Jakarta – Seoul Sister City

SIMILARITIES AND DISSIMILARITIES

Searching new opportunity and lesson from Sister City

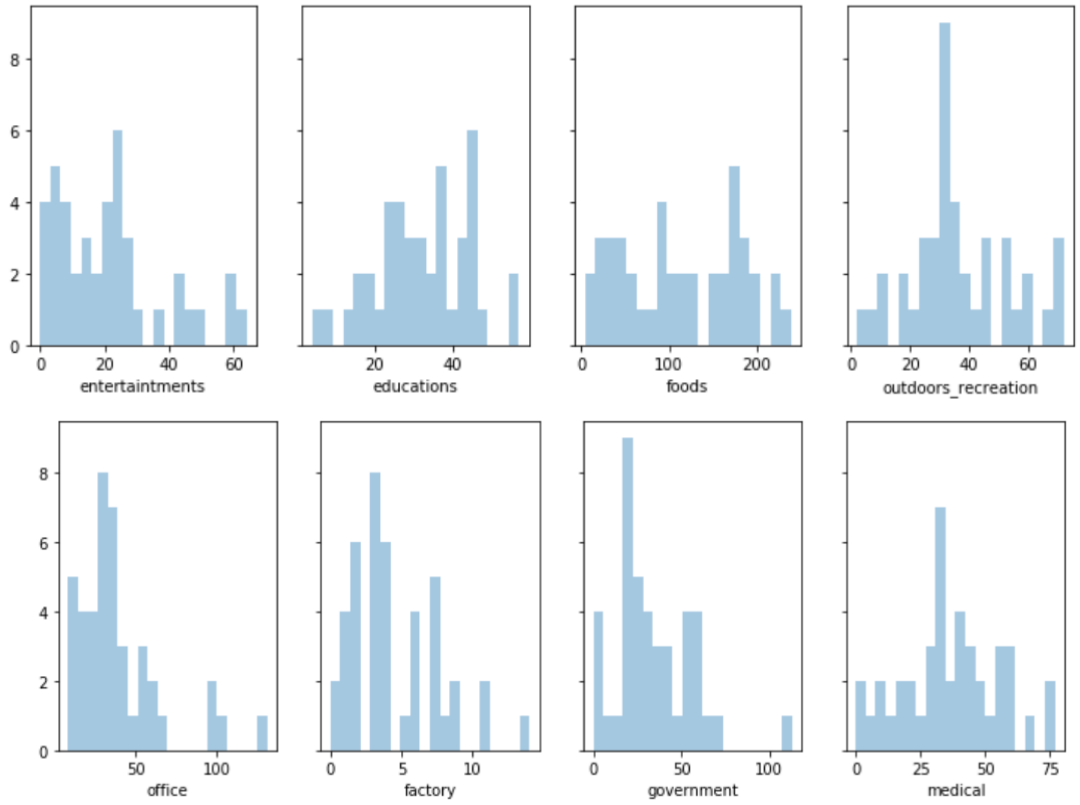
- Jakarta and Seoul have become sister city since 1984
- Looking for new opportunity and lesson learned from Seoul by finding similarities and dissimilarities between two cities.
- By looking into dissimilarities expecting to find area of improvements.

Data Acquisition and Cleaning

- List of Jakarta neighborhoods scrapped from Wikipedia page.
- List of Seoul neighborhoods scrapped from Wikipedia page.
- Longitude and Latitude fetched with Nominatim geocoder.
- Neighbourhood venue fetched through Foursqare explorer API calls.
- Remove unnecessary fields and irrelevant neighbourhoods.
- In total 42 rows for Jakarta and 24 rows for Seoul, both contains 9 features.

Methodology

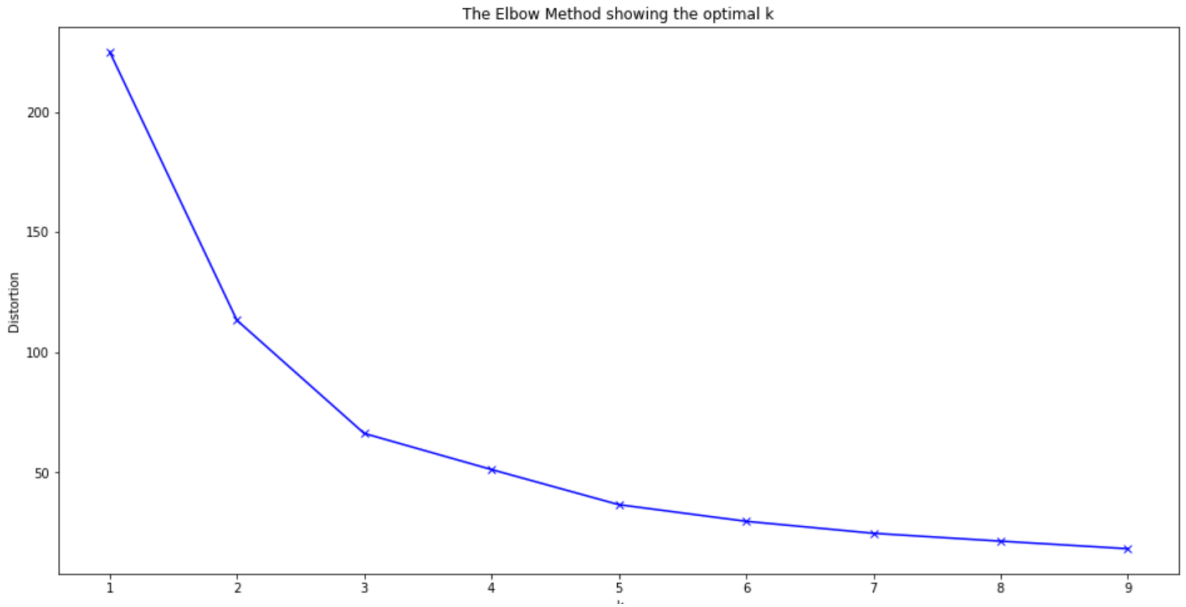
Exploratory Data Analysis



	Latitude	Longitude	entertaintments	educations	foods	outdoors_recreation	office	factory	government	medical	shopping_service
count	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000	42.000000
mean	-6.207038	106.836402	21.500000	32.023810	113.095238	36.214286	39.119048	4.523810	34.357143	36.619048	54.428571
std	0.060899	0.054728	16.653206	12.202535	67.321511	17.150035	26.714573	3.179493	22.355030	18.818715	29.285094
min	-6.330008	106.701594	0.000000	4.000000	4.000000	2.000000	8.000000	0.000000	0.000000	0.000000	5.000000
25%	-6.247659	106.800587	8.250000	23.000000	50.500000	27.000000	21.250000	2.000000	19.250000	24.750000	31.250000
50%	-6.193588	106.832902	21.000000	33.000000	109.000000	32.000000	32.000000	4.000000	29.500000	34.500000	51.000000
75%	-6.160590	106.870367	26.000000	42.000000	172.750000	45.000000	45.250000	6.750000	51.250000	49.500000	73.750000
max	-6.117265	106.944454	64.000000	57.000000	239.000000	72.000000	131.000000	14.000000	113.000000	77.000000	121.000000

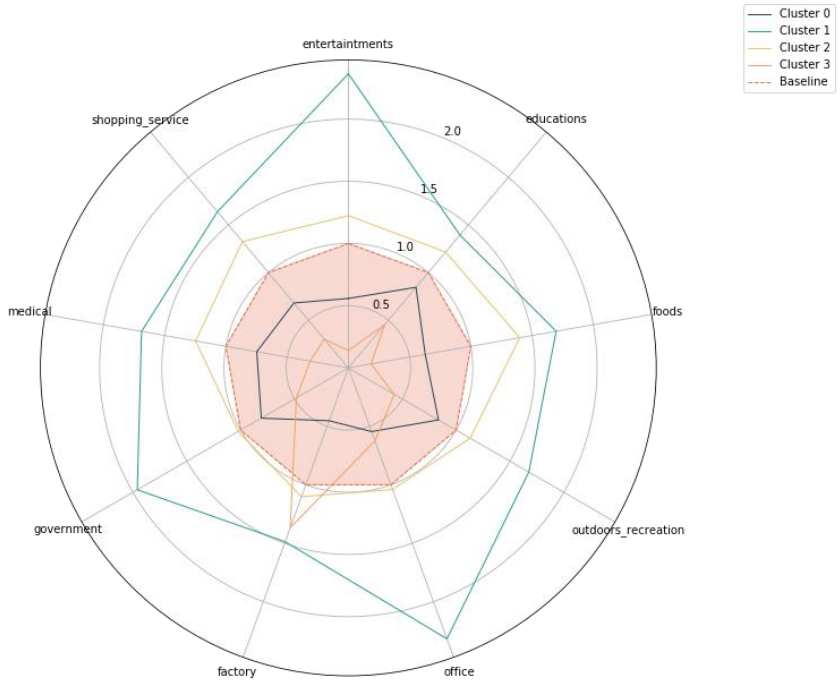
From the EDA we found that feature engineering is needed. The numeric data are scaled using standard scaler.

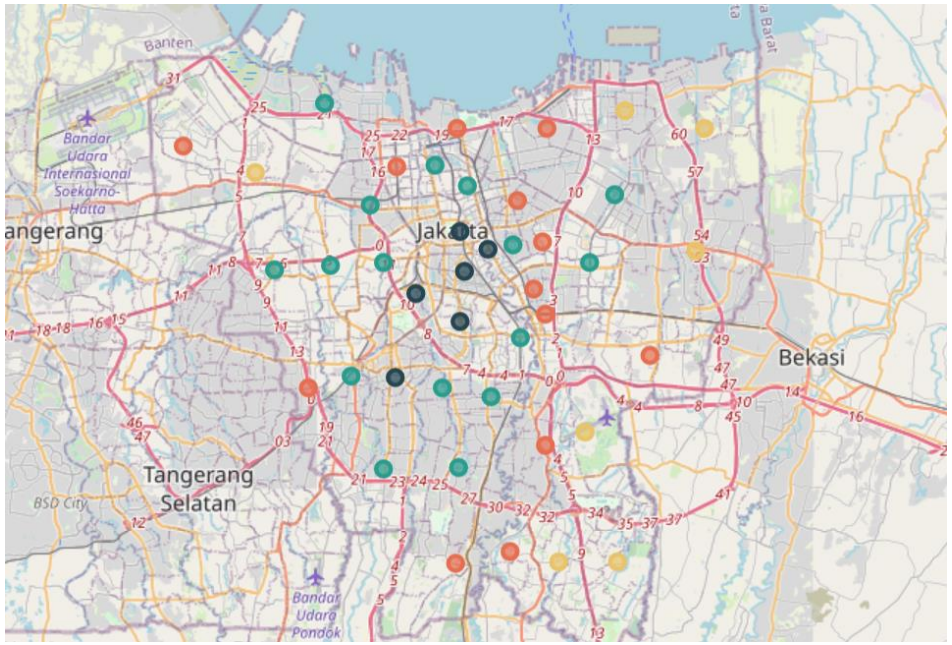
Find Optimal Clusters



- Optimal cluster according to elbow chart is k=3 but I decided to use k=4 in order to get more distinctive clusters.
- Same number of clusters for both Jakarta and Seoul

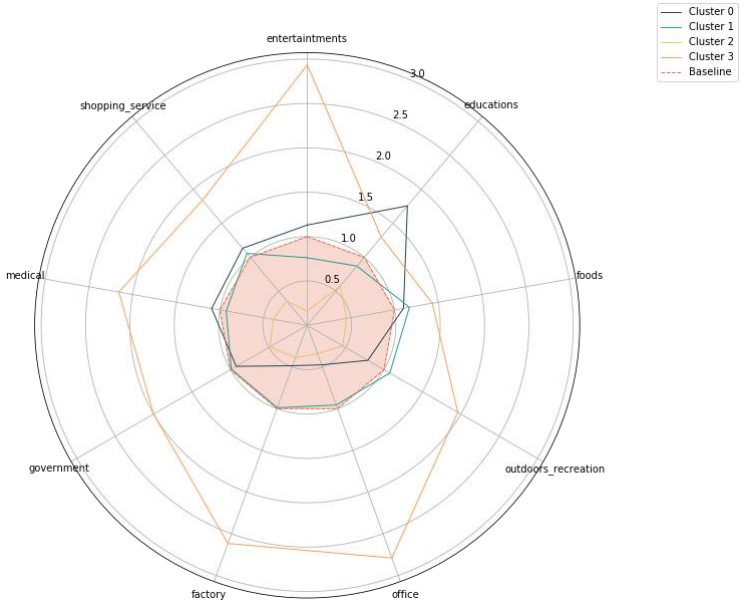
Jakarta Neighborhood Segmentation

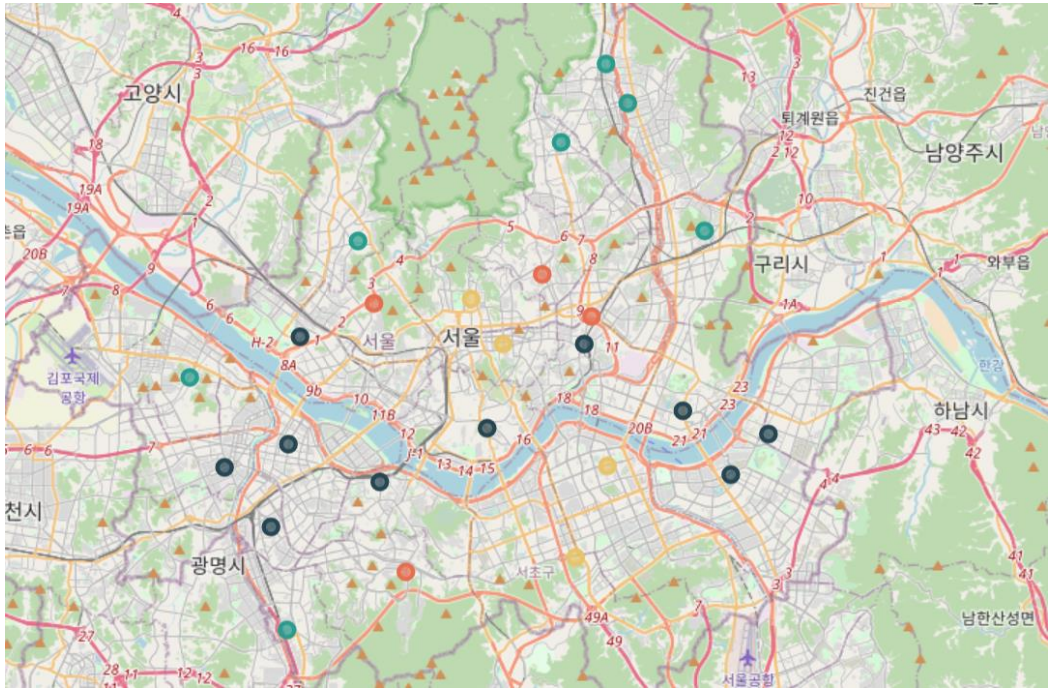




Jakarta neighborhood segmented into 4 cluster: sub-urban cluster, city center, average cluster and industrial cluster.

Seoul Neighborhood Segmentation





Seoul neighborhood segmented into 4 cluster: education cluster, average cluster, sub-urban and city center.

Conclusion and Ideas

- By segmenting Jakarta and Seoul neighborhoods into clusters with Kmeans, we can find similarities and dissimilarities between two cities.
- The quality of the study can be improved by :
  - Acquire data from more credible source
  - Using more granular venue categories