Restaurant location problem

1. Introduction

This report is part of the final assignment for the Applied Data Science Capstone project. In this report will be presented a business problem concerning the best location of a restaurant in the city of Chicago-IL. In the next subsections, the context of the problem will be described, and the research question will be proposed. After the introduction, it will be presented the process to acquire and clean the data (Section 2), the research method to answer the research question (Section 3), the main results (Section 4), the discussion of the results (Section 5) and the conclusion (Section 6).

1.1 Background:

One of the most beautiful parks in Chicago,IL, is the Millennium Park and he is also one of the most positively reviewed The park is one of the most famous parks in IL, attracting millions of visitors a year.

Due the large number of visitors the park receives every day, an investor would like to know if the surrounding of the park would be a good place to build his restaurant. To answer this question, the investor will open his restaurant if there is a region, next to the park, with not too many competitors. If such area is available, he also wants to know if it is better to open an Italian, a vegetarian or a Japanese restaurant, the types of food he already invests in.

2. Data acquisition and cleaning:

To answer the question proposed in the business problem, some location information will be required. First, it is necessary the latitude and longitude of the park. Due to its size, the center of the park will be used as reference for its location. The latitude and longitude were already identified using Google maps and is 41.882222, -87.621370. With the latitude and longitude of the park, the closest food venues will be identified through Foursquare API, with a radius of 2500 meters. This large radius was chosen due the size of the park. With a 2500 radius, we will assure that at least a five-block radius surrounding the park will be considered when looking for food venues.

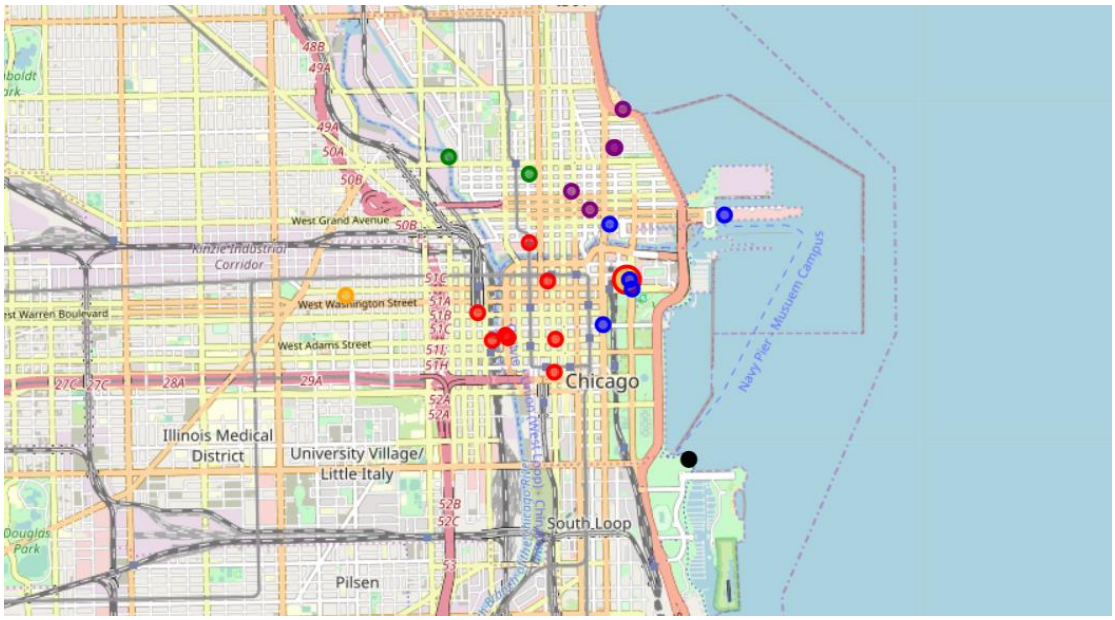
After collecting the data, the data will be reviewed and cleaned. A cluster analysis will be performed using the latitude and longitude of the venues to identify regions with lower density of food venues. The clusters will be plotted through using folium to make the visualization easier. If there is a low density region, the categories of the venues will be reviewed to see what are the most popular food types and a recommendation of a type of restaurant will be made based on the other types of restaurants available in the region selected.

3. Results:

In this section, the main results of the cluster analysis will be presented using the folium module as a representation tool. First, all the venues identified were plotted in a map for a first analysis.

To further explore the regions, the cluster analysis was performed considering six clusters.

Figure 1: Location of the food venues considering their clusters



With Figure 1 is possible to confirm The Millennium park is filled with food venues from the north (Mixing the purple and green clusters shows that the north is to far from the park and is already filled) combining it with blue cluster we deduce that opening a food venue in the north is not possible.

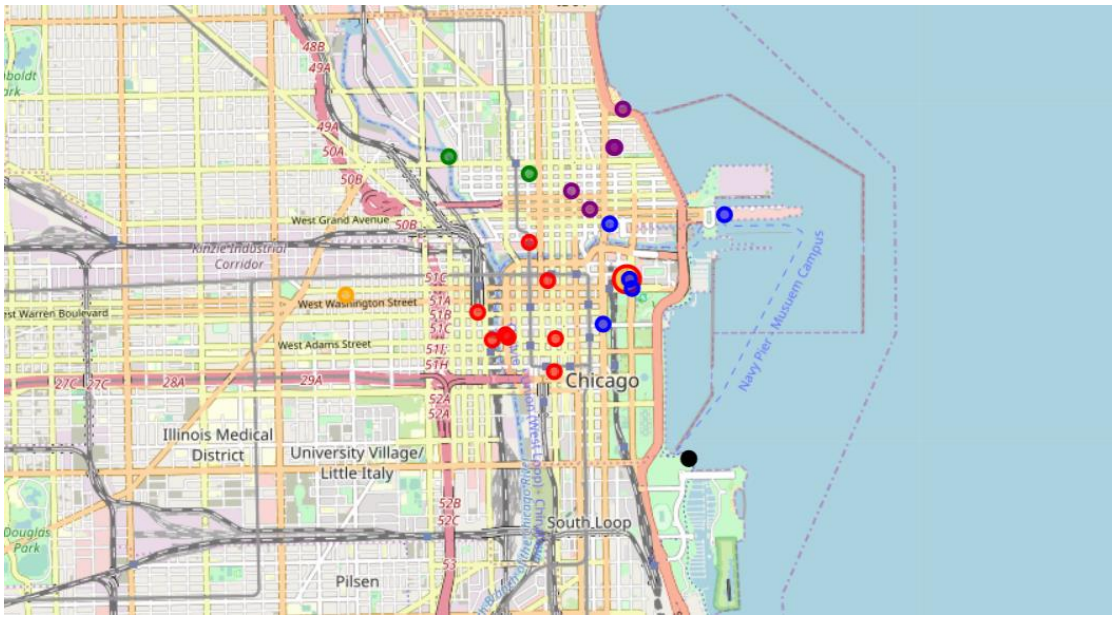
If we look the west (red cluster) we could see that its also filled with food venues which might not be a good option but further investigation upon that shows us most of this venues are not restaurants:

Figure 2:food venues in the red cluster

0	Atrium Food Court	Food Court	100 W Randolph, Concourse	btw Clark & Lake	41.885163	-87.631639	[{'label': 'display', 'lat': 41.88516275046533...	911	60601	US	Chicago	IL	United States	[100 W Randolph Concourse (btv Clark & Lake),..
2	Revival Food Hall	Food Court	125 S Clark St	W Adams St	41.879560	-87.630550	[{'label': 'display', 'lat': 41.87956, 'lng': ...	816	60603	US	Chicago	IL	United States	[125 S Clark St (W Adams St, Chicago, IL 6060.
10	Chicago Union Station Food Court	Food Court	225 S Canal St	at Adams St.	41.879339	-87.638738	[{'label': 'display', 'lat': 41.879339217638, ...	1474	60606	US	Chicago	IL	United States	[225 S Canal St (a Adams St., Chicago, IL 60.
11	Bridgeport Pasty - Food Truck	Food Truck	NaN	NaN	41.879691	-87.636715	[{'label': 'display', 'lat': 41.87969136866029...	1302	60608	US	Chicago	IL	United States	[Chicago, IL 60608 United States
16	Ogilvie Food Court	Food Court	500 W Madison St	at Ogilvie Transportation Center	41.882125	-87.640672	[{'label': 'display', 'lat': 41.88212517500138...	1599	60661	US	Chicago	IL	United States	[500 W Madison S (at Ogilvie Transportation C.
19	Mediterranean Express Food Truck	Food Truck	NaN	NaN	41.879865	-87.637026	[{'label': 'display', 'lat': 41.879865, 'lng':...	1323	60606	US	Chicago	IL	United States	[Chicago, IL 60606 United States
21	The Mart Food Court	Food Court	222 Merchandise Mart Plz	NaN	41.888869	-87.634052	[{'label': 'display', 'lat': 41.88886897769535...	1285	60654	US	Chicago	IL	United States	[222 Merchandise Mart Plz, Chicago IL 60654, .
22	Coco's southern style soul food	Southern / Soul Food Restaurant	426 S Clark St	NaN	41.876320	-87.630755	[{'label': 'display', 'lat': 41.87632, 'lng': ...	1018	60605	US	Chicago	IL	United States	[426 S Clark St Chicago, IL 60605 United Sta.

By further investigating figure 2 we can deduce that opening a Japanese or Italian restaurant in the red cluster might be a good idea, while the restaurant might suffer from the fast food venues nearby it will thrive on the close proximity to the park and on the lack of restaurants in the cluster.

If we look the the south (black cluster):



We can see that the black cluster is also lacking in food venues and we should investigate further upon that:

Figure 3:Black cluster venues.

```
#Displaying the blue cluster
food_filtered.loc[food_filtered['cluster'] == 5,:]
```

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	neighborh
4	Bubblenet Food Court	Food Court	1200 S Lake Shore Dr	@ Shedd Aquarium	41.867925	-87.61326	[{'label': 'display', 'lat': 41.86792457273813...	1727	60605	US	Chicago	IL	United States	[1200 S Lake Shore Dr (@ Shedd Aquarium), Chic...	

Upon further looking into this place its rating is only 5.4 and it is also far right inside the park which means it will not be competition to us and even furthermore we can see that this place is a fast food place and the black cluster will be the most beneficial for us to open our new restaurant and it doesn't matter even which one.

6. Conclusion:

In this project, an investor wanted to open a restaurant next to the millennium park in Chicag,IL,to take advantage of the number of tourists that visit the parks every year. A cluster analysis was used to identify the regions next to the park that had the lowest density of food venues, and the South region were identified as good option. The black cluster region had a lower density of food venues. In the South, the investor can choose between Italian , vegetarian and Japanese , because there aren't any restaurants available nearby.