Applying Unsupervised Machine Learning Technology to Determine the Best Location for Opening a Steakhouse in Chicago

Author: Kui Shen

This project demonstrates a systematic approach on how to apply unsupervised machine learning technology (i.e., KMeans clustering methodology) to determine the best location for opening a steakhouse restaurant in Chicago.

Applied Data Science Capstone Project

5/5/2021

Executive Summary

Background of Chicago Steakhouse

- Chicago has been at the heart of high-quality American meats since 1870.
- Chicago steakhouse attract hundreds and thousands of locals and tourists around the global year-round.
- A new competitive modern steakhouse in Chicago can be both lucrative and culture-enriching.

Business Problem/Challenge

- For potential steakhouse owners, it is usually difficult to choose the best location for launching the business.

Purpose of the Project

- To determine and recommend the best location for opening a steakhouse in Chicago through the utilization of Unsupervised Machine Learning (UML) Clustering Methodology - KMeans.

Procedure of the Project

- Acquire Chicago community data:
  - Wikipedia – A Chicago community list along with their associated population, area, and population density, etc.
  - Python GeoPy Library – Geographical coordinates of the Chicago communities
  - Foursquare API – Venue information
- Data cleaning and pre-processing
- Segmentation based on Chicago community venue information
- Clustering the Chicago communities using KMeans
- Results refining and final answer developing
- Discussion, conclusion and future direction

1 - Data Acquisition, Cleaning, and Pre-processing

- The city of Chicago consists of 284 neighborhoods, which combined to form its 77 official communities.
- Neighborhood and community data are from websites:
  - https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago
  - https://en.wikipedia.org/wiki/Community_areas_in_Chicago
- Community data is more complete and ready for use than Neighborhood data. So it is selected for this project.
- Chicago community data read into Jupyter Notebook:

	Number[8]	Name[8]	2017 population[9]	Area (sq mi.)[10]	Area (km2)	2017 populationdensity (/sq mi.)	2017 populationdensity (/km2)
0	01	Rogers Park	55062	1.84	4.77	29925.00	11554.11
1	02	West Ridge	76215	3.53	9.14	21590.65	8336.20
2	03	Uptown	57973	2.32	6.01	24988.36	9648.06
3	04	Lincoln Square	41715	2.56	6.63	16294.92	6291.50
4	05	North Center	35789	2.05	5.31	17458.05	6740.59
5	06	Lake View	100470	3.12	8.08	32201.92	12433.23
6	07	Lincoln Park	67710	3.16	8.18	21427.22	8273.10
7	08	Near North Side	88893	2.74	7.10	32442.70	12526.20
8	09	Edison Park	11605	1.13	2.93	4235.40	1635.30
9	10	Norwood Park	37089	4.37	11.32	8487.19	3276.92

1 - Data Acquisition, Cleaning, and Pre-processing (Cont'd)

- **Chicago communities augmented with corresponding geographical coordinates using Python GeoPy libraries:**

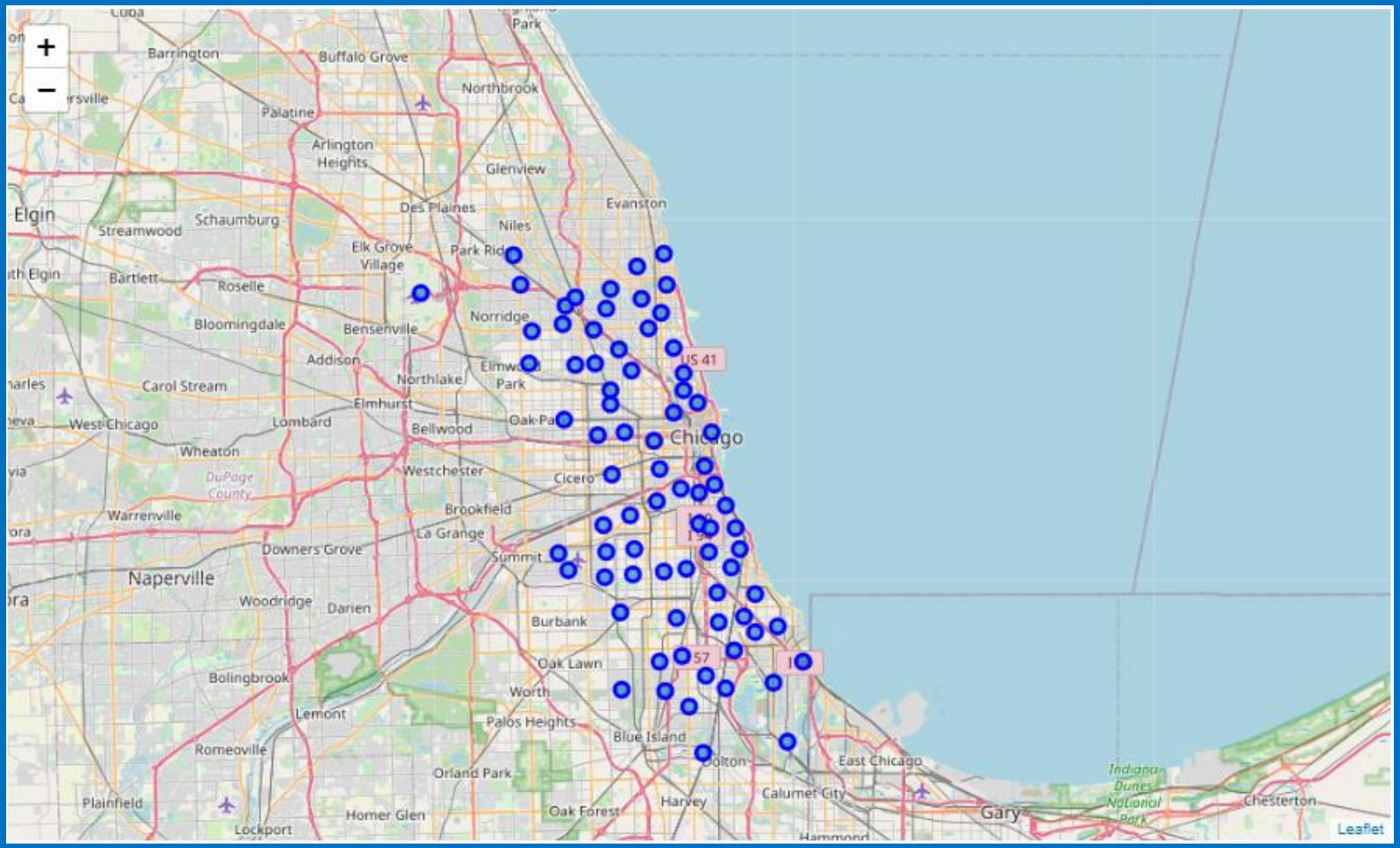
	Number[8]	Name[8]	2017 population[9]	Area (sq mi.) [10]	Area (km2)	2017 populationdensity (/sq mi.)	2017 populationdensity (/km2)	Latitude	Longitude
0	01	Rogers Park	55062	1.84	4.77	29925.00	11554.11	42.00897	-87.66619
1	02	West Ridge	76215	3.53	9.14	21590.65	8336.20	41.99948	-87.69266
2	03	Uptown	57973	2.32	6.01	24988.36	9648.06	41.96538	-87.66936
3	04	Lincoln Square	41715	2.56	6.63	16294.92	6291.50	41.97583	-87.68914
4	05	North Center	35789	2.05	5.31	17458.05	6740.59	41.95411	-87.68142
5	06	Lake View	100470	3.12	8.08	32201.92	12433.23	41.93982	-87.65682
6	07	Lincoln Park	67710	3.16	8.18	21427.22	8273.10	41.92184	-87.64744
7	08	Near North Side	88893	2.74	7.10	32442.70	12526.20	41.90034	-87.63433
8	09	Edison Park	11605	1.13	2.93	4235.40	1635.30	42.00789	-87.81399
9	10	Norwood Park	37089	4.37	11.32	8487.19	3276.92	41.98572	-87.80664

- **Chicago community data cleaned and pre-processed:**

	Community	Population	Area	Pop_Density	Latitude	Longitude
0	Rogers Park	55062	1.84	29925.00	42.00897	-87.66619
1	West Ridge	76215	3.53	21590.65	41.99948	-87.69266
2	Uptown	57973	2.32	24988.36	41.96538	-87.66936
3	Lincoln Square	41715	2.56	16294.92	41.97583	-87.68914
4	North Center	35789	2.05	17458.05	41.95411	-87.68142
5	Lake View	100470	3.12	32201.92	41.93982	-87.65682
6	Lincoln Park	67710	3.16	21427.22	41.92184	-87.64744
7	Near North Side	88893	2.74	32442.70	41.90034	-87.63433
8	Edison Park	11605	1.13	4235.40	42.00789	-87.81399
9	Norwood Park	37089	4.37	8487.19	41.98572	-87.80664

1 - Data Acquisition, Cleaning, and Pre-processing (Cont'd)

- **Chicago communities are plotted using Folium map:**



2 - Segmentation Based on Community Venue Information

○ Chicago communities segmentation – get all venues using Foursquare API: 1 = yes; 0 = no:

	Community	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	American Restaurant	Amphitheater	Antique Shop	Arcade	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Basketball Court	Beach	Bed & Breakfast	Beer Bar	Beer Garden	Beer Store	Big Box Store	Bike Rental / Bike Share	Bistro	Boat or Ferry	Bookstore	C
0	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Rogers Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

○ Chicago communities segmentation – calculate mean() of venues per community:

	Community	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	American Restaurant	Amphitheater	Antique Shop	Toy / Game Store	Track	Train Station	Transportation Service	Tunnel	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Albany Park	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.047619	0.000000	0.0
1	Archer Heights	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.050000	0.0	0.0	0.0	0.0	0.0	0.050000	0.000000	0.0
2	Armour Square	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.090909	0.0	0.0	.0	0.0	0.045455	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0
3	Ashburn	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0
4	Auburn Gresham	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0
5	Austin	0.076923	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.076923	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.076923	0.000000	0.0
6	Avalon Park	0.066667	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0
7	Avondale	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.032258	0.0	0.0	0.0	0.0	0.032258	0.0	0.0	0.0	0.0	0.0	0.000000	0.032258	0.0
8	Belmont Cragin	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0
9	Beverly Hills	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0

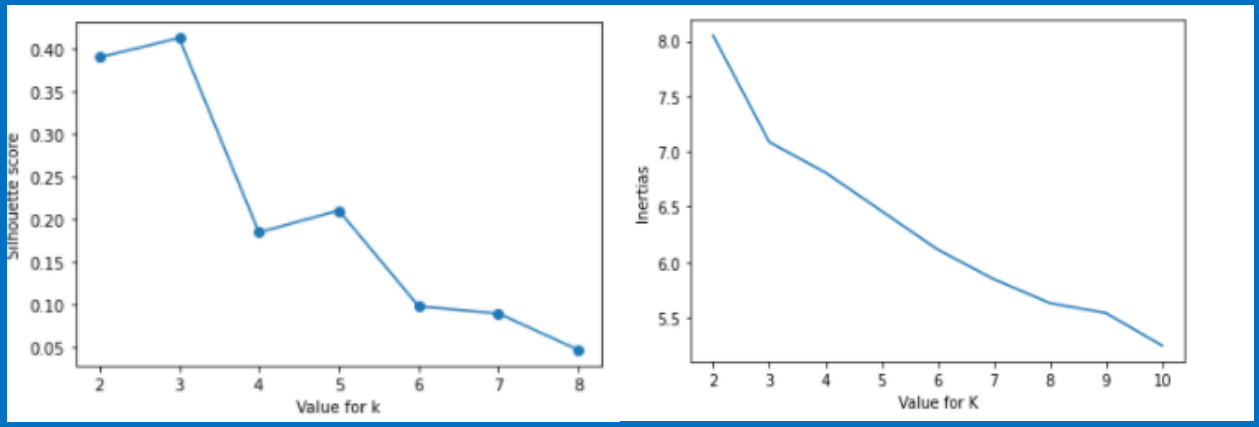
2 - Segmentation Based on Community Venue Information (Cont'd)

- **Chicago communities segmentation – calculate Top-10 venues per community:**

	Community	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albany Park	Mexican Restaurant	Karaoke Bar	Bank	Taco Place	Seafood Restaurant	Park	Dive Bar	Discount Store	Grocery Store	Korean Restaurant
1	Archer Heights	Mexican Restaurant	Mobile Phone Shop	Grocery Store	Pharmacy	Bar	Bank	Gas Station	Sandwich Place	Candy Store	Gym / Fitness Center
2	Armour Square	American Restaurant	Bar	Asian Restaurant	Business Service	Sandwich Place	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store	College Rec Center
3	Ashburn	Home Service	Light Rail Station	Automotive Shop	Pizza Place	Construction & Landscaping	Cosmetics Shop	Nightclub	Optical Shop	Office	Noodle House
4	Auburn Gresham	Park	Discount Store	Basketball Court	Convenience Store	ATM	Nightclub	Optical Shop	Office	Noodle House	Non-Profit
5	Austin	Bus Station	Park	Intersection	Liquor Store	Grocery Store	Train Station	Gym	Donut Shop	ATM	Wings Joint
6	Avalon Park	Pizza Place	Burger Joint	Fast Food Restaurant	Boutique	ATM	Grocery Store	Diner	Cajun / Creole Restaurant	Sandwich Place	Food
7	Avondale	Food Truck	Park	Chinese Restaurant	Gym	Road	Rental Car Location	Storage Facility	Sandwich Place	Supermarket	Bar
8	Belmont Cragin	Mexican Restaurant	Pizza Place	Grocery Store	Theater	Café	Laundromat	Field	Currency Exchange	Latin American Restaurant	Chinese Restaurant
9	Beverly Hills	Flower Shop	Coffee Shop	Park	Platform	Pizza Place	Dessert Shop	New American Restaurant	Office	Noodle House	Non-Profit

3 - Clustering Communities Using KMeans

- **Determine number of clusters for KMeans model - Silhouette Score and Elbow Method:**



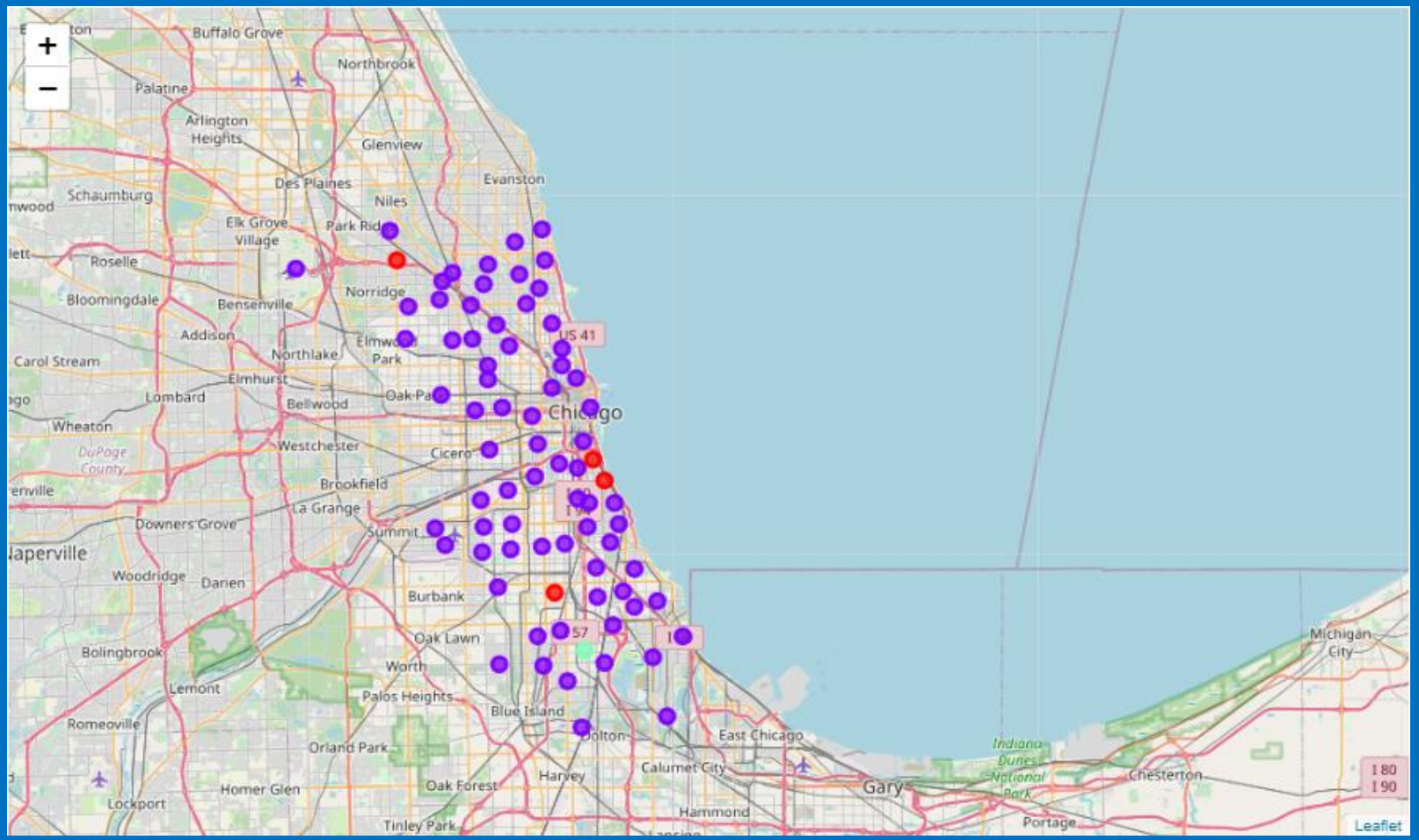
Silhouette Score plotting (left) showing k=3 is the best **Elbow Method plotting (right) showing k=3 is the best**

- **Determine number of clusters for KMeans model using Silhouette Score and Elbow Method:**

	Community	Population	Area	Pop_Density	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Co
13	Albany Park	51992	1.92	27079.17	41.96829	-87.72338	1	Mexican Restaurant	Karaoke Bar	Bank	Taco Place	Seafood Restaurant	Park
56	Archer Heights	13142	2.01	6538.31	41.81154	-87.72556	1	Mexican Restaurant	Mobile Phone Shop	Grocery Store	Pharmacy	Bar	Bank
33	Armour Square	13455	1.00	13455.00	41.83458	-87.63189	1	American Restaurant	Bar	Asian Restaurant	Business Service	Sandwich Place	Chine Resta
69	Ashburn	43792	4.86	9010.70	41.74785	-87.70995	1	Home Service	Light Rail Station	Automotive Shop	Pizza Place	Construction & Landscaping	Cosm Shop
70	Auburn Gresham	46278	3.77	12275.33	41.74319	-87.65504	0	Park	Discount Store	Basketball Court	Convenience Store	ATM	Night
24	Austin	95260	7.15	13323.08	41.88775	-87.76363	1	Bus Station	Park	Intersection	Liquor Store	Grocery Store	Train Statio
44	Avalon Park	9985	1.25	7988.00	41.74507	-87.58816	1	Pizza Place	Burger Joint	Fast Food Restaurant	Boutique	ATM	Groce Store
20	Avondale	37368	1.98	18872.73	41.93925	-87.71125	1	Food Truck	Park	Chinese Restaurant	Gym	Road	Renta Locat
18	Belmont Cragin	79910	3.91	20437.34	41.92802	-87.75384	1	Mexican Restaurant	Pizza Place	Grocery Store	Theater	Café	Launc
71	Beverly Hills	20822	3.18	6547.80	41.71201	-87.6709	1	Flower Shop	Coffee Shop	Park	Platform	Pizza Place	Desse Shop

3 - Clustering Communities Using KMeans (Cont'd)

- **Chicago community clustering result (3 clusters) plotted using Folium map:**



3 - Clustering the Communities Using KMeans (Cont'd)

- **Chicago Community Cluster #0 (showing a portion) – Recreation Paradise (not chosen):**

	Community	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
70	Auburn Gresham	0	Park	Discount Store	Basketball Court	Convenience Store	ATM	Nightclub	Optical Shop	Office	Noodle House	Non-Profit
34	Douglas	0	Park	Bus Station	Shopping Mall	Train Station	Insurance Office	Nightclub	Optical Shop	Office	Noodle House	Non-Profit
9	Norwood Park	0	Park	Gay Bar	Bus Station	Pakistani Restaurant	Other Great Outdoors	Optical Shop	Office	Noodle House	Non-Profit	Nightclub
35	Oakland	0	Park	Public Art	Track	ATM	New American Restaurant	Office	Noodle House	Non-Profit	Nightclub	National Park

- **Chicago Community Cluster #1 (showing a portion) – City Dwelling Wander Land (chosen):**

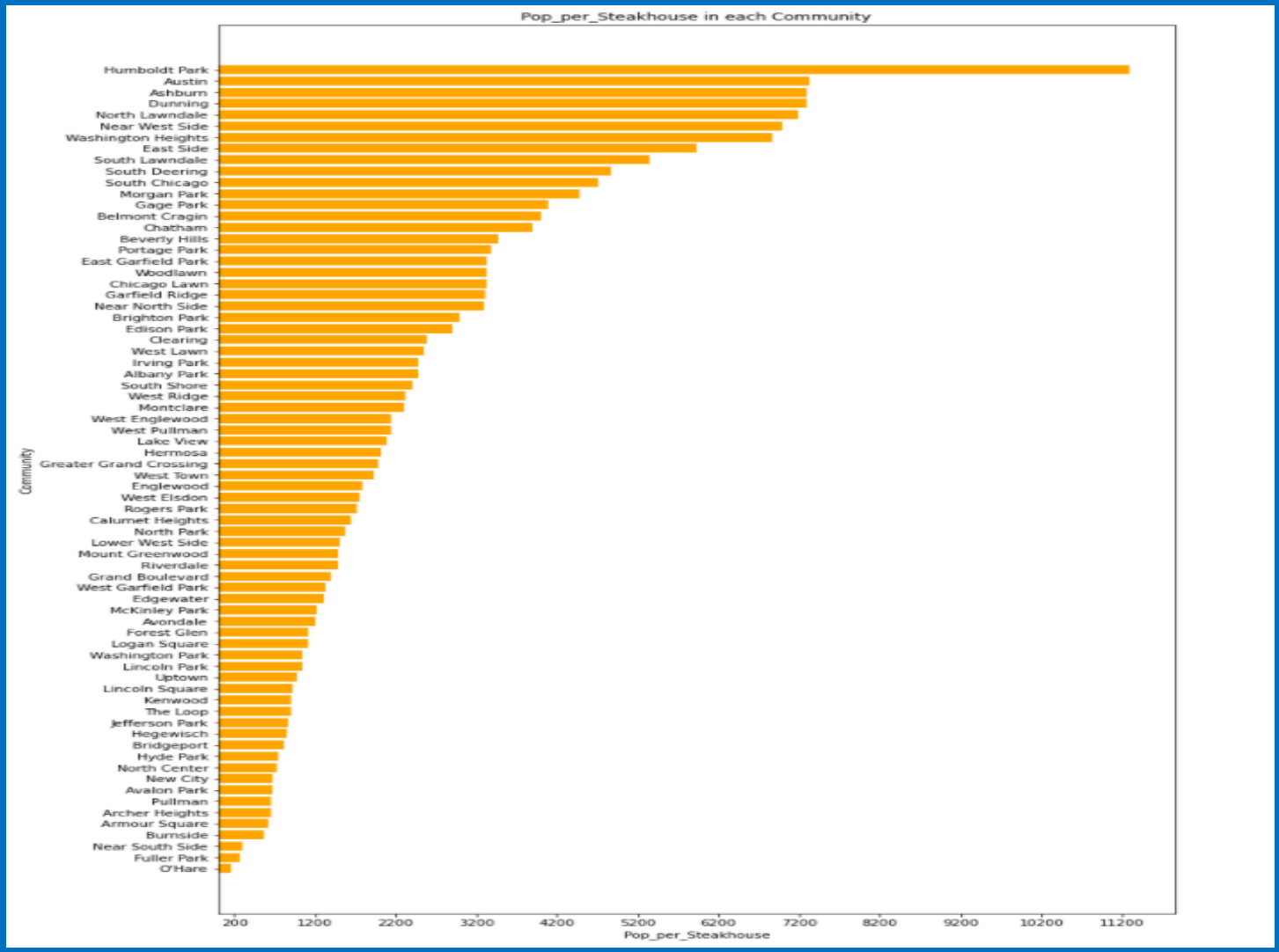
	Community	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
13	Albany Park	1	Mexican Restaurant	Karaoke Bar	Bank	Taco Place	Seafood Restaurant	Park	Dive Bar	Discount Store	Grocery Stor
56	Archer Heights	1	Mexican Restaurant	Mobile Phone Shop	Grocery Store	Pharmacy	Bar	Bank	Gas Station	Sandwich Place	Candy Store
33	Armour Square	1	American Restaurant	Bar	Asian Restaurant	Business Service	Sandwich Place	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store
69	Ashburn	1	Home Service	Light Rail Station	Automotive Shop	Pizza Place	Construction & Landscaping	Cosmetics Shop	Nightclub	Optical Shop	Office
24	Austin	1	Bus Station	Park	Intersection	Liquor Store	Grocery Store	Train Station	Gym	Donut Shop	ATM
44	Avalon Park	1	Pizza Place	Burger Joint	Fast Food Restaurant	Boutique	ATM	Grocery Store	Diner	Cajun / Creole Restaurant	Sandwich Place
20	Avondale	1	Food Truck	Park	Chinese Restaurant	Gym	Road	Rental Car Location	Storage Facility	Sandwich Place	Supermarket
18	Belmont Cragin	1	Mexican Restaurant	Pizza Place	Grocery Store	Theater	Café	Laundromat	Field	Currency Exchange	Latin American Restaurant
71	Beverly Hills	1	Flower Shop	Coffee Shop	Park	Platform	Pizza Place	Dessert Shop	New American Restaurant	Office	Noodle House

- **Chicago Community Cluster #2 – Wild Wild West (not chosen):**

	Community	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
48	Roseland	2	Intersection	ATM	Other Great Outdoors	Optical Shop	Office	Noodle House	Non-Profit	Nightclub	New American Restaurant	Pakistani Restaurant

4 - Results Refining and Final Answer Developing

- **Chicago communities cluster 1 – population per steakhouse on bar chart:**



4 - Results Refining and Final Answer Developing (Cont'd)

- **Chicago community segmentation, clustering, and refining results - 4 final communities:**

	Community	Steakhouse	Population	Cluster Labels	Pop_per_Steakhouse
10	Bridgeport	41	33637	1	820
33	Hyde Park	36	26827	1	745
40	Logan Square	66	73046	1	1106
59	Rogers Park	32	55062	1	1720

- **Rogers Park has the highest population over # of steakhouse ratio (1,720) among the four communities:**
  - **It has the least number of steakhouse;**
  - **It has the second largest population.**
- **For steakhouse business, these facts combined could indicate:**
  - **Least competition in Rogers Park;**
  - **Largest growth potential in Rogers Park.**

5 - Discussion, Conclusion, and Future Direction

- According to this analysis, Rogers Park community is the best location for opening a steakhouse in Chicago. It is recommended to the business problem owner.
- Python (sklearn) KMeans as a UML clustering model is relatively simple and effective, guaranteed to converge, and reusable. It also has limitations such as hard to determine the # of clusters, still requires considerable efforts to deliver an answer after clustering is done.
- In reality, one needs to consider many more factors in determining where to open a steakhouse – local demographics, religion, tradition, regulation, etc.
- Technically, one needs to evaluate other options and possibilities before diving into a project only because he/she knows one particular tool. It is always preferred to develop a holistic view and strive for the best solution possible.


THANK YOU