Abstract

As part of the IBM & Coursera Applied Data Science Specialization, it is required to do a small research using the taught data science methodologies and data sources. The data science methodologies and data sources that had more emphasis were, Geographical Data with the Python package Folium, and the usage of third party APIs for retrieval of location data.

The problem that this little research tackles is about Food Venues, especially Mexican Food venues in the city of Naha, Okinawa Japan which is the place I wish to live in the not so distant future. One of my plans while living there is Co-owning a Mexican Food restaurant.

Keywords: *Data Science; Foursquare; GIS; Google Maps API; Capstone, Python.*

1 Introduction

In this project we will try to find an optimal location for a restaurant.

Specifically, this report will be targeted to stakeholders (mainly Myself) interested in opening a Mexican Taqueria in Naha City, Okinawa, Japan. The focus of this is looking for locations that are not already crowded with restaurants. We are also particularly interested in areas with no Mexican restaurants in vicinity. We would also prefer locations as close to the IT COLLEGE as possible, this because IT people love tacos and because I'd love to work as a teacher there and have tacos after class, but only assuming that first two conditions are met. We will use a data science approach to select the most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that taking a decision using the top neighborhoods will be simplyfied.

All of this is part of the final project for the Applied Data Science Specialization offered by IBM Through Coursera.
[ICa]

2 Data

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

- Number of existing restaurants in the neighborhood (any type of restaurant)
- Number of and distance to Mexican restaurants in the neighborhood, if any
- Distance of neighborhood from IT College

Following the project example, I decided to also use regularly spaced grid of locations, centered around city center, to define our neighborhoods. In fact I will mostly use the same approach and step by step analysis, due to time constraint and not having any other particulary interesting problem to solve using the foursquare API which is required for this project.

[ICb]

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 the IT college will be obtained using Google Maps API geocoding.

[Deva]
[Devb]

2.1 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 aprox. 6x6 killometers centered around Naha city center.
Let's first find the latitude & longitude of IT College, using specific, well known address and Google Maps geocoding API.
Now let's create a grid of area candidates, equaly spaced, centered around city center and within 3km from IT College. 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).

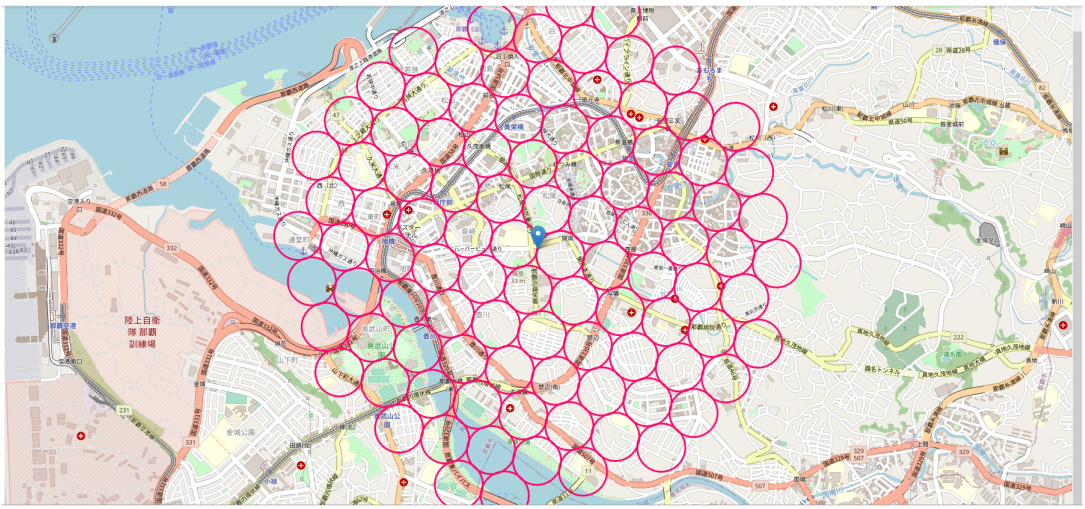

Figure 1: Candidate Locations.

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 - coffe shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Mexican restaurant' category, as we need info on Mexican restaurants in the neighborhood.

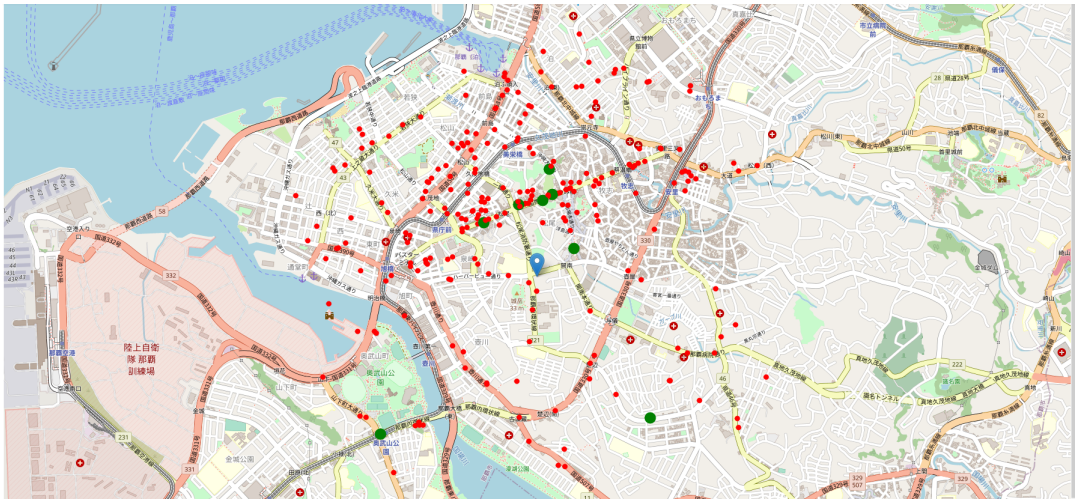

Figure 2: Restaurants (red), Mexican Restaurants (green)

2.2 Methodology

In this project we will direct our efforts on detecting areas of Naha that have low restaurant density, particularly those with low number of Mexican 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 3km from IT College. We have also identified Mexican restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Naha - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Mexican 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 Mexican restaurants in radius of 400 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 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=250m: 2.24

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Mexican restaurant
0	Japan, 〒900-0031 Okinawa, Naha, Wakasa, 3-chōm...	26.223398	127.677228	8.031182e+06	1.421792e+07	2962.283631	0	2030.395708
1	2-chōme-14-14 Wakasa, Naha, Okinawa 900-0031, ...	26.221176	127.674926	8.031782e+06	1.421792e+07	2869.690630	1	1906.502905
2	1-chōme-10-13 Wakasa, Naha, Okinawa 900-0031, ...	26.218955	127.672624	8.032382e+06	1.421792e+07	2900.883367	2	1966.468896
3	1-chōme-8-14 Tsuji, Naha, Okinawa 900-0037, Japan	26.216733	127.670322	8.032982e+06	1.421792e+07	3052.068857	3	2195.278231
4	3 Chome-25 Maejima, Naha, Okinawa 900-0016, Japan	26.224933	127.682816	8.030282e+06	1.421844e+07	2868.262748	1	1620.091745
5	Japan, 〒900-0016 Okinawa, Naha, Maejima, 3-chō...	26.222711	127.680513	8.030882e+06	1.421844e+07	2570.395143	2	1527.894788
6	Hotel Roco Inn Okinawa, 1-chōme-27-11 Matsuyam...	26.220489	127.678211	8.031482e+06	1.421844e+07	2388.918414	2	1442.368041
7	Japan, 〒900-0032 Okinawa, Naha, Matsuyama, 1-c...	26.218267	127.675908	8.032082e+06	1.421844e+07	2350.942617	1	1397.380464
8	1-chōme-3-1 Kume, Naha, Okinawa 900-0033, Japan	26.216046	127.673606	8.032682e+06	1.421844e+07	2463.114125	2	1595.280149
9	1-chōme-16-16 Nishi, Naha, Okinawa 900-0036, J...	26.213824	127.671304	8.033282e+06	1.421844e+07	2706.830469	4	1963.966738

Figure 3: Computed Restaurant distances.

Let's calculate the distance to Mexican restaurant from every area candidate center (not only those within 300m - we want distance to closest one, regardless of how distant it is). (Go to Linked Jupyter Notebook for details) On average Mexican restaurant can be found within 1200m from every area center candidate.

Let's crete a map showing heatmap / density of restaurants and try to extract some meaningfull info from that.

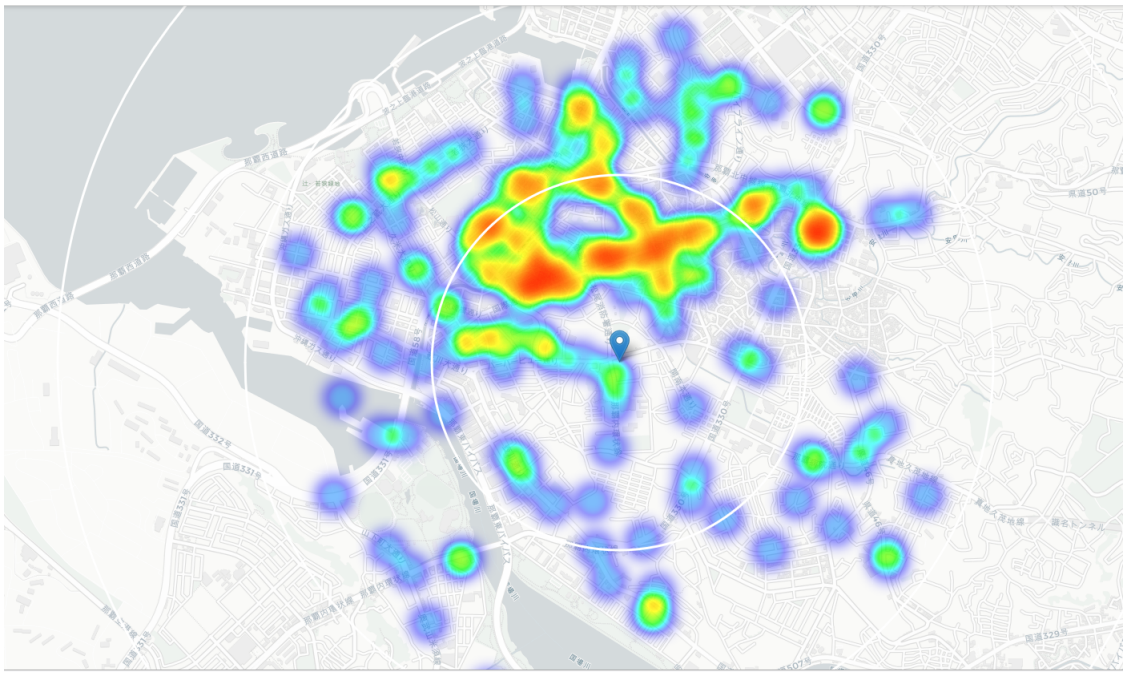

Figure 4: Heatmap of Restaurants in the area.

Looks like a few pockets of low restaurant density closest to Naha center can be found south, south-east and east from IT College.

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

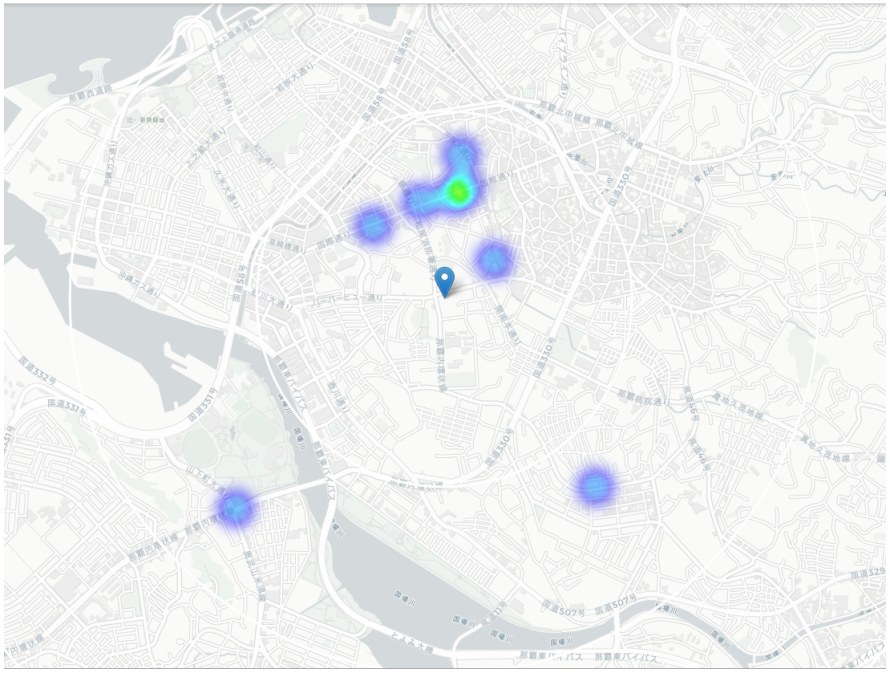

Figure 5: Heatmap of Restaurants in the area.

This map is so 'cold' (Mexican restaurants represent a subset of 8% of all restaurants in Naha) but it also indicates higher density of existing restaurants directly north and west

from IT College, with closest pockets of lowrestaurant 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 IT College - we will move the center of our area of interest and reduce it's size to have a radius of 2.5km. This places our location candidates mostly in boroughs Asahimachi and Matsuo

3.1 Asahimachi & Matsuo

Analysis with Google earth show Asahimachi and Matsuo as beautiful, interesting, highly populated and concurrent in the work hours.

This places also sport multiple recreation and shopping sites.

Asahimachi is just a bridge away from the Onoyama park and sports complex, also seems to have an activity boost for the nightlife thanks to some more turistic places nearby and a pachinko parlor inside the block.

Matsuo has a big park in it and offers many food places without being overcrowded with them like the surrounding areas, and most relevant a few blocks away you can finde the Naha Kokusai Dori Shopping Street.

Asahimachi

Figure 6: Asahimachi map view.

Figure 7: Asahimachi satellite view.

Figure 8: Asahimachi street view.

Matsuo

Figure 9: Matsuo map view.

Figure 10: Matsuo satellite view.

Figure 11: Matsuo street view.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 170m appart), Now let's calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use radius of 250 meters) and distance to closest Mexican restaurant.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Mexican restaurant
0	26.217613	127.678577	8.031852e+06	1.421888e+07	5	949.563464
1	26.216984	127.677924	8.032022e+06	1.421888e+07	6	958.213950
2	26.216354	127.677272	8.032192e+06	1.421888e+07	4	996.231487
3	26.215725	127.676620	8.032362e+06	1.421888e+07	1	1000.000000
4	26.215095	127.675967	8.032532e+06	1.421888e+07	2	1000.000000
5	26.214466	127.675315	8.032702e+06	1.421888e+07	2	1000.000000
6	26.213837	127.674663	8.032872e+06	1.421888e+07	1	1000.000000
7	26.217918	127.680025	8.031632e+06	1.421902e+07	4	841.663362
8	26.217289	127.679373	8.031802e+06	1.421902e+07	7	806.288048
9	26.216659	127.678720	8.031972e+06	1.421902e+07	8	805.979911

Figure 12: Candidates.

It is Important to also filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters, and no Italian restaurants in radius of

400 meters.

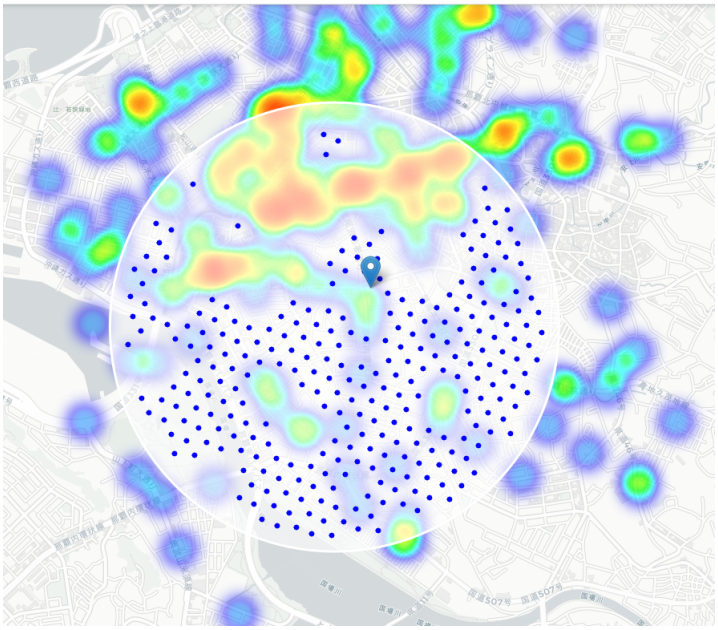

Figure 13: Grid of candidate areas.

Now to test how well we are covering the interested ares lets make a union of the previous heat map with a new one for the proposed locations.

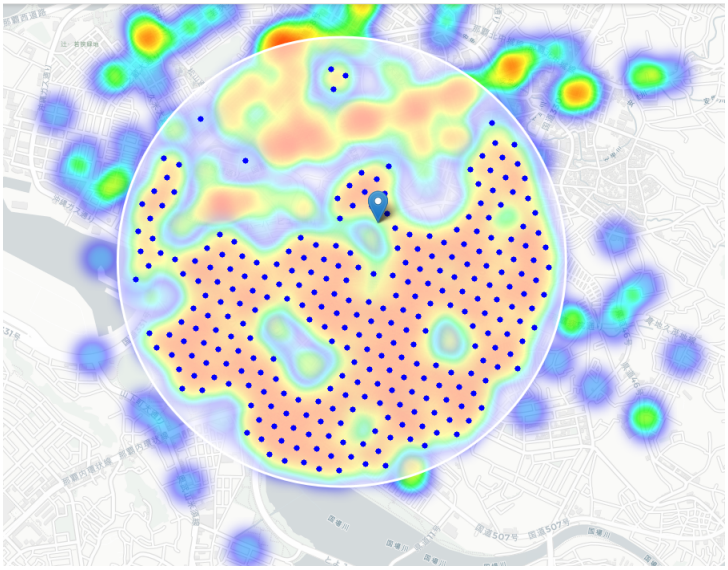

Figure 14: Union Heatmap current restaurant locations and candidate locations.

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. I decided to go for 50 clusters, thats a bit too much but, thats the best approximate number to cover the vast areas of interest, covering any point and avoid to much overlappin as well

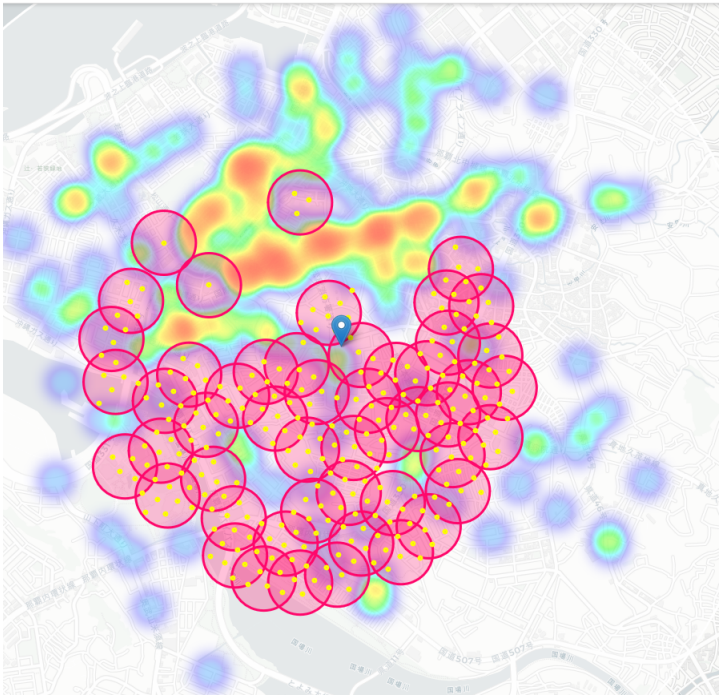

Figure 15: Clusters.

Not bad - 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 those zones on a city map without heatmap, using shaded areas to indicate our clusters:

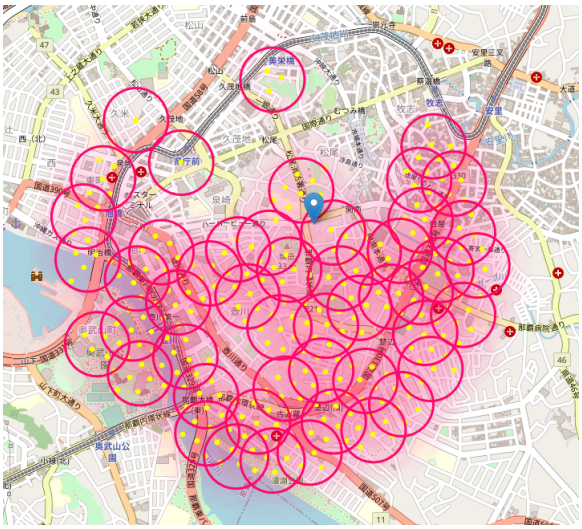

Figure 16: Shaded Clusters for better visibility.

3.2 Asamichi and Matsuo clusters

Let's zoom in on candidate areas near Asahimachi:

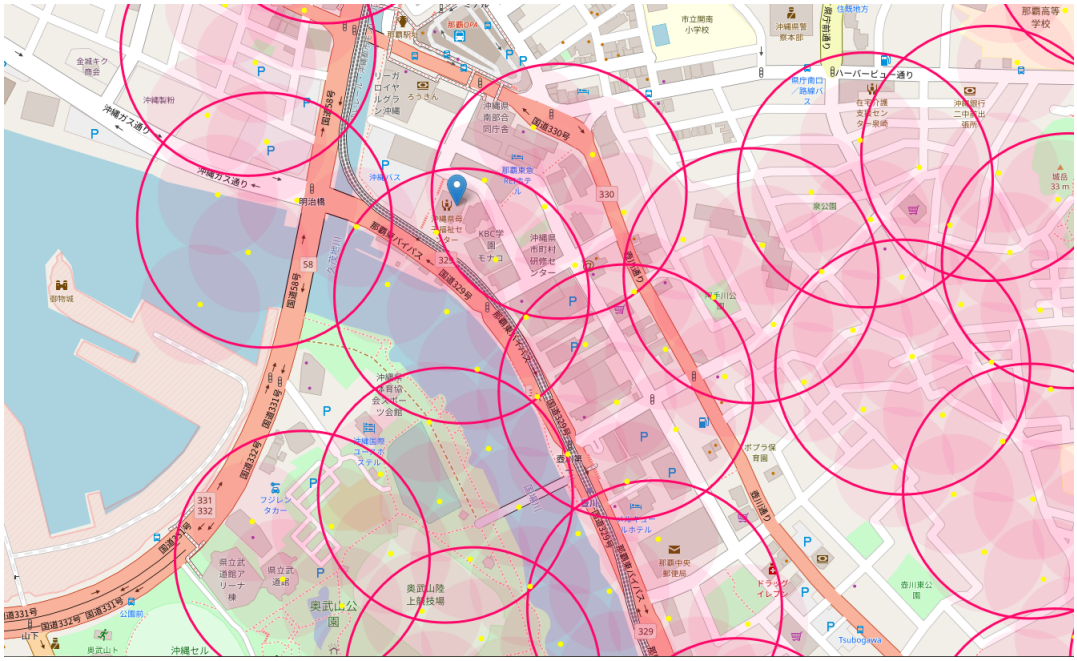

Figure 17: Viable locations near Asahimachi.

Let's zoom in on candidate areas near Matsuo:

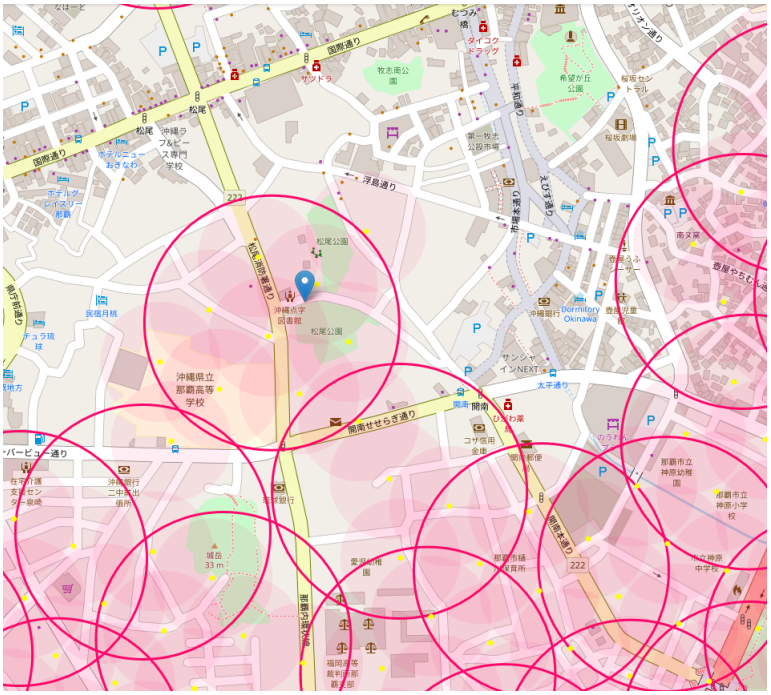

Figure 18: Viable locations near Matsuo.

3.3 Final map with addresses

Finaly, 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
============================================================

1-chōme-21-27 Higawa, Naha, Okinawa 900-0022        => 0.6km from IT College
44 Ōnoyamachō, Naha, Okinawa 900-0026               => 2.2km from IT College
3-chōme-14-8 Kohagura, Naha, Okinawa 900-0024       => 2.2km from IT College
1-chōme-2-15 Yorimiya, Naha, Okinawa 902-0064       => 1.3km from IT College
Japan, 〒900-0034 Okinawa, Naha, Higashimachi, 8-８ (有)寿 Arakaki.Bld => 2.1km from IT College
3-chōme-6-3 Sobe, Naha, Okinawa 900-0023            => 0.9km from IT College
1-chōme-17-5 Tsuboya, Naha, Okinawa 902-0065        => 1.1km from IT College
2-chōme-19-11 Sobe, Naha, Okinawa 900-0023          => 1.4km from IT College
Japan, 〒900-0025 Okinawa, Naha, Tsubogawa, 3-chōme, 那覇バスターミナル => 1.5km from IT College
1-chōme-6-33 Makishi, Naha, Okinawa 900-0013        => 1.4km from IT College
Japan, 〒900-0024 Okinawa, Naha, Kohagura, 3-chōme-23, 国道329号線 Shinnahao Bridge => 2.2km from IT College
1-chōme-13-16 Matsuo, Naha, Okinawa 900-0014        => 0.3km from IT College
1-chōme-13-1 Kohagura, Naha, Okinawa 900-0024       => 2.0km from IT College
116 Yogi, Naha, Okinawa 902-0076                    => 1.8km from IT College
114-6 Asahimachi, Naha, Okinawa 900-0029            => 1.8km from IT College
2-chōme-35-17 Sobe, Naha, Okinawa 900-0023          => 1.6km from IT College
1-chōme-20-44 Higawa, Naha, Okinawa 900-0022        => 0.9km from IT College
Japan, 〒900-0035 Okinawa, Naha, Tondōchō, 1, 国道58号  => 2.2km from IT College
1-chōme-18-22 Yogi, Naha, Okinawa 902-0076          => 1.7km from IT College
3-chōme-11-13 Sobe, Naha, Okinawa 900-0023          => 1.0km from IT College
1-chōme-6-23 Sobe, Naha, Okinawa 900-0023           => 0.5km from IT College
Wing Chateau Kohagurakoenmae, 3-chōme-5-38 Kohagura, Naha, Okinawa 900-0024 => 1.9km from IT College
Japan, 〒900-0022 Okinawa, Naha, Higawa, 2-chōme-14, 樋川2-14-6 座安アパート1F => 0.9km from IT College
1 Chome-28 Tsuboya, Naha, Okinawa 902-0065          => 1.4km from IT College
2-chōme-3-17 Tsubogawa, Naha, Okinawa 900-0025      => 1.7km from IT College
1-chōme-10-21 Yorimiya, Naha, Okinawa 902-0064      => 1.4km from IT College
2-chōme-1-24 Sobe, Naha, Okinawa 900-0023           => 1.0km from IT College
2-chōme-101-3 Izumizaki, Naha, Okinawa 900-0021     => 1.1km from IT College
51番地 Ōnoyamachō, Naha, Okinawa 900-0026             => 2.0km from IT College
1-chōme-6-1 Kohagura, Naha, Okinawa 900-0024        => 1.6km from IT College
2-chōme-9-10 Kohagura, Naha, Okinawa 900-0024       => 1.9km from IT College
1-chōme-2-9 Yogi, Naha, Okinawa 902-0076            => 1.2km from IT College
2-chōme-16-1 Kume, Naha, Okinawa 900-0033           => 2.0km from IT College
26-50 Ōnoyamachō, Naha, Okinawa 900-0026            => 2.3km from IT College
112-27 Asahimachi, Naha, Okinawa 900-0029           => 1.5km from IT College
1-chōme-6-8 Yogi, Naha, Okinawa 902-0076            => 1.5km from IT College
1-chōme-1-1-1 Kumoji, Naha, Okinawa 900-0015        => 1.4km from IT College
3-chōme-22-51 Makishi, Naha, Okinawa 900-0013       => 1.4km from IT College
Japan, 〒900-0021 Okinawa, Naha, Izumizaki, 2-chōme-１７-１５ 泉崎荘 => 0.8km from IT College
1-chōme-7-1 Yorimiya, Naha, Okinawa 902-0064        => 1.6km from IT College
1-chōme-30-9 Higawa, Naha, Okinawa 900-0022         => 1.0km from IT College
```

Figure 19: Found Addresses.

We have created 50 addresses representing centers of zones containing locations with low number of restaurants and no Mexican restaurants nearby, all zones being fairly close to city center (all less than 4km from IT College, and about half of those less than 2km from IT College). Although zones are shown on map with a radius of 500 meters (pink 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.

Figure 20: Viable addresses.

4 Results and Discussion

Our analysis shows that although there is a great number of restaurants in Naha (in our initial area of interest which was 6x6km around IT College), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from IT College. but our attention was focused on Adamaichi and Matsuo 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. 3x3km around from IT College) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Mexican 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 50 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and

Mexican 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 Berlin center but not crowded with existing restaurants (particularly Mexican) - 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.

5 Conclusion

Purpose of this project was to identify Naha areas close to center with low number of restaurants (particularly Mexican restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Mexican restaurant. By calculating restaurant density distribution from Foursquare data we have first identified addresses that justify further analysis (Adamaichi and Matsuo), and then generated extensive collection of all other 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 decission 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.

More Data and Time is required.

References

[Deva] Foursquare Developer. Foursquare developer portal and api reference. `https://developer.foursquare.com/`. Accessed: 2020-05-29.

[Devb] Google Developer. Capstone example project notebook. `https://developers.google.com/maps/documentation`. Accessed: 2020-05-29.

[ICa] IBM and Coursera. Applied data science specialization. `https://www.coursera.org/specializations/applied-data-sciencefour`. Accessed: 2020-05-29.

[ICb] IBM and Coursera. Capstone example project notebook. `https://cocl.us/coursera_capstone_notebook`. Accessed: 2020-05-29.