Analysis of immigrant population and small businesses

June, 2019

Applied Data Science Capstone

by Alizamin Jafarli

Problem

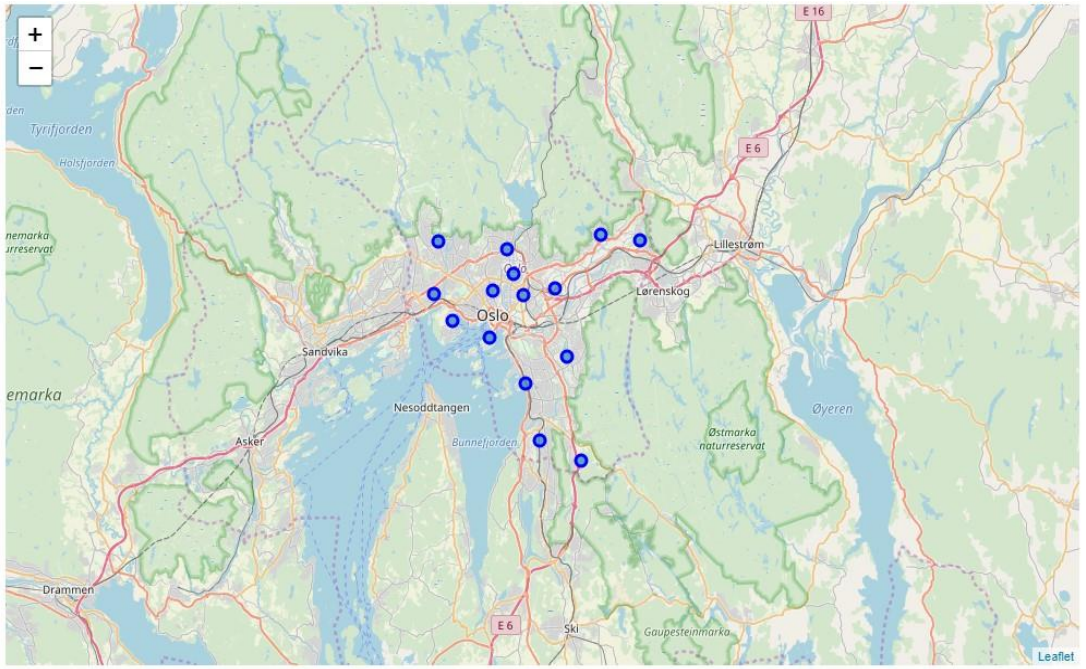
- The book “Statup Migrants” by Nicolai Strøm-Olsen and Maria Amelie
- Do immigrants contribute to small businesses?
- Locational analysis of immigrant population and venues in Oslo boroughs – any pattern?

Data, numbers taken from:

- Wikipedia -> List over Oslo boroughs
- GeoPy -> Latitude and longitude coordnates of boroughs
- Statistics Bank of Oslo Municipality -> Immigrant population
- Norwegian Statistics Bureau -> Social security receivers
- Carto.com -> json coordinates for boroughs
- Foursquare API -> Venue data

Data

	Borough	Population	AreaInKm2	Latitude	Longitude	ImmigrantsTotal	NumberVenues
0	Alna	49 358	137	59.929854	10.817046	26723	7
1	Bjerke	31 973	77	59.941395	10.829208	14405	10
2	Frogner	58 283	83	59.909640	10.687961	16498	5
3	Gamle Oslo	54 575	75	59.899237	10.734767	21839	9
4	Grorud	27 525	82	59.962343	10.875290	14011	5
5	Grünerløkka	58 906	48	59.925471	10.777421	21141	14
6	Nordre Aker	50 724	136	59.953638	10.756412	9518	7
7	Nordstrand	51 169	169	59.870880	10.780353	9201	6
8	Sagene	43 131	31	59.938273	10.765849	11332	27
9	St. Hanshaugen	38 109	36	59.927950	10.738958	10666	30
10	Stovner	32 850	82	59.959292	10.924499	19021	4
11	Søndre Nordstrand	38 925	184	59.835944	10.798496	21779	7
12	Ullern	33 463	94	59.925818	10.665132	6717	4
13	Vestre Aker	48 605	166	59.958300	10.670319	8726	5
14	Østensjø	49 968	122	59.887563	10.832748	13106	1



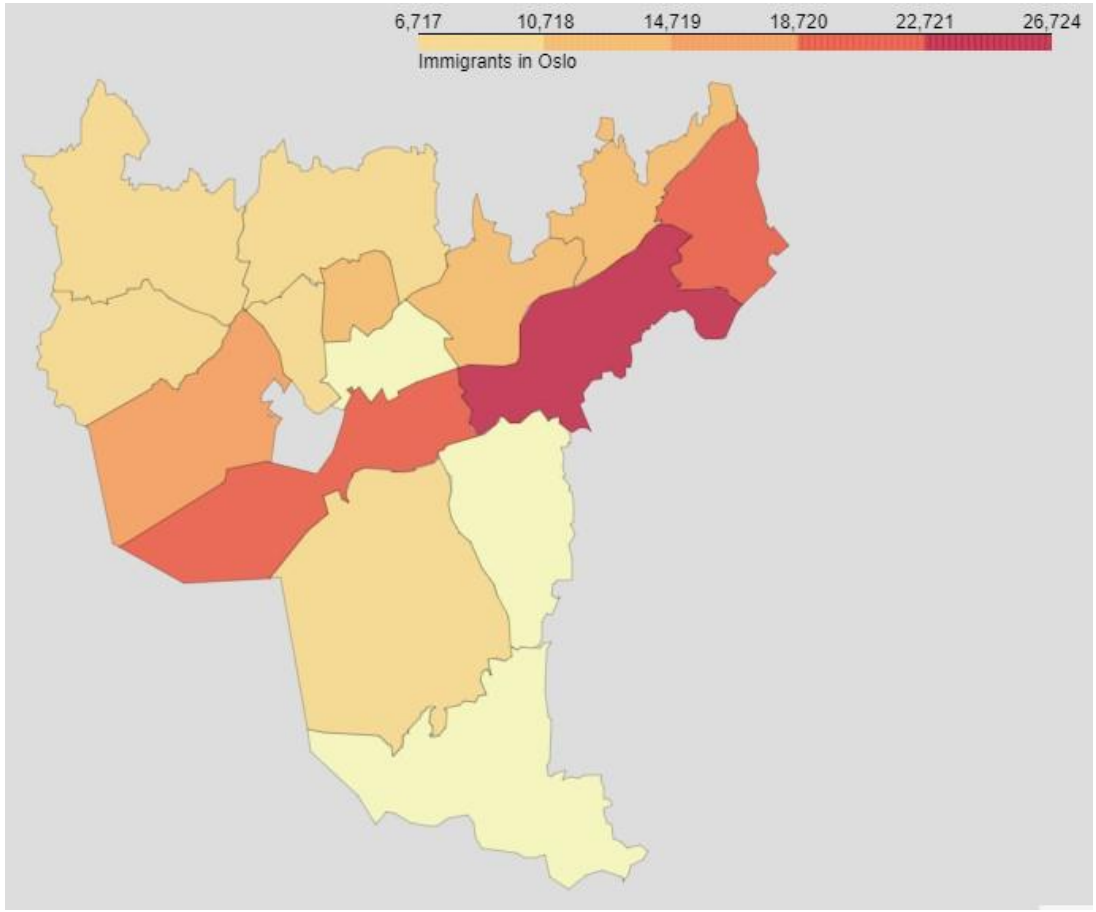
Methodology

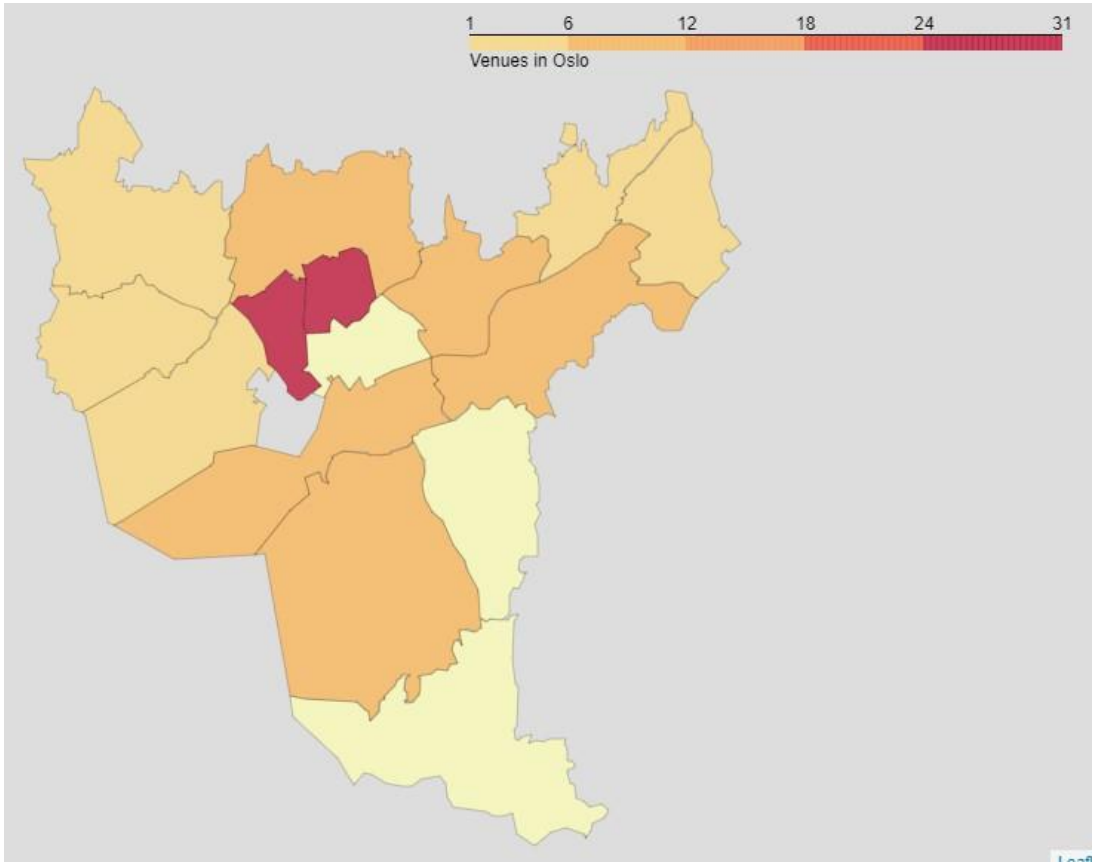
- Visualizations (histograms, folium maps, choropleth)
- Unsupervised machine learning (clustering)

- Notice:
  - Most small businesses started by immigrants are usually not present online, especially on Foursquare
  - Among the venues that Foursquare returns, there are many common outside areas like parks, bus / tram / railway stations

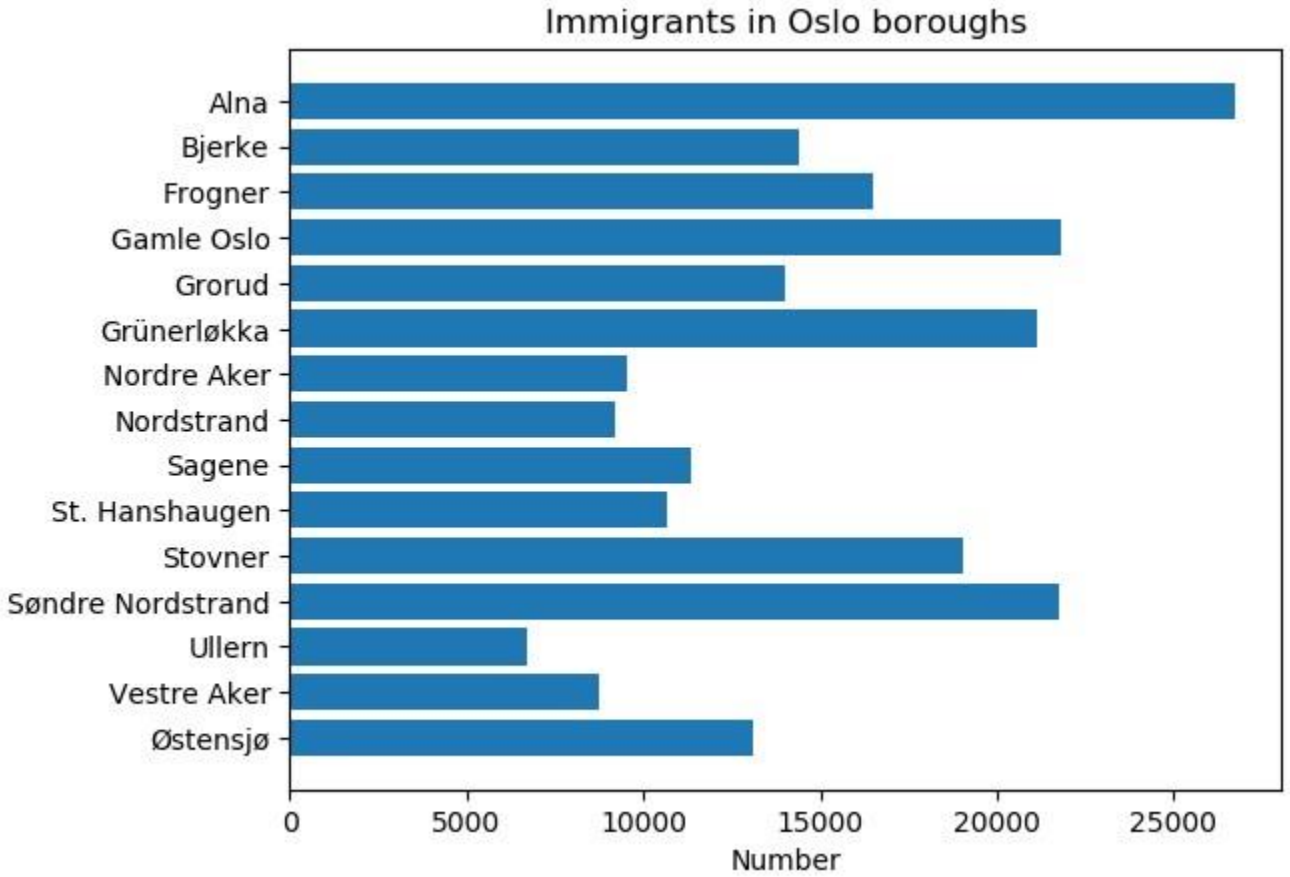
Choropleth maps

(Problem with some of borough coordinates)

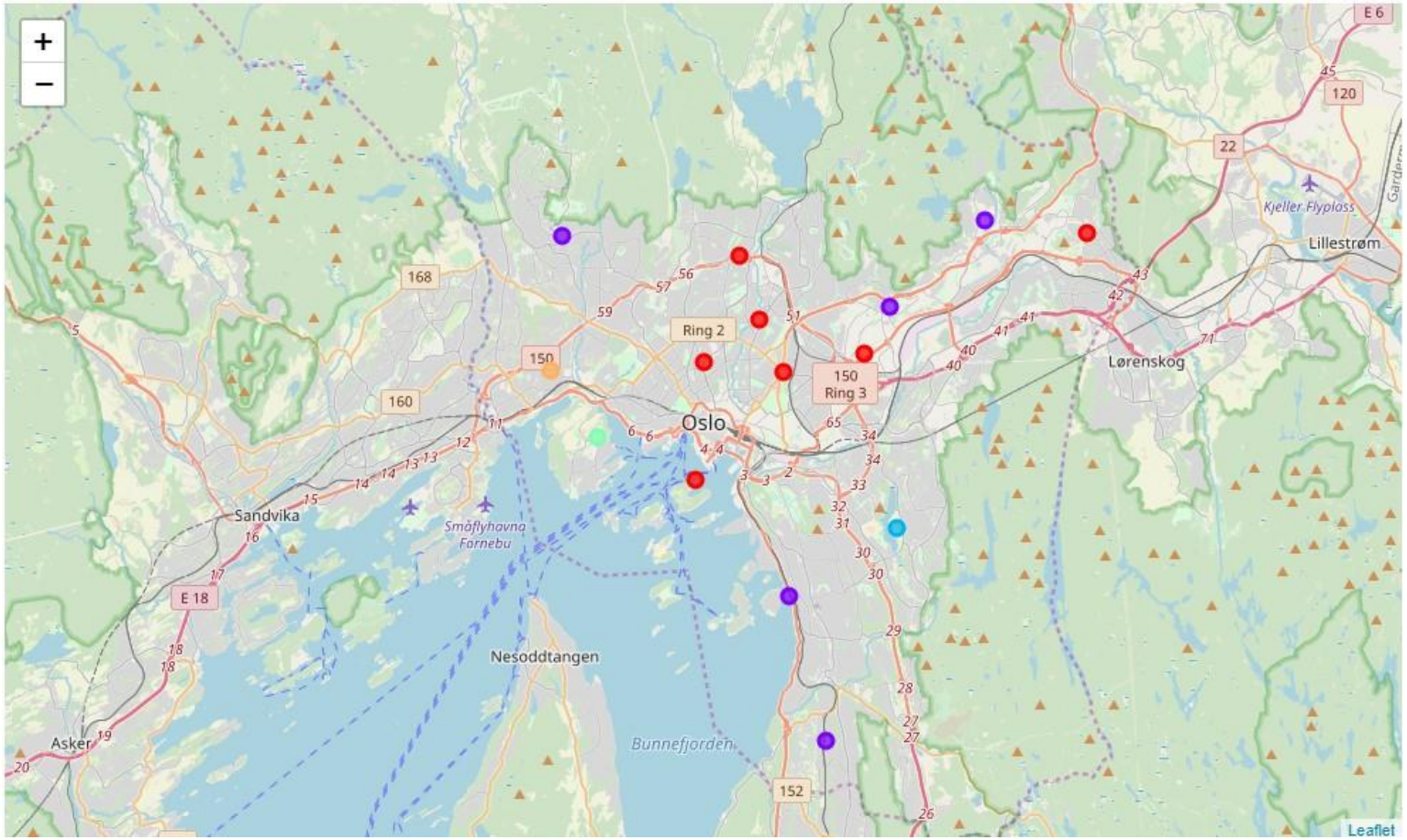




Histogram


Immigrants in Oslo boroughs
Alna
Bjerke
Frogner
Gamle Oslo
Grorud
Grünerløkka
Nordre Aker
Nordstrand
Sagene
St. Hanshaugen
Stovner
Søndre Nordstrand
Ullern
Vestre Aker
Østensjø
0
5000
10000
15000
20000
25000
Number

Clustering results



Interpretation

- Cluster 5: We had only 4 venues returned: 2 rail stations, 1 bus station and flower shop. Dog run and farm are just random outputs which is not present in Ullern. So, it should be no surprise why stations were clustered together.
- Cluster 4: 5 venues for this borough. The borough is rich with museums and nature which might be reason for a separate clustering.
- Cluster 3: We had only one venue for Østensjø, and that one venue weighs heaviest as a result. This is small sample problem.
- Cluster 2: Groceries abound in this cluster but mix of immigrant heavy and light boroughs.
- Cluster 1: Bakeries rule in this cluster, and mainly immigrant heavy eastern boroughs (Alna, Stovner) and central boroughs (Gamle Oslo, Grunerløkka, Sagene, St. Hanshaugen) in this cluster except Nordre Aker.

Observations

- The intersection immigrant population and venues is the central part of Oslo
- How much of these small businesses are due to natural tendency for businesses to be in the city center or the engagement of immigrant population?

Intersection between immigrant population and venues is the central part of Oslo

Table 1. Immigrant population and total venues

N	Immigrant Population	Most Venues
1	Alna (east)	St. Hanshaugen (centre west)
2	Gamle Oslo (centre)	Sagene (centre)
3	Søndre Nordstrand (east south)	Grunerløkka (centre)
4	Grunerløkka (centre)	Bjerke (centre east)
5	Stovner (east)	Gamle Oslo (centre)
6	Frogner (west, many foreign companies)	Alna (east)
7	Bjerke (east)	Nordre Aker (west)
8	Grorud (east)	Søndre Nordstrand (east south)

How about eastern regions like Alna, Østensjø, Grorud and Stovner do not have centrality bias?

- 6 out of 8 top rows under Immigrant Population and Top Social Security Receivers coincide

Figure 13. Total social security receivers in Oslo boroughs

N	Immigrant Population	Top Social Security Receivers
1	Alna (east)	Gamle Oslo (centre)
2	Gamle Oslo (centre)	Grunerløkka (centre)
3	Søndre Nordstrand (east south)	Sagene (centre)
4	Grunerløkka (centre)	Søndre Nordstrand (east south)
5	Stovner (east)	Frogner (west)
6	Frogner[13] (west)	St. Hanshaugen (centre west)
7	Bjerke (east)	Stovner (east)
8	Grorud (east)	Alna (east)

The higher welfare state, the less incentive to start a private business?

- The correlation between social security receivers and immigrant population is high enough: 0.60

	AreaInKm2	Latitude	Longitude	ImmigrantsTotal	NumberVenues	SocialSecurity
AreaInKm2	1.000000	-0.429327	-0.045028	-0.006455	-0.632242	-0.469866
Latitude	-0.429327	1.000000	0.096902	-0.235440	0.123390	-0.218143
Longitude	-0.045028	0.096902	1.000000	0.402237	-0.149014	0.027611
ImmigrantsTotal	-0.006455	-0.235440	0.402237	1.000000	-0.109468	0.601718
NumberVenues	-0.632242	0.123390	-0.149014	-0.109468	1.000000	0.366228
SocialSecurity	-0.469866	-0.218143	0.027611	0.601718	0.366228	1.000000

Conclusion

- Some patterns, but hard to call it a causality
- Some clustering of smal business in the centre, but that is a natural tendency everywhere
- In eastern boroughs where immigrant population populated
  - Small sample
    - Simply less online presence, or
    - Less motivation to start private business due to welfare state -> Positive correlation