

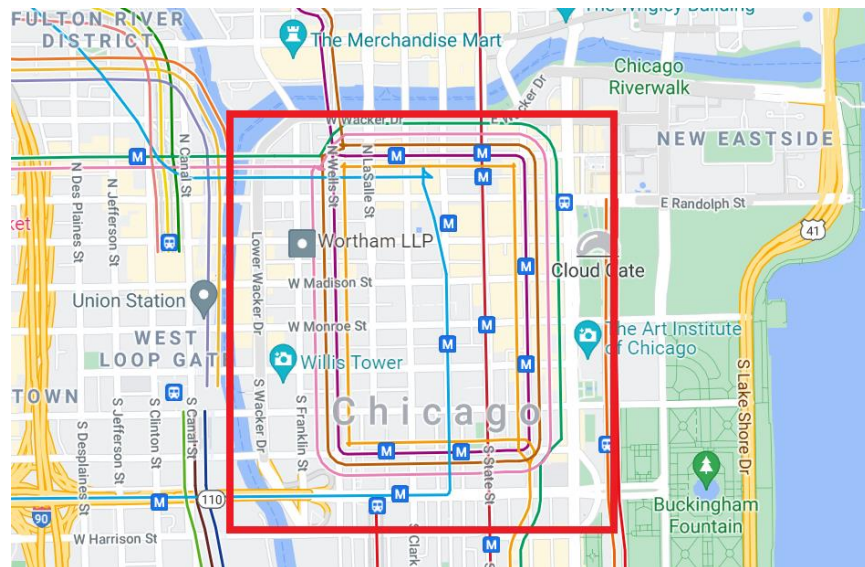
Capstone Project - The Battle of Neighborhoods (Week 2)

--- The best locations to open a new café at Chicago loop

1. Background & Problem Introduction

1.1 Background

Chicago, the third biggest city in the United States, is an international hub for finance, culture, industry, commerce, and transportation with about 2.7 million population according to the data in 2019. As one of the best large cities, it set a record of 58 million tourist visitors in 2018. Because of its unique geographic location, people from nearby states will also come for better career opportunities, which makes Chicago downtown a highly developed, and even crowded business district. The Loop is the center of downtown. It is named so because all the subway lines turn and go across this area, which merged into a rail loop. Please see the google map screenshot below for details.



1.2 Problem & who would be interested

High traffic flow and high population density mean more business opportunities. I believe many small business entrepreneurs or chain-store owners are interested in starting a new business in this area. For example, a café, which could offer quick breakfast, brunch, and coffee chat places.

There are already many café stores existing in the Loop. Is it possible to find a spot that is close to traffic hubs or commercial centers but not crowded with other café stores? Only under that scenario, the new store will have enough market shares to survive and win from the intensive competition.

2. Data

2.1 Data that we need

Based on my assumptions and thoughts in the background & problem section, below are the datasets that we need for our analysis:

- All the existed café stores' Latitude and longitude data
- Latitude and longitude data of traffic hubs and important commercial centers (Like Chicago City Hall and malls)
- Population density of the Loop area (either density data by blocks or a simple density map)

2.2 Data source

For the café stores Latitude and longitude data, I will log in to my personal developer Foursquare account and then make calls to the Foursquare API according to what we have learned during this capstone course. In the earlier assignment, we also learned how to transform the raw Foursquare JSON data into a Pandas data frame, including all the detailed info we need, like store name, latitude, and longitude. Please see the screenshot below. Now we can use this skill to acquire the data frames we need.

Get relevant part of JSON and transform it into a *pandas* dataframe

```
[17]: # assign relevant part of JSON to venues
venues = results['response']['venues']

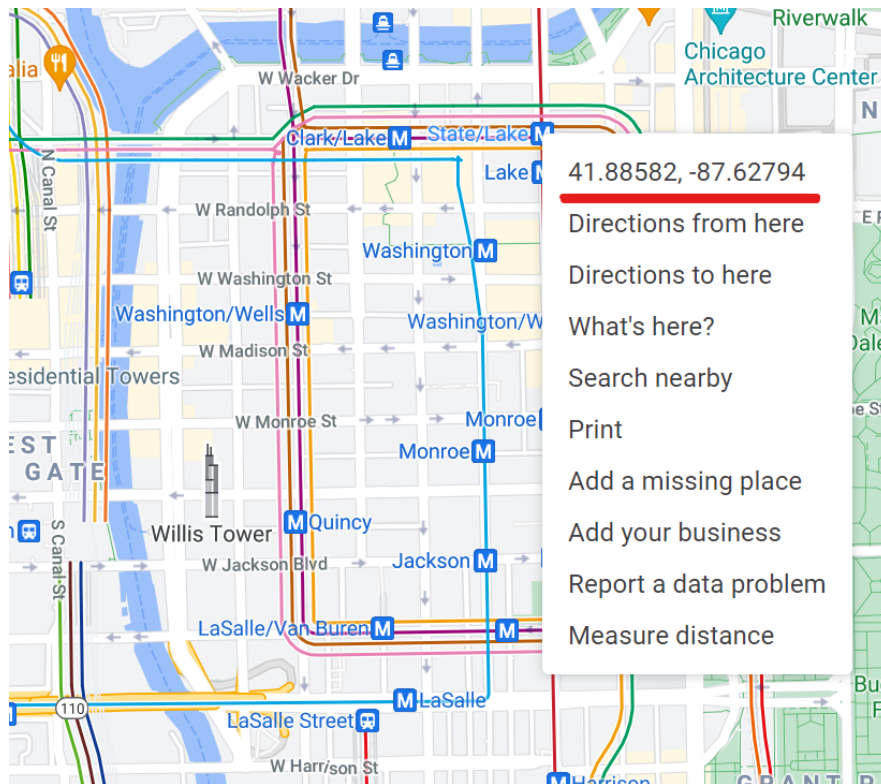
# transform venues into a dataframe
dataframe = json_normalize(venues)
dataframe.head()
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:5: FutureWarning: pandas.io.json.
son_normalize is deprecated, use pandas.json_normalize instead

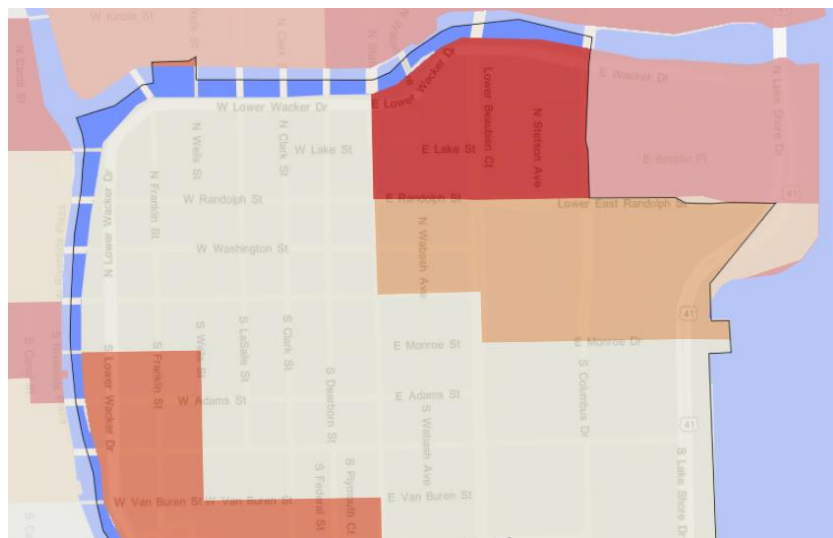
	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	loc
0	4fa862b3e4b0ebff2f749f06	Harry's Italian Pizza Bar	{('id': '4bf58dd8d48988d1ca941735', 'name': 'P...	v- 1620605014	False	225 Murray St	40.715218	-74.014739	
1	4f3232e219836c91c7bfde94	Conca Cucina Italian Restaurant	{('id': '4d4b7105d754a06374d81259', 'name': 'F...	v- 1620605014	False	63 W Broadway	40.714484	-74.009806	

With the help of the Folium Python library, we could easily visualize all the geographic data so that the current café business status will be concluded.

I found them on third-party websites in terms of population and geographic information for the city hot spots. Since there are only 5 city hot spots that we will mark out, let's get their latitudes and longitudes through Google Map at <https://www.google.com/maps>. It will be quick and simple. The most recent version of Google Map has a function for seeing the latitude and longitude by right-clicking. For example, for the State/Lake station, if we right-click it, the first line of the pop-up will be its latitude and longitude. Please see the screenshot below, which I have highlighted in red marker:



The population density map is from '<https://statisticalatlas.com/>', which has population density info for every district in the US. Unfortunately, I did not find any website with the data I need to do web scraping. But at least we have a map to refer to. Please see the picture below for what it looks like.



The darker area has a higher population density

3. Methodology & Analysis

The overall procedure will be:

- Use Foursquare query to get the details of existed café and Starbucks store, then transfer them into a whole Pandas data frame
- Get the best K value with elbow method for K-Means clustering
- Cluster and visualize all the stores in a Folium map, then mark all the city hot spots
- Combine the final Folium output with the population density map then spot potential areas we are looking for

3.1 Foursquare to data frame

Foursquare is the main platform I use in this project. With my personal developer account, I can apply search queries to get info of venues I need, then use Python libraries to download and convert the results to data frames. The reason why I searched both café and Starbucks is that when I check the output of keyword 'café', Starbucks is not included. However, they are also competitors of our new store. So, I did the search query twice and merged both outputs into a whole data frame.

```
In [11]: df1 = df_cafe.loc[:,['id','name','location.lat','location.lng']]
df2 = df_star.loc[:,['id','name','location.lat','location.lng']]
df_all = pd.concat([df1, df2], ignore_index=True, sort=False)
print(df_all.shape)
df_all.head()
```

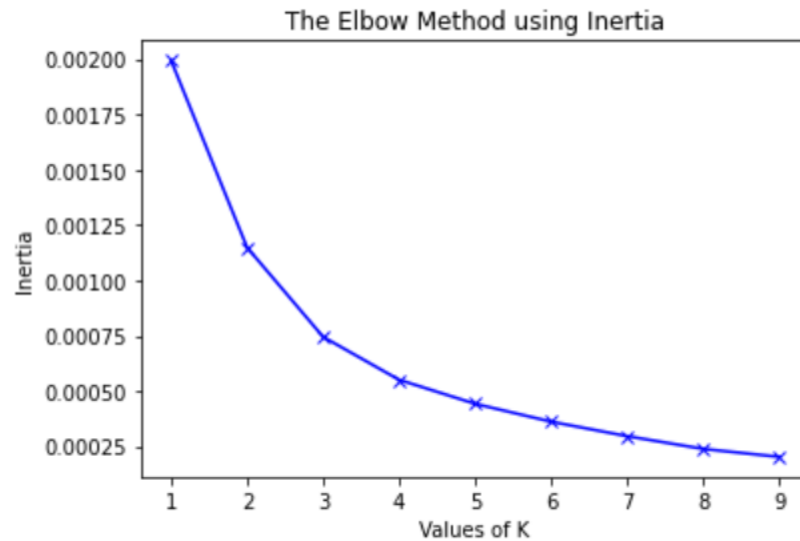
(90, 4)

Out[11]:

	id	name	location.lat	location.lng
0	5d54c94335b9e6000829a2ce	Hanabusa Cafe	41.881831	-87.626545
1	4ac3492df964a5201f9b20e3	Corner Bakery Cafe	41.885883	-87.632910
2	4b589a40f964a520ab6028e3	Alonti Café & Catering	41.883219	-87.633370
3	4b9e3467f964a520afd136e3	Jazzman's Café & Bakery	41.885042	-87.631631
4	4e4c6980bd413c4cc6692d63	Alonti Market Cafe	41.881320	-87.630132

3.2 Elbow method for the optimal K value

To find any similarities or patterns among existing stores, I will use K-Means clustering for grouping based on the latitudes and longitudes. The latitudes and longitudes could be seen as Y values and X values, respectively. So, the first step is to find the optimal K value with the Elbow method. We learned the theory of the Elbow method in this course. In terms of exact steps, I referred to an article on GeeksforGeeks (<https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>). The result seems not obvious, but we could still tell the elbow point is 4.



3.3 Visualization

After clustering, I inserted a new column Cluster Labels to indicate the group number of each store. Create the Folium map. I marked the Chicago city hall with a large solid red dot and then added all the stores in different group colors, also 4 big subway transfer stations, and a Macy's mall with dark non-filled spots.

```
In [17]: # set number of clusters as 4
kclusters = 4

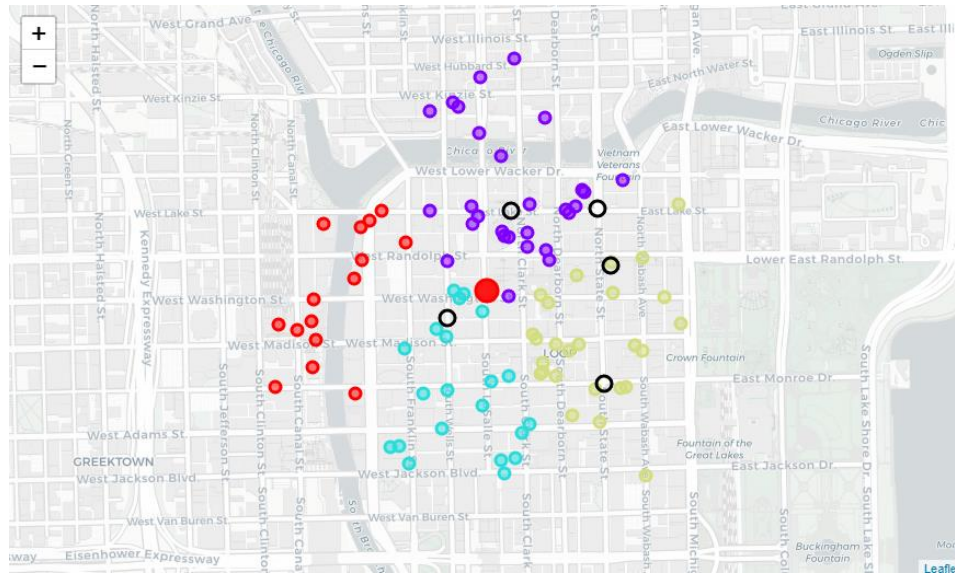
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(df_geo)

# insert a new column indication the lable of each store
df_all.insert(0, 'ClusterLabels', kmeans.labels_)
df_all.head()
```

Out[17]:

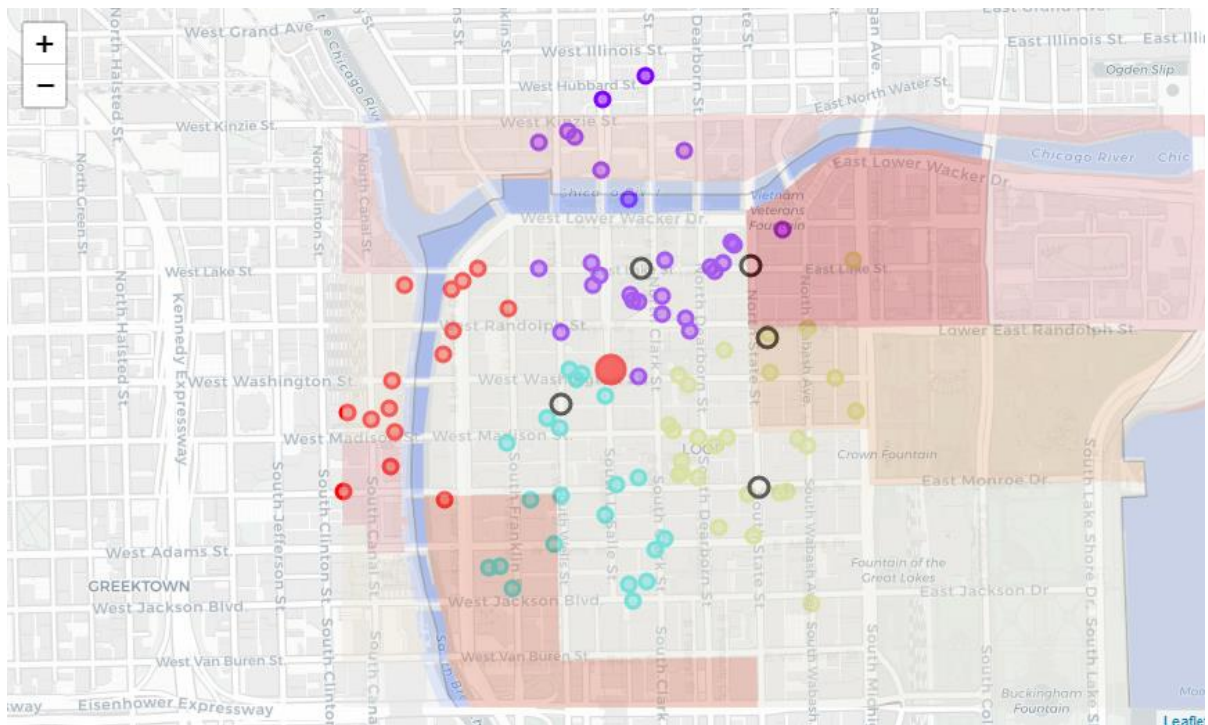
	ClusterLabels	id	name	location.lat	location.lng
0	2	5d54c94335b9e600829a2ce	Hanabusa Cafe	41.881831	-87.626545
1	0	4ac3492df964a5201f9b20e3	Corner Bakery Cafe	41.885883	-87.632910
2	1	4b589a40f964a520ab6028e3	Alonti Café & Catering	41.883219	-87.633370
3	0	4b9e3467f964a520afd136e3	Jazzman's Café & Bakery	41.885042	-87.631631
4	2	4e4c6980bd413c4cc6692d63	Alonti Market Cafe	41.881320	-87.630132

New data frame with Cluster Labels



3.4 Combine the maps

As mentioned in the data section, the only resource I have is the Loop population density map. The darker area has higher population density. If we try to analyze with two separate pictures, it will be quite hard to conclude and not accurate. So, I used Photoshop the app to combine them together. Make the Folium map bottom layer, then the density map second layer with 50 percent transparency. Below is my final output.



Final Output

4. Results

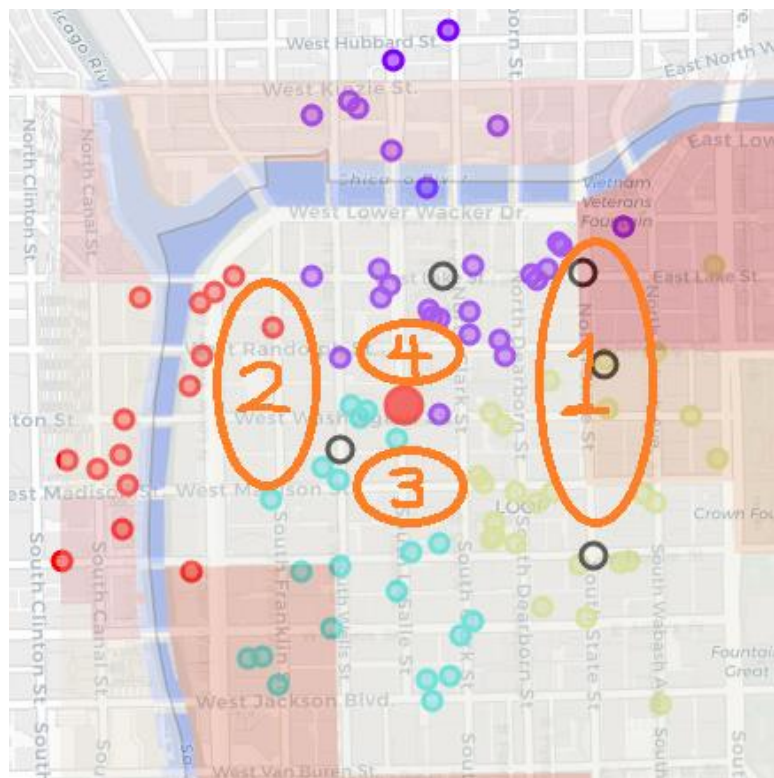
The final output of our analysis process is the pic above. The first thing I noticed is that all the venues are separated into 4 groups, which happened to be 4 directions around the Chicago city hall (large red dot). The red group stays further from the city hall than the other three groups, spreading along the Chicago River. Second, it is obvious that many existing stores are located around 3 city hot spots on the left side, which are the city hall, Washington/Wells station, and Clark/Lake station. This makes sense in facts because there are more traffic volume and office buildings beside hot spots. The third point is a gap in Group Purple. I believe it indicates that West Lower Wacker Dr is not an appropriate area for a café store. The top right and bottom left areas have higher density in terms of population density than the center.

5. Discussion & Recommendation

In my opinion, the key criteria of a good candidate area for opening a new café are:

- Close to city hubs like big subway transfer stations, malls, and the city hall.
- Not crowded with many existed stores.
- The higher the population density, the better.
- Surrounded by buildings instead of on a high traffic-volume street, for example 'West Lower Wacker Dr' in the map.

Let us find and circle the areas that meet the criteria above.



As we can see, the most recommended area for a new café store is Area 1. It is close to State/Lake subway station and Macy's. Not many stores existing, belongs to a high population density area. This is true from my personal experience since I worked in the Loop from 2017 to 2020.

The second most recommended area would be Circle 2 in the picture. The only reason why it is not as good as Area 1 is the population density. Despite that factor, Area 2 is a great location to consider. At last, Area 3 and 4 are small but still recommended. Those two intersections are good candidate places.

6. Conclusion & What could be improved

In this project, I used Foursquare as the main data resource practiced its search queries. With the help of the Folium library and the K-Means algorithm, I was able to visualize and find the pattern of existing café stores to prove my assumption. By utilizing Photoshop, I was able to take population density into consideration as another key factor. I am confident to conclude that there are still many potential and candidate places in the Loop. If any stakeholders are interested in opening a new café store, my final output picture can be a good reference for decisions.

As for things that could be improved, If I could find accurate data of population density by blocks or average traffic volume by blocks, I will know better how people are located. So that I could give quantitative results and tell the strength of each recommendation.

I believe there are available business sites for rent or sell in the Loop that we can find online. Filter and find appropriate ones for a café store. Combine that result with our output in the Recommendation section, we may be able to tell stakeholders which spot to choose instead of recommended areas.