

Analysis of socio economic conditions among Norway communes

Using Data science approach

By : Logesh G

Introduction & Problem description

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 and needy are taken well care of. As the economy grows, business activities increase, Norway needs more skilled and unskilled immigrants in their workforce. Despite high standard of living, educated society and appreciable law and order, Norwegian society is often criticized for not being open when it comes to integration with immigrants. On the other hand Norwegian society is a moral, equal and egalitarian (classless, horizontal) society with individuals constitutional rights held high.

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 attempt to find answers to are :

1. Why some municipalities operate profitably and few do not?
2. Does the immigrant population adds strength to local economy of municipalities?
3. Does education level play a role?
4. How does crime rate distribute 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?
6. If I were to move to Norway, which commune or locality would I choose?

This project and the report attempts to provide better understanding of socio economic conditions in Norway, based on realistic figures. It is targeted for people who wants to know more about Norway or to those living in Norway to know more about the municipalities they live in.

Data and tools

As mentioned Norway's administration is divided into 18 counties and 422 Municipalities called Communes. We need at least few key indicators for each commune.

Norwegian Statistics department is well efficient in maintaining reliable statistics. They are published openly as well.

1. Commune wise indicators

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 municipality looks as this <https://www.ssb.no/kommunefakta/agdenes> . Like this there are 422 municipalities.

Each page contains demographics, income, education, religion and commune's profitability.

I used algorithmic scrapping - 'Beautiful soup' method to go to each webpage and scrape necessary information out of each commune page.

2. Crime data

Crime rate is one of the important indicators to assess well-being at living place. Unfortunately, there is no crime data available in commune pages. However, statistics department has a table available that contains total number of registered complaints per police district.

Utilizing data manipulation and data wrangling, we can find offences per 1000 inhabitants registered at police. We can then separate them into commune wise figures.

3. Geo Co ordinates

We need geo coordinates to create visual impact of our analysis. Geo Coordinates of each municipalities or their centroids are not directly available. I was able to find a table with postcode and their coordinates. The problem is each municipality have several post codes.

Using data manipulation techniques, we can extract coordinates per postcodes into commune wise then find the centroid for each municipality.

4. Geojson

Geojson file is necessary to link our findings into geographical spread. Geonorge is the official body for creating and maintaining maps in Norway. From their page, it was possible to extract polygons for each commune and was made into geo json file.

Methodology

Stage1:

Understanding the requirement and what is needed is the first step. To assess the living conditions in a particular country, country, city or locality, following indicators are needed.

1. Name of commune/official code
2. Population
3. Education level
4. Income per capita
5. Expenditure per capita
6. Debt per capita
7. GDP if its country else profitability if its commune
8. Number of immigrants
9. Immigrants origin (top10 countries)
10. Crime rate
11. Location information

These indicators are not exhaustive but enough to assess and analyze living conditions.

Our goal is to collect these data from various sources and create a master data table with all of them for each commune.

Data preparation took lot of time and effort in data manipulation, wrangling, aggregating and merging them based on commune codes.

Some assumptions are made such as:

There is no direct information available on total immigrant population. The commune wise webpage had a bar chart of top10 countries contributing to immigration but not total.

However, it had a statistic on number of people registered with Norwegian church and those with other faiths. Norway being a Christian country, number of people with other faiths gives indirectly close figure to total immigrants.

Stage 2:

Once data is prepared we can cluster or segment them based on chosen features. In our case I have decided to use K means algorithm for clustering. I have used features such as population, percentage of university graduates, total immigrants, income per capita, debt per capita, communes profitability figure and average offence per thousand inhabitants (average of police complaints over last 10 years)

There are several algorithms and methods to choose for this purpose. The reason for choosing k means is because of its simplicity, the K-means is vastly used for clustering in many data science applications, especially useful if we need to quickly discover insights from unlabeled data. Our kommune data is a good candidate to be clustered using this method.

Some real-world applications of k-means:

Customer segmentation

Understand what the visitors of a website are trying to accomplish

Pattern recognition

Machine learning

Data compression

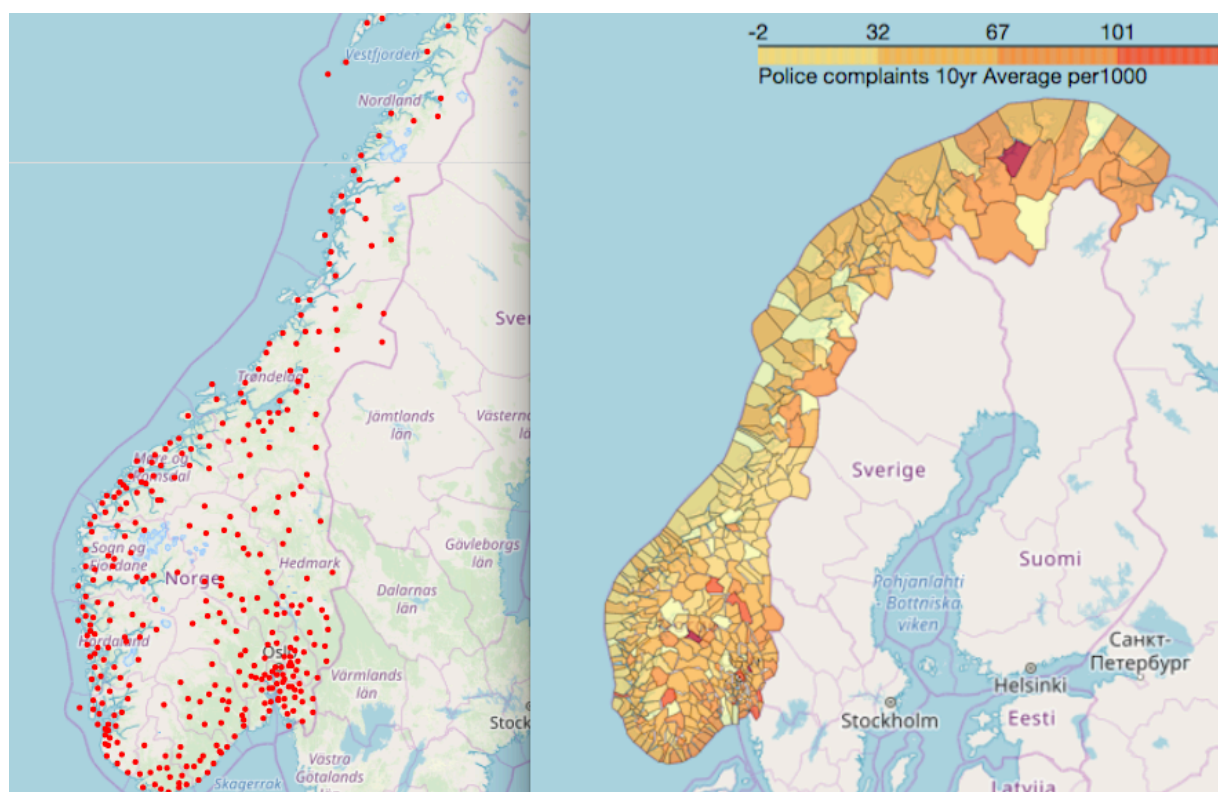
Stage3:

As we aggregate the master table from various data sources, we constantly visualize them to get better understanding of our data.

When they are clustered we examine them using pie charts, scatter plots and map overlay using folium.

Results

As we have data from various sources, each of them were in different formats. After tedious data wrangling , the dataset was ready for clustering. Here is how the data looked before clustering.



Map on the left shows centroids of each commune with average offence/1000 (crime registered). Map on the right shows same information on choropleth map with communes.

Crime rate is one of the many indicators we have. We use many indicators as feature variables in clustering. The communes were clustered into five groups.

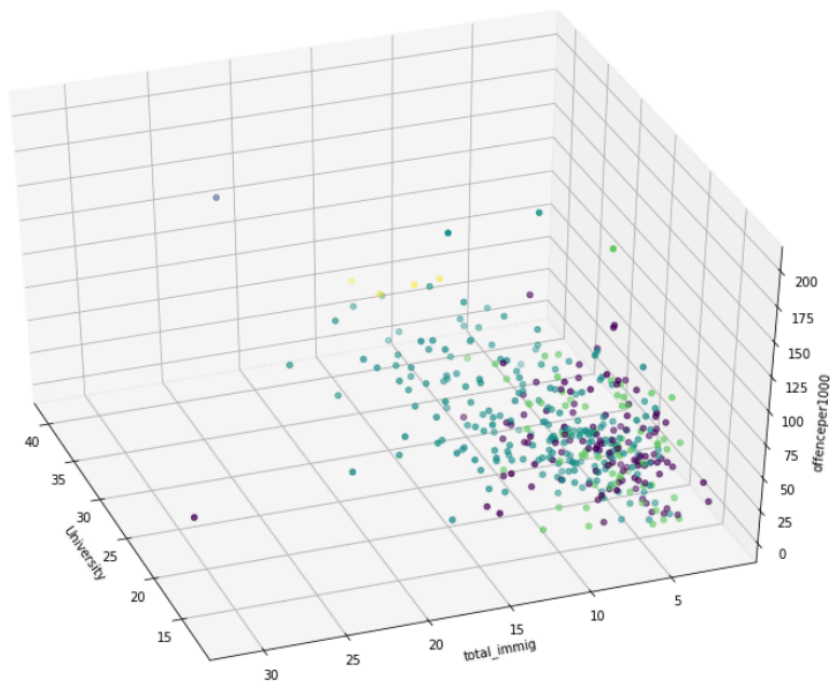
Initial observation on clustered data looks as this :



By observing mean value of population, we see Oslo as single most cluster with high average population 76% (Note: It doesnot mean 76% of total population lives in Oslo. It is mean population per cluster! Likewise, other values in pie chart are mean values per cluster, percentile applied). Population has played a main role in clustering. 20% is represented by a cluster containing four communes Bergen, Stavanger, Trondheim and Bærum. Remaining 3 clusters have very low mean population (2500, 3100 and 15000 respectively).

Urbanization is centered around Oslo, Bergen, Stavanger and Trondheim.

Other 3 indicators [education](#), [crime](#) and [immigration](#) looks proportionally distributed.



Scatter plot showing above 3 indicators with cluster labels (dot colour). We can immediately observe:

Except for one or two communes, offence per 1000 inhabitants is very low - just less than an average of 50 complaints over past 10 years. Based on it we can say Norway generally is a very safe country with low crime.

There is no correlation between total immigrants and crime rate. Instead police complaints gets lower as the immigrant population rate increases.

Looks like at least 20-25% of the total population tend to have university degrees. Ofcourse primary education in Norway is compulsory, so everyone would be literate. Those that donot have university degrees would still have vocational training or occupational education or simply they do not constitute adult population in this figure.

With one fourth of people with university degrees, we can call Norway an educated society.

As the immigrant rate increases we see university education rate increases aswell. Possible explanation would be that most immigrants to Norway might have university education or might acquire university education. Also, most immigrants would have come as a worker in some kind of occupations in oil, shipping, fishing and IT industries. They tend to have education as pre-requisite.

Role of education and immigration:

This made me ponder on a thought - The associated costs by government and the time to produce a graduate would be far more expensive for Norway than any other country. Calculating Net present value (NPV) on expenses over 20 years would amount to few

millions per graduate. By virtue of business activity and economic power Norway is able to get skilled, educated immigrants. This adds significantly not just to economic growth but also harvests social benefits associated with education.

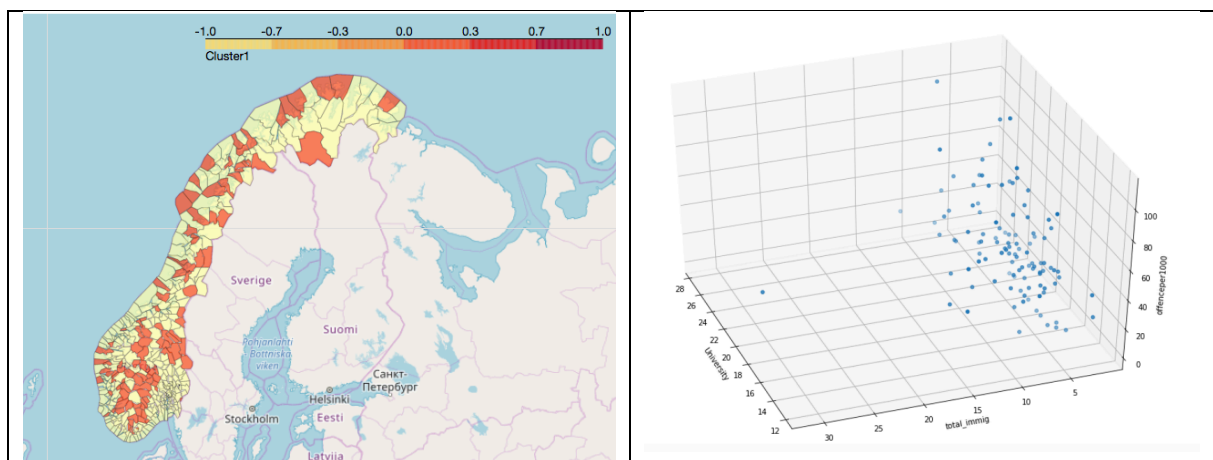
A country does invest lot of resources into education for several years, why not have a system where they train kids right from their childhood in a particular occupation? Reason is education creates lot of other indirect social benefits, for example, an educated person understands law and order, talks interestingly when you sit next in a bus, understands or resolves social problems quicker and so on. This is what Economists call indirect intangible social benefit - we cannot quantify them using physical measures but they add value to society.

There must be a portion of unskilled immigrants too, yet they tend to contribute to economy in long run, in a way that is more beneficial than the cost of acquiring or accommodating them.

Cluster analysis

If we analyze individual clusters:

Cluster 1:



Cluster 1 contains 112 communes

Mean population	2547
University graduates%	18
Immigrants %	6
Kommunes profitability %	3
Per capita income	127126
Debt per capita	74847
Offence per 1000 inhabitants (10yrs avg)	41

Cluster 2:

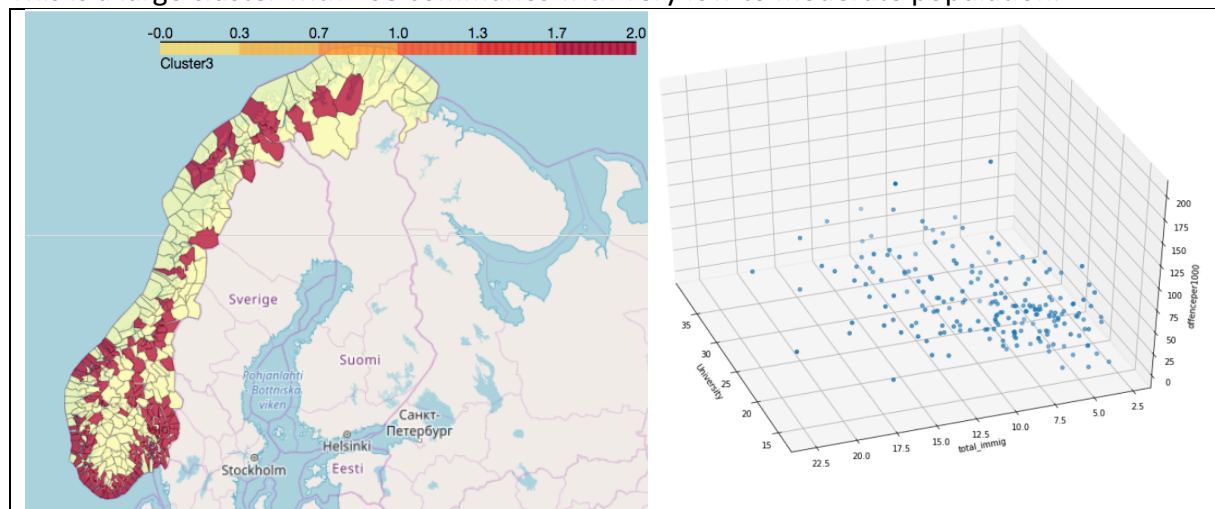
Cluster 2 contains only one commune Oslo. With its relatively large population it could well be an outlier with reference to other communes.

(Map and scatter not shown because it is only commune in the cluster)

Mean population	681067
University graduates%	41 (very high education rate)
Immigrants %	22 (largest immigrant population)
Kommunes profitability %	5
Per capita income	91624
Debt per capita	39969 (interesting: very low debt)
Offence per 1000 inhabitants (10yrs avg)	131

Cluster 3:

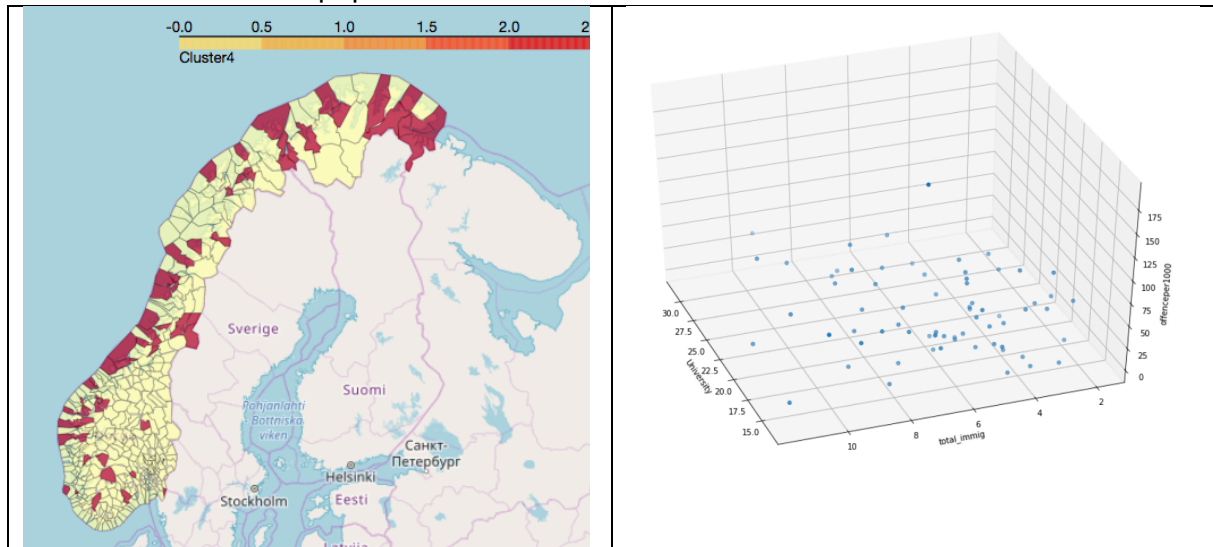
This is a large cluster with 208 communes with very low to moderate population.



Mean population	15251
University graduates%	21
Immigrants %	9
Kommunes profitability %	1 - lowest
Per capita income	86425
Debt per capita	74005
Offence per 1000 inhabitants (10yrs avg)	51

Cluster 4:

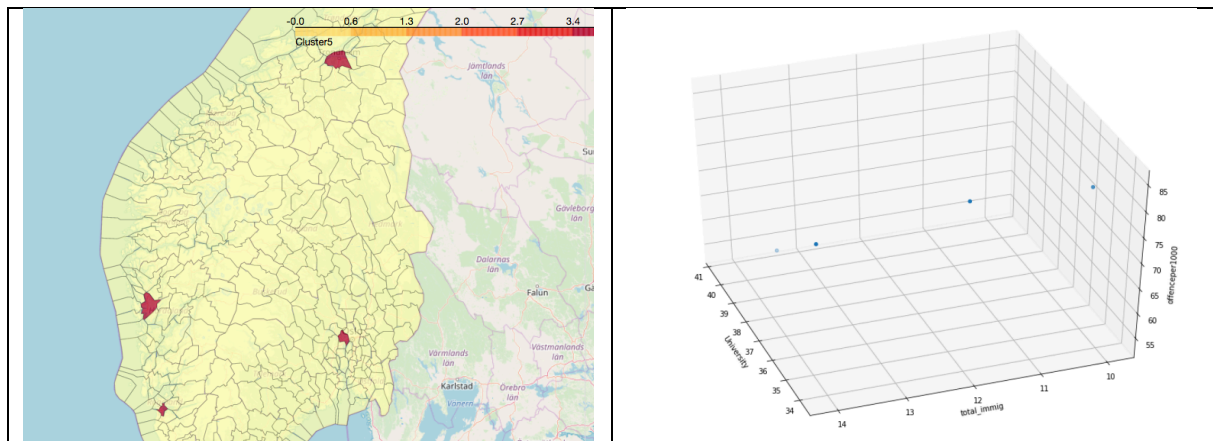
68 communes with low population



Mean population	3199
University graduates%	19
Immigrants %	6
Kommunes profitability %	5 - high profitability
Per capita income	130298
Debt per capita	137023 (interesting: debt is more than income)
Offence per 1000 inhabitants(10yrs avg)	39

Cluster 5:

4 communes with high population over a million with high profitability , high income and low debt.



Mean population	184557
University graduates%	36
Immigrants %	12
Kommunes profitability %	4 - high profitability
Per capita income	80488

Debt per capita	59694
Offence per 1000 inhabitants(10yrs avg)	75

More observations:

- 3 out 4 clusters are operating at profit more than the inflation rate which is 3 - 3.5%.
- Oslo has high education, immigration, income but very low debt!! Next high
- population cluster 5 with large immigrant population has low per capita debt. Would it be that immigrants tend to avail less loans? Something might be of interest to banking sector.
- Over 200 communes (50%) have low population and very low average profitability. Probably they are spending more on public services for low population. In economics they call it 'cross subsidization'.

Conclusion:

The input data set we have created has few more variables than those used. Some more indicators could also be gathered from other sources. However, we have used main indicators such that, based on our analysis, one can get a better understanding of overall socioeconomic conditions prevailing in Norway. It can also help decide which area or commune could one can choose by cross comparing their needs and expectations with facts generated in this report.

If I were to choose which commune to choose, I would choose cluster 2 (Oslo) or cluster 5 (Stavanger, Bergen, Bærum, Trondheim) , which already am living in .

Jupyter Notebook with codes are available at:

https://github.com/Loghse/Norway_kommune_macro_analysis/blob/master/Norway_kommunes.ipynb

Further considerations:

As mentioned I have kept my focus on education, immigration and crime rate when discussing with the help of plots and maps. Discussing other variables such as income, debt, expenditure with visual tools could have been also useful.

The crime data used in this analysis is total offences registered at police. It could range from bicycle theft, driving offence, domestic violence to as high as break in theft and attempt to murder. There is a whole range of offences which could have been explored individually, but would be an extensive exercise.

I have omitted 44 communes due to lack of completeness in their data. They are smaller communes and I can only assume with minimal impact over the results.

Country wise immigrant population with top10 countries Sweden, Lithuania, Poland, Germany, Eritrea, Iraq, Somalia, Philippines, Pakistan and Vietnam were available. Total of these figures were used as cluster features, instead of each of them. In theory, population with immigrant population from each country and how they interact with other variables could be explored. I chose not to perform such an exercise out of ethical considerations, mainly not to hurt sentiments of any particular group. The development of a nation, its economy and well-being of its inhabitants is a collective group work. Irrespective of what origin one comes from, any law-abiding citizen engaged in meaningful occupation should be respected by the society and protected by law. Norwegian society is good at it and should be respected for doing that.

While no analysis could be 100% accurate, this analysis might also have some errors, bias or deviations. I have done my best to keep them minimum.

Anyone who wish to do exploratory analysis in this area, might find my codes and methods useful. So make the codes and data available at my github repository.