IDENTIFYING THE BEST NEIGHBORHOODS TO MOVE TO IN CHICAGO

Applied Data Science Capstone Project

THERE ARE SEVERAL THINGS TO CONSIDER WHEN FINDING A NEIGHBORHOOD TO MOVE TO:

- Crime rates
- House prices
- Proximity to work
- Proximity to other venues of interest

For a particular family, interests are:

- School rankings
- Proximity to parks, martial arts schools, and music schools

We gather the most recent data in all these categories and determine the best neighborhoods (community areas) this family can move to in Chicago.

DATA COLLECTION AND CLEANING

1. Geographic boundaries of Chicago community areas (Chicago Data Portal)
2. Wikipedia data on population and population density of community areas
3. Wikipedia data for community area - neighborhood relation
4. Latitude and longitude information for community areas (Geopy geocoder library)
5. Crime Data (Chicago Data Portal – SODA API)
6. Zillow Housing Data (ZILLOW – research data)
7. School Ranking Data (Chicago Data Portal)
8. Venue data (Foursquare API)
9. Driving distance to Downtown Chicago (Microsoft Bing Maps Distance Matrix API)

DATA COLLECTION AND CLEANING

- Group crime data (all incidents reported since the beginning of 2019 till 11/27/2020) by community areas
  - Huge data set -> use SODA API with an SQL-like query to extract relevant data only
- Estimate community area median house price s using Zillow single value house price estimates for neighborhoods
  - Chicago neighborhoods are not official, thus different sources list different number of neighborhoods
  - Community area – neighborhood relation available in Wikipedia
  - Wikipedia neighborhoods (246) do not exactly match Zillow neighborhoods (200)
- Estimate community area median school scores
  - School data has ranking and latitude longitude information, no association with community areas
  - Write a function that determines whether a latitude-longitude pair belongs to a community area using the geojson geometry file
  - Convert categorical school rankings to numerical scores

DATA COLLECTION AND CLEANING

- Number of venues of interest within each community area
  - Determine venue category ids for parks, martial arts schools, music schools
  - Use Foursquare API with the search endpoint.
- Driving distance to Downtown Chicago (job location)
  - Create a developer account for Microsoft Bing Maps Distance Matrix API
  - Provide latitudes-longitudes of community areas as origins
  - Provide latitude-longitude of Downtown Chicago as the destination
  - Set Travelmode to Driving in the url

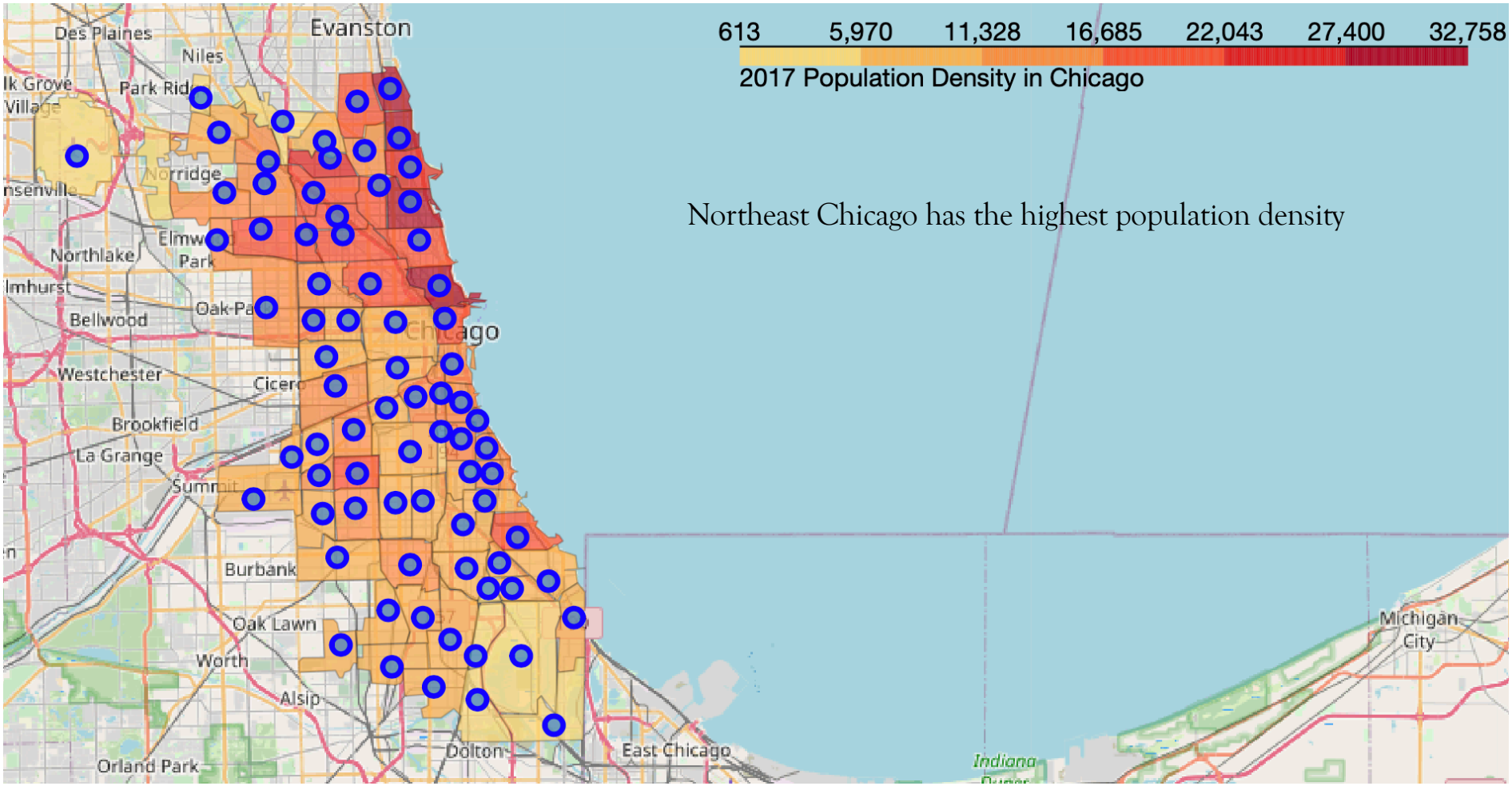
METHODOLOGY

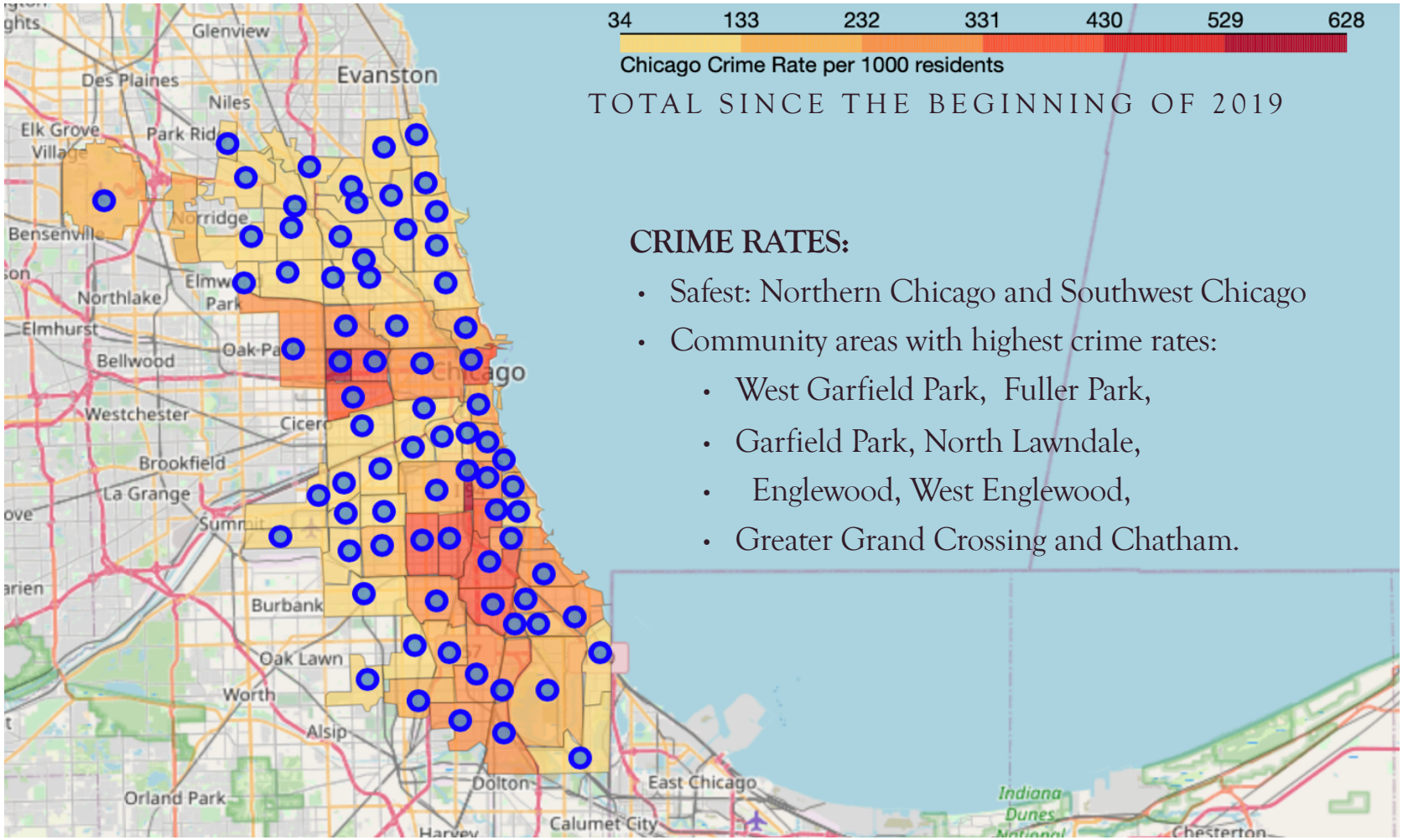
Explatory Data Analysis

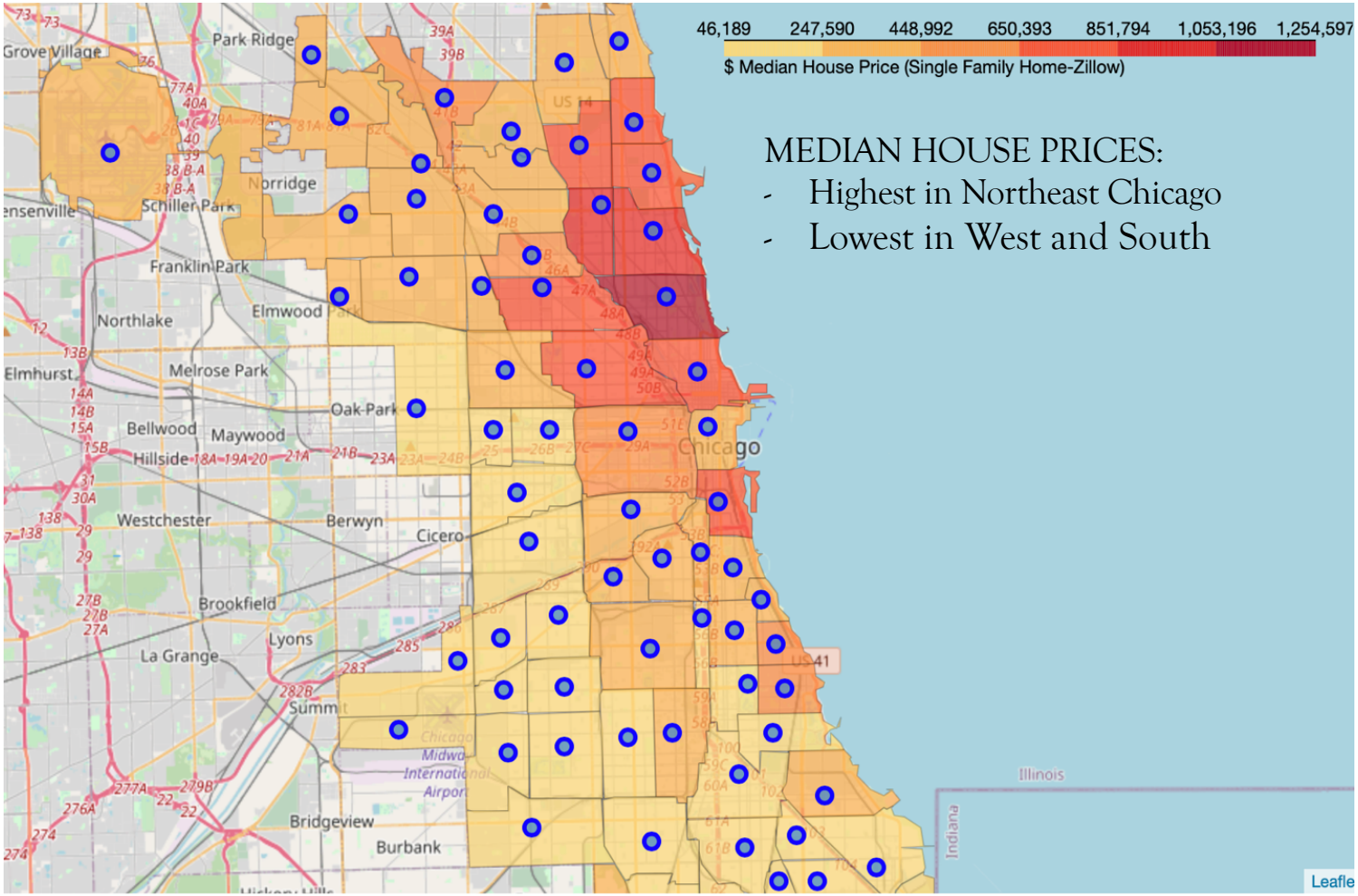
Folium Choropleth maps for each feature

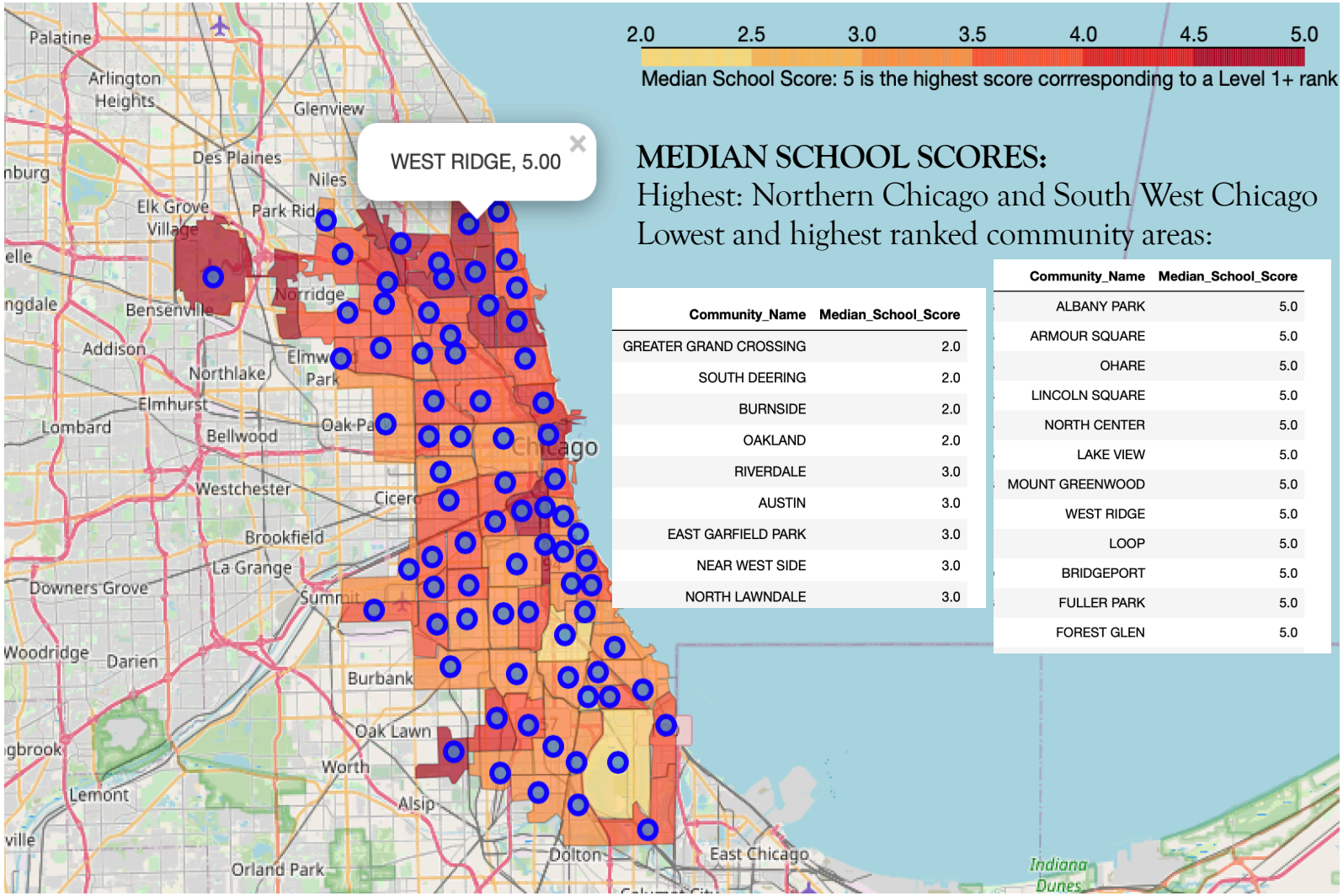
Histograms

Kmeans Clustering


613
5,970
11,328
16,685
22,043
27,400
32,758
2017 Population Density in Chicago
Northeast Chicago has the highest population density
Evanston
Des Plaines
Niles
Park Ridge
Norridge
Northlake
Elmwood Park
Oak Park
Bellwood
Westchester
Cicero
Brookfield
La Grange
Summit
Burbank
Oak Lawn
Worth
Alsip
Orland Park
Dolton
East Chicago
Michigan City
Indiana


34
133
232
331
430
529
628
Chicago Crime Rate per 1000 residents
TOTAL SINCE THE BEGINNING OF 2019
CRIME RATES:
• Safest: Northern Chicago and Southwest Chicago
• Community areas with highest crime rates:
• West Garfield Park, Fuller Park,
• Garfield Park, North Lawndale,
• Englewood, West Englewood,
• Greater Grand Crossing and Chatham.
Glenview
Evanston
Des Plaines
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Northlake
Elmhurst
Bellwood
Westchester
Cicero
Brookfield
La Grange
Burbank
Oak Lawn
Worth
Alsip
Orland Park
Dolton
Calumet City
East Chicago
Indiana Dunes


46,189
247,590
448,992
650,393
851,794
1,053,196
1,254,597
$ Median House Price (Single Family Home-Zillow)
MEDIAN HOUSE PRICES:
- Highest in Northeast Chicago
- Lowest in West and South
Park Ridge
Norridge
Schiller Park
Franklin Park
Elmwood
Northlake
Elmhurst
Melrose Park
Oak Park
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
Lyons
La Grange
Bridgeview
Burbank
Chicago
Illinois
Indiana
Leafle



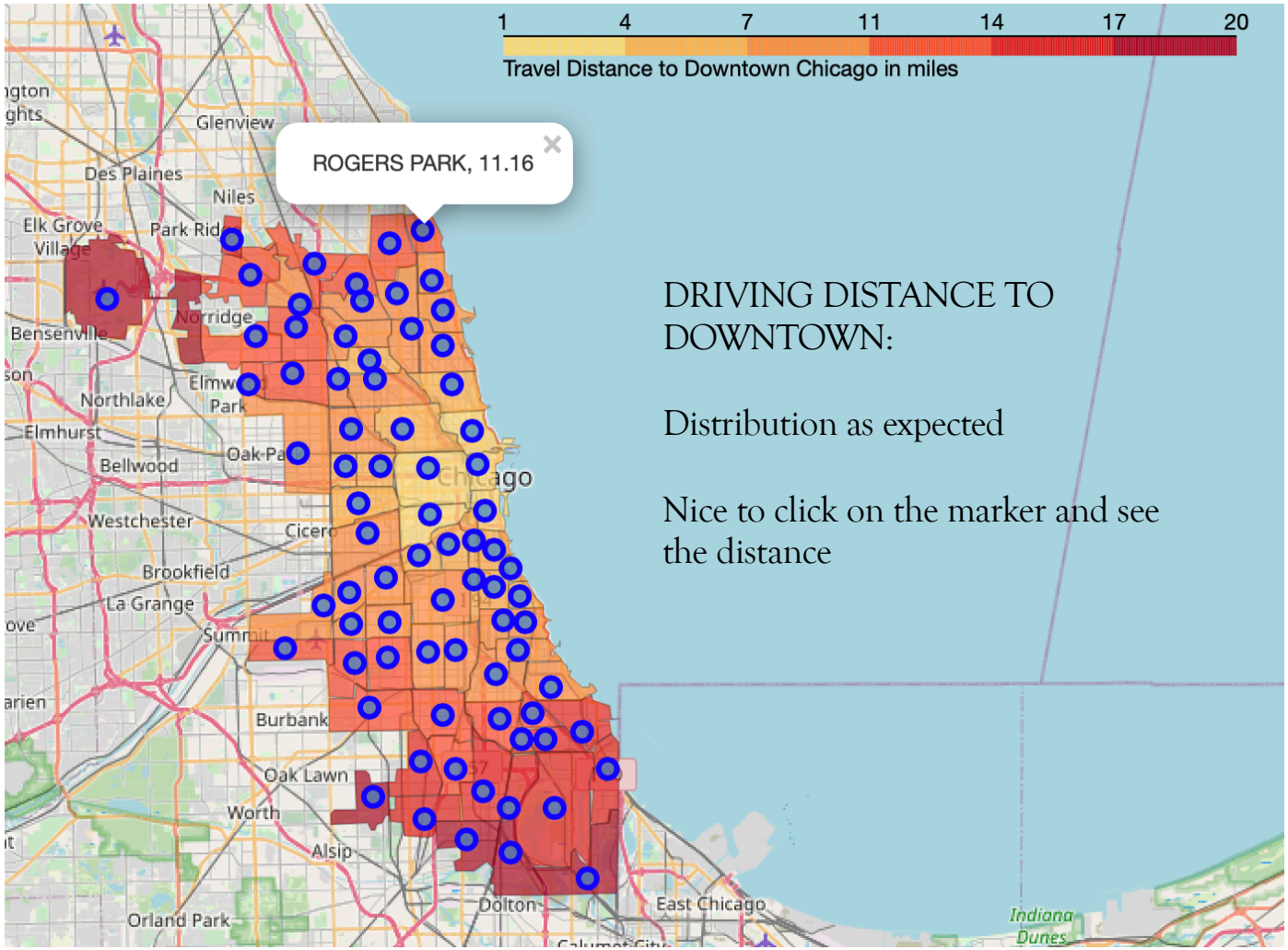
MEDIAN SCHOOL SCORES:

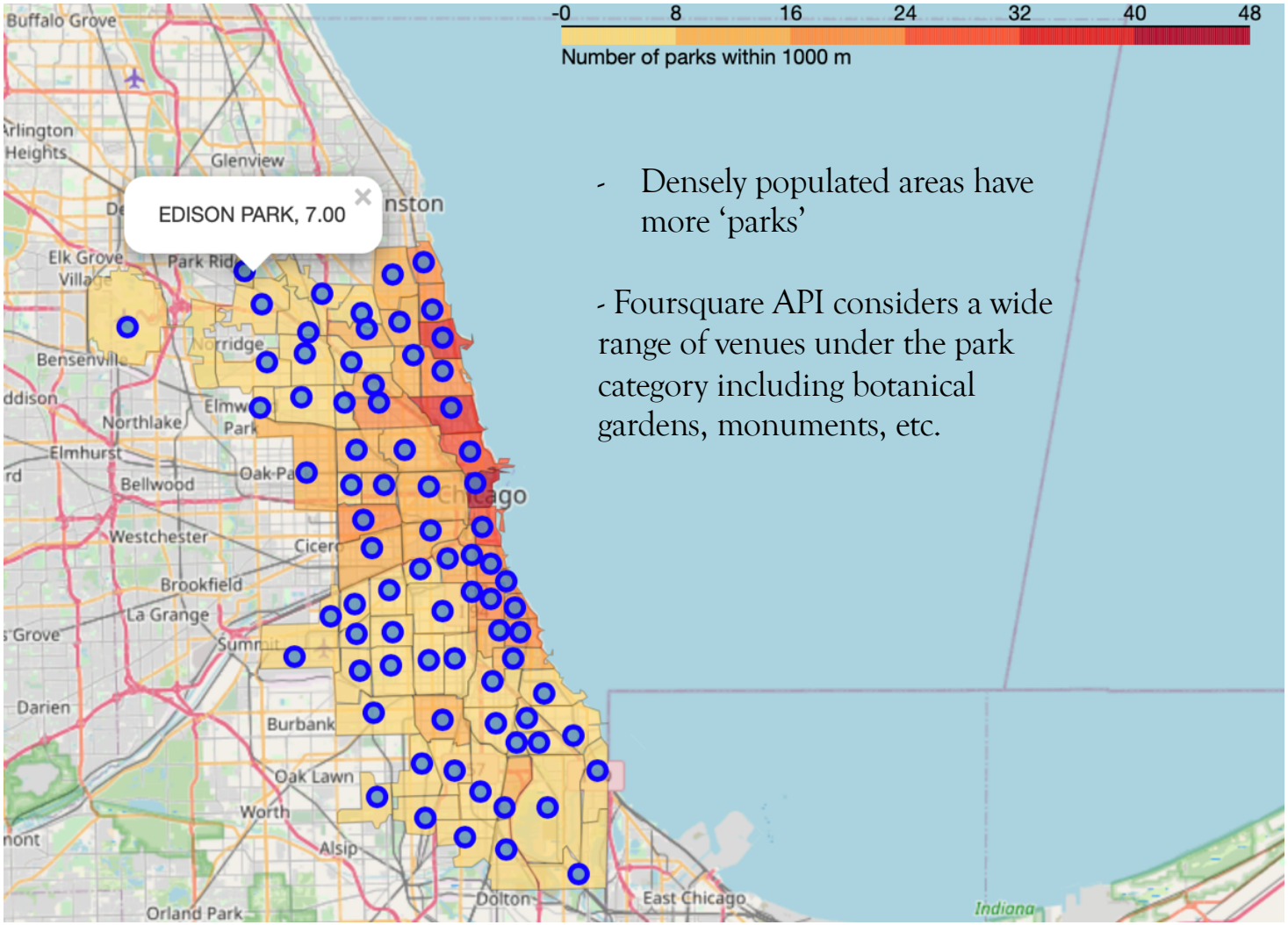
Highest: Northern Chicago and South West Chicago

Lowest and highest ranked community areas:

Community_Name	Median_School_Score
GREATER GRAND CROSSING	2.0
SOUTH DEERING	2.0
BURNSIDE	2.0
OAKLAND	2.0
RIVERDALE	3.0
AUSTIN	3.0
EAST GARFIELD PARK	3.0
NEAR WEST SIDE	3.0
NORTH LAWNDALE	3.0

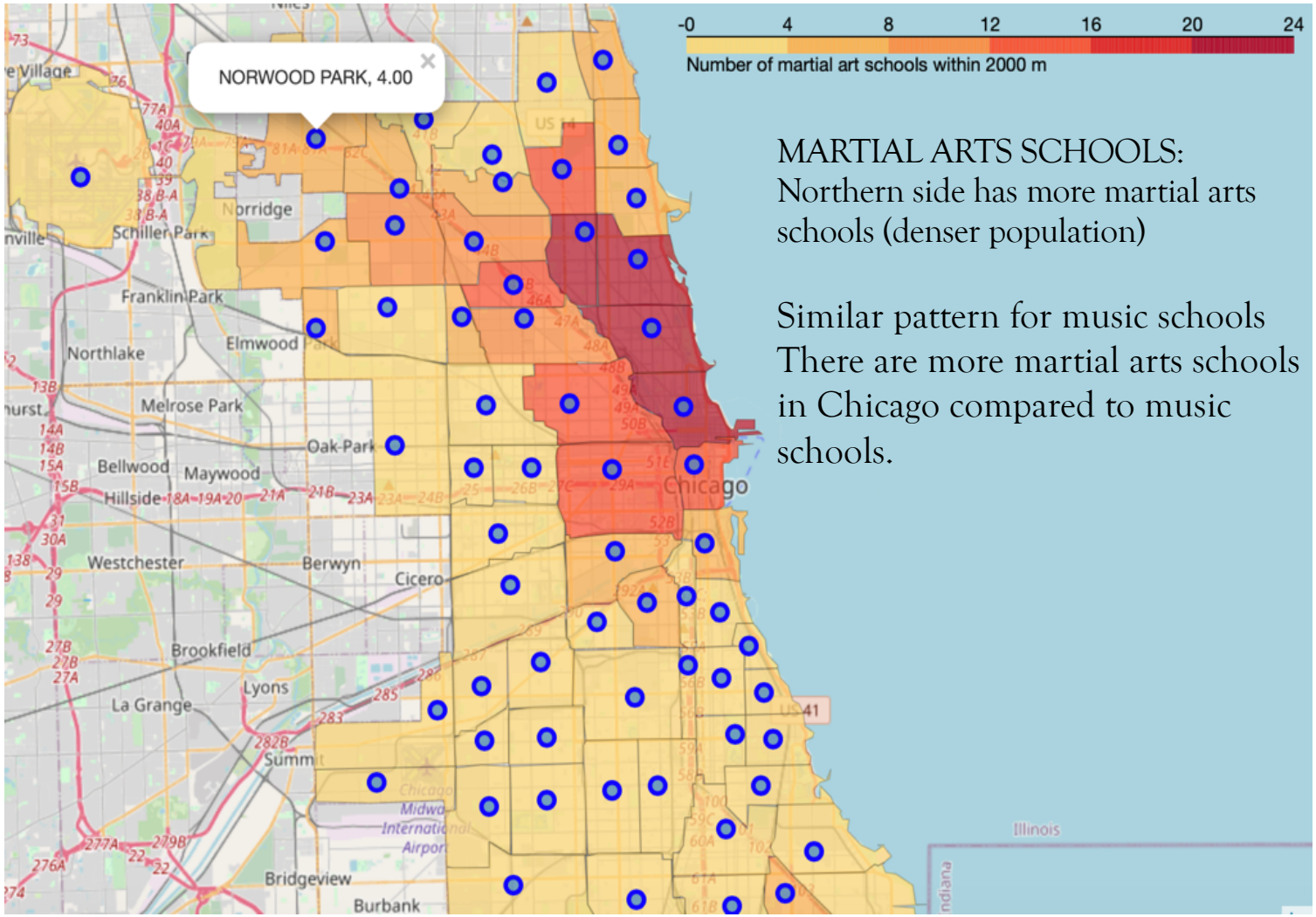
Community_Name	Median_School_Score
ALBANY PARK	5.0
ARMOUR SQUARE	5.0
OHARE	5.0
LINCOLN SQUARE	5.0
NORTH CENTER	5.0
LAKE VIEW	5.0
MOUNT GREENWOOD	5.0
WEST RIDGE	5.0
LOOP	5.0
BRIDGEPORT	5.0
FULLER PARK	5.0
FOREST GLEN	5.0


1
4
7
11
14
17
20
Travel Distance to Downtown Chicago in miles
ROGERS PARK, 11.16
Glenview
Des Plaines
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Northlake
Elmhurst
Bellwood
Westchester
Cicero
Brookfield
La Grange
Burbank
Oak Lawn
Worth
Alsip
Orland Park
Dolton
East Chicago
Indiana Dunes
DRIVING DISTANCE TO DOWNTOWN:
Distribution as expected
Nice to click on the marker and see the distance



- Densely populated areas have more ‘parks’

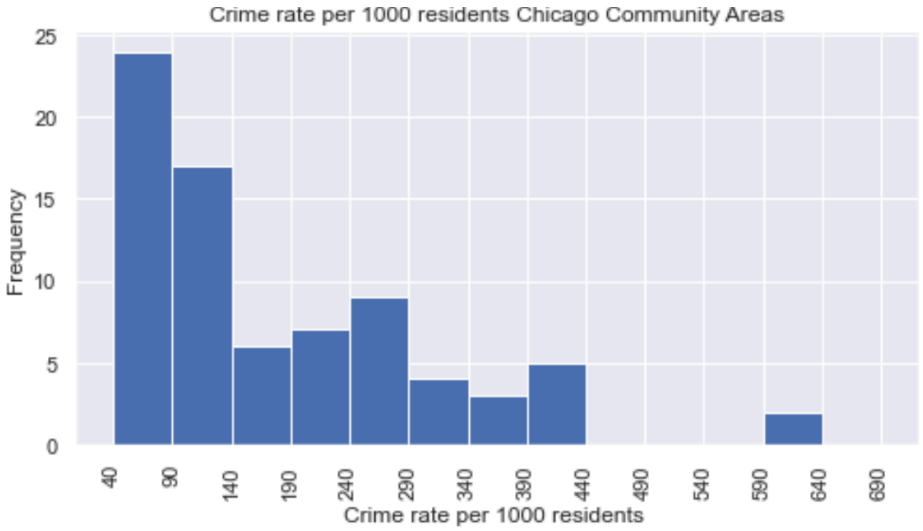
- Foursquare API considers a wide range of venues under the park category including botanical gardens, monuments, etc.

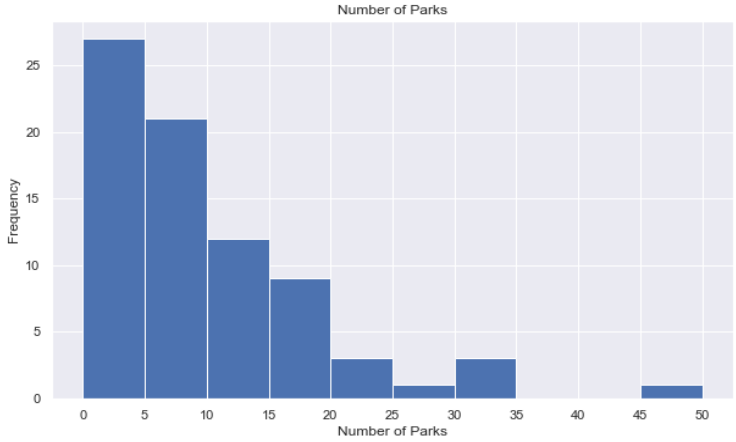


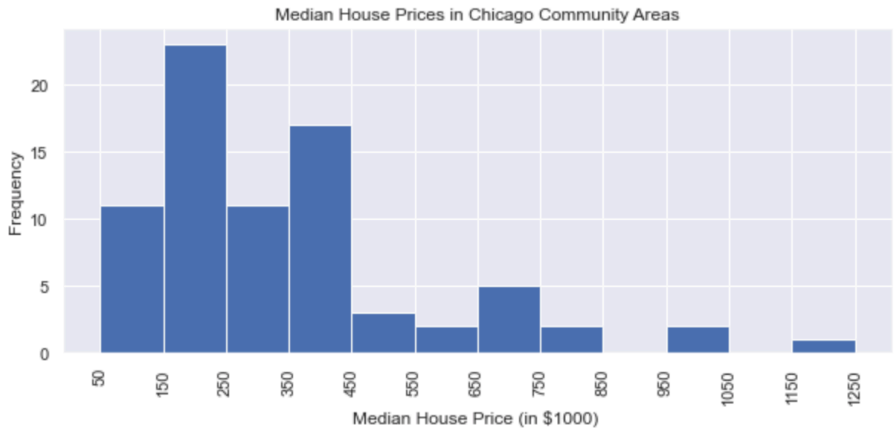
MARTIAL ARTS SCHOOLS:
Northern side has more martial arts schools (denser population)

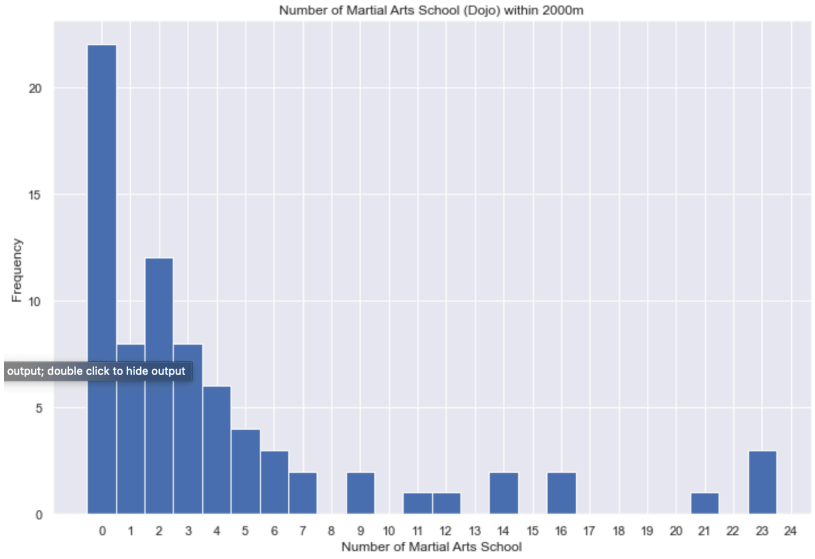
Similar pattern for music schools
There are more martial arts schools in Chicago compared to music schools.

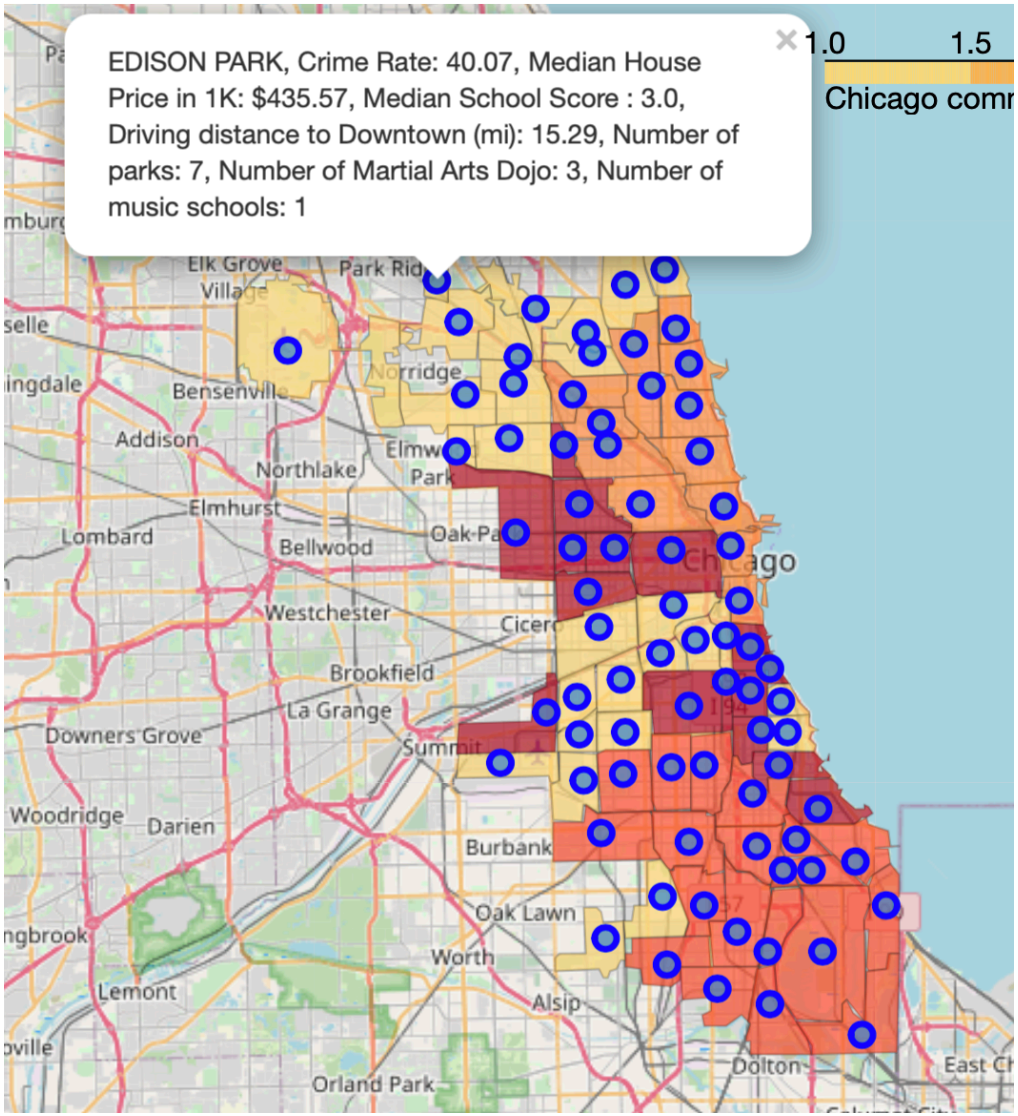
Most of these features have skewed distributions – quite natural











KMEANS clustering

- Skewed features: Take the logarithm to normalize
- Use standard scaler (obtain z scores)
- Consider importance of features:

Primary features: Crime rate, Median House Price Median School Score

Secondary features: Driving distance, number of venues of interest

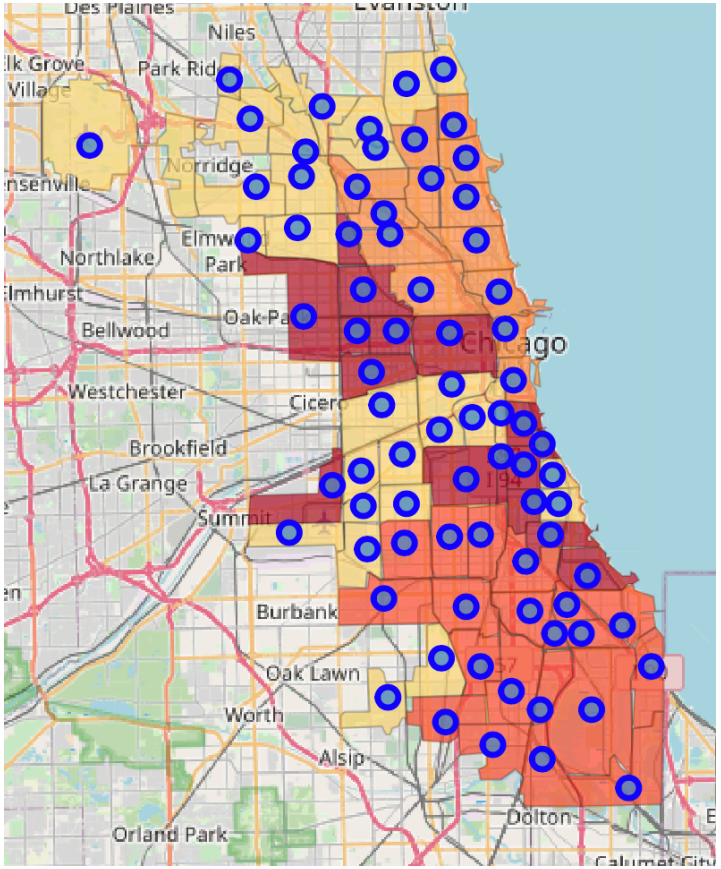
1st level : use primary features in KMeans, obtain 2 clusters – use elbow method and silhouette scores to determine K=2

2nd level: Use secondary features in Kmeans for both of the primary clusters - use elbow method and silhouette scores to determine K=2

Obtained 4 clusters, each represented by a different color on the map.

Map also has markers for each community area listing feature values

The family can compare community areas using this map

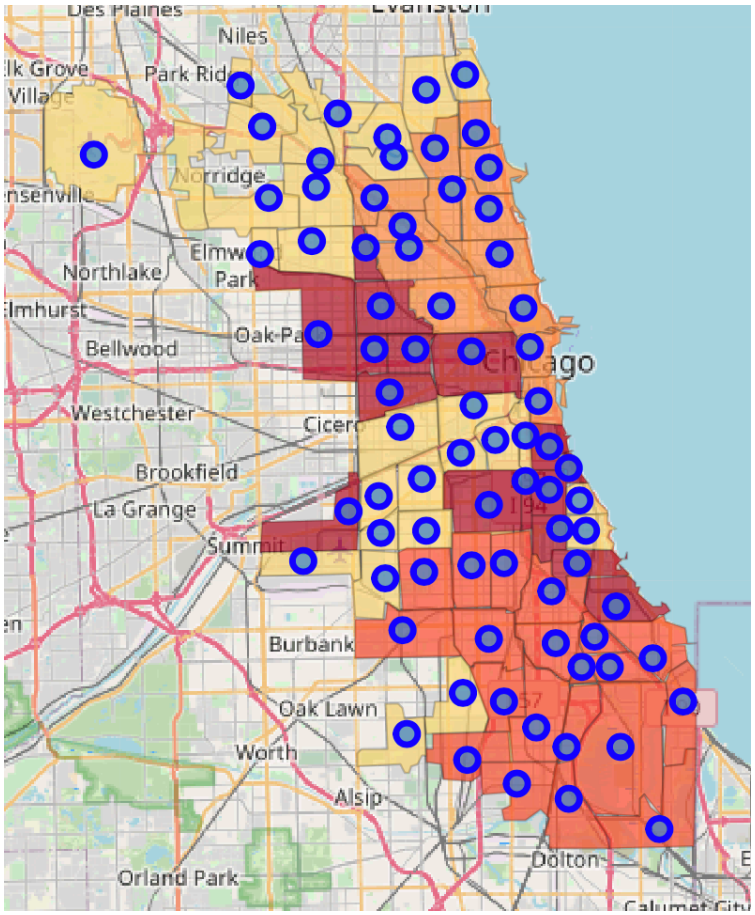


Yellow cluster (1):

- lower crime rates (most are below 100 and all are below 145 per 1000 residents).
- median house prices below $450000 - *except Kenwood and Hyde Park , $640,000, at the shore*
- median school scores mostly at or above 4 – *except Gage Park, Clearing (3.5)*
- number of music schools or martial arts dojos within 2km is 0 or 1- *except Jefferson Park and Portage Park*

Orange cluster (2):

- lower crime rates. most below 120, the exceptions are the community areas very close to Downtown such as Loop (424.64) and Near North Side (214.99)
- median house prices start at $400K and go as high as $1200K. The community areas in the north west region of this cluster, **Irving Park and Avondale,** have the lowest house prices in this cluster.
- median school scores are all at or above 4 with only one exception of 3.5 (Near South Side)
- parks, music schools, and martial arts dojos are abundant in these community areas.



Dark orange cluster (3):

- northern part of this cluster -> pretty high crime rates (Englewood: 425, Greater Grand Crossing: 378)
- Lowest house prices –most under $150K
- The median school scores are all at or below 3 – *except Englewood and Eastside (3.5)*
- some parks, no music schools, very few have martial arts schools

Darkest cluster (4):

- high crime rates- most above 200.
- Hermosa and Garfield Ridge are the safest community areas in this cluster with crime rates below 100.
- median house prices close to $200K and above going as high as $500K.
- median school scores are 3 or below except for Humboldt Park (4) and West Garfield Park (3).
- The community areas in this cluster at the top side have in general good number of parks, music schools and martial art dojos. The southern parts of the cluster lack those venue types except for the parks.

RESULTS

- **the community areas in the Orange Cluster (2) are the best if the family can afford it.**
- Especially **Irving Park** and **Avondale** can be good choices as they have lower house prices.
- **The next best cluster is the Yellow Cluster (1).** The family can especially consider **the community areas in this cluster that are at the boundary of the Orange Cluster.**

We can also filter the values in our data frame to come up with a short list. Let's use the following criteria:

- 1. median house price is less than $700,000
- 2. crime rate less than 230 (per 1000 residents considering total number of crimes in since 2019)
- 3. median school score is greater than or equal to 4
- 4. more than 2 parks
- 5. at least one music school and one martial arts school
- 6. driving distance less than 11 miles

Jefferson Park is the best choice based on these criteria:

Community Area	Crime Rate	Median House Price	Median School Score	Driving Distance to Downtown (mi)	Number of Parks within 1km	Number of Martial Arts Schools within 2 km	Number of Music Schools within 2 km
Jefferson Park	66.73	$312,732.5	4.5	10.58	5	7	2
Irving Park	81.75	$427,607.0	4	8.38	4	9	3
Avondale	99.76	$486,231.0	4	7.19	15	12	6
West Town	141.62	$692,204.5	4	5.21	11	14	3

CONCLUSION

- Collected most recent crime and median house price data
- Collected data from several sources (web scraping, use of APIs: Foursquare, Microsoft Bing Maps Distance Matrix API, SODA API) and combined at the community area level
- Created choropleth maps and histograms for Explatory Data Analysis
- Used Kmeans clustering to form clusters and interpreted them
- Accounted for importance of features in clustering using a 2-step clustering approach focusing on primary and secondary features
- Generated a map that the family/user can use to compare community areas
- Suggestied clusters of community areas to look into (1-orange, 2-yellow at the orange boundary)
- Short listed some community areas (Jefferson Park, Irving Park, Avondale, West Town)
- Valuable results to any person who is considering to move to Chicago and people/stakeholders who would want to know more about Chicago community areas for other purposes including city/business planning.
- Got many ideas for new projects (many data sets, APIs)