Lebanese restaurant in Sydney

1. Introduction

1.1. Background

Lebanese cuisine is an ancient one and part of the cuisine of the Levant. Many dishes in Lebanese cuisine can be traced back thousands of years and were influenced by different foreign civilizations that held power each in an era where the most recent one was the French after World War 1, therefore it is loved and intended by a lot of people from different cultures and nationalities.

Sydney is a multinational city which has a lot of Lebanese people residing in it, therefore an idea of a Lebanese restaurant can be very appealing and profitable for any businessman with an investing mind.

1.2. Problem

The challenge here is to find the most suitable location for this restaurant to be successful and profitable. A bad location choice can be crucial and lead to the restaurant closing, therefore I am going to use machine learning and data science and study different neighborhoods and venues in order to get the best location possible.

2. Data

2.1. Data Acquisition

Our data is very straight forward, we only need to define our center of interest and the rest will be calculated and drawn.

Our center location is somewhere in the crowded area of the city center.
Once it is defined, all we need is Google Maps API to get the Latitude and Longitude of different locations. We also need the number of restaurants and their type and location in every neighborhood which will be obtained using Foursquare API. For this we need to do some calculations and conversions between Cartesian coordinates and X/Y coordinates in meters.

2.2. Neighborhood Candidates

We create a grid of area candidates, equally spaced, centered around city center and within ~6km from Darling Harbor. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in 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 Folium map.

We can remove candidates that happen to be above the water or included in a big park or over any dead area for better processing time. Our final number of candidates will be 252.

Here is a list of addresses for centers of our candidates ranged from 100 to 120:

```
[8]: addresses[100:120]
```

```
[8]: 112                      2 Harris St, Pyrmont NSW 2009
     113          White Bay Cruise Terminal, Balmain NSW 2041
     114                  9/25-29 Palmer St, Balmain NSW 2041
     115                       35 Arthur St, Balmain NSW 2041
     116       Paringa Reserve, Elliott St, Balmain NSW 2041
     117    Peppercorn Reserve, St Georges Cres, Drummoyne...
     118              76A St Georges Cres, Drummoyne NSW 2047
     119                 2 Lawson St, Bondi Junction NSW 2022
     120              5R Oxford St, Centennial Park NSW 2021
     121                     66 Queen St, Woollahra NSW 2025
     122                  354 Oxford St, Paddington NSW 2021
     123                    51 Gipps St, Paddington NSW 2021
     124    Oxford St after South Dowling St, Darlinghurst...
     125                   13 Waine St, Surry Hills NSW 2010
     126    Suite 308, Level 3, Fayworth House, 379 - 383 ...
     127                        145N Day St, Sydney NSW 2000
     128                 120-140 Pyrmont St, Pyrmont NSW 2009
     129                       33A John St, Pyrmont NSW 2009
     130       Glebe Island Bridge, Bank St, Pyrmont NSW 2009
     131                       Port Access Rd, Rozelle NSW 2039
     Name: Address, dtype: object
```

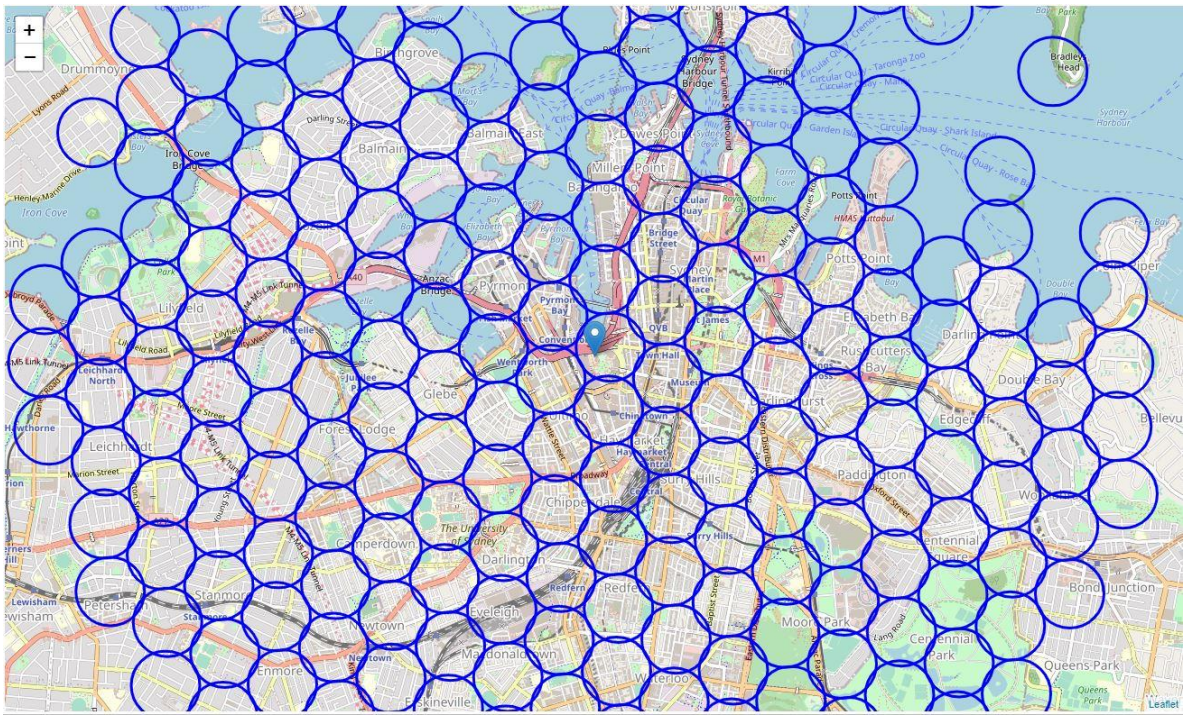
So now we have all the data needed to continue with our study and our data-frame looks like this:

```
[9]: df_locations.head(10)
```

[9]:

	Address	Latitude	Longitude	X	Y	Distance from center
0	Unnamed Road, Mosman NSW 2088	-33.844531	151.238421	4.675661e+06	-1.524362e+07	5909.949238
1	12 Musgrave St, Mosman NSW 2088	-33.842097	151.232963	4.676361e+06	-1.524362e+07	5735.634228
2	121 Cremorne Rd, Cremorne Point NSW 2090	-33.839662	151.227505	4.677061e+06	-1.524362e+07	5643.358929
3	59 Wycombe Rd, Neutral Bay NSW 2089	-33.837227	151.222049	4.677761e+06	-1.524362e+07	5637.153537
4	8 Premier St, Neutral Bay NSW 2089	-33.834792	151.216593	4.678461e+06	-1.524362e+07	5717.298313
5	St Leonards Park, North Sydney NSW 2060	-33.832356	151.211137	4.679161e+06	-1.524362e+07	5880.263600
6	Bradleys Head Walking Track, Mosman NSW 2088	-33.852302	151.244472	4.674561e+06	-1.524301e+07	5800.000000
8	2 Milson Rd, Cremorne Point NSW 2090	-33.847432	151.233554	4.675961e+06	-1.524301e+07	5242.136969
9	26 Milson Rd, Cremorne Point NSW 2090	-33.844997	151.228096	4.676661e+06	-1.524301e+07	5086.255990
10	158A Kurraba Rd, Kurraba Point NSW 2089	-33.842561	151.222639	4.677361e+06	-1.524301e+07	5023.942675

And here is our grid with all our neighborhoods in circular shape surrounding Darling Harbor

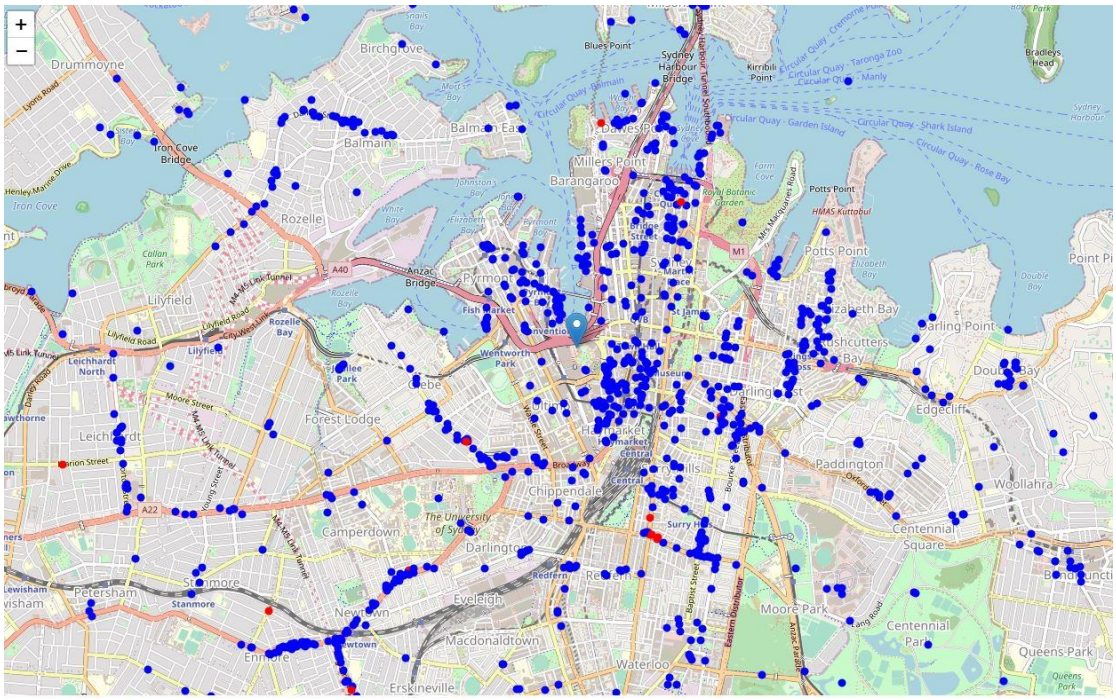


Now that we have our location candidates, we use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in '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 'Lebanese restaurant' category, as we need info on Lebanese restaurants in the neighborhood.

We have a total of 1050 restaurants which only 17 are Lebanese restaurant or 1.62%.

Here we see all the collected restaurants in our area of interest on map as blue dots and also the Lebanese restaurants in Red.



3. Methodology

3.1. Concept

We will direct our efforts on detecting areas of Sydney that have average restaurant density, particularly those with low number of Lebanese restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from Sydney center (Darling Harbor). We have also identified Lebanese restaurants (according to foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Sydney - we will use heat-maps to identify a few promising areas close to center with low number of restaurants in general (and no Lebanese restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on 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 less than two and no more than four restaurants in radius of 250 meters, and we want locations without Lebanese restaurants in radius of 1000 meters. We will present 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.

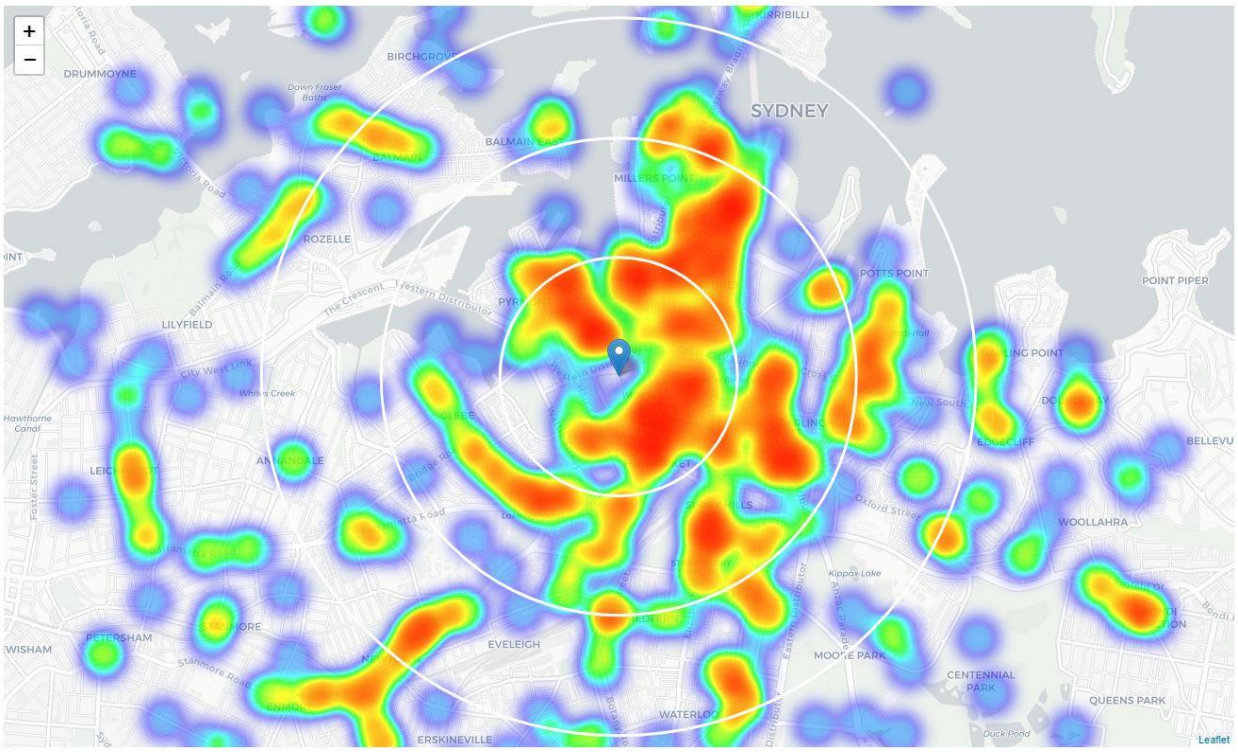
3.2. Analysis

For this step we need to calculate the number of restaurants in each area, and the distance to the nearest Lebanese restaurant and add them as columns to our dataframe to have something like this:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Lebanese restaurant
0	Unnamed Road, Mosman NSW 2088	-33.844531	151.238421	4.675661e+06	-1.524362e+07	5909.949238	1	3995.819017
1	12 Musgrave St, Mosman NSW 2088	-33.842097	151.232963	4.676361e+06	-1.524362e+07	5735.634228	0	3781.408136
2	121 Cremorne Rd, Cremorne Point NSW 2090	-33.839662	151.227505	4.677061e+06	-1.524362e+07	5643.358929	0	3589.414193
3	59 Wycombe Rd, Neutral Bay NSW 2089	-33.837227	151.222049	4.677761e+06	-1.524362e+07	5637.153537	2	3394.704188
4	8 Premier St, Neutral Bay NSW 2089	-33.834792	151.216593	4.678461e+06	-1.524362e+07	5717.298313	0	3338.283809
5	St Leonards Park, North Sydney NSW 2060	-33.832356	151.211137	4.679161e+06	-1.524362e+07	5880.263600	2	3426.990088
6	Bradleys Head Walking Track, Mosman NSW 2088	-33.852302	151.244472	4.674561e+06	-1.524301e+07	5800.000000	0	4057.659163
8	2 Milson Rd, Cremorne Point NSW 2090	-33.847432	151.233554	4.675961e+06	-1.524301e+07	5242.136969	0	3321.077179
9	26 Milson Rd, Cremorne Point NSW 2090	-33.844997	151.228096	4.676661e+06	-1.524301e+07	5086.255990	1	3127.623939
10	158A Kurraba Rd, Kurraba Point NSW 2089	-33.842561	151.222639	4.677361e+06	-1.524301e+07	5023.942675	1	2915.863374

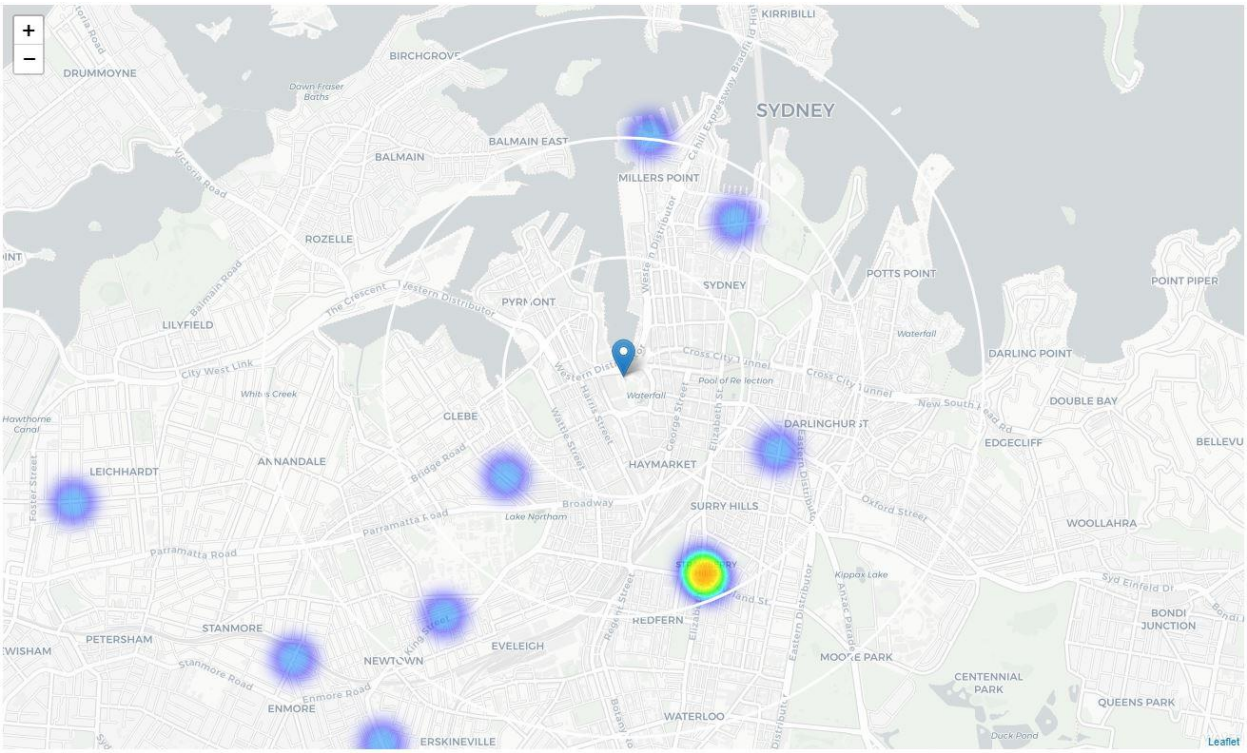
The average distance to closest Lebanese restaurant from each area center is a shy of 2000m. That is not so close which can help us get more promising candidates.

Let's take a look at the figure showing heat-map / density of restaurants and try to extract some meaningful info from that. Also, let's show a few circles indicating distance of 1km, 2km and 3km from Darling Harbor.



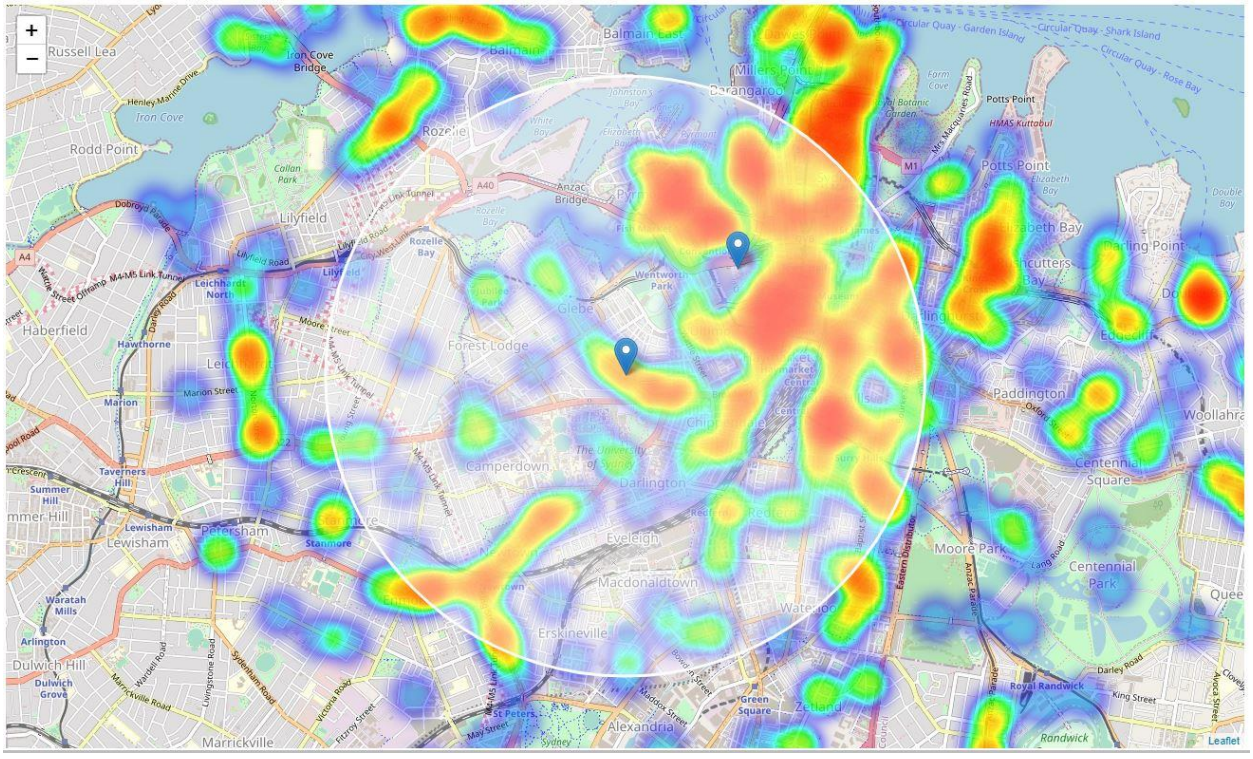
Looks like a few pockets of low restaurant density closest to city center can be found north, south, south-west and west from Darling Harbor.

Here is another heat-map showing heat-map/density of Lebanese restaurants only.



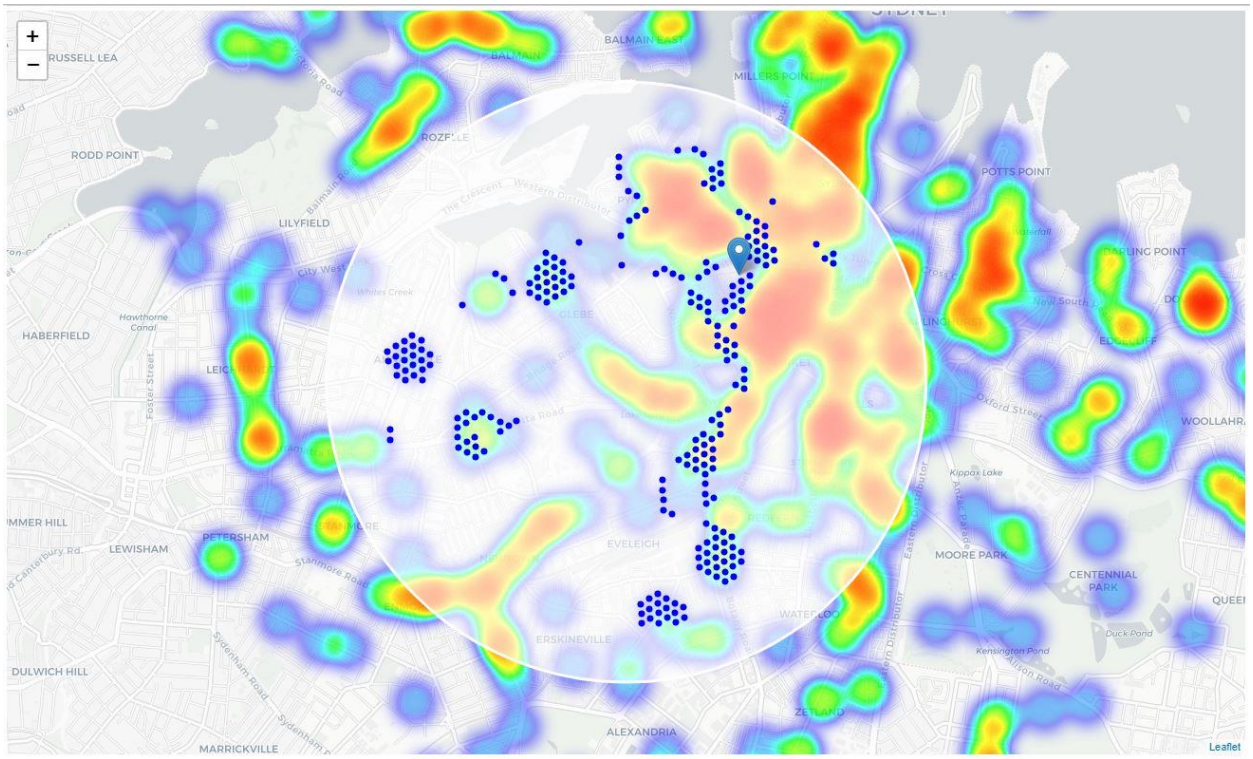
As expected this map is very clear (Lebanese restaurants represent a subset of ~2% of all restaurants in Sydney) but we can see that Lebanese restaurants are scattered around Darling Harbor with no specific direction.

Knowing that Parramatta and the area South West to Darling Harbor has the biggest Lebanese community in Sydney, we can move the center of our area of interest south west of Darling Harbor and reduce its size to have a radius of 2.5km to have the following map



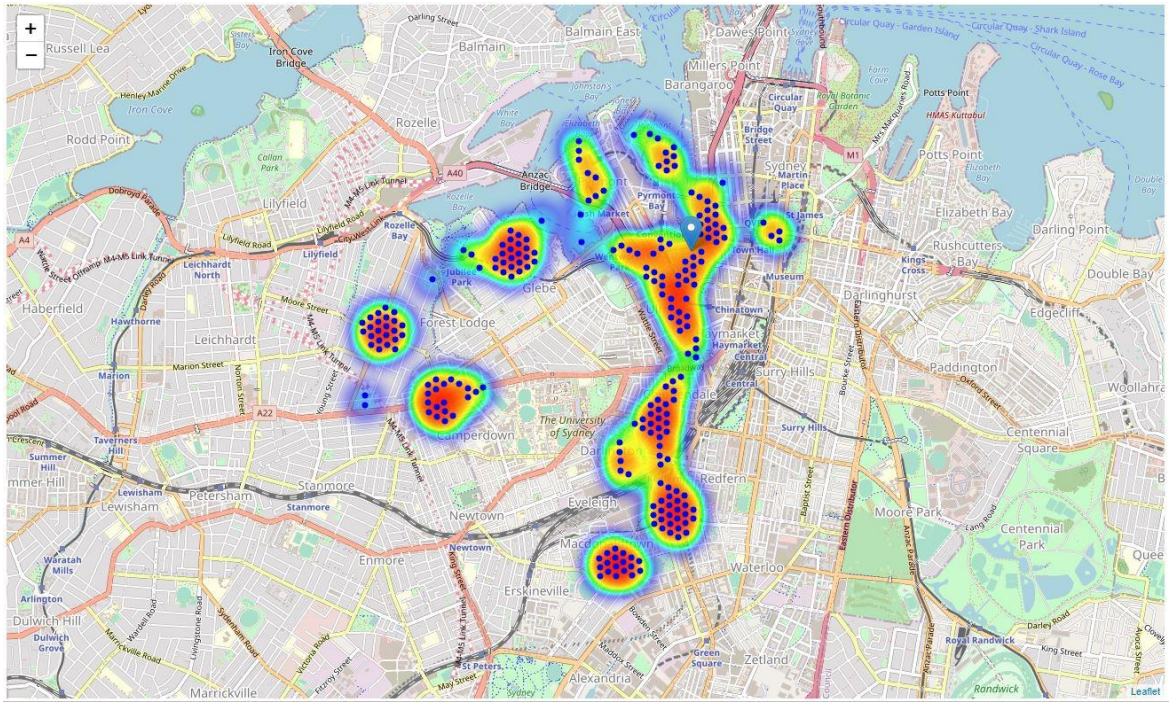
This is our new area of interest, our final candidates should be close to an area where there are at least 2 restaurants and at most 4. This way we know that our area contains restaurants and thus it is not a dead or deserted area on one hand and on the other hand it is not crowded with restaurants.

To show our potential restaurants location we created a denser grid of locations, 100m apart from each other and restricted to our new area of interest. Here is how it looks like on a map:



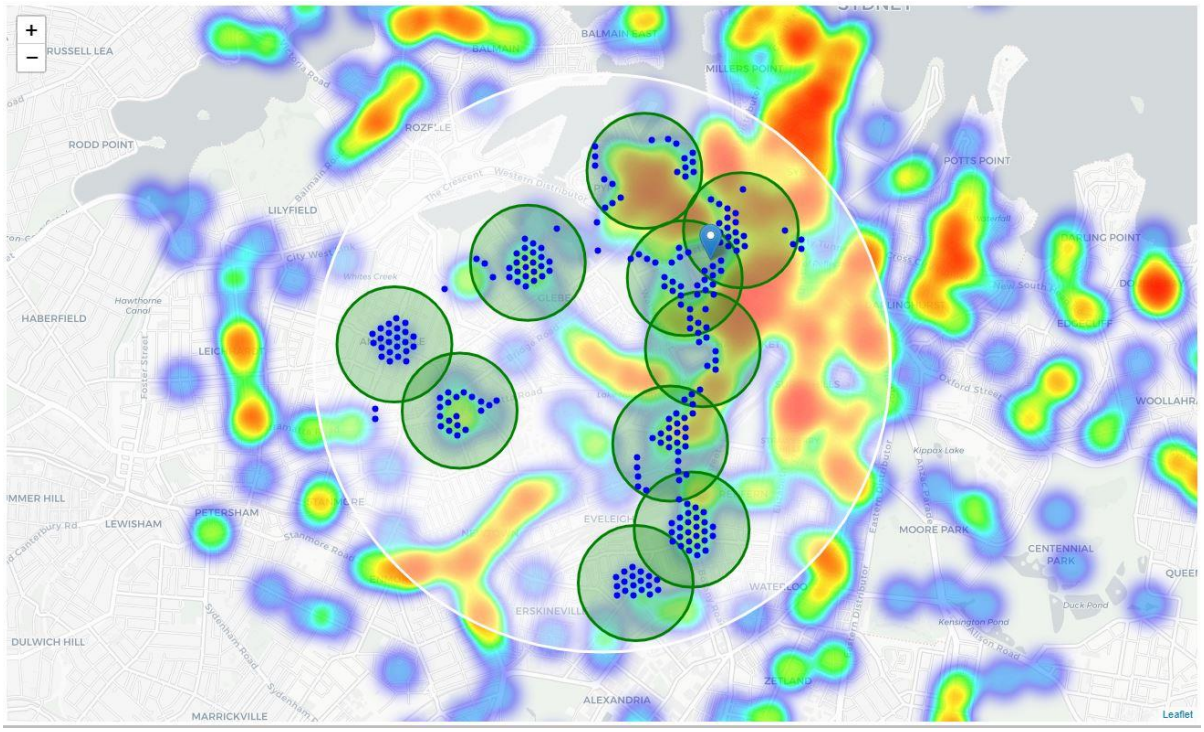
We now have a bunch of locations fairly close to Darling Harbor and we know that each of those locations has no more than four restaurants in radius of 250m, and no Lebanese restaurant closer than 1000m. Any of those locations is a potential candidate for a new Lebanese restaurant, at least based on nearby competition.

Let's now show those good locations in a form of heat-map:

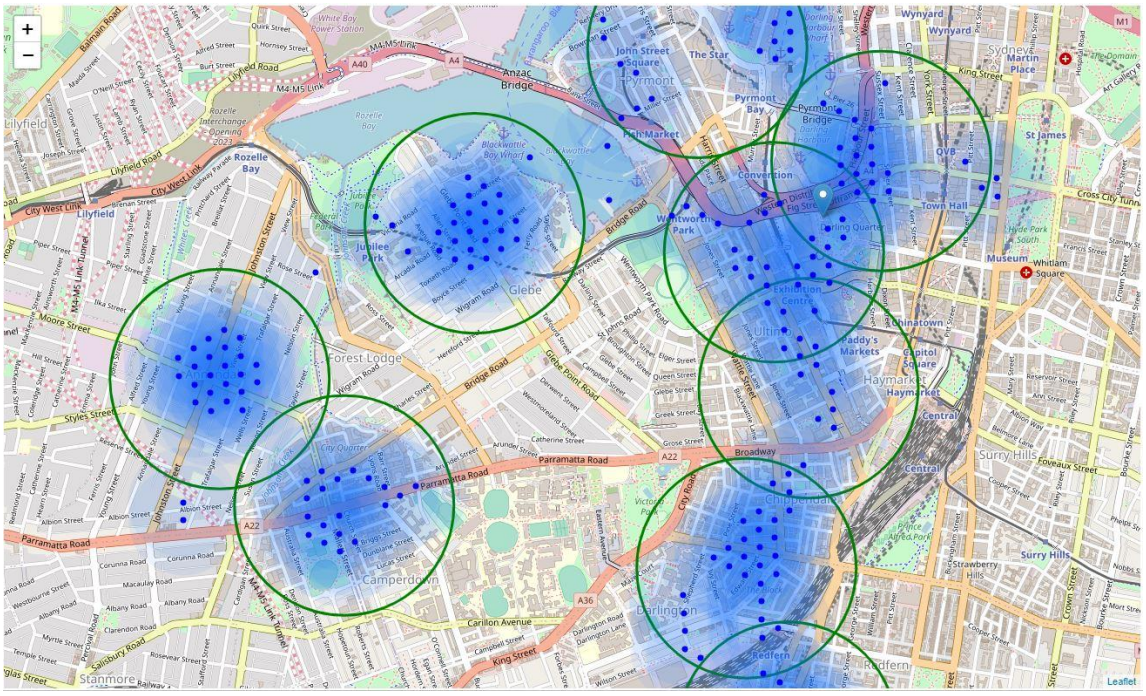


What we have now is a clear indication of zones with low number of restaurants in vicinity, and no Lebanese restaurants at all nearby.

Let us 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. A choice of 10 clusters will be a good one since we can visually divide them into 10 areas. Let's see how it will look like on a map



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



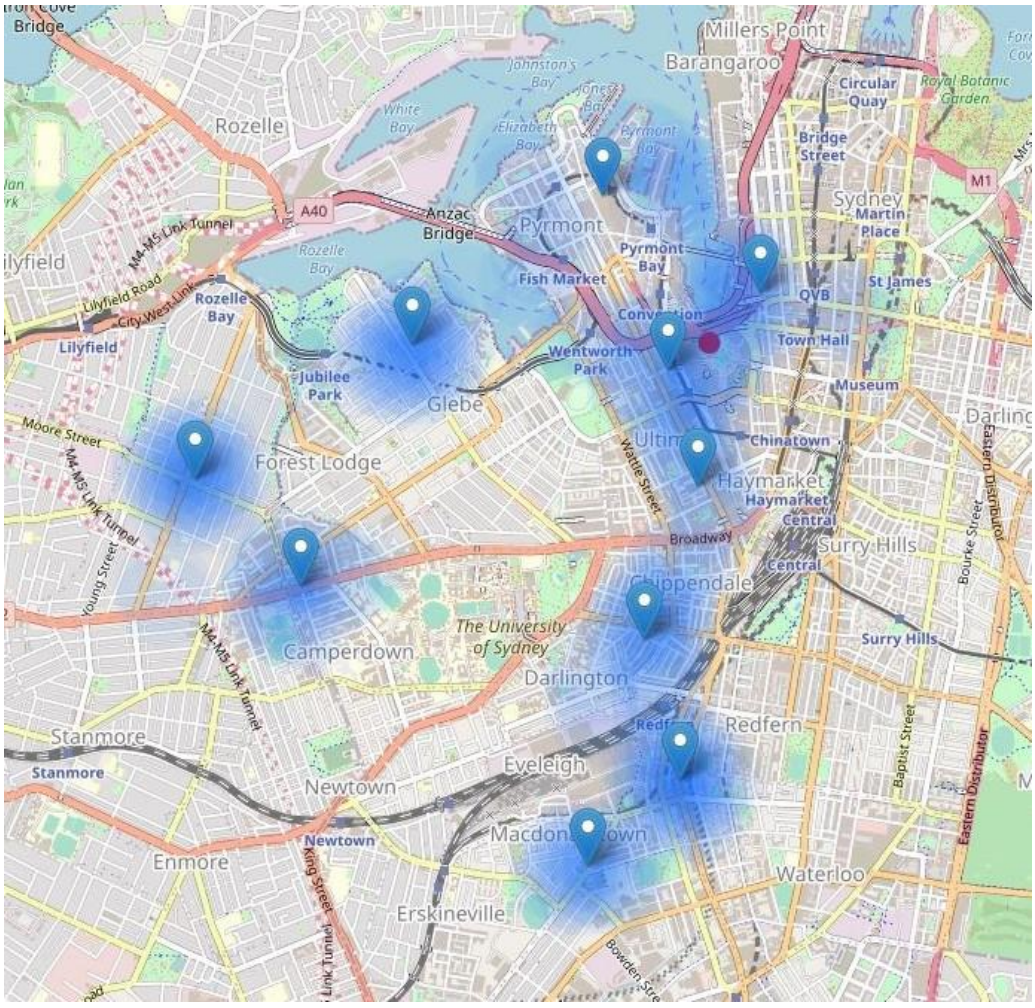
Now we have points with altitudes and longitudes, we can use reverse geocoding to get the real addresses which can be presented to stakeholders.

Here's a list of those addresses

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

140 A22, Annandale NSW 2038                          => 3.1km from Darling Harbor
27 Vine St, Darlington NSW 2008                      => 2.0km from Darling Harbor
Levels 20 & 21 Tower 2 Darling Park 201 Sussex Street, Sydney NSW 2000 => 0.5km from Darling Harbor
397 Glebe Point Rd, Glebe NSW 2037                   => 1.9km from Darling Harbor
77 Mitchell Rd, Alexandria NSW 2015                  => 3.5km from Darling Harbor
113 Booth St, Annandale NSW 2038                     => 3.5km from Darling Harbor
111A Quarry St, Ultimo NSW 2007                      => 0.3km from Darling Harbor
80 Pyrmont St, Pyrmont NSW 2009                      => 1.2km from Darling Harbor
23 Botany Rd, Waterloo NSW 2017                      => 2.8km from Darling Harbor
731 Harris St, Ultimo NSW 2007                       => 0.9km from Darling Harbor
```

This concludes our analysis. We have created 10 addresses representing centers of zones containing locations with low number of restaurants and no Lebanese restaurants nearby, all zones being fairly close to city center (all less than 4km from Darling Harbor, and about half of those less than 2km from Darling Harbor). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.



4. Results and Discussion

Our analysis shows that although there is a great number of restaurants in Sydney (1000+ in our initial area of interest which was 12x12km around Darling Harbor), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north, west and south from Daring Harbor, so we focused our attention to areas in south-west, corresponding to areas near Parramatta road. Our areas of interest offer a combination of popularity among Lebanese, 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 approx. 5x5km south-west from Darling Harbor) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than four restaurants in radius of 250m and those with a Lebanese restaurant closer than 1000m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

5. Conclusion

Purpose of this project was to identify Sydney areas close to center with low number of restaurants (particularly Lebanese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Lebanese restaurant. By calculating restaurant density distribution from foursquare data we have generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on 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 park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.