

Average Monthly Income & Venues Data Analysis of Thailand

Chusana Nuntanart

**A1 Background and Problem**

Thailand the country at the center of the Southeast Asian Indochinese with population about 68 million people, composed of 77 provinces, was the most visited country in Southeast Asia in 2013, according to the World Tourism Organization.

As a resident of Thailand, I would like to study the average monthly income in each provinces of Thailand therefore both government and investor might use the information for making long-term strategic planing.

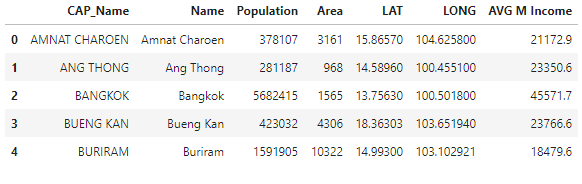
In order to forecast the average monthly income, we could use the venues data because in each provinces if people has higher monthly income, the consumable ability will be higher then the investment in each venues will be higher and vice versa.

**A2 Description of the data**

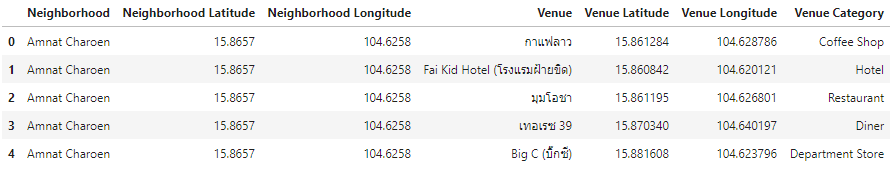
* Name of each provinces from Wikipedia
  + Cleaning and Making initial dataframe
* The average monthly income from Thailand government open data.
  + Matching with the dataframe
* Latitude and Longitude from Google.
  + Matching with the dataframe
* Common venues from Four Square for each provinces of Thailand.
  + Ranking the most common venues in each provinces
  + Merging with the dataframe
* Thailand Geojson file from Github/Apisit for choropleth map.
  + Solving the unmatched province name.

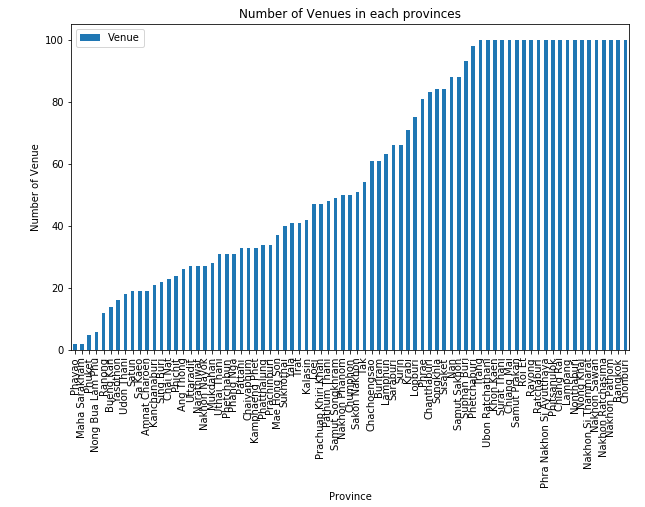
**B.Methodology**

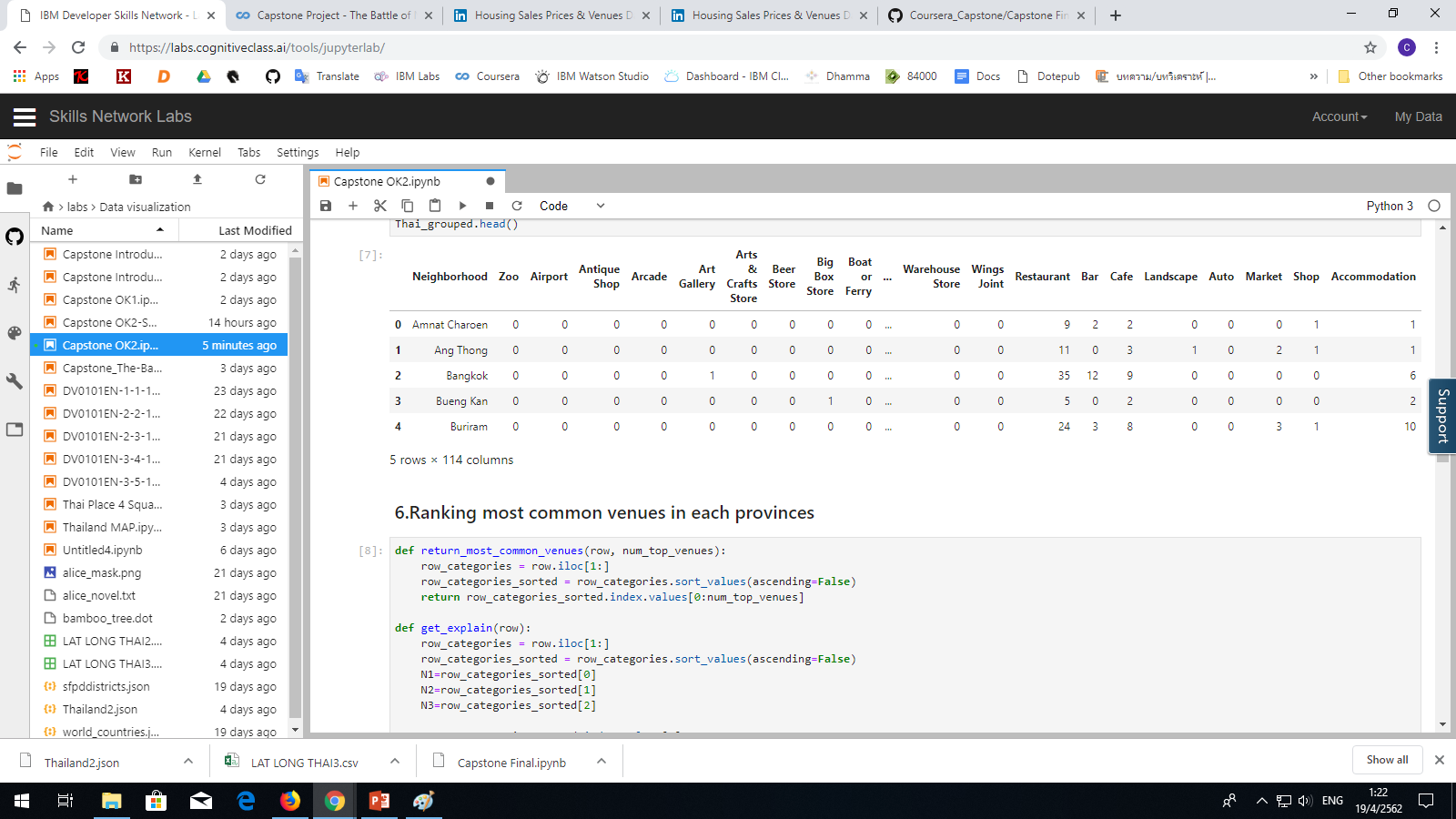
1. Load prepared data

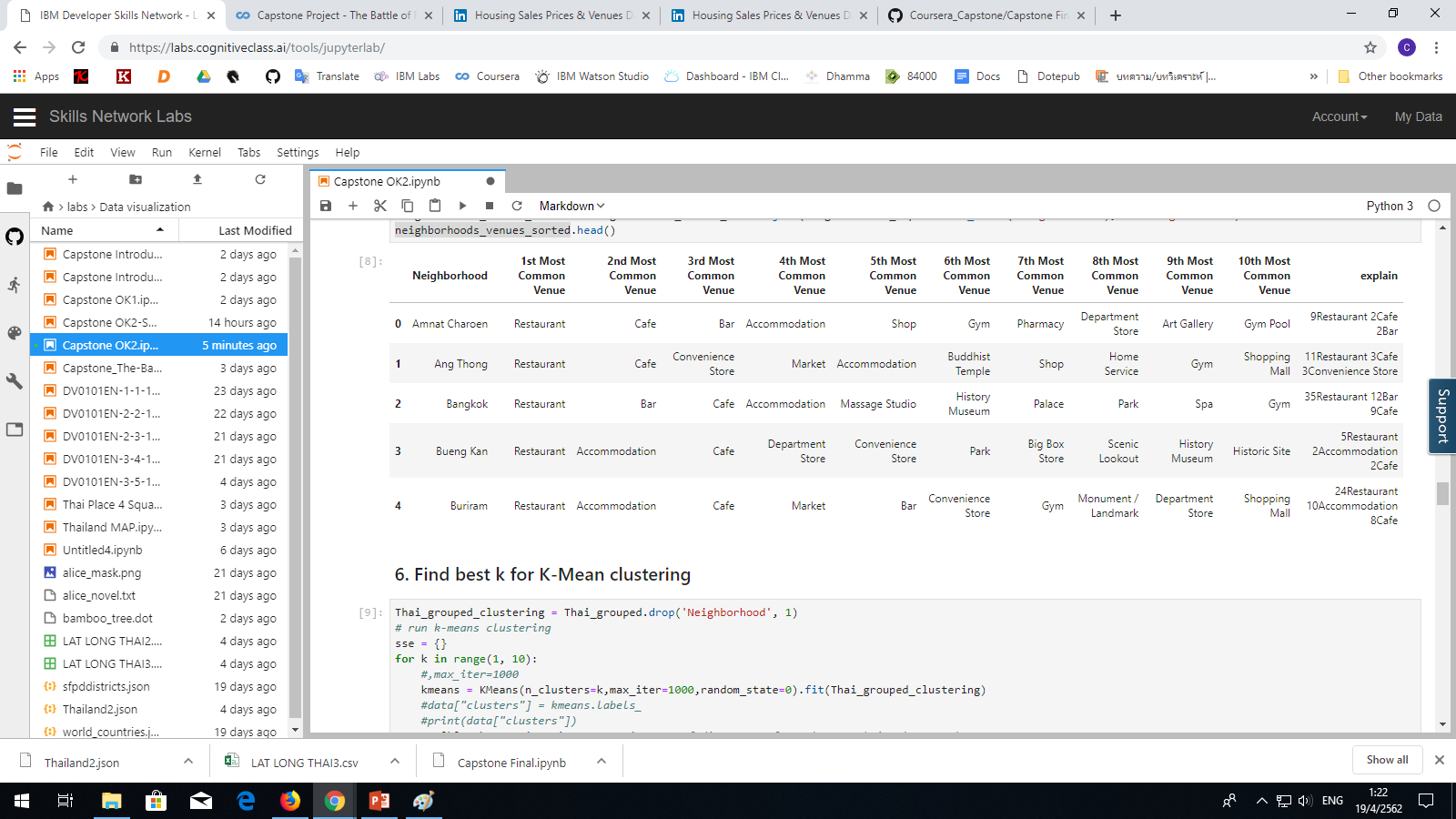
****

2.Get Venue data from Four Square

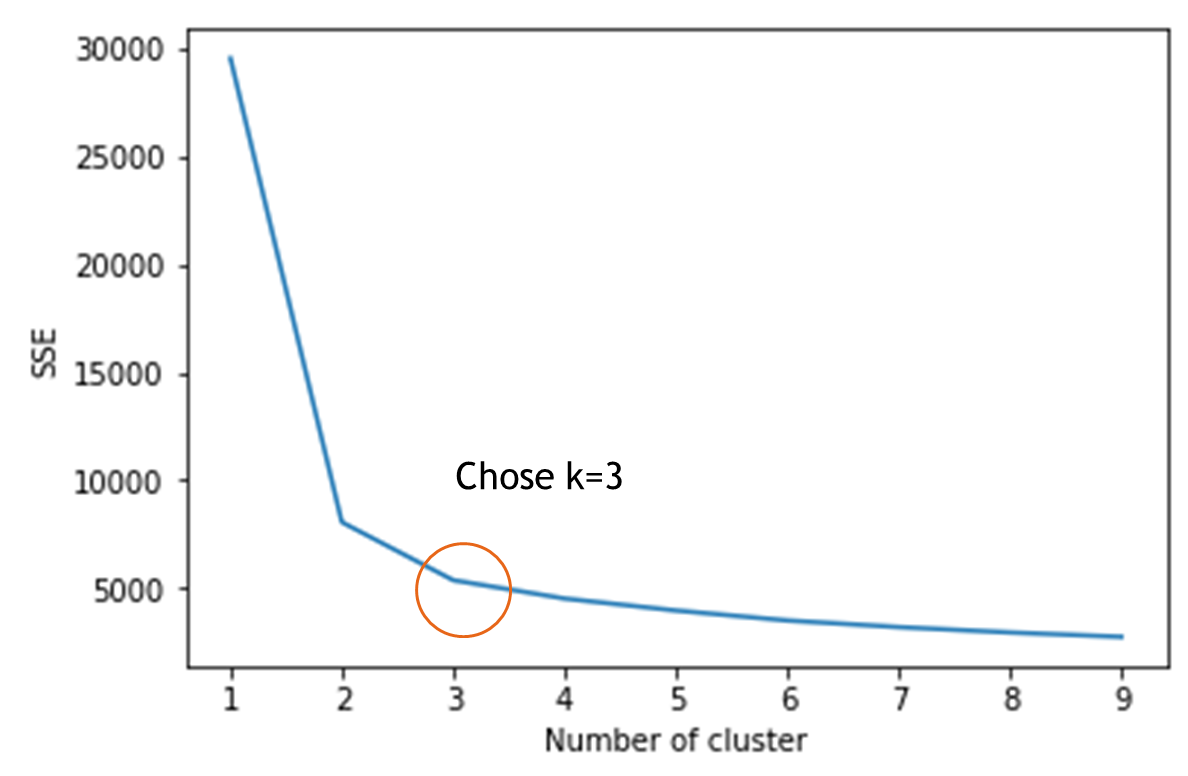
****

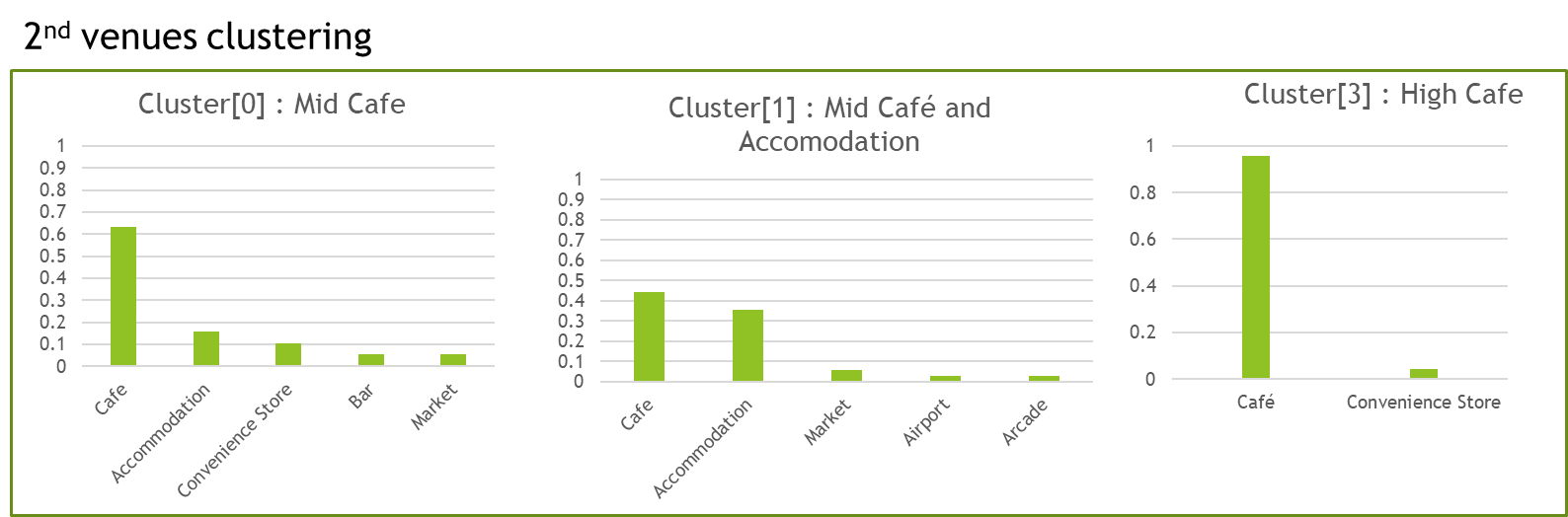
3.Check Number of venues in each provinces 

4.Summary data from Four square

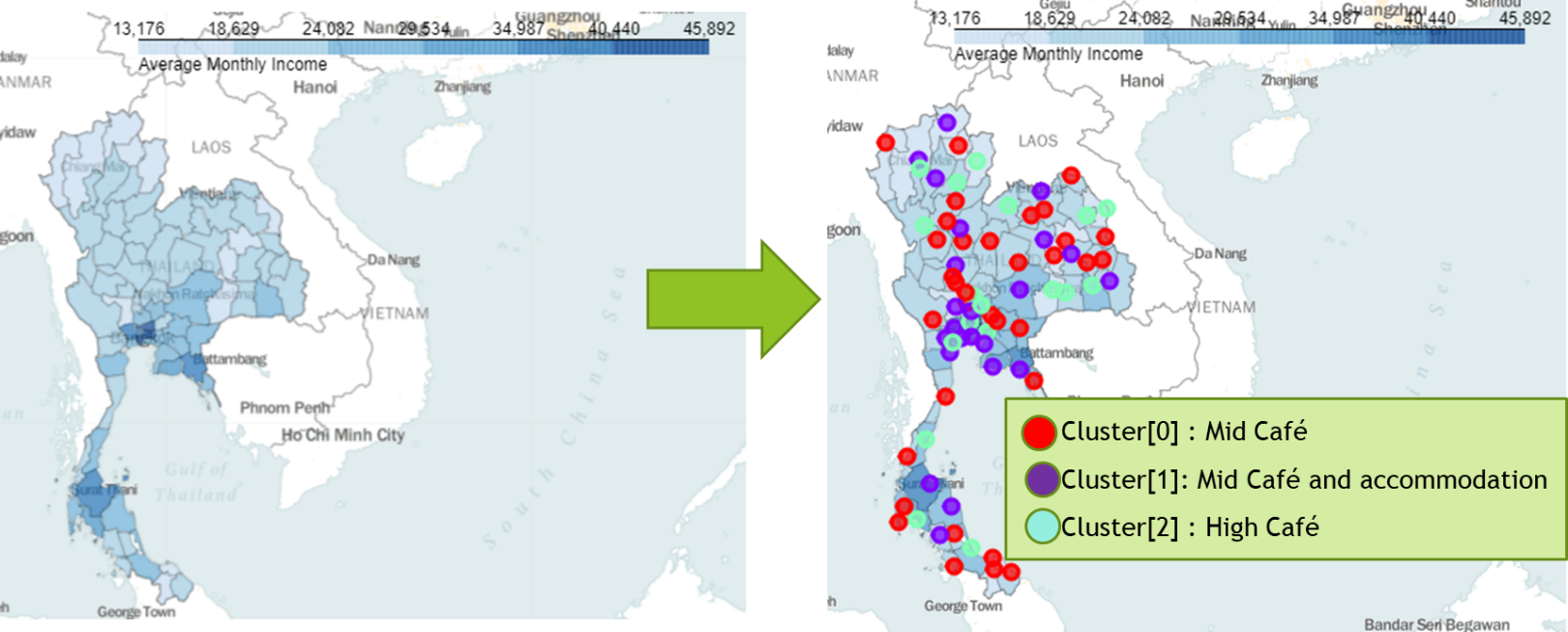
5. Ranking most common venues in each provinces

6. Find best k for K-Mean clustering



7. Define Cluster : But the 1st Most Common Venue is Restaurant so we check the 2nd venues to plot graph

**C.Results**

****

**D. Discussion**

First after we get venues data from Four square we can see that some provinces has very low venues and some provinces has reach limit 100 venues because there are high and low density of population in Thailand.

Then we do the ranking and found that the restaurant is the most common venues in each country, we can also see it from the grouping there are many kind of restaurant in Thailand.

Then we try to find the best K in K-Mean clustering by elbow method then we chose k=3 to do the clustering

After we did clustering the results we can see that the restaurants is the highest common venues in every provinces,then the cafe and accomodation, so we can see that we can define the name by

- Cluster[0] : Medium Cafe

- Cluster[1] : Medium Cafe and Accomodation

- Cluster[2] : High Cafe

Finaly we map each group of provinces based on each cluster, We can see that at the Bangkok capital provinces of ThailandThere are many "Cluster[2] : High Cafe" in that area and also Bangkok has the most average monthly income in Thailand.

**E.Conclusion**

As a result, the higher average monthly income province tend to has many restaurant and cafe, we may observe the growth or expansion of the province by observe new restaurant or cafe in each sub district then predict the expansion of its.

Both government and investor also can use this analysis to make a better strategy to manage city or provide facility service or even invest in land by further analysis in each sub-district.