SEGMENTING AND CLUSTERING NEIGHBORHOODS IN WEST JAKARTA

Coursera Capstone Project IBM Applied Data Science

Rizal Mahifa

Introduction

DKI Jakarta is a metropolitan area in Indonesia. West Jakarta is one of the city in the DKI Jakarta area. Many people who come to West Jakarta for continue their education at university, work and others. To support these needs, many people are looking for a place to live that is in the vicinity of these needs. So people need a quick overview of which neighborhoods they will choose in West Jakarta that can facilitate their activities. Based on this problem, clustering is one approach that can solve this problem. With clustering, we can see which neighborhoods in West Jakarta have residential areas or apartments around universities, workplaces, places to eat and others. Especially for those who are Muslims can see the area of residence or apartment they will choose, have a halal eating places and places of worship around the residence or the apartment or not.

Data

First, I would like to parse neighborhoods which are districts in West Jakarta at the link https://en.wikipedia.org/wiki/West_Jakarta. Next, every neighborhood I will determine latitude and longitude using a geocoder. Then, for each neighborhood in West Jakarta, I will look for apartments in the neighborhood using Foursquare api and determine the latitude and longitude of the apartment. The final step in this stage to collecting data is to determine the nearest venues (1 km from the apartement) using the foursquare api with the conditions already mentioned in the previous business problem.

This table is show the neighborhoods in West Jakarta. I get the latitude and longitude using geocoder.

	Neighborhoods	Residence Type	Residence Name	Venue	LatNeigh	LongNeigh	LatRes	LongRes	LatVen	LongVen
0	Cengkareng	Apartement	City Garden Apartemen	University	-6.152899	106.744718	-6.154850	106.741723	-6.155088	106.749034
1	Grogol Petamburan	Apartement	Kost susilo 2a	University	-6.164188	106.788317	-6.165551	106.792227	-6.167022	106.791305
2	Grogol Petamburan	Apartement	Kost susilo 2a	University	-6.164188	106.788317	-6.165551	106.792227	-6.167662	106.791945
3	Grogol Petamburan	Apartement	Kost susilo 2a	University	-6.164188	106.788317	-6.165551	106.792227	-6.167377	106.791000
4	Grogol Petamburan	Apartement	Kost susilo 2a	University	-6.164188	106.788317	-6.165551	106.792227	-6.167662	106.791461

Methodology

- **Web scraping Wikipedia page for neighborhood list**

  In this step, I do web scraping from https://en.wikipedia.org/wiki/West_Jakarta using BeautifulSoap library in notebook to get neighborhoods in West Jakarta.

- **Get latitude and longitude coordinates using Geocoder**

  After first step done, next I use Geocode for get each latitude and longitude coordinates for neighborhood of West Jakarta.

- **Use Foursquare API to get apartment data for each neighborhood and to get venue data for each apartment.**

  In this step, I use Foursquare API to get apartment data and latitude and longitude from apartment and for each apartment, I use Foursquare API for get venue for the apartment.

- **Group data by neighborhoods and apartment and taking the mean of the frequency.**

  In this step, group the data based on neighborhood and apartment, then taking the mean of the venue frequency

- **Perform clustering using K-Means**

  Using K-Mean clustering to cluster venue based on neighborhood and apartment

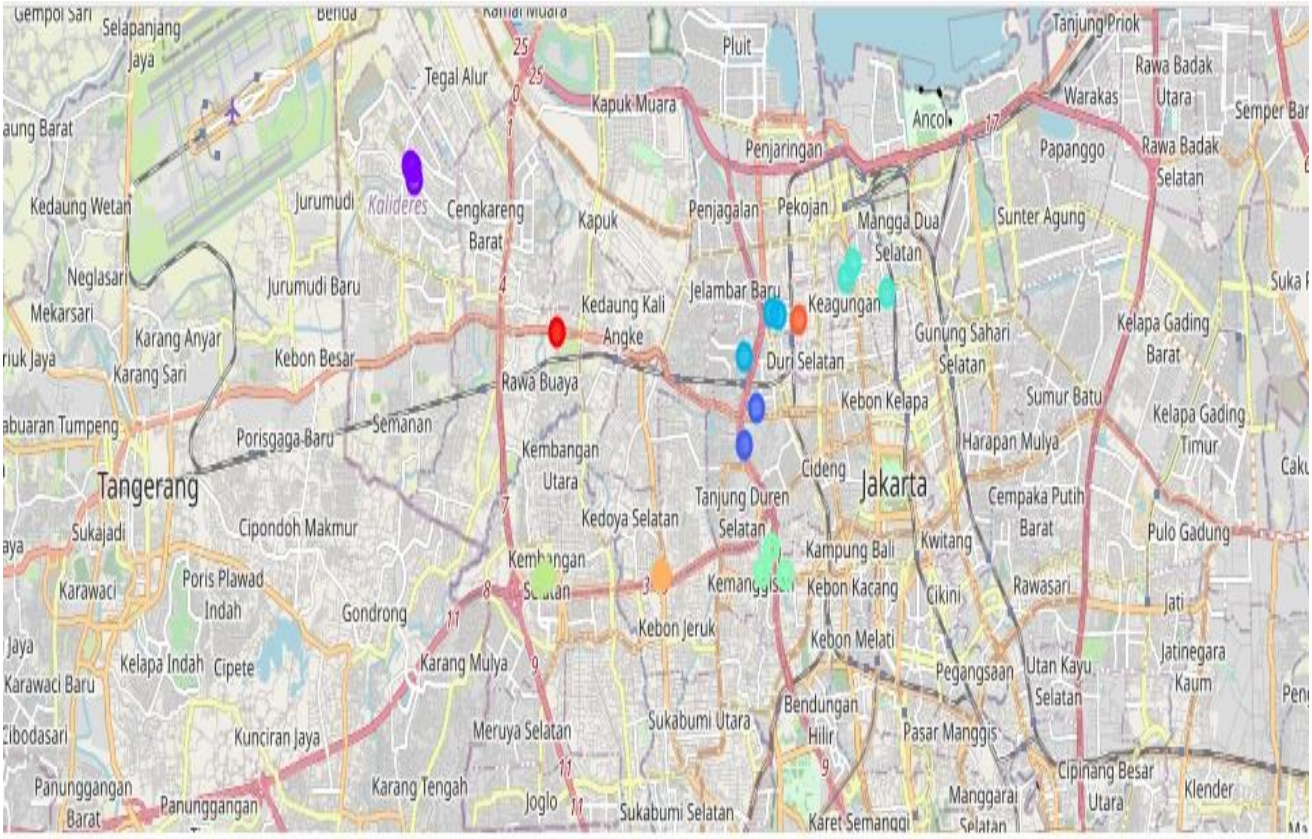
- **Filter venue category by University, Food and Office**

  Filter the grouped data after apply K-Means by University, Food and Office venue

- **Visualize the cluster in map using Folium**

  Visualize the data on the map using Folium to see the cluster

Result



Based on our analysis above, we can build conclusions that will be useful to aid people who want visit the West Jakarta, Indonesia for living and continue their study at University, looking for job or all of it.

From clustering, there are top three clusters which is can be choose from eight cluster:

- Cluster 3 : For people who want has a job and studying at university
- Cluster 1 : For people who only looking for a job or already have a job
- Cluster 2 : For people who only want to continue their study at university or already studying at University

Conclusion

The aim of this project is to explore the residence in West Jakarta that people choose to live and carry out their activities such as lecturing, eating and working. These residence have been identified using Foursquare and have been plotted on the map. The map reveals that there are three main areas can be choose: Cluster 3, Cluster 1 and Cluster 2. Based on their needs, people can choose between three clusters.

- If people who want has a job and studying at university, they can choose apartment in cluster 3
- If people who only want has a job, they can choose apartment in cluster 1
- If people who only want has studying at university, they can choose apartment in cluster 2