Helping parents choose the perfect neighborhood for their needs

Nadav Tzuk

June 22, 2019

1. Introduction

1.1. Background

Not so many years ago, while I and my wife were looking for an apartment, we had difficulties to decide where to live. We knew which city we cared for most, but choosing the area was much more complicated. On the one hand, we wanted to be close by to cofee houses and bars, but on the other hand we were expecting our first born, so we needed to live nearby health care and education facilities. In those days we had to base our decisions on friend's suggestions and looking into maps. I thought, wouldn't it be great to have an automated system to analyze which areas are suttible for people to live ?

1.2. Problem and Interset

Parents that are looking for an apartment have to gather a lot of information to support their decision. A machine learning model can help to choose the most suitable place to live.

2. Data

Parents looking for a place to live, usually prefer to live nearby the most needed facilities, such as: education for their children, health, shopping and recreation. Thus, The data needed to help parents select a preferable living place, can be separated into three main categories:

1. List of neighborhoods – first of all, a list of all Chicago's neighborhood's names is needed. It will be aquiered from a wikipedia web page. Off course, any city neighborhood's list can be used, even a list of cities in a specific country.
2. Coordinates - Using geolocator to download the coordinates for every neighborhood.

3. Forsquare - Using forsquare to find a list of places in each neighborhood, divided into four categories:
    a. Education: Daycare, Preschool, Elementary School, High School
    b. Health: Pharmacy, Hospital, Doctors Office, Dentists Office
    c. Shopping: Grocery Store, Baby Store, Shopping Mall, Shopping Plaza
    d. Recreation: Coffee Shop, Bar, Gym, Park

2.1. Data cleaning

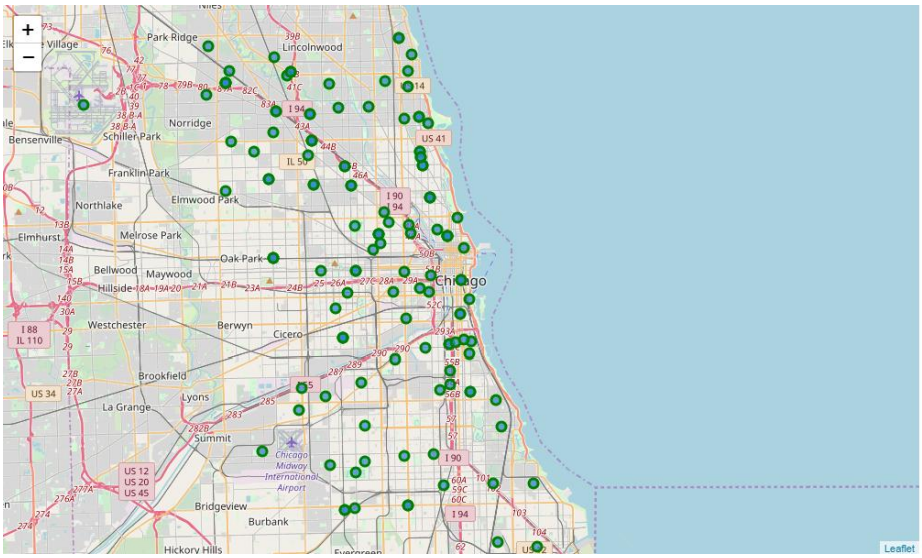
In order to locate a list of neighborhoods in Chicago, I used the data in a wikipedia web page. This page had a lot of irellevant information needed to be cleaned using python code. Moreover, The geo locator did not find five of the neighbothoods, therefor I removed them from the data.

3. Methodology

3.1. Exploratory Data Analysis

There are 159 neighborhoods in Chicago (164 before cleaning). For each neighborhood I searched for places that are walking distance, therefor a 500 hundred meters radius was chosen for the search with forsquare.

3.1.1. Map of Chicago whith neighborhoods green markers



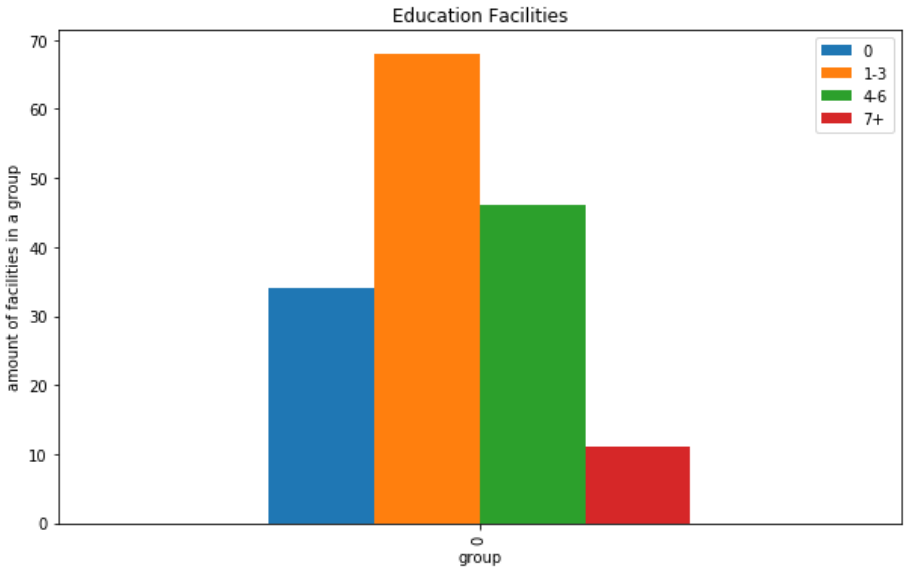
3.1.2. Descriptive statistics

	Education	Health	Shopping	Recreation
count	159.000000	159.000000	159.000000	159.000000
mean	2.691824	13.371069	2.226415	22.735849
std	2.330181	19.227339	2.508024	28.788722
min	0.000000	0.000000	0.000000	0.000000
25%	1.000000	2.000000	0.000000	5.000000
50%	2.000000	6.000000	1.000000	10.000000
75%	5.000000	18.000000	3.000000	27.500000
max	9.000000	155.000000	14.000000	136.000000

3.1.2.1. Education

In average, there are 2.69 Education facilities is every neighborhood.

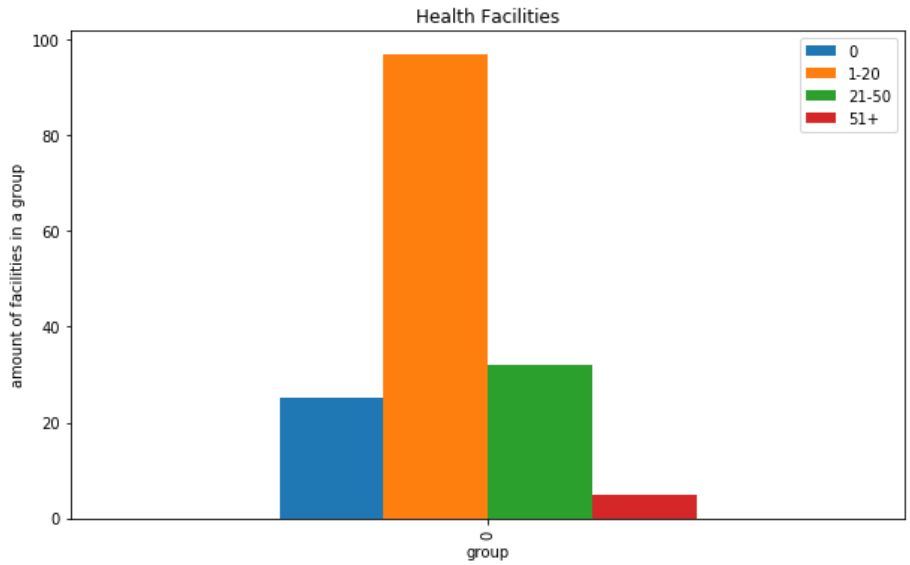
34 neighborhoods don't have any education facilities, and 11 neighborhoods have at least 7 facilities.



3.1.2.2. Health

In average, there are 13.37 Health facilities is every neighborhood.

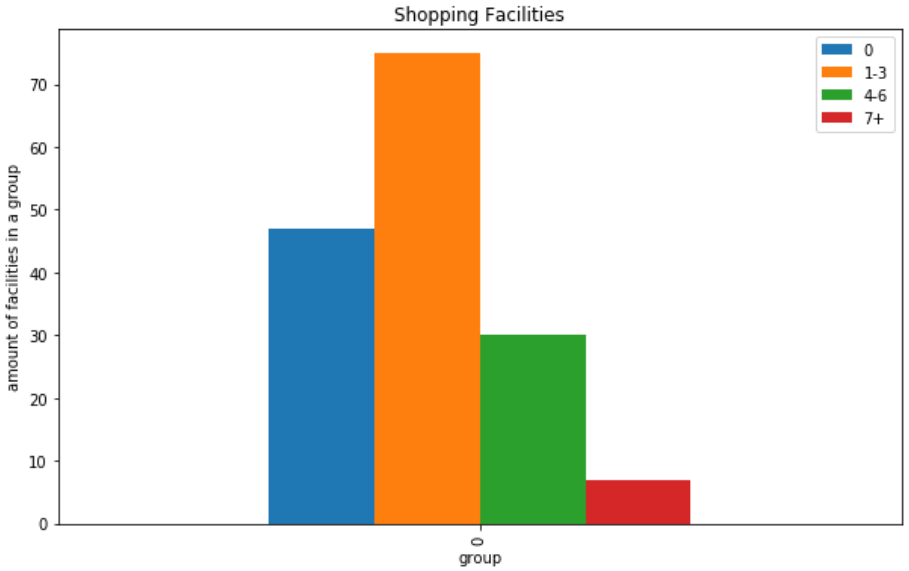
25 neighborhoods don't have any Health facilities, and 5 neighborhoods have at least 51 facilities.



3.1.2.3. Shopping

In average, there are 2.23 Shopping facilities is every neighborhood.

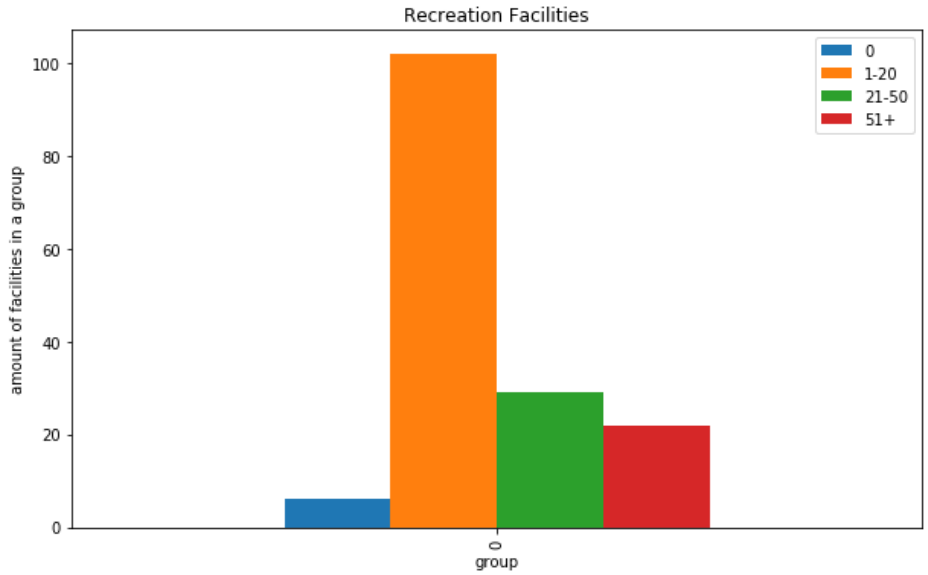
47 neighborhoods don't have any Shopping facilities, and 7 neighborhoods have at least 7 facilities.



3.1.2.4. Recreation

In average, there are 22.74 Recreation facilities is every neighborhood.

6 neighborhoods don't have any Recreation facilities, and 22 neighborhoods have at least 51 facilities.



3.2. Machine learning model

In order to make the desicion which neighborhood is better, it is easier to divide the 159 neighborhoods into main groups. The k-means model was chosen for this unsupervised task. Parents who would like to chose place to live will ne able to do so according to their prefences.

The model chosen will split the neighborhoods into three clusters, because two clusters are few and for more then three clusters one cluster has only one neighborhood therefor we don't get more information by adding clusters.

4. Results

The results of the model show three clusters:

4.1. low amount of facilities - the cluster has 103 neighborhoods included. In this cluster, in average, there are 2.17 Education facilities, 4.98 health facilities, 0.92 shopping facilities and 7.94 recreation facilities.

4.2. Middle amount of facilities - the cluster has 36 neighborhoods included. In this cluster, in average, there are 3.42 Education facilities, 18.72 health facilities, 3.97 shopping facilities and 29.63 recreation facilities.

4.3. High of amount facilities - the cluster has 20 neighborhoods included. In this cluster, in average, there are 4.05 Education facilities, 46.95 health facilities, 5.8 shopping facilities and 86.5 recreation facilities.

Labels	Education	Health	Shopping	Recreation
0	3.416667	18.722222	3.972222	29.638889
1	2.174757	4.980583	0.922330	7.941748
2	4.050000	46.950000	5.800000	86.500000

Labels	Education	Health	Shopping	Recreation
0	36	36	36	36
1	103	103	103	103
2	20	20	20	20

5. Discussion and Conclusion

In the background I raised a problem many parents encounter, where to live ? parents usually care for facilities in a walking distance. In this paper I suggested to enquire the neighborhoods in chicago in order to divide them into three groups, each depending on the number of facilities (Education, Health, Shopping and Recreation). According to these results, parents can easily choose the neighborhood they prefer in accordance to the facilities in the preferred cluster.

For future studies, many more properties of a neighborhood can ne taken into account in order to help the , such as rent prices, socio economic status, age of population and more.