MAKATI NEIGHBORHOODS: WHERE TO LOCATE BEFORE AND AFTER THE SUBWAY IS BUILT

Applied Data Science Capstone

By Juanito Mercado

Outline

- Introduction
- Data
- Methodology
- Results
- Discussion
- Conclusion

Introduction

- Metro Manila has one of the worst traffic environments in the world
- Makati (part of Metro Manila) is building the country’s first subway
- Where in Makati should a restaurant be located now?
- Where to locate once the subway is operational?

Data

- Makati Zip Codes to get Makati Neighborhoods' Geo Coordinates

  Source - https://en.wikipedia.org/wiki/List_of_ZIP_codes_in_the_Philippines

	City	ZipCode	Neighborhood	Latitude	Longitude
0	Makati City	1200	Makati CPO (Inc, Buendia Up To	14.561608	121.014653
1	Makati City	1203	San Antonio Village (Inc. Malu)	14.563281	121.012607
2	Makati City	1204	La Paz-Singkamas-Tejeros	14.568549	121.008594
3	Makati City	1205	Sta. Cruz	14.567455	121.015539
4	Makati City	1206	Kasilawan	14.576348	121.014462

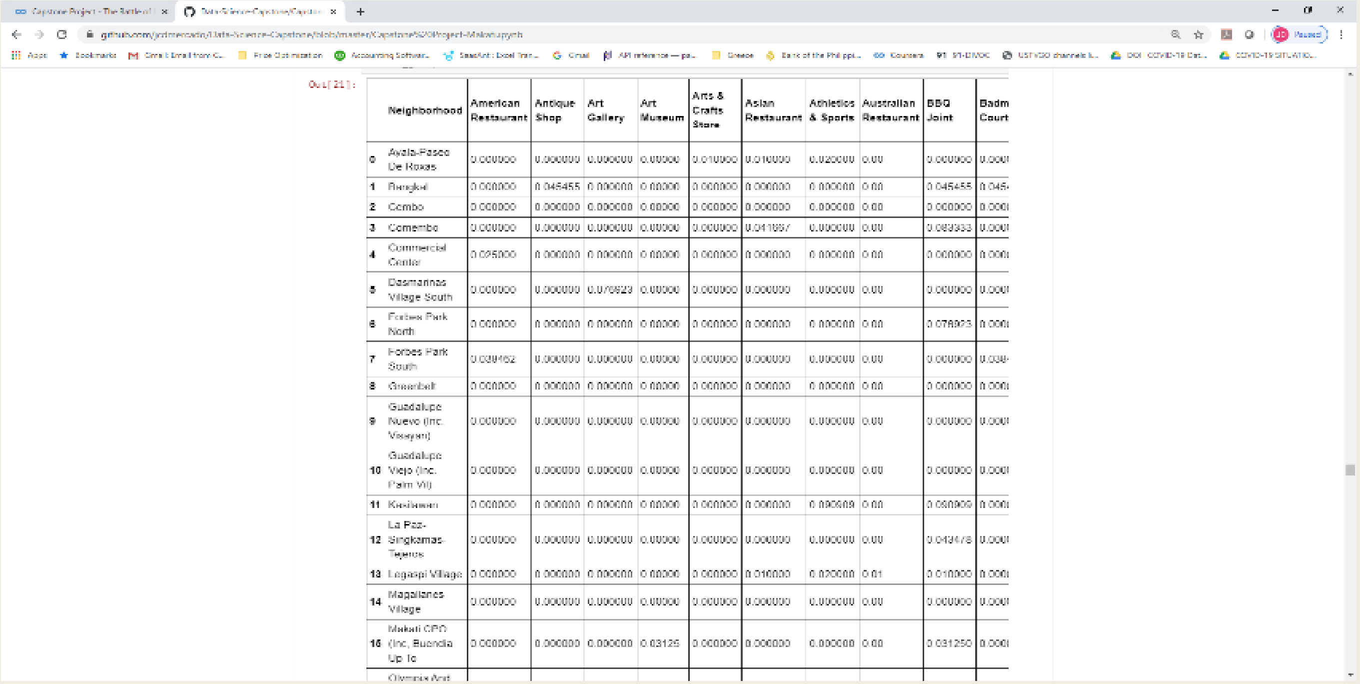
Data

- Nearby venues can be gathered from Foursquare.com

	name	categories	lat	lng
0	Ayala Triangle Gardens	Park	14.556471	121.023204
1	Banapple Pies & Cheesecakes	Restaurant	14.556634	121.023619
2	The Peninsula Manila	Hotel	14.555066	121.025466
3	Escolta	Filipino Restaurant	14.555485	121.025509
4	Little Flour Café	Café	14.557978	121.021919

Data

- Foursquare venue data grouped by Neighborhood
- Frequencies of each venue type
- This serves as the input for K-Means Clustering.


Out[21]:

	Neighborhood	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Badm Court
0	Ayala-Paseo De Roxas	0.000000	0.000000	0.000000	0.00000	0.010000	0.010000	0.020000	0.00	0.000000	0.000
1	Bangkal	0.000000	0.045455	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.045455	0.045
2	Cembo	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
3	Comembo	0.000000	0.000000	0.000000	0.00000	0.000000	0.041667	0.000000	0.00	0.083333	0.000
4	Commercial Center	0.025000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
5	Dasmarinas Village South	0.000000	0.000000	0.076923	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
6	Forbes Park North	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.076923	0.000
7	Forbes Park South	0.038462	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.038
8	Greenbelt	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
9	Guadalupe Nuevo (Inc. Visayan)	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
10	Guadalupe Viejo (Inc. Palm Vil)	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
11	Kasilawan	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.090909	0.00	0.090909	0.000
12	La Paz-Singkamas-Tejeros	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.043478	0.000
13	Legaspi Village	0.000000	0.000000	0.000000	0.00000	0.000000	0.010000	0.020000	0.01	0.010000	0.000
14	Magallanes Village	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00	0.000000	0.000
15	Makati CPO (Inc, Buendia Up To	0.000000	0.000000	0.000000	0.03125	0.000000	0.000000	0.000000	0.00	0.031250	0.000
	Olympia And										

Data

- Identify proposed Makati Subway Stations then get Geo Coordinates
  - Source - https://businessmirror.com.ph/2019/10/30/makati-subway-project-gets-additional-332-million-initial-funding
  - Geo Coordinates obtained from Google Maps
- Compute the shortest distances (in meters) of each neighborhood to any of the stations

	Station	Latitude	Longitude
0	EDSA-Ayala	14.55093	121.02883
1	Ayala Triangle	14.55671	121.02281
2	Makati Central Park	14.56215	121.01494
3	Police Headquarters	14.56356	121.01524
4	Circuit City	14.57305	121.01946
5	Makati City Hall	14.57081	121.02728
6	Rockwell	14.56324	121.03580
7	Guadalupe	14.56741	121.04542
8	University of Makati	14.56397	121.05575
9	Ospital ng Makati	14.54681	121.06176

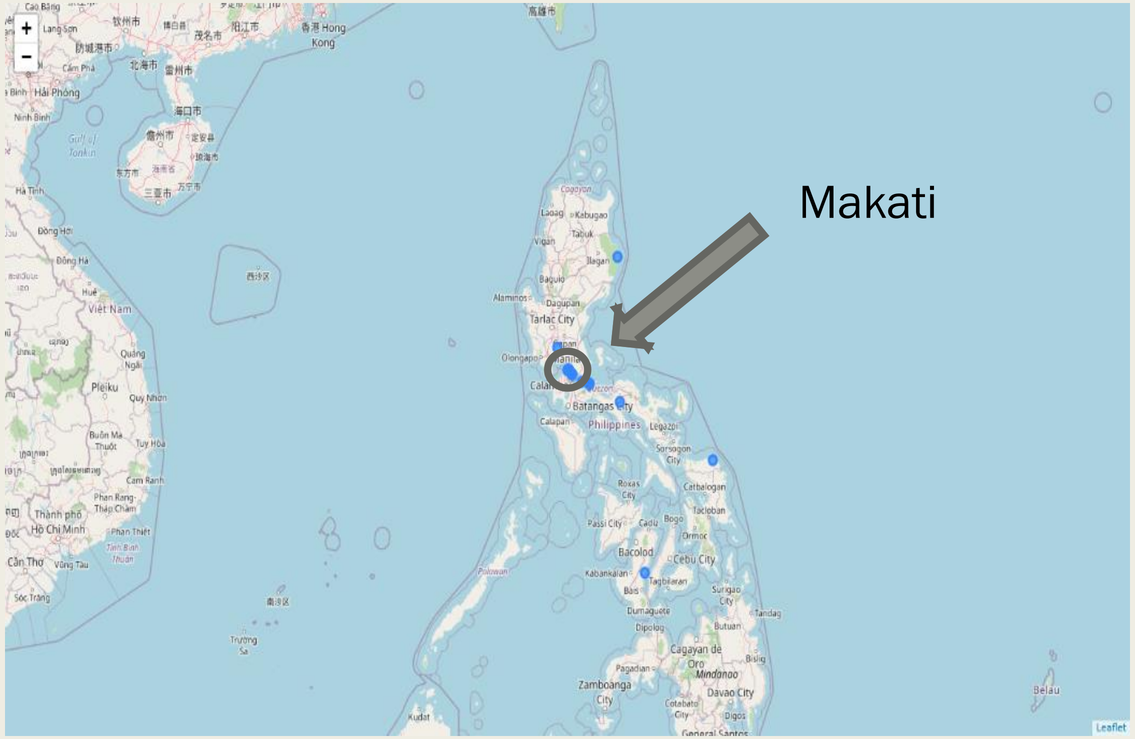
Neighborhood	Distance
Makati CPO (Inc, Buendia Up To	67.475320
San Antonio Village (Inc. Malu)	280.839332
La Paz-Singkamas-Tejeros	904.242772
Sta. Cruz	432.163349
Kasilawan	650.561473

Methodology

- Find nearby venues to Makati neighborhoods using Foursquare
- Generate a table of frequencies of top venue types of each Makati neighborhood
- Use K-Means Clustering to group similar neighborhoods and gain insight on types of venues that thrive in those clusters
  - *Display square distances for k from 1 to 9 and use Elbow Method to find optimal k*
- Add Subway station proximity as a factor and repeat K-Means Clustering and compare clusters to original clusters to gain insights
  - *Display square distances for k from 1 to 9 and use Elbow Method to find optimal k*

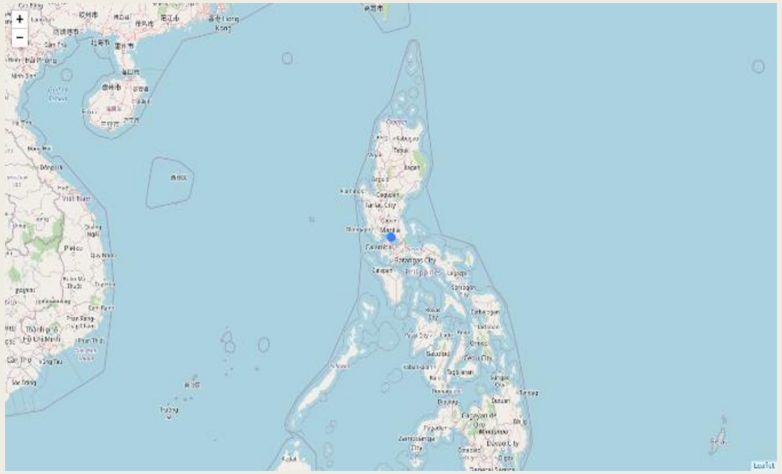
Results

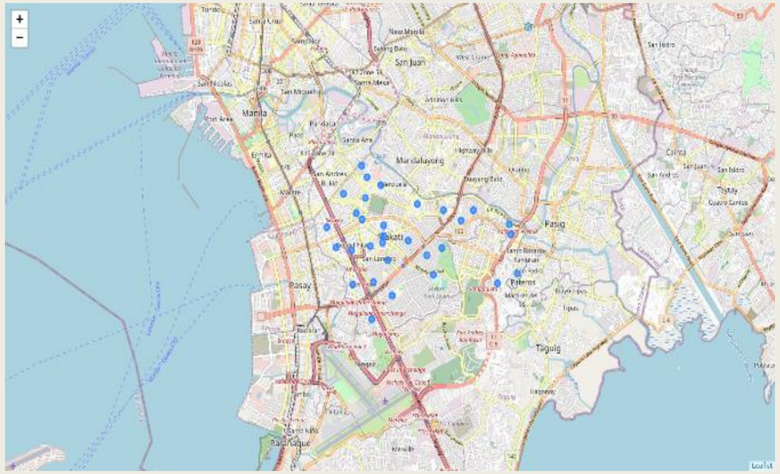
- Makati Zip Code Geo Coordinates from Pgeocode were not useable
- Many coordinates given were outside Makati and not even in Metro Manila
- For those that were in Makati, the coordinates were not accurate.
- Many neighborhoods also shared zip codes even if far apart



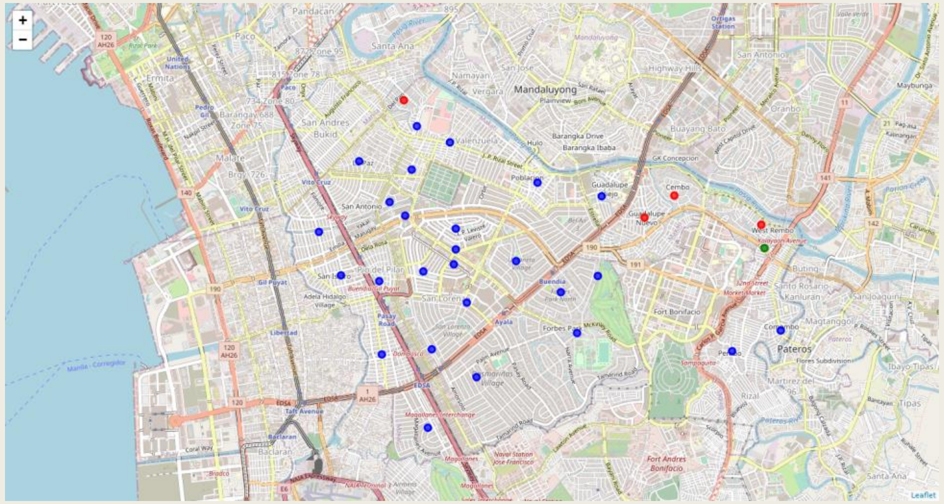
Results

- Coordinates were corrected using Google Maps
- Redundant Zip Codes pointing to the same location were removed
- This new list of neighborhoods with coordinates were used to get nearby venues from Foursquare.
- The top venue types were counted for each neighborhood.
- K-Means Clustering was performed.





Results
3 Clusters

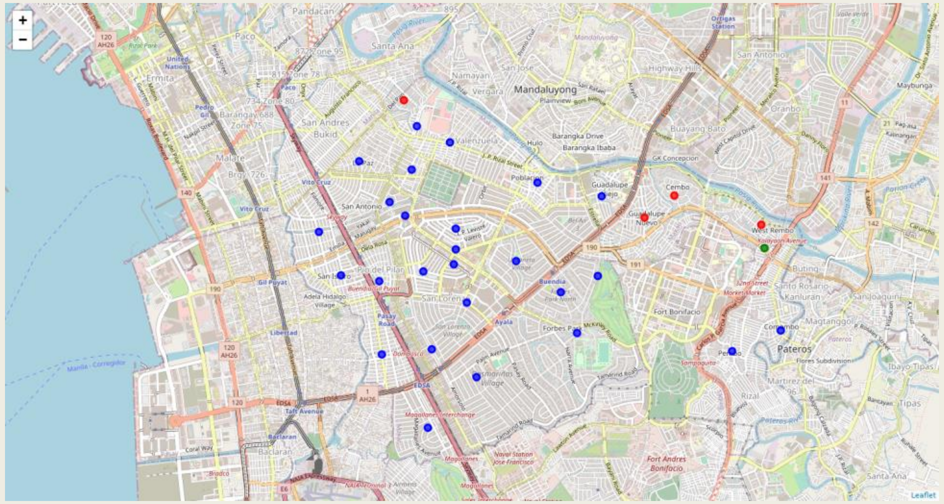


- **Cluster 1**: Ayala-Paseo De Roxas, Bangkal, Comembo, Commercial Center, Dasmarinas Village South, Forbes Park North, Forbes Park South, Greenbelt, Guadalupe Viejo, La Paz-Singkamas-Tejeros, Legaspi Village, Magallanes Village, Makati CPO, Olympia And Carmona, Palanan, Pasong Tamo 2000 Up & Ecology V, Pembo, Pio Del Pilar, Poblacion, Rembo (East) & Malapad Na Bato, Salcedo Village, San Antonio Village, San Isidro, San Lorenzo Village, Sta. Cruz, and Urdaneta Village

- **Cluster 2**: Cembo, Guadalupe Nuevo, Kasilawan and Rembo (West)

- **Cluster 3**: Pinagkaisahan-pitogo

Results
3 Clusters



- Without using economic factors in the clustering, the clusters still grouped the most developed (**Cluster 1**), least developed (**Cluster 3**), and in between (**Cluster 2**)
- **Cluster 1** includes all commercial, office areas and gated subdivisions. Restaurants, cafes, and spas are the most common venues here.
- **Cluster 2** has convenience stores and fast food restaurants as the most common venues.
- **Cluster 3** has the least commercial activity where the most common venue types are: Food (neighborhood food stall), Soccer Field, Intersection, and Basketball Court.

Results with Subway Proximity
5 Clusters

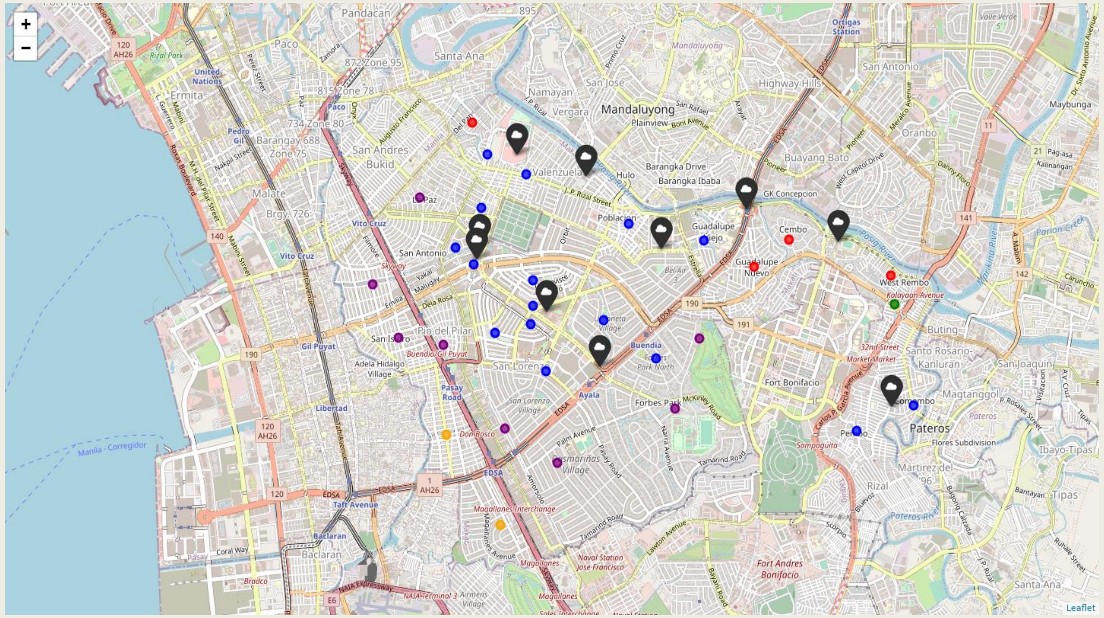
- **Cluster 1**: Ayala-Paseo De Roxas, Comembo, Commercial Center, Forbes Park North, Greenbelt, Guadalupe Viejo, Legaspi Village, Makati CPO, Olympia And Carmona, Pembo, Poblacion, Salcedo Village, San Antonio Village, San Lorenzo Village, Sta. Cruz, and Urdaneta Village

- **Cluster 2**: Cembo, Guadalupe Nuevo, Kasilawan and Rembo (West)

- **Cluster 3**: Pinagkaisahan-pitogo

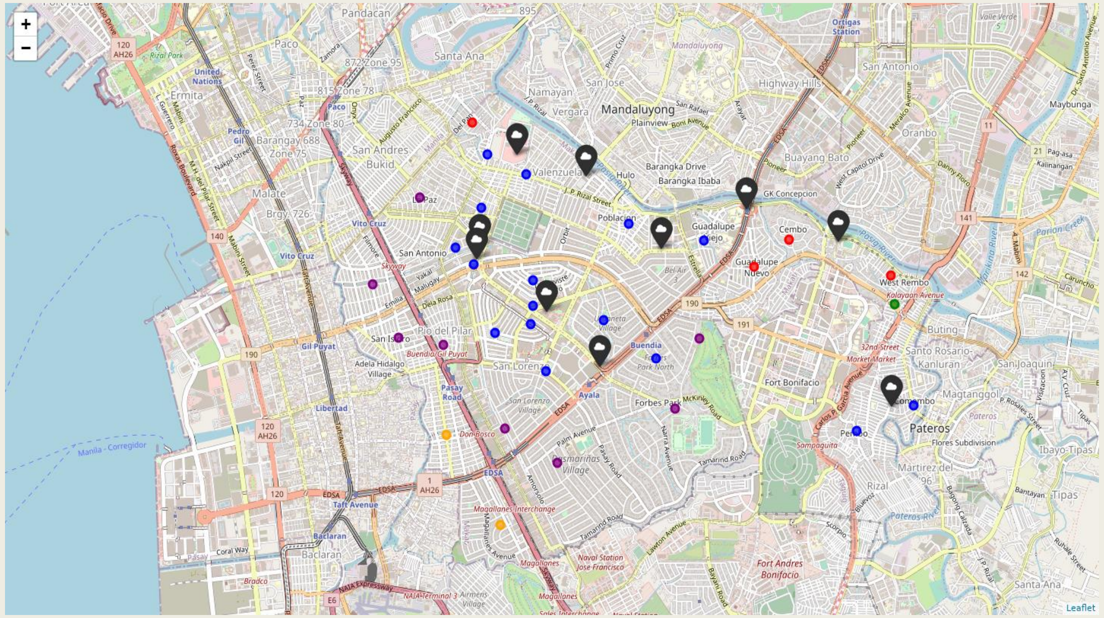
- **Cluster 4**: Dasmarinas Village South, La Paz-Singkamas-Tejeros, Palanan, Pasong Tamo 2000 Up & Ecology V, Pio Del Pilar, Rembo (East) & Malapad Na Bato, Forbes Park South, and San Isidro

- **Cluster 5**: Magallanes Village and Bangkal



Results with Subway Proximity
5 Clusters

- **Cluster 2** and **Cluster 3** of original clustering are maintained as separate clusters.
- Cluster 1 now broken down into 3 separate clusters (1, 4 and 5).
  - *Most retained as a cluster (**Cluster 1**) - neighborhoods that were in original Cluster 1 that are also closest to the stations.*
  - ***Cluster 4** are neighborhoods from original Cluster 1 that are not that close to the stations.*
  - ***Cluster 5** are neighborhoods from original Cluster 1 that are farthest from the stations.*



Discussion

- Use original clustering in this way:
  - *To find a location for a restaurant, browse through Cluster 1*
  - *Look for neighborhoods where most common venues are not restaurants.*
  - *Locations for cafes and spas can be found in the same way.*
  - *To find a fastfood or convenience store location, browse through Cluster 2 as well*
- New clustering relevant if restaurant to be built appeals to subway-riding customers
- For such restaurants, location options in new Cluster 1 is ideal
- Clusters 4 and 5 only considered if neighborhoods can generate enough demand or restaurant’s market brings cars or uses Uber/Grab

Discussion

- Improvements:
  - *identify distance thresholds deemed the same to restaurant patrons*
    - Define closeness to a station
    - Is 200 meters treated the same as 20 meters?
    - Is 2 kilometers treated the same as 20 kilometers?
    - Using the new clustering as an example, Clusters 4 and 5 (orange and purple) might be grouped together while Cluster 1 (blue) might need to be subdivided further.
  - *look into machine learning to predict profitability based on type of venue, neighborhood and demographic data of the market.*
    - Actual distance of the venue to the subway stations can be included but since there are no subways yet in the Philippines, it will be difficult to find its relationship to profitability.

Conclusion

- Using alternative paid sources may increase the possible breadth and depth of this analysis.
- Foursquare and Google Maps offer more detailed location data and reviews that can help complete the picture.
- With better data for all major cities in the Philippines, we can study other neighborhoods and venues and how the building of transportation infrastructure changes the way they are related and valued.
- With data science, better models lead to more informed decision-making.