

Capstone Project – The Battle of Neighborhoods | Finding a Better Place Angeles City, Pampanga, Philippines

1. Introduction:

The purpose of this Capstone Project is to help people in exploring better facilities around their neighborhood.

Angeles City, Pampanga, where I stayed for 14 years, I was assigned in this neighborhood because of my work from previous company. The city is a 1st class highly urbanized city in the province of Pampanga, where it is geographically situated but remains politically independent. It is located in the region of Central Luzon, Philippines.

Due to new development in the region. Angeles City will be the main hub to Mega City in Central Luzon, with Clark International Airport under construction and is expected to be finished this year 2020.

Problem statement: Prospects of a Business establishment specially restaurant, close to office areas, and urban areas in Angeles City, Philippines.

During the daytime, especially in the lunch and dinner hours, office areas provide huge opportunities for restaurants. Reasonably priced Php150 – Php250 pesos shops are usually always full during the lunch hours (11 am — 2 pm) and, given this scenario, we will go through the benefits and pitfalls of opening a lunch cum dinner restaurant in highly dense urban places.

The core of Angeles City is made of 33 barangay but, we will only concentrate on top 10 most commercial/business district barangay.

Target Audience

1. Business personnel who wants to invest or open a restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of office workers in Angeles City and adjacent cities during lunch hours.
2. Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
3. New graduates, to find reasonable lunch/dinner place close to office.
4. Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it, and, finally be able to tell a story out of it.

2. Data Preparation:

- 2.1. Get the barangay names, Population and Land Area of each barangay from city population and government website in the Philippines.

First make use of barangay population table of Angeles City page from City Population. I have used Pandas, Lxml and BeautifulSoup4 library to create the initial data-frame. And use same library for land area of each barangay from Philippines government website. For a clean and understandable data-frame, I renamed some of the barangay for data-frame merging. Using 10 barangay has commercial/business district, this will be use to get the most venues within 2 kilometers radius of each area later on.

	Barangay	Area Sqm	Population2005	Population2010	Population2015
0	Agapito del Rosario	139,855	3,102	2,313	3,230
1	Amsic	1,521,162	3,474	7,736	14,379
2	Anunas	3,946,577	7,817	15,213	20,911
3	Balibago	2,648,768	27,914	32,291	40,087
4	Capaya	2,564,281	6,661	8,280	8,870
5	Claro M. Recto	179,228	5,691	4,741	3,981
6	Cuayan	4,139,180	1,645	4,852	10,363
7	Cutcut	4,082,166	22,462	21,601	27,843
8	Cutud	2,125,116	2,230	16,531	23,177
9	Lourdes Northwest	440,460	11,429	10,450	9,896
10	Lourdes Sur	210,199	5,943	4,772	4,797
11	Lourdes Sur East	206,663	5,880	4,656	4,741
12	Malabanias	2,625,601	19,169	23,034	33,174
13	Margot	2,312,225	2,768	3,606	5,239

	Barangay	Area Sqm
1	Agapito del Rosario	139,855
2	Amsic	1,521,162
3	Anunas	3,946,577
4	Balibago	2,648,768
5	Capaya	2,564,281
6	Claro M. Recto	179,228
7	Cuayan	4,139,180
8	Cutcut	4,082,166
9	Cutud	2,125,116
10	Lourdes Northwest	440,460
11	Lourdes Sur	210,199
12	Lourdes Sur East	206,663

2.2. Get the coordinates of the major 10 barangay.

Some of the coordinates returned by Geopy library are wrong and some are missing using Nominatim and I have to correct using their barangay hall address and correct also the naming of each barangay and get the correct coordinates using Google API.

	Barangay	Area Sqm	Population2005	Population2010	Population2015	Geo_data	Latitude	Longitude
0	Agapito del Rosario	139,855	3,102	2,313	3,230	Agapito del Rosario Barangay Hall, Angeles, Pa...	15.141864	120.590848
1	Amsic	1,521,162	3,474	7,736	14,379	Amsic Barangay Hall, Angeles, Pampanga, Philip...	15.155718	120.571175
2	Anunas	3,946,577	7,817	15,213	20,911	Anunas Barangay Hall, Angeles, Pampanga, Phil...	15.155556	120.553550
3	Balibago	2,648,768	27,914	32,291	40,087	Balibago Barangay Hall, Angeles, Pampanga, Phi...	15.166296	120.590899
4	Capaya	2,564,281	6,661	8,280	8,870	Capaya Barangay Hall, Angeles, Pampanga, Phil...	15.145857	120.616836
5	Claro M. Recto	179,228	5,691	4,741	3,981	Claro M. Recto Barangay Hall, Angeles, Pampang...	15.146734	120.591412
6	Cuayan	4,139,180	1,645	4,852	10,363	Cuayan Barangay Hall, Angeles, Pampanga, Phil...	15.146772	120.548620
7	Cutcut	4,082,166	22,462	21,601	27,843	Cutcut Barangay Hall, Angeles, Pampanga, Phil...	15.137750	120.581956
8	Cutud	2,125,116	2,230	16,531	23,177	Cutud Barangay Hall, Angeles, Pampanga, Philip...	15.172503	120.627258

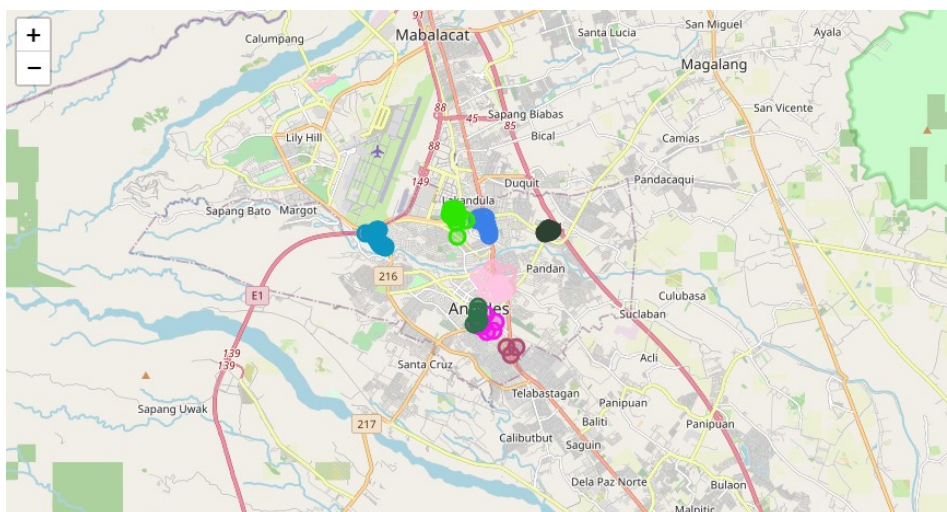
2.3. Foursquare Data

Finally I make use of Foursquare API to obtain the most common venues within 2 kilometers of each 10 barangay.

	Barangay	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Balibago	15.166296	120.590899	Velvet VIP Lounge	15.165926	120.589953	Lounge
1	Balibago	15.166296	120.590899	Fortune Hong Kong Seafood Restaurant	15.165954	120.590142	Chinese Restaurant
2	Balibago	15.166296	120.590899	Yisan Korean Restaurant	15.167593	120.589138	Buffet
3	Balibago	15.166296	120.590899	Pigs and Pints	15.163581	120.589602	Cocktail Bar
4	Balibago	15.166296	120.590899	Jollibee	15.167092	120.589937	Fast Food Restaurant
...
95	Pulung Maragul	15.160082	120.608663	Pizza Hut	15.162700	120.609000	Pizza Place
96	Pulung Maragul	15.160082	120.608663	Gerry's Grill	15.163822	120.609764	Filipino Restaurant
97	Pulung Maragul	15.160082	120.608663	Toll House	15.163223	120.609634	Snack Place

3. Exploring the Data and Major 10 barangay of Angeles City

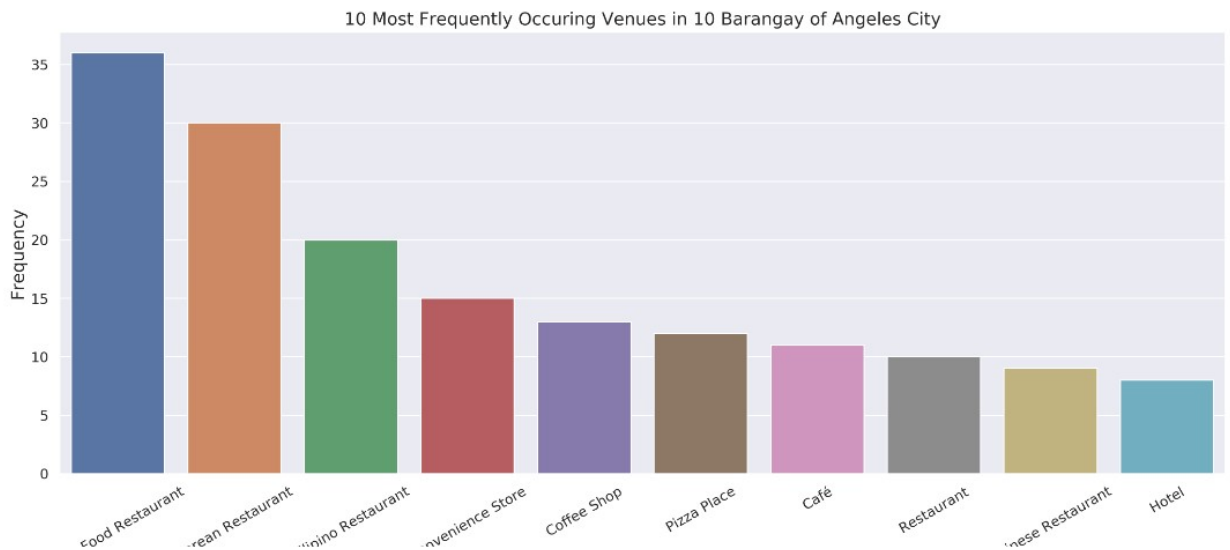
From the Foursquare data, we could see that there are 84 unique categories, but for this analysis, We will concentrate in Restaurant Category. As the focus is on 10 Major Barangay in Angeles City,(Anunas, Balibago, Claro M. Recto, Lourdes Northwest, Malabanias, Pulung Maragul, Salapungan, San Nicolas, Santo Domingo, and Santo Rosario). We found that there are 147 restaurants among the 341 top venues in these 10 Barangay. I have used Folium library to plot a leaflet map of only these restaurants in these 10 Major Barangay of Angeles City. See map below.



We have found out that

Fast food restaurants top the charts of most common venues in the 10 Barangay, followed by Korean restaurants and Filipino restaurants

A plot of the 10 most frequent venues in these 10 Barangay.



Then we obtain information about the top 5 venues of each 10 barangay. And proceed as follows

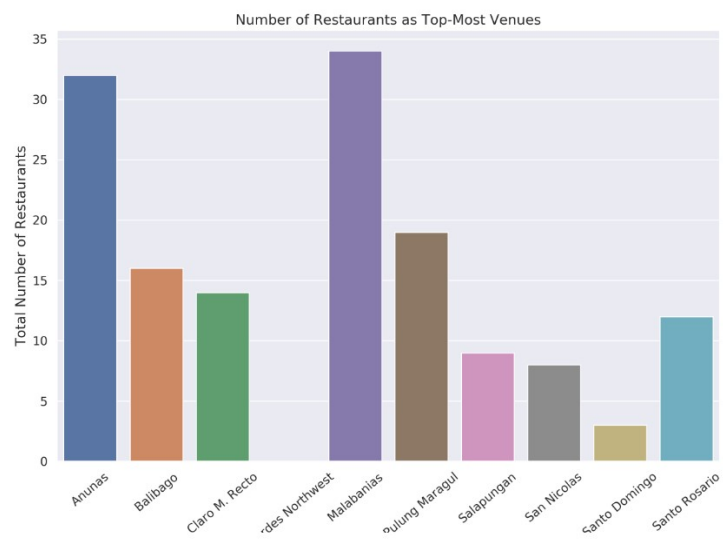
Create a data-frame with pandas one hot encoding for the venue categories.

Use pandas groupby on Barangay column and obtain the mean.

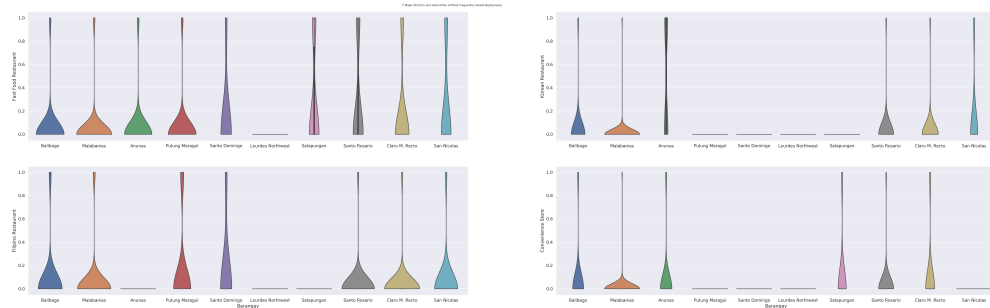
Transpose the data-frame at step 2 and arrange in descending order.

----Anunas----		
	venue	freq
0	Korean Restaurant	0.45
1	Convenience Store	0.08
2	Mexican Restaurant	0.06
3	Hotel	0.06
4	Fast Food Restaurant	0.06
----Balibago----		
	venue	freq
0	Chinese Restaurant	0.10
1	Filipino Restaurant	0.08
2	Convenience Store	0.08
3	Hotel	0.05
4	Asian Restaurant	0.05
----Claro M. Recto----		
	venue	freq
0	Fast Food Restaurant	0.17
1	Convenience Store	0.10
2	BBQ Joint	0.10
3	Restaurant	0.10
4	Coffee Shop	0.07
----Lourdes Northwest----		
	venue	freq
0	Hot Dog Joint	0.25
1	Café	0.25
2	Market	0.25
3	Flea Market	0.25

And explore which barangay has the highest number of restaurants as the most common venue and the plot below.



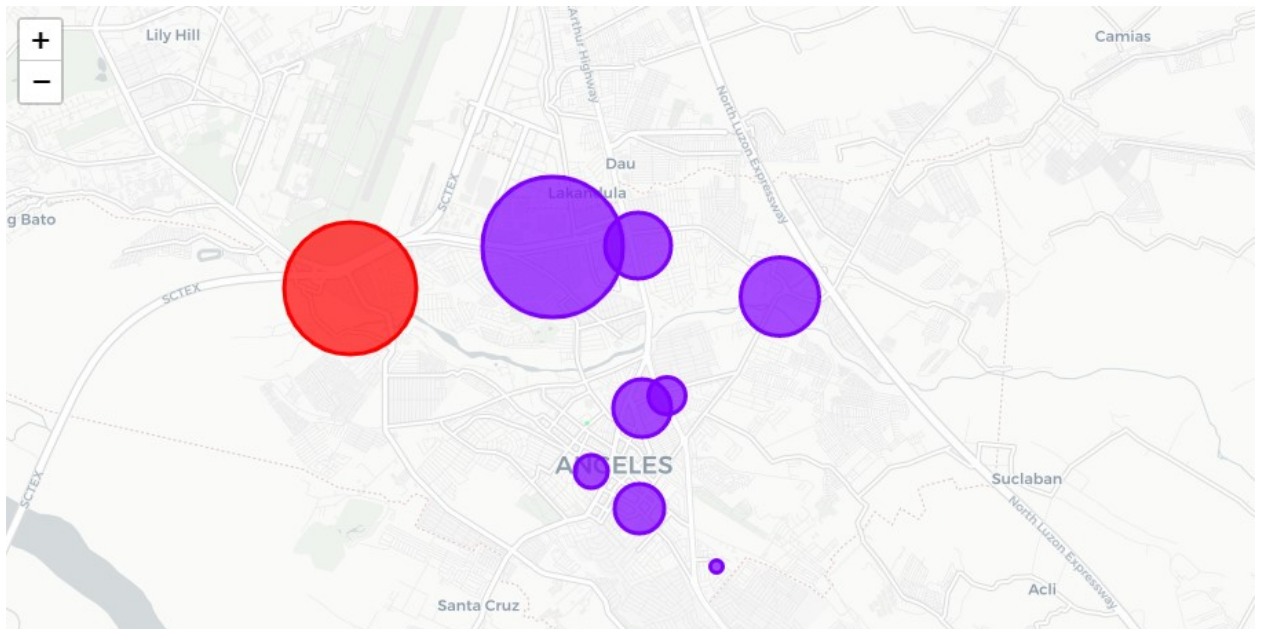
Since the focus is mostly on the competitions provided by restaurants on these barangay, I also explored how most common restaurant categories are distributed in each barangay and the violin plot below of categorical variables give the answer.



From here we see that lots of Fast food restaurant in most of the barangay except Lourdes Northwest. And Barangay Anunas don't have any Filipino restaurant.

3.1 Exploring the Data and Major 10 barangay of Angeles City

We try to cluster these 10 barangay based on the frequency of venue categories and use K-Means clustering. Our expectation would be based on the similarities of venue categories, these barangay will be clustered. Using K-Means algorithm from Scikit-learn library we obtain 3 clusters as shown below.



Here the radius of the circles represent the number of restaurants as most common venue for the corresponding barangay and, we have seen that it is maximum for Barangay Malabantias (34) followed by Anunas (32) and no restaurant in Barangay Lourdes Northwest.

For the most common venues this clustering makes a complete sense as most of the barangay are dominated by Fast Food restaurant, Filipino restaurant, Chinese restaurant falls under the purple cluster whereas barangay Anunas dominated by Korean restaurant and Convenience store falls under red cluster. And Lourdes Northwest with light blue cluster dominated with Market and Café.

4. Results

The results of the exploratory data analysis and clustering are summarized

- Fast food restaurants top the charts of most common venues in 10 Major barangay.
- Barangay Anunas are dominated by Korean restaurants and convenience store. Don't have any Filipino restaurant nearby.
- Lourdes Northwest don't have any restaurant but have market and café.
- Barangay Malabantias and Pulung Maragul dominated with restaurants because near Commercial Malls like SM Malls and Ayala Malls.
- The clustering was based only on the most common venues on each barangay. Most of the group with Filipino and Fast food restaurant are clustered, second is Korean restaurant and convenience store are clustered and the last one with no restaurant but café and market.

5. Discussion

According to the analysis, Barangay Lourdes Northwest will provide least competition for an upcoming lunch restaurant as no restaurant nearby, only market and café.

From web-scraped data, Barangay Santo Domingo, could potentially be a target for starting quality restaurants as its near the border of Angeles City and common venue is café also.

Some drawbacks of this analysis are – we don't have any data for the average land price and the clustering is completely based on the most common venues obtained from Foursquare data. Since the distance of the venues from any public utility vehicles, number of potential customers, benefits and drawbacks of Barangay Lourdes Northwest, could all play a major role and thus, this analysis is definitely far from being conclusory. However, it gives us

preliminary information on possibilities of opening restaurants around the 10 Major barangay in Angeles City.

If we have other data that will help with our analysis and adding other clustering techniques the result could vary.

6. Conclusion

We have got a small glimpse of how to real life data –science projects look like. Using some frequent python libraries we get the data using scrap web-data, use Foursquare API to explore the 10 major barangay in Angeles City, and use Google API for the geolocation of each barangay. We saw the results of segmentation of barangay using folium leaflet map. Some of drawbacks and chance for improvements to represent even more analysis in real life business problem. Since the analysis were mostly concentrated on the possibilities of opening a restaurants. Some of the results obtained are surprisingly what I have expected after 14 years of staying in Angeles City.