

1. A description of the problem and a discussion of the background. (15 marks)

I am resident of Norway for over 12 years. Norway is one of the Scandinavian countries, with total population of 5,328,212 (as of 2018).

The total number of immigrant population is 765,108, roughly 14% of total population. Norway is divided into 18 counties and 422 municipalities. With low population, good income, Norwegians do enjoy highest standard of living. The social security system is Norway is said to be well developed - people pay high tax and in turn less privileged are taken well care of.

In this backdrop, I would like to do an exploratory analysis on socio economic conditions in Norway across various municipalities. Some of the interesting questions I would like to raise and find answers to are :

1. Why some municipalities operate profitably and few donot?
2. Does the immigrant population adds strength to local economy of municipalities?
3. Does education level play a role?
4. How does crime rate distributed across municipalities? Are there any links to education, income and immigration?
5. Based on our findings could we recommend a municipality or a region to someone immigrating to Norway?

2. A description of the data and how it will be used to solve the problem. (15 marks)

I see the main purpose of the project is to apply learned skills and tools in previous courses, which include,

1. Data scrapping
2. Data wrangling
3. Run appropriate ML algorithms
4. Use co ordinates
5. Data visualization

I have to try to engage all the above in my work. And the problem and data should support the intention.

Scrapping and data manipulation :

Norwegian statistics department operates a webpage. The basic facts of each municipalities are available in their webpage <https://www.ssb.no/kommunefakta/>.

The data for each municipalities looks as this <https://www.ssb.no/kommunefakta/agdenes> . Like this there are 422 municipalities.

Each page contains demographics, income, education, religion and net income.

We can use beautiful soup method to scrape those details along with municipality name and code.

The crime data is available separately in a tabular form. It is actually called “ offence registered at police “. The table has various types of police complaints registered per police district. We may have to integrate them and link to municipalities.

Geo co ordinates

Geo Coordinates of each municipalities need to be found. I was able to find a table with postcode and their co ordinates. The problem is each municipality have several post codes. We need to link post codes to municipality name so that each municipality will have co ordinates.

Machine Learning

I intend to use Kmeans clustering to cluster municipalities based on socio economic indicators.

Geo location

At some point I need to display cluster labels on a folium map. To do that I need to create a geo json file to link my labels to co ordinates in the map.

I have run some preliminary test and the result I try to get looks like this.

