Recommending the ideal location to open a restaurant

Daniel Boey

June 7, 2021

1. Introduction

1.1 Background

Penang is an island-state off the northwestern coast of Peninsular Malaysia. It comprises two halves – Penang Island, where the capital city of George Town is located, and a strip of mainland Peninsula named Seberang Perai (formerly Province Wellesley). Penang's beaches are nice, though a little lackluster when compared to those in some other Malaysian states, but this is more than compensated for by the island's rich multicultural history dating back to the beginnings of British colonization in the 18th century and is full of Malay, Chinese, Indian and European influences. Penang is also well known domestically and in Singapore for being the "food paradise" of Malaysia.

1.2 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 Penang, Malaysia.

We will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Japanese restaurants in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

1.3 Interest

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

2. Data acquisition

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 Japanese 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:

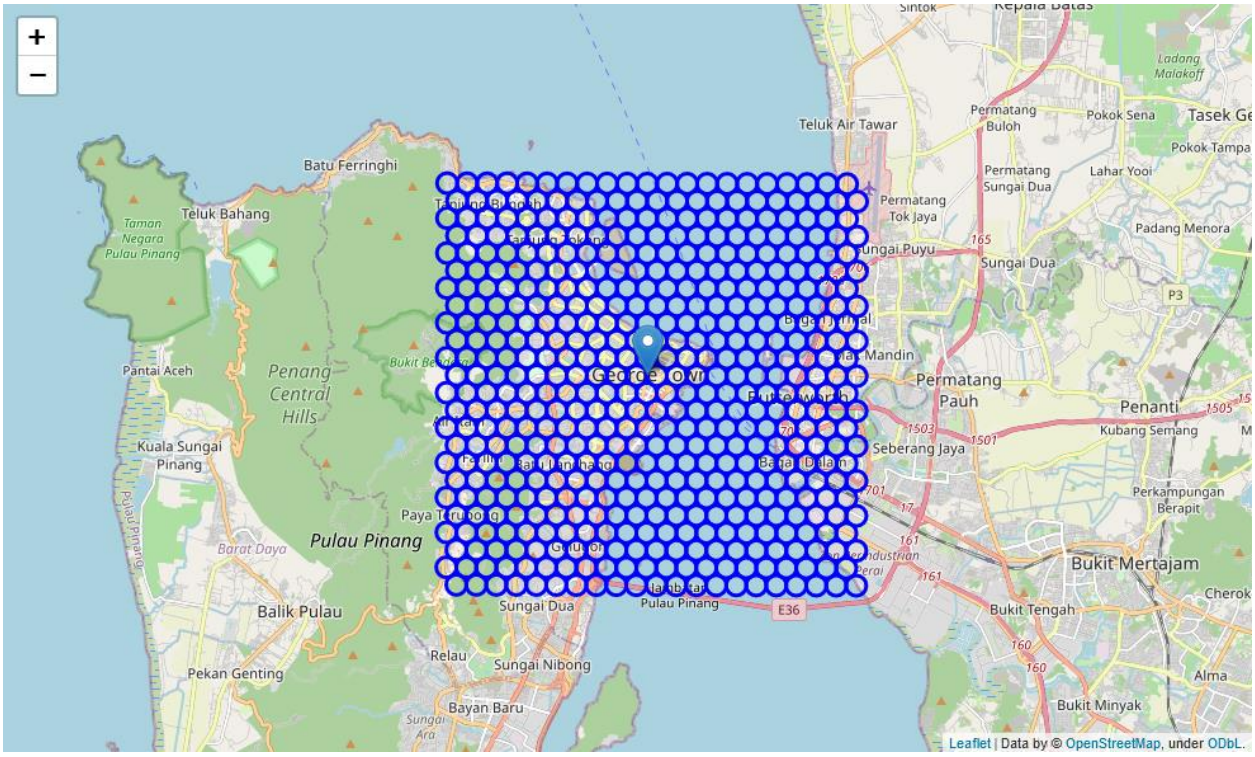
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of Penang center will be obtained using Google Maps API geocoding of Penang's main city, George Town

3. Exploratory Data Analysis

We will 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 12x12 kilometers centered around George Town. 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 the map.

We offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all its neighbors – and we have ourselves a hexagonal grid of cells. Here's a visualization of the city center location and candidate neighborhood centers:



We'll then use Google Maps API to approximate addresses of those locations. We'll then place them all into a Pandas dataframe and clean up any invalid addresses:

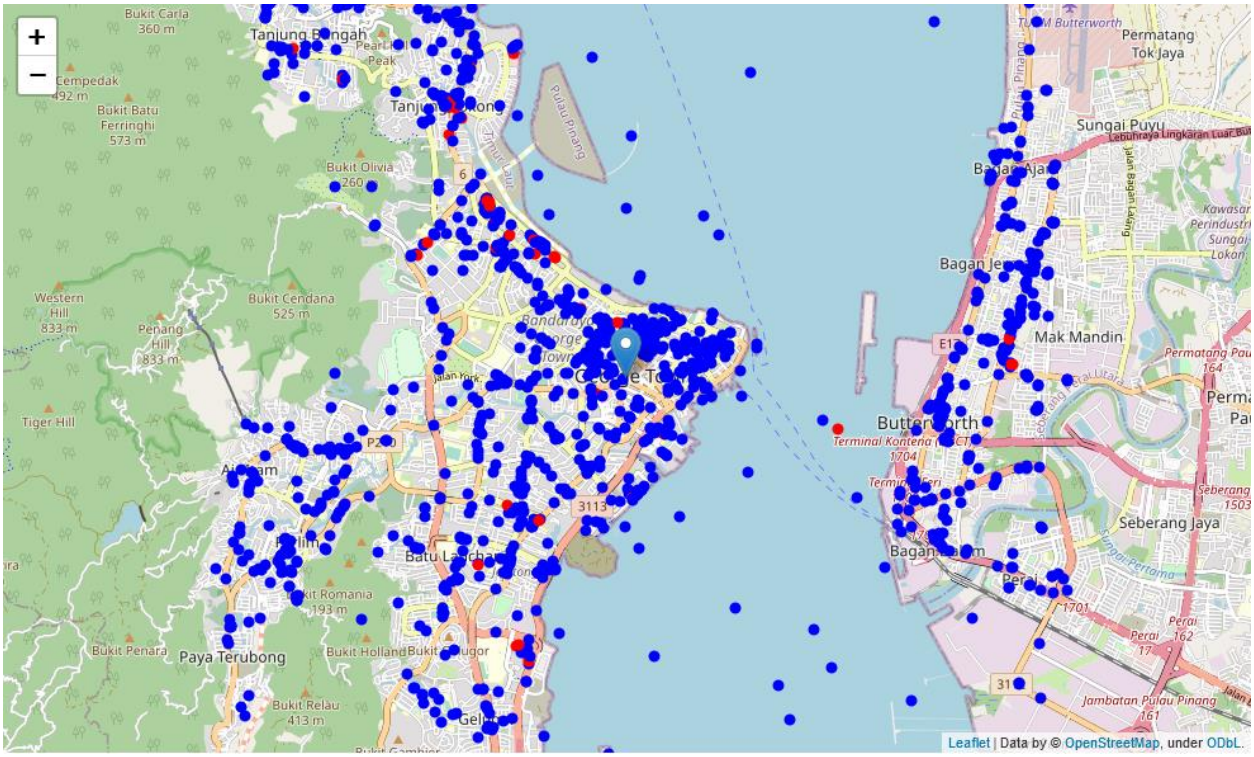
	Address	Latitude	Longitude	X	Y	Distance from center
0	Mukim 13 Paya Terubong, Penang	5.357847	100.277194	641514.479933	592367.674215	8448.076704
1	130, Halaman Bukit Gambir 4, 11700 Gelugor, Pu...	5.357836	100.282608	642114.479933	592367.674215	8055.432949
2	52, Halaman Bukit Gambir 9, Sunway Bukit Gambi...	5.357824	100.288022	642714.479933	592367.674215	7689.603371
3	1, Halaman Bukit Gambir, Sunway Bukit Gambier,...	5.357813	100.293436	643314.479933	592367.674215	7354.590403
4	Doping Control Centre, 11700 Gelugor, Pulau Pi...	5.357801	100.298850	643914.479933	592367.674215	7054.785610
5	10, Lengkok Sastera, 11800 Gelugor, Pulau Pinang	5.357790	100.304265	644514.479933	592367.674215	6794.850992
6	6, 11900 Gelugor, Pulau Pinang	5.357778	100.309679	645114.479933	592367.674215	6579.513660
7	Unnamed Road, 11700, 11700 Gelugor, Penang	5.357767	100.315093	645714.479933	592367.674215	6413.267498
8	Jambatan Pulau Pinang, Gelugor, Pulau Pinang	5.357755	100.320507	646314.479933	592367.674215	6300.000000
9	Jambatan Pulau Pinang, Gelugor, Pulau Pinang	5.357743	100.325921	646914.479933	592367.674215	6242.595614

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 'Japanese restaurant'

category, as we need info on Japanese restaurants in the neighborhood. Foursquare API has given us thus far:

```
Total number of restaurants: 1299
Total number of Japanese restaurants: 51
Percentage of Japanese restaurants: 3.93%
Average number of restaurants in neighborhood: 2.25
```

Plotting these restaurants in our area of interest on the map, and coloring Japanese restaurants in red:



We noticed that there are still wayward points of interest in the middle of ocean, probably from incorrect GPS information captured by Foursquare. We'll keep this in mind when we're recommending a location unless we are aiming for a floating restaurant.

This concludes our exploratory phase – we're now ready to use this data for analysis and to produce the report on optimal locations for a new Japanese restaurant!

4. Methodology

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

After collecting and explored the data, in this 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 a radius of 250 meters, and we want locations without Japanese restaurants in a 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.

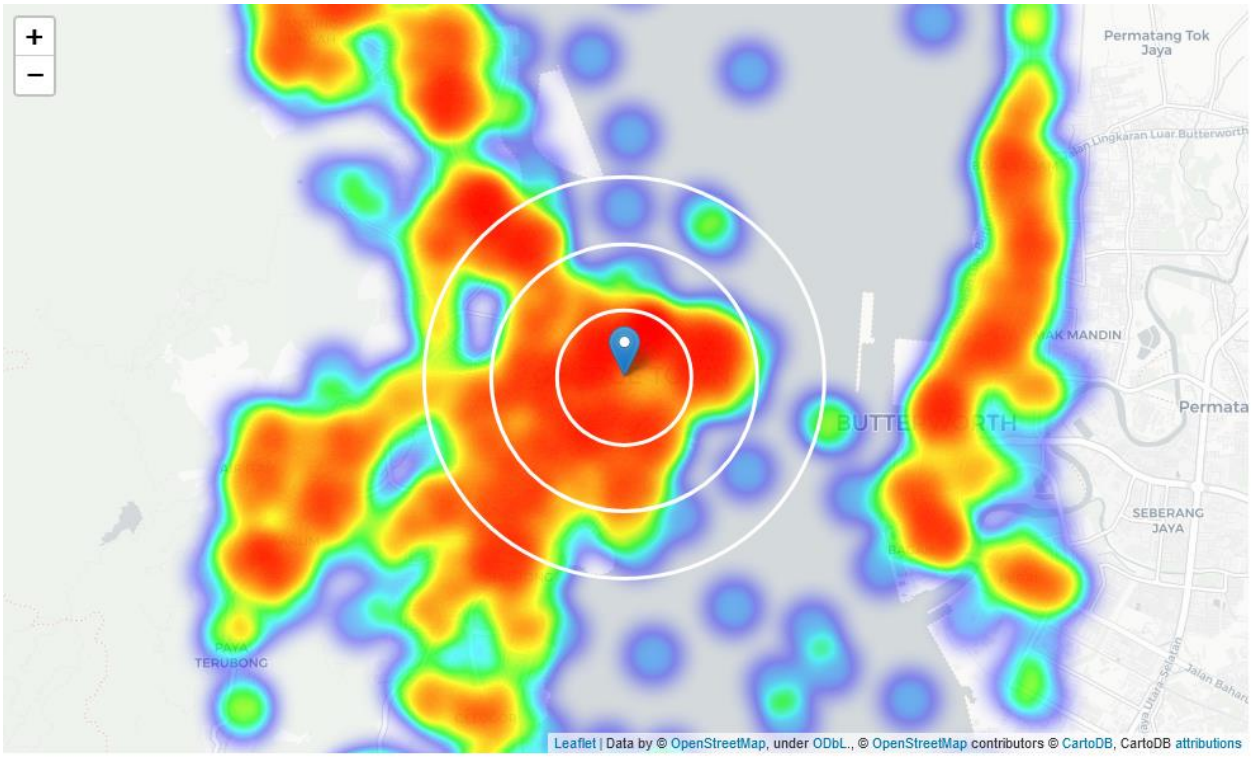
4.1 Analysis & Clustering

Count of number of restaurants in every area candidate and distance of closest Japanese restaurant to area center:

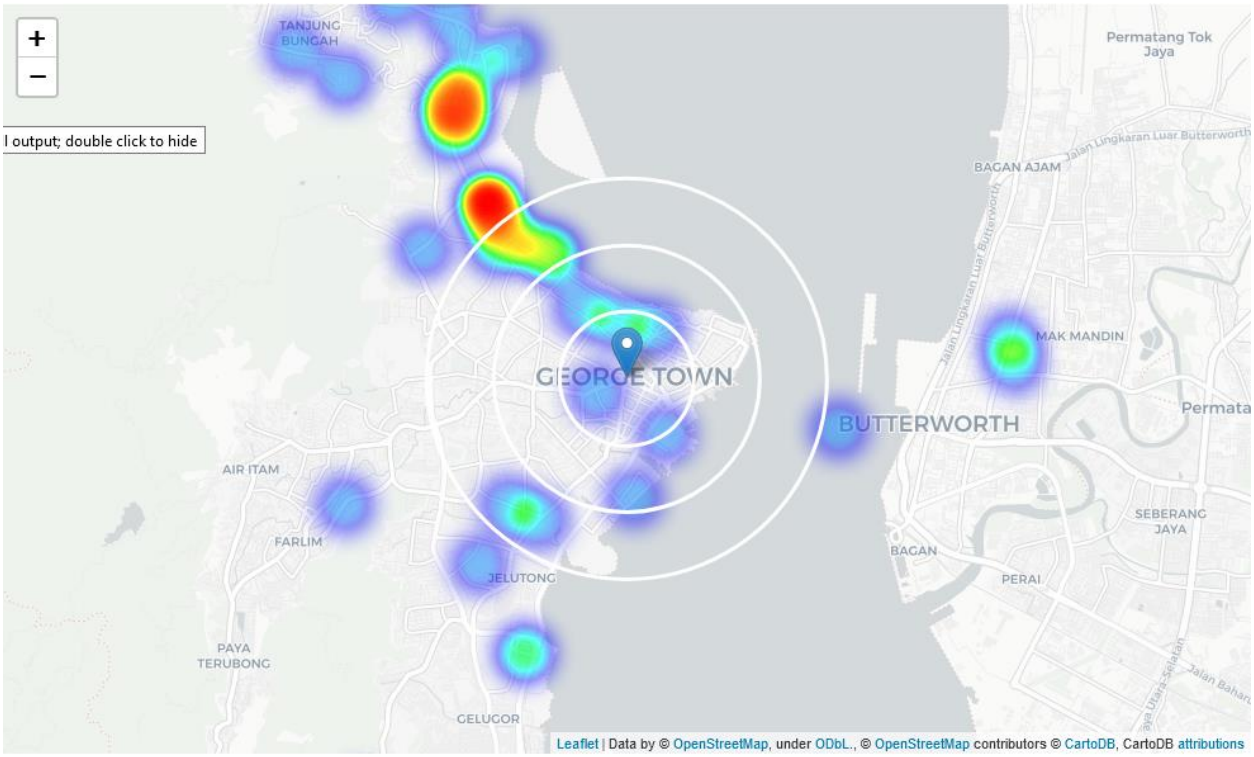
```
Average number of restaurants in every area with radius=300m: 2.25

Average distance to closest Japanese restaurant from each area center: 1861.4336775880297
```

By creating a heatmap showing the density of restaurants, we can observe pockets of low restaurant density closest to the city center to the west of George Town:



Doing the same but a heatmap showing the density of Japanese restaurants only:



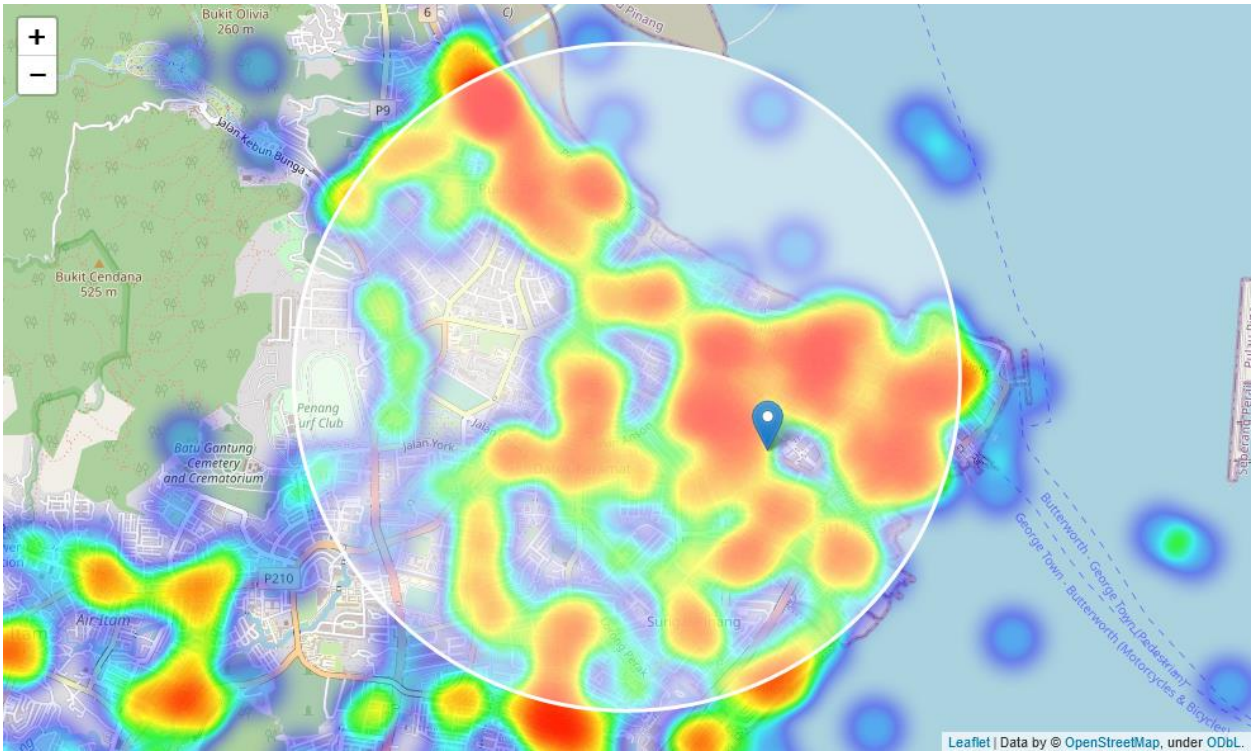
This map is not so 'hot' (Japanese restaurants represent a subset of ~4% of all restaurants in George Town, Penang) but it also indicates higher density of existing Japanese restaurants directly northwest from George Town, with closest pockets of **low Japanese restaurant density positioned west, southwest, south, southeast and east from city center**.

Based on this information we will now focus our analysis on areas **west from George Town center** - we will move the center of our area of interest and reduce its size to have a radius of **2.5km**.

4.2 Vicinities of Jalan Logan

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to city center and well connected, the location appears to justify further analysis.

Let's define a new, narrower region of interest, which will include low-restaurant-count parts of Jalan Logan closest to George Town:



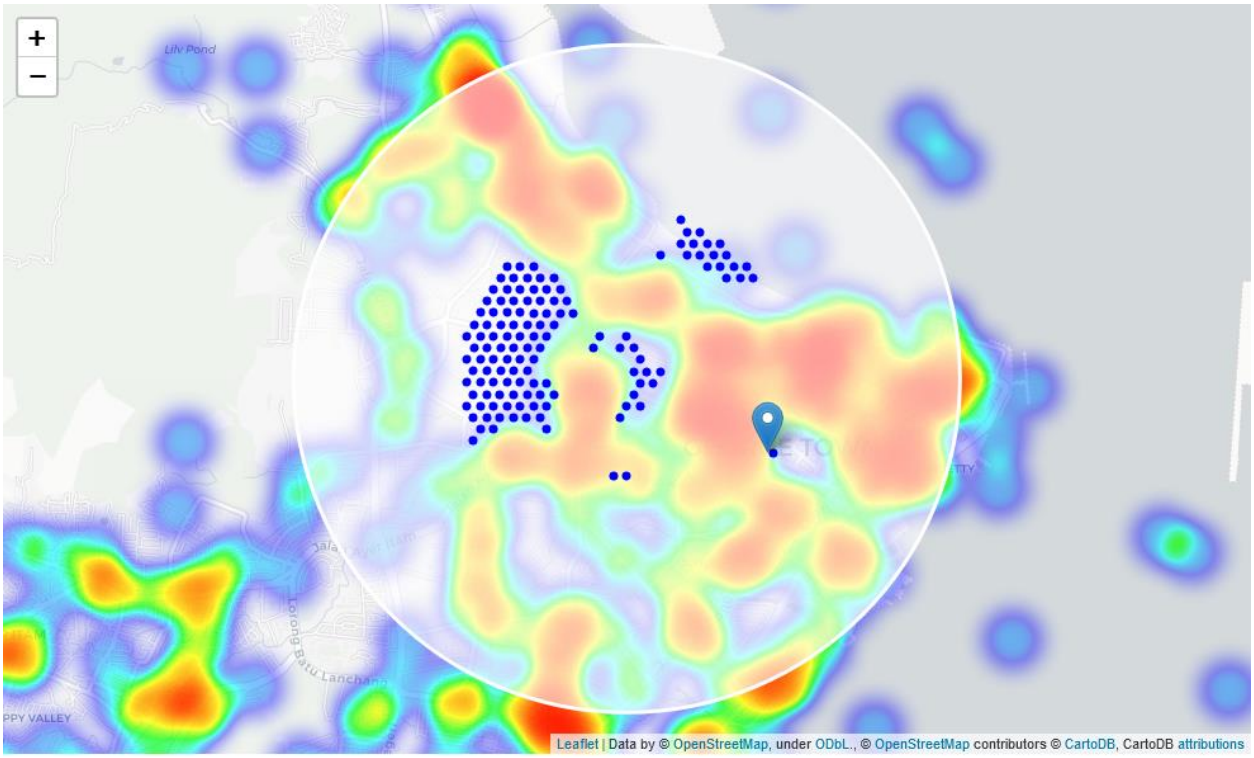
First, we calculate the number of restaurants in the vicinity (we used a radius of 250 meters) and distance to closest Japanese restaurant:

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Japanese restaurant
0	5.412588	100.310247	645164.479933	598428.057122	6	1615.229149
1	5.412586	100.311150	645264.479933	598428.057122	9	1515.289746
2	5.412584	100.312052	645364.479933	598428.057122	8	1415.358902
3	5.412582	100.312955	645464.479933	598428.057122	7	1315.438570
4	5.412580	100.313857	645564.479933	598428.057122	7	1215.531341
5	5.412578	100.314759	645664.479933	598428.057122	7	1115.640735
6	5.412577	100.315662	645764.479933	598428.057122	8	1015.771656
7	5.412575	100.316564	645864.479933	598428.057122	8	915.931147
8	5.412573	100.317467	645964.479933	598428.057122	9	816.129693
9	5.412571	100.318369	646064.479933	598428.057122	2	716.383616

and then keeping only locations with no more than two restaurants in a radius of 250 meters, we'll end up with:

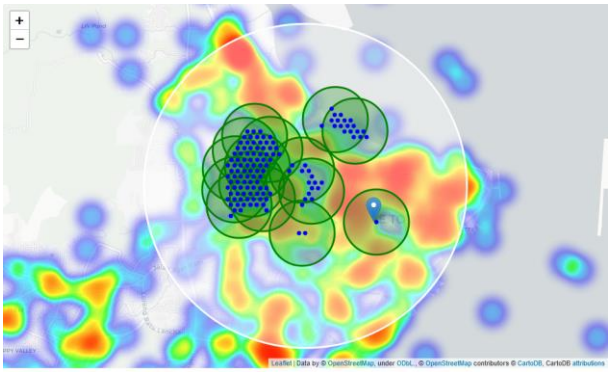
```
Locations with no more than two restaurants nearby: 156
Locations with no Japanese restaurants within 400m: 328
Locations with both conditions met: 128
```

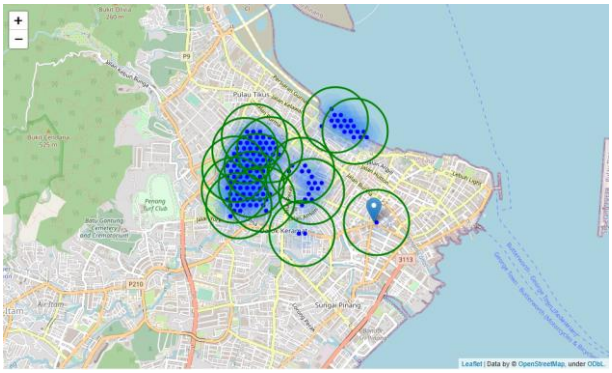
Overlaying these potential locations onto the heatmap:



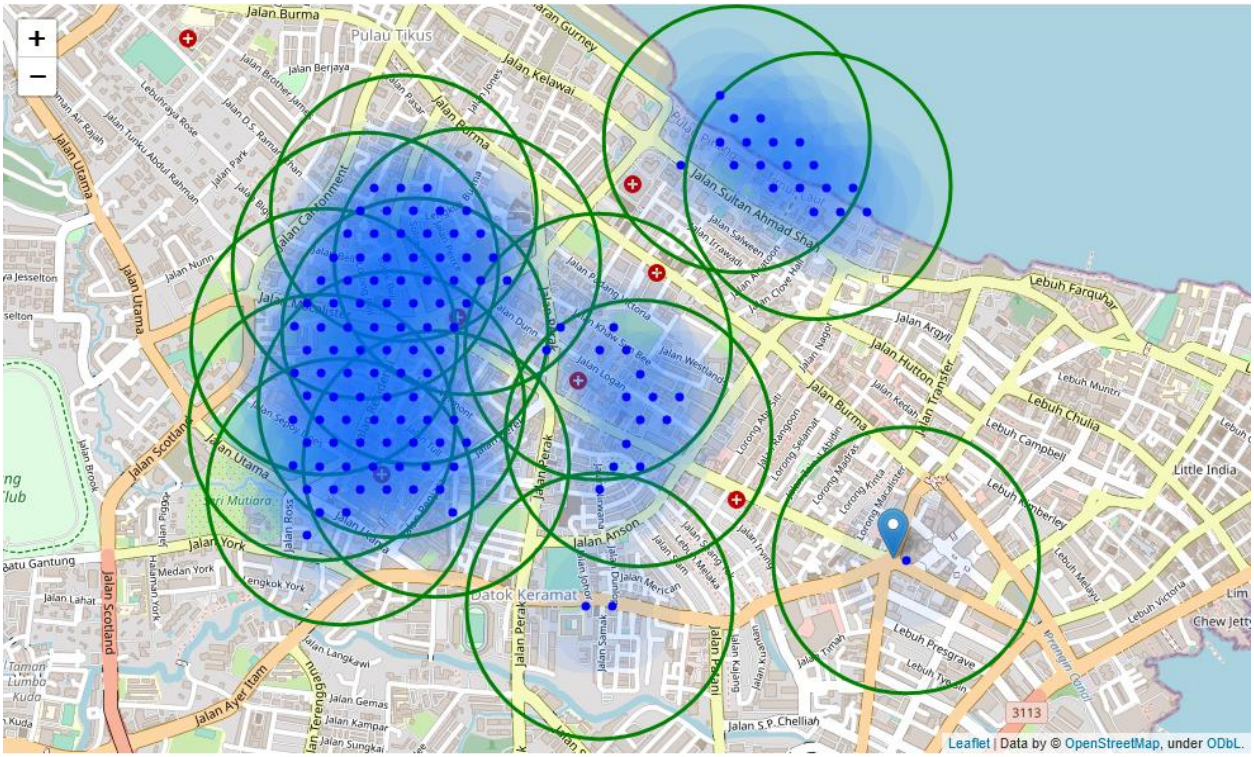
Reverse heatmap of the potential locations only:

Clustering these locations with kMeans Machine Learning algorithm to create centers of zones containing good locations. These zones, with their centers' addresses will be the final result of our analysis (with and without headmap):





Zoomed into Jalan Logan vicinity:



By geocode-reversing these candidate area centers and getting their addresses, these are what we can recommend to our stakeholders for further analysis:

```
48,B,C,D, Jln Sultan Ahmad Shah, 10050 George Town, Pulau Pinang => 1.4km from George Town
1, Jalan Tull, George Town, 10450 George Town, Pulau Pinang => 2.0km from George Town
1, Jalan Pierce, George Town, 10350 George Town, Pulau Pinang => 1.9km from George Town
233, Jalan Macalister, George Town, 11400 George Town, Pulau Pinang => 1.1km from George Town
18-1-1, Casabella APT, Pulau Tikus, 10350 George Town, Penang => 2.3km from George Town
65, Jalan Utama, George Town, 10450 George Town, Pulau Pinang => 2.1km from George Town
219, Jalan Magazine, George Town, 10300 George Town, Pulau Pinang => 0.1km from George Town
4, Jalan Turf Club, 10450 George Town, Pulau Pinang => 2.2km from George Town
111, Jalan Macalister, George Town, 10450 George Town, Pulau Pinang => 2.3km from George Town
395, Jalan Dato Keramat, George Town, 10150 George Town, Pulau Pinang => 1.1km from George Town
12, Jalan Khaw Sim Bee, George Town, 10400 George Town, Pulau Pinang => 1.4km from George Town
15, Lebuhraya Codrington, Pulau Tikus, 10350 George Town, Pulau Pinang => 2.3km from George Town
58, Jln Sultan Ahmad Shah, 10050 George Town, Pulau Pinang => 1.7km from George Town
219, Jalan Macalister, George Town, 10400 George Town, Pulau Pinang => 2.0km from George Town
34, Jalan Berek, George Town, 10450 George Town, Pulau Pinang => 1.8km from George Town
```

5. Results & Conclusions

Our analysis shows that although there is a great number of restaurants in Penang (~1300 in our initial area of interest which was 12x12km around George Town), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north, northeast, east, southeast and south from Georgetown, so we focused our attention to areas west, corresponding to the vicinities of Jalan Logan. Vicinity of Jalan Logan offers 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 narrower area of interest (covering approx. 5x5km west from George Town) 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 a Japanese restaurant closer than 400m 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 Japanese 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 Penang city 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.

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.