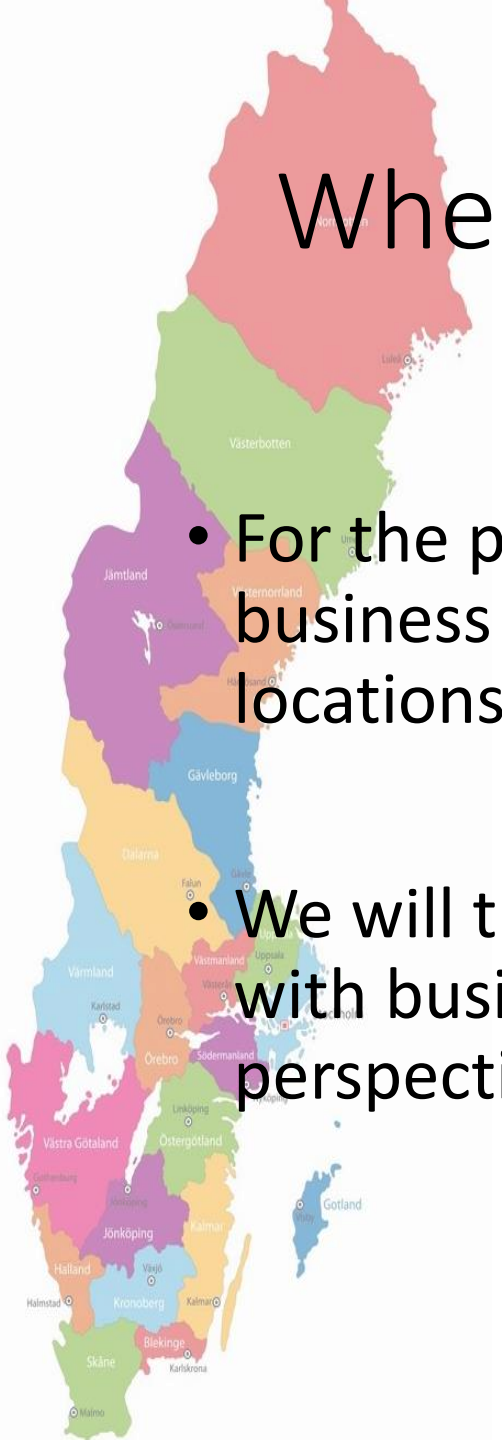
Analysis of the Swedish communes business development potential.

Dmitry Zakharov

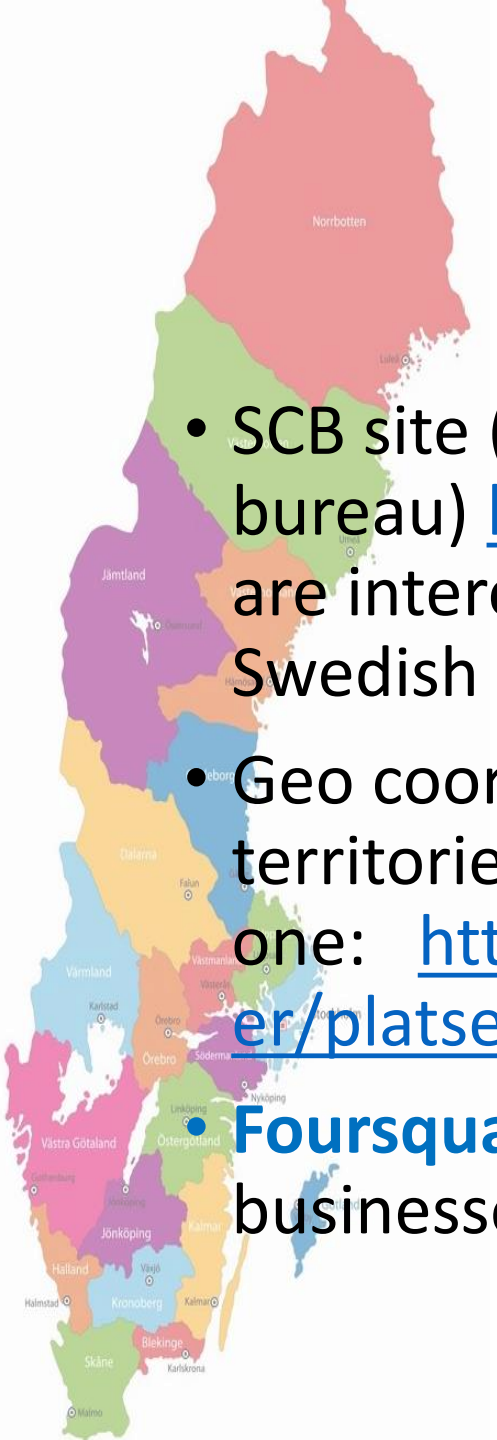
March 2020

Where to start a new business in Sweden?



- 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.

- 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.

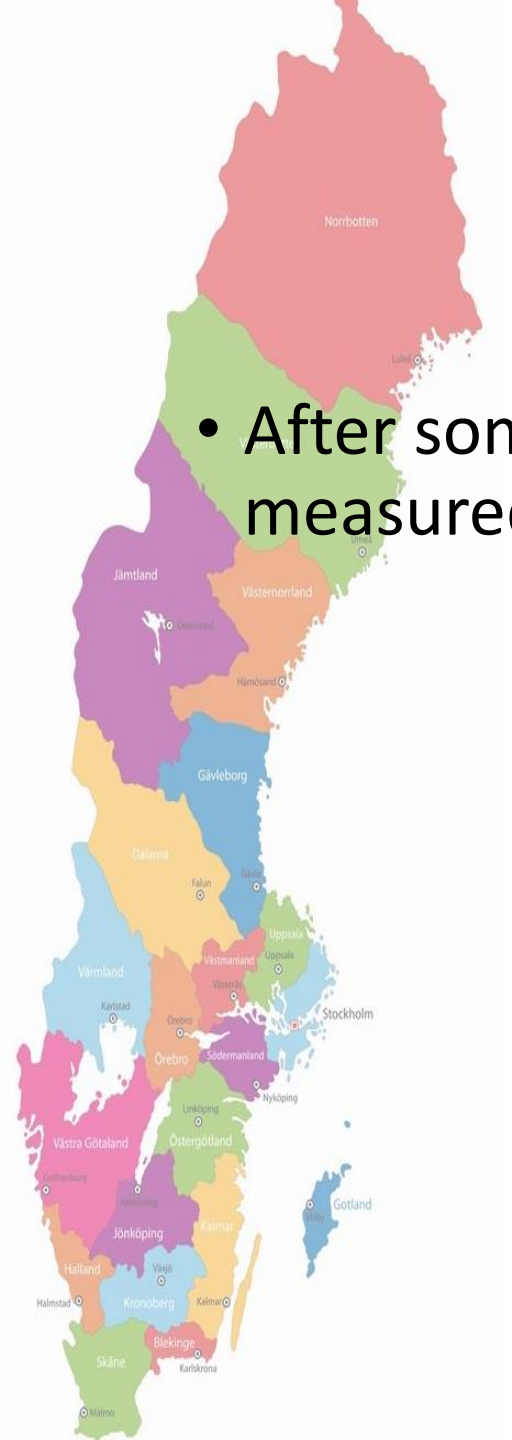


Data to be used

- SCB site (Swedish statistics bureau) http://www.statistikdatabasen.scb.se/pxweb/en/ssd/ . We are interested in the data on population and average income for all Swedish communes.
- Geo coordinates for all Swedish communes and the squares of their territories. There is a lot of ready files with this data, we chose this one: https://raw.githubusercontent.com/peterdalle/svensktext/master/platser/kommuner.csv
- **Foursquare** service that we will use to measure the quantity of businesses around the communal center cities.

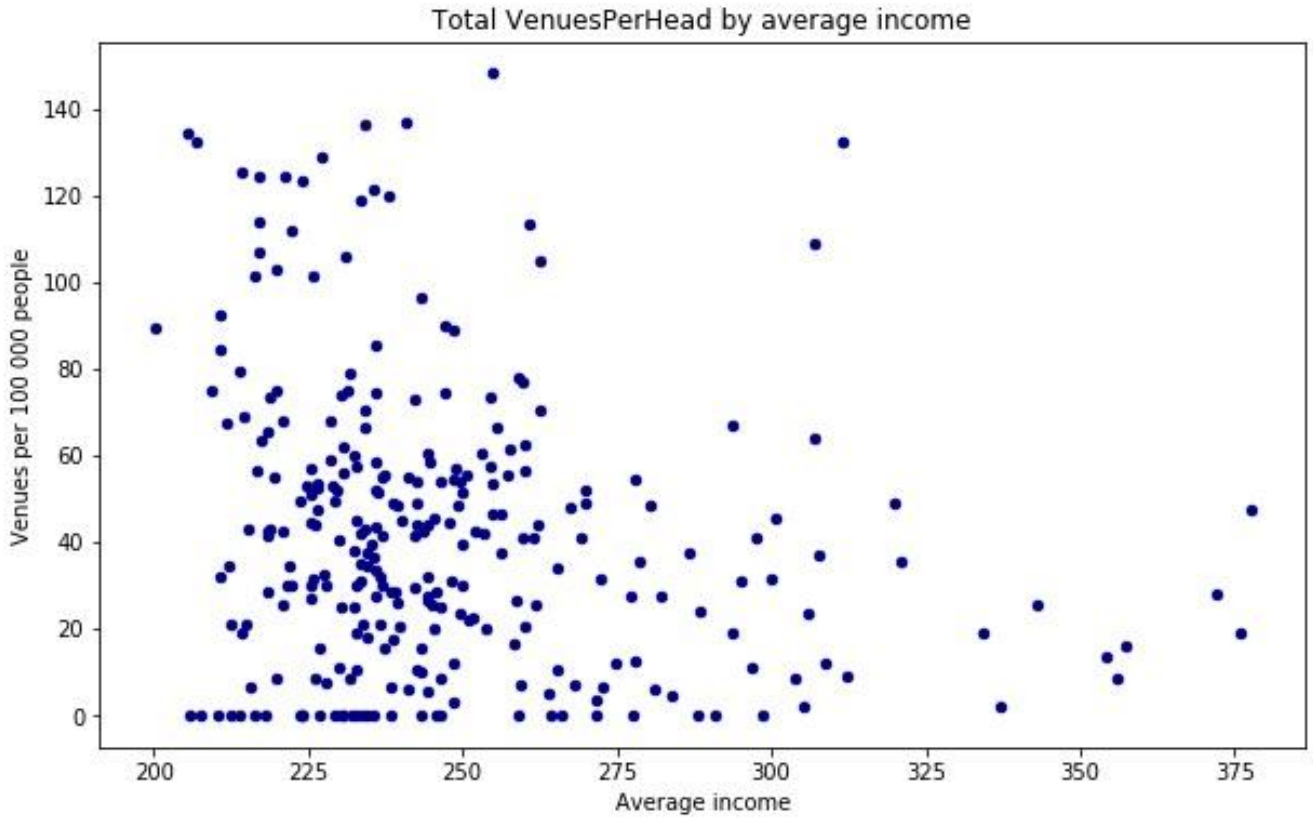
Methodology.

- We assume that important for us will be the following data:
  - Commune name
  - Average income in the commune
  - Commune population
  - Commune area
  - Commune central city latitude
  - Commune central city longtitude
- To this we add info on quantity of businesses from **Foursquare**
- We will calculate communes saturation with businesses (venues per person) and assume that highest business potential have communes with lowest business saturation and highest average income.



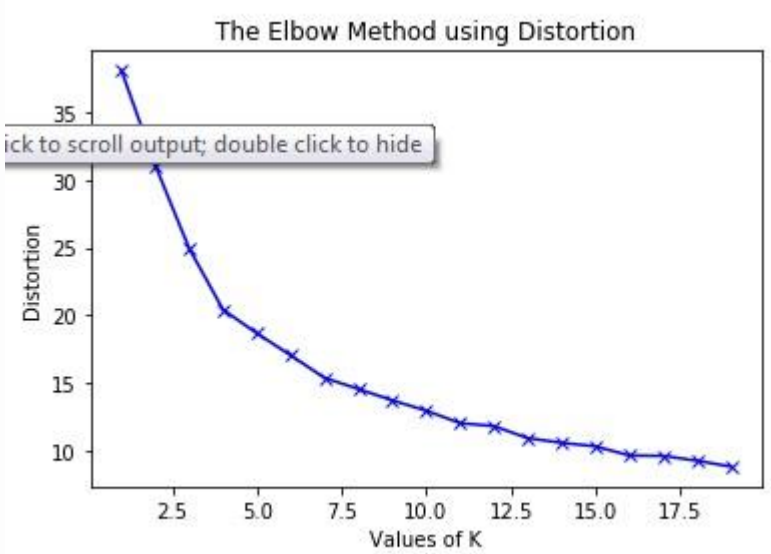
Data for analysis.

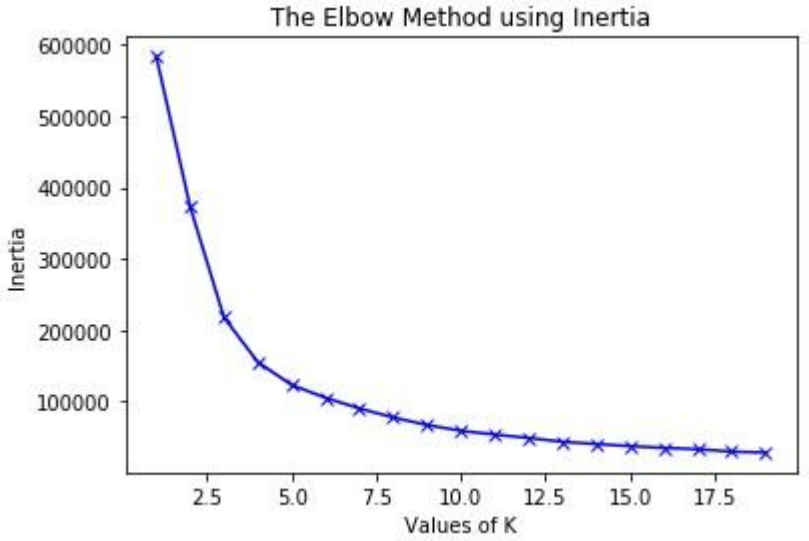
- After some data preprocessing, we can plot all Swedish communes as measured by their business saturation vs average income.



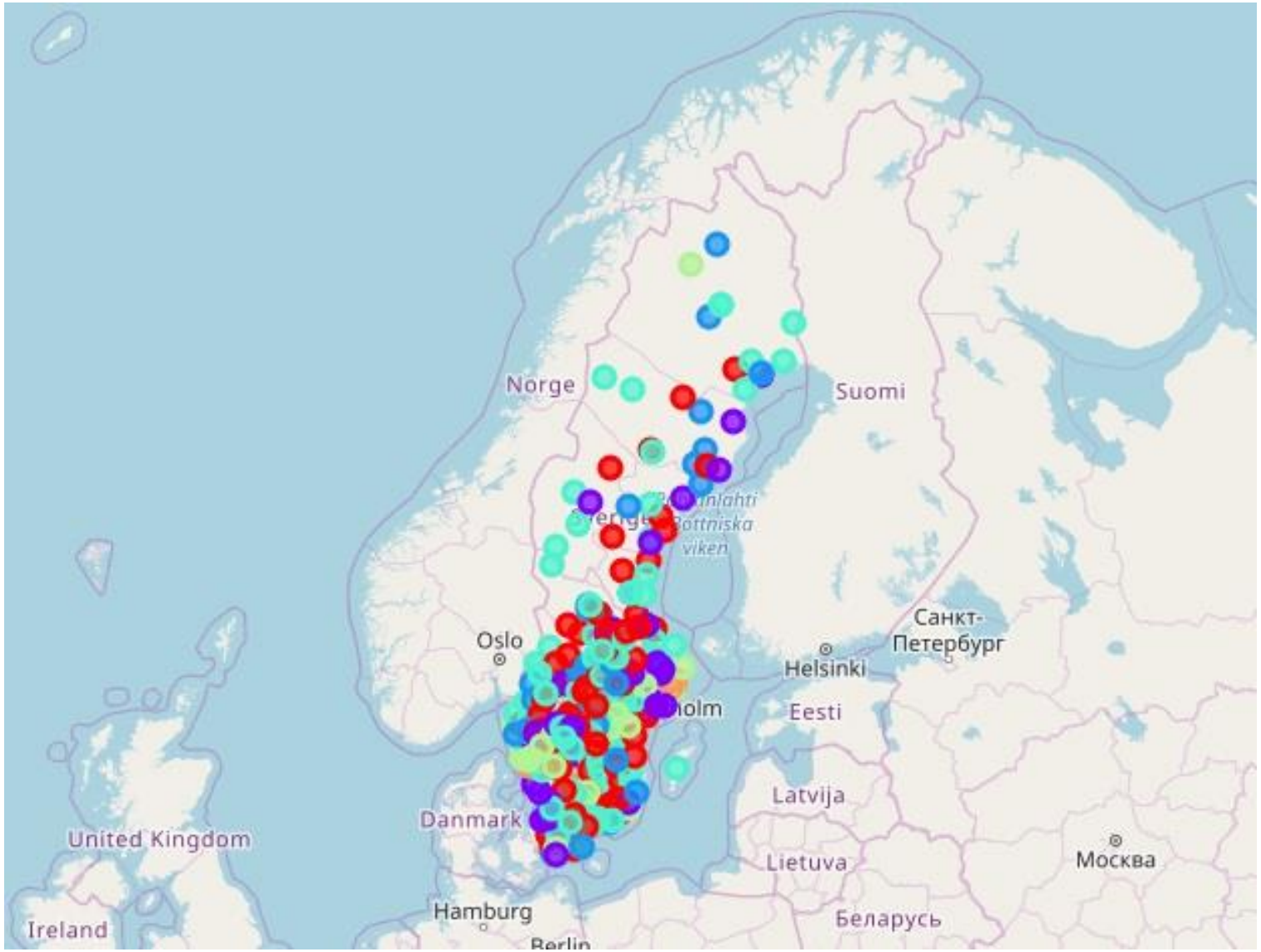
K-means clustering: how many clusters?

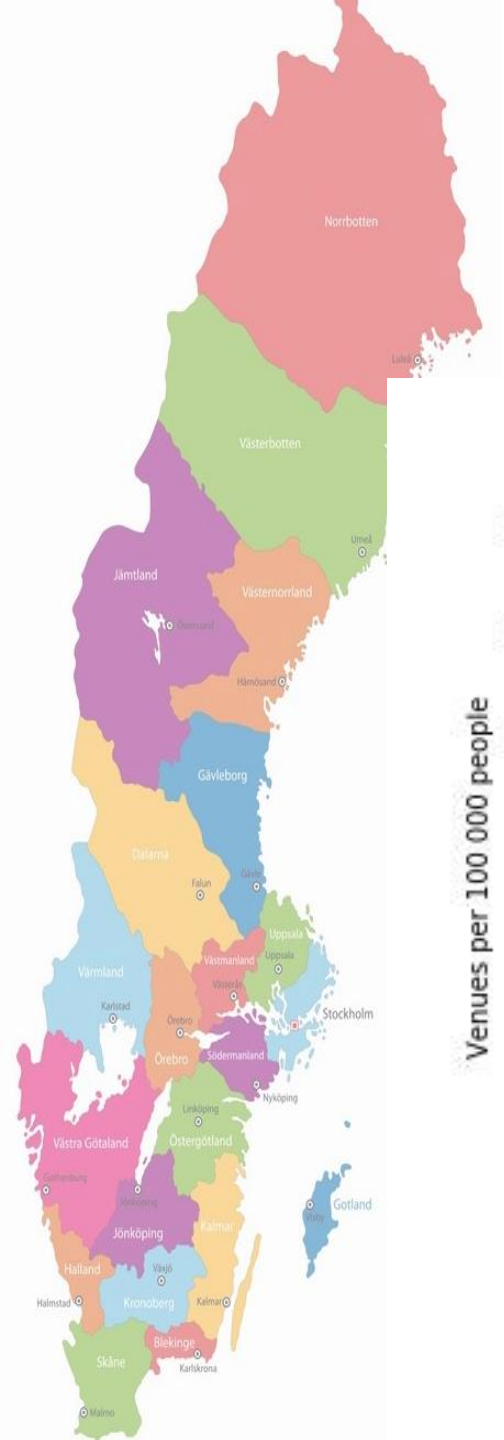
- Using Elbow method, we can see that the reasonable quantity of clusters lies in 4-6 range. We choose 6.



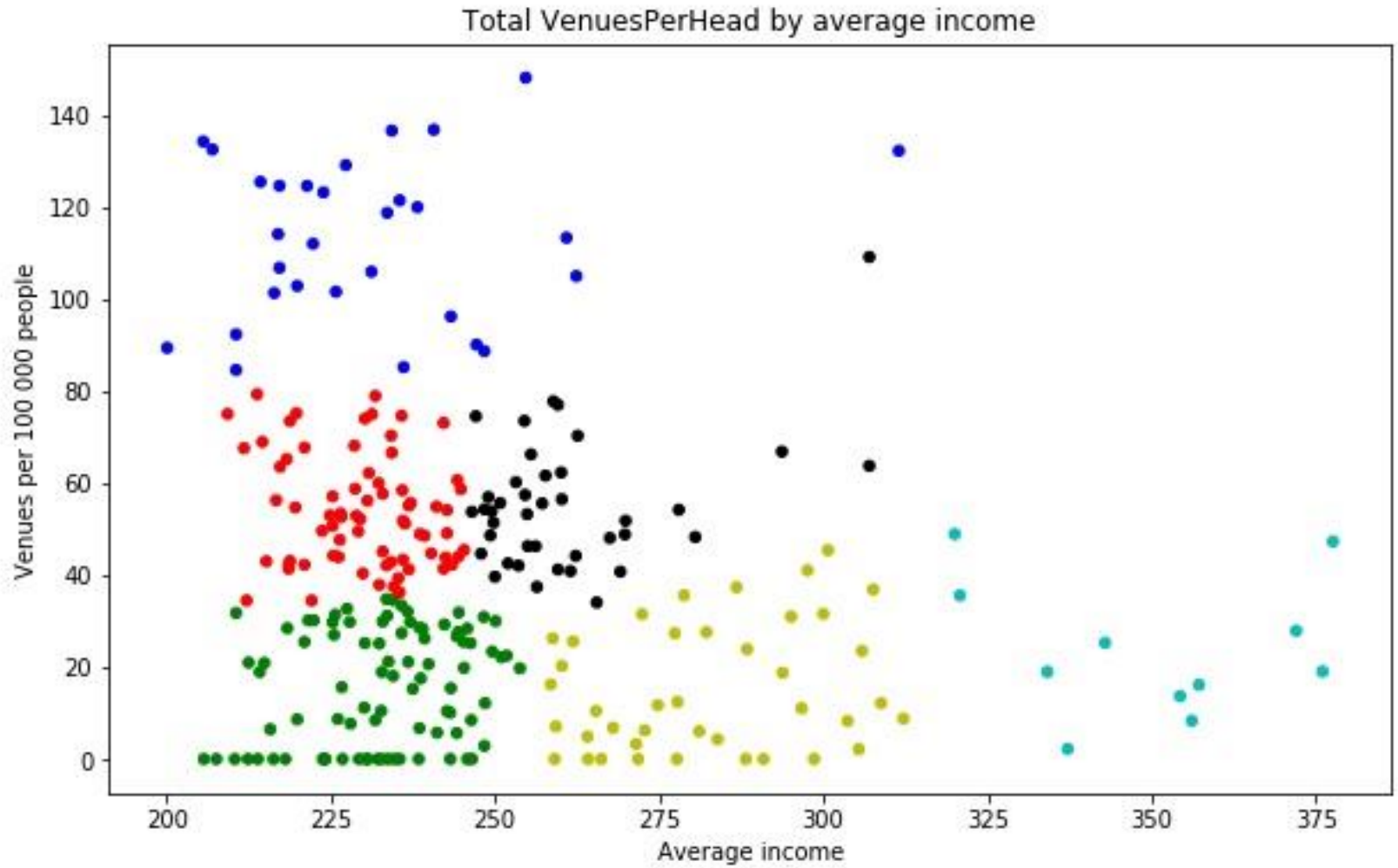


Clustered communes on the map.

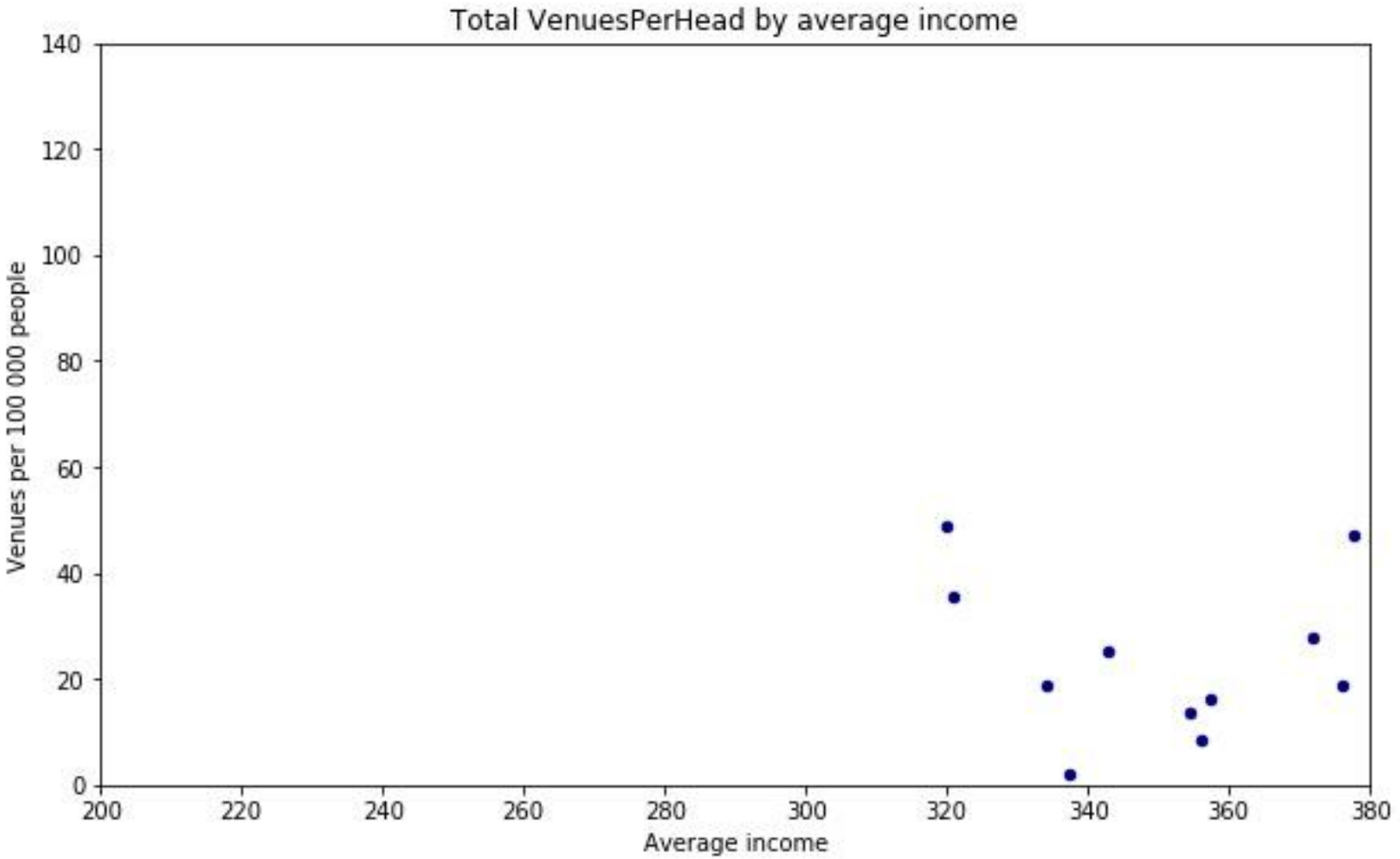


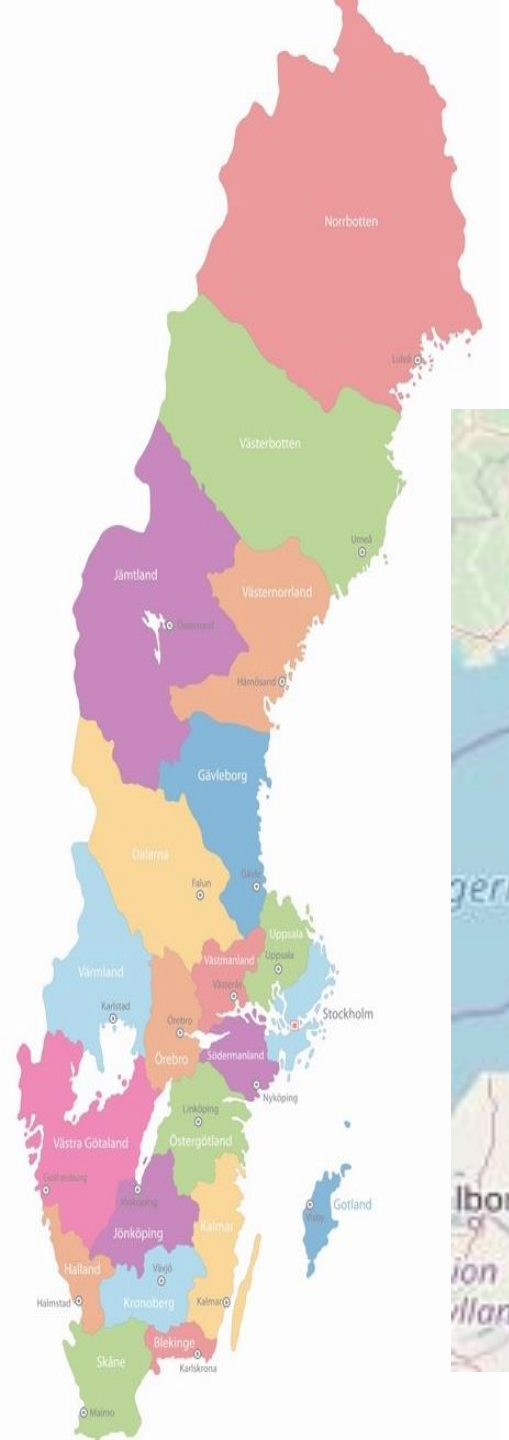


Clustered communes on Saturation/Average Income plot

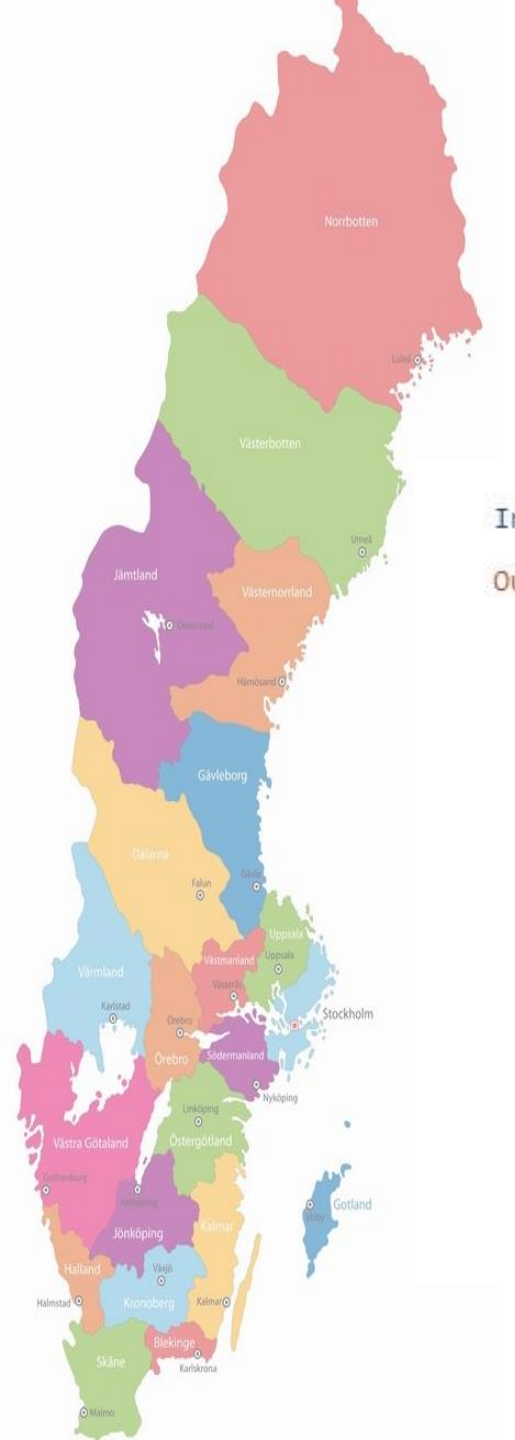


Our cluster of choice: highest income, lowest business saturation





Cluster of choice on the map.



The list of communes with highest business potential

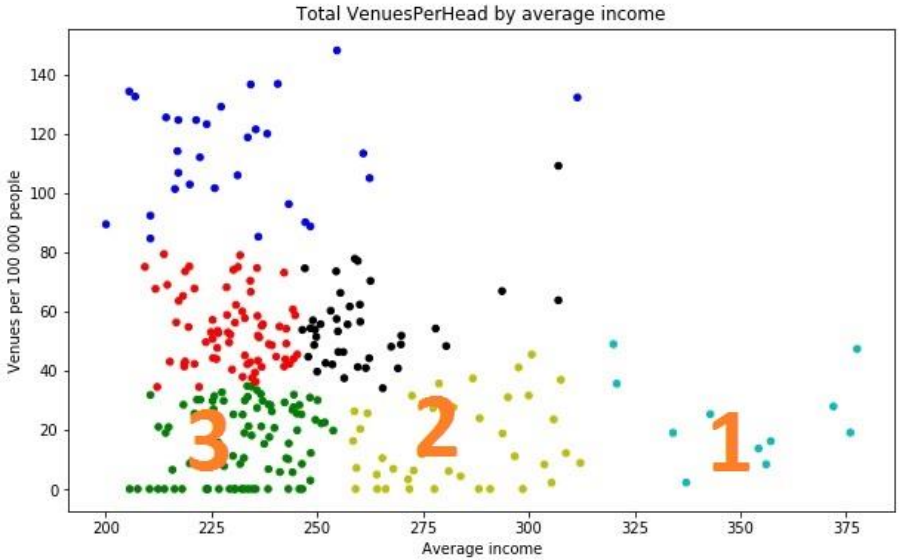
```
In [46]: df_se_filt2.loc[df_se_filt2['Cluster Labels'] == 5]
```

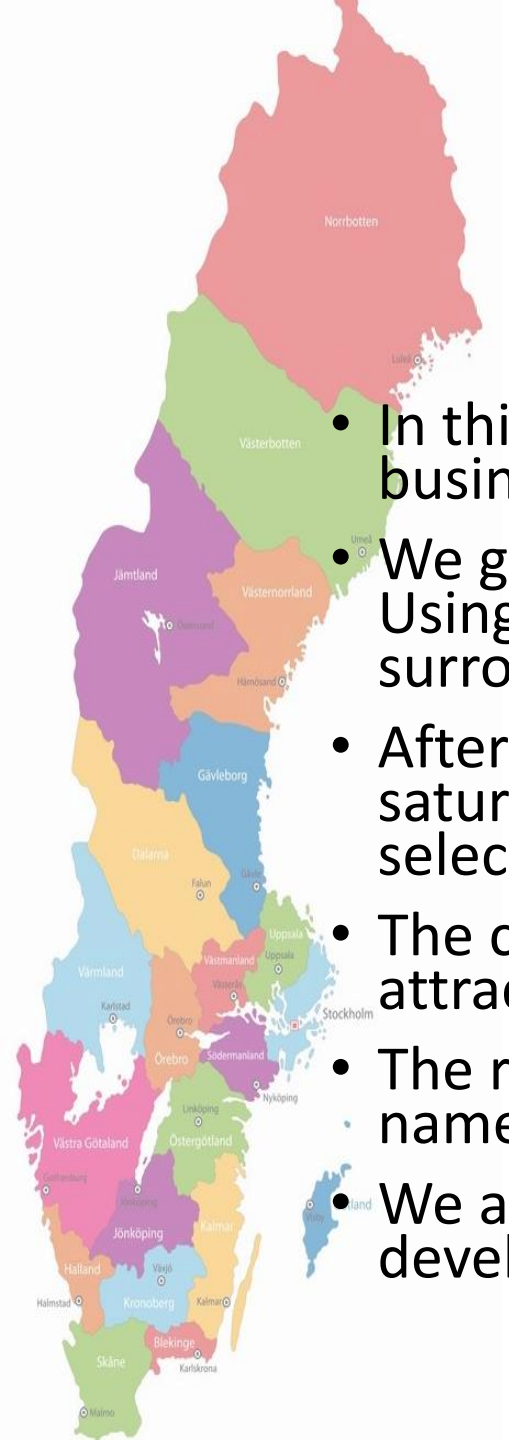
Out[46]:

	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

Some questions and future directions.

- It would be good to enable additional processing of venue categories in order to:
  - Exclude the categories that might be irrelevant to our analysis like airports, busstations, stadiums etc.
  - To make it possible to run the analysis in the specified field of interest (i.e. restaurant, hotel, etc.)
- 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 is interesting to analyze why the communes with the highest business saturation lie mostly in the low-income area.
- 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.