

Applied Data Science

Coursera Capstone project- The Battle of the Neighborhoods

Table of Contents

<u>INTRODUCTION: BUSINESS PROBLEM</u>	1
<u>DATA</u>	2
NEIGHBORHOOD CANDIDATES	2
FOURSQUARE	3
<u>METHODOLOGY</u>	4
<u>ANALYSIS</u>	5
BUDERIM AND SIPPY DOWNS	9
RESULTS AND DISCUSSION	12
CONCLUSION	13
Figure 1 Candidate neighborhood centers	3
Figure 2 Location of all restaurants	4
Figure 4 Heatmap of restaurants.....	7
Figure 5 Heatmap of Japanese restaurants.....	8
Figure 6 Japanese restaurants density	9
Figure 7 Good location for Japanese restaurant	10
Figure 8 Zones of good location	11
Figure 9 Addresses of center clusters.....	12

Introduction: 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 a **Japanese restaurant** in **Sunshine Coast**, Queensland, Australia.

Since there are lots of restaurants in Sunshine Coast, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

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

Data

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 Italian restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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**
- coordinate of Berlin center will be obtained using **Google Maps API geocoding** of well-known Sunshine Coast location (Mountain Creek)

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 approx. 12x12 kilometers centered around Berlin city center.

Let's first find the latitude & longitude of Berlin city center, using specific, well known address and Google Maps geocoding API.

Coordinate of Mountain Creek Queensland, Australia: [-26.699, 153.101]

Now let's create a grid of area candidates, equally spaced, centered around city center and within ~6km from Mountain Creek. 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. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Coordinate transformation check

```
-----  
Mountain Creek center longitude=153.133, latitude=-26.799  
Mountain Creek center UTM X=4881095.616965571, Y=-16211842.854695398  
Mountain Creek center longitude=153.133, latitude=-26.798999999999999
```

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 it's neighbors**.

364 candidate neighborhood centers generated.

Let's visualize the data we have so far: city center location and candidate neighborhood centers:

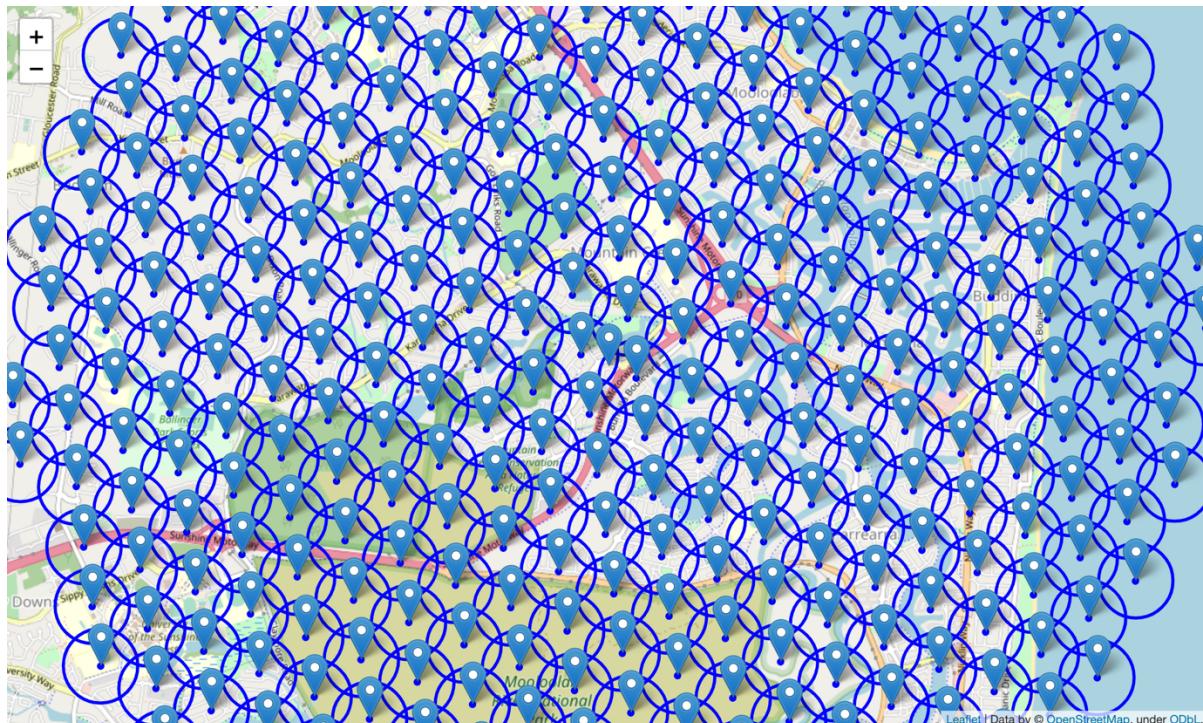


Figure 1 Candidate neighborhood centers

OK, 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 ~6km from Mountain Creek.

Let's now use Google Maps API to get approximate addresses of those locations.

Reverse geocoding check

```
-----  
Address of [-26.699, 153.101] is: 3 Figtree Ct, Mountain Creek QLD 4557  
, Australia
```

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 '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 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

Total number of restaurants: 85

Total number of Japanese restaurants: 3

Percentage of Japanese restaurants: 3.53%

Average number of restaurants in neighborhood: 0.2857142857142857

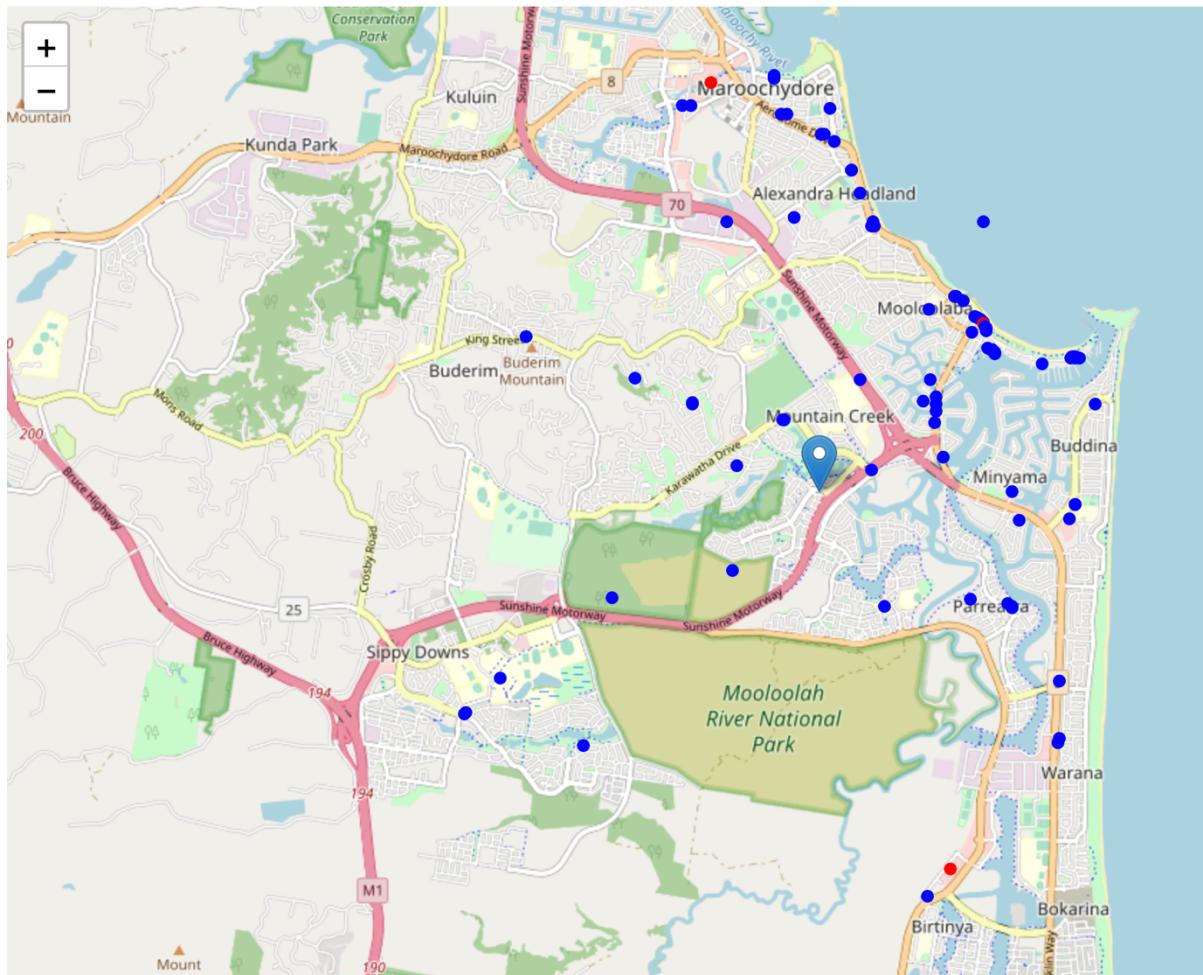


Figure 2 Location of all restaurants

Looking good. So now we have all the restaurants in area within few kilometers from Mountain Creek, and we know which ones are Japanese restaurants! We also know which restaurants exactly are in 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 Japanese restaurant!

Methodology

In this project we will direct our efforts on detecting areas of Sunshine Coast that have low restaurant density, particularly those with low number of Japanese 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 Sunshine Coast center** (Mountain Creek). We have also **identified Japanese restaurants** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Sunshine Coast - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and no Italian 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 more than two restaurants in radius of 250 meters**, and we want locations **without Japanese restaurants in radius of 3000 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.

Analysis

Let's perform some basic explanatory 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: 0.2857142857142857

Table 1 Number of restaurants in every area candidates

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	Queensland	-26.665515	153.131698	4.888155e+06	1.622887e+07	5992.495307	0
1	256 Alexandra Parade, Alexandra Headland QLD 4572	-26.663889	153.127215	4.888755e+06	1.622887e+07	5840.376700	0
2	251 Alexandra Parade, Alexandra Headland QLD 4572	-26.662264	153.122732	4.889355e+06	1.622887e+07	5747.173218	0
3	2 Aerodrome Rd, Maroochydore QLD 4558	-26.660639	153.118249	4.889955e+06	1.622887e+07	5715.767665	0
4	2 Aerodrome Rd,	-26.659014	153.113767	4.890555e+06	1.622887e+07	5747.173218	0

	Maroochydore QLD 4558							
5	54 Alexandra Parade, Maroochydore QLD 4558	-26.657388	153.109285	4.891155e+06	-1.622887e+07	5840.376700	0	0
6	40 Alexandra Parade, Maroochydore QLD 4558	-26.655763	153.104803	4.891755e+06	-1.622887e+07	5992.495307	0	0
7	Pacific Blvd, Buddina QLD 4575	-26.671441	153.136857	4.887255e+06	-1.622835e+07	5855.766389	0	0
8	Pacific Blvd, Buddina QLD 4575	-26.669816	153.132373	4.887855e+06	-1.622835e+07	5604.462508	0	0
9	Queensland	-26.668190	153.127890	4.888455e+06	-1.622835e+07	5408.326913	0	0

OK, now let's calculate the **distance to nearest Japanese restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).

Table 2 Distance to nearest Japanese restaurants

	Address	Latitude	Longitude	X	Y	Distance from center	Distance to Japanese restaurant
0	Queensland	-26.665515	153.131698	4.888155e+06	-1.622887e+07	5992.495307	0 2521.481119
1	256 Alexandra Parade, Alexandra Headland QLD 4572	-26.663889	153.127215	4.888755e+06	-1.622887e+07	5840.376700	0 2484.070096
2	251 Alexandra Parade, Alexandra Headland QLD 4572	-26.662264	153.122732	4.889355e+06	-1.622887e+07	5747.173218	0 2589.081199
3	2 Aerodrome Rd, Maroochydore QLD 4558	-26.660639	153.118249	4.889955e+06	-1.622887e+07	5715.767665	0 2820.652170
4	2 Aerodrome Rd, Maroochydore QLD 4558	-26.659014	153.113767	4.890555e+06	-1.622887e+07	5747.173218	0 3151.002360

5	54 Alexandra Parade, Maroochydore QLD 4558	-26.657388	153.109285	4.891155e+06	-1.622887e+07	5840.376700	0	0	2640.718775
6	40 Alexandra Parade, Maroochydore QLD 4558	-26.655763	153.104803	4.891755e+06	-1.622887e+07	5992.495307	0	0	2067.299941
7	Pacific Blvd, Buddina QLD 4575	-26.671441	153.136857	4.887255e+06	-1.622835e+07	5855.766389	0	0	2383.606066
8	Pacific Blvd, Buddina QLD 4575	-26.669816	153.132373	4.887855e+06	-1.622835e+07	5604.462508	0	0	2101.027151
9	Queensland	-26.668190	153.127890	4.888455e+06	-1.622835e+07	5408.326913	0	0	1966.482214

OK, so **on average Japanese restaurant can be found within ~3500m** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's create a map showing **heatmap / density of restaurants** and try to extract some meaningful info from that. Also, let's show **borders of Sunshine Coast boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Mountain Creek.

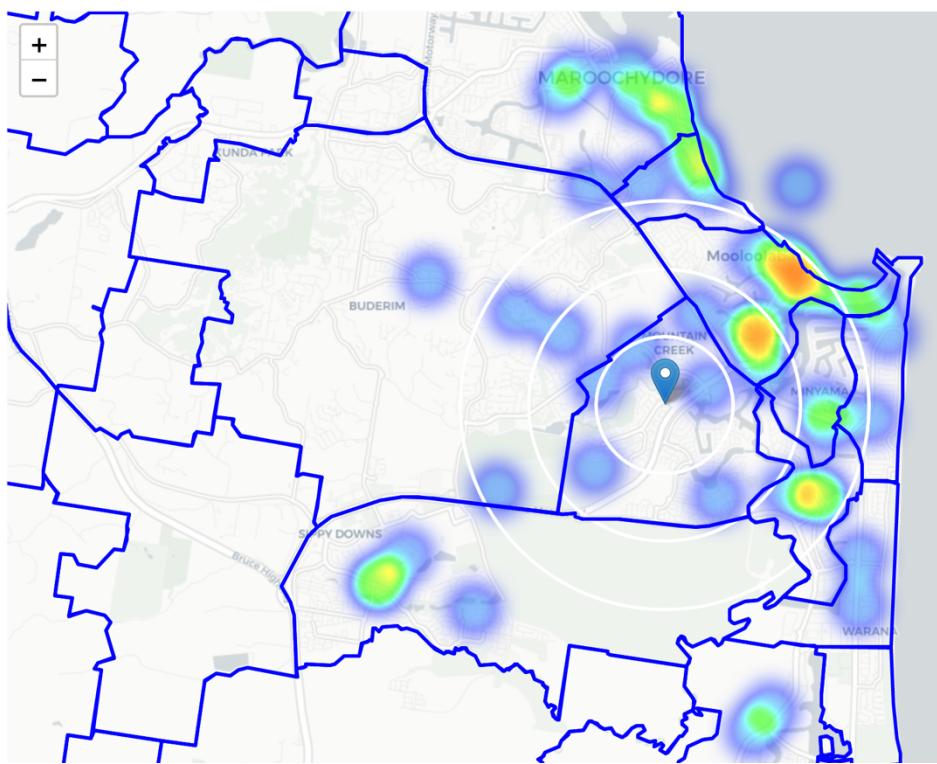


Figure 3 Heatmap of restaurants

Looks like a few pockets of low restaurant density closest to city center can be found **south, south-east and east** from **Mountain Creek**.

Let's create another heatmap map showing **heatmap/density of Japanese restaurants** only.

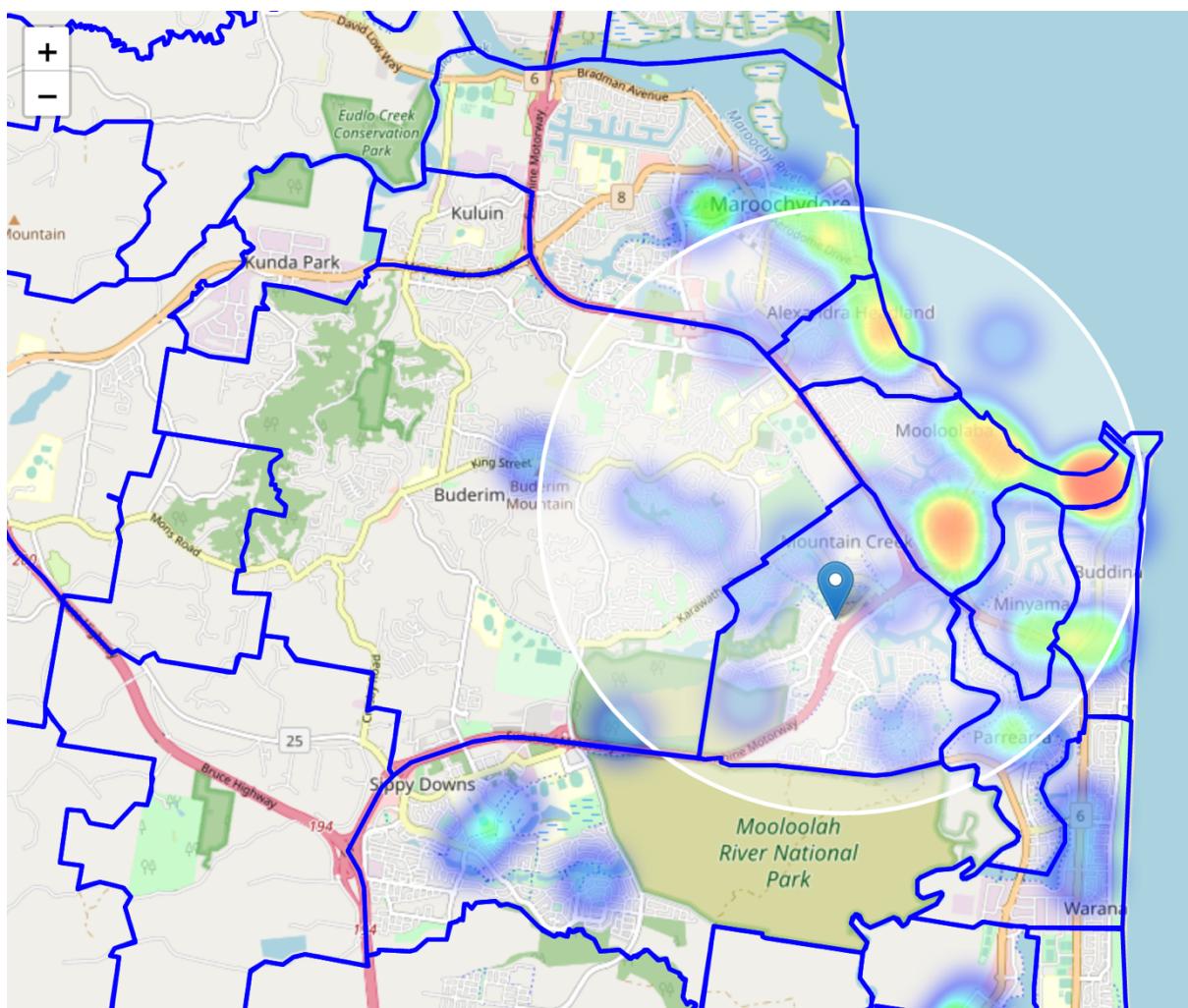


Figure 4 Heatmap of Japanese restaurants

This map is not so 'hot' (Japanese restaurants represent a subset of ~2% of all restaurants in Sunshine Coast) but it also indicates higher density of existing Japanese restaurants directly north and east from Mountain Creek, with closest pockets of **low Japanese restaurant density positioned west, north-west and south from city center**.

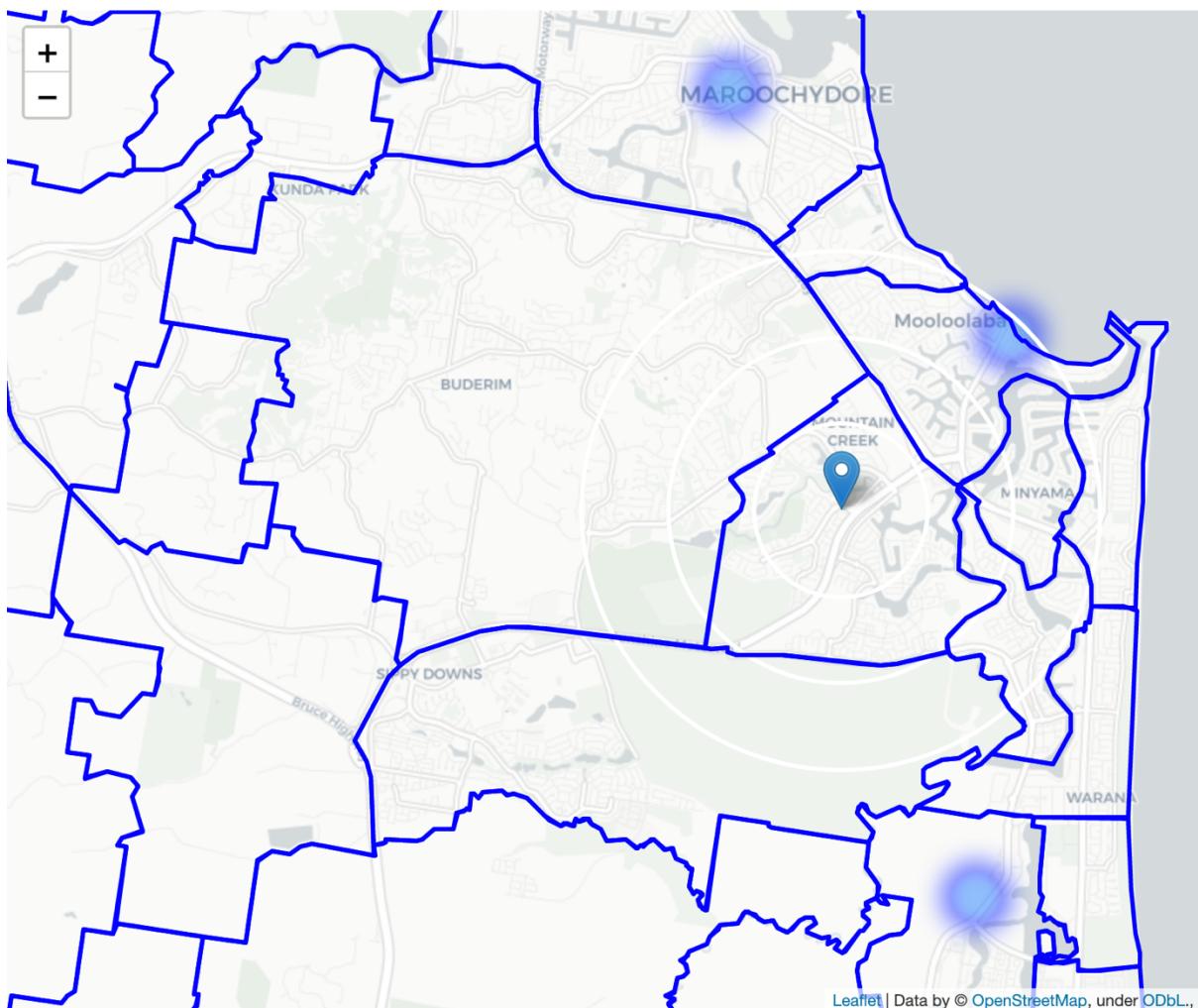


Figure 5 Japanese restaurants density

Based on this we will now focus our analysis on areas *west, north-west, south and east from Sunshine Coast center* - we will move the center of our area of interest and reduce it's size to have a radius of **3.0km**. This places our location candidates mostly in boroughs **Buderim and Sippy Downs** (another potentially interesting borough is **Miyama** with large low restaurant density south-east from city center, however this borough is less interesting to stakeholders as it's mostly residential and less popular with tourists).

Buderim and Sippy Downs

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to city center and well connected, those boroughs appear to justify further analysis.

Let's define new, narrower region of interest, which will include low-restaurant-count parts of Buderim and Sippy Downs closest to Mountain Creek.

OK. Let us now **filter** those locations: we're interested only in **locations with no more than two restaurants in radius of 250 meters**, and **no Japanese restaurants in radius of 3000 meters**. Let's see how this looks on a map.

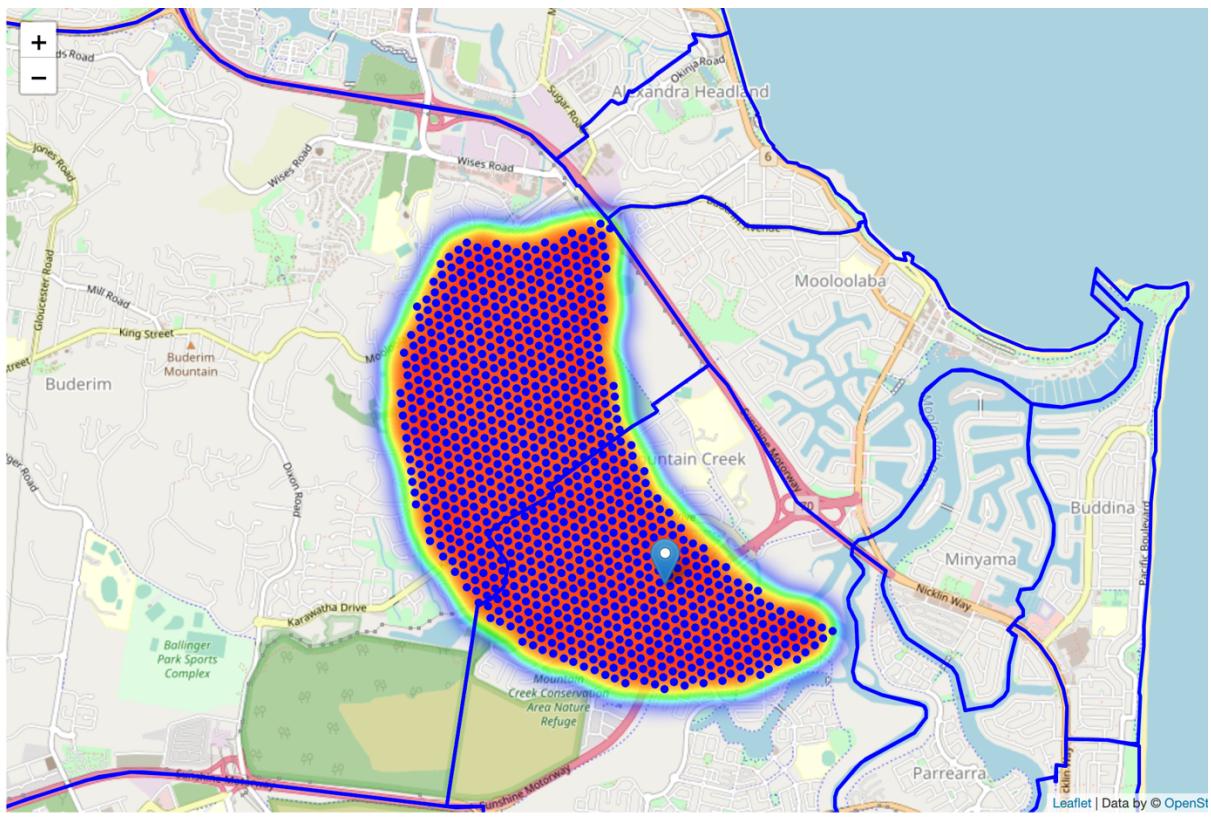


Figure 6 Good location for Japanese restaurant

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and *no* Japanese 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.

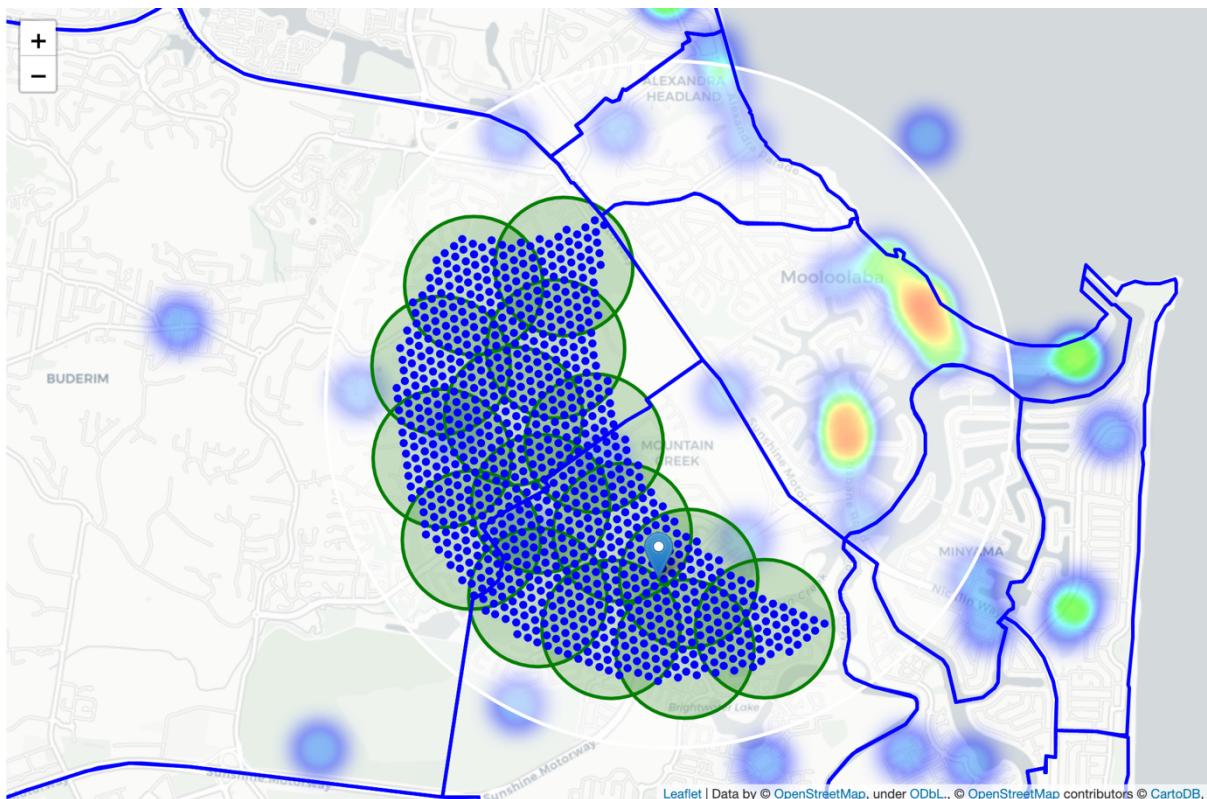


Figure 7 Zones of good location

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

98 Bundilla Blvd, Mountain Creek QLD 4557 in Creek	=> 0.3km from Mounta
11 Alfriston Dr, Buderim QLD 4556 in Creek	=> 2.0km from Mounta
15 Munbilla Cl, Mountain Creek QLD 4557 in Creek	=> 1.1km from Mounta
13 Spruce Ct, Buderim QLD 4556 in Creek	=> 2.2km from Mounta
1B Molakai Dr, Mountain Creek QLD 4557 in Creek	=> 1.4km from Mounta
10 Troywood Cres, Buderim QLD 4556 in Creek	=> 3.1km from Mounta
36 Carlisle Cct, Mountain Creek QLD 4557 in Creek	=> 0.7km from Mounta
Karawatha Dr at Hibiscus Retirement Resort, Buderim QLD 4556	=> 1.7km f rom Mountain Creek
30 Yew Ct, Buderim QLD 4556 in Creek	=> 2.8km from Mounta
40 Cootamundra Dr, Mountain Creek QLD 4557 in Creek	=> 1.1km from Mounta
25 Saratoga Dr, Mountain Creek QLD 4557 in Creek	=> 0.5km from Mounta
52 Glenfields Blvd, Mountain Creek QLD 4557 in Creek	=> 0.6km from Mounta
Unit 29/23 Elsa Wilson Dr, Buderim QLD 4556 in Creek	=> 2.3km from Mounta

15 Haddys Cl, Mountain Creek QLD 4557
in Creek
25 Elizabeth Daniels Way, Buderim QLD 4556
in Creek

=> 1.3km from Mounta
=> 3.0km from Mounta

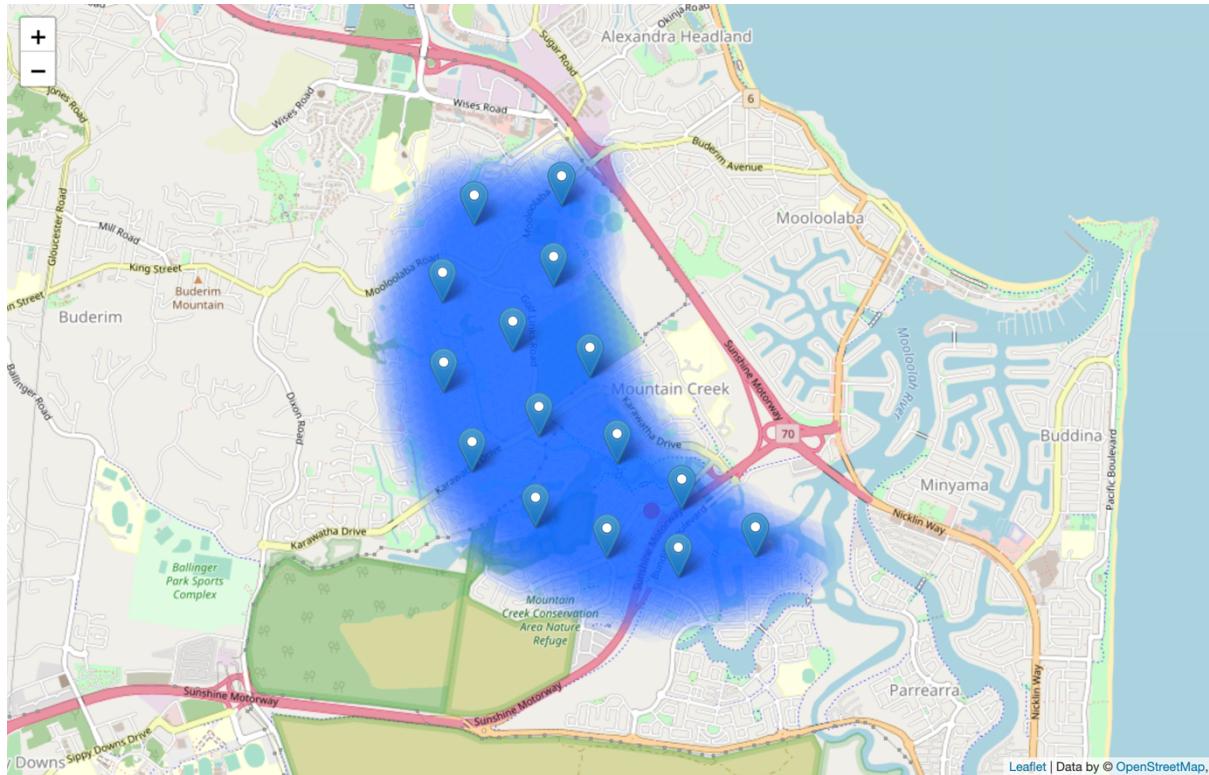


Figure 8 Addresses of center clusters

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 4km from Mountain Creek, and about half of those less than 2km from Mountain Creek). 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. Most of the zones are located in Buderim and Sippy Downs boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a small number of restaurants in Sunshine Coast (~80 in our initial area of interest which was 12x12km around Mountain Creek), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and east from Mountain Creek, so we focused our attention to areas south, north-west and west, corresponding to boroughs Buderim, Sippy Downs and south-east corner of central Mountain Creek borough. Another borough was identified as potentially interesting (Minyama Berg, south-east from Mountain Creek), but our attention was focused on Buderim and Sippy Downs which 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 approx. 5x5km south-east from Mountain Creek) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Japanese restaurant closer than 3000m 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.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Sunshine Coast center but not crowded with existing restaurants (particularly Japanese) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Sunshine Coast areas close to center with low number of restaurants (particularly Japanese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general boroughs that justify further analysis (Buderim and Sippy Downs), and then 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.