Income and Venue Data Analysis of Chicago

Ting Gong

Jun 13, 2021

OUTLINE

- Introduction
- Data used in the analysis
- Methodology
- Results
- Discussion
- Conclusion
- Reference

INTRODUCTION

Chicago is the third largest city in the United States with its nearly 9 million population. The entire city includes 77 community areas and are highly demographically diverse.

I like the community that I live in because of the convenience and the similar lifestyle of other community members. For example, there are many restaurants in the neighborhood, and the grocery stores are in five-minute walking distance. Park and recreational areas also are just a block away. Community members are well-educated and have decent jobs and are at a relatively high-income level.

However, I may need to move to other areas due to career development. I am curious about the similarity of different community areas in Chicago and how income levels distribute across Chicagoland. By knowing the similar communities as my current one, I may better decide the possible communities for me to live in the future.

IBM Developer

SKILLS NETWORK

Data used in the analysis

- First, I used Requests and BeautifulSoup to webscrape the community areas in Chicago from Wikipedia page. It contains the name of total 77 community areas in Chicago.

- Second, I used community name to get the geographic coordinates of each community. Next, I used geographic coordinates to get venue information for each community by accessing Foursquare API.

- Third, I downloaded "Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012" from City of Chicago data portal. The file includes Per Capita Income from 2008 to 2012 of each community. I used this data to group communities into 4 income levels.

- Last, I downloaded "chicago-community-areas.geojson" from kaggle.com. The geojson file contains the geocodes of the boundaries of each community in Chicago. I used the geojson file to create a choropleth map that color-codes the income levels of community.

IBM Developer

SKILLS NETWORK

METHODOLOGY

1. Obtain community information and locations

- **Created a main data frame that contains the community name and the geographic coordinates.**
  - First, webscrape "List of community areas" table from Wikipedia page[1], then convert it to a Dataframe.
  - Second, use **Geopy** liabrary to get the geographic coordinates of each community.
  - Third, clean and process data and concatenate all information into one dataframe

	Community	Latitude	Longitude
0	Rogers Park	42.009037	-87.676849
1	West Ridge	42.003548	-87.696243
2	Uptown	41.966630	-87.655546
3	Lincoln Square	41.968512	-87.688653
4	North Center	41.956107	-87.679160

METHODOLOGY
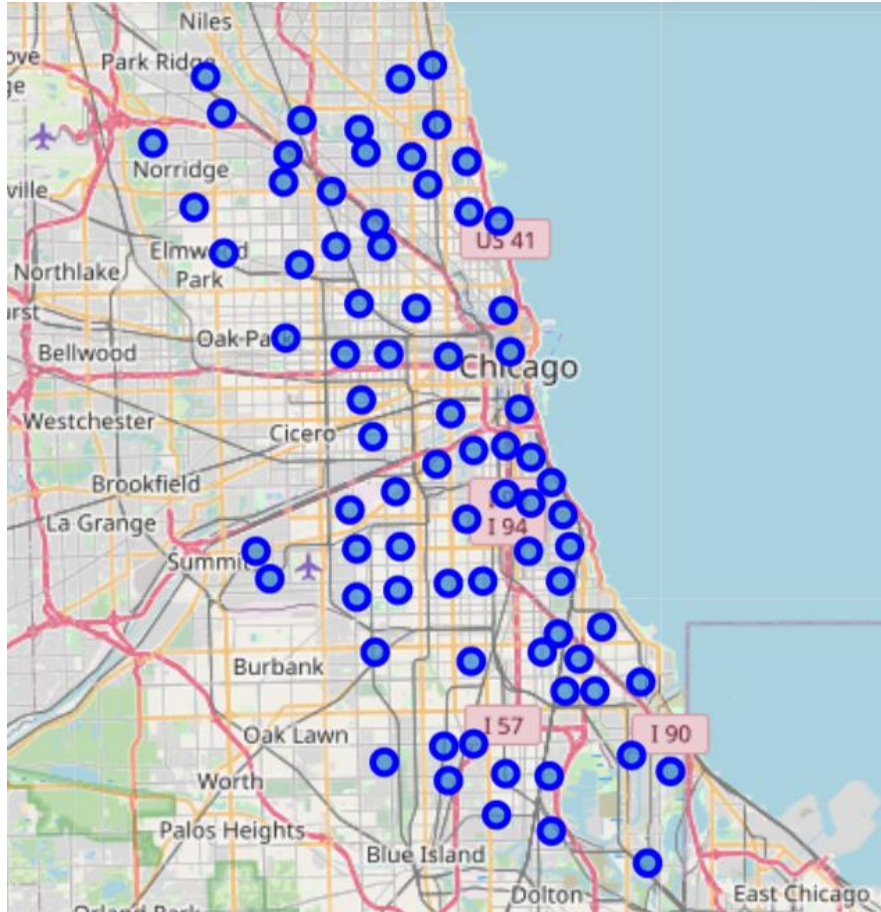
- Use **Folium** to create a map that visualize the location of communities

METHODOLOGY

2. Explore venues in each community

- **Utilize Foursquare API to explore communities**
  - **Obtain credential to use Foursquare API [2]**
  - **Set search radius**

  As we know, the area of Chicago is 234 square mile (606 square km). The city is divided into 77 communities. On average, the area of each community is 3 square mile (7.8 square km). Thus, the radius of each community is approx. 0.98 mile (1.58 km). To maximize the coverage but also reduce overlap, I set the radius to 1500 meters.

IBM **Developer**

SKILLS NETWORK

METHODOLOGY

After removing duplicates, Foursquare API returned 4279 unique venues.

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rogers Park	42.009037	-87.676849	Bark Place	42.010080	-87.675223	Pet Store
1	Rogers Park	42.009037	-87.676849	El Famous Burrito	42.010421	-87.674204	Mexican Restaurant
2	Rogers Park	42.009037	-87.676849	Mind Crusher Tattoo	42.003801	-87.672525	Tattoo Parlor
3	Rogers Park	42.009037	-87.676849	Taqueria & Restaurant Cd. Hidalgo	42.011634	-87.674484	Mexican Restaurant
4	Rogers Park	42.009037	-87.676849	Morse Fresh Market	42.008087	-87.667041	Grocery Store
...	...	...	...	...	...	...	...
4274	Edgewater	41.982814	-87.673951	Pizza Hut Express	41.991173	-87.682488	Pizza Place
4275	Edgewater	41.982814	-87.673951	Baskin-Robbins	41.995824	-87.670531	Ice Cream Shop
4276	Edgewater	41.982814	-87.673951	Claddagh Ring	41.975941	-87.687097	Bar
4277	Edgewater	41.982814	-87.673951	Starbucks	41.990917	-87.683370	Coffee Shop
4278	Edgewater	41.982814	-87.673951	LensCrafters	41.983662	-87.657648	Optical Shop

4279 rows × 7 columns

IBM Developer

SKILLS NETWORK

METHODOLOGY

I grouped by the community to examine the number of venues in each community and 338 unique venue categories are returned.

Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Albany Park	53	53	53	53	53	53
Archer Heights	64	64	64	64	64	64
Armour Square	80	80	80	80	80	80
Ashburn	54	54	54	54	54	54
Auburn Gresham	64	64	64	64	64	64
...	...	...	...	...	...	...
West Lawn	37	37	37	37	37	37
West Pullman	23	23	23	23	23	23
West Ridge	57	57	57	57	57	57
West Town	100	100	100	100	100	100
Woodlawn	40	40	40	40	40	40

77 rows × 6 columns

METHODOLOGY

3. Analyze venues in each community

- **Discover the top venues in each community**

  - Set dummy values to each unique venue category

  - Group rows by community and by taking the mean of the frequency of occurrence of each category

 - Sort the number of top venues in descending order.

IBM Developer

SKILLS NETWORK

METHODOLOGY

	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	Middle Eastern Restaurant	Grocery Store	Donut Shop	Sandwich Place	Pizza Place	Mexican Restaurant	Convenience Store	Discount Store	Diner	Chinese Restaurant
1	Archer Heights	Mobile Phone Shop	Sandwich Place	Bank	Mexican Restaurant	Fast Food Restaurant	Seafood Restaurant	Gas Station	Shoe Store	Donut Shop	Video Game Store
2	Armour Square	Baseball Stadium	Sandwich Place	Chinese Restaurant	Pizza Place	Train Station	Bar	Shipping Store	Grocery Store	Coffee Shop	Mobile Phone Shop
3	Ashburn	Park	Fast Food Restaurant	Pizza Place	Sandwich Place	Bus Station	Donut Shop	Mexican Restaurant	Intersection	American Restaurant	Seafood Restaurant
4	Auburn Gresham	Pharmacy	Lounge	Fast Food Restaurant	Discount Store	Southern / Soul Food Restaurant	American Restaurant	Sandwich Place	Grocery Store	Seafood Restaurant	Cosmetics Shop

IBM Developer

**SKILLS NETWORK**

METHODOLOGY

4. Cluster Neighborhoods using K-Means

- Find out the optimal K using Elbow Method
  - I use correlation because the Canberra function gives a clearer view of elbow break point.

- Merge cluster labels with the top 10 venues data frame

- Examine top venues in each cluster

The Elbow Method showing the optimal k

IBM Developer

SKILLS NETWORK
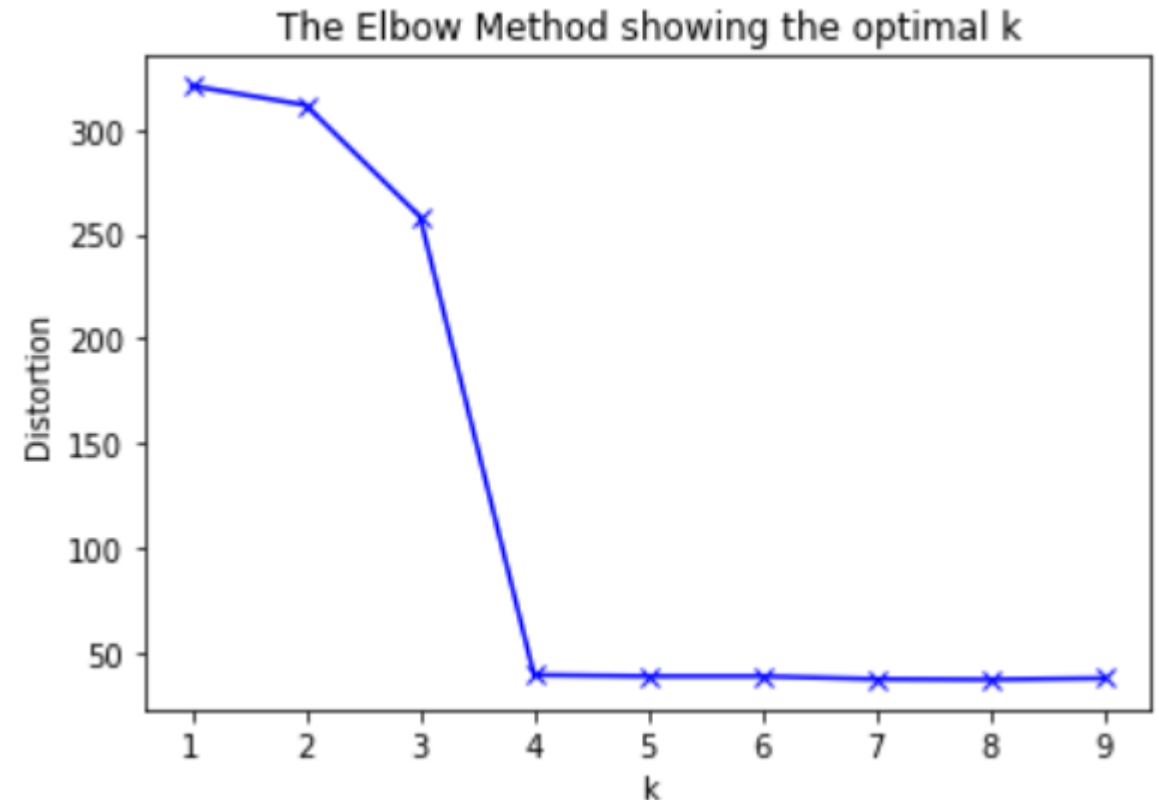
METHODOLOGY
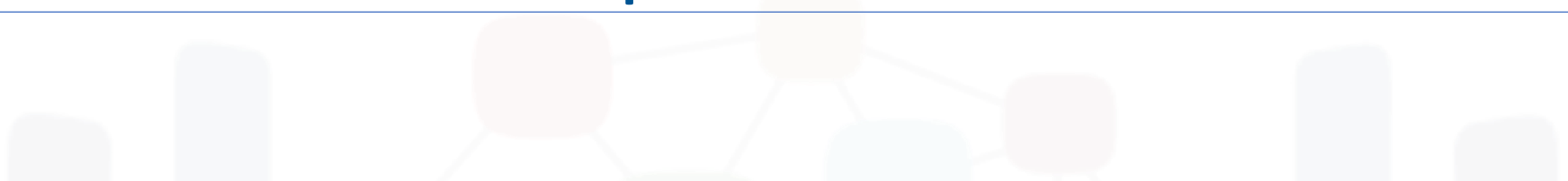
- Visualize the clusters on the map
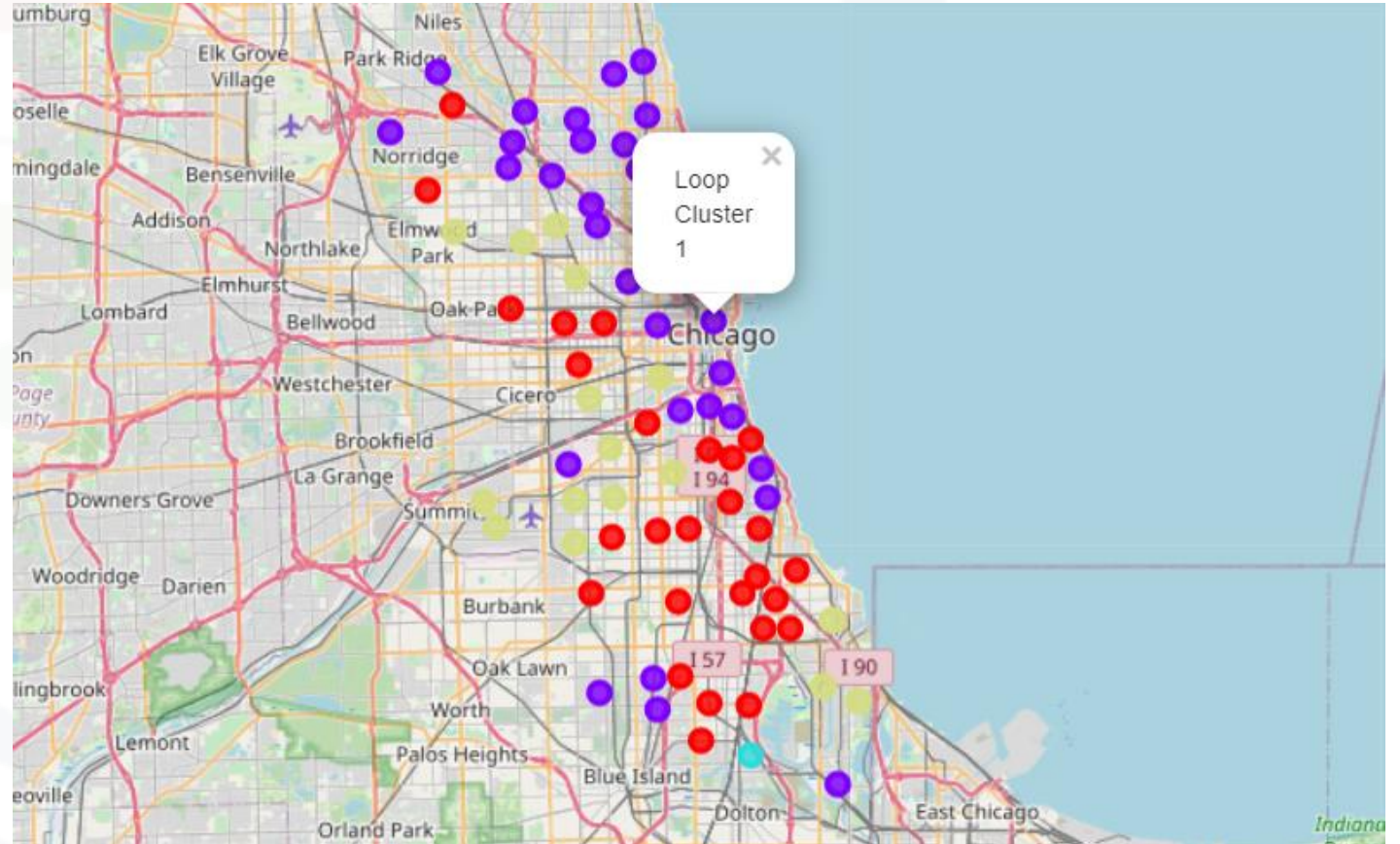
METHODOLOGY – Merge cluster labels with Top 10

Community	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 Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Rogers Park	42.009037	-87.676849	1	Sandwich Place	Pizza Place	Mexican Restaurant	Park	African Restaurant	Thai Restaurant	Café	Diner	Grocery Store	Asia Restaura
West Ridge	42.003548	-87.696243	1	Indian Restaurant	Grocery Store	Pakistani Restaurant	Pharmacy	Pizza Place	Convenience Store	Korean Restaurant	Donut Shop	Sandwich Place	Middl Easter Restaura
Uptown	41.966630	-87.655546	1	Vietnamese Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Theater	Chinese Restaurant	Breakfast Spot	Beach	Grocery Store	Discou Stor
Lincoln Square	41.968512	-87.688653	1	Bar	Pizza Place	Park	Gym	Brewery	Gourmet Shop	American Restaurant	Sandwich Place	Coffee Shop	Caf
North Center	41.956107	-87.679160	1	Bar	Dive Bar	Coffee Shop	Salon / Barbershop	Yoga Studio	Café	Brewery	Video Store	Mobile Phone Shop	Tha Restaura

IBM Developer

SKILLS NETWORK

METHODOLOGY – Top venues in each cluster

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 Mos...
0	ParkFast Food RestaurantSandwich PlaceFast Foo...	ATMClothing StoreDiscount StoreFoodFoodParkPar...	Sandwich PlaceSalon / BarbershopFast Food Rest...	BarPizza PlaceFried Chicken JointSandwich Plac...	Fried Chicken JointIce Cream ShopGrocery Store...	Sushi RestaurantDiscount StoreATMLiquor StoreC...	Sports F Coursel...
1	Sandwich PlaceIndian RestaurantVietnamese Rest...	Pizza PlaceGrocery StoreCoffee ShopPizza Place...	Mexican RestaurantPakistani RestaurantPizza Pl...	ParkPharmacySushi RestaurantGymSalon / Barbers...	African RestaurantPizza PlaceTheaterBreweryYog...	Thai RestaurantConvenience StoreChinese Restau...	C Restaura SpotAm...
2	Grocery Store	Baseball Field	Train Station	Pier	Park	Martial Arts School	Y
3	Grocery StoreMexican RestaurantGrocery StoreMe...	Fast Food RestaurantFast Food RestaurantMexica...	Mexican RestaurantPizza PlacePharmacyGrocery S...	Mobile Phone ShopDonut ShopVideo StoreSandwich...	BakeryPharmacyDinerDonut ShopTaco PlaceArt Gal...	Pizza PlaceDiscount StorePizza PlaceVideo Stor...	Pa Plac Res

METHODOLOGY

- By observing the top venues in each cluster, now we can add label names of each cluster.

  - Cluster 0 : "Fast Food Venues"

  - Cluster 1 : "Multiple Social Venues"

  - Cluster 2 : "Pier & Park"

  - Cluster 3 : "Mexican Venues"

- I also created a label indicating the top 3 venues of each community for future use on map.

METHODOLOGY

Cluster Names

	Clusters	Labels
0	0	Fast Food Venues
1	1	Multiple Social Venues
2	2	Pier & Park
3	3	Mexican Venues

Top 3 venues in community

	Community	Join
0	Albany Park	4 Donut Shop, 3 Grocery Store, 3 Middle Easter...
1	Archer Heights	5 Mobile Phone Shop, 3 Bank, 3 Mexican Restaurant
2	Armour Square	5 Baseball Stadium, 5 Chinese Restaurant, 5 Pi...
3	Ashburn	5 Park, 4 Fast Food Restaurant, 4 Pizza Place
4	Auburn Gresham	5 Pharmacy, 4 Discount Store, 4 Fast Food Rest...

IBM Developer

SKILLS NETWORK

METHODOLOGY

5. Obtain Per Capita Income of communities in Chicago

- Download "Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012"[3] from City of Chicago data portal.

- Read it into Pandas data frame.

	Community	Per Capita Income
0	Rogers Park	23939
1	West Ridge	23040
2	Uptown	35787
3	Lincoln Square	37524
4	North Center	57123

IBM Developer

SKILLS NETWORK

METHODOLOGY

- Use histogram to visualize the distribution of income

- Group communities into four levels of income.
  - Below $30k, Low Income
  - $30k-$50k, Medium-Low Income
  - $50k-$70k, Medium Income
  - Above $70k, High Income


Average Income in Range
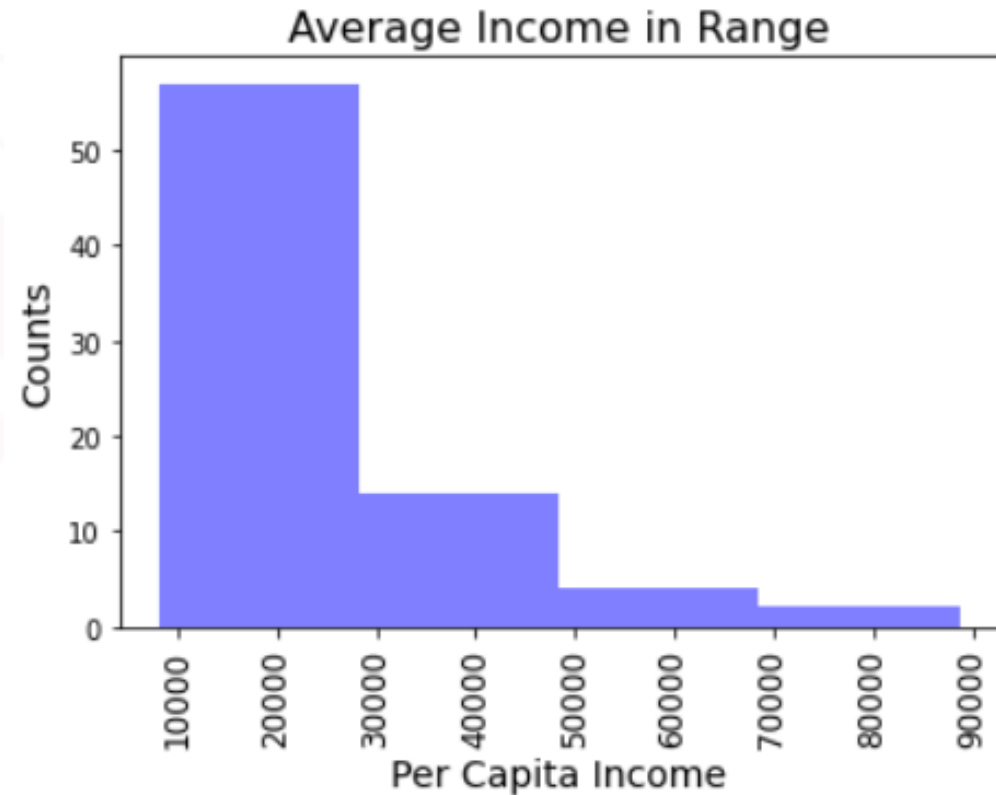
METHODOLOGY

Finally, merge all new variables with main clustered data frame.

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	Join	Labels	Income Level
1	Sandwich Place	Pizza Place	Mexican Restaurant	Park	African Restaurant	Thai Restaurant	Café	Diner	Grocery Store	Asian Restaurant	7 Sandwich Place, 5 Mexican Restaurant, 5 Park	Multiple Social Venues	Low Income
1	Indian Restaurant	Grocery Store	Pakistani Restaurant	Pharmacy	Pizza Place	Convenience Store	Korean Restaurant	Donut Shop	Sandwich Place	Middle Eastern Restaurant	15 Indian Restaurant, 4 Grocery Store, 3 Pakis...	Multiple Social Venues	Low Income
1	Vietnamese Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Theater	Chinese Restaurant	Breakfast Spot	Beach	Grocery Store	Discount Store	8 Vietnamese Restaurant, 7 Coffee Shop, 5 Pizz...	Multiple Social Venues	Medium Low Income
1	Bar	Pizza Place	Park	Gym	Brewery	Gourmet Shop	American Restaurant	Sandwich Place	Coffee Shop	Café	7 Bar, 4 Brewery, 4 Gym	Multiple Social Venues	Medium Low Income
1	Bar	Dive Bar	Coffee Shop	Salon / Barbershop	Yoga Studio	Café	Brewery	Video Store	Mobile Phone Shop	Thai Restaurant	4 Bar, 3 Coffee Shop, 3 Dive Bar	Multiple Social Venues	Medium Income

IBM Developer

SKILLS NETWORK

RESULTS

Loop / Multiple Social Venues-Medium Income / 8 Hotel, 4 Bar, 4 Park
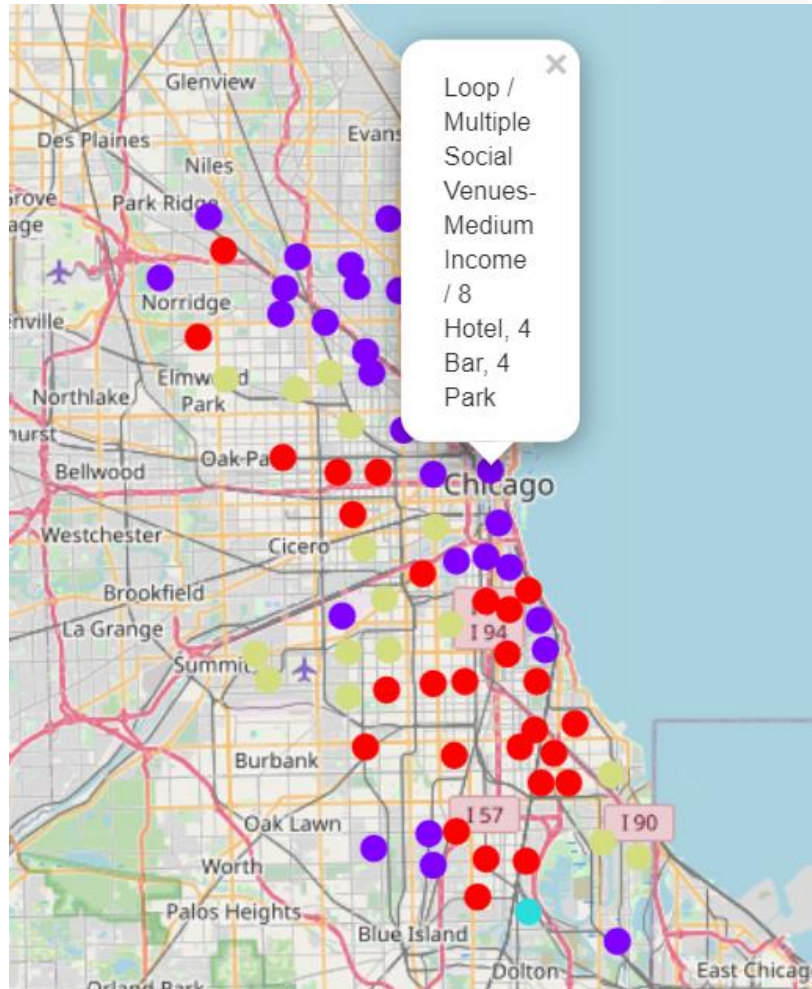
- Visualize the clusters with income levels on the map.
- The label contains:
  - Community name
  - Cluster name
  - Income level
  - Top 3 venue category in the community

RESULTS

- Create a choropleth map indicating income levels
  - Use "chicago-community-areas.geojson"[4] as the map layer
  - The color from light to dark represents the income levels from low to high.

IBM **Developer**

SKILLS NETWORK
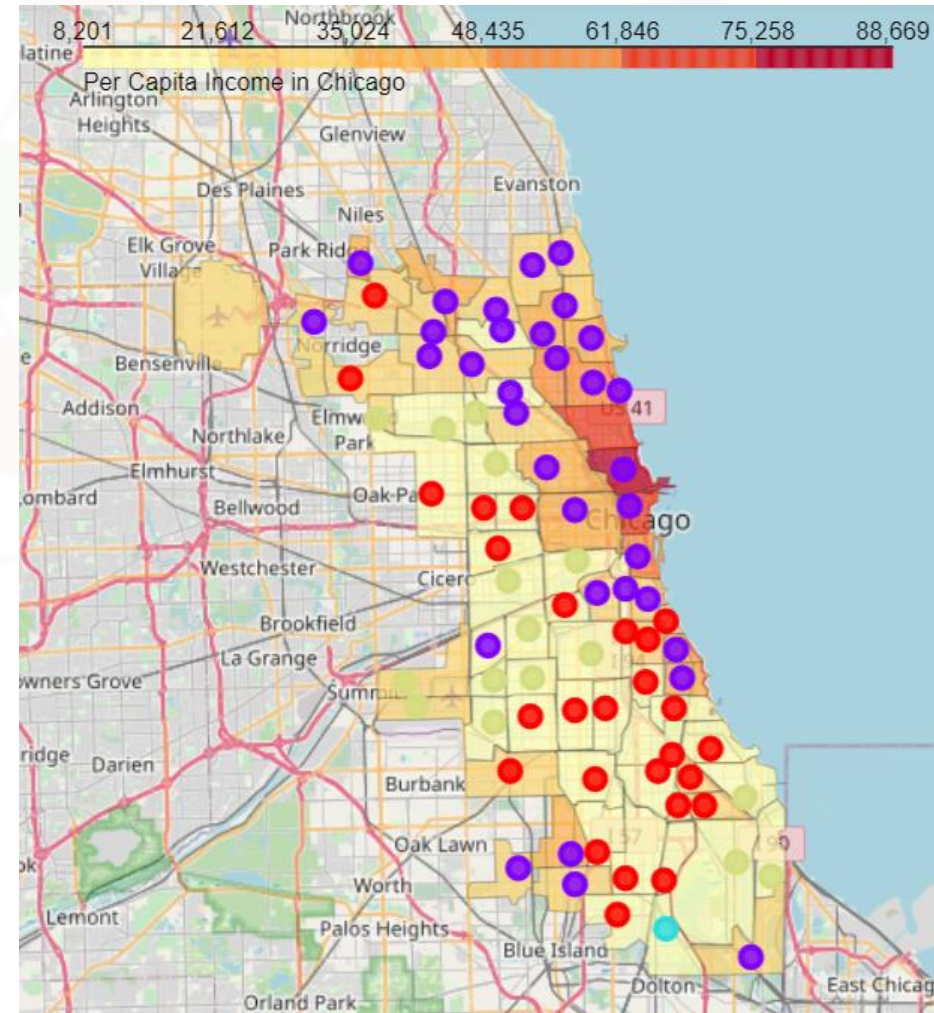
DISCUSSION

- Chicago is a large city with residents with diverse background. The analysis only used venue information to cluster communities regardless the demographic background. On the other hand, Foursquare API, as default, only returned 100 venues within the radius of each coordinates, which excluded many venues in seven communities including Loop, Near North Side, Near South Side, Near West Side, Rogers Park, Uptown, and West Town. Some of these communities are highly influenced by their demographic and economical factors. The limit of data may skew the accuracy of clustering.

DISCUSSION

- Per Capita Income from City of Chicago represents the income levels from 2008 – 2012. Income has increased nationwide in the past years. Per Capita Income for Chicago in 2019 was $40,144, which has increased $6,664 (19.5%) from $33,590 in 2012[5]. The data used in the analysis may not accurately present the current income level in Chicago. However, the distribution of income in Chicago remains reliable.

- To get more accurate results, further analysis maybe needed to take demographics, house value, and crime into consideration.

CONCLUSION

- Communities with multiple social venues are mainly located on the northside of Chicago.

- Fast food venues are mainly located on the southside of Chicago.

- Mexican venues are on the west of Chicago.

- Communities with medium to high income are on the northeast of Chicago along the lake. The income decreases from east to west.

- Fast food venues and Mexican venues are in the low-income communities.

- To conclude, I will remain in the communities with multiple social venues. In addition, the house price and crime data will be further considered.

Reference

[1] Wikipedia: List of community areas

https://en.wikipedia.org/wiki/Community_areas_in_Chicago

[2]Foursquare API

https://developer.foursquare.com

[3] "Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012" from City of Chicago data portal.

https://data.cityofchicago.org/Health-Human-Services/Per-Capita-Income/r6ad-wvtk

[4] "chicago-community-areas.geojson" from kaggle.com.
https://www.kaggle.com/doyouevendata/chicago-community-areas-geojson

[5] Historical Real Per Capita Income for Chicago
https://www.deptofnumbers.com/income/illinois/chicago

IBM Developer

SKILLS NETWORK