

# City Exploration Recommendation

Muhammad Faris Herlansyah

# The Importance of Recommending First-time Visitors

- ▶ First-time visitors may have no idea where to designate parts of the city they would like to explore mainly. Here we take an example of the city of Yogyakarta, located in Indonesia.
- ▶ Yogyakarta is a very popular tourist destination due to the various kinds of attractions they offer. For example, there are cultural attractions, e.g., the Keraton of Yogyakarta and Yogyakarta's two city squares, as well as low-cost culinary spots spread in the city.
- ▶ Because of the variety in kinds of attractions, there is needed a recommender system to recommend first-time visitors with no idea where to explore.



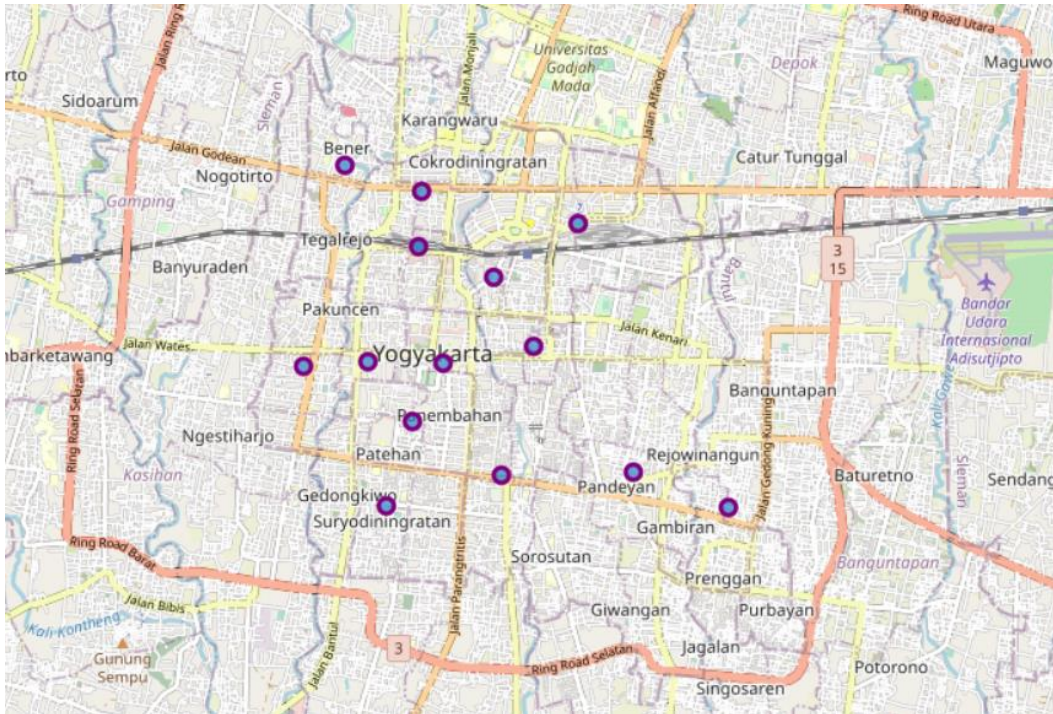
# Data Acquisition and Cleaning

- ▶ Districts data scraped from [kodepos.nomor.net](http://kodepos.nomor.net)
- ▶ Geographical coordinates of each district obtained using Nominatim
- ▶ Venues data obtained using Foursquare API
- ▶ There is obtained fourteen districts of Yogyakarta with its respective geographical coordinates
- ▶ Venues data contained 373 rows, meaning that there are 373 recorded venues in Yogyakarta, and 97 columns, meaning that there are 97 types of venues

# Geographical Coordinates of Yogyakarta's Fourteen Districts

1	Kode POS	Kecamatan	Kota	Lat_Kecamatan	Long_Kecamatan
0	55211	Danurejan	Yogyakarta	-7.792842	110.371795
1	55271	Gedongtengen	Yogyakarta	-7.789338	110.363376
2	55221	Gondokusuman	Yogyakarta	-7.786791	110.381157
3	55121	Gondomanan	Yogyakarta	-7.802395	110.366112
4	55231	Jetis	Yogyakarta	-7.783297	110.363649
5	55171	Kotagede	Yogyakarta	-7.818311	110.397941
6	55131	Kraton	Yogyakarta	-7.808799	110.362726
7	55141	Mantrijeron	Yogyakarta	-7.818067	110.359731
8	55151	Mergangsan	Yogyakarta	-7.814734	110.372558
9	55261	Ngampilan	Yogyakarta	-7.802183	110.357603
10	55111	Pakualaman	Yogyakarta	-7.800395	110.376249
11	55241	Tegalrejo	Yogyakarta	-7.780455	110.355073
12	55161	Umbulharjo	Yogyakarta	-7.814378	110.387374
13	55251	Wirobrajan	Yogyakarta	-7.802624	110.350447

# Visualization of Yogyakarta's Fourteen Districts



- ▶ We can see some districts have their own characteristics, e.g.:
- ▶ Districts having railway stations
- ▶ Districts near cultural centers
- ▶ Districts primarily dominated by residential areas

# Venues Data Obtained using Foursquare API

- We can see culinary spots and hostel there, which is common in a touristy city.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Danurejan	-7.792842	110.371795	Laura's Backpackers	-7.790985	110.370988	Hostel
1	Danurejan	-7.792842	110.371795	Jagung Bakar Menthel	-7.792026	110.373640	Snack Place
2	Danurejan	-7.792842	110.371795	BAKPIAKU Gajah Mada (pusat)	-7.797168	110.372700	Bakery
3	Danurejan	-7.792842	110.371795	Bakpia Citra Premium	-7.793618	110.368021	Snack Place
4	Danurejan	-7.792842	110.371795	Rumah Makan Murni 83	-7.791199	110.372408	Asian Restaurant
5	Danurejan	-7.792842	110.371795	Sate Ayam Podomoro	-7.795822	110.369001	Asian Restaurant
6	Danurejan	-7.792842	110.371795	Indomaret Jl.Hayam Wuruk	-7.795520	110.372715	Department Store
7	Danurejan	-7.792842	110.371795	Spesial Sambal Mentah "Bu Saring"	-7.792306	110.373451	Asian Restaurant
8	Danurejan	-7.792842	110.371795	warung makan pak parno sate kambing dan tongseng	-7.792491	110.373380	Diner
9	Danurejan	-7.792842	110.371795	Kedai Kopi Mataram	-7.792888	110.367511	Coffee Shop

# Dummy Encoding

- ▶ It helps us see the occurrences of every listed venue in each district.

	District	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bakery	Bar	Basketball Stadium	Batik Shop	Beach	Bed & Breakfast	Beer Garden
0	Danurejan	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Danurejan	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Danurejan	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	Danurejan	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Danurejan	0	0	0	0	0	1	0	0	0	0	0	0	0	0

# Frequency of Occurrence of Each Venue

	District	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bakery	Bar	Basketball Stadium	Batik Shop	Beach	Bed & Breakfast
0	Danurejan	0.000000	0.000000	0.000000	0.000000	0.062500	0.250000	0.000000	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000
1	Gedongtengen	0.000000	0.000000	0.000000	0.000000	0.000000	0.025641	0.000000	0.000000	0.051282	0.000000	0.000000	0.000000	0.025641
2	Gondokusuman	0.030303	0.030303	0.000000	0.000000	0.000000	0.060606	0.090909	0.030303	0.000000	0.000000	0.000000	0.000000	0.000000
3	Gondomanan	0.000000	0.018868	0.056604	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.018868	0.018868	0.018868
4	Jetis	0.000000	0.000000	0.000000	0.000000	0.000000	0.106061	0.000000	0.045455	0.000000	0.015152	0.000000	0.000000	0.000000
5	Kotagede	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Kraton	0.000000	0.000000	0.000000	0.000000	0.033333	0.066667	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000
7	Mantrijeron	0.000000	0.000000	0.160000	0.000000	0.000000	0.040000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Mergangsan	0.000000	0.000000	0.000000	0.000000	0.000000	0.083333	0.041667	0.083333	0.000000	0.000000	0.000000	0.000000	0.000000
9	Ngampilan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.111111
10	Pakualaman	0.000000	0.000000	0.000000	0.000000	0.000000	0.047619	0.047619	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000
11	Tegalrejo	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.083333	0.000000	0.000000	0.000000	0.000000	0.000000
12	Umbulharjo	0.000000	0.000000	0.000000	0.133333	0.066667	0.133333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Wirobrajan	0.000000	0.000000	0.000000	0.058824	0.058824	0.058824	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.000000

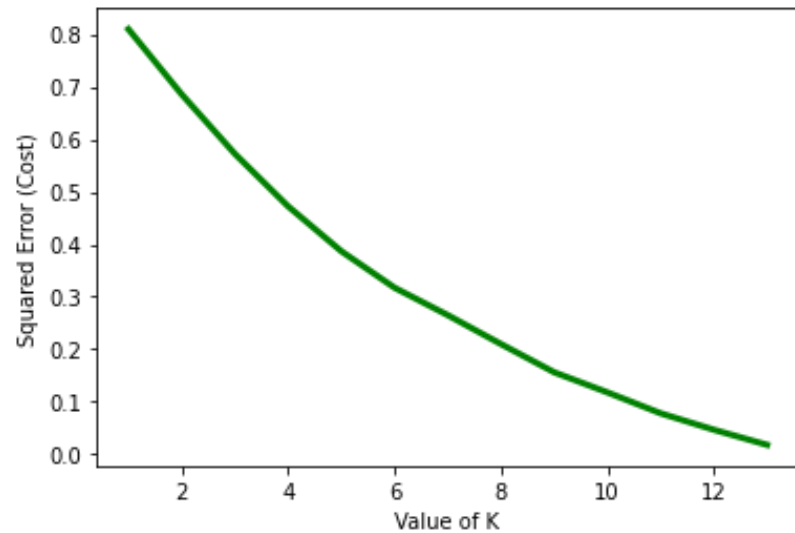


# Most Common Venues

- ▶ As we may have predicted, culinary spots dominate in some districts

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Danurejan	Asian Restaurant	Snack Place	Bakery	Department Store	Diner
1	Gedongtengen	Hotel	Indonesian Restaurant	Noodle House	Coffee Shop	Café
2	Gondokusuman	Food Truck	Indonesian Restaurant	BBQ Joint	Diner	Breakfast Spot
3	Gondomanan	Indonesian Restaurant	Historic Site	Clothing Store	Art Gallery	Chinese Restaurant
4	Jetis	Hotel	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Food Truck

# Clustering Method



- ▶ Using K-Means to cluster districts.
- ▶ Apply elbow method to find the appropriate number of clusters.
- ▶ The appropriate value of  $k$  is located at the “elbow” of the curve. In the picture on the left, the best value of  $k$  is probably 5.

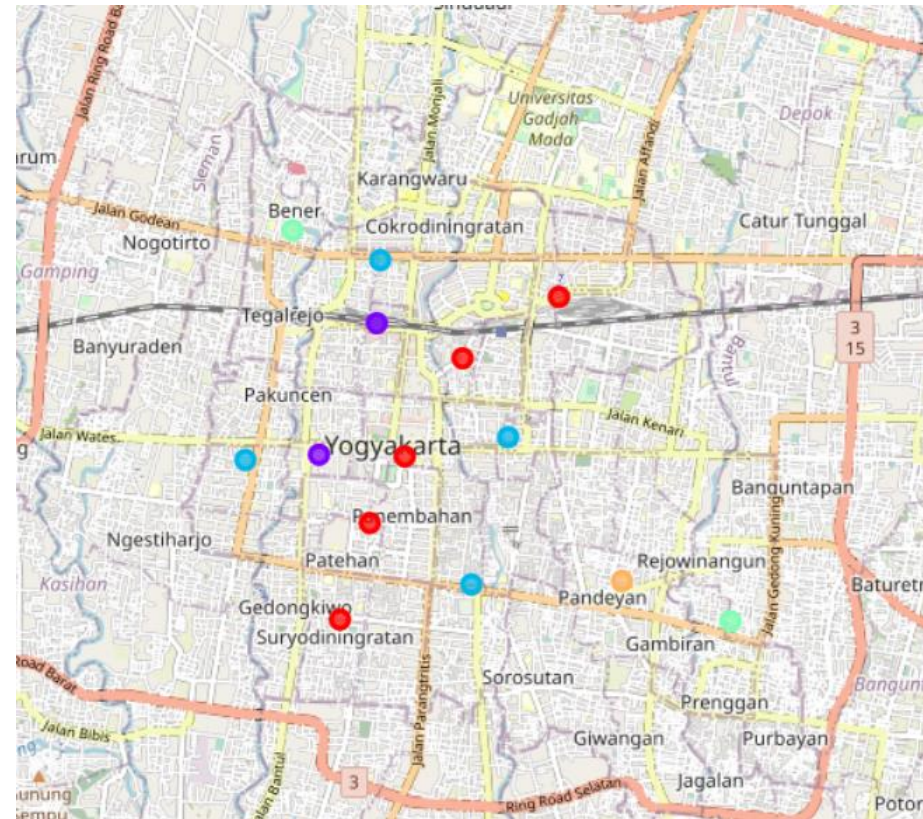
# Clustering Method

- ▶ Apply  $k = 5$ , build the model, then see the resulted clusters and label the districts with corresponding cluster labels

	Kode POS	Kecamatan	Kota	Lat_Kecamatan	Long_Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	55211	Danurejan	Yogyakarta	-7.792842	110.371795	0	Asian Restaurant	Snack Place	Bakery	Department Store	Diner
1	55271	Gedongtengen	Yogyakarta	-7.789338	110.363376	1	Hotel	Indonesian Restaurant	Noodle House	Coffee Shop	Café
2	55221	Gondokusuman	Yogyakarta	-7.786791	110.381157	0	Food Truck	Indonesian Restaurant	BBQ Joint	Diner	Breakfast Spot
3	55121	Gondomanan	Yogyakarta	-7.802395	110.366112	0	Indonesian Restaurant	Historic Site	Clothing Store	Art Gallery	Chinese Restaurant
4	55231	Jetis	Yogyakarta	-7.783297	110.363649	2	Hotel	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Food Truck

# Clustering Result

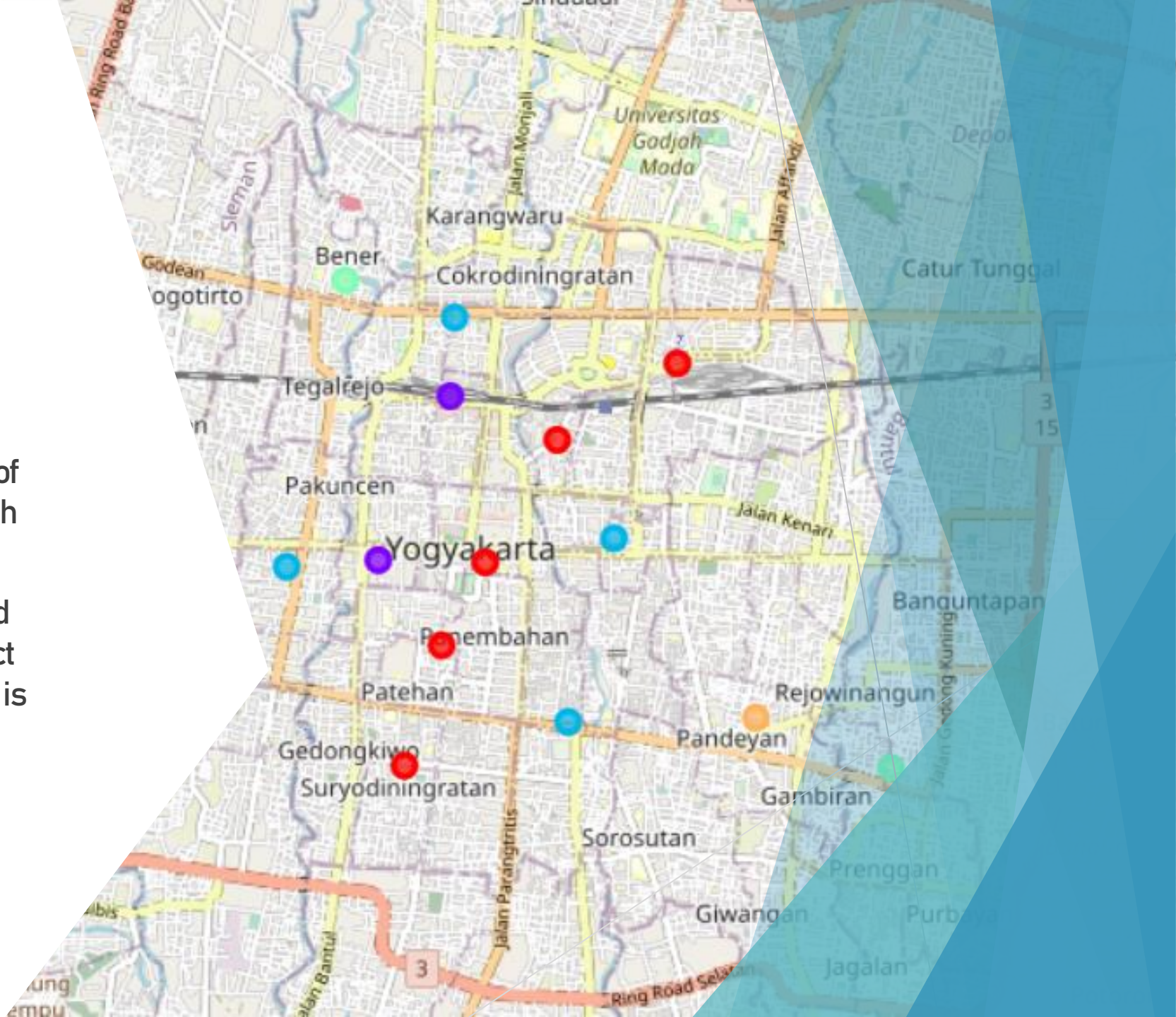
- ▶ Visualization of clusters on a Yogyakarta map is on the right.
- ▶ Cluster colors:
  - ▶ Red : cluster 0
  - ▶ Violet : cluster 1
  - ▶ Blue : cluster 2
  - ▶ Green : cluster 3
  - ▶ Orange : cluster 4





# Cluster Analysis and Result

- ▶ Cluster 0 is dominated by cultural and historical venues. This is supported by the fact that, for example, the seat of the Sultanate of Yogyakarta, and the North and South Squares, are in Cluster 0.
- ▶ Cluster 1 is dominated by hotels and hostels. This is supported by the fact that the Yogyakarta railway station is in Cluster 1.





# Cluster Analysis and Result

- ▶ Cluster 2 is dominated by light-bite culinary spots with a significant prevalence of hotels. Cluster 2's location, being proximate Cluster 0 and Cluster 1, is a right place to be dominated by places to hang out.
- ▶ Cluster 3, located in the extreme northwest and southeast, is dominated by convenience and grocery stores. We can recommend visitors to go there to buy goods.
- ▶ Cluster 4, the smallest, is dominated by food stalls.



# Conclusion

- ▶ Although not very accurate, we can build clustering models to help recommend visitors about right places to explore when they visit a particular city.
- ▶ The model can be improved by using larger data sets containing a larger number of venues. The data obtained may not be the latest, while Yogyakarta is a growing city, hence we need a more updated data to make a better recommendation.