

Coursera Capstone: Battle of the Neighbourhoods

Presentation by Lyubomir Ahtapodov

Outline

- ▶ Introduction
- ▶ Data
- ▶ Exploratory Data Analysis
- ▶ K-Means Clustering
- ▶ Results & Discussion
- ▶ Conclusion

Introduction

- ▶ Sports are important for wellbeing and preventativ healthcare
- ▶ In big, hectic cities, the location of any commercialvenue is extremely important



- ▶ Question: What would be a good location for a sporting goods shop in Chicago, IL?

Data

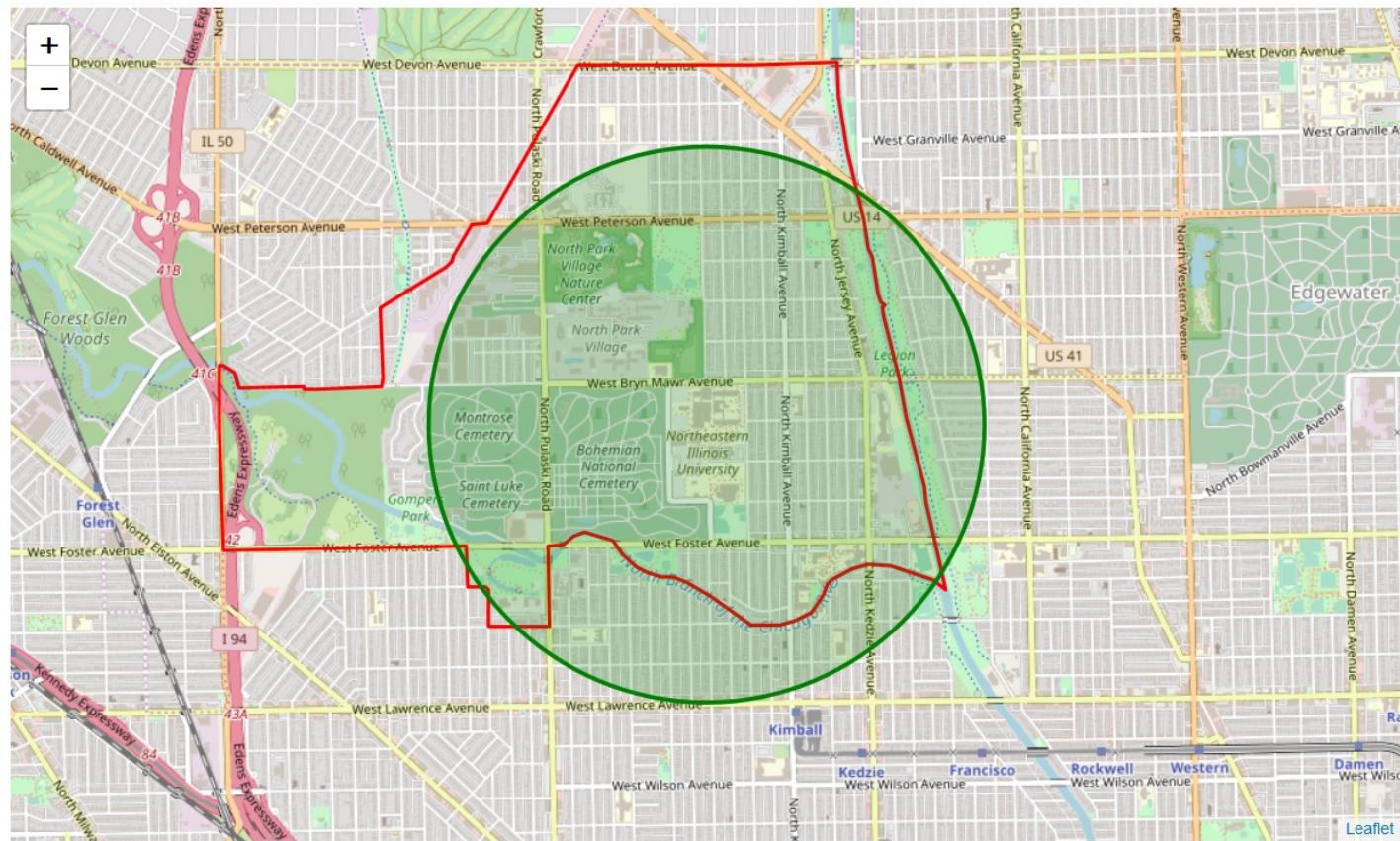
- ▶ The data was obtained from following two sources:
 - Geographical coordinates of the community area boundaries were retrieved from the City of Chicago's official site
 - Venue information was obtained from Foursquare



Exploratory Data Analysis

- From the CA boundary geographical data, the centre point of each CA and a best matching circular vicinity was computed.

Example for North Park community area



Exploratory Data Analysis (contd.)

- ▶ Next, the venue information for each CA was requested from Foursquare.
- ▶ Venues whose occurrence correlates with that of a sporting goods shop were shortlisted using pairwise correlation analysis

Correlation analysis

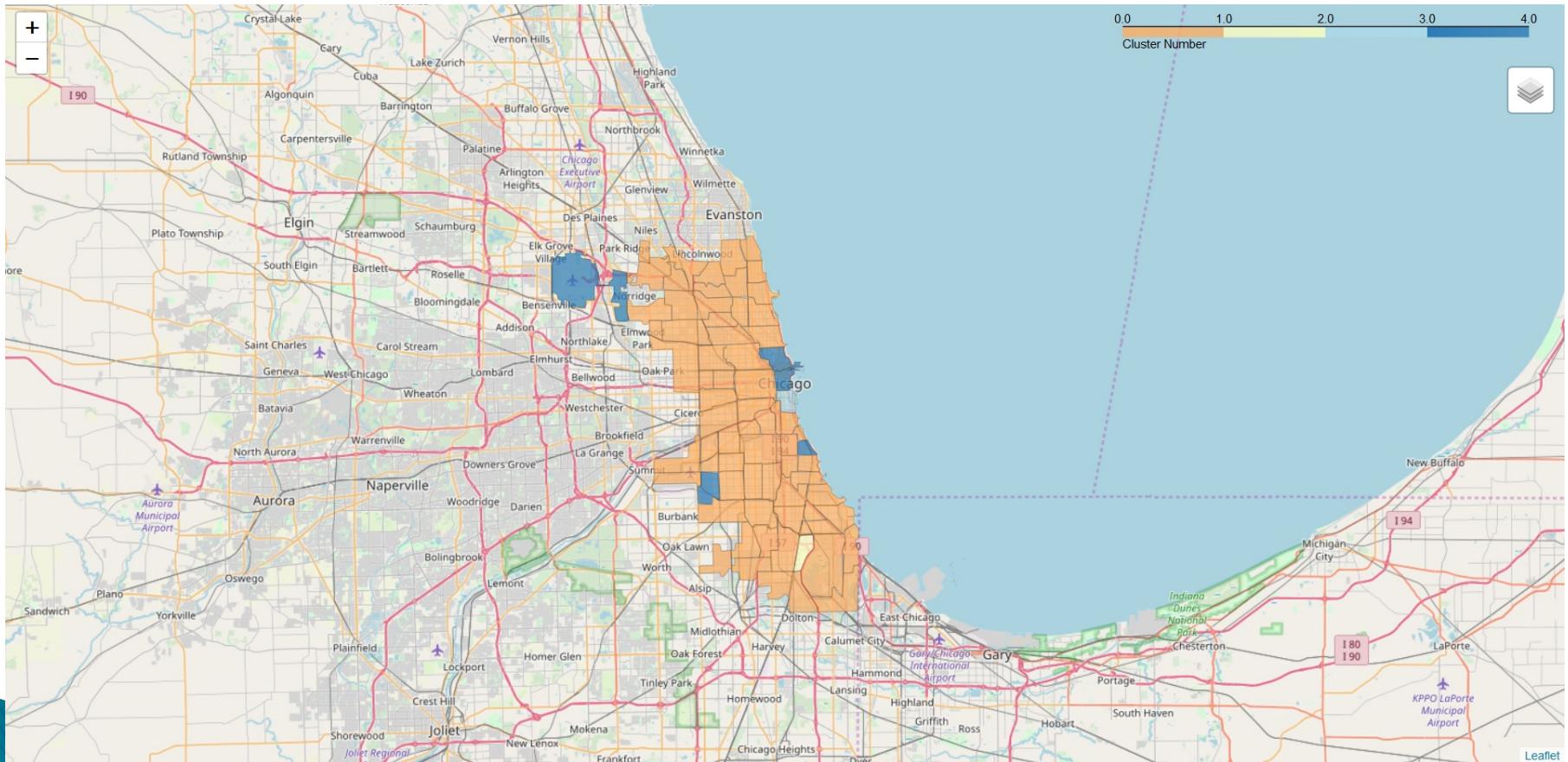
One-hot encoded venue data

CA Name	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Waterfront	Weight Loss Center	Whisky Bar	Wine Bar
0 ALBANY PARK	0.000000	0.000000	0.000000	0.012658	0.000000	0.000000	0.00	0.000000	0.012658	...	0.000000	0.000000	0.0	0.000000
1 ARCHER HEIGHTS	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000000
2 ARMOUR SQUARE	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.014493	...	0.000000	0.000000	0.0	0.000000
3 ASHBURN	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.033333	...	0.000000	0.016667	0.0	0.000000
4 AUBURN GRESHAM	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.0	0.014085
5 AUSTIN	0.030000	0.000000	0.000000	0.000000	0.010000	0.000000	0.00	0.000000	0.020000	...	0.000000	0.000000	0.0	0.000000
6 AVALON PARK	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000000
7 AVONDALE	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.0	0.010000

Venue Type	Correlation Coeff
128 Food Court	0.693692
131 Football Stadium	0.640939
148 Golf Course	0.614094
35 Big Box Store	0.411420
163 History Museum	0.409262
0 ATM	0.406466
294 Snack Place	0.395163
59 Cafeteria	0.390556
32 Bed & Breakfast	0.390556
72 Circus	0.390556

K-Means Clustering

- ▶ The shortlisted venues were used to cluster Chicago's CA's using the K-Means algorithm (with machine learning)



Results and Discussion

- ▶ Each cluster was scored: $Score = \sum_{venues} P_{avg}(venue) \cdot C_{corr}(venue)$
Cluster 0 (the largest) scored worst, all others scored equally
- ▶ Community areas and sporting goods shops were counted for each cluster

Cluster Labels	Sporting Goods Shop	CA Name	
0	0	9	70
1	1	1	1
2	2	1	1
3	3	3	5

- ▶ In Cluster 3, three CA's were found to have no sporting goods shop despite it being shared most favourable:
Kenwood, Loop, West Lawn

Conclusion

- ▶ Based on the present analysis, the CA's of Clusters 1, 2 and 3 are the most favourable locations for a sporting goods shop. These are: **Kenwood, Loop, Near North Side, Near South Side, O'Hare, Pullman, West Lawn**
- ▶ Of these three CA's deserve special attention as they lack a sporting goods shop at present: **Kenwood, Loop, West Lawn**