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, located in the Luzon area, is one of the largest cities in terms of land area and population in the Philippines [2].

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 should build. It would also help them decide as to where to build the restaurant as location may be a factor because of aspects 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, both Google and ArcGIS and their services were used.

ArcGIS is an online geocoding service that turns addresses into coordinates [5]. Google's geocoding API is another service used in this study. It provides geocoding services similar to arcGIS. It allows the developers to convert an address to geographic coordinates that can then be used to plot on a map [15]. 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 Google and arcGIS. The search is first done by using the location name as the parameter. If Google does not give a result or if it cannot find the location, the search is then done by using the zip code of the location. If by then there is still no result, arcGIS would then be used for retrieving the latitude and longitude by passing the zip code of the location.

	Code	Location	Latitude	Longitude
0	1105	Alicia	14.661434	121.025268
1	1102	Amihan	14.632086	121.068624
2	1106	Apolonio Samson	14.655931	121.007706
3	1106	Baesa	14.671225	121.011409
4	1116	Bagbag	14.696551	121.032464

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 it 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 restaurant to be used 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 also made which 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 to the weighted data frame.

Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common
Alicia	Filipino Restaurant	Juice Bar	Fast Food Restaurant	Japanese Restaurant	Chinese Rest
Amihan	Filipino Restaurant	Burger Joint	Bakery	Chinese Restaurant	
olonio Samson	Bakery	Café	BBQ Joint	Filipino Restaurant	Pizza
BF Homes	Filipino Restaurant	Snack Place	Korean Restaurant	Asian Restaurant	
Baesa	Fast Food Restaurant	Snack Place	Chinese Restaurant	BBQ Joint	Filipino Rest

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

3. Modeling

3.1 - Algorithm

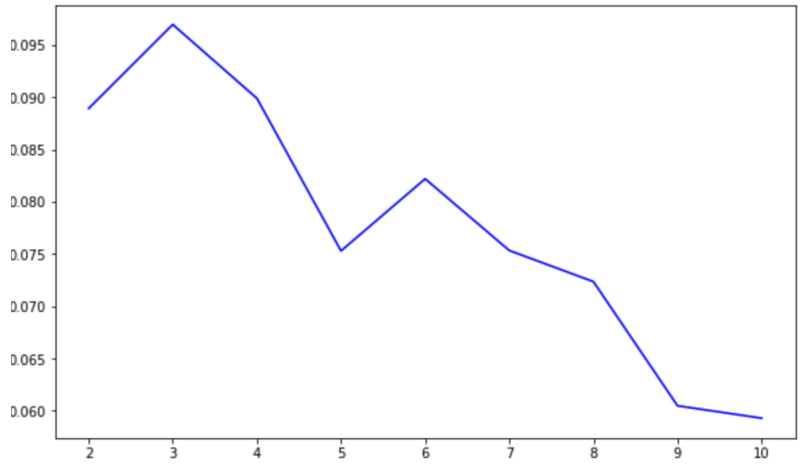

Figure 4 - Elbow Method on choosing K

The machine learning algorithm that will be used for modeling is K-means Clustering. This 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 or together and 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

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 average silhouette method. It is commonly done with a line plot of each point in the dataset, this method determines the degree of separation between the clusters, it measures how far clusters are from each other. A high average silhouette score signifies that clusters are not close to each other and are separated well[10]. In the line plot, we would choose the value of K with the highest average 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 three.

3.3 - Modeling

	Location	Afghan Restaurant	American Restaurant	Arcade	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	...	Tapas Restaurant	Tea Room	Tex-Mex Restaurant	Restau
0	Alicia	0.0	0.020833	0.0	0.0	0.041667	0.0	0.041667	0.000000	0.020833	...	0.000000	0.000000	0.0	
1	Amihan	0.0	0.000000	0.0	0.0	0.040816	0.0	0.000000	0.020408	0.081633	...	0.020408	0.000000	0.0	
2	Apolonio Samson	0.0	0.000000	0.0	0.0	0.055556	0.0	0.111111	0.000000	0.138889	...	0.000000	0.027778	0.0	
3	BF Homes	0.0	0.000000	0.0	0.0	0.051282	0.0	0.025641	0.000000	0.051282	...	0.000000	0.051282	0.0	
4	Baesa	0.0	0.000000	0.0	0.0	0.000000	0.0	0.095238	0.000000	0.000000	...	0.000000	0.047619	0.0	

Figure 5 - Partial Data Frame containing the Cluster Labels

With three being the number of clusters or K, K-means clustering is performed again with the new value of K. The cluster labels were retrieved. They were then 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.

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.

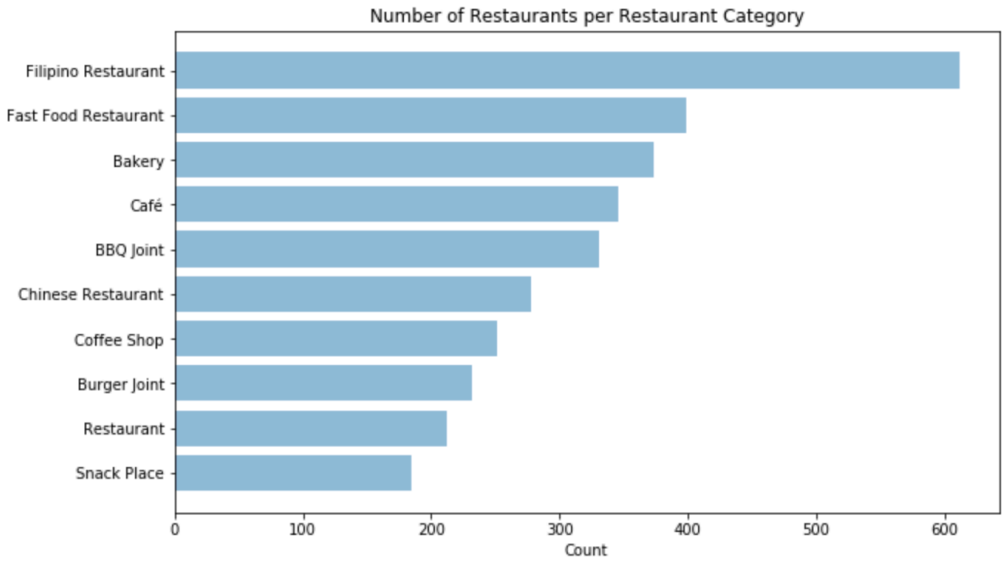

Figure 6 - Restaurant Category Frequency Bar Chart

Based on the chart in Figure 6, the restaurant category that is most common among all the locations are Filipino restaurants. This outcome has been expected as Filipinos tend to love dining in places where they are familiar with the cuisine.

The second most common are fast food restaurants. One reason may be because there is a large percentage of Filipinos still living below the national poverty line [12] and fast food restaurants in the Philippines offer a much cheaper dining experience relative to other types of restaurants. Also, they offer their food in a much faster rate as compared to other types of restaurants which makes them a good choice for fast dining. Because of this, some Filipinos would prefer them over the other options, especially the workers in the BPO industry which has an annual growth rate of 17% and produced 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 depicts the cluster in which the location belongs to.

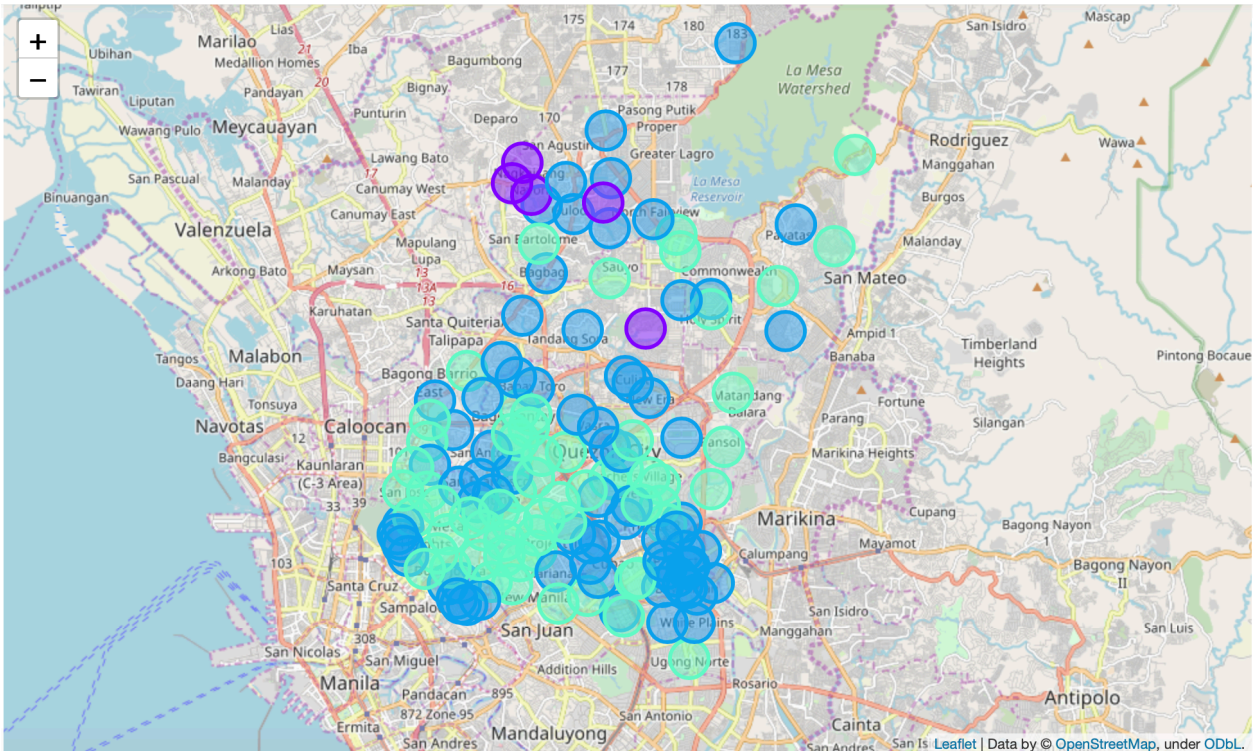

Figure 7 - Leaflet Map with Markers color coded by cluster membership

Based on the map in Figure 7, Cluster 1, which makes up only 5 of the locations, is colored purple. Cluster 2 being blue and Cluster 3 being sky-blue. 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.

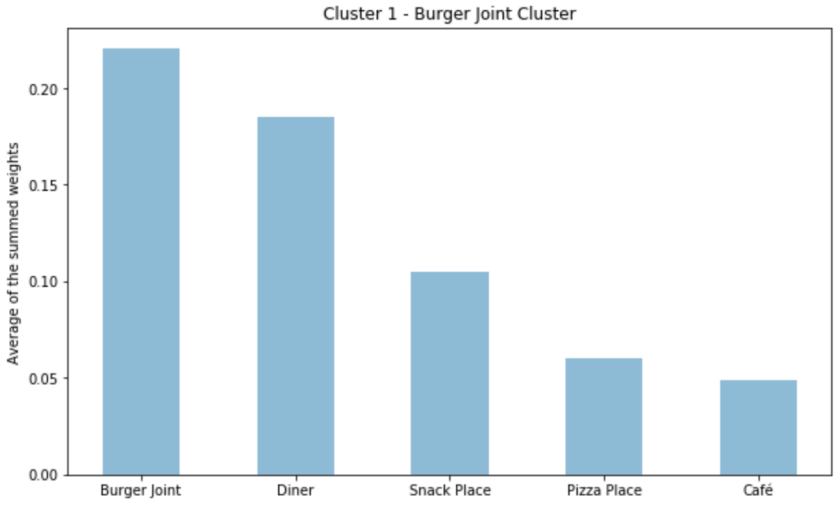

Figure 8 - Restaurant Category frequency in Cluster 1

In Cluster 1, it is apparent that Burger Joints are the most frequent restaurant categories. The other restaurant categories in the top five also show a bit of a similarity with one another as Snack Place 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.

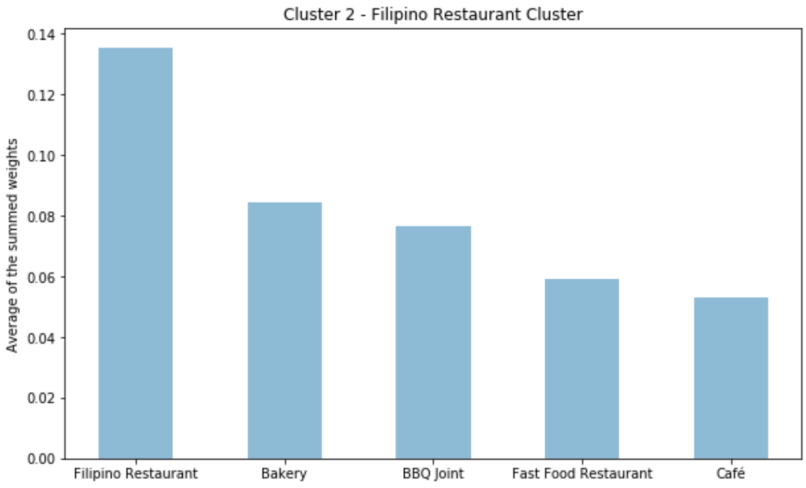

Figure 9 - Restaurant Category Frequency in Cluster 2

Cluster 2, also labeled as Filipino Restaurant Cluster, shows 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.

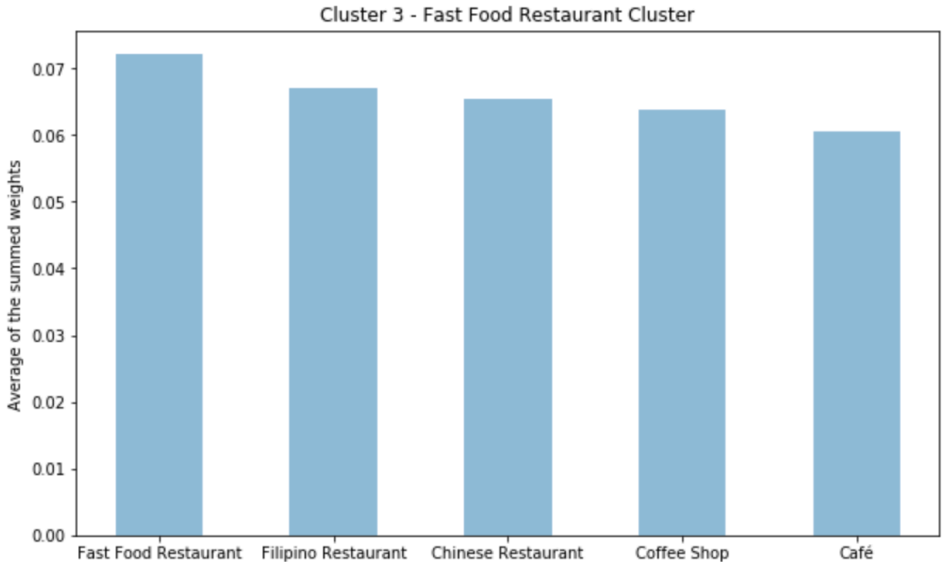

Figure 10 - Restaurant Category Frequency in Cluster 3

In Cluster 3, Fast Food Restaurant is the most common category found in the locations within the cluster. However, the top five restaurant categories within the cluster show minimal difference when compared to each other in terms of the average summed weights. The locations within the cluster may indicate that it is a busy area where there are many people of diverse ethnicities that prefer a wider variety of cuisines.

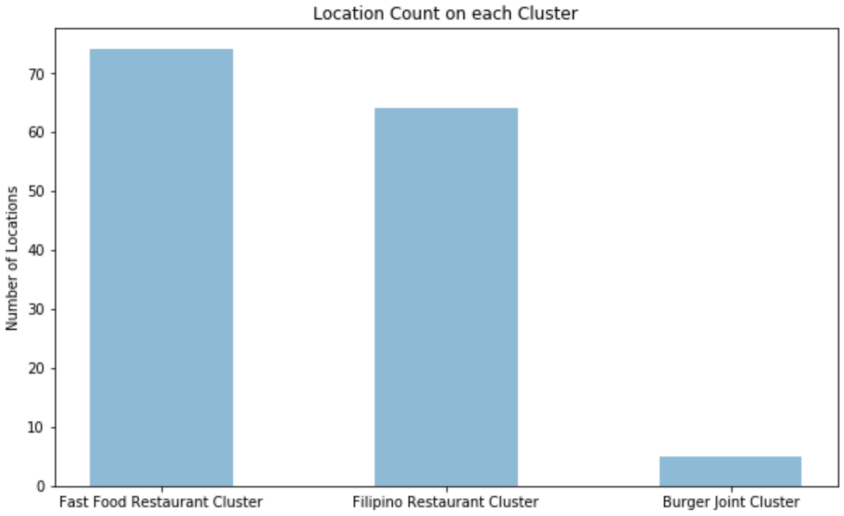

Figure 11 - Cluster Frequency within Quezon City

Based on the data shown in Figure 11, the cluster that has the highest amount of locations within it is the Fast Food Restaurant Cluster based on the results given by the model. There is a clear gap between the second highest cluster and the cluster next to it in terms of frequency because the Burger Joint Cluster only has five locations within it as shown in Figure 7. Majority of the locations within Quezon City prefer to have Filipino Restaurants. 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, Google's geocoding services, 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 have 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 have 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 Google's geocoding service or 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 and machine learning algorithms. Foursquare's API may not also be the best in terms of collecting restaurants in the Philippines, thus, 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

[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?fbclid=IwAR1VdklGWMKqF9nENgAIHqRIHimpZq2Yh7qmo-bjyEla14Ww-R-ajqldBs4
[13] https://www.flatworldsolutions.com/articles/growth-bpo-industry-philippines.php
[14] https://python-visualization.github.io/folium/
[15] https://developers.google.com/maps/documentation/geocoding/start