



PYTHON DATA SCIENCE CAPSTONE PROJECT PRESENTATION

Title: *Battle of the Neighborhoods – Finding the
right neighborhood in Chicago*

Author: *Dayli Steinhoff*

Date: *April/2020*

A stylized graphic of a city skyline in shades of blue and white, featuring various skyscrapers and a Ferris wheel, positioned on the left side of the slide.

AGENDA

 *Introduction*

 *Data Acquisition and Cleaning*

 *Methodology*

 *Results*

 *Discussion and Conclusion*



INTRODUCTION

(FROM PROBLEM TO SOLUTION)

Finding the right neighborhoods in Chicago

Chicago is considered one of the most popular cities in the US. This case study targets to support „young families“ willing to move to Chicago by finding a suitable neighborhood. For new families moving into this city it is **hard to find the proper place** to live because of the **large number of neighbourhoods** and also because of **lack of (customized) information** for the target group.

starting
point

Target Audience

Young families (professionals) looking for a **safe environment** to grow up **their children**, close to **good public schools**, but at the same time having the flexibility to chose places close to their **interests** (e.g. sport, restaurants etc.)



Criteria



Low
criminality



Schools with
good ratings



Good locations
around

Result

Recommendation (**pre-selected neighborhoods**) based on different data sources:

- Filter out criminality hot spots
- Detection of areas with top ranked schools
- After applying a pre-selection of neighborhoods based on criminality and public schools, group the neighborhoods in similar cluster according to amenities.

DATA ACQUISITION AND CLEANING



Source: Crime locations (community area number), type of crime and date (year) were scrapped from Chicago Crime Report Data

(<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>)

Information: This set contained a partial view from reported incidents of crime from 2001 to the present. Due to a large size only a part could be downloaded. More than 350k rows and 22 columns were the basis for plotting trend.

Additional 56k delicts from January to April 2020 were added for analysing the current situation.

Main remarks on data processing: The year 2001 was excluded due to low number of observations for the trend analysis and neighbourhood names were enhanced by using Chicago Public Schools data source.



Source: Chicago Public Schools and Ratings were obtained from Chicago Public Schools - Progress Report

(<https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t>)

Information: This set contained 79 columns and 566 rows, whereby each row contains the information for one school. For the analysis I have used only a limited set of metrics based on its relevance for the project and skip the remaining information. This set also contained a full view with neighbourhoods, ZIP codes and coordinates needed during the project.

Main remarks on data processing: Datatype needed to be transformed into floating for the k-mean algorithm and later on normalized. 513 schools were derived out of 566 rows having a full set of minimum information for the analysis.



Source: API request from Foursquare using the coordinates from the sources above and inserting it into

'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{ }&radius={}&limit={}' and requesting a json file.

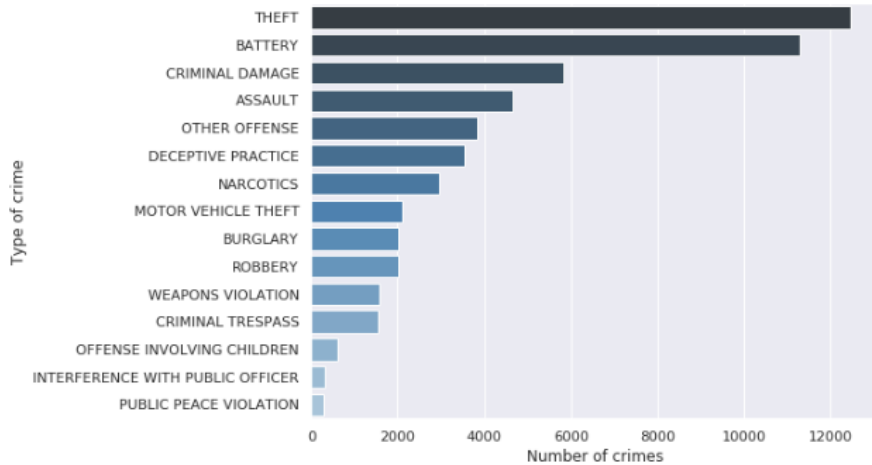
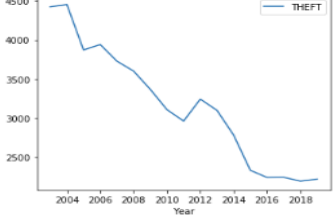
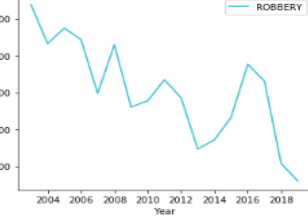
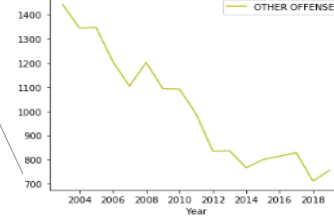
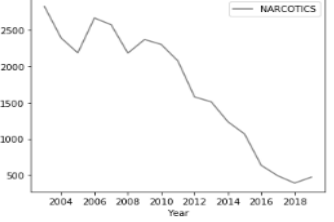
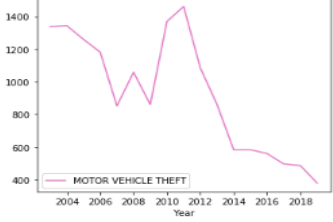
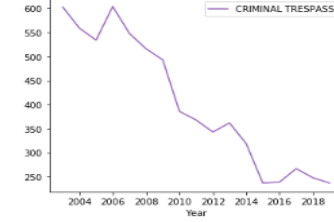
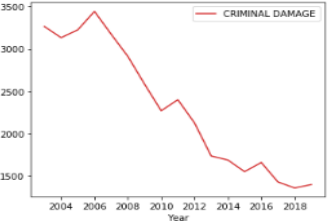
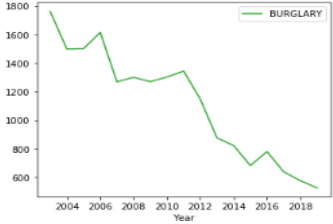
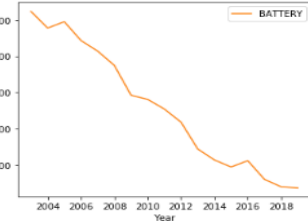
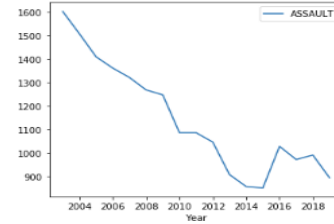
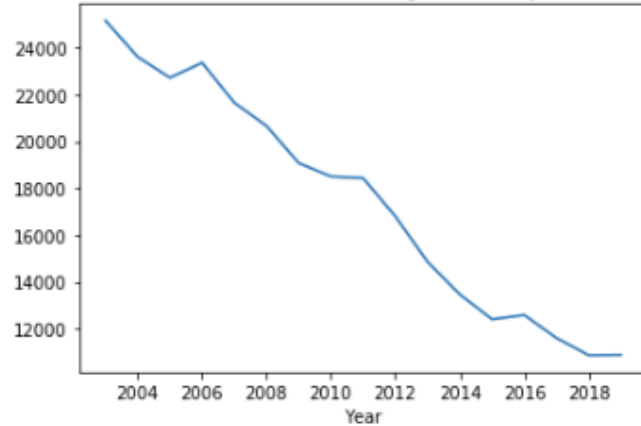
Information: Selected limit 100 and 500m radius. 1044 venues were extracted for Chicago with 200 unique categories.

Main remarks on data processing: Use of one hot encoding for processing categorical variables and convert them into numeric for running the k-mean algorithm.



INSIGHTS ABOUT THE CRIME SITUATION IN CHICAGO

Number of Crime in Chicago over the years

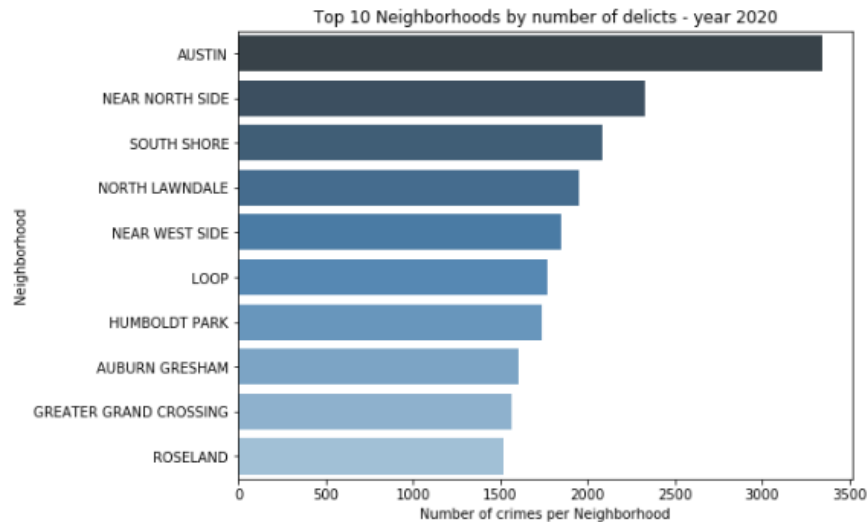


Remarks

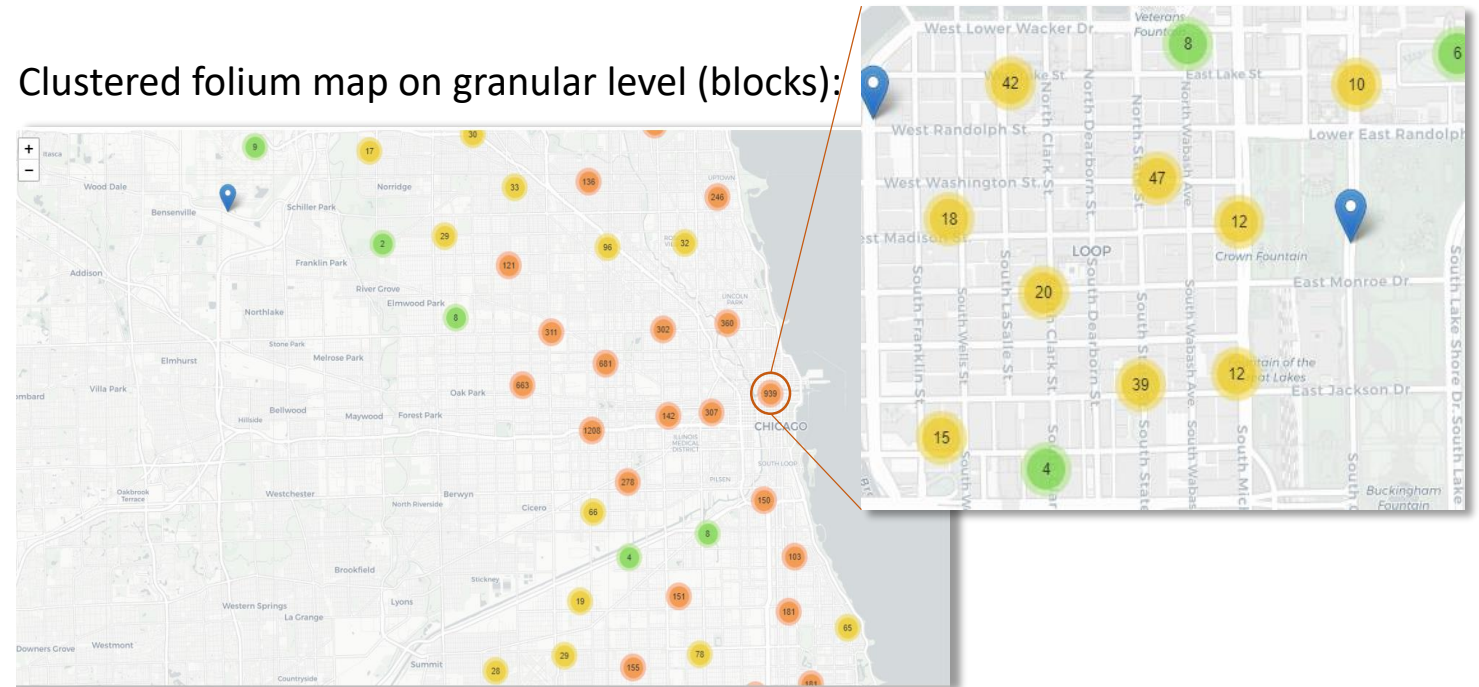
- Declining number of delicts in Chicago in the last years
- This positive declining trend is benefiting from especially lower number of delicts related to theft, narcotics, battery and criminal damage among others
- Theft and battery are the most common type of crimes followed by criminal damage and assault

HOT-SPOTS TO AVOID BASED ON NUMBER OF DELICTS IN 2020

High level ranking on criminal incidents:



Clustered folium map on granular level (blocks):

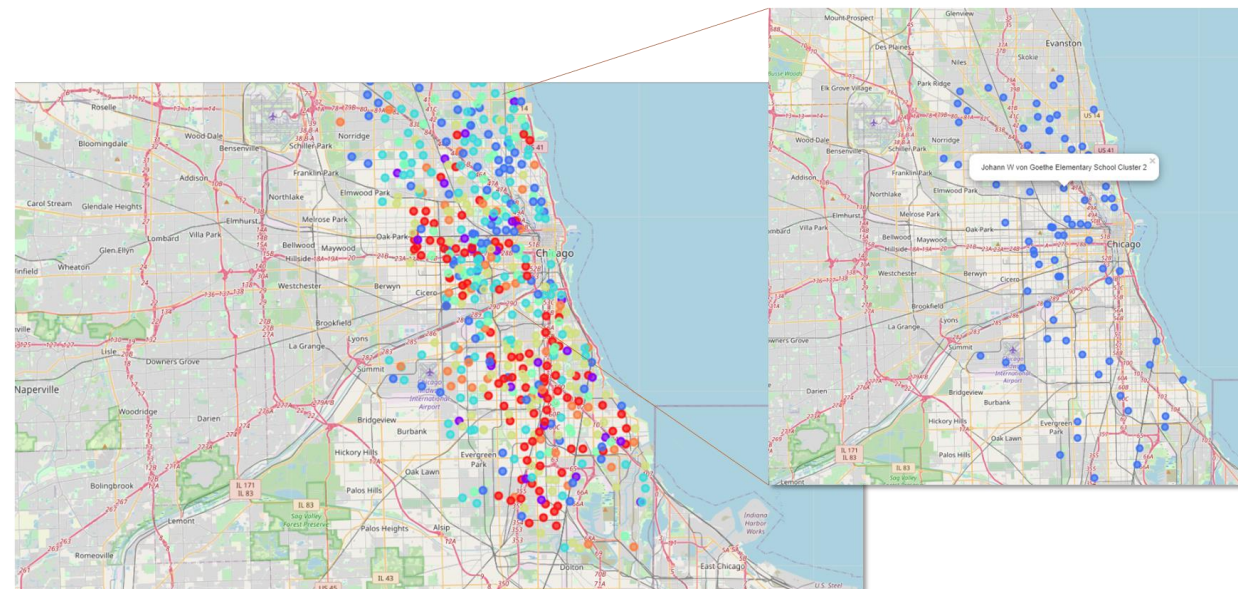


Remarks

- First recommendation: Young families should avoid the areas with higher number of incidents detected during the analysis like Austin, Near North Side, South Shore and North Lawndale.
- In case that the users are still considering moving into one of these neighbourhoods, please go into a more granular level and look the location by blocks.

PUBLIC SCHOOLS IN CHICAGO

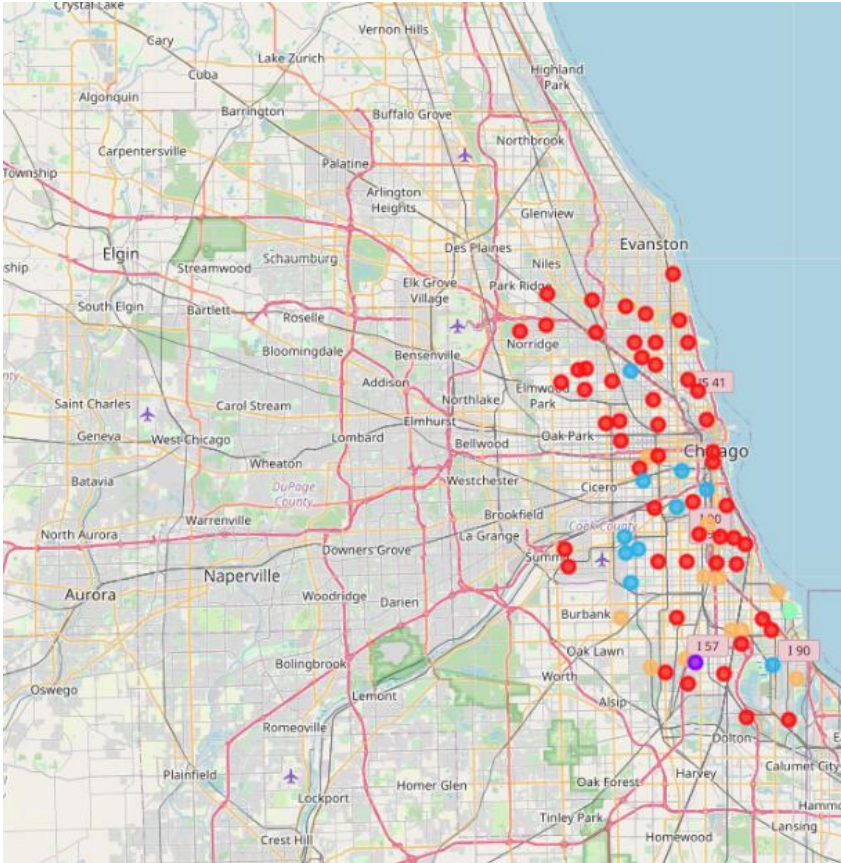
BUILDING CLUSTERS BASED ON RATING OF SELECTED METRICS



Remarks

- The best school cluster is the Cluster 3 having top ratings in all categories.
- There are 85 top public schools which are located in 45 different neighbourhoods whereby we can find already 5 top rated schools in Lake view, West Town .
- The new cluster with top schools excluding the areas with high criminality. Now the user have a total of 73 public in 39 neighbourhoods

EXPLORING THE VENUES OF THE DIFFERENT NEIGHBORHOODS



- **Cluster 0:** neighbourhoods with restaurants, coffee shops and stores. Fast food like hot dogs and donuts are very common in this area. Before filter: 53 neighbourhoods, after filter 28 potential neighbourhoods (e.g. **Lake View, West Town, North Center** and **Albany Park**).
- **Cluster 1:** After removing the locations with criminality and excluding the areas not included in the top ten, this cluster remains empty. Before the filters, this cluster had a basketball court which might be interesting for basketball fans.
- **Cluster 2:** mainly typical for the Mexican and Latin American restaurants but also parks and banks. Before filter: 10 neighbourhoods, after filter 6 potential neighbourhoods (e.g. **South Lawndale** and **Gage Park**).
- **Cluster 3:** After removing the locations with criminality and excluding the areas not included in the top ten, this cluster remains empty. Before the filters, this cluster show Yoga studios and event locations in South Chicago.
- **Cluster 4:** peculiarity of farmers markets but also Yoga studios. Before and after filter 5 potential neighbourhoods (e.g. **Chattam** and **Beverly**).

DISCUSSION AND CONCLUSION (FURTHER DIRECTIONS)

RECOMMENDATION

- Despite of the declining number of delicts in Chicago in the last years, I will recommend young families to avoid the areas detected during the analysis like **Austin, Near North Side, South Shore** and **North Lawndale**
- Places like **Lake View, West Town, North Center** and **Albany Park** were having even more than one top-rated public school and seem very attractive for parents looking for high quality public schools.
- These neighbourhoods were allocated to **Cluster 0** having a focus on restaurants, coffee shops and stores.

FURTHER DIRECTIONS

- Include cost of living as a new criteria
- Consider potential job locations and optimize the distance



THANK YOU