

The Battle of Neighborhoods, Minneapolis: Latin American restaurant

Joao Vitor Coelho

November 18, 2020

1. Introduction

1.1. Background

The Minneapolis-St. Paul metropolitan area is the 16th largest in the United States, and a major financial and cultural center of the Upper Midwest region of the country. With more than 3.5 million people in the Metropolitan Statistical Area and a favorable economic environment, Minneapolis has seen a steady population growth in recent decades - and, just as other metropolitan areas of the United States, it has experienced a significant growth in its Hispanic/Latino population, which represents more than 10.5% of the population in Saint Paul and almost 9% of the population in Minneapolis. However, the metropolitan area still lags behind many of its American counterparts with respect to options of Latin American restaurants and other establishments that cater to Latin American consumers/

1.2. Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Latin American restaurant in Minneapolis-St. Paul, Minnesota.

Since there are lots of restaurants in Minneapolis and Saint Paul we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in

areas with no Latin American restaurants in the vicinity. We would also prefer locations as close to downtown Minneapolis as possible, assuming that the first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Interest

Business owners interested in opening a new Latin American restaurant in the Minneapolis-St. Paul area would be very interested in finding the optimal location for such a restaurant, in order to maximize the exposure to potential customers and, consequently, their revenue. An optimal location for a restaurant would also make it more accessible for their immediate suppliers and other supply chain operators.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Latin American restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use a regularly spaced grid of locations, centered around Downtown Minneapolis, to define our neighborhoods.

The following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinates of Downtown Minneapolis will be obtained using Google Maps API geocoding of well known Minneapolis location (Government Plaza)

Neighborhood candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approximately 20x20 kilometers centered around Downtown Minneapolis.

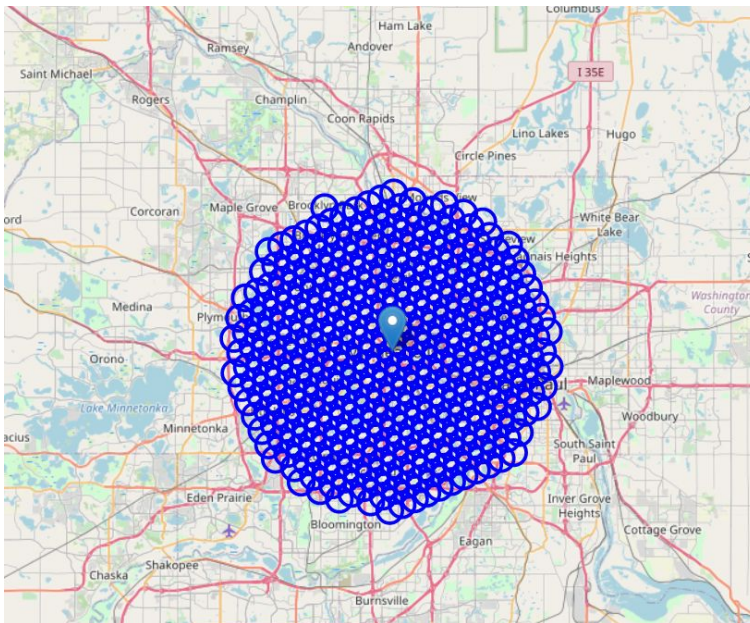
First, we need to find the latitude & longitude of Downtown Minneapolis, using a specific, well-known address and Google Maps geocoding API. Afterwards, we are able to create a grid of area candidates, equally spaced, centered around Downtown and within ~20km from Government Plaza. Our neighborhoods will be defined as circular areas with a radius of 1000 meters, so our neighborhood centers will be 2000 meters apart.

To accurately calculate distances we need to create our grid of locations in the Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on the Folium map. So let's create functions to convert between the WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Coordinate transformation check

```
-----
Downtown Minneapolis longitude=-93.2661600000001, latitude=44.9768693
Downtown Minneapolis UTM X=-4695922.652236367, Y=11936800.847093288
Downtown Minneapolis longitude=-93.26615999999993, latitude=44.97686930000011
```

Let's create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all its neighbors. Consequently, we can visualize the data we have so far, showing our central location and candidate neighborhood centers:



We now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to its neighbors is exactly the same) and within ~20km from Government Plaza. Let's now use Google Maps API to get approximate addresses of those locations.

```
[ '5100 E River Rd, Minneapolis, MN 55421',
  '6407 N Bryant Ave, Minneapolis, MN 55430',
  '7042 N Humboldt Ave, Minneapolis, MN 55430',
  '2024 Sugarloaf Trail, Minneapolis, MN 55444',
  '1502 Sibley Memorial Hwy, Mendota Heights, MN 55120',
  'MN-5, St Paul, MN 55111',
  'Hiawatha MDOT Trail, Minneapolis, MN 55417',
  '5305 Hwy 55, Minneapolis, MN 55417',
  '4508 43rd Ave S, Minneapolis, MN 55406',
  '3824 37th Ave S, Minneapolis, MN 55406',
  '3017 E 31st St, Minneapolis, MN 55406',
  '2500 East 25th Street, Minneapolis, MN 55406',
  '400 21st Ave S, Minneapolis, MN 55455',
  '420 SE Main St, Minneapolis, MN 55414',
  '316 5th Ave NE, Minneapolis, MN 55413',
  '1700 NE 2nd St, Minneapolis, MN 55413',
  'Heron Rookery, 28th Ave NE, Minneapolis, MN 55418',
  '4028 1/2 N Washington Ave, Minneapolis, MN 55412',
  '4800 N Bryant Ave, Minneapolis, MN 55430',
  '5446 N Humboldt Ave, Minneapolis, MN 55430']
```

Looking good. Let's now place all this into a Pandas dataframe.

	Address	Latitude	Longitude	X	Y	Distance from center
0	499 Kingston Ave E, St Paul, MN 55117	44.993960	-93.079906	-4.701923e+06	1.191775e+07	19974.984355
1	2155 Sloan St, Maplewood, MN 55117	45.006078	-93.087772	-4.699923e+06	1.191775e+07	19467.922334
2	128 Lake St, Little Canada, MN 55117	45.018199	-93.095643	-4.697923e+06	1.191775e+07	19157.244061
3	2946 Rice St, St Paul, MN 55113	45.030321	-93.103519	-4.695923e+06	1.191775e+07	19052.558883
4	266 Owasso Blvd N, St Paul, MN 55126	45.042446	-93.111400	-4.693923e+06	1.191775e+07	19157.244061
5	519 Shoreview Park Rd, Shoreview, MN 55126	45.054572	-93.119286	-4.691923e+06	1.191775e+07	19467.922334
6	4140 Reiland Ln, Shoreview, MN 55126	45.066701	-93.127177	-4.689923e+06	1.191775e+07	19974.984355
7	775 Bradley St, St Paul, MN 55130	44.964899	-93.078965	-4.705923e+06	1.191948e+07	20000.000000
8	1181 Westminster St, St Paul, MN 55130	44.977011	-93.086827	-4.703923e+06	1.191948e+07	19078.784028
9	108a Biglow Ln, St Paul, MN 55117	44.989126	-93.094693	-4.701923e+06	1.191948e+07	18330.302780

The last step for now persist this data into a local file.

Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

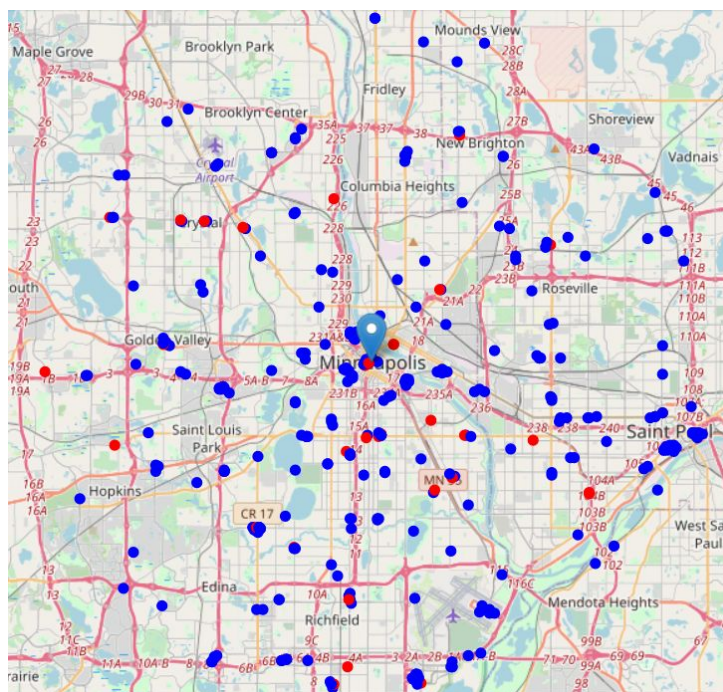
We're interested in venues in the 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those.

So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Latin American Restaurant' category, as we need info on Latin American restaurants in the neighborhood.

Transforming our Foursquare API restaurant data into numpy, we can generate the following list of restaurants:

```
Total number of restaurants: 372
Total number of Latin American restaurants: 44
Percentage of Latin American restaurants: 11.83%
Average number of restaurants in neighborhood: 1.0163934426229508
```

As we can see, almost 12% of the restaurants of our list using Foursquare data are Latin American restaurants. After gathering all the restaurant data, we are able to create a Folium map in order to see all the collected restaurants in our area of interest, while also showing Latin American restaurants in different colors.



Looking good. So now we have all the restaurants in an area within a few kilometers from Government Plaza, and we know which ones are Latin American restaurants! We also know which restaurants exactly are in the vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Latin American restaurant!

3. Methodology

In this project we will direct our efforts on detecting areas of the Minneapolis-St. Paul region that have a low restaurant density, particularly those with low numbers of Latin American restaurants. We will limit our analysis to areas ~20km around Downtown Minneapolis.

In the first step, we have collected the required data: location and type (category) of every restaurant within 20km from Downtown Minneapolis (Government Plaza). We have also identified Latin American restaurants (according to Foursquare categorization).

The second step in our analysis will be calculation and exploration of “restaurant density” across different areas of the Twin Cities region - we will use heat maps to identify a few promising areas close to center with low number of restaurants in general (and no Latin American restaurants in vicinity) and focus our attention on those areas.

In the third and final step, we will focus on the most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Latin American restaurants in radius of 400

meters. We will present a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final “street level” exploration and search for optimal venue location by stakeholders.

4. Exploratory Analysis

Let's perform some basic exploratory data analysis and derive some additional info from our raw data. First let's count the number of restaurants in every area candidate:

Average number of restaurants in every area with radius=300m: 1.0163934426229508

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	499 Kingston Ave E, St Paul, MN 55117	44.993960	-93.079906	-4.701923e+06	1.191775e+07	19974.984355	0
1	2155 Sloan St, Maplewood, MN 55117	45.006078	-93.087772	-4.699923e+06	1.191775e+07	19467.922334	0
2	128 Lake St, Little Canada, MN 55117	45.018199	-93.095643	-4.697923e+06	1.191775e+07	19157.244061	1
3	2946 Rice St, St Paul, MN 55113	45.030321	-93.103519	-4.695923e+06	1.191775e+07	19052.558883	3
4	266 Owasso Blvd N, St Paul, MN 55126	45.042446	-93.111400	-4.693923e+06	1.191775e+07	19157.244061	1
5	519 Shoreview Park Rd, Shoreview, MN 55126	45.054572	-93.119286	-4.691923e+06	1.191775e+07	19467.922334	0
6	4140 Reiland Ln, Shoreview, MN 55126	45.066701	-93.127177	-4.689923e+06	1.191775e+07	19974.984355	0
7	775 Bradley St, St Paul, MN 55130	44.964899	-93.078965	-4.705923e+06	1.191948e+07	20000.000000	0
8	1181 Westminster St, St Paul, MN 55130	44.977011	-93.086827	-4.703923e+06	1.191948e+07	19078.784028	0
9	108a Biglow Ln, St Paul, MN 55117	44.989126	-93.094693	-4.701923e+06	1.191948e+07	18330.302780	0

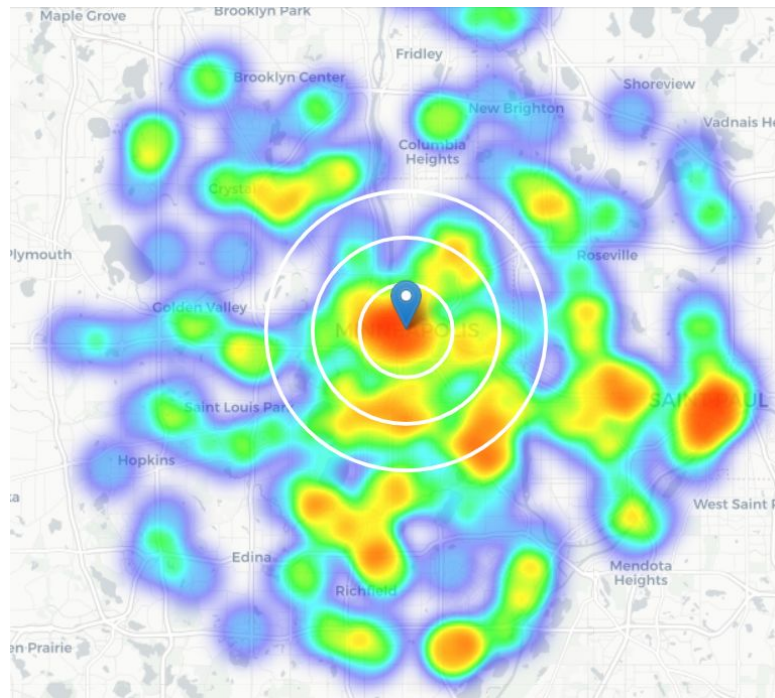
At this point, we are able to calculate the distance to the nearest Latin American restaurant from every area candidate center (not only those within 300m - we want the distance to the closest one, regardless of how distant it is).

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Latin American restaurant
0	499 Kingston Ave E, St Paul, MN 55117	44.993960	-93.079906	-4.701923e+06	1.191775e+07	19974.984355	0	5751.914871
1	2155 Sloan St, Maplewood, MN 55117	45.006078	-93.087772	-4.699923e+06	1.191775e+07	19467.922334	0	3753.030584
2	128 Lake St, Little Canada, MN 55117	45.018199	-93.095643	-4.697923e+06	1.191775e+07	19157.244061	1	1756.687918
3	2946 Rice St, St Paul, MN 55113	45.030321	-93.103519	-4.695923e+06	1.191775e+07	19052.558883	3	294.391444
4	266 Owasso Blvd N, St Paul, MN 55126	45.042446	-93.111400	-4.693923e+06	1.191775e+07	19157.244061	1	2255.522157
5	519 Shoreview Park Rd, Shoreview, MN 55126	45.054572	-93.119286	-4.691923e+06	1.191775e+07	19467.922334	0	4253.010003
6	4140 Reiland Ln, Shoreview, MN 55126	45.066701	-93.127177	-4.689923e+06	1.191775e+07	19974.984355	0	6252.104283
7	775 Bradley St, St Paul, MN 55130	44.964899	-93.078965	-4.705923e+06	1.191948e+07	20000.000000	0	3076.635592
8	1181 Westminster St, St Paul, MN 55130	44.977011	-93.086827	-4.703923e+06	1.191948e+07	19078.784028	0	3560.919969
9	108a Biglow Ln, St Paul, MN 55117	44.989126	-93.094693	-4.701923e+06	1.191948e+07	18330.302780	0	4888.211890

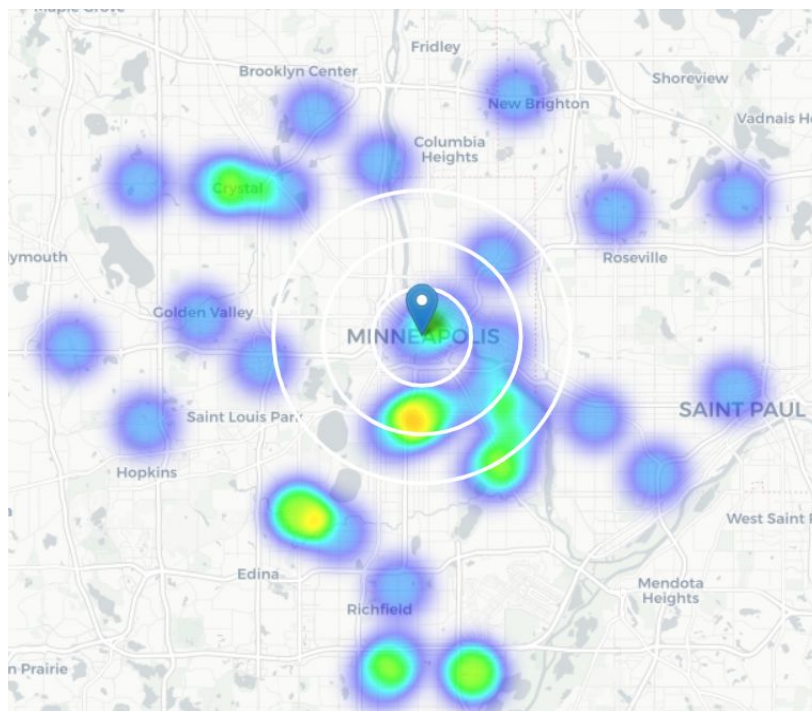
```
print('Average distance to closest Latin American restaurant from each area center:', df_locations['Distance to Latin American restaurant'].mean())
```

Average distance to closest Latin American restaurant from each area center: 3184.0708926606726

On average, a Latin American restaurant can be found within ~3200m from every area center candidate. There is a significant distance between each center candidate, but we still need to filter our areas more carefully. After this, we are able to generate a heat map using Folium showing the areas with the highest density of restaurants.



Looks like there are a few pockets of low restaurant density closest to Downtown Minneapolis. At this point, we are able to generate another heat map, now showing heat map/density of Latin American restaurants only.



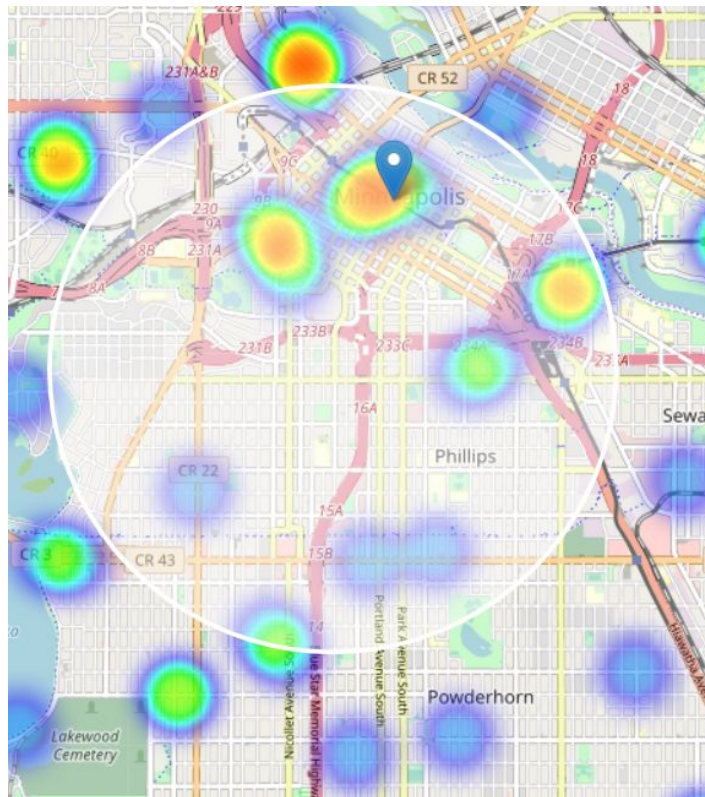
This map is not as hot as the previous one (Latin American restaurants represent a subset of ~12% of all restaurants in the Twin Cities) but it also indicates higher density of existing Latin American restaurants directly north and west from Government Plaza, with closest pockets of low Latin American restaurant density positioned east, south-east and south from city center.

Based on this we will now focus our analysis on areas south-west, south, south-east and east from Downtown Minneapolis - we will move the center of our area of interest and reduce its size to have a radius of 2.5km. This places our location candidates mostly in the Downtown West/Loring Park and Downtown East parts of Minneapolis, which are more interesting to stakeholders given their central location in the city and their mixed business-residential compositions. It is also important to point out that these areas concentrate important attractions of the city for both tourists and locals.

Downtown West/Loring Park and Downtown East

In addition to being neighborhoods that are close to the central business district of Minneapolis, these areas concentrate a number of important city attractions for both tourists and locals, such as the Minneapolis Convention Center, Loring Park, Gold Medal Park, US Bank Stadium and the University of Minnesota.

Let's define a new, more narrow region of interest, which will include low-restaurant-count parts of Downtown West/Loring Park and Downtown East that are closer to Government Plaza. With that region in mind, we can generate another heat map highlighting our focus region:



Not bad - this nicely covers all pockets of low restaurant density in Downtown West/Loring Park, South Minneapolis and Downtown East that are close to Downtown Minneapolis.

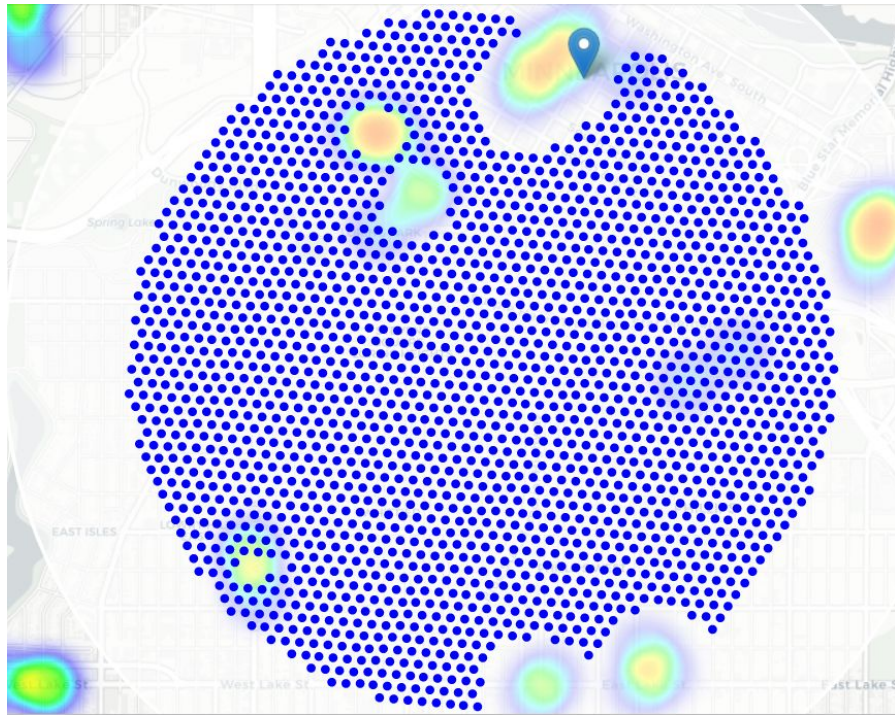
Let's also create a new, more dense grid of candidate locations restricted to our new region of interest (let's make our location candidates 100m apart). With 2261 candidate neighborhood centers generated, we can now calculate two most important things for each location candidate: number of restaurants in the vicinity (we'll use a radius of 250 meters) and distance to the closest Latin American restaurant.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Latin American restaurant
0	44.970308	-93.251486	-4.697473e+06	1.193580e+07	0	2003.007087
1	44.970913	-93.251882	-4.697373e+06	1.193580e+07	0	1921.399443
2	44.966736	-93.250043	-4.698023e+06	1.193589e+07	0	2437.645937
3	44.967341	-93.250439	-4.697923e+06	1.193589e+07	0	2347.734257
4	44.967946	-93.250836	-4.697823e+06	1.193589e+07	0	2258.670974
5	44.968551	-93.251232	-4.697723e+06	1.193589e+07	0	2170.560526
6	44.969156	-93.251629	-4.697623e+06	1.193589e+07	0	2083.523800
7	44.969761	-93.252026	-4.697523e+06	1.193589e+07	0	1997.701142
8	44.970366	-93.252422	-4.697423e+06	1.193589e+07	0	1913.255937
9	44.970971	-93.252819	-4.697323e+06	1.193589e+07	0	1830.378843

Let us now filter those locations: we're interested only in locations with no more than two restaurants in a radius of 250 meters, and with no Latin American restaurants in a radius of 400 meters.

```
Locations with no more than two restaurants nearby: 2161
Locations with no Latin American restaurants within 400m: 2101
Locations with both conditions met: 2058
```

We are now able to see how this would look on a map.

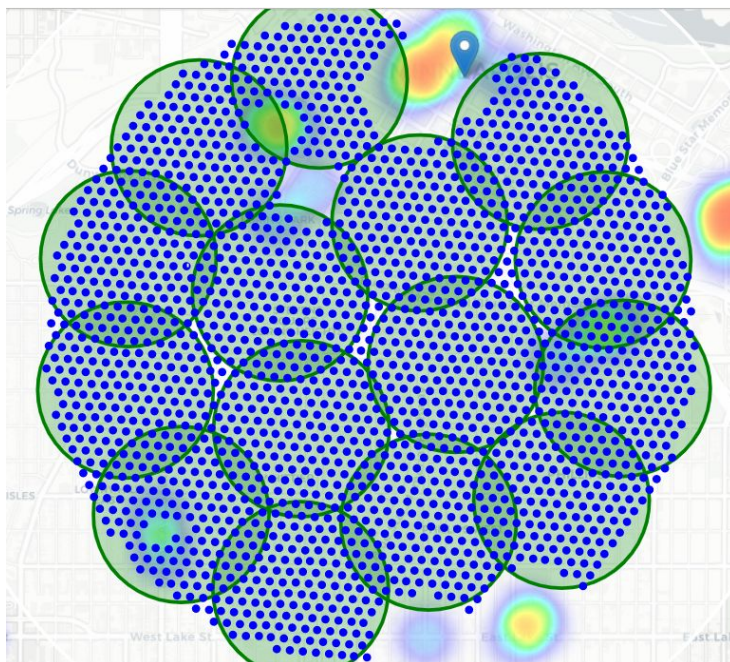


Looking good. We now have a bunch of locations fairly close to Government Plaza (mostly in Downtown West/Loring Park, Downtown East and South Minneapolis), and we know that each of those locations has no more than two restaurants in a radius of 250 meters, and no Latin American restaurant closer than 400 meters. Any of those locations is a potential candidate for a new Latin American restaurant, at least based on nearby competition.

We are able to show those good locations in a form of heat map on Folium:

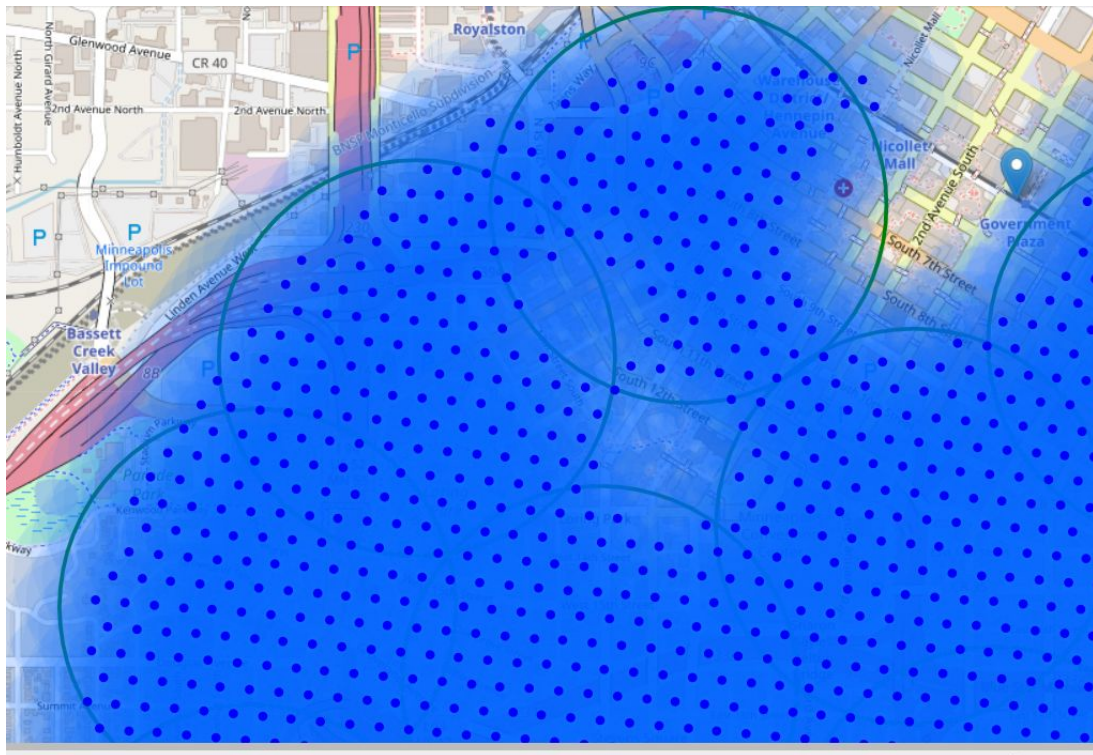


Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and no Latin American restaurants at all nearby. We can now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.



Our clusters represent groupings of most of the candidate locations and their cluster centers are placed nicely in the middle of the zones full with candidate locations. Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find out the best possible locations based on neighborhood specifics. Let's see those zones on a city map without a heat map, using shaded areas to indicate our clusters:

Let's zoom in on candidate areas in Downtown West/Loring Park:



We can now do the same for and candidate areas in Downtown East:

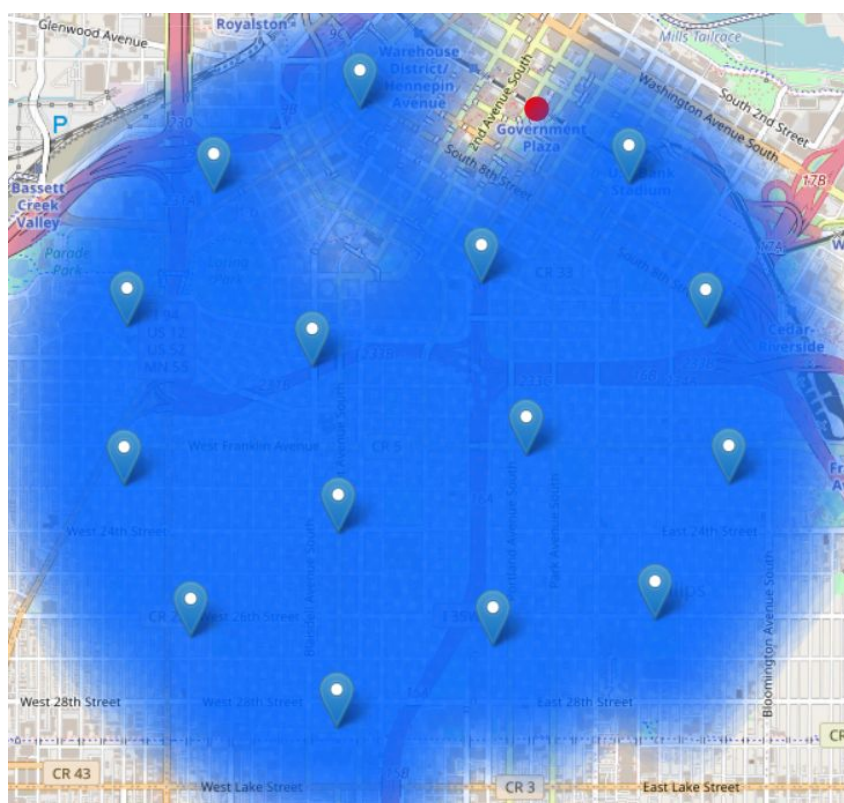


Finally, let's reverse geocode those candidate area centers to get the addresses which can be presented to stakeholders.

```
=====
Addresses of centers of areas recommended for further analysis
=====
```

824 Hennepin Ave, Minneapolis, MN 55403	=> 1.1km from Government Plaza
611 E Franklin Ave, Minneapolis, MN 55404	=> 2.2km from Government Plaza
2623 Garfield Ave, Minneapolis, MN 55408	=> 4.0km from Government Plaza
101 1/2 Oak Grove St, Minneapolis, MN 55403	=> 2.2km from Government Plaza
1012 E 26th St, Minneapolis, MN 55404	=> 3.3km from Government Plaza
1231 S 9th St, Minneapolis, MN 55404	=> 1.7km from Government Plaza
900 W 22nd St, Minneapolis, MN 55405	=> 3.5km from Government Plaza
1101 S 4th Ave, Minneapolis, MN 55404	=> 1.2km from Government Plaza
24th St W & Nicollet Ave S, Minneapolis, MN 55404	=> 3.0km from Government Plaza
2837 Nicollet Ave, Minneapolis, MN 55408	=> 4.1km from Government Plaza
2121 14th Ave S, Minneapolis, MN 55404	=> 2.7km from Government Plaza
530 Chicago Ave, Minneapolis, MN 55415	=> 0.8km from Government Plaza
2637 S 5th Ave, Minneapolis, MN 55408	=> 3.4km from Government Plaza
27 Groveland Terrace, Minneapolis, MN 55403	=> 2.9km from Government Plaza
15 16th St N, Minneapolis, MN 55403	=> 2.1km from Government Plaza

This concludes our analysis. We have created 15 addresses representing centers of clusters containing locations with low number of restaurants and no Latin American restaurants nearby - all zones being fairly close to Downtown Minneapolis (all less than 5km from Government Plaza, and four of those less than 2km from Government Plaza). Although zones are shown on map with a radius of ~500 meters (green circles), their shapes are actually very irregular and their centers and addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Downtown West/Loring Park and Downtown East, which we have identified as interesting due to being popular with locals and tourists alike, fairly close to downtown and well connected by public transport.



5. Results and Discussion

Our analysis shows that although there is a significant number of restaurants in Minneapolis-St. Paul (in our initial area of interest which had a radius of 20 kilometers around Government Plaza), there are pockets of low restaurant density fairly close to Downtown Minneapolis. Highest concentration of restaurants was detected south and northeast from Government Plaza, so we focused our attention to areas southwest and east, corresponding to Downtown West/Loring Park and Downtown East. Our attention was focused on Downtown West/Loring Park and Downtown East because these two areas offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approximately 5 kilometers around Government Plaza, especially in southwestern areas) we first created a dense grid of location candidates (spaced 100 meters apart); those locations were then filtered, so that those with more than two restaurants in a radius of 250 meters and those with an Latin American restaurant closer than 400 meters were removed.

These location candidates were then clustered to create zones of interest, which contained the greatest number of location candidates. Addresses of the centers of these zones were also generated using reverse geocoding to be used as markers or starting points, in order to get a more detailed local analysis based on other factors.

The result of all this is a list of 15 zones, containing the largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and particularly Latin American restaurants. This, of course, does not imply that these

zones are actually optimal locations for a new restaurant: the purpose of this analysis was only to provide info about areas close to Downtown Minneapolis but not crowded with existing restaurants (particularly Latin American ones); it is possible that there is a very good reason for the small number of restaurants in any of these areas, especially reasons which would make them unsuitable for a new restaurant regardless of the lack of competition in the area. Therefore, recommended zones should be considered only as a starting point for more detailed analysis which could eventually result in an optimal location, which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6. Conclusion

The purpose of this project was to identify areas in the Minneapolis-St. Paul metropolitan area that were close to Downtown Minneapolis with a low number of restaurants (especially Latin American restaurants) in order to help stakeholders in narrowing down the search for optimal location for a new Latin American restaurant in the area. By calculating restaurant density distribution from Foursquare API data, we have first identified parts of the metropolitan area that could justify further analysis (Downtown West/Loring Park and Downtown East), and then generated an extensive collection of locations which satisfied some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing the greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for a final exploration by stakeholders.

The final decision on the optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to parks or bodies water), levels of noise and proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood, among other driving factors that can help stakeholders to determine a new business location.