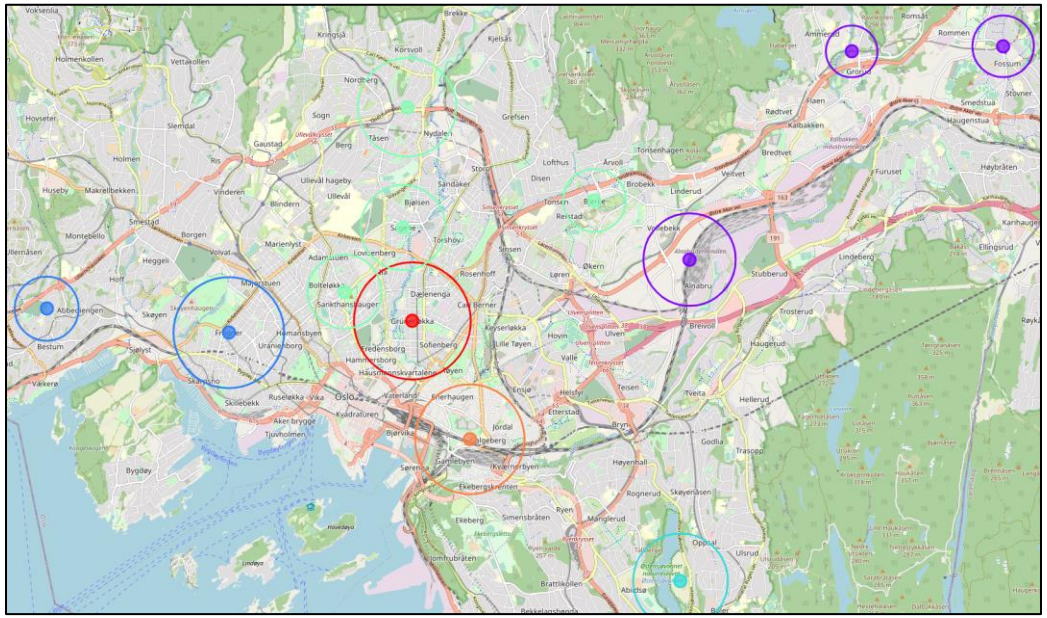
Oslo, Norway

Finding a place to live

-the power of analytics

Joachim Eriksson

20.02.2021



One of the most sought-after places to live in Norway is Oslo with its cultural diversity, great business opportunities and wide offers in different activities and parks.

However, it is quite difficult to know where to live as areas differ in activity offers and people density.

So, how do we make it easier to choose where to live in Oslo?

Agenda

Data

Wikipedia – Boroughs and population.

GeoPy – coordinates.

Foursquare – venues, activities and parks.

Methodology

Feature Engineering.

Dimensionality Reduction (PCA).

K-Means Clustering Algorithm.

Analysis

Visualizing clusters on a map.

Analyzing differences between boroughs.

Results

Classified differences of boroughs on a map.

Conclusion

How do we find a place to live?

Data used for this analysis consist of GeoPy, Foursquare and Wikipedia to be able to find venues and population densities for each borough

- Table of boroughs with population in Oslo from Wikipedia.
- GeoPy to retrieve coordinates for each borough.
- Foursquare API for its venues and parks.

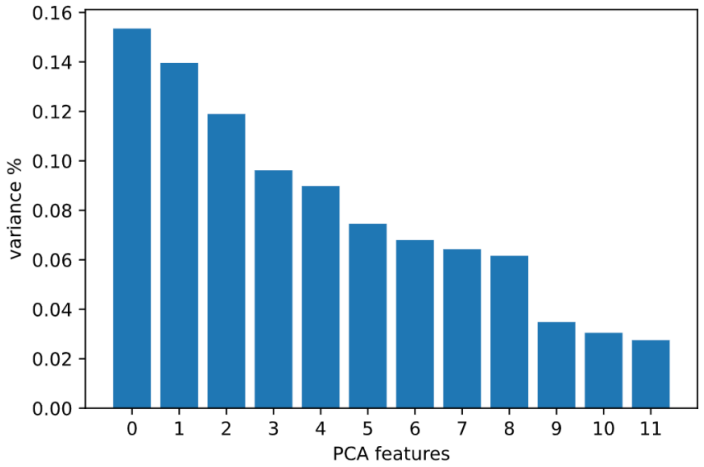
	index_col	Borough	venue_name	venue_category	venue_lat	venue_lng
423	8	Sagene	Bombay Cuisine	Indian	59.929423	10.760781
331	6	Nordre Aker	Mat & Mer	Deli / Bodega	59.940046	10.759173
12	0	Alna	en til! pizza	Pizza	59.942593	10.814542
593	11	Søndre Nordstrand	Godt Brød	Bakery	59.907029	10.757393
716	14	Østensjø	Ulvøya	Beach	59.869724	10.772059
596	11	Søndre Nordstrand	Østre Greverud Idrettshall	Athletics & Sports	59.773655	10.815591
28	0	Alna	Harald Huysman Karting	Racetrack	59.919503	10.836280
357	7	Nordstrand	Fiskevollbukta	Beach	59.842372	10.777234
711	14	Østensjø	Fuglen Coffee Roasters Oslo	Coffee Shop	59.906185	10.774646
279	5	Grünerløkka	East Kitchen	Asian	59.920755	10.757422

Feature engineering was done using the one-hot encoding method together with PCA analysis to extract most important features

- One-hot encoded features with frequency (15 rows, 127 features)

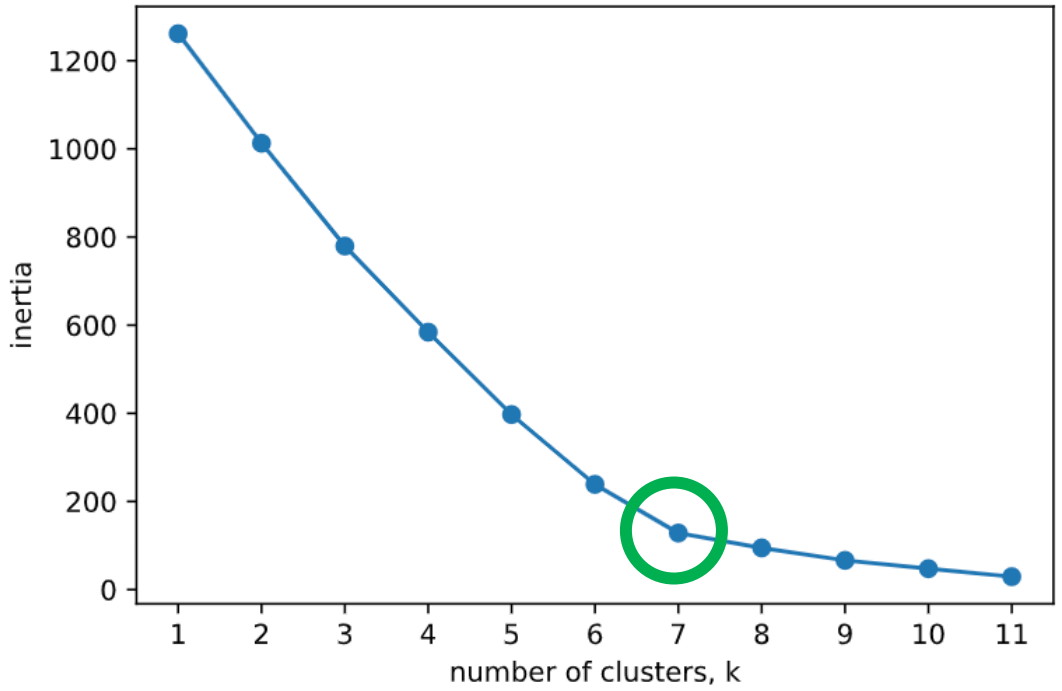
	Borough	index_col	venue_lat	venue_lng	Advertising Agency	Amphitheater	Apparel	Art Gallery	Art Museum	Arts & Crafts	...	Theme Park	Theme Restaurant	Track	Trail	Train Station	Wat
0	Alna	0	59.930589	10.821124	0.0	0.0	0.00	0.00	0.00	0.02	...	0.0	0.00	0.0	0.0	0.0	
1	Bjerke	1	59.939794	10.787494	0.0	0.0	0.00	0.00	0.00	0.00	...	0.0	0.00	0.0	0.0	0.0	
2	Frogner	2	59.920573	10.714704	0.0	0.0	0.00	0.00	0.02	0.00	...	0.0	0.00	0.0	0.0	0.0	
3	Gamle Oslo	3	59.908831	10.739617	0.0	0.0	0.02	0.04	0.00	0.00	...	0.0	0.02	0.0	0.0	0.0	
4	Grorud	4	59.944661	10.858087	0.0	0.0	0.00	0.00	0.00	0.02	...	0.0	0.00	0.0	0.0	0.0	

- After applied PCA features went from 127 to 12 with 95% of variance explained



For the clustering K-Means was used, but before running the algorithm from scikit-learn, the optimal K had to be found for optimal number of clusters.

K=7 was deemed the most optimal using the knee method



Cluster Labels on dataset

	Cluster Labels	Borough	Population	Square KM	Borough ID	lat	lng
0	1	Alna	49801.0	137.0	12	59.932417	10.835276
1	4	Bjerke	33422.0	77.0	9	59.940668	10.808725
2	2	Frogner	59269.0	83.0	5	59.922224	10.706649
3	6	Gamle Oslo	58671.0	75.0	1	59.907349	10.773927
4	1	Grorud	27707.0	82.0	10	59.961424	10.880549
5	0	Grünerløkka	62423.0	48.0	2	59.923856	10.757889
6	4	Nordre Aker	52327.0	136.0	8	59.953638	10.756412
7	3	Nordstrand	52459.0	169.0	14	59.864561	10.786143
8	4	Sagene	45089.0	31.0	3	59.936887	10.755306
9	4	St. Hanshaugen	38945.0	36.0	4	59.927950	10.738958
10	1	Stovner	33316.0	82.0	11	59.962140	10.922823
11	3	Søndre Nordstrand	39066.0	184.0	15	59.835944	10.798496
12	2	Ullern	34569.0	94.0	6	59.925567	10.655798
13	5	Vestre Aker	50157.0	166.0	7	59.958300	10.670319
14	3	Østensjø	50806.0	122.0	13	59.887563	10.832748

Results

When finding a place to live we need to look at the differences between boroughs.
Beneath you can see cluster 0 through 2 with differences in venues and population density

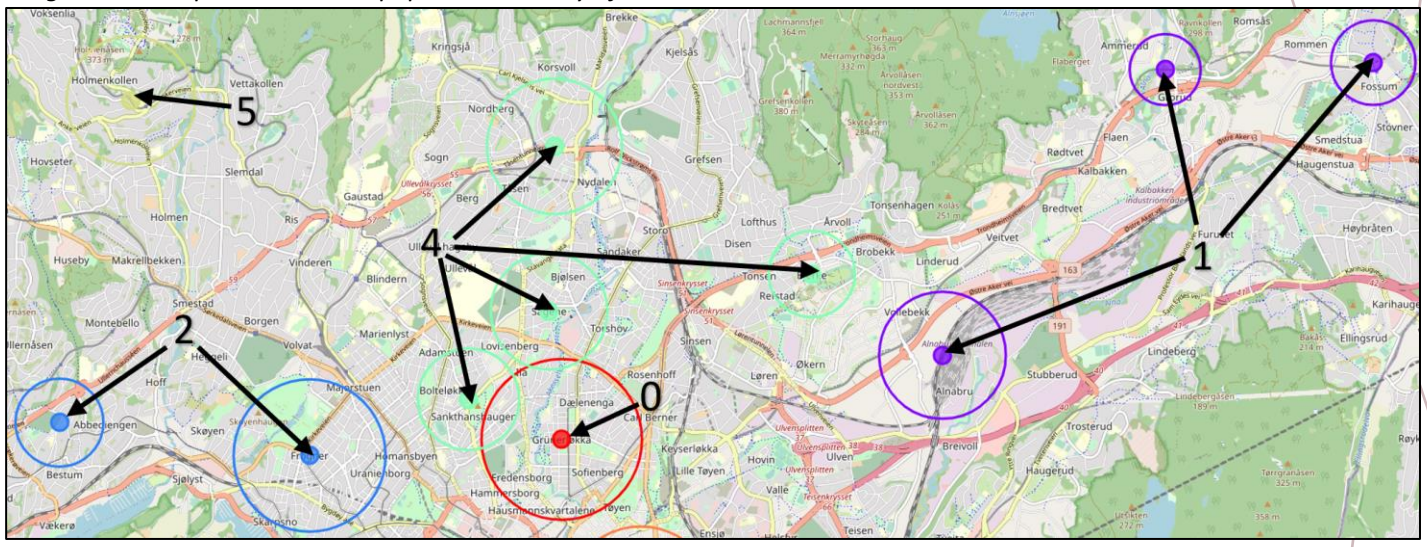
[n] *= cluster ID*

[0] Grünerløkka has bars, coffee shops, parks, cafés, Italian, French and other exotic food types. Overall, this borough is not tilted towards any specific category, and seems to have many different options.

[1] The boroughs Alna, Grorud and Stovner are areas with the most grocery stores and supermarkets, hotels and cafés. Grorud and Stovner are one of the areas with the lowest population densities with 4,0 and 4,8 percent of total respectively. It is likely that these burrows are more quiet places to live.

[2] Frogner and Ullern are the highest in Gyms and hotels, Frogner having a high population density and lower for Ullern.

Ring sizes correspond to relative population density of Oslo.



The rest of the clusters 3 through 6 we see examples of beaches, ski areas and golf courses

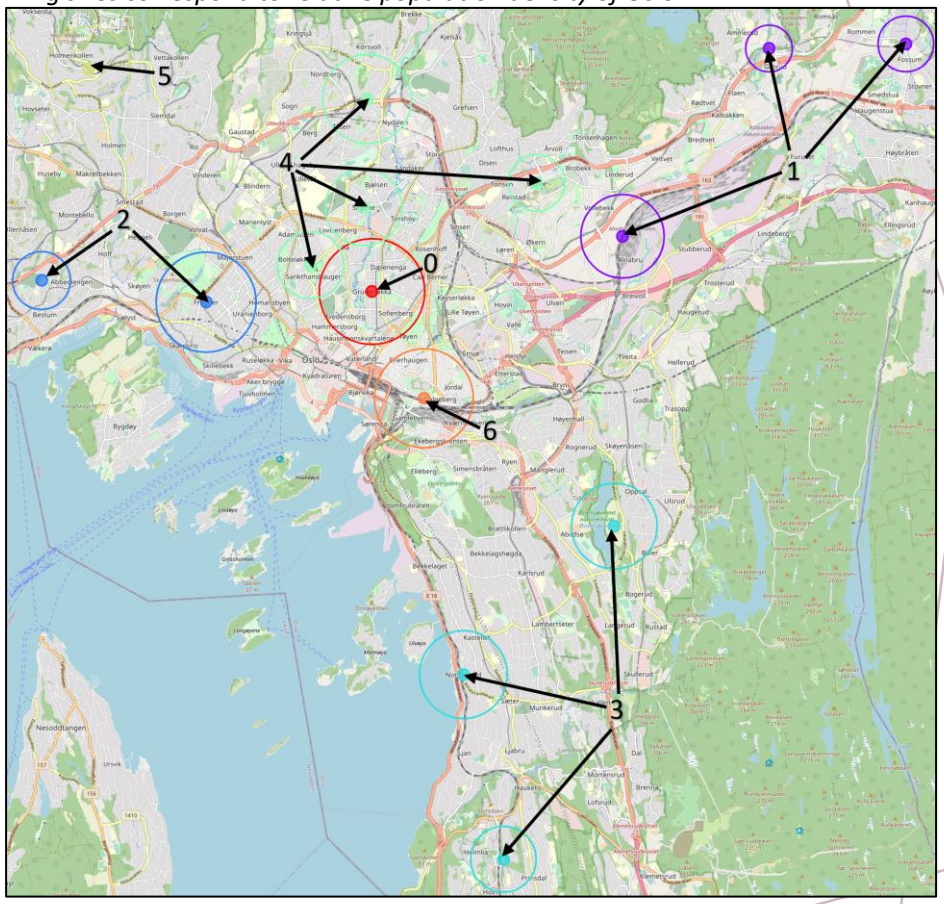
[3] Nordstrand, Søndre Nordstrand and Østensjø have access to beaches and lodges. There are also parks nearby. The density is high for Norsdatand and Østensjø around 7,5% and 5,7 for Søndre Nordstrand.

[4] Bjerke, Nordre Aker, Sagene and St. Hanshaugen are all dense in coffee shops, bakeries, restaurants, parks and gyms. They are also middle tier in population density – so this has something for all.

[5] Vestre Aker has hotels, ski areas and golf courses and medium population density.

[6] Gamle Oslo is the 3rd most population dense area and the highest in number of bars. It also has parks and record shops.

Ring sizes correspond to relative population density of Oslo.



So, how do we make it easier to choose where to live in Oslo?

As we can see on the map clustering is great for making distinctions and helps us separate the boroughs.

By analyzing the venues, activities, population density, parks and beaches it makes finding a new place to live a lot easier.

Where would you like to live?