Segmenting and Clustering Neighborhoods in West Jakarta

Coursera Capstone Project

IBM Applied Data Science Capstone

Business Problem

- Many people are looking for a place to live that is in the vicinity of their needs.
- Objective : To analyze and select the best apartment which is based on some considered based on people needs.
- Business question :

  If someone come to West Jakarta, which area and apartment they can choose for living and support their needs?

Data

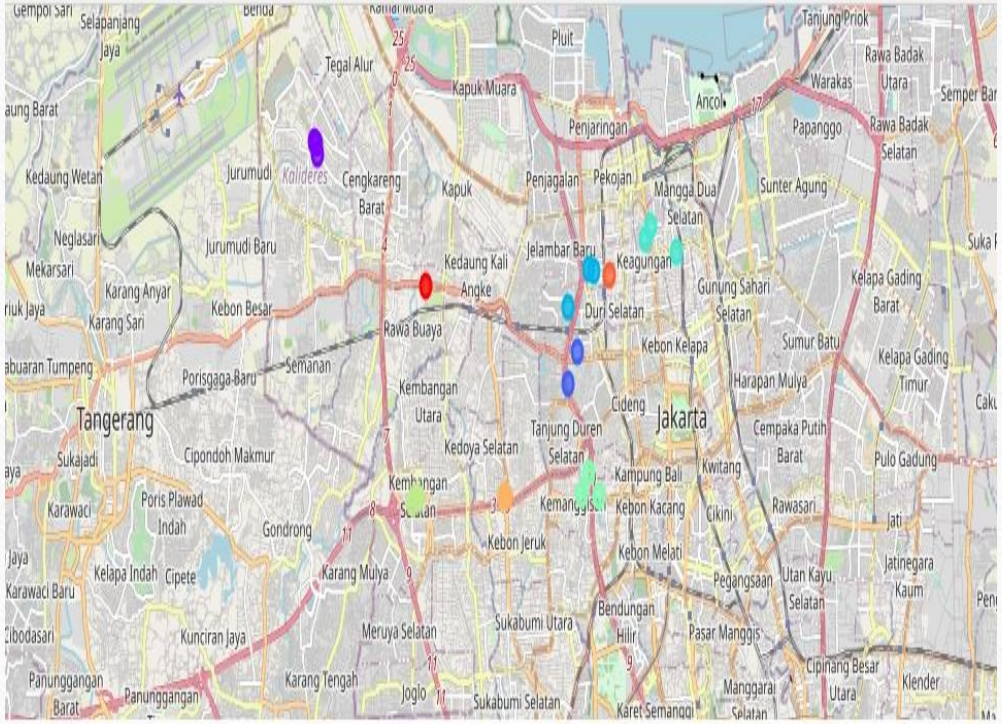
- Data Required
  - List of Neighborhoods in West Jakarta.
  - Latitude and Longitude coordinates of the neighborhoods.
  - Venue data , particularly data related to the problem (Food Place, University and Office).
- Source of Data
  - Wikipedia for neighborhoods.

    https://en.wikipedia.org/wiki/West_Jakarta.
  - Geocoder package for latitude and longitude coordinates.
  - Foursquare API for venue data.

Methodology

- Web scraping Wikipedia page for neighborhood list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get apartment data for each neighborhood and to get venue data for each apartment.
- Group data by neighborhoods and apartment and taking the mean of the frequency.
- Perform clustering using K-Means
- Filter venue category by University, Food and Office
- Visualize the cluster in map using Foium

Result



- 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 want has a job
  - Cluster 2 : For people who only want has studying at university

Conclusion

- Answer to business question :
  - 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