

Analysis of the Swedish communes business development potential.

Dmitry Zakharov

March 2020

Introduction.

In this work we are interested to estimate some aspects of the investment potential for Swedish communes.

For the potential entrepreneurs thinking where to start some sort of a business in Sweden it is important to understand what geographical locations have more potential for the development.

The question is important as there is no clear answer to the questions like: which communes (cities) are better for investing: “rich” or “poor”, “big” or “small”?

It is obvious that there are rich and active communes (like capital cities), but they might be oversaturated with businesses which results in high competition and, hence, low profitability of new businesses.

On the other hand, “small” and “quiet” communes might be missing small businesses and, even though level of economical activities might be low there, the competition level is also significantly lower than in big cities.

We will try to analyze the level of “saturation” of Swedish communes with businesses and try to segment them according to the perspectives of new business development.

The question is important for those potential entrepreneurs that are thinking about the best locations to start a new business in Sweden.

Data that will be used to solve the problem.

To analyze the problem we will start with the following data sources:

a. SCB site (Swedish statistics bureau) <http://www.statistikdatabasen.scb.se/pxweb/en/ssd/> where it is possible to find the data on population, income and a lot of other statistics for Sweden. In particular, we are interested in the data on population and average income for all Swedish communes. We downloaded this data from SCB site and put it here for further processing: https://github.com/Dmitry-Za/Coursera_IBM_DataScience_Final/blob/master/se%20medelinkomst%20population.xlsx

b. We are also interested to have geographical coordinates for all Swedish communes and the squares of their territories. There is a lot of ready files with this data, in particular we found this file that contains all the necessary data and can be used for our purposes: <https://raw.githubusercontent.com/peterdalle/svensktext/master/platser/kommuner.csv>

c. Foursquare service that we will use to measure the quantity of businesses around the communal center cities.

Methodology.

In the scope of our project, we will read the required data, remove unnecessary information and combine all the data into one dataframe convenient for further processing.

We assume that important for us will be the following data:

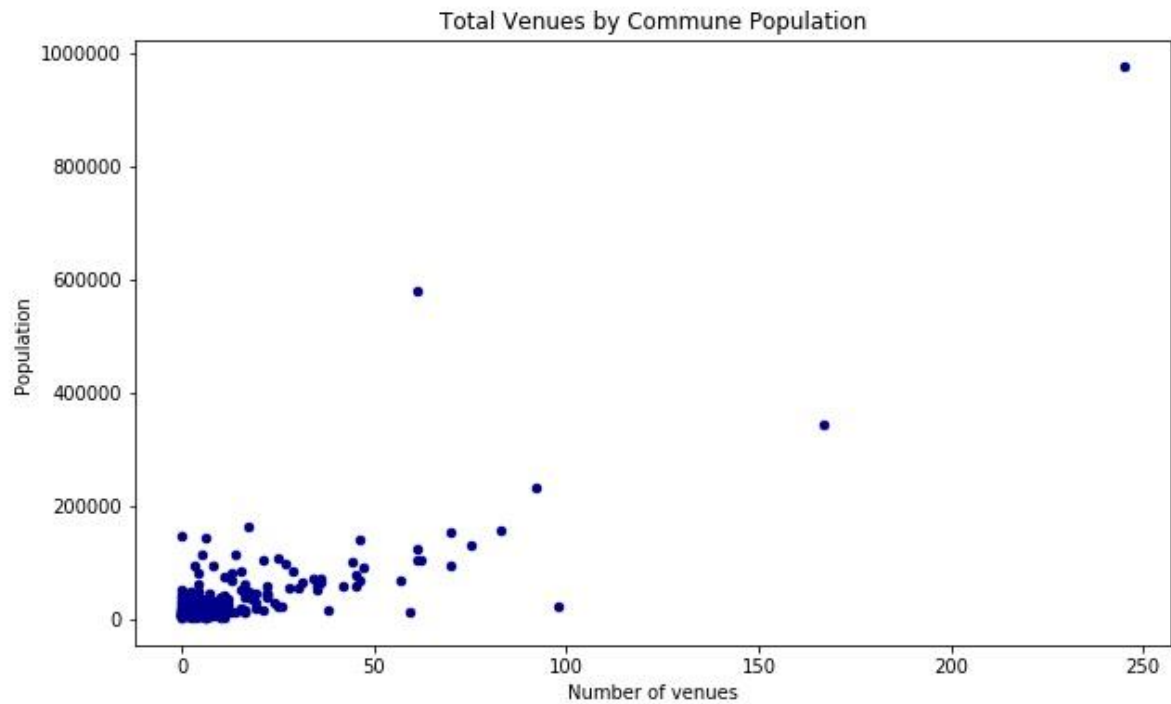
1. Commune name
2. Average income in the commune
3. Commune population
4. Commune area
5. Commune central city latitude
6. Commune central city longitude

Having this data, we can show all the communes on the map:



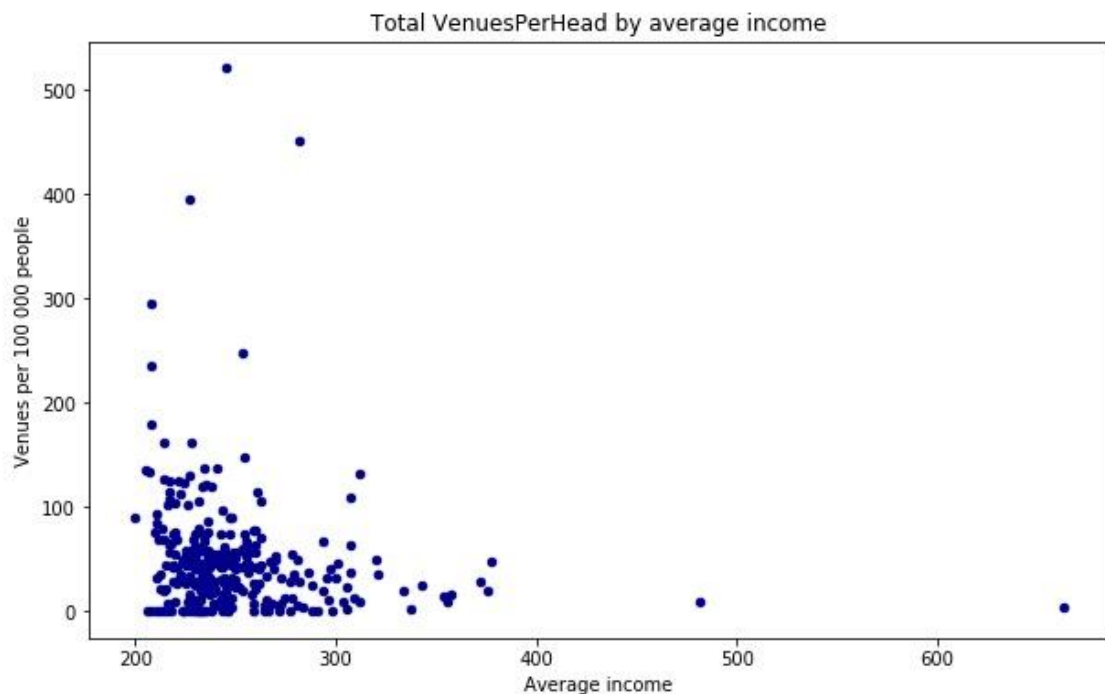
For each commune we will explore the area around the central city using the Foursquare, extract the info on Total quantity of venues in the area and add it to our dataframe. We will assume that this information reflect the quantity of service businesses existing in the area.

Then we will plot the quantity of venues against the commune population to check that the result looks reasonable assuming that higher population requires higher quantity of venues.



Since we are interested in the business potential of the communes, we will calculate saturation of communes with businesses, i.e. quantity of venues for 100 000 of population.

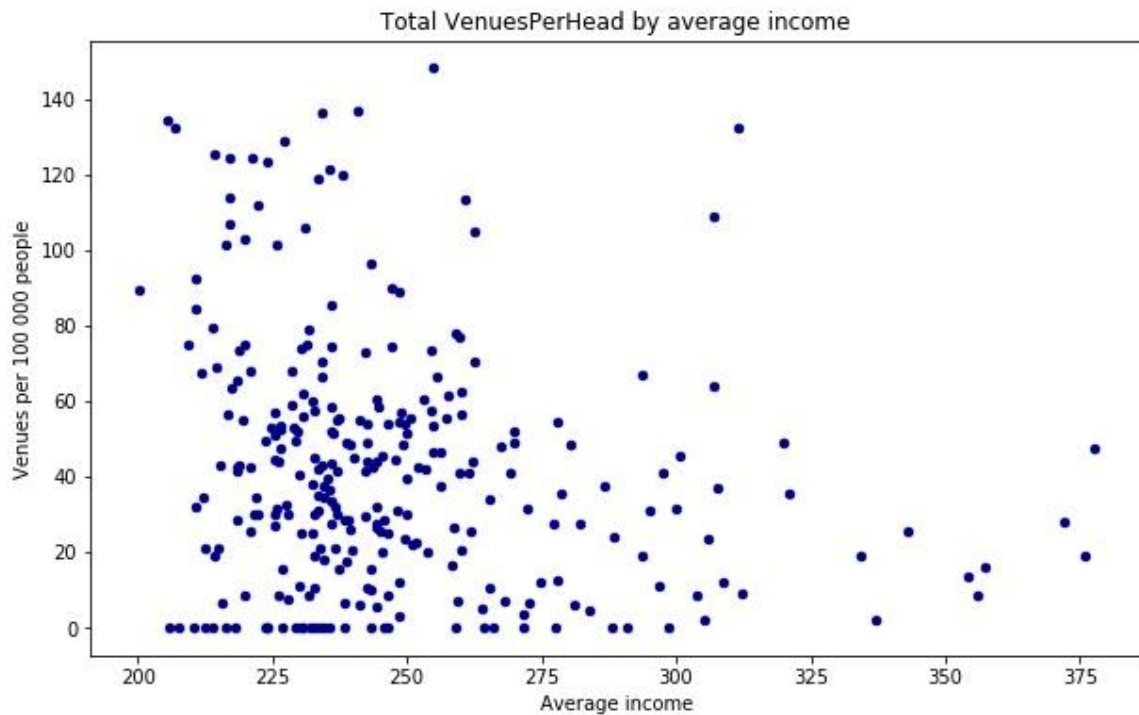
We will also assume that higher average income in the commune provides higher potential for the business success. To analyze this further, we will scatterplot Average Income vs Saturation with businesses.



We will assume that undersaturated communes with high average income are the best for new business development.

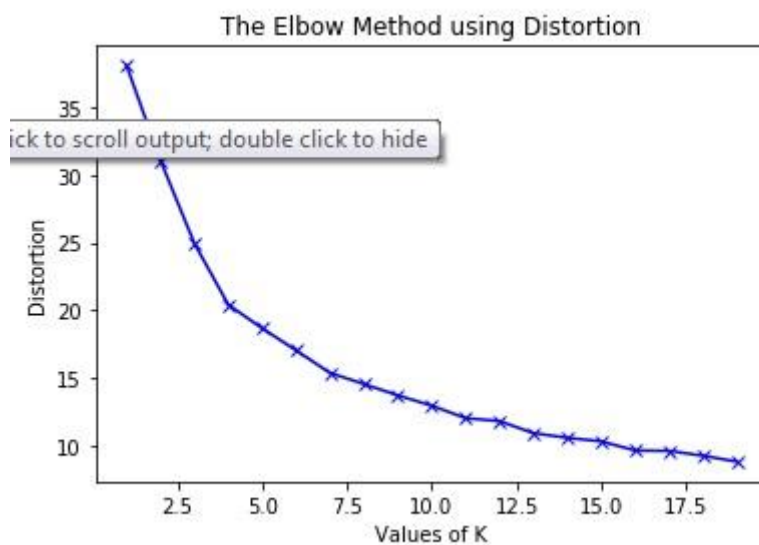
In the drawing above we can see a number of definitely oversaturated communes with comparatively low income that we can exclude from our analysis (those with >150 venues for 100 000 population). We can also see 2 extremely high income communes with very low saturation. Investigation shows that these 2 communes are the residence areas for the very rich people who obviously do not need any businesses in their living quarters, so we exclude these 2 communes also.

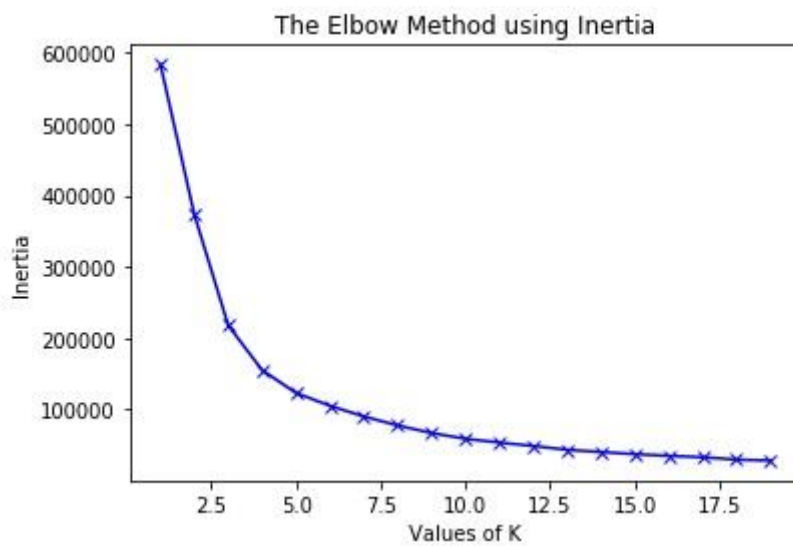
The rest we will analyze:



We will divide these communes into a number of clusters to choose the best for our purposes.

To determine the number of clusters we will use Elbow method using both distortion and inertia.



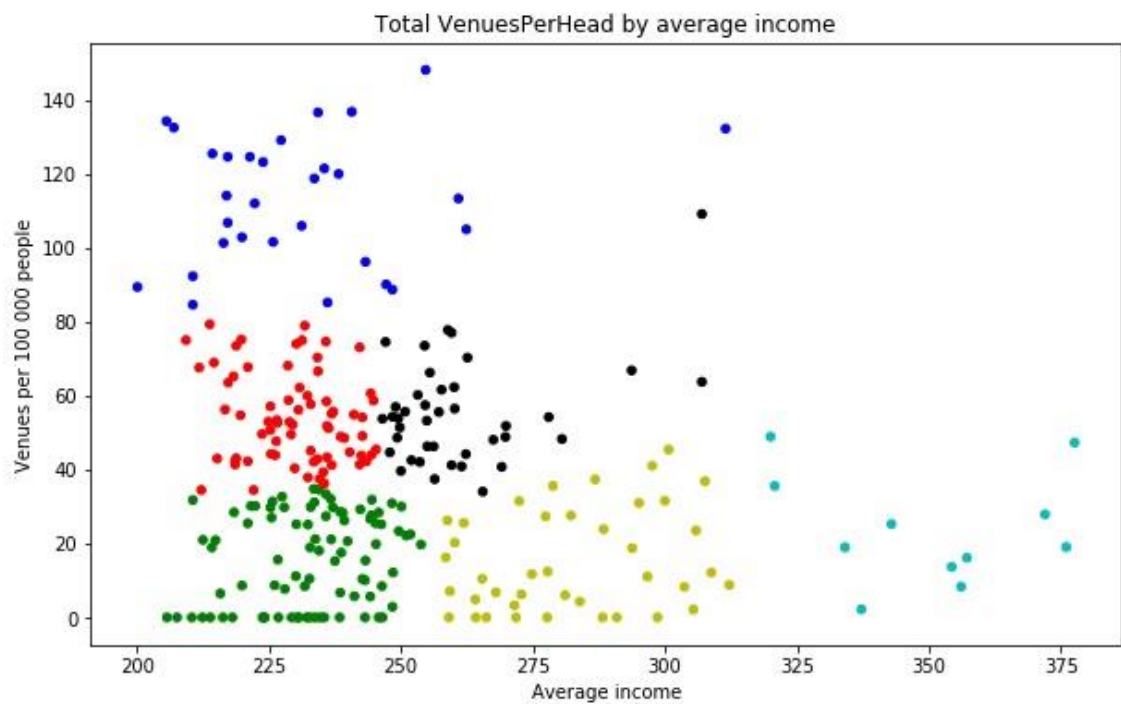


Based on the above, we will set the number of clusters at 6.

After using k-means method to cluster, we can show clustered communes on the map:



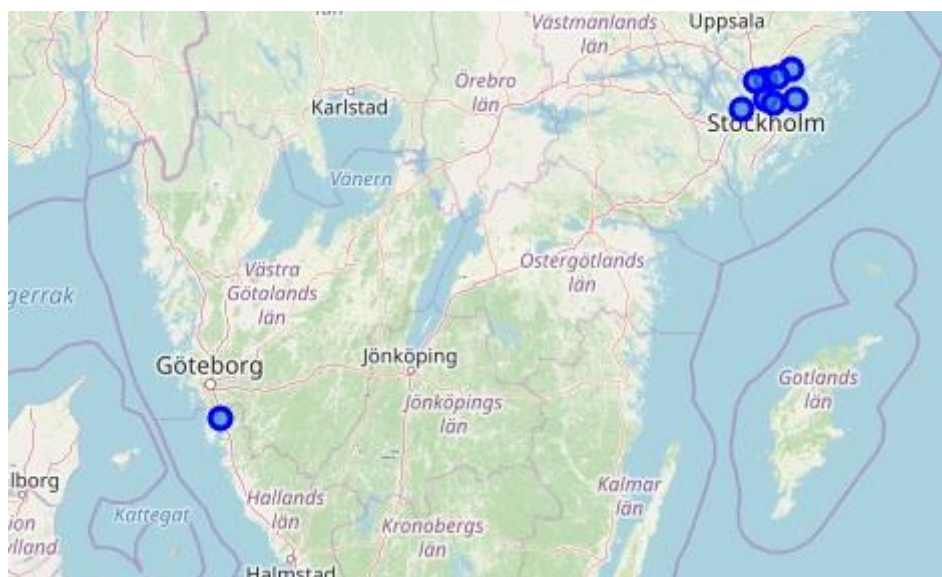
And we can also see the results of clustering on Saturation/Average Income plot:



Results.

Obviously, the right-bottom cluster with the highest average income and low saturation is the most promising according to our analysis methodology and can be recommended for investments into new businesses.

On the map the communes from this cluster look as follows:



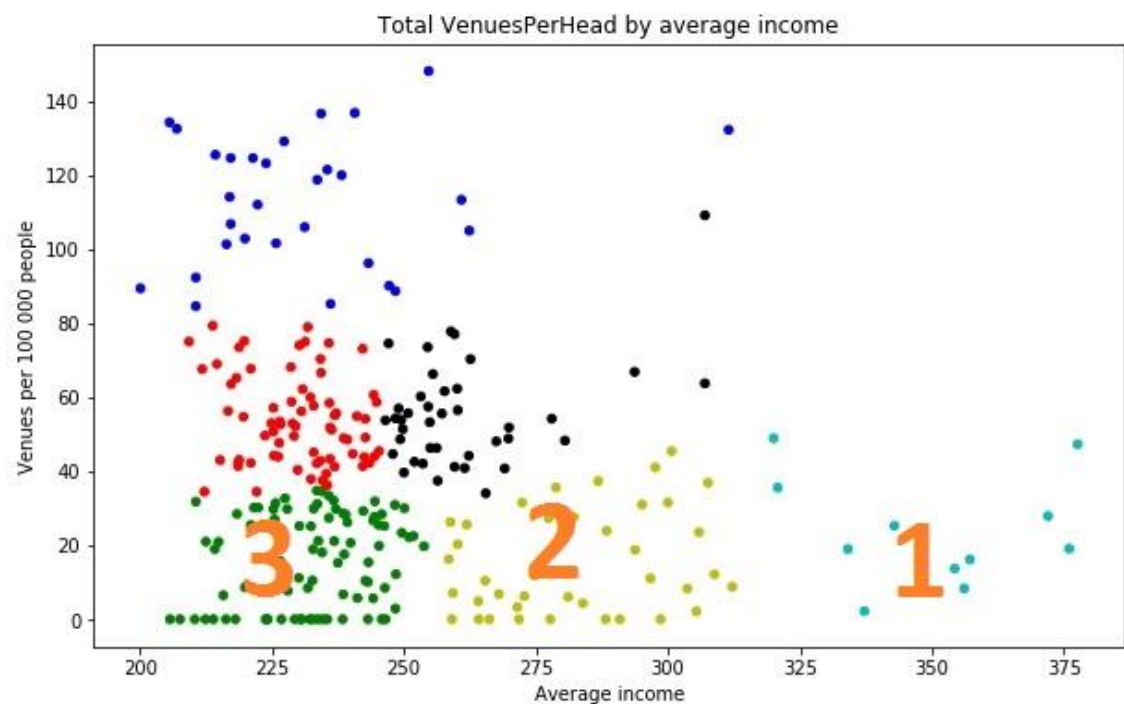
The full list of these communes is here:


```
In [46]: df_se_filt2.loc[df_se_filt2['Cluster Labels'] == 5]
```

```
Out[46]:
```

	Cluster Labels	Code	Av_income	Population	name	area	lat	lon	NumVen	VenPer1000
2	5	117	337.2	45574	Österåkers kommun	554.65	59.500058	18.352485	1	2.194234
3	5	120	320.0	45000	Värmdö kommun	2980.99	59.333333	18.383333	22	48.888889
5	5	125	372.0	28690	Ekerö kommun	384.53	59.279834	17.790225	8	27.884280
13	5	160	377.6	71874	Täby kommun	71.22	59.441900	18.070330	34	47.305006
15	5	163	334.1	73857	Sollentuna kommun	57.96	59.439110	17.941480	14	18.955549
16	5	180	342.9	974073	Stockholms kommun	214.12	59.329324	18.068581	246	25.254781
18	5	182	376.0	105189	Nacka kommun	128.46	59.307903	18.156042	20	19.013395
22	5	187	356.1	12003	Vaxholms kommun	106.85	59.452788	18.183603	1	8.331251
98	5	1233	354.3	36628	Vellinge kommun	705.62	55.470893	13.019990	5	13.650759
103	5	1262	357.2	24834	Lomma kommun	90.20	55.670667	13.077576	4	16.106950
133	5	1384	320.8	84395	Kungsbacka kommun	1472.93	57.503556	12.082334	30	35.547130

Discussion section.



It is interesting to note that although our recommended cluster #1 is the best according to our analysis methodology, there are some questions and observations that could be addressed in the scope of future and deeper analysis.

1. It would be good to enable additional processing of venue categories in order to:
 - a. Exclude the categories that might be irrelevant to our analysis like airports, busstations, stadiums etc.
 - b. To make it possible to run the analysis in the specified field of interest (i.e. to evaluate the business potential for opening a new restaurant, a new hotel, a new clothes boutique etc. rather than looking at all businesses taken together.)

2. It would be interesting to look in more detail into clusters #2 and #3 and compare them to cluster #1. The question of why cluster #1 has so low saturation with businesses is important. It might be that there exist some external obstacles to opening new businesses in these communes, in this case cluster #2 or even #3 might appear to be of higher interest.
3. It is interesting to analyze why the communes with the highest business saturation lie mostly in the low-income area.
4. It is interesting to note that the communes from cluster #1 lie exclusively in the south of Sweden. It might be promising to introduce some geographical distribution parameters into our model.

Conclusion.

In this project we analyzed the potential of Swedish communes for the new business development.

We gathered data on location, population and average income of the communes. Using Foursquare, we added information on the quantity of venues in the surroundings.

After processing the data, we scatterplotted it with Average Income vs Business saturation, clustered the results into 6 clusters (based on Elbow analysis) and selected the best cluster with lowest business saturation and the highest income.

The communes from this cluster we can recommend as the most investment-attractive for starting new businesses.

The results were presented on the geographical map and as a list of commune names.

We also noted some potential areas for further investigation and for the further development of our model.