

Analysis of Chicago Neighborhood for Potential Food Truck Locations

Maryna Bort

May 1st, 2020

1. Introduction/Business Problem

This project is focusing on helping a new client with finding the best location for a new food truck business called "Hot Potatoez".

The client is interested in expanding the business to Chicago, IL. The client would like to see a short list of neighborhoods where food trucks potentially would be in high demand.

Potential locations would be locations that are not crowded with a lot of restaurants or fast food chains. One of the most preferable locations would be parks or any other outdoor public places.

2. Data

This project requires two pieces of data.

The first piece is the list of all Chicago neighborhoods and their latitude and longitude. This piece of data will be taken from the website called Data.Opendatasoft. The data is in Geojson file.

The second piece of data will be acquired through Foursquare. This piece of data will provide the most common venues for each Chicago neighborhood. For example, a Chicago neighborhood Chrysler Village (Lat: 41.7764, Lon: -87.7518) has an airport terminal, airport service and heliport as the most common venues, which do not look like the best food truck locations.

After the two pieces of data are combined, the results will be clustered into sets of a number of neighborhoods that have appropriate most common venues for a food truck business. Each cluster will be examined to see if the neighborhoods in this cluster will be potential candidates for a food truck.

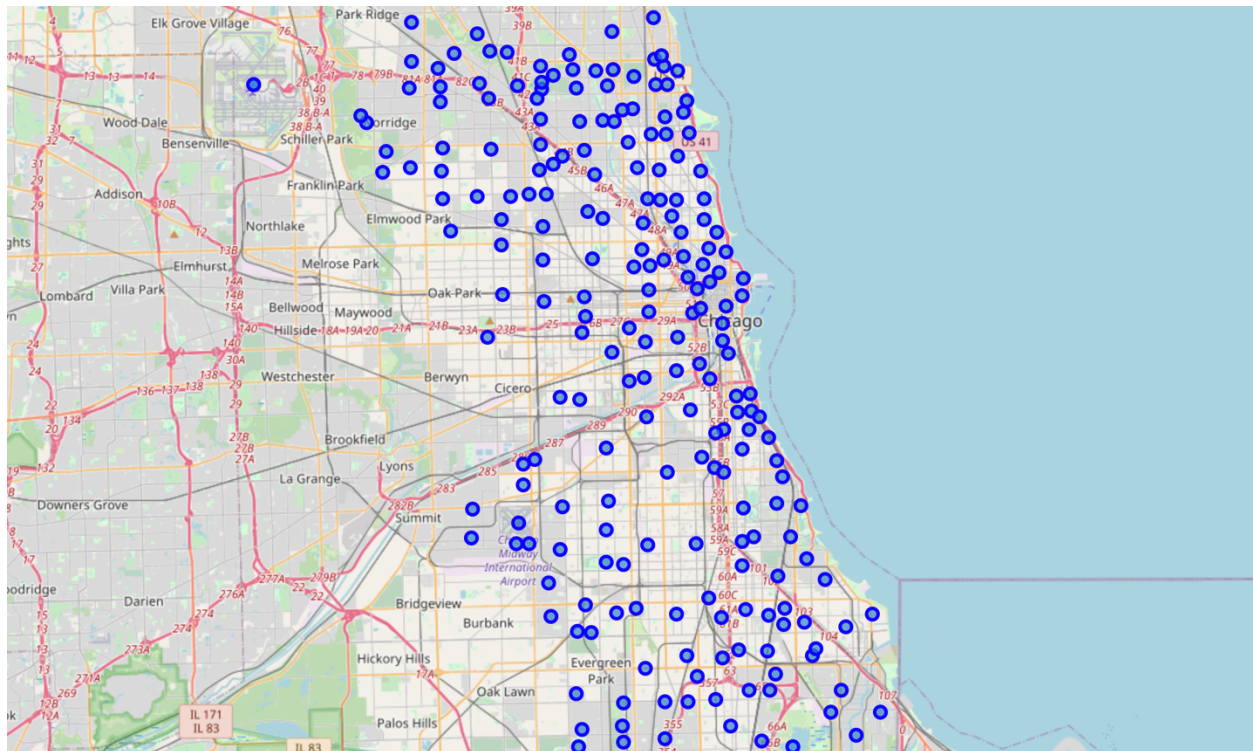
3. Methodology

The first step is to import all needed libraries. The libraries that we need include: pandas, numpy, folium, matplotlib, Nominatim, geopy and sklearn.

After that, we load and explore our data. When looking at the data, it becomes evident, that all the relevant data is in the *features* key, which is basically a list of the neighborhoods. We need to extract a list of neighborhoods along with their longitude and latitude. After that is done, all this information is put into a pandas dataframe which looks as the following image:

	Neighborhood	Latitude	Longitude
0	Heart of Italy	41.848657	-87.684331
1	The Bush	41.745141	-87.539765
2	Rogers Park	42.009614	-87.670109
3	Clearing (E)	41.776564	-87.744548
4	South Loop	41.860771	-87.625568

Next, we use geopy to get the coordinates of Chicago to later create a map of the city. We also use the list of neighborhoods to superimpose them on top of the city map. The result is the following:



The next step is to explore the Chicago neighborhoods using the Foursquare API. This requires creating the GET request, making the GET request and returning only relevant information for each nearby venue. After this is done, we get a new dataframe called “Chicago_venues”, which contains information about venues in each of the Chicago neighborhoods based on their longitude and latitude.

The next step is to analyze our data. To do that, we will find the frequency of occurrence of each venue category for each of the neighborhoods. We can see the top 5 most common venues per neighborhood:

```

----Archer Heights----
      venue  freq
0 Mexican Restaurant 0.29
1           Bank     0.12
2           Pharmacy 0.12
3       Dessert Shop 0.06
4           Market 0.06

----Ashburn----
      venue  freq
0  Cosmetics Shop 0.25
1 Construction & Landscaping 0.25
2    Automotive Shop 0.25
3   Light Rail Station 0.25
4             ATM    0.00

```

All this information is put into a pandas dataframe along with the top 10 most common venues for each of the Chicago neighborhoods. The dataframe looks as the following:

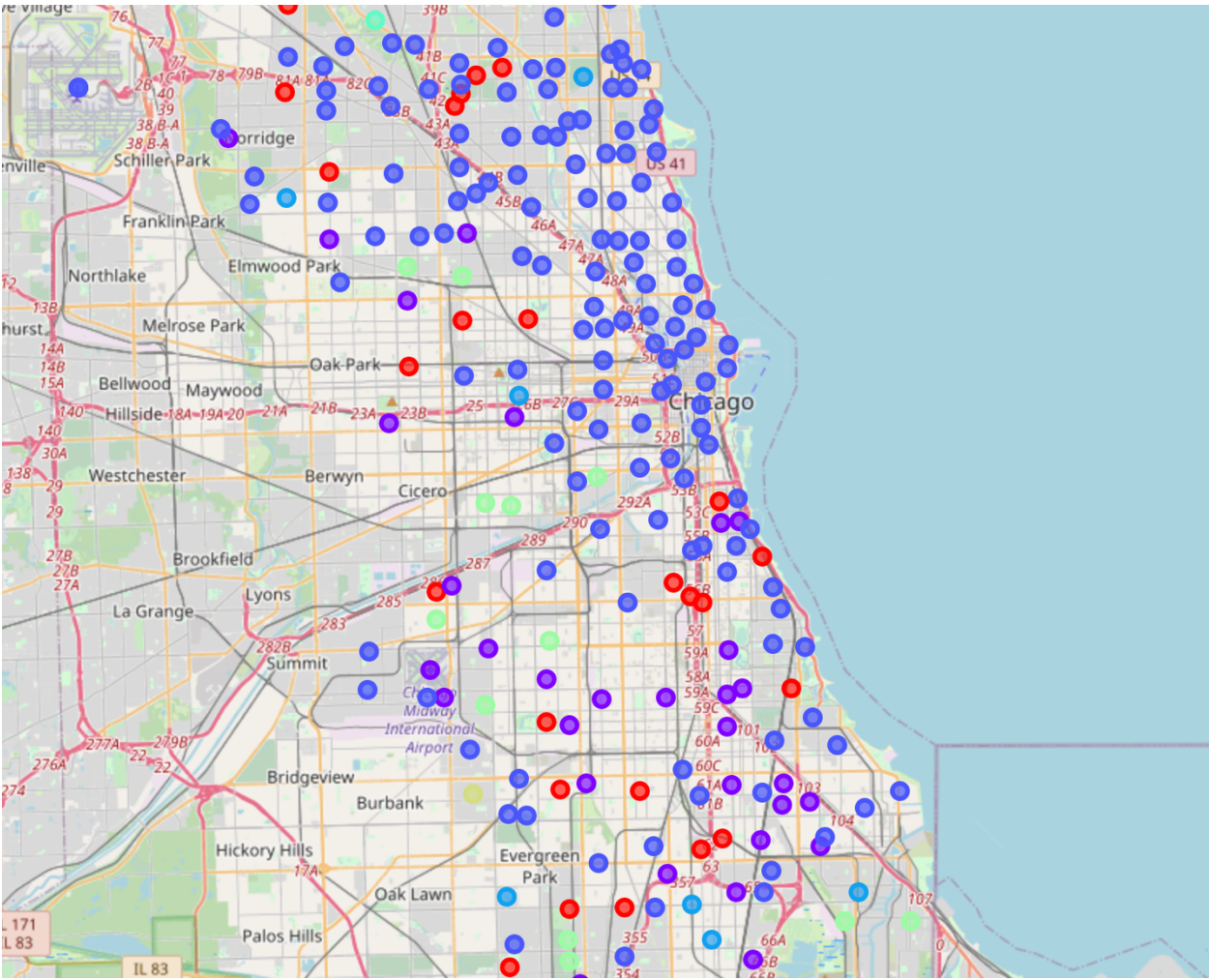
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albany Park	Bus Station	Sandwich Place	Coffee Shop	Thai Restaurant	Thrift / Vintage Store	Sushi Restaurant	Chinese Restaurant	Paper / Office Supplies Store	Train Station	Discount Store
1	Altgeld Gardens	Park	Field	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Falafel Restaurant	Farm	Farmers Market
2	Andersonville	Sushi Restaurant	Breakfast Spot	Pizza Place	Coffee Shop	Middle Eastern Restaurant	Burger Joint	Café	Bakery	Grocery Store	Southern / Soul Food Restaurant
3	Arcadia Terrace	Gym	Middle Eastern Restaurant	Bakery	Vietnamese Restaurant	Burger Joint	Greek Restaurant	Sushi Restaurant	Eastern European Restaurant	Turkish Restaurant	Nature Preserve
4	Archer Heights	Mexican Restaurant	Pharmacy	Bank	Cosmetics Shop	Market	Nightclub	Chinese Restaurant	Sandwich Place	Dessert Shop	Seafood Restaurant

The next step is to cluster our neighborhoods based on their most common venues. For this task, we use K-Means clustering. K-Means clustering is an unsupervised machine learning algorithm. It creates clusters of data points based on how similar points are within a cluster and how different they are compared to the clusters around. For this project we set the number of clusters to 10. It seems as any value less than 10 would be not enough as we have more than 200 neighborhoods, and values above 10 would create too many clusters that would have a very small number of neighborhoods per cluster.

After that, we create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood, it looks as the following:

	Neighborhood	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
0	Heart of Italy	41.848657	-87.684331	2.0	Italian Restaurant	Mexican Restaurant	Pizza Place	Thrift / Vintage Store	Sandwich Place	Food Truck	Fast Food Restaurant	Shoe Store	Gas Station	Bank
1	The Bush	41.745141	-87.539765	2.0	Concert Hall	Zoo Exhibit	Filipino Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Falafel Restaurant	Farm	Farmers Market
2	Rogers Park	42.009614	-87.670109	2.0	Mexican Restaurant	Pizza Place	American Restaurant	Donut Shop	Theater	Chinese Restaurant	Bakery	Train Station	Bar	Grocery Store
3	Clearing (E)	41.776564	-87.744548	1.0	Hotel	American Restaurant	Pizza Place	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Mexican Restaurant	Martial Arts Dojo	Taco Place	Taxi Stand
4	South Loop	41.860771	-87.625568	2.0	Park	Grocery Store	Coffee Shop	Bar	Dog Run	Steakhouse	Donut Shop	Bank	Breakfast Spot	Salon / Barbershop

We can also add the resulting clusters to the map of Chicago:



4. Results and Discussion

Based on our analysis, we can see that the neighborhoods in Chicago are very diverse. Our k-means clustering analysis did not produce a clear winner. Although the neighborhoods within each cluster still have a good amount of differences, the results might be used to see what list of neighborhoods should be considered above the rest.

For instance, if we look at the cluster #2, we can see that this cluster is probably not the best set of neighborhoods that would work for our client's food truck business. The reason for that is that this set of neighborhoods have a lot of restaurants, fast food chains, pizza spots, lounges and food joints. We think that it would be hard to compete in such neighborhoods with such dense population of food places. For this reason, this cluster is excluded from the list of potential candidates. The rest of the clusters except 1 and 8 are also excluded from the list of potential neighborhoods.

If we look at clusters 1 and 8, we can see that the most common venues are parks, soccer fields, pools and beaches along with zoo exhibits, farmers markets and bike trails. Based on that, we can say that these two clusters would be the best choice for food truck locations.

5. Conclusion

The purpose of this project was to help a client with expanding his food truck business called "Hot Potatoez" to Chicago.

The process included collecting a list of neighborhoods, their longitude and latitude; retrieving venue information from the FourSquare website; using the two previous pieces of information to find the most common venues for each neighborhood; after that clustering (using k-means clustering) our neighborhoods to produce a set of neighborhoods that would be the best potential candidates for the food truck business.

The results of our analysis, allow us to conclude that clusters 1 and 8 are our choice for potential food truck locations. These two clusters have their most common venues as parks, zoo exhibits, soccer fields, bike trails, beaches, pools and farmers markets. We think that these public venues would fit our business most and also the lack of restaurants and other food places will help our client not to suffer from competition.