

IBM Data Science Coursera Capstone Project

Introduction: Business Problem

Jakarta is the capital of Indonesia with a population of 10,5 million, and is the heart of the second densest metropolitan area in the world behind Tokyo, Japan. Having hosted ASEAN games recently back in 2018, it has witnessed heavy investment in transportational infrastructure with the opening of the first MRT line in Indonesia just this year.

Given that, the city will surely see more growth of which information regarding the lay of the land will be invaluable for investors and entrepreneurs to make strategic decision for investment or choosing locations for business operation.

This project will attempt to explore patterns of subdistricts within Jakarta by categorizing them into clusters in order to identify existing trends within neighborhoods of Jakarta. From there on, recommendations can be made on which category of neighborhood will be most suitable for a certain type of venue to be opened.

The result of this project is aimed at general entrepreneur but may be most useful for entrepreneurs on the food and beverage sector given that location can be the deciding factor for a success.

Data

To analyze trends in Jakarta's subdistrict, the list of subdistrict is obtained from [Jakarta subdistrict wikipedia page](#).

Venue queries will then be made by subdistricts using FourSquare APIs. The resulting data regarding venue category will be used to observe commonality between subdistricts. The commonality clusters can then provide insight on which type of venue will thrive better on which cluster. K-means clustering algorithm will be used to find pattern between the subdistricts.

In summary, the following data is required to meet the objective:

- Subdistricts of Jakarta
 - Coordinates of these subdistricts
 - Trending Venues on the area
 - Venue categories
-

Methodology

Given that our objective is to generally categorize the subdistricts, we will use K-means clustering algorithm to categorize each of the subdistricts within Jakarta.

A one-hot encoding will be done on the venue dataframe and it will be grouped by subdistrict. The encoding will return venue categories as column per subdistrict, which will then be grouped to provide weighting of venue type occurrence on each subdistrict.

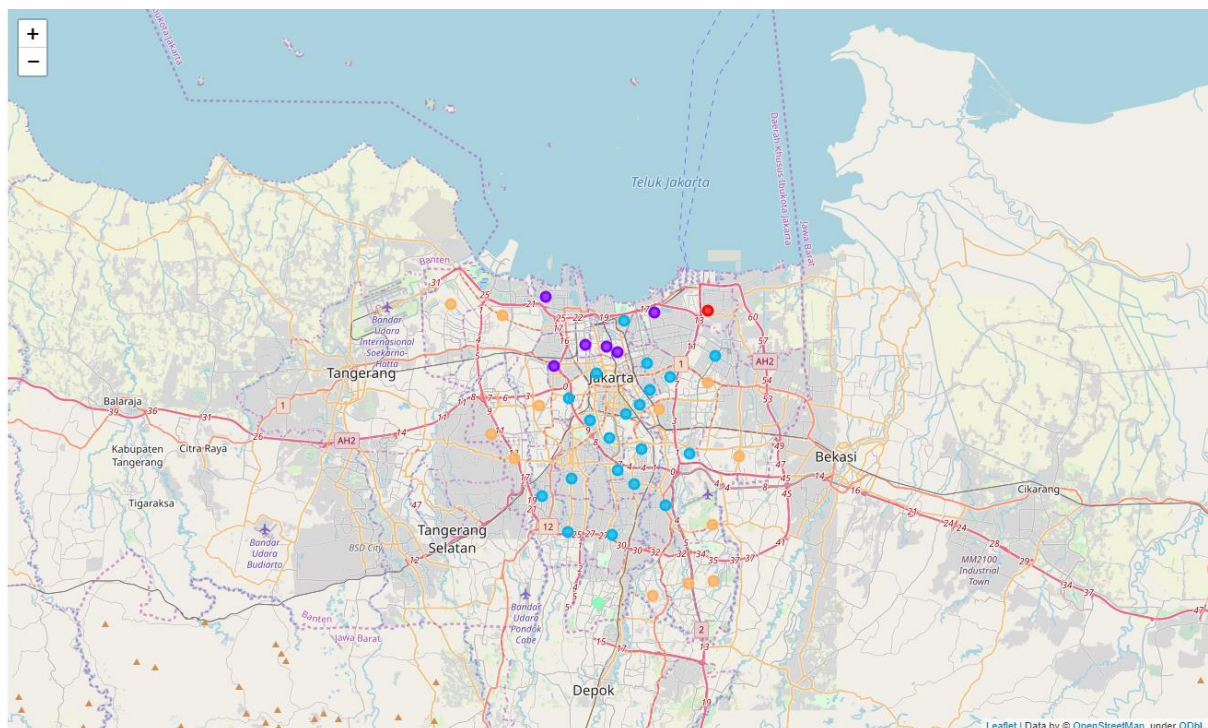
The encoded dataframe will be further filtered into top venues before the K-means clustering algorithm will be run over it. This will return cluster labels over the subdistricts.

The clusters will be observed one by one manually to determine its content.

Recommendation will be made based on the clustering.

Analysis

Map of Clusters



Cluster 1

Cluster 1 contains a higher concentration of noodle house and chinese restaurant. Located north of Jakarta, this might be the cluster for Chinese restaurants.

Subdistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Grogol Petamburan	Noodle House	Coffee Shop	Chinese Restaurant	Clothing Store	Asian Restaurant	Seafood Restaurant	Restaurant	Indonesian Restaurant	Steakhouse	Korean Restaurant
Taman Sari	Chinese Restaurant	Noodle House	Asian Restaurant	Hotel	Seafood Restaurant	Bakery	Coffee Shop	Fast Food Restaurant	Hotel Bar	Steakhouse
Tambora	Chinese Restaurant	Noodle House	Asian Restaurant	Hotel	Coffee Shop	Bakery	Food Truck	Indonesian Restaurant	Steakhouse	Massage Studio
Sawah Besar	Chinese Restaurant	Noodle House	Hotel	Seafood Restaurant	Coffee Shop	Indonesian Restaurant	BBQ Joint	Asian Restaurant	Café	Restaurant
Penjaringan	Seafood Restaurant	Chinese Restaurant	Noodle House	Coffee Shop	Café	Indonesian Restaurant	Bakery	Restaurant	Balinese Restaurant	Vegetarian / Vegan Restaurant
Tanjung Priok	Asian Restaurant	Chinese Restaurant	Noodle House	Seafood Restaurant	Indonesian Restaurant	Pizza Place	Beach	Massage Studio	Café	Restaurant

Cluster 2

With the most member, cluster 2 seems to contain a good amount of coffee shops and hotels. Located central-south of the city, it is expected to have a high concentration of places to hang out and stay.

Subdistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Palmerah	Coffee Shop	Asian Restaurant	Indonesian Restaurant	Hotel	Convenience Store	Food Truck	Café	Chinese Restaurant	Pizza Place	Clothing Store
Cempaka Putih	Indonesian Restaurant	Pizza Place	Coffee Shop	Fast Food Restaurant	Hotel	Asian Restaurant	Café	Restaurant	Indonesian Meatball	Food Truck
Gambir	Hotel	Indonesian Restaurant	Coffee Shop	Seafood Restaurant	Asian Restaurant	Fast Food Restaurant	Camera Store	Chinese Restaurant	Restaurant	Soup Place
Johar Baru	Indonesian Restaurant	Hotel	Pharmacy	Seafood Restaurant	Pizza Place	Restaurant	Convenience Store	BBQ Joint	Furniture / Home Store	Coffee Shop
Kemayoran	Hotel	Indonesian Restaurant	Indonesian Meatball	Asian Restaurant	Food Court	Seafood Restaurant	Chinese Restaurant	Convenience Store	Cupcake Shop	Gym
Senen	Indonesian Restaurant	Coffee Shop	Hotel	Café	Fast Food Restaurant	Bookstore	Restaurant	Asian Restaurant	Bakery	Chinese Restaurant
Tanah Abang	Coffee Shop	Hotel	Indonesian Restaurant	Japanese Restaurant	Restaurant	Lounge	Building	Italian Restaurant	Chinese Restaurant	Breakfast Spot
Cilandak	Coffee Shop	Asian Restaurant	Café	Food Truck	Indonesian Restaurant	Fast Food Restaurant	Steakhouse	Motorcycle Shop	Chinese Restaurant	Bakery
Kebayoran Baru	Indonesian Restaurant	Coffee Shop	Japanese Restaurant	Hotel	Noodle House	Food Truck	Sushi Restaurant	BBQ Joint	Breakfast Spot	Burger Joint
Kebayoran Lama	Coffee Shop	Indonesian Restaurant	Clothing Store	Café	Bakery	Shopping Mall	Pizza Place	Sushi Restaurant	Japanese Restaurant	Dessert Shop
Mampang Prapatan	Coffee Shop	Restaurant	Hotel	Asian Restaurant	Food Truck	Steakhouse	Indonesian Restaurant	Bakery	Fast Food Restaurant	Japanese Restaurant
Pancoran	Coffee Shop	Asian Restaurant	Convenience Store	Café	Fast Food Restaurant	Indonesian Restaurant	Supermarket	Noodle House	Hotel	Pizza Place
Pasar Minggu	Indonesian Restaurant	Food Truck	Coffee Shop	Japanese Restaurant	Convenience Store	Chinese Restaurant	Noodle House	Café	Bakery	Fast Food Restaurant
Setiabudi	Coffee Shop	Hotel	Italian Restaurant	Indonesian Restaurant	Chinese Restaurant	Japanese Restaurant	Fast Food Restaurant	BBQ Joint	Indonesian Meatball	Lounge
Tebet	Indonesian Restaurant	Coffee Shop	Clothing Store	Dessert Shop	Café	Bakery	Japanese Restaurant	Restaurant	Pizza Place	Food Truck
Jatinegara	Coffee Shop	Indonesian Restaurant	Asian Restaurant	Fast Food Restaurant	Food Truck	Café	Donut Shop	Spa	Shopping Mall	Movie Theater
Kramat Jati	Fast Food Restaurant	Indonesian Restaurant	Restaurant	Bakery	Asian Restaurant	Seafood Restaurant	Shopping Mall	BBQ Joint	Fried Chicken Joint	Donut Shop
Menteng	Indonesian Restaurant	Hotel	Coffee Shop	Betawinese Restaurant	Fast Food Restaurant	Café	Steakhouse	History Museum	Food Truck	Sandwich Place
Kelapa Gading	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	Bakery	Shopping Mall	Korean Restaurant	Sushi Restaurant	Japanese Restaurant	BBQ Joint
Pademangan	Seafood Restaurant	Hotel	Noodle House	Chinese Restaurant	Indonesian Restaurant	Coffee Shop	Theme Park	Theme Park Ride /	Asian Restaurant	General Entertainment

Cluster 3

Containing only 1 subdistrict, cluster 3 seems to be in its own group located at the south of the city. This might be due to the low count of trending venues of 10. Nothing much can be gathered from this cluster.

Subdistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Jagakarsa	Convenience Store	Burger Joint	Food Truck	Theme Park	Restaurant	Noodle House	Fast Food Restaurant	College Residence	Other Great Outdoors	Flower Shop

Cluster 4

Having Indonesian restaurants as its common venue, cluster 4 have higher count of fast food and convenience store in general. The subdistricts within this cluster are more scatered, being dispersed on around the border surrounding cluster 2.

Subdistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Cengkareng	Noodle House	Asian Restaurant	Café	Pizza Place	Bar	Bistro	Seafood Restaurant	Indonesian Restaurant	Steakhouse	Gym
Kalideres	Noodle House	Asian Restaurant	Coffee Shop	Pizza Place	Fast Food Restaurant	Food Court	Food Truck	Convenience Store	Japanese Restaurant	Café
Kebon Jeruk	Asian Restaurant	Convenience Store	Indonesian Restaurant	Coffee Shop	Pizza Place	Steakhouse	Noodle House	Café	Seafood Restaurant	Fast Food Restaurant
Kembangan	Shop & Service	Japanese Restaurant	Snack Place	Soup Place	Gas Station	Plaza	Diner	Park	Convenience Store	Indonesian Restaurant
Pesanggrahan	Food Truck	Noodle House	Fast Food Restaurant	Indonesian Restaurant	Convenience Store	Food Court	Café	Electronics Store	Asian Restaurant	Bakery
Cipayung	Indonesian Restaurant	Convenience Store	Garden	Arcade	High School	Pizza Place	Asian Restaurant	Food & Drink Shop	Noodle House	Food Stand
Ciracas	Fast Food Restaurant	Coffee Shop	Noodle House	Convenience Store	Asian Restaurant	Indonesian Restaurant	Water Park	Rest Area	Garden	Dim Sum Restaurant
Duren Sawit	Indonesian Restaurant	Fast Food Restaurant	Salon / Barbershop	Convenience Store	Food Truck	Noodle House	Indonesian Restaurant	Seafood Restaurant	Asian Restaurant	Department Store
Makasar	Indonesian Restaurant	Fast Food Restaurant	Golf Course	Noodle House	Pizza Place	Asian Restaurant	Bookstore	Monument / Landmark	Food Truck	Music Venue
Matraman	Indonesian Restaurant	Asian Restaurant	Pizza Place	Convenience Store	Coffee Shop	Pharmacy	Seafood Restaurant	Café	Fast Food Restaurant	Noodle House
Pasar Rebo	Indonesian Restaurant	Fast Food Restaurant	Seafood Restaurant	Dessert Shop	Baseball Stadium	Food Truck	Noodle House	Convenience Store	Factory	Café
Pulo Gadung	Noodle House	Indonesian Restaurant	Coffee Shop	Food Truck	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Café	Soup Place	Convenience Store

Cluster 5

Similar with cluster 3 with only 1 subdistrict, cluster 5 seems to be in its own group located at the north east of the city. This might be due to the low count of trending venues of 15. Nothing much can be gathered from this cluster.

Subdistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Koja	Indonesian Meatball	Indonesian Restaurant	Pizza Place	Restaurant	Convenience Store	Satay Restaurant	Bakery	High School	Government Building	Food & Drink Shop

Result and Discussion

Groupings as a result of K-means clustering algorithm tallies with how Jakarta historically develops. Having most of cluster 1, containing a high count of chinese restaurant, at the north side of the city fits the chinatown part of the city. Cluster 2 being the dominant type of subdistrict which is located in the middle also fits the reality. North eastern part being quite sparse in trending venue also fits the reality that the area is more of an industrial area, thus having less venues.

There are definite limitation with using the FourSquare API as the 100 venues limit might skew the result of the more densely populated subdistrict. Also, some subdistricts have low count of venues that it might be considered to be insufficient in determining its characteristics. It might also be the case that FourSquare user base are skewed to the foodie type, which might explain the limited trending venues on the north east part of the city. For most of the subdistricts, restaurants and coffee shops are the dominant venue type with cluster 2 having more variation in terms of cuisine.

Conclusion

Opening of new western restaurant may be best done in cluster 5 where there are less of such restaurant to compete. Business which does not rely on foot traffic may choose to locate themself in the north east of Jakarta.