

# Popular Type of Restaurants in Quezon City

Joshua Louis D. Isanan

August 14, 2019

## 1. Introduction

### 1.1 Background

The Philippines is a South East Asian country which brings millions of visitors every year, the country is also known to have many high-quality restaurants [1]. Quezon City is one of the largest cities in terms of land area and population in the Philippines [2], located in the Luzon area.

### 1.2 Problem

While it is true that there are many restaurants in the Philippines, especially in the Luzon area. There is still a lack of data as to which restaurant categories are prevalent in a lot of areas in the Philippines, especially in larger and more populous areas such as Quezon City.

### 1.3 Interest

For people who are planning to build restaurants in Quezon City, data such as which restaurant categories are doing well, or more frequently occurring would aid their decision making in the future as to what type of restaurants they'll try to build. It would also help them decide as to where to build the restaurant as location may be a factor because of things such as competition and natural resources nearby. A research shows that a restaurant's mortality is influenced by location, size and affiliation [3]. Which further proves the point that location may be a helpful factor in the decision-making of future restaurant owners.

## 2. Data Preparation

### 2.1 Data Sources

To get the locations and districts within Quezon City, zip codes were used. The source for the zip codes within Quezon City were scraped from Lhiza's Corner which is a site that advertises a german school but at the same time collects information from the Philippines [4]. To retrieve the latitude and longitude to pinpoint the location of a district using only the zip codes, ArcGIS and their services were used. ArcGIS is an

online geocoding service that turns addresses into coordinates [5]. To retrieve the nearby restaurants and their respective categories using the latitudes and longitudes, Foursquare's Places API was used. Foursquare is a location data and intelligence company and an independent location data platform [6].

## 2.2 Data Wrangling and Web Scraping

Using beautifulsoup4, which is a library that assists in scraping information from web pages [7], zip codes from Quezon City were retrieved along with their location names. Pandas, which is a data analysis tool that offers data structures such as series and tables [8], were used to create a data frame of zip codes and location data provided by arcGIS.

	Code	Location	Latitude	Longitude
0	1105	Alicia	14.651771	121.025405
1	1102	Amihan	14.629034	121.058258
2	1106	Apolonio Samson	14.662860	121.006515
3	1106	Baesa	14.662860	121.006515
4	1116	Bagbag	14.690154	121.036538

Figure 1 - Partial Data Frame of Location, Post codes, Latitudes and Longitudes.

Foursquare's API was then used to retrieve nearby restaurants within the given radius. After Foursquare's API was used, it was then shown that one location did not have any restaurants near it. That said location was Fairview, and was then removed from the data frame. Using Foursquare's API, a data frame of every restaurant was then created. One-hot encoding was then performed on the data frame of every restaurants for use in modeling later, then the data frame was grouped by Location so that each category column would have their weighted representation in frequency.

Location	American Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar	Bistro	Breakfast Spot	Bubble Tea Shop	...	Tapas Restaurant	Tea Room	Tex-Mex Restaurant	Thai Restaurant	Rest
0 Alicia	0.0	0.000000	0.040000	0.0	0.040000	0.000000	0.0	0.0	0.040000	...	0.0	0.080000	0.0	0.0	
1 Amihan	0.0	0.041667	0.062500	0.0	0.125000	0.000000	0.0	0.0	0.020833	...	0.0	0.062500	0.0	0.0	
2 Apolonio Samson	0.0	0.000000	0.000000	0.0	0.000000	0.052632	0.0	0.0	0.052632	...	0.0	0.000000	0.0	0.0	
3 BF Homes	0.0	0.000000	0.029412	0.0	0.058824	0.000000	0.0	0.0	0.029412	...	0.0	0.058824	0.0	0.0	
4 Bagbag	0.0	0.037037	0.037037	0.0	0.037037	0.000000	0.0	0.0	0.000000	...	0.0	0.000000	0.0	0.0	

Figure 2 - Partial Data Frame with weighted restaurant category frequency

Another data frame was made as well that contains locations with their respective top five most frequent restaurant categories within the radius of the location. This data frame will be used later after the modeling process. Cluster labels will be appended to this data frame as it is much easier to read and manipulate than having the cluster label be appended weighted data frame.

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alicia	Filipino Restaurant	Fast Food Restaurant	Tea Room	Snack Place	Chinese Restaurant
1	Amihan	Fast Food Restaurant	Bakery	Tea Room	Food Truck	BBQ Joint
2	Apolonio Samson	Café	Pizza Place	Filipino Restaurant	Chinese Restaurant	Fast Food Restaurant
3	BF Homes	Filipino Restaurant	Snack Place	Korean Restaurant	Comfort Food Restaurant	Coffee Shop
4	Bagbag	Filipino Restaurant	Fast Food Restaurant	Diner	Donut Shop	Fish & Chips Shop

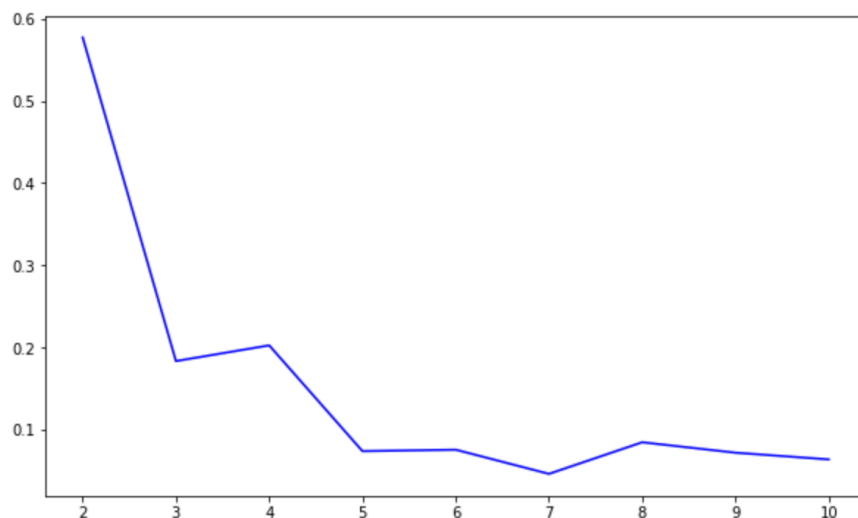
*Figure 3 - Partial Data Frame with Location names and the top 5 most frequent restaurant categories*

### 3. Modeling

#### 3.1 - Algorithm

The machine learning algorithm that will be used for modeling is K-means Clustering. This machine learning algorithm is supervised and is widely-used among the Data Science community. The main objective of K-means is to group data points that are similar, together, to discover underlying patterns. K-means aims to have a fixed number of k clusters in the dataset[9].

#### 3.2 - Choosing the right K



*Figure 4 - Elbow Method on choosing K*

A common problem when doing K-means clustering is knowing what the optimum number of K is. Though there is no definite answer to this question, there are however, some methods that will aid in choosing what K to use in modeling.

One of the common methods in choosing K is called the elbow method. It is commonly done with a line plot of each point in the dataset, the plot will look like an arm and the “elbow” on that arm is the value of k that is optimal[10]. For our score basis, we will be using silhouette score. Figure 4 shows the plot with the x-axis being the number of K and the silhouette score being the values in the y-axis. The number of k chosen based on the plot is five.

### 3.3 - Modeling

With five being the number of clusters or K, K-means clustering is performed again with the new value of K. The cluster labels were retrieved, modified so that the labels will start at 1 instead of 0, then appended in the data frame containing the location and the five most common restaurant categories within the radius of the location.

	index	Code	Location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	0	1105	Alicia	14.651771	121.025405	1	Filipino Restaurant	Fast Food Restaurant	Tea Room	Snack Place	Chinese Restaurant
1	1	1102	Amihan	14.629034	121.058258	4	Fast Food Restaurant	Bakery	Tea Room	Food Truck	BBQ Joint
2	2	1106	Apolonio Samson	14.662860	121.006515	1	Café	Pizza Place	Filipino Restaurant	Chinese Restaurant	Fast Food Restaurant
3	4	1116	Bagbag	14.690154	121.036538	5	Filipino Restaurant	Fast Food Restaurant	Diner	Donut Shop	Fish & Chips Shop
4	5	1109	Bagong Buhay	14.620135	121.063451	4	BBQ Joint	Burger Joint	Asian Restaurant	Dim Sum Restaurant	Restaurant

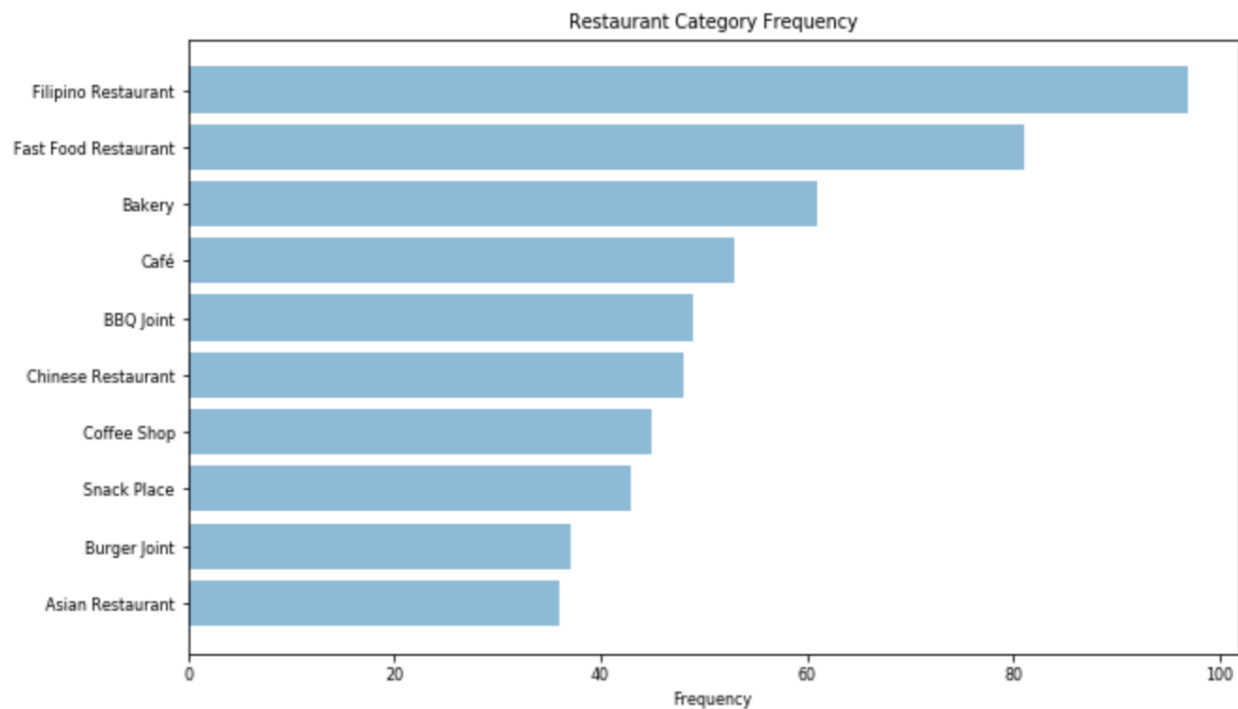
*Figure 5 - Partial Data Frame containing the Cluster Labels*

## 4. Results and Discussion

### 4.1 Restaurant Category Frequency

Using the data frame containing all the nearby restaurants in each district in Quezon City, Matplotlib, a Python plotting library used for data visualization[11] was used to create a horizontal bar chart for which restaurant categories are most common in all of the districts in Quezon City.

Based on the chart in Figure 6, the restaurant category that’s most common among all the locations are Filipino restaurants. Though this outcome has been expected as Filipinos tend to love dining in places where they are familiar with the cuisine. The second most frequent are fast food restaurants. There is a large percentage of Filipinos still living below the national poverty line[12], fast food restaurants in the Philippines

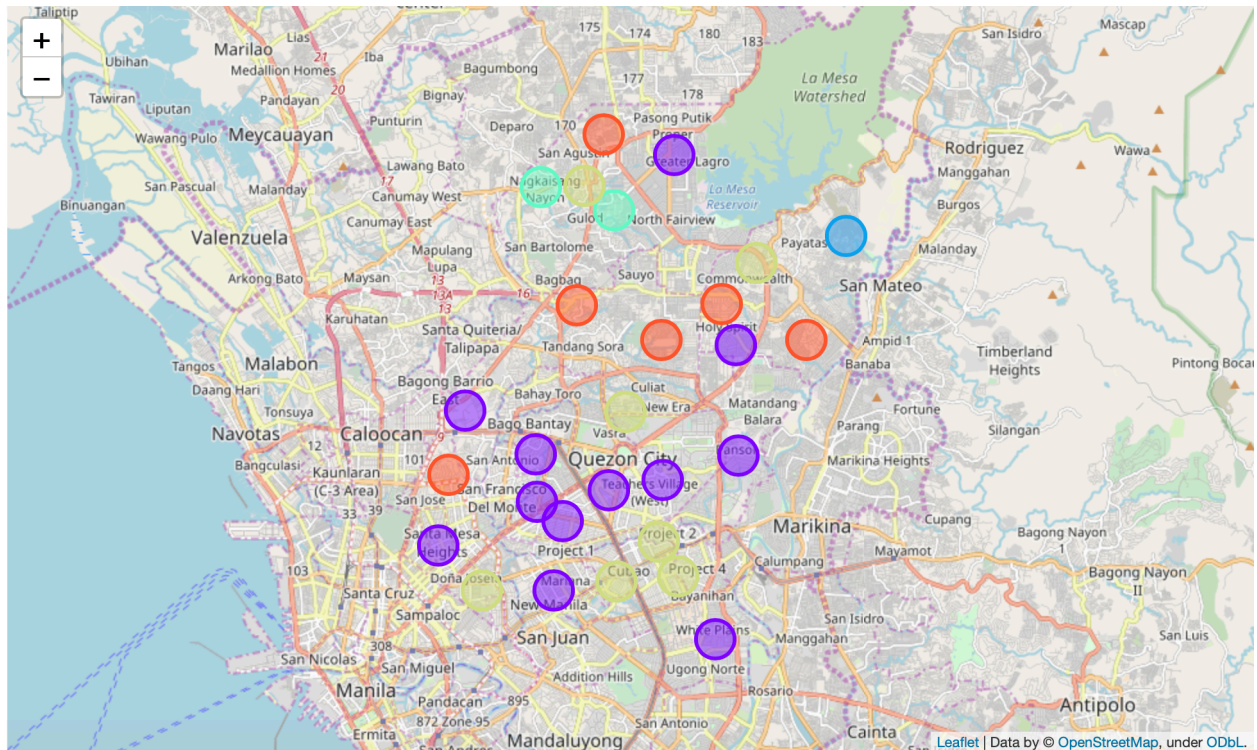


*Figure 6 - Restaurant Category Frequency Bar Chart*

offer a much cheaper dining experience relative to other type of restaurants. Fast food restaurants also offer their food in a much faster rate as compared to other type of restaurants which make them a good choice for fast dining, which some Filipinos could prefer over the other options, especially the workers in the BPO industry which has an annual growth rate of 17% and could produce 1.3 million new jobs on 2016[13]. Bakeries are also in the same case as Fast Food Restaurants, they provide a fast and relatively cheap way to dine in the Philippines and that may be a factor as to why there are a lot of bakeries within the area.

## 4.2 Clustering Results

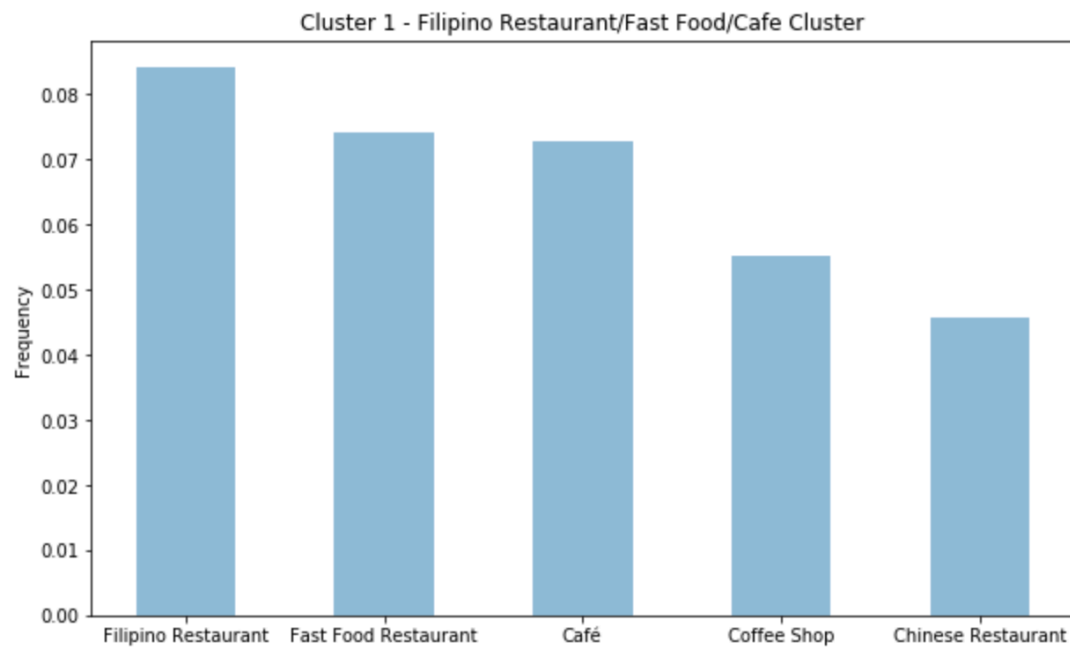
Using Folium, a Python library that is used to visualize data on a leaflet map[14]. The locations along with their corresponding latitudes and longitudes, are plotted on the map, the color of the circle marker, depicting the cluster in which the location belongs to.



*Figure 7 - Leaflet Map with Markers color coded by cluster membership*

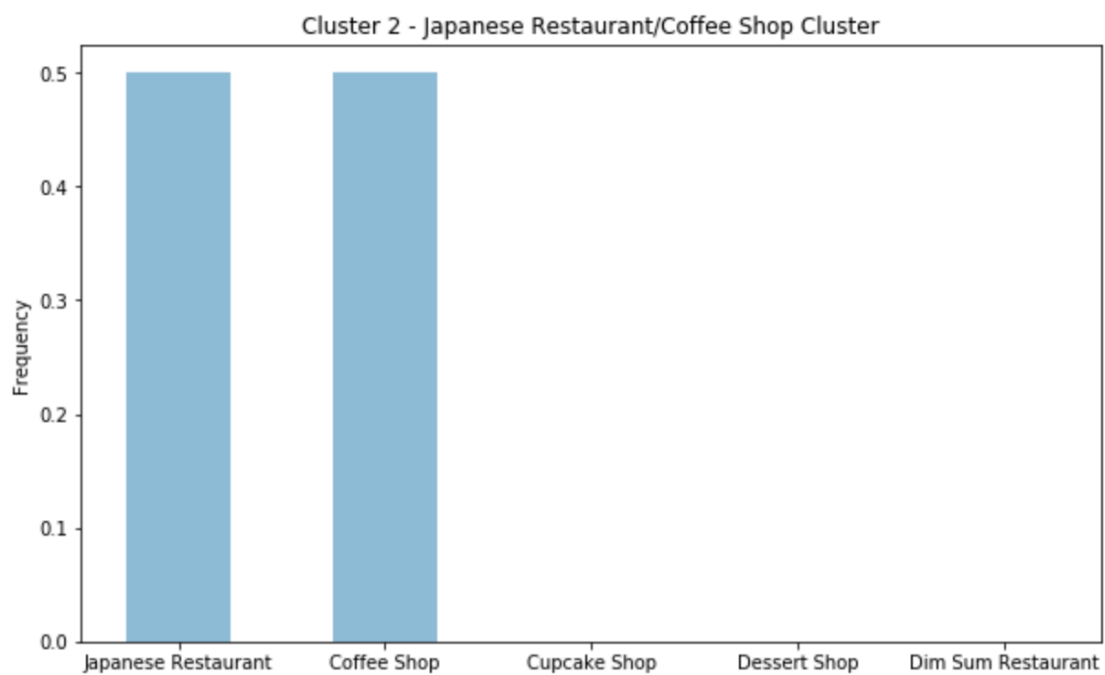
Based on the map in Figure 7, Cluster 1, which makes up majority of the circle markers, are colored purple. Cluster 2, only having one member inside of the cluster, being blue. Cluster 3, which makes up only two of the locations or circle markers, being sky-blue. Cluster 4 being a darker shade of yellow and Cluster 5 being red. We can see that based on the map plot alone and the cluster labels, the amount of information conveyed is not enough to make assumptions.

It is necessary to show what each cluster represents in terms of restaurant category frequencies within it.



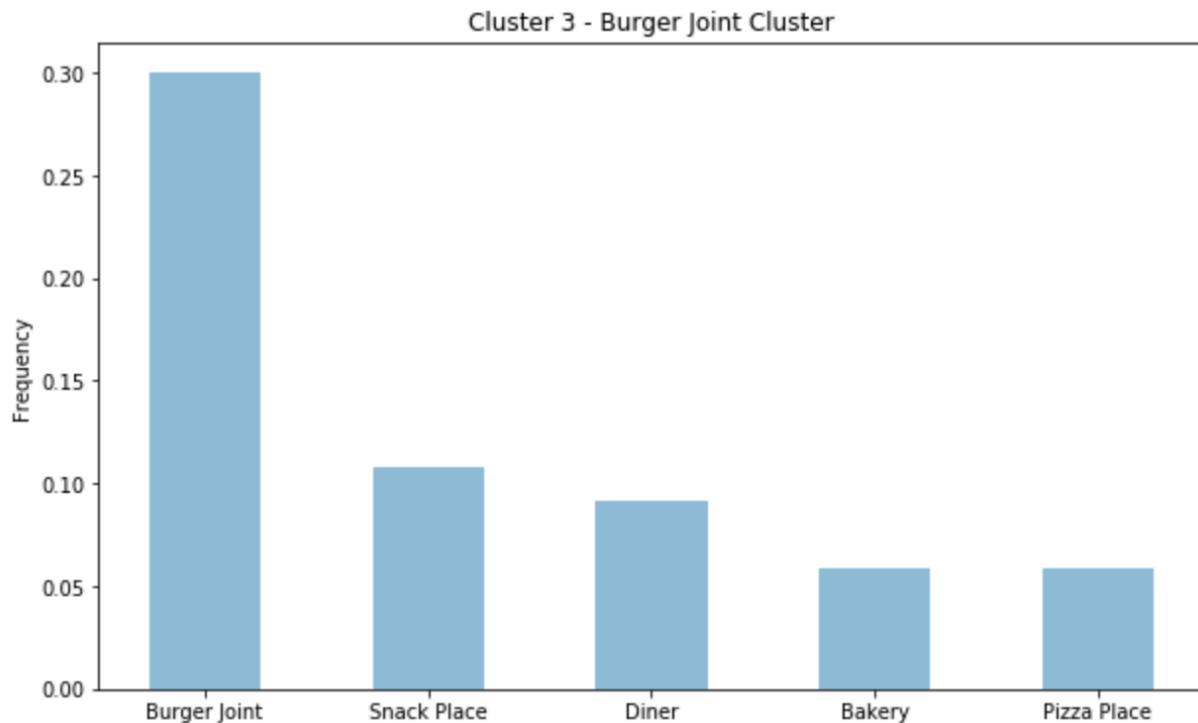
*Figure 8 - Restaurant Category frequency in Cluster 1*

In Cluster 1, or Filipino Restaurant/Fast Food/Cafe Cluster. It is shown that Filipino Restaurants, Fast Food Restaurants and Cafes make up the most in terms of frequency of restaurant category.



*Figure 9 - Restaurant Category Frequency in Cluster 2*

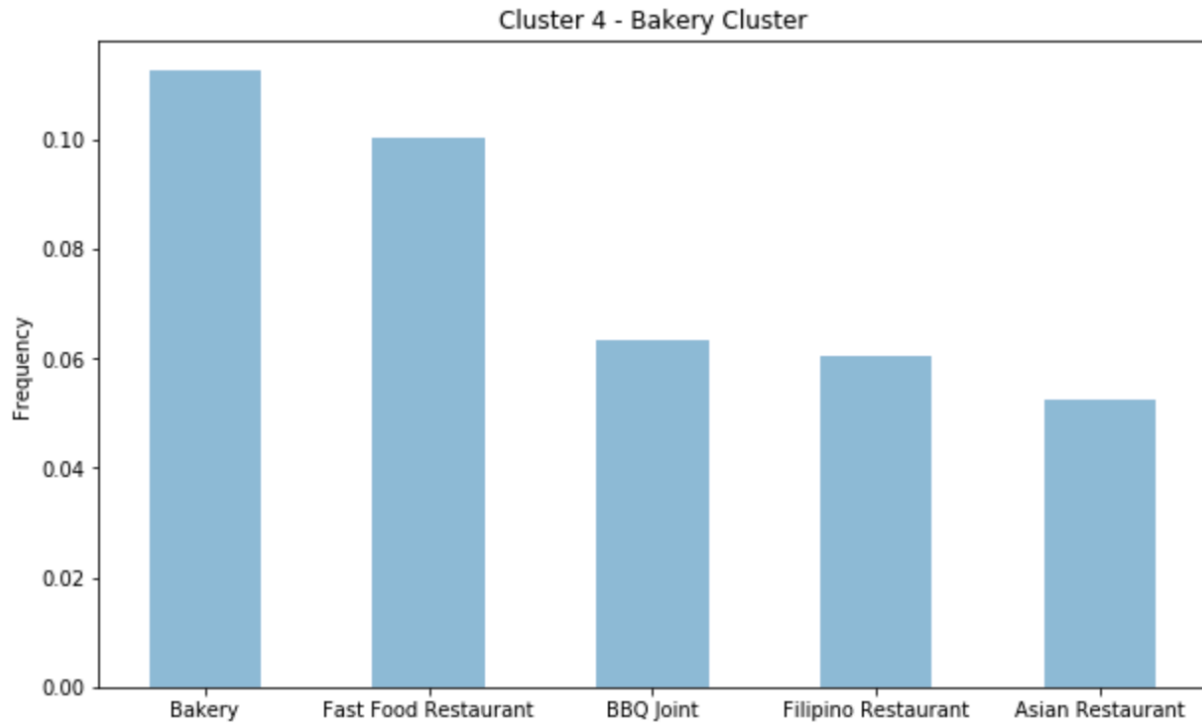
Cluster 2, which is the Japanese Restaurant / Coffee shop cluster. There is only one location that is a member of this cluster. The location associated with Cluster 2 is close to the sea, it may be associated or is a factor as to why a Japanese Restaurant is within the location, as Japanese Restaurants usually serve seafood in their dishes. Though that assumption may not reflect the actual truth because there is a lack of restaurants within the area.



*Figure 10 - Restaurant Category Frequency in Cluster 3*

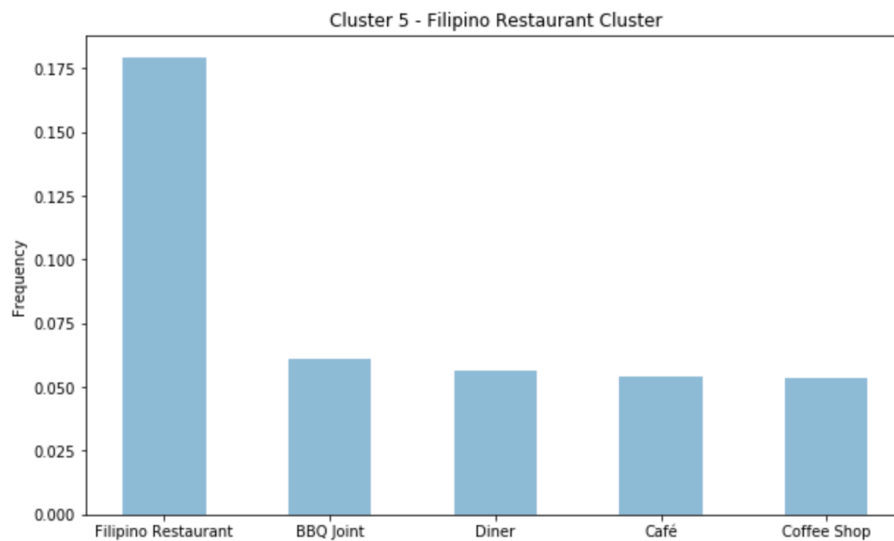
In Cluster 3, it is apparent that Burger Joints are the most frequent restaurant categories. The other restaurant categories in the top five also shows a bit of a similarity with each other as Snack Place, Bakery and Pizza Place are places where people eat snacks. The people in these locations may prefer going to food establishments that serve lighter and a more snack-friendly dishes.





*Figure 11 - Restaurant Category Frequency in Cluster 4*

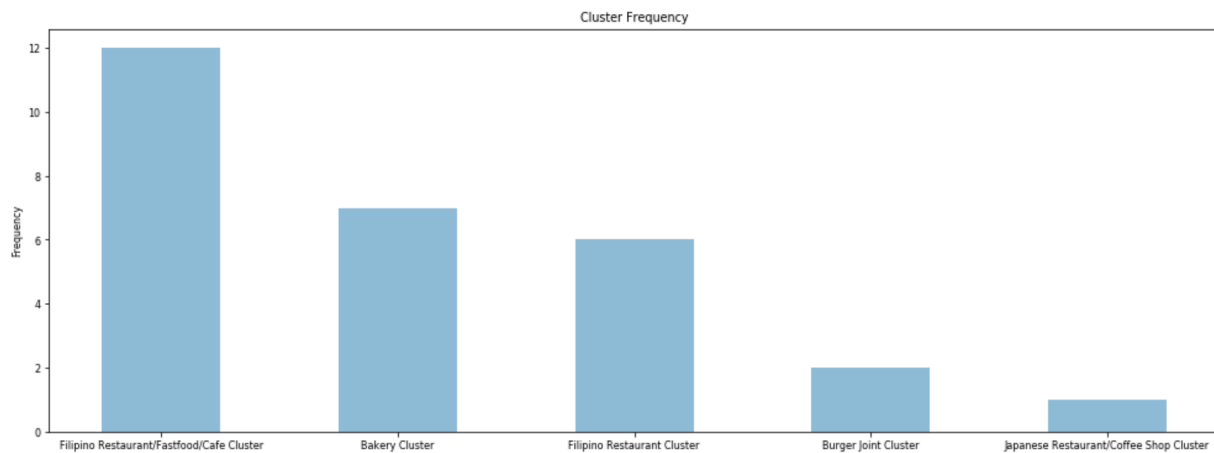
Cluster 4, which is the Bakery Cluster. This cluster shows that Bakeries and Fast Food Restaurants are the most frequent restaurant categories in these locations. Bakeries and Fast Food restaurants are known to serve their customers meals in a much faster rate as compared to other restaurant types.



*Figure 12 - Restaurant Category Frequency in Cluster 5*

Cluster 5, also labeled as Filipino Restaurant Cluster, show a much higher preference in Filipino Restaurants as it has a higher frequency difference as compared with the next category which is the Barbecue Joint. The people in these locations may heavily prefer a more traditional meal or local cuisine as compared to other types of restaurants.

With each cluster having different sets of top five restaurant categories that is most common within it. A visualization of which clusters make up the majority of the cluster population within Quezon City.



*Figure 13 - Cluster Frequency within Quezon City*

Based on the data shown in Figure 13, the first cluster which is the Filipino Restaurant/ Fastfood/ Cafe Cluster is shown to have the biggest frequency of population within Quezon City using the model. There is a clear gap between the top cluster and the cluster next to it in terms of frequency. The top cluster in the ranking and the third cluster in the ranking has the same top most frequent restaurant type which is Filipino Restaurant. Majority of the locations within Quezon City prefer to have Filipino Restaurants as it is the most populous and most frequent restaurant type among majority of the clusters. While other types of restaurants such as Bakeries and Fast Food are also common within Quezon City.

## 5. Conclusion

### 5.1 Conclusion

The study has been able to successfully retrieve restaurant category data within Quezon City with the aid of the Foursquare API, arcGIS' geocoder and the post code data for Quezon City locations. It has also been able to create a model using the K-means algorithm to create clusters within Quezon City based on the means or weights of

restaurant categories in each location. The results and the models could be used to aid in decision-making for future restaurant owners, like where they would build the restaurant, and if the location that they've picked would be ideal for the type of restaurant they would like to build. It could also help them decide as to what type of restaurant they would build if they've already picked a location.

## 5.2 Future Studies

Though the model was able to create results that could be of use for analysis, it could still be improved with better data, such as a more accurate location provider than arcGIS' geocoder. With expert opinion, the parameters that were used in modeling may have not been the optimal choices. Future studies may look into other methods of modeling using the k-means algorithm or other clustering algorithms or machine learning algorithms. Foursquare's API may not also be the best in terms of collecting restaurants in the Philippines so future research on this could use other location data platforms in retrieving restaurants.

## 6. References

- [1] Top 10 Restaurants in the Philippines  
<https://traveltips.usatoday.com/top-10-restaurants-philippines-100801.html>
- [2] [http://www.wpro.who.int/philippines/areas/urban\\_health/urban\\_poor/rup\\_quezon\\_cityprofile.pdf](http://www.wpro.who.int/philippines/areas/urban_health/urban_poor/rup_quezon_cityprofile.pdf)
- [3] Parsa, H & Self, John & Sydnor, Sandra & Jin Yoon, Hae. (2011). Why Restaurants Fail? Part II - The Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis. Journal of Foodservice Business Research. 14. 360-379. 10.1080/15378020.2011.625824.
- [4] <http://www.Lhiza.com>
- [5] ArcGIS Online Geocoding Service <https://geocode.arcgis.com/arcgis/>
- [6] Foursquare <https://developer.foursquare.com/>
- [7] beautifulsoup4 <https://pypi.org/project/beautifulsoup4/>
- [8] Pandas <https://pandas.pydata.org/>
- [9] Understanding K-means Clustering in Machine Learning  
<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

[10] Determining Optimal Number of Clusters: 3 Must Know Methods

<https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/>

[11] <https://matplotlib.org/>

[12] [https://psa.gov.ph/poverty-press-releases/nid/138411?](https://psa.gov.ph/poverty-press-releases/nid/138411?fbclid=IwAR1VdklGWMKqF9nENgAIHqRIHimpZq2Yh7qmo-bjyEla14Ww-R-ajqldBs4)

[fbclid=IwAR1VdklGWMKqF9nENgAIHqRIHimpZq2Yh7qmo-bjyEla14Ww-R-ajqldBs4](https://psa.gov.ph/poverty-press-releases/nid/138411?fbclid=IwAR1VdklGWMKqF9nENgAIHqRIHimpZq2Yh7qmo-bjyEla14Ww-R-ajqldBs4)

[13] <https://www.flatworldsolutions.com/articles/growth-bpo-industry-philippines.php>

[14] <https://python-visualization.github.io/folium/>