Cluster District in Bangkok

Narongchai Jaroonpipatkul

May 05, 2021

2. Data acquisition and cleaning

2.1 Data sources

- I scrape common districts in Bangkok data from wikipedia.
  (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok)
- I use statistics data about area in each district in Bangkok and calculate density population.
  (http://www.bangkok.go.th/upload/user/00000130/Logo/statistic/stat%202015(ENG).pdf)
- I used Forsquare API to get the venues of each district in Bangkok.
  (https://developer.foursquare.com/)

2.2 Data cleaning

After searching for data on many websites, some websites have data not match. So, I decide to select from the most reliable source.

There are not too many public data related to demographic and social parameters. In many case, I setup my on table.

2.3 Feature selection

After data cleaning, there were 50 districts with 164 venue categories. I have data on the population and area size of each district. Then, I calculate population density and use for built model.