Location for Mediterranean restaurant in Chicago

Faisal Farooqi

February 2021

1. Introduction

1.1 Background

I am a big fan of Mediterranean food. More than a decade ago I introduced a colleague at work to this cuisine and he immediately took a liking to it. So much so that a few months later he decided to open a restaurant and asked me for help with the menu and selection of a location.

I was more than willing guidance on the menu but had no idea how to recommend a location. Our office was located in the suburbs of Chicago, IL and there were multiple restaurants with similar offering within 10-minute driving distance. Similar situation existed in the vicinity of my home where I could pick up a kebab dinner within a short driving distance. This was a time when data science had not become popular and using data to make a selection never crossed mind. We resorted to scouting locations the hard way which was driving to them and assessing the neighborhood. After wasting considerable time and resources my colleague dropped the idea of opening a restaurant.

Many years later I now feel that I am in a better position to suggest a location for opening a restaurant if someone asks me for help. Data science tools have evolved, python libraries for data handling are well developed and a multitude of data sources are available to make an informed decision.

1.2 Problem

Restaurant industry has been severely impacted by the covid pandemic. Local regulations in almost all countries have limited in-store dining to some extent and the state of Illinois (where Chicago is located) is no different. In the current situation anyone looking to open a restaurant would prefer a location with good foot traffic. The usual spots like city centers, sports arenas and movie theaters are not good candidates either due to covid related restrictions on attendance.

The aim of this report is to not only suggest a location for a Mediterranean restaurant in the city of Chicago but to also consider the influence of covid related factors and select a location which is suitable for customers to carry out the food, for example close to a bus or train station.

2. Data sources and usage

2.1 Sources of data

The primary source of data is the Foursquare location data available via *Place API* here. This API allows query of data based on location and other optional parameters.

City of Chicago can be considered a grid due to the road network generally aligned in East-West and North-South directions. This grid can be used to define regions and evaluate restaurants within each region but would require extensive work to achieve evenly distributed regions and to calculate latitude/longitude of region centers. A better source for the purpose of creating geographical regions is the zip codes used by the United States Postal service. List of zip codes used in United States is available here. This data includes longitude and latitude of each zip code along with GeoJSON format map description.

A map of zip codes which belong to city of Chicago can be found here and the list of Chicago zip codes is here.

Chicago Transit Authority (CTA) maintains a website which contains route maps, station location and rider statistics. Data can be downloaded in various formats.

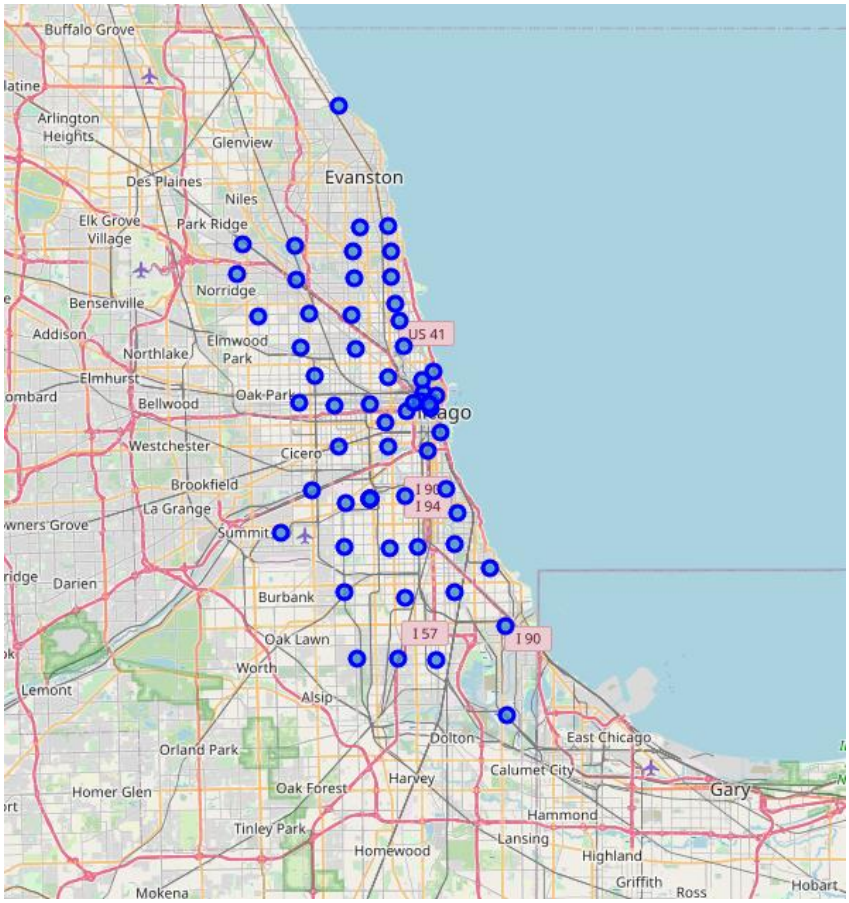
2.2 Data cleaning

Multiple Foursquare API sub-categories within the *Food* category can be considered as Mediterranean cuisine. To narrow response results the following sub-categories need to be selected:

- Mediterranean Restaurant
- Kebab Restaurant
- Turkish Restaurant
- Middle Eastern Restaurant
- Greek Restaurant
- Falafel Restaurant

Zip code longitude/latitude are available via *opendatasoft* need to be filtered to select Chicago zip codes. There are multiple ways to achieve this, first is to filter on state and city. Second method is to merge the table from *opendatasoft* with list of Chicago zip code obtained from other sources listed above. 'Timezone' and 'Daylight savings time flag' fields provided by *opendatasoft* are redundant for the analysis of this report and can be dropped.

43,191 zip codes were found in the data downloaded from *opendatasoft.* Filtering them for state of Illinois and then for city of Chicago resulted in 85 zip codes. These are plotted in the next diagram, blue dots represent the approximate center of the area covered by a zip code:



Dataframe for Chicago zip code looks like the following after dropping extra columns.

	Zip	City	State	Latitude	Longitude
0	60668	Chicago	IL	41.811929	-87.68732
1	60646	Chicago	IL	41.995331	-87.76010
2	60654	Chicago	IL	41.888627	-87.63538
3	60647	Chicago	IL	41.921126	-87.70085
4	60606	Chicago	IL	41.882582	-87.63760
...	...	...	...	...	...
80	60633	Chicago	IL	41.655423	-87.55365
81	60638	Chicago	IL	41.787982	-87.77380
82	60683	Chicago	IL	41.811929	-87.68732
83	60632	Chicago	IL	41.809299	-87.71050
84	60603	Chicago	IL	41.880446	-87.63014

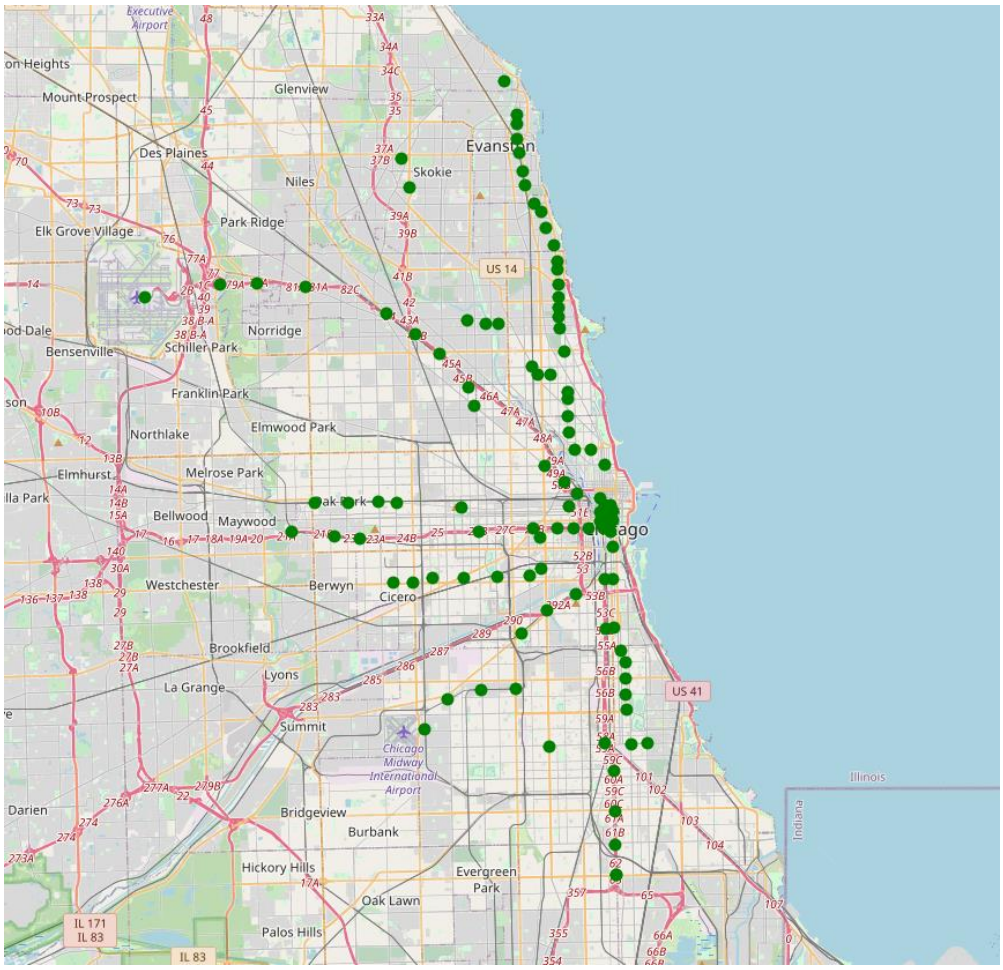
Data for one of the zip code towards the north side of the city has incorrect co-ordinates at it shows up outside the city boundaries. Location of this zip code was modified manually.

Within data available from CTA we shall only consider train stations and ignore bus stops for this project. Train stations cover a larger surface area which results in better chances of finding an open lease when compared to bus stops. From the data available for train stations only 'STATION_NAME' and 'Location' fields are relevant, the remaining columns can be dropped. There are duplicate entries for some stations due to the fact they have entrances on different streets, these need to be removed so that only unique entries are left.

Data downloaded from CTA website contained 300 station entries, some of them were duplicate as stations have multiple entrances and each was listed separately. After removing duplicate entries, dropping redundant columns and creating new columns to hold Latitude and Longitude separately we got the following dataframe with 109 unique stations:

	STATION_NAME	Latitude	Longitude
0	18th	41.857908	-87.669147
1	35th/Archer	41.829353	-87.680622
2	35th-Bronzeville-IIT	41.831677	-87.625826
3	43rd	41.816462	-87.619021
4	51st	41.80209	-87.618487
...	...	...	...
104	Chicago	41.896075	-87.655214
105	63rd	41.780536	-87.630952
106	Cermak-McCormick Place	41.853115	-87.626402
107	Wilson	41.964273	-87.657588
108	Washington/Wabash	41.88322	-87.626189

Next diagram shows the location of CTA train stations within the city. Each green dot represents a station.



2.3 Data usage

Once data collected from various sources is cleaned up the first step is to setup a grid of geographical regions covering city of Chicago via zip codes boundaries. After that general restaurant density and Mediterranean restaurant density should be computed for each zip code. This can be done based on distance or by matching the zip code from results returned by Foursquare API. A heatmap can be used to visualize the results.

Train station locations can be matched up with zip codes via location information to enhance the heat maps to select candidate areas suitable for opening a new restaurant. Use of clustering may or may not be required depending on the results obtained via heat map examination. If no areas stand out then clustering can be used to incorporate additional features returned by Foursquare API (e.g ratings) to form regions which can serve as candidate for a new Mediterranean restaurant.

3. Methodology

First step is to check the percentage of Mediterranean restaurants against the total restaurants in the defined area. Foursquare API was used to look up all restaurants and Mediterranean restaurants within a radius of 500m. This radius value was found out to be low and not many restaurants were returned in query results.

Radius was eventually increased to 2500m which provided better coverage between center spots of zip codes and resulted in more hits against the criteria defined for Foursquare API.

```
Total number of restaurants: 3204
Total number of Mediterranean restaurants: 89
Percentage of Mediterranean restaurants: 2.78%
Average number of restaurants in neighborhoods: 79.96470588235294
```

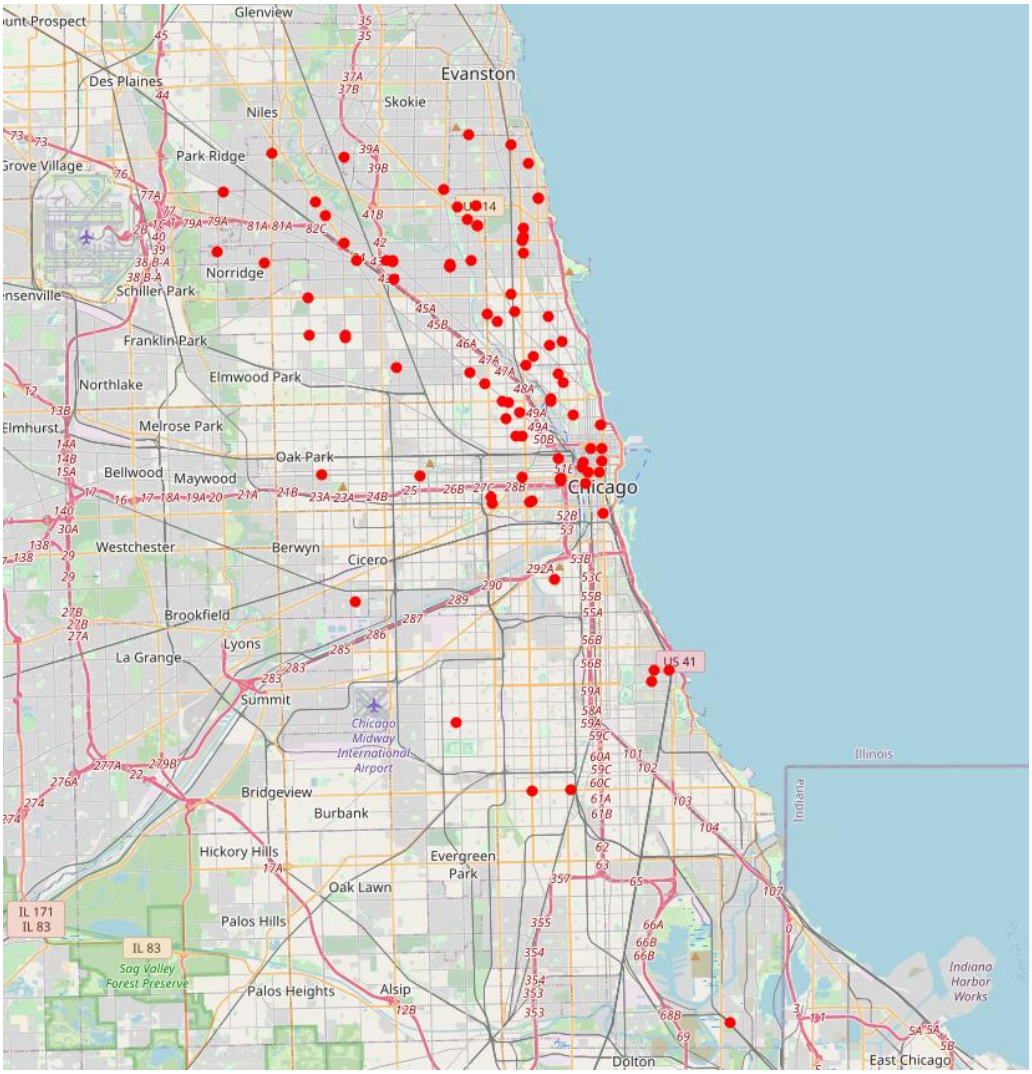
Less than 3% of restaurants found are classified as Mediterranean according to the categories selected for this analysis. Overall, there are about 80 restaurants on average in each zip code, this is not as dense as expected in a large city.

Some of the restaurants classified as Mediterranean were validated by printing out part of the list created during analysis.

```
List of Mediterranean restaurants
---------------------------
('524f47017e48b2ca0ee15ee1', 'NOK', 41.99221350729494, -87.78258099086783, '6075 N Milwaukee Ave (at N Elston Ave), Chicago 60646', 1892, True)
('51c74958498ea67eb8f65a6b', "Brandy's Gyros", 41.986706727822856, -87.77736404315814, '5777 N Milwaukee Ave (btwn N Austin & W Ardmore), Chicago 60646', 1721, True)
('4e35aa3b2271b2cf067a445d', 'Gourmetpita', 42.010501861572266, -87.76712799072266, '5549 Touhy Ave (Central), Skokie 60077', 1786, True)
('577563d5498e8d2eb29bd25d', 'Ema', 41.89091768250545, -87.63095663368316, '74 W. Illinois St. (Clark), Chicago 60654', 1167, True)
('5065ca63e4b03c090a0fd8b5', 'Naf Naf Grill', 41.88327556720619, -87.63556332317656, '309 W Washington St (btwn N Franklin St and N Wacker Dr), Chicago 60606', 548, True)
('5d14b83b3f9ff700234d9f45', 'The Purple Pig', 41.89079113341187, -87.62501120567322, '444 N Michigan Ave, Chicago 60611', 1227, True)
('5806b79538fa3fe281b7170b', 'Roti Modern Mediterranean', 41.8860482, -87.6249484, '80 East Lake Street (Michigan Ave.), Chicago 60601', 757, True)
('50c0f3cce4b00a170e2828c1', 'Benjyehuda', 41.88140598249916, -87.63239323865992, '10 S LaSalle St (at W Madison St), Chicago 60602', 215, True)
('5b1854154a7aae002cbe8e62', 'Aba', 41.886852162003066, -87.64888478200851, '302 N Green St (Fulton St), Chicago 60607', 638, True)
('50baac29e4b0f3b7954934ec', 'Masada', 41.921994378701974, -87.6974575242601, '2206 N California Ave (btwn W Palmer St & W Lyndale St), Chicago 60647', 297, True)
('5821410fd0d8ce77a3f93698', 'Zaatar MedGrill', 41.9175043804195, -87.68972235923317, '2459 W Armitage Ave, Chicago 60647', 1006, True)
('4a4aaa0ff964a5200fac1fe3', 'Artopolis Bakery, Cafe and Agora', 41.87765448246344, -87.64725275816552, '306 S Halsted St (at W Jackson Blvd), Chicago 60661', 539, True)
('4a9c6450f964a520053720e3', 'Greek Islands', 41.87892161094905, -87.64732805417756, '200 S Halsted St (at Adams St), Chicago 60661', 417, True)
('54526cc8498e4c349557502e', 'Naf Naf Grill', 41.88126702765826, -87.6260884025218, '28 S Wabash Ave (Monroe), Chicago 60603', 348, True)
('4a805101f964a520f0f41fe3', 'Athena Greek Restaurant', 41.87887126409897, -87.64730910579813, '212 S Halsted St (btwn W Adams St & W Jackson Blvd), Chicago 60661', 421, True)
('525aed9c11d22db7dc07bb1d', "Sander's Family Restaurant & Pancake House", 41.82800018696912, -87.76072917918232, 'Skokie', 1769, True)
('4c5df1b5857ca5931579cecb', "Mr. Zee's", 41.9527518395677, -87.78684101298944, '3960 N Narragansett Ave (at W Irving Park Rd), Chicago 60634', 1223, True)
('59efc308da2e00304625f3e3', 'Jerusalem Grill', 41.937474, -87.786018, '3142 N Narrangasett Ave, Chicago 60634', 1167, True)
('51c0b461498e32096df5faed', 'The Nile Restaurant', 41.79522707950626, -87.59725841734928, '1162 E 55th St, Chicago 60615', 1889, True)
('4ad68d06f964a520c70721e3', "Cedar's Mediterranean Kitchen", 41.79992882764553, -87.59621672128615, '1206 E 53rd St (at S Woodlawn Ave), Chicago 60615', 523, True)
...
Total: 89
```

Next step is to visualize the location of the selected restaurants around Chicago's neighborhood and observe if clustering is required to group them and isolate candidate locations for a new restaurant.

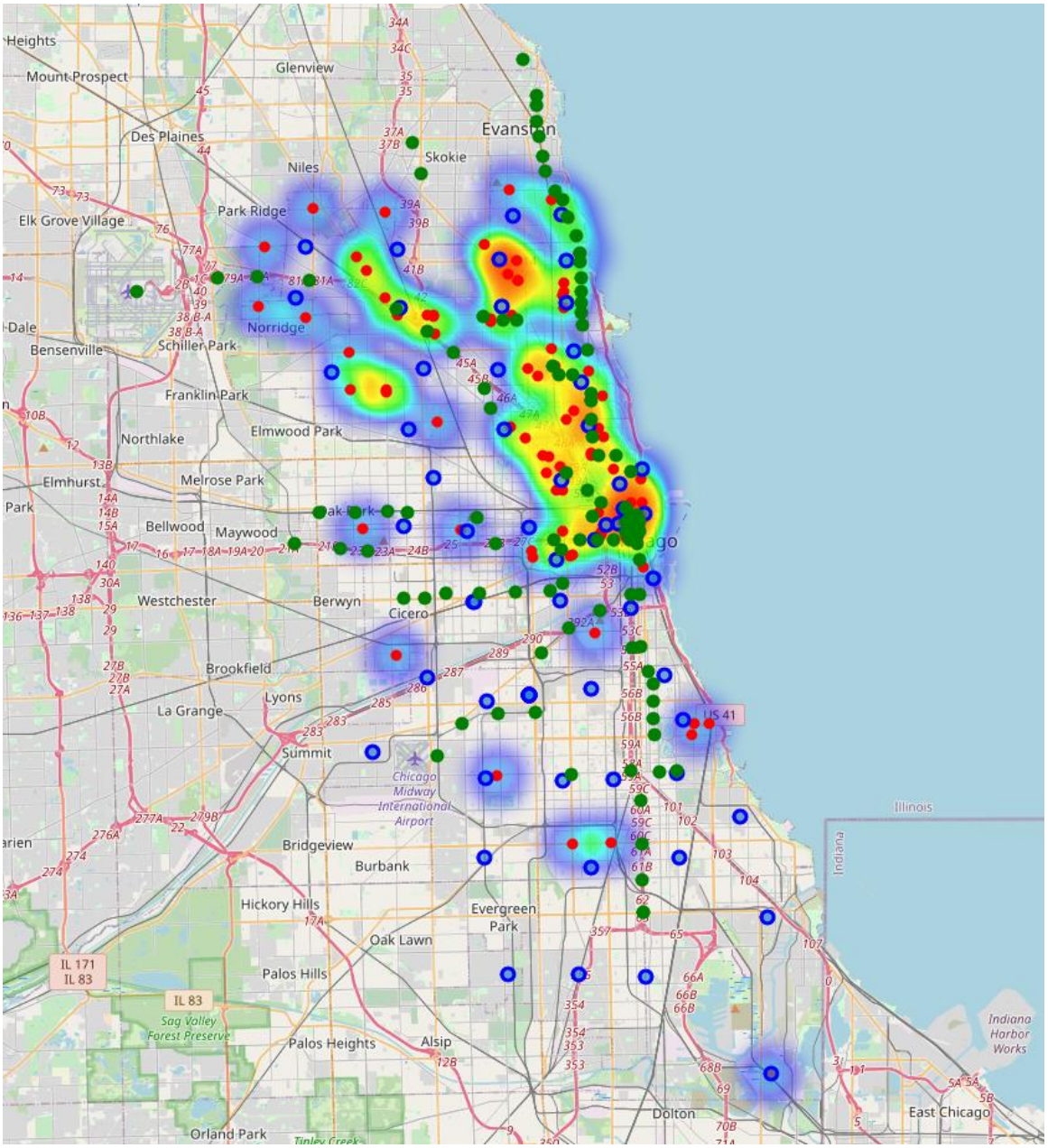
Next diagram shows the city of Chicago with location of Mediterranean restaurants identified by Foursquare data. Each red dot represents a Mediterranean restaurant.



Mediterranean restaurants are concentrated around the North and North-West neighborhoods. Vast areas of the city exist with no restaurants. Instead of machine learning heatmap would be a better tool exclude areas with comparatively higher density of Mediterranean restaurants.

Next and last step is to create heat maps of existing Mediterranean restaurants and superimpose them with zip code markers and train station locators. The resulting map can be used to identify a few spots around the city which can serve as candidates to open a new restaurant.

Heatmap produced with combined data is shown next. Red markers are for Mediterranean restaurants, green for train stations and blue represent center of area served by zip code.


Heights
Mount Prospect
Glenview
Des Plaines
Evanston
Skokie
Niles
Park Ridge
Elk Grove Village
Norridge
Dale
Bensenville
Schiller Park
Franklin Park
Northlake
Elmwood Park
Elmhurst
Melrose Park
Park
Bellwood
Maywood
Oak Park
Westchester
Berwyn
Cicero
Brookfield
Lyons
La Grange
Summit
Chicago Midway International Airport
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
IL 171
IL 83
Sag Valley Forest Preserve
Palos Heights
Alsip
Orland Park
Dolton
East Chicago
Indiana Harbor Works
Illinois
Indiana
US 41

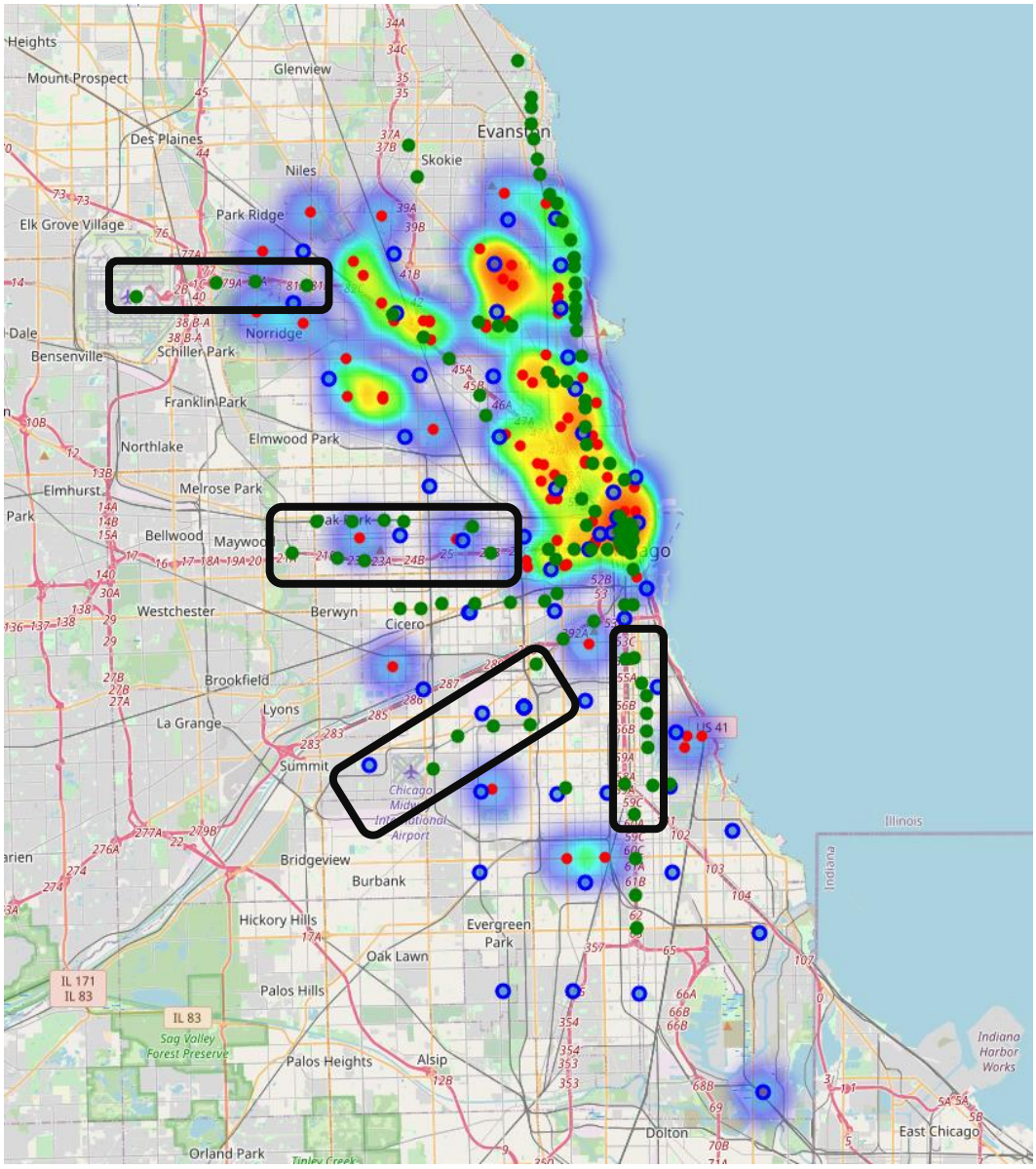
4. Results

Heatmap of existing Mediterranean restaurants clearly shows that they are concentrated in the downtown area and North/North-West neighborhoods of the city. South and West neighborhoods on the other hand are not well represented.

Stations along two train lines in the South and West neighborhoods do not have any Mediterranean restaurants in the vicinity. First is the Green line which covers both West and South neighborhoods and second is the Orange line which runs South-West towards Midway airport. Vicinity of any station on these lines outside downtown will be a good candidate to open a new Mediterranean restaurant.

Finally, if one is looking for a spot in the North neighborhoods then Blue line stations near O'Hare airport can be considered as they do not have Mediterranean restaurants on premise but there are competitors outside walking range.

Map below shows the areas identified enclosed in black lines.



5. Discussion

Chicago is a large city but density of overall existing restaurants is low outside the downtown area. Percentage of Mediterranean restaurants is miniscule when taking into account the population of the city.

While the limited concentration of Mediterranean restaurant made the analysis simple it is possible that Foursquare does not have ample data for the city of Chicago in general or specifically for Mediterranean restaurants. Additional data sources such as *google* or *yelp* can be used to compare and validate the results.

Another factor to consider is the population of city neighborhoods. Use census data to target demographic who would prefer to eat Mediterranean food.

6. Conclusion

Aim of this report was to identify candidate location for opening a Mediterranean restaurant in the city of Chicago. Existing Mediterranean restaurants were obtained using Fourquare API for Chicago neighborhoods defined by zip codes. Density of all restaurants was studies and percentage of Mediterranean restaurants was examined. A visual representation of existing Mediterranean restaurants was prepared and was superimposed with train stations locations. A heatmap was used to exclude city areas and neighborhoods already overcrowded. Finally, candidate locations were identified and grouped in three sets.

Final decision on restaurant location will be made by stakeholders based on characteristics of the local area. Other factors like population density, crime, parking, real estate availability, prices, social and economic dynamics of every neighborhood etc. should be considered.