Capstone Project:

Restaurants in Metro Manila

August 05, 2021

A. Introduction: Business Problem

A1. Background

In this project, we would like to know what is the best possible place for a restaurant in Metro Manila. It is officially known as National Capital Region (NCR) and is the capital of the Philippines.

There are a whopping 21,017 hotel and restaurant establishments from a 2000 census done by the Philippine government. NCR is comprised of 897 barangays. These are grouped into six congressional districts. For the purpose of this project, we will group the restaurants by district.

So as part of this, we will list and visualize all major parts of NCR.

A2. Problem

Question that can be asked – What is the best location in NCR for a restaurant? – Which area have a large number of restaurants? – Which of the areas have the least number of restaurants? – Which is the best place to stay if I prefer Korean food? – What places have high rating of restaurants?

A3. Target

- Entrepreneurs who are interested in finding the best spot to open a restaurant.

- People who would like to know which locality is best to eat a specific food (ie. Korean).

- People who would like to eat in a restaurant based on the rating.

B. Data Acquisition and Cleaning

B1. Data Sources

For this project we need the following data:

1. NCR restaurant data that contains list of localities, hotel name, rating along with their latitude and longitude.

Data source: Zomato Kaggle Dataset: "https://www.kaggle.com/jeppbautista/zomato-dataset-for-metro-manila-restaurants?select=location.csv"

Description: This dataset contains the required information. We will use this dataset to explore various locality of NCR.

2. Nearby places in each locality of NCR.

Data source: FourSquare API: "https://developer.foursquare.com/"

Description: By using this API we will get all the venues in each neighborhood.

B2. Data Cleaning

The Zomato Kaggle Dataset has three different data consisting of the location, the ratings, and the cuisines each restaurant have. Each csv file shares an 'id' column. These datasets are then downloaded in the data assets column in Watson Studio.

As the location, ratings, and type of cuisine are necessary for the analysis of the project, the data downloaded were combined into one table. I combined the three datasets by the 'id' column they all share. The data frame ended up having 6830 restaurant samples and 35 columns.

There are several problems with the datasets. First, each dataset has many columns that are not necessary for the analysis. To solve this, I created a new data frame consisting of only the 'id', 'locality', 'longitude', 'latitude', 'cuisines', 'aggregate_rating', 'rating-text', 'votes'. I did this by including only the restaurant sample that has a defined longitude. I ended up having the same number of restaurant samples and 8 columns.

Second, there are NaN and zero values present in the data frame. To solve the zero values, I removed any restaurant samples that has an aggregate rating of 0.0. I ended up having 5962 restaurant sample left. To make sure that I have no NaN values in my data frame, I used the function df2.isnull().any() to check each column (df2 is the main data frame). Seeing as the cuisines column returned a boolean value True, I then looked for the row that contains a null value in its cuisines column. After figuring out the index, I then dropped the entire row from the data frame. This is because I need only restaurants with known cuisines for this project.

Third, the names of the restaurants were unavailable. Each restaurant only has a corresponding id number for identification. There is not much I can do to solve this.

After fixing these problems, I started on my data analysis. Also, before that, I fixed my column names to capital letters to make them neat.

C. Exploratory Data Analysis

C1. Visualize Metro Manila and its localities.

I initially created a map using the folium library to show the restaurants in Metro Manila. I gathered the latitude and longitude of Metro Manila as it is necessary to build a map. Before I started the entire analysis though, I imported and downloaded all necessary libraries needed for this project. Below is the map:

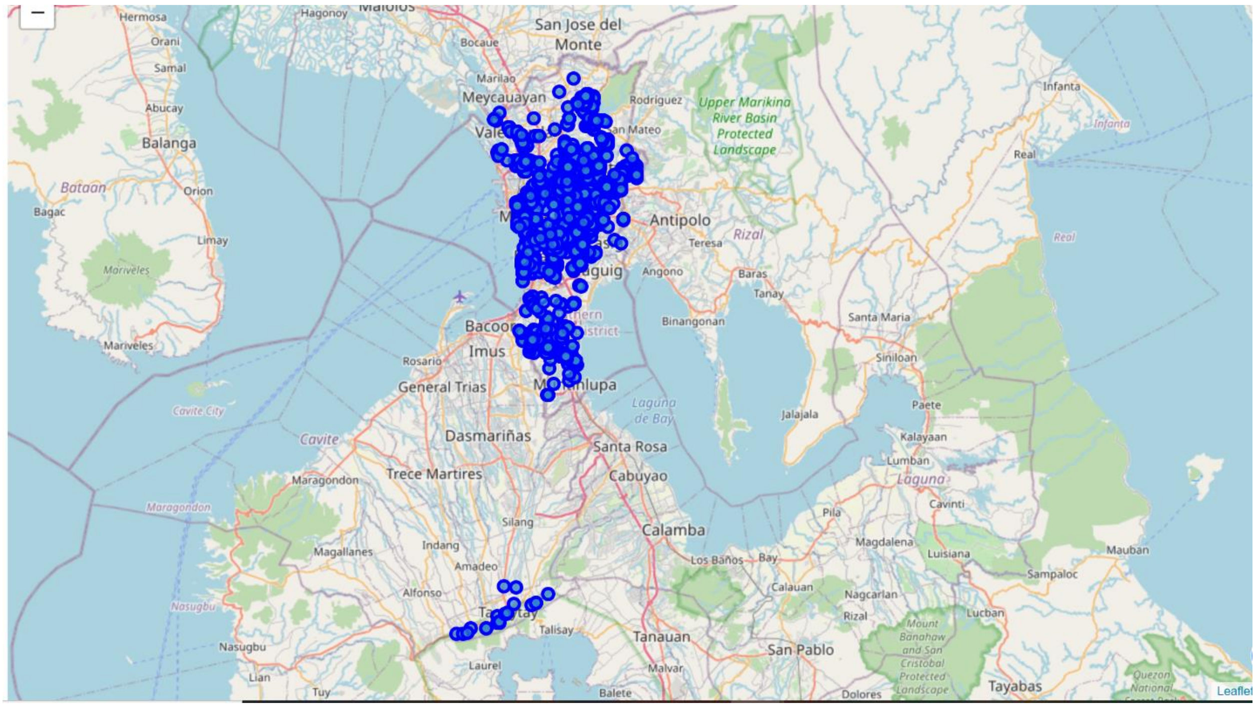

Figure 1: Map of Metro Manila Restaurants

From the map generated, we can see that most of the restaurants in Metro Manila are in situated around mostly the same area. In fact, there is only around 23 restaurants along Tagaytay.

C2. Approach

- Collect the Metro Manila data from Zomato Kaggle dataset.
- Using FourSquare API I will find all venues for each neighborhood.
- Filter out all venues that are nearby by locality.
- Using aggregate rating for each restaurant to find the best places.
- Visualize the ranking of neighborhoods using Folium library.

D. Results

D1.1 What places have the restaurants in Metro Manila?

To figure out which locality have the best rated restaurants in Metro Manila, I first group the samples by their locality. Then I took the mean of the aggregate rating of each locality and choose ten localities that has the largest mean. In order to be able to plot my results, I imported matplotlib.pyplot library. I then plotted this using a bar plot. The plot below displays the result:

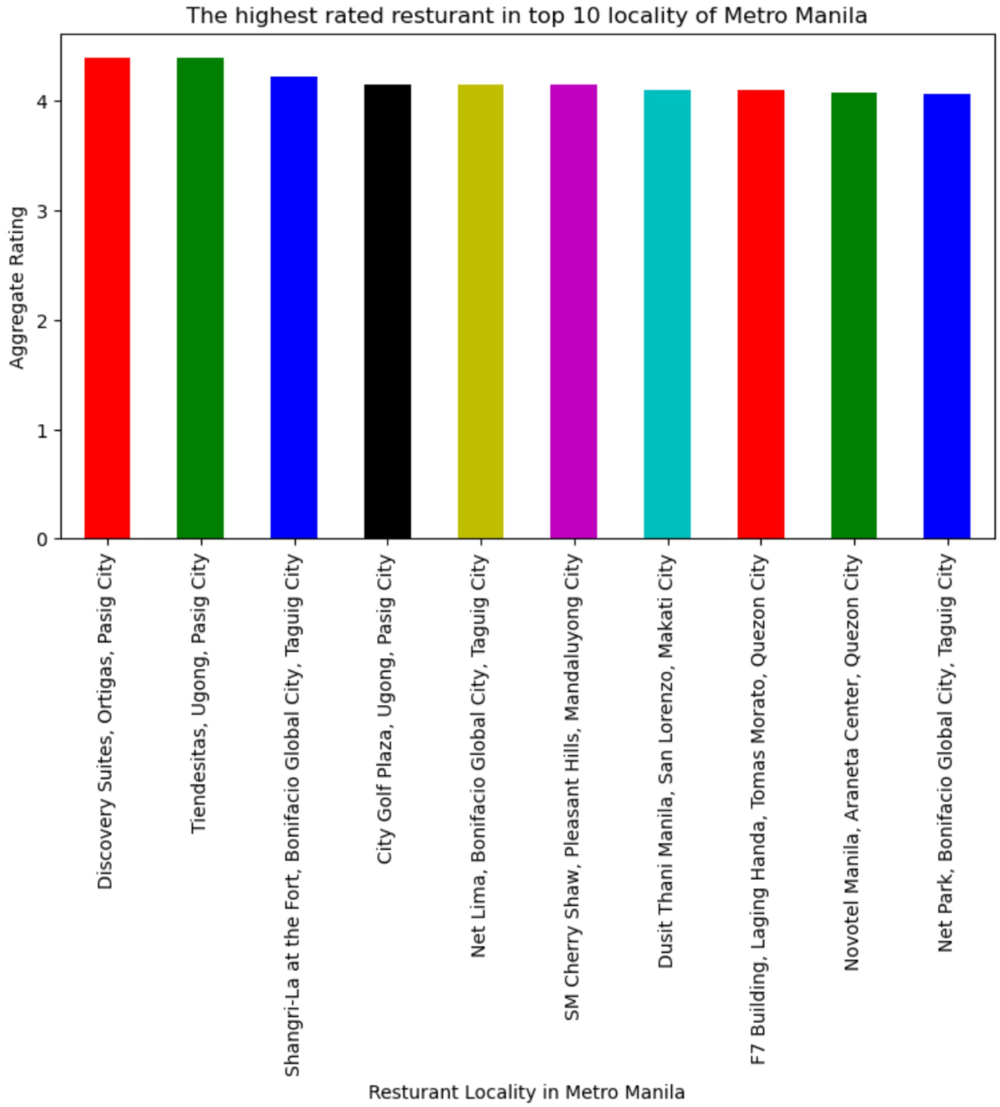

Figure 2: Plot of the Highest Rated Restaurants in Metro Manila

We can see that the best restaurants are located in Ortigas and Ugon; both located in Pasig City.

D1.2 What places have the worst rated restaurant in Metro Manila?

Using the same steps as in D1.1, except instead of choosing the top 10 highest aggregate rating mean I took the top 10 smallest mean.

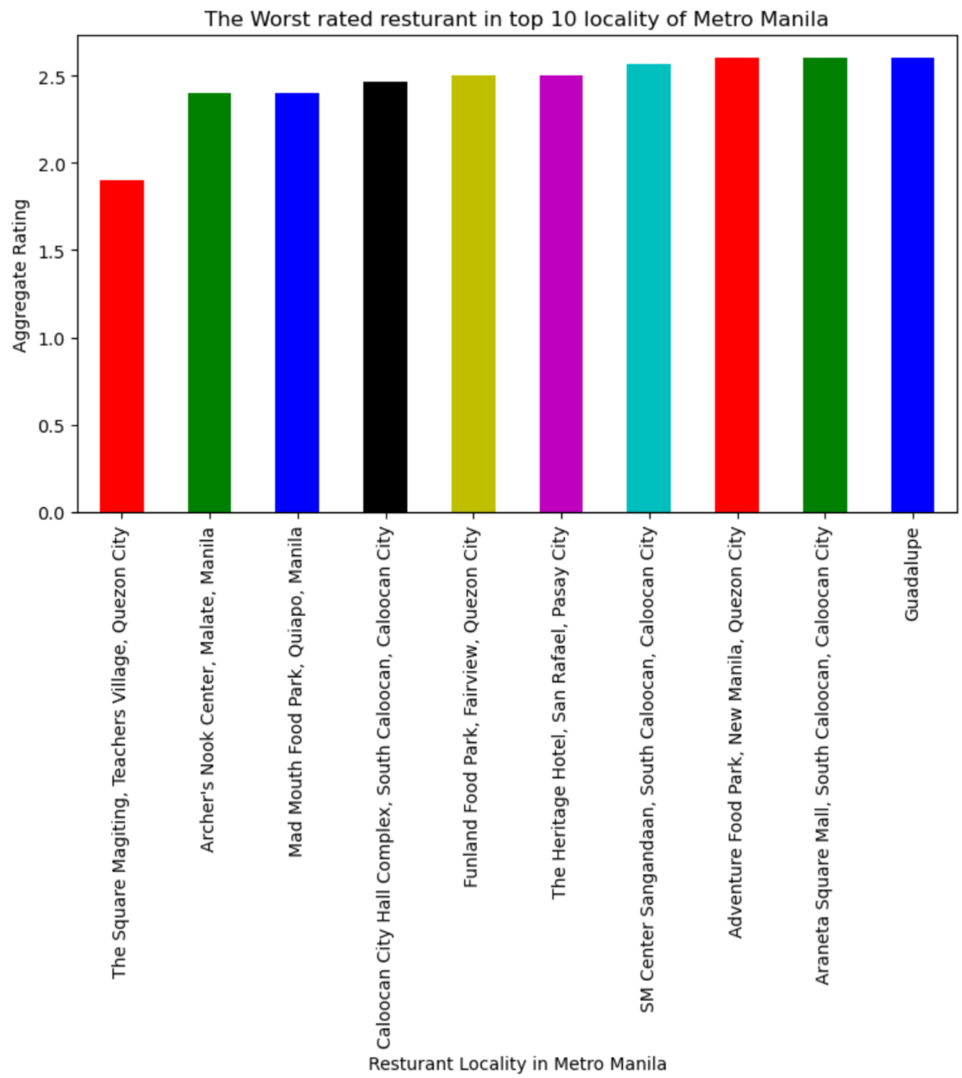

Figure 3: Plot of the Worst Rated Restaurants in Metro Manila

We can see that the worst rated restaurant in Metro Manila is located at The Square Magiting, Teachers Village, Quezon City.

D1.3 Which locality has the greatest number of restaurants?

To calculate how many restaurants are in each locality, I initially group the samples by their locality and then use the count function to count the number of ID. I then took the ten largest number of counts and plotted it using a bar plot.

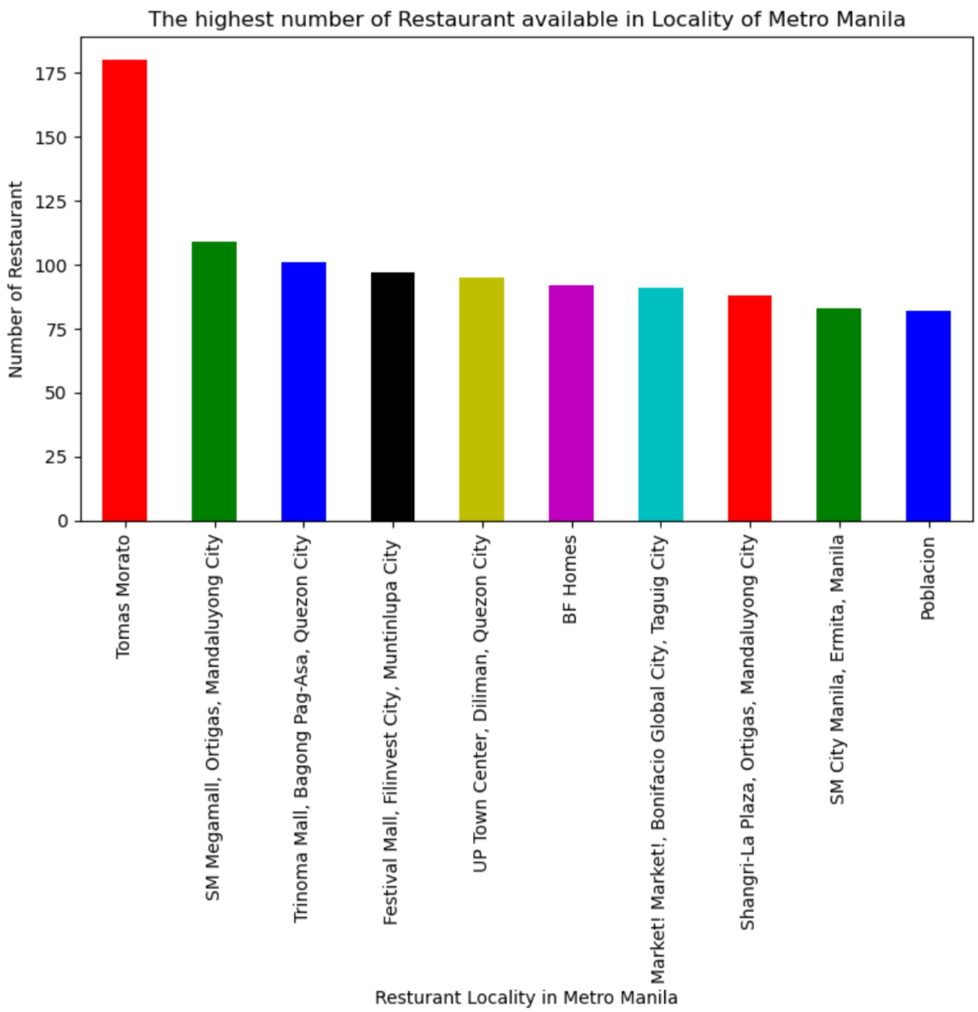

Figure 4: Plot of the Greatest Number of Restaurants in Metro Manila

If someone wants a lot of restaurant options to choose from, they should go to Tomas Morato.

D1.4 Which place does not have a lot of restaurants?

Again, the same process as D1.3, but now taking the ten smallest number of counts.

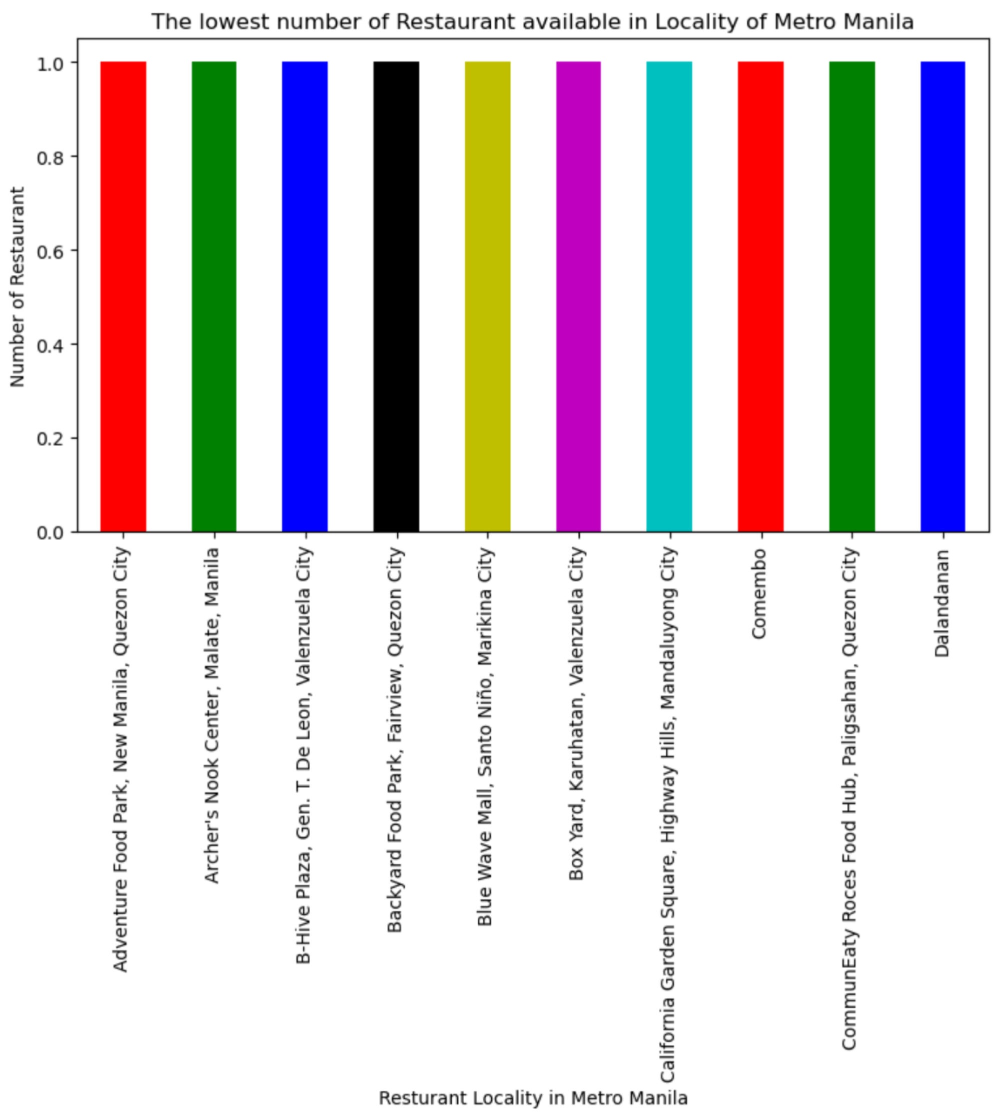

Figure 5: Plot of the Lowest Number of Restaurants in Metro Manila

As seen on the graph, each locality only has one restaurant available. Hence, if a person wants a lot of options, they should not go to any of these places.

D1.5 What are the best places for a Korean restaurant?

To answer this, I initially made a data frame containing of restaurants that has ‘Korean’ listed in its cuisine column. I then grouped these restaurants by their locality and counted how many restaurants are in each locality. I choose the first five localities that has the greatest number of Korean restaurants and plotted it.

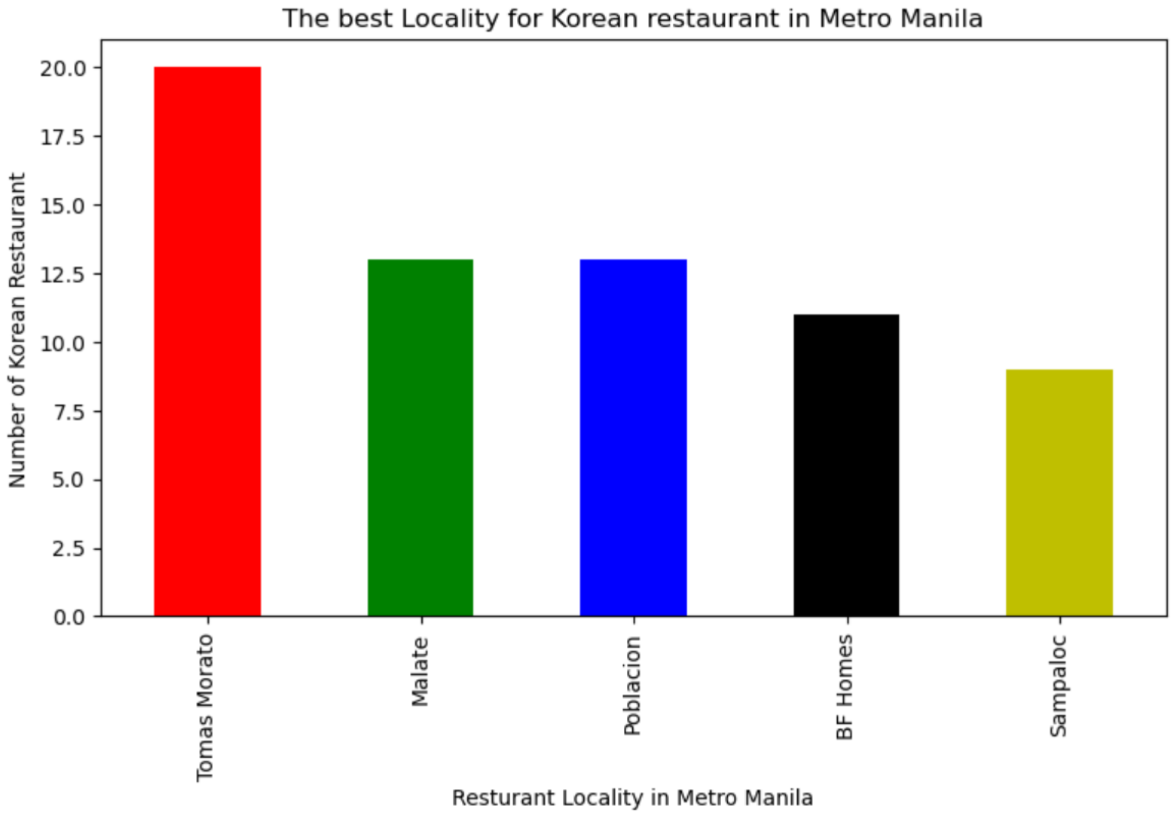

Figure 6: Plot of the Greatest Number of Korean Restaurants

We can see that if a person wants many choices of Korean restaurants, then Tomas Morato is the best place to go.

D1.6 What place has the best rated Korean restaurants?

Using the same process as D1.5, instead of counting the number of restaurants in each locality I took the mean of the aggregate rating. I then choose the first five highest rated locality and then plotted it.

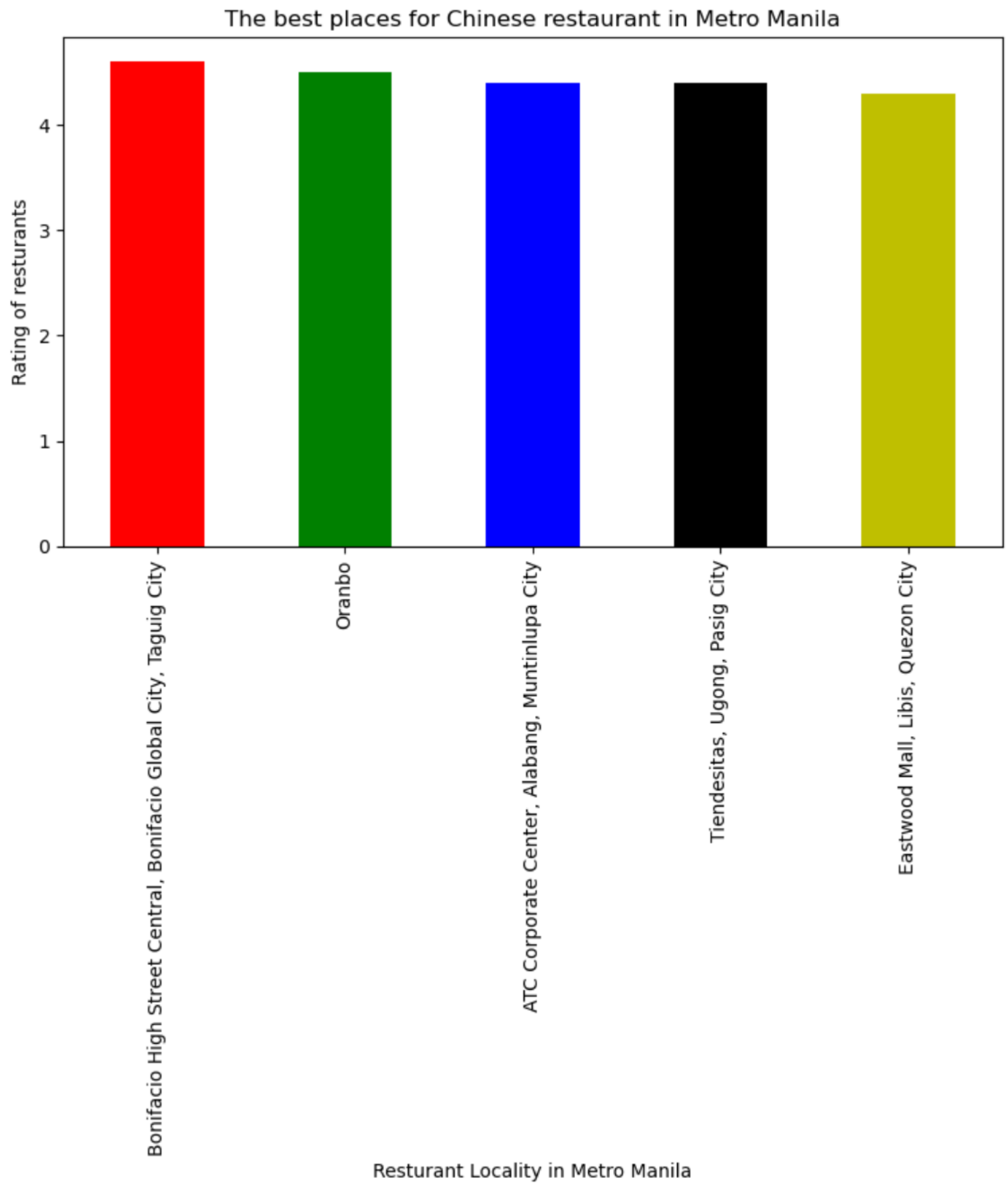

Figure 7: Plot of the Lowest Number of Korean Restaurants

Compared to the result I got from C2.6 where Tomas Morato has the greatest number of Korean restaurants, it does not have the highest rated Korean restaurants. In fact, Bonifacio High Street Central in Taguig City has the most rated Korean restaurants.

D2. Data Transformation

In order to find the venues for each locality, I utilized the FourSquare API to explore venues in Metro Manila and segment them.

Before that though, I want to make a new data frame that where the restaurant samples are grouped by their locality, the mean aggregate rating of each locality, the cuisines available, the ratings each restaurant got, the sum of votes each locality have, and the mean latitude and longitude of each locality. To make sure there are no zero values, I removed the restaurant samples that has an aggregate rating of 0. The new data frame consists of 446 localities and 8 columns. Below is the data frame that I made:

```
df_final.head()
```

Out[41]:

	Locality	Lat	Lng	No_of_Restaurant	Cusines	Agg_Rating	Comments	No_of_Votes
0	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	2	Chinese, Asian	3.400000	Good, Average	88
1	238 Center, Banawe, Quezon City	14.643305	121.000312	6	Chinese, Japanese, Tea, Bubble Tea, Steak, Ame...	3.466667	Excellent, Good, Average	716
2	A.Venue Mall, Poblacion, Makati City	14.566090	121.029942	6	Chinese, Tea, Desserts, Coffee, Middle Eastern...	3.683333	Good, Very Good, Average	555
3	AG New World Manila Bay Hotel, Malate, Manila	14.573424	120.982572	4	Chinese, Japanese, Filipino, Korean, Chinese, ...	3.700000	Good, Very Good, Average	171
4	ATC Corporate Center, Alabang, Muntinlupa City	14.424588	121.028302	7	Korean, Grill, Filipino, Mexican, Japanese, Su...	3.957143	Very Good, Good	972

Figure 8

After I defined my FourSquare credentials and version, I then created a function that would take the nearby venues to all localities in Metro Manila. The function requires the names of each locality, and the latitude and longitude. I have set the radius of search to 500 meter around the centre of each neighborhood and limit to 100 venues. It returns a data frame as the one below:

In [46]:
```
Metro_Manila_venues.head()
```

Out[46]:

	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	Pepper Lunch Express	14.602815	120.974099	Japanese Restaurant
1	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	S&R New York Style Pizza	14.603680	120.973579	Pizza Place
2	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	Sincerity Café & Restaurant	14.603601	120.973514	Chinese Restaurant
3	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	Dairy Queen	14.604480	120.972193	Ice Cream Shop
4	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	Happy Lemon	14.603501	120.973700	Bubble Tea Shop

Figure 9

From this, I got that there are 343 venues returned by FourSquare in Metro Manila.

I then analyzed each locality to see what kind of venue they have available. I did this by using dummy variables. Whenever a locality does not offer that specific category, it would be given a 0 value and whenever it caters a specific category it would be given 1. I ended up having a data frame as the one below:

```
In [52]: Metro_Manila_grouped.head(5)
```

Out[52]:

	Locality	Accessories Store	Airport	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium
0	168 Shopping Mall, Binondo, Manila	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.035294	0.0	0.0
1	238 Center, Banawe, Quezon City	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	A.Venue Mall, Poblacion, Makati City	0.000000	0.0	0.0	0.0	0.017857	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.017857	0.0	0.0
3	AG New World Manila Bay Hotel, Malate, Manila	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	ATC Corporate Center, Alabang, Muntinlupa City	0.019608	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Figure 10

I could only include a small portion of the data frame as there are 344 columns and 446 localities in the data frame. After this, I then took the top five most common venue in each locality and put it into a data frame.

```
In [56]: Locality_venues_sorted.head(5)
```

Out[56]:

	Locality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	168 Shopping Mall, Binondo, Manila	Chinese Restaurant	Fast Food Restaurant	Bakery	Pizza Place	Bubble Tea Shop	Japanese Restaurant	Grocery Store	Donut Shop	Asian Restaurant	Shopping Mall
1	238 Center, Banawe, Quezon City	Chinese Restaurant	Dessert Shop	Coffee Shop	Steakhouse	Dumpling Restaurant	Gas Station	Filipino Restaurant	Basketball Court	Spa	Bubble Tea Shop
2	A.Venue Mall, Poblacion, Makati City	Korean Restaurant	Café	Hostel	Coffee Shop	Fast Food Restaurant	Spa	Pizza Place	Middle Eastern Restaurant	Bar	Filipino Restaurant
3	AG New World Manila Bay Hotel, Malate, Manila	Japanese Restaurant	Bar	Korean Restaurant	Chinese Restaurant	Bakery	Middle Eastern Restaurant	Spa	Shabu-Shabu Restaurant	Restaurant	Clothing Store
4	ATC Corporate Center, Alabang, Muntinlupa City	Coffee Shop	Japanese Restaurant	Korean Restaurant	Café	Spa	Department Store	Dessert Shop	Filipino Restaurant	Burger Joint	Donut Shop

Figure 11

D3. Clustering

Now that I prepared all the data I need, I can use the k-Means clustering algorithm to group similar localities. However, before I perform k-Means I first have to know the appropriate number of clusters I would need. To run an evaluation to select the best k, the number of categories in the algorithm, I used the Elbow Method. Below is a figure of the evaluation.

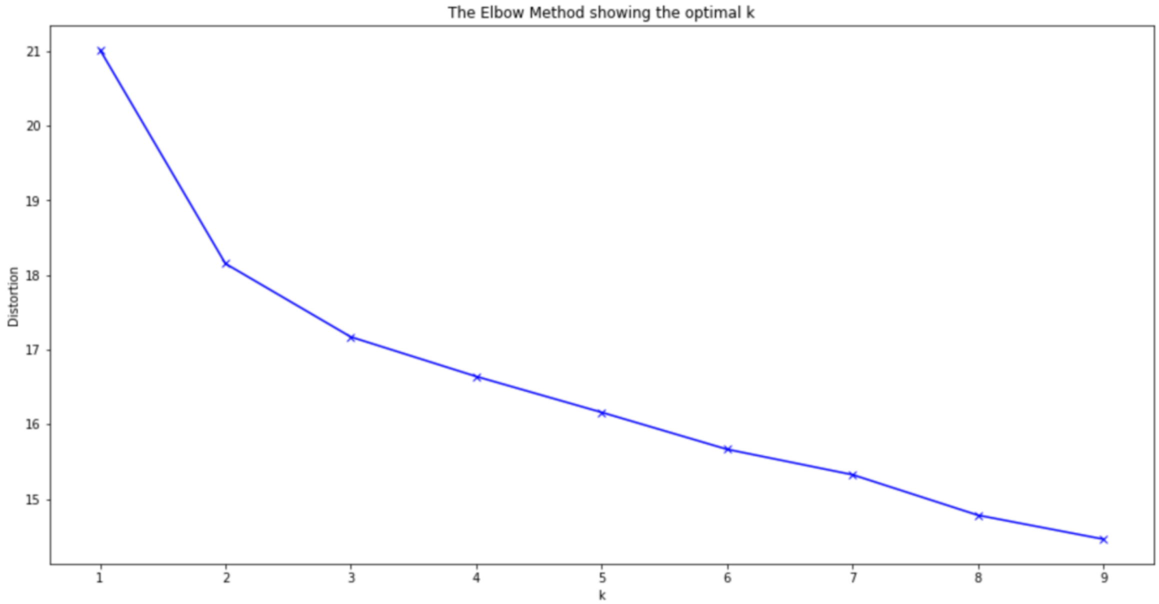

Figure 12

We can observe that the “elbow” is the number 3 which is optimal for this case. Now we can run a K-Means using as n_clusters the number 3. After we I performed the k-means clustering, I then created a new data frame including the clustering labels as well as the top ten venues for each locality. This new data frame was made by merging data frames of Figure 8 and Figure 11 with the addition of cluster labels.

	Locality	Lat	Lng	No_of_Restaurant	Cusines	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	168 Shopping Mall, Binondo, Manila	14.604273	120.973656	2	Chinese, Asian	3.400000	Good, Average	88	2	Chinese Restaurant	Fast Food Restaurant	Bakery	Pizza Place	Bubble Tea Shop	Japanese Restaurant	Grocery Store	Donut Shop	Asian Restaurant	Shopping Mall
1	238 Center, Banawe, Quezon City	14.643305	121.000312	6	Chinese, Japanese, Tea, Bubble Tea, Steak, Ame...	3.466667	Excellent, Good, Average	716	2	Chinese Restaurant	Dessert Shop	Coffee Shop	Steakhouse	Dumpling Restaurant	Gas Station	Filipino Restaurant	Basketball Court	Spa	Bubble Tea Shop
2	A.Venue Mall, Poblacion, Makati City	14.566090	121.029942	6	Chinese, Tea, Desserts, Coffee, Middle Eastern...	3.683333	Good, Very Good, Average	555	0	Korean Restaurant	Café	Hostel	Coffee Shop	Fast Food Restaurant	Spa	Pizza Place	Middle Eastern Restaurant	Bar	Filipino Restaurant
3	AG New World Manila Bay Hotel, Malate, Manila	14.573424	120.982572	4	Chinese, Japanese, Filipino, Korean, Chinese, ...	3.700000	Good, Very Good, Average	171	0	Japanese Restaurant	Bar	Korean Restaurant	Chinese Restaurant	Bakery	Middle Eastern Restaurant	Spa	Shabu-Shabu Restaurant	Restaurant	Clothing Store
4	ATC Corporate Center, Alabang, Muntinlupa City	14.424588	121.028302	7	Korean, Grill, Filipino, Mexican, Japanese, Su...	3.957143	Very Good, Good	972	0	Coffee Shop	Japanese Restaurant	Korean Restaurant	Café	Spa	Department Store	Dessert Shop	Filipino Restaurant	Burger Joint	Donut Shop

Figure 13

Finally, I visualized the resulting clusters using the Folium library.

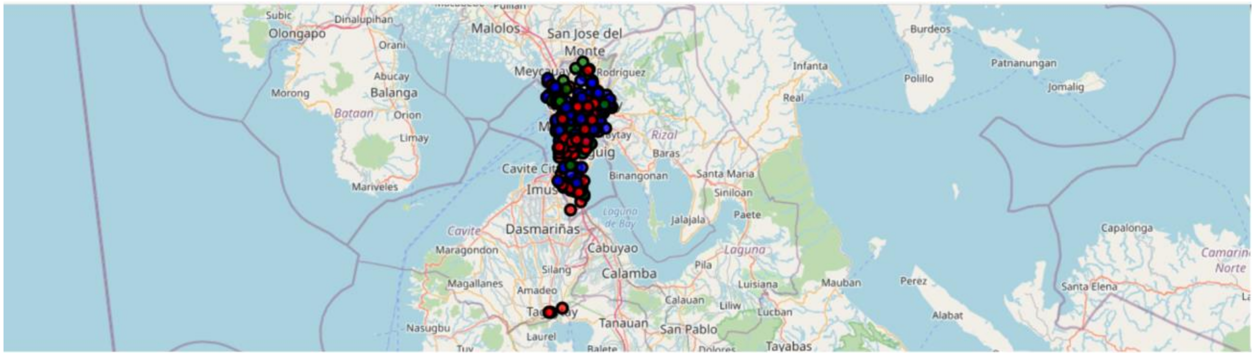

Figure 14

I also examined each cluster to determine the discriminating venue categories that distinguish each cluster.

```
In [63]: ## Cluster 1
Metro_Manila_merged.loc[Metro_Manila_merged['Cluster Labels'] == 0, Metro_Manila_merged.columns[[1] + list(range(5, Metro_Manila_merged.shape[1]))]].head(5)
```

Out[63]:

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	14.566090	3.683333	Good, Very Good, Average	555	0	Korean Restaurant	Café	Hostel	Coffee Shop	Fast Food Restaurant	Spa	Pizza Place	Middle Eastern Restaurant	Bar	Filipino Restaurant
3	14.573424	3.700000	Good, Very Good, Average	171	0	Japanese Restaurant	Bar	Korean Restaurant	Chinese Restaurant	Bakery	Middle Eastern Restaurant	Spa	Shabu-Shabu Restaurant	Restaurant	Clothing Store
4	14.424588	3.957143	Very Good, Good	972	0	Coffee Shop	Japanese Restaurant	Korean Restaurant	Café	Spa	Department Store	Dessert Shop	Filipino Restaurant	Burger Joint	Donut Shop
5	14.420648	3.233333	Good, Average	89	0	Café	Hotel Bar	Filipino Restaurant	Yoga Studio	Spa	Steakhouse	Diner	Burger Joint	Dessert Shop	Lounge
6	14.573472	3.375000	Good, Average	283	0	Coffee Shop	Japanese Restaurant	Breakfast Spot	Wings Joint	Hotel	Ice Cream Shop	Filipino Restaurant	Bakery	Supermarket	Chinese Restaurant

Figure 15: Cluster 1

Coffee Shops are the most recommended venues in Cluster 1 with 99 available. There are also 286 localities.

```
In [64]: ## Cluster 2
Metro_Manila_merged.loc[Metro_Manila_merged['Cluster Labels'] == 1, Metro_Manila_merged.columns[[1] + list(range(5, Metro_Manila_merged.shape[1]))]].head(5)
```

Out[64]:

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	14.656873	2.600000	Average	7	1	Fast Food Restaurant	Chinese Restaurant	Supermarket	Bookstore	Japanese Restaurant	Mobile Phone Shop	Filipino Restaurant	Donut Shop	Intersection	Arcade
29	14.686639	2.900000	Average	5	1	Fast Food Restaurant	Convenience Store	Bakery	Soup Place	Casino	Creperie	Filipino Restaurant	Eastern European Restaurant	Electronics Store	Empanada Restaurant
36	14.533019	2.950000	Average	11	1	Fast Food Restaurant	Hotel	Fried Chicken Joint	Diner	Filipino Restaurant	Casino	Convenience Store	Food Truck	Flea Market	Donut Shop
56	14.651056	2.466667	Poor, Average	23	1	Fast Food Restaurant	Convenience Store	Burger Joint	Bakery	Diner	Bubble Tea Shop	Bookstore	Mobile Phone Shop	Café	Chinese Restaurant
74	14.547407	2.900000	Average	4	1	Fast Food Restaurant	Convenience Store	Tapas Restaurant	Diner	Gym	Gym / Fitness Center	Pharmacy	Event Space	Drugstore	Dumpling Restaurant

Figure 16: Cluster 2

Fastfood restaurants are the most recommended here with 33 restaurants available. There are also 42 localities in Cluster 2.

```
In [65]: ## Cluster 3
Metro_Manila_merged.loc[Metro_Manila_merged['Cluster Labels'] == 2, Metro_Manila_merged.columns[[1] + list(range(5, Metro_Manila_merged.shape[1]))]].head(5)
```

Out[65]:

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	14.604273	3.400000	Good, Average	88	2	Chinese Restaurant	Fast Food Restaurant	Bakery	Pizza Place	Bubble Tea Shop	Japanese Restaurant	Grocery Store	Donut Shop	Asian Restaurant	Shopping Mall
1	14.643305	3.466667	Excellent, Good, Average	716	2	Chinese Restaurant	Dessert Shop	Coffee Shop	Steakhouse	Dumpling Restaurant	Gas Station	Filipino Restaurant	Basketball Court	Spa	Bubble Tea Shop
7	14.593377	3.650000	Very Good, Good, Average	1491	2	Convenience Store	Bar	Japanese Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Fast Food Restaurant	Spa	Bakery	BBQ Joint	Café
14	14.427325	3.500000	Very Good, Average, Good	436	2	Spa	Restaurant	Convenience Store	Filipino Restaurant	Basketball Court	Food & Drink Shop	Beer Garden	Badminton Court	BBQ Joint	Korean Restaurant
18	14.692027	3.118182	Good, Average	91	2	Convenience Store	Spa	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Asian Restaurant	Dumpling Restaurant	Filipino Restaurant	Pharmacy	Pizza Place

Figure 17: Cluster 3

Convenience stores are abundant here with 29 available. It is followed by 21 fast-food restaurants and 19 Chinese restaurants. There are 118 localities in Cluster 3.

E. Discussion

Looking at the unique elements in the '1st Most Common Venue' column of each cluster, it can be seen that Cluster 1 has the greatest number of Korean restaurants with 9 available. Cluster 1 is also where coffee shops/cafes are most abundant. If a person is not demanding of the type of cuisine they eat, places in Cluster 1 would be best. It has Filipino, Japanese, Chinese, Korean, Italian, and Seafood restaurants. Cluster 2 would also be viable if they do not mind fast food restaurants instead. Cluster 3 also has a lot of fast-food restaurants, but unlike Cluster 2 where it the only viable option is a Chinese restaurant, Cluster 2 has more other options. It has a Filipino, Chinese, Thai, Korean, Asian, Steakhouse, and Italian restaurants.

F. Conclusion

To iterate, the best rated restaurants in Metro Manila are both in Pasig City, particularly in Ortigas and Ugong. The worst rated restaurant is in Teachers Village, Quezon City. Tomas Morato has the greatest number of restaurants and also has the greatest number of Korean restaurants. And Bonifacio Global City in Taguig City has the best rated Korean restaurant.

Foursquare data and K-means algorithm, alongside with folium mapping tools can be powerful tools helping people narrow down the places that caters a restaurant according to their tastes. With many options to choose from a little help from a data scientist can really make the process much smoother.