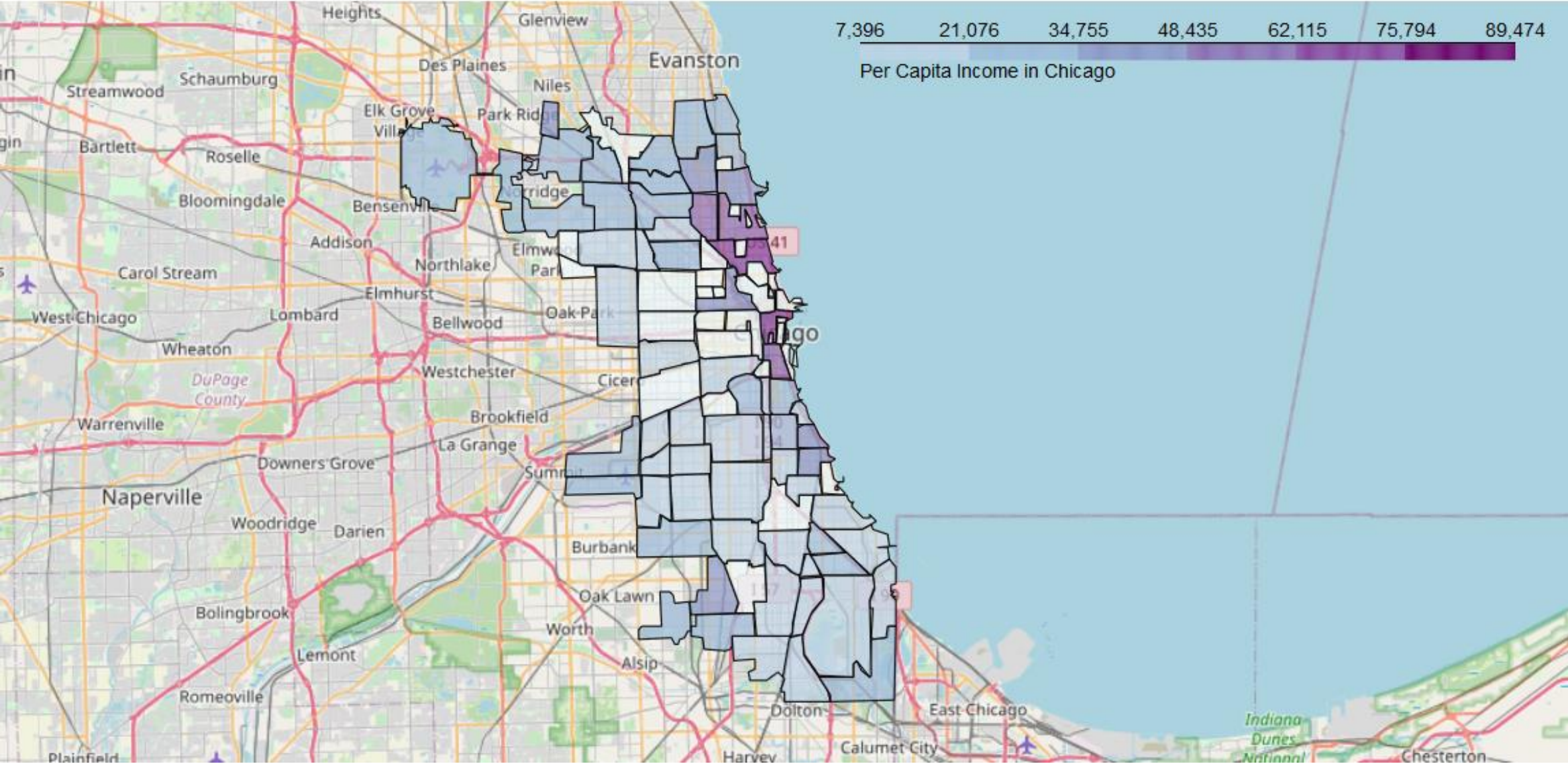# Chicago neighborhoods with high/low number of existing daycare facilities

10 Neighborhoods with the largest number of daycares

	COMMUNITY AREA NAME	count
22	Englewood	35
5	Austin	33
43	North Lawndale	22
4	Auburn Gresham	20
36	Logan Square	19
51	Roseland	19
15	Chicago Lawn	19
38	Lower West Side	17
3	Ashburn	16
54	South Shore	16

10 Neighborhoods with the smallest number of daycares

	COMMUNITY AREA NAME	count
1	Archer Heights	3
21	Edison Park	2
23	Fuller Park	2
13	Calumet Heights	2
37	Loop	2
34	Lincoln Park	2
2	Armour Square	2
46	Oakland	2
12	Burnside	1
9	Beverly	1

Per capita income in Chicago neighborhoods

Per Capita Income in Chicago

7,396　21,076　34,755　48,435　62,115　75,794　89,474

Taking closer look at Chicago neighborhoods with highest/lowest per capita income: Reference table with neighborhoods names

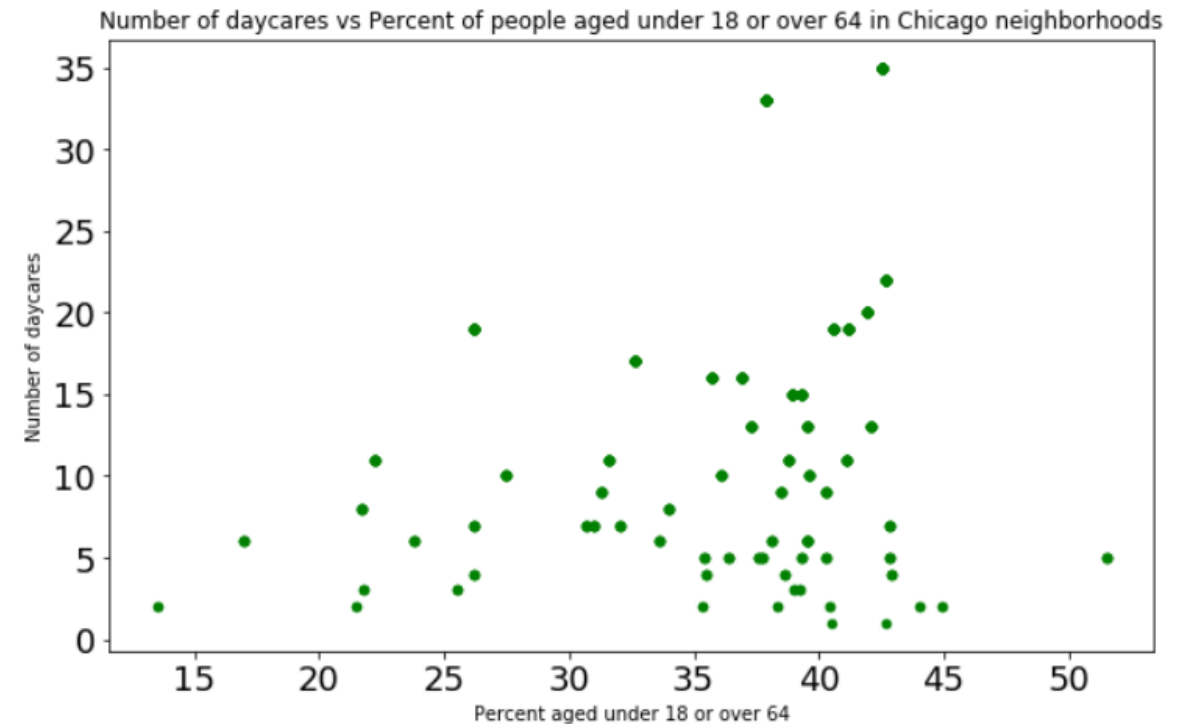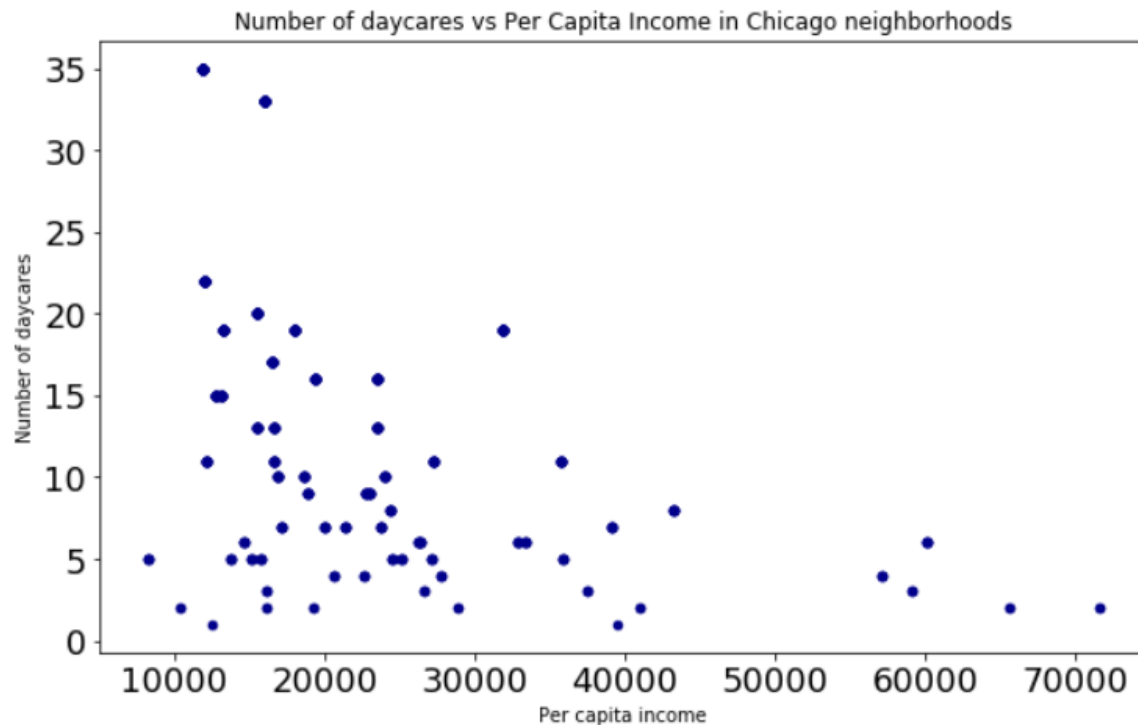Top 5 Chicago neighborhoods with **highest per capita income**

COMMUNITY AREA NAME	PERCENT OF HOUSING CROWDED	PERCENT HOUSEHOLDS BELOW POVERTY	PERCENT AGED 16+ UNEMPLOYED	PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA	PERCENT AGED UNDER 18 OR OVER 64	PER CAPITA INCOME	HARDSHIP INDEX
Near North Side	1.9	12.9	7.0	2.5	22.6	88669	1.0
Lincoln Park	0.8	12.3	5.1	3.6	21.5	71551	2.0
Loop	1.5	14.7	5.7	3.1	13.5	65526	3.0
Lake View	1.1	11.4	4.7	2.6	17.0	60058	5.0
Near South Side	1.3	13.8	4.9	7.4	21.8	59077	7.0

Top 5 Chicago neighborhoods with **lowest per capita income**

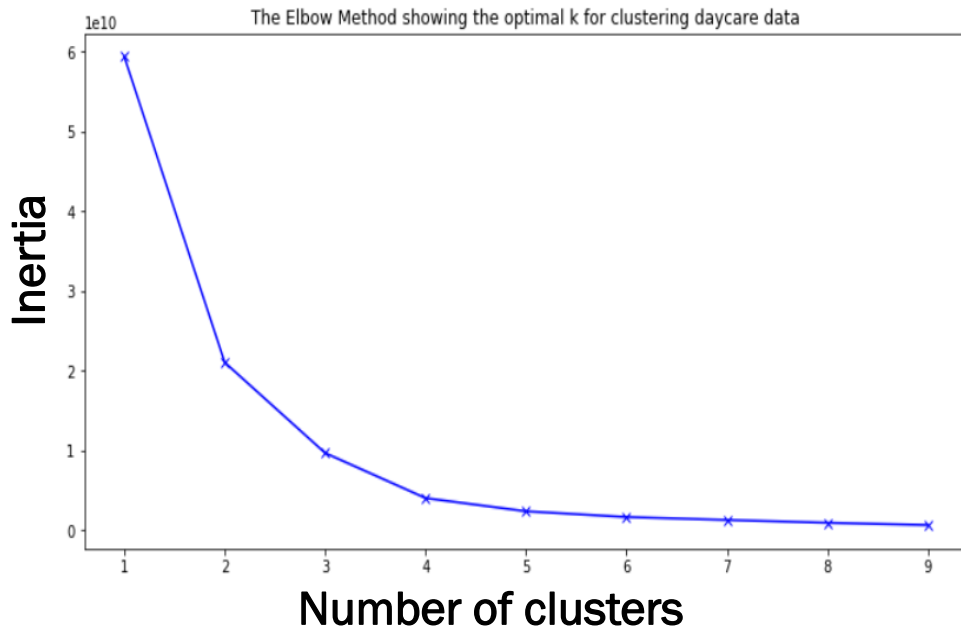| COMMUNITY AREA NAME	PERCENT OF HOUSING CROWDED	PERCENT HOUSEHOLDS BELOW POVERTY	PERCENT AGED 16+ UNEMPLOYED	PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA	PERCENT AGED UNDER 18 OR OVER 64	PER CAPITA INCOME	HARDSHIP INDEX
West Englewood	4.8	34.4	35.9	26.3	40.7	11317	89.0
West Garfield Park	9.4	41.7	25.8	24.5	43.6	10934	92.0
Fuller Park	3.2	51.2	33.9	26.6	44.9	10432	97.0
South Lawndale	15.2	30.7	15.8	54.8	33.8	10402	96.0
Riverdale	5.8	56.5	34.6	27.5	51.5	8201	98.0

Impact of socioeconomic factors on the number of daycares per neighborhood: Lower income neighborhoods has larger number of daycares

Number of daycares vs Per Capita Income in Chicago neighborhoods

Number of daycares vs Percent of people aged under 18 or over 64 in Chicago neighborhoods

K-means clustering machine learning method

The Elbow Method showing the optimal k for clustering daycare data

(x-axis: Number of clusters, y-axis: Inertia)

- In order to perform the K-means clustering, we define the optimal number of clusters k.
- Calculate inertia for different number of clusters and analyze the Elbow plot.

- Optimal k = 4

- Analysis of each resulting cluster:

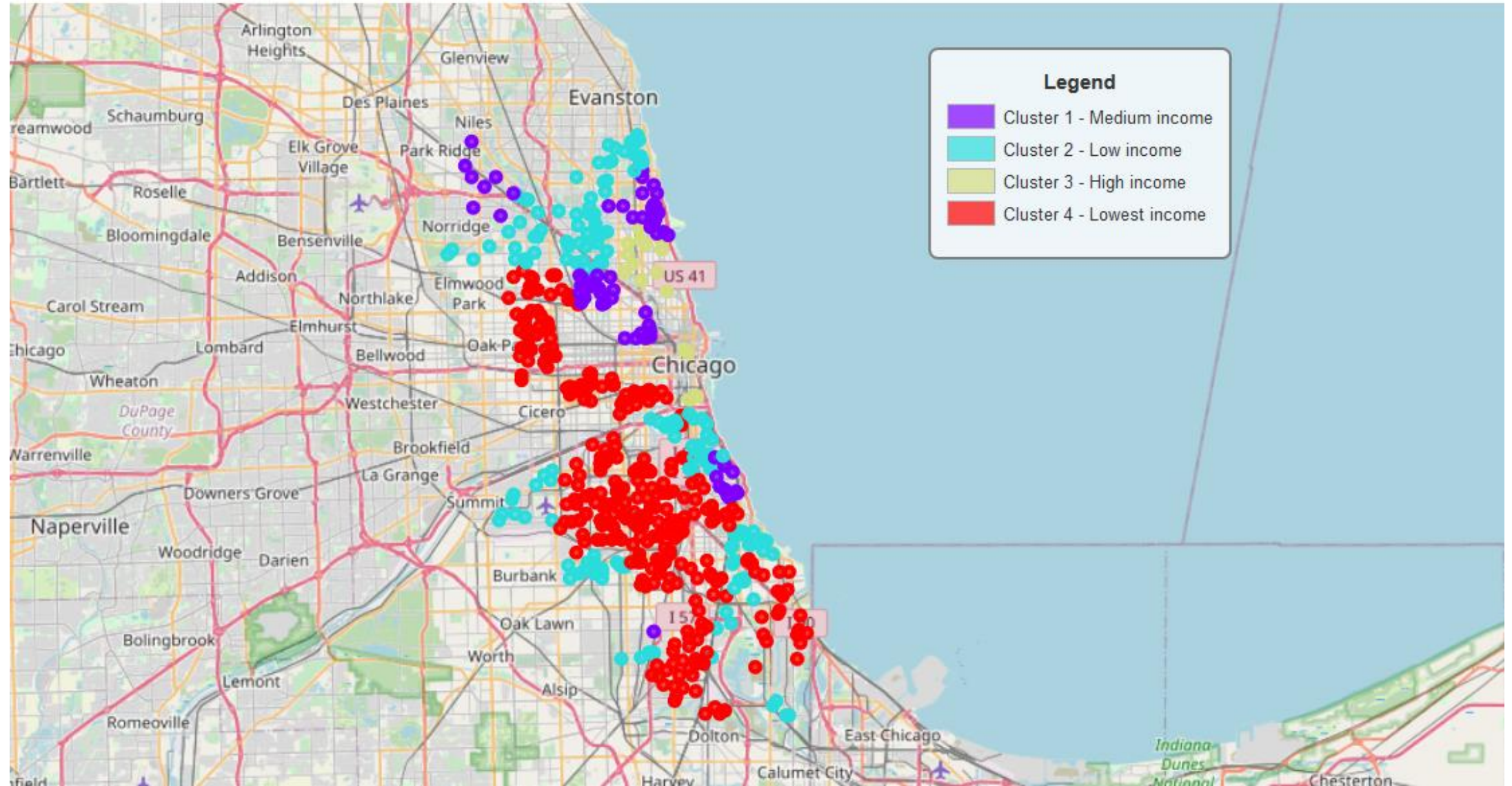 cluster 1 = 35735 average income, 68 data points

  cluster 2 = 23598 average income, 159 data points

  cluster 3 = 61189 average income, 17 data points

  cluster 4 = 14767 average income, 313 data points

Daycare data clustering, based on per capita income and number of daycares per neighborhood

Legend
- Cluster 1 - Medium income
- Cluster 2 - Low income
- Cluster 3 - High income
- Cluster 4 - Lowest income

Conclusions and future direction

- The goal of this project was to provide insight on daycares in Chicago, Illinois. The results of analysis showed that there are a lot of existing daycare businesses. The data about daycares on the Foursquare platform, which was required to be used in this project, is very limited (36 data points) and clearly does not represent the actual business landscape.

- Supplementing this dataset with the data from the city of Chicago allowed to perform a more detailed analysis. Impact of socioeconomic factors on the density of daycares per neighborhood was evaluated using k-means clustering machine learning algorithm.

- For a person, interested in starting a new daycare business, the results of this analysis will help estimate the acceptable price point for a given neighborhood. Low income neighborhoods will not be receptive to a daycare with high weekly payments. By looking at the daycares in each of the four data clusters, one can make a decision whether to pursue a new daycare business venue in a given neighborhood.

- Further analysis can be performed on the types of existing daycare businesses (full time vs part time) as well as the age of children those businesses serve.