Finding the best place to open a coffee shop

in Oslo

A Data scientist point of view

Hadrien GRIMAUD

April 27, 2021

1. Introduction

1.1 Motivation

Oslo is the capital and most populous city of Norway. Although for Scandinavian countries, populous means often less than a million citizens (Oslo alone is around 700.000 citizens, and little more than 1 million with all the suburbs).

Oslo is divided into 15 districts or boroughs (called "Bydel") of almost equal population size (around 50.000 people for most of them), and each one divided in numerous neighborhoods. This partition is relatively recent and were created the 1st of January 2004.

Oslo neighborhoods - Map - *Wikipedia*

This is also the economic and financial center of Norway, as well as an important center for maritime industries and trade in Europe. According to the annual report of best cities in the world for happiness and well-being **[6]**, Oslo is in the Top10 since several years, and Norway is also among the top 5 happiest country in the world, so it is a good place to live.

Here are some pictures of the town:

Oslo neighborhoods – trees in the city and fjord (sea) - *Pixabay©*

Oslo neighborhoods – the Opera - *Pixabay©*

Due to the small population and high level of income of Norwegian people, there are many possibilities for business. Indeed, Norwegians lives in one of the richest countries in the world **[7]** with a Gross Domestic Product (GDP) per capital > 90K$, and an exceedingly small population, less than 5,5M (all Norwegians together are not even third of NY population !!) for a quite large country of 385 207 km2 which for example is larger than Germany (populated with more than 83 million people).

It is also (like all Scandinavian countries) a highly prized destination for sustainable expats **[8]**, which should increase in the future years.

The almost only negative factor to go in Norway is cold climate (which in the next decades, because of global warming, should be also another advantage). However, as climate in Norway in still generally cold, people use to go in coffee shops, which are very popular, and where you can find hots drinks and nice pastries (this characteristic is not only because of cold climate, but probably an important factor). That is why I will focus my analysis on coffee shops more than restaurants. It is not a too complicated business (easier than a restaurant where you have many losses due to expired stock for example) and it could be a good start for an expat with not a high level of education, or also a pastry or bakery-shop chain, wanting to develop her business in North Europe.

Oslo neighborhoods – A coffee shop - *Pixabay©*

1.2 Problem

Data that might be suitable for a coffee shop business may include density of coffee shop/restaurant, also density of shops (many people doing shopping contribute to an increasing demand for drinks and snacks), density of people, average income of people per district, …

As we are asked to use Foursquare API for this analysis, we will use the number/density of different kind of venues.

For the clustering of neighborhood, we will use the most common venues.

2. Data acquisition and preprocessing

2.1 Data sources

For our analysis of neighborhoods of Oslo, I took mainly the data from Wikipedia by scrapping, a sometimes verify some points on the Oslo city website **[3]**

[1] https://no.wikipedia.org/wiki/Delbydeler_i_Oslo (thanks to Google Translate)

[2] https://en.wikipedia.org/wiki/Oslo

[3] https://www.oslo.kommune.no/ (thanks to Google Translate)

To find the coordinates of each neighborhoods to geolocalize them, I used the python library Geopy, which allows to find latitude and longitude for a given address. For example

```
address = 'Oslo, NO'

geolocator = Nominatim(user_agent="oslo_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Oslo are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Oslo are 59.9133301, 10.7389701.
```

We could also have simply used Google Maps, but as we needed to do it for all neighborhoods (more than 100) it was much faster to do it with a little Python programming.

As the purpose of this capstone was to use the **Foursquare API**, I used it for my Exploratory Data Analysis (EDA) to get the most common venues.

[4] https://developer.foursquare.com

And finally, all the pictures in our report are free of charges and come from Pixabay.

[5] https://pixabay.com

Others general REFERENCES

[6] https://worldhappiness.report/ed/2020/

[7] https://howmuch.net/articles/richest-countries-in-the-world

[8] https://www.internations.org/expat-insider/

2.2 Data preprocessing

Data scrapped from Wikipedia was unfortunately in a raw format, due to the lack of a (nice) table consisting of all boroughs and neighborhoods in Oslo.

Here is the list of different realized operations to have a clean data in a Pandas dataframe. Pictures are often easier than a long text, I will illustrate each step:

- **Scrapping boroughs of Oslo** (20 items, removing irrelevant data to keep 16 neighborhoods, as according to Wikipedia and Oslo website [3]
- **Scrapping neighborhoods** (raw format), officially 92 items

	Borough	Neighborhoods
0	Gamle Oslo	[Lodalen, Grønland, Enerhaugen, Nedre Tøyen, Kampen, Vålerenga, Helsfyr]
1	Grünerløkka	[Grünerløkka vest, Grünerløkka øst, Dælenenga, Rodeløkka, Sinsen, Sofienberg, Hasle-Løren]
2	Sagene	[Iladalen, Sagene, Bjølsen, Sandaker, Torshov]
3	St. Hanshaugen	[Hammersborg, Bislett, Ila, Fagerborg, Lindern]
4	Frogner	[Bygdøy, Frogner, Frognerparken, Majorstuen nord, Majorstuen syd, Homansbyen, Uranienborg, Skillebekk]
5	Ullern	[Ullernåsen, Lilleaker, Ullern, Montebello-Hoff, Skøyen]
6	Vestre Aker	[Røa, Holmenkollen, Hovseter, Holmen, Slemdal, Grimelund, Vinderen]
7	Nordre Aker	[Disen, Myrer, Grefsen, Kjelsås, Korsvoll, Tåsen, Nordberg, Ullevål hageby]
8	Bjerke	[Veitvet, Linderud, Økern, Årvoll]
9	Grorud	[Ammerud, Rødtvet, Nordtvet, Grorud, Romsås]
10	Stovner	[Vestli, Fossum, Rommen, Haugenstua, Stovner, Høybråten]
11	Alna	[Furuset, Ellingsrud, Lindeberg, Trosterud, Hellerudtoppen, Tveita, Teisen]
12	Østensjø	[Manglerud, Godlia, Oppsal, Bøler, Skullerud, Abildsø]
13	Nordstrand	[Ljan, Nordstrand, Bekkelaget, Simensbråten, Lambertseter, Munkerud]
14	Søndre Nordstrand	[Holmlia syd, Holmlia nord, Prinsdal, Bjørnerud, Mortensrud, Bjørndal]
15	Sentrum	[Sentrum]

- **Duplicate boroughs rows** according to the number of corresponding neighborhoods

	Borough	Neighborhood
0	Gamle Oslo	Lodalen
1	Gamle Oslo	Grønland
2	Gamle Oslo	Enerhaugen
3	Gamle Oslo	Nedre Tøyen
4	Gamle Oslo	Kampen
5	Gamle Oslo	Vålerenga
6	Gamle Oslo	Helsfyr
7	Grünerløkka	Grünerløkka vest

After this job, we needed to find latitude/longitude coordinates of each neighborhood, using Geopy package with Nominatim geolocator. That is putting some address on format 'borough, neighborhood, Town, COUNTRY) we get the coordinates. Example for the Lodalen neighborhood:

```
# Check on example with a neighborhood
address = 'Lodalen, Gamle Oslo, Oslo, NO'

geolocator = Nominatim(user_agent="test_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(location)
print(latitude,longitude)
```

```
Lodalen, Gamle Oslo, Oslo, Norge
59.90466055 10.77869333319595
```

It does not work for a few neighborhoods: Nedre Tøyen, Grünerløkka vest & øst, Majorstuen nord & syd, Holmlia nord &syd. Fortunately, it is quite easy to fix these problems:

For Nedre Tøyen, we just have to replace 'Nedre Tøyen' by 'Tøyen' and it is ok. For the other neighborhoods, there are just part of the same neighborhoods, so we do not have to make a difference between north & south, east and west, as the geolocation will do later.

Finally, we get a dataframe consisting of 90 rows corresponding to distinct neighborhoods:

	Borough	Neighborhood
0	Gamle Oslo	Lodalen
1	Gamle Oslo	Grønland
2	Gamle Oslo	Enerhaugen
3	Gamle Oslo	Nedre Tøyen
4	Gamle Oslo	Kampen
5	Gamle Oslo	Vålerenga
6	Gamle Oslo	Helsfyr
7	Grünerløkka	Grünerløkka vest
8	Grünerløkka	Grünerløkka øst
9	Grünerløkka	Dælenenga
10	Grünerløkka	Rodeløkka
11	Grünerløkka	Sinsen
12	Grünerløkka	Sofienberg
13	Grünerløkka	Hasle-Løren
14	Sagene	Iladalen

Focus on the Nominatim geolocator

Nominatim is a search engine for OpenStreetMap data. You may search for a name or address (forward search) or look up data by its geographic coordinate (reverse search). Each result comes with a link to a details page where you can inspect what data about the object is saved in the database and investigate how the address of the object has been computed.

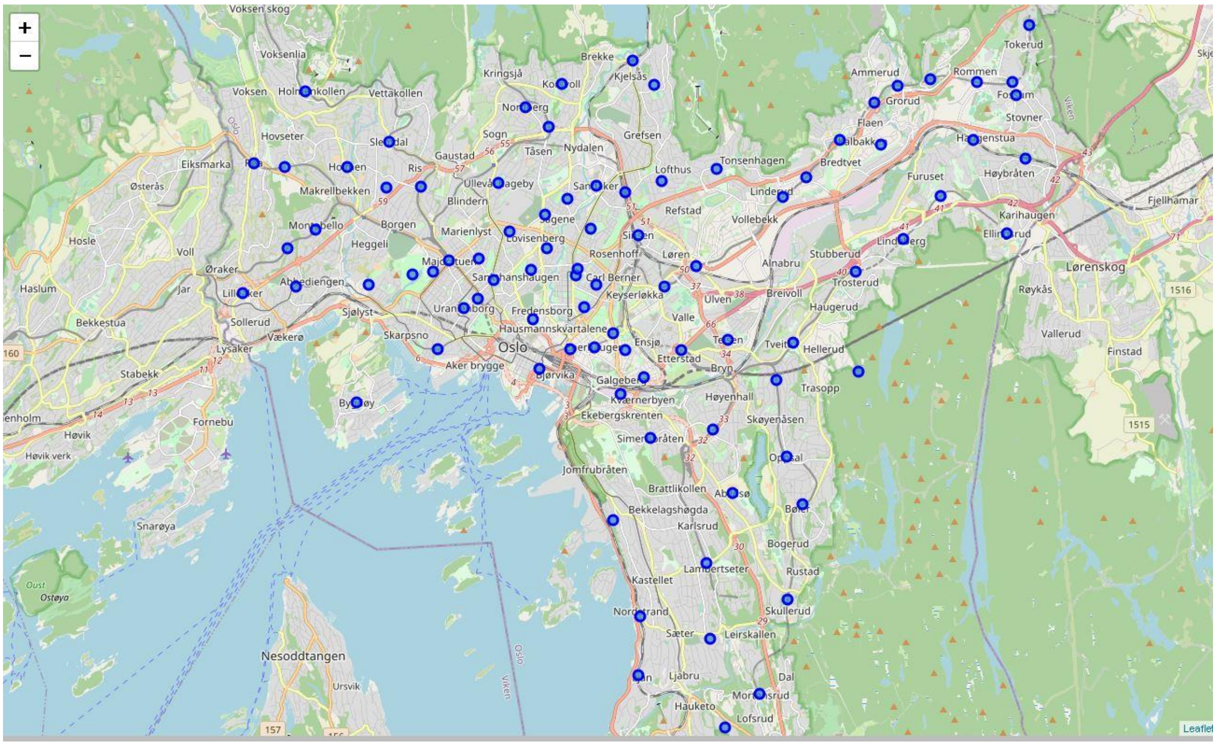
You can find the OpenStreetMap on the following link if you want to walk around Oslo (or any other location):

https://www.openstreetmap.org/search?query=oslo#map=10/59.9715/10.7227

OpenStreetMap is very famous, as it is the geolocator of many apps from world-wide tech companies like Facebook, Apple, Amazon and others (see Wikipedia) , except Google which has his own geolocation API (used in Google Map)

For the specific case of Norway, all Norwegian official addresses has been released to the public for free (see https://wiki.openstreetmap.org/wiki/Addresses#Norway) and can be find on this website : https://www.kartverket.no/en/api-and-data

Using Folium library, we create a first dynamic map to visualize our data:



Each blue circle represents a neighborhood of Oslo. To view the interactive version of this map, look at my GitHub notebook here:

3. Methodology

3.1 Exploratory Data Analysis (EDA) part 1: Foursquare API

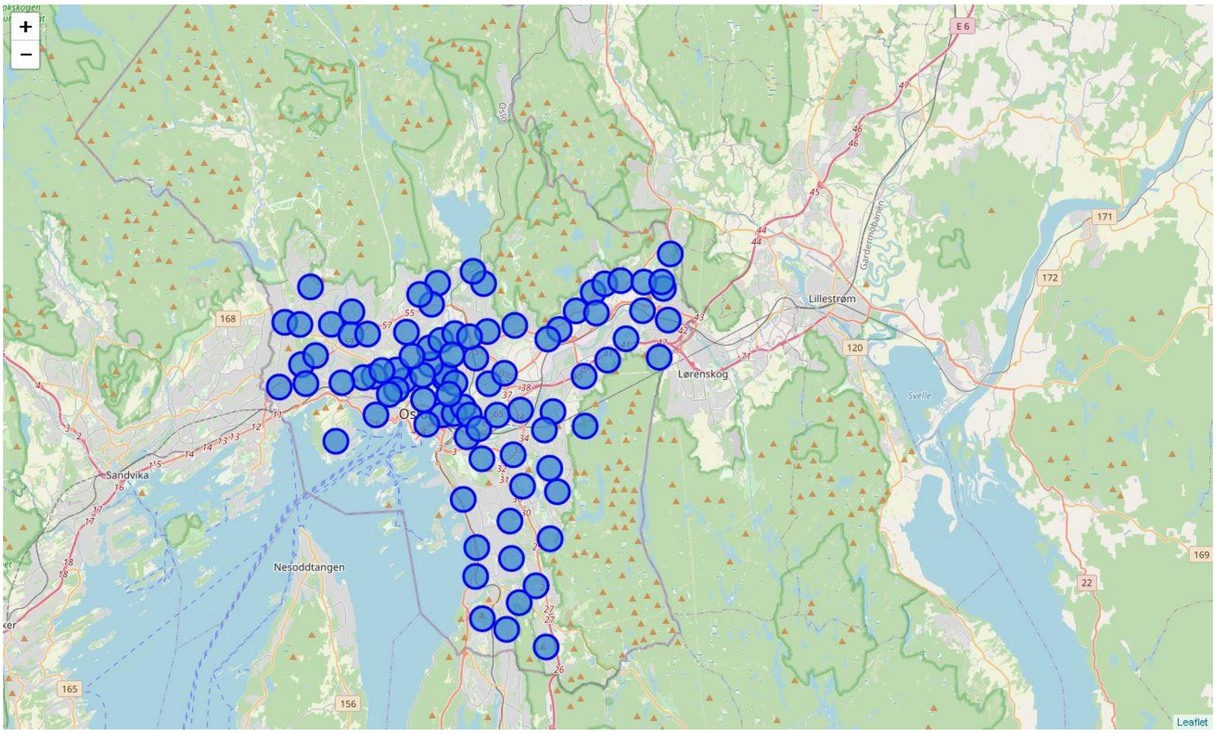
We are now able to determine our relevant features e.g., numbers of coffee shops and related by neighborhoods using Foursquare API (and also numbers of shopping stores).

First, we will use Foursquare to retrieve all venues of interest for each neighborhood. We must specify a radius in meters around each neighborhood point (see map above). But the problem is:

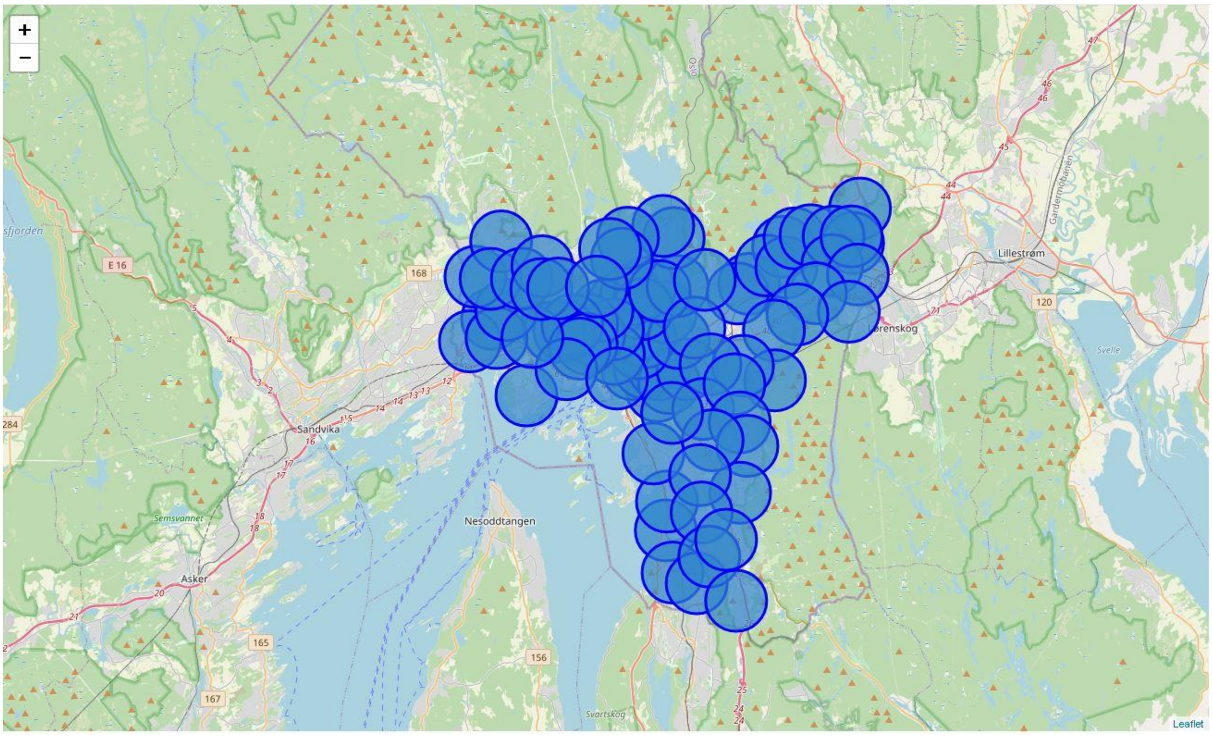
Which should be the right radius if we want to get all the results from Foursquare API?

This radius we use will greatly depends on the size of the city. If we took for example radius = 500 m, what ensure us we will obtain all the venues in Oslo? And not only a part of total venues.

Let see for example the area covered by radius of 500m around each neighborhood coordinate:



Here we will miss a lot of venues! But this is what we have done in our preceding analysis concerning NY and Toronto. This gives a good idea of the division of venues in Oslo, but this a big approximation. That is why in our analysis, we are a going to recover the entire set of venues. For this matter, we must make a **covering** of Oslo's neighborhoods. We then increase the radius of our research to R = 1250m, that is circles of 2,5km diameter:



Here we cover all the town. But this large covering has a disadvantage: as many circles of research intersect each other, the Foursquare API requests will give us many **duplicates**, e.g. some identical venues many times. But this problem will be easily fixed by considering the lat/lon coordinates, which are the same for all duplicates. Moreover, some circles go beyond Oslo, so we will also have some **outliers** to manage (but with little effort)

Let see what our request will send back: **3927 venues**

But if we make a query, with for example, the famous Maaemo restaurant in Oslo (a gastronomic restaurant, which was awarded two stars in the Michelin Guide https://guide.michelin.com/en)

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Gamle Oslo	Lodalen	59.904661	10.778693	Maaemo	59.910431	10.760291	Scandinavian Restaurant
87	Gamle Oslo	Grønland	59.912795	10.760161	Maaemo	59.910431	10.760291	Scandinavian Restaurant
195	Gamle Oslo	Enerhaugen	59.913294	10.769124	Maaemo	59.910431	10.760291	Scandinavian Restaurant
319	Gamle Oslo	Tøyen	59.915757	10.775795	Maaemo	59.910431	10.760291	Scandinavian Restaurant
416	Gamle Oslo	Kampen	59.912731	10.780247	Maaemo	59.910431	10.760291	Scandinavian Restaurant
1085	Grünerløkka	Sofienberg	59.920539	10.765193	Maaemo	59.910431	10.760291	Scandinavian Restaurant
1716	St. Hanshaugen	Hammersborg	59.918449	10.746358	Maaemo	59.910431	10.760291	Scandinavian Restaurant
3889	Sentrum	Sentrum	59.909331	10.748758	Maaemo	59.910431	10.760291	Scandinavian Restaurant

We got it several times, and we remark another problem: we no longer know in which neighborhood he belongs! Dropping duplicates will give us the right number of venues, but we have to get back their address (with borough and neighborhood) using the **reverse encoder** of Geopy: in the previous case of the Maamo restaurant, this gives us the following result:

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="oslo_reverse")
location = geolocator.reverse("59.910431, 10.760291")
#location.address
location.raw['display_name']
```

```
'Oslo Z, Annette Thommessens plass, Grønland, Gamle Oslo, Oslo, Norge'
```

That is, borough is Gamle Oslo, and neighborhood is Grønland (by the way, this address is not exact, as if we check the restaurant's address on Google, it is 200m below… so recall that Foursquare is not updated continuously, of course)

We also have some outliers like Marka (which mean “forest” in Norvegian) and surrounds Oslo, as well as some other administrative districts like Stovner or Lørenskog, which are located at the East of Oslo. Once these duplicates and outliers removed, we get:

- 1353 venues
- 16 boroughs (as expected)
- 166 neighborhoods

That means our administrative division with 90 neighborhoods was not enough accurate, as here we get 166 different neighborhoods! There are in fact subdivision of our administrative 90 neighborhoods, and commonly used for an accurate address. So, we will continue our analysis.

3.2 Exploratory Data Analysis (EDA) Part 2: Features Selection

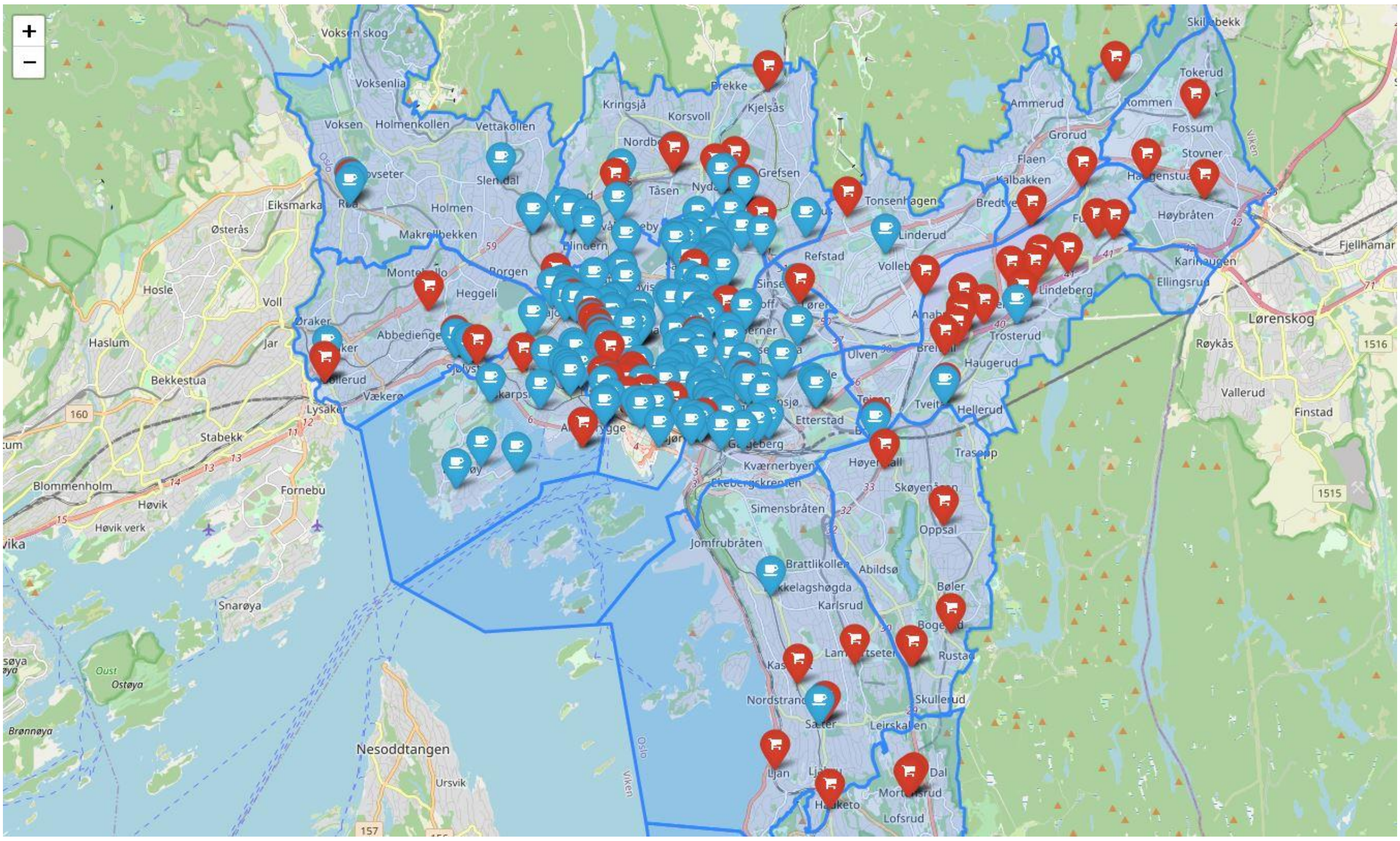
We now define what we mean by coffee shop (and similar) as well as shopping list, according to the results of our Foursquare API requests:

list_coffee_shop = ['Coffee Shop', 'Bar', 'Café', 'Bakery', 'Cocktail Bar', 'Pub', 'Brewery', 'Bistrot', 'Creperie', 'Dessert Shop']

list_shopping = ['Shopping Mall', 'Furniture / Home Store', 'Train Station', 'Clothing Store', 'Electronics Store', 'Sporting Goods Shop', 'Bookstore', 'Theater', 'Art Gallery']

list_global = list_coffee_shop + list_shopping

The coffee shop list defines our features, that is the venues we want to localize, and extract some interesting insights. We get here 216 venues of interest (= number of elements of list_global) Remember we are looking for neighborhoods where there are a few coffee shops, and many shops if possible. To achieve this goal, let see first our data distribution (To view the interactive version of this map, plug my [GitHub link] into Jupyter Nbviewer [https://nbviewer.jupyter.org/])



list_coffee_shop items are in blue (with a coffee cup symbol)

list_shopping items are in red (with a shopping-cart symbol)

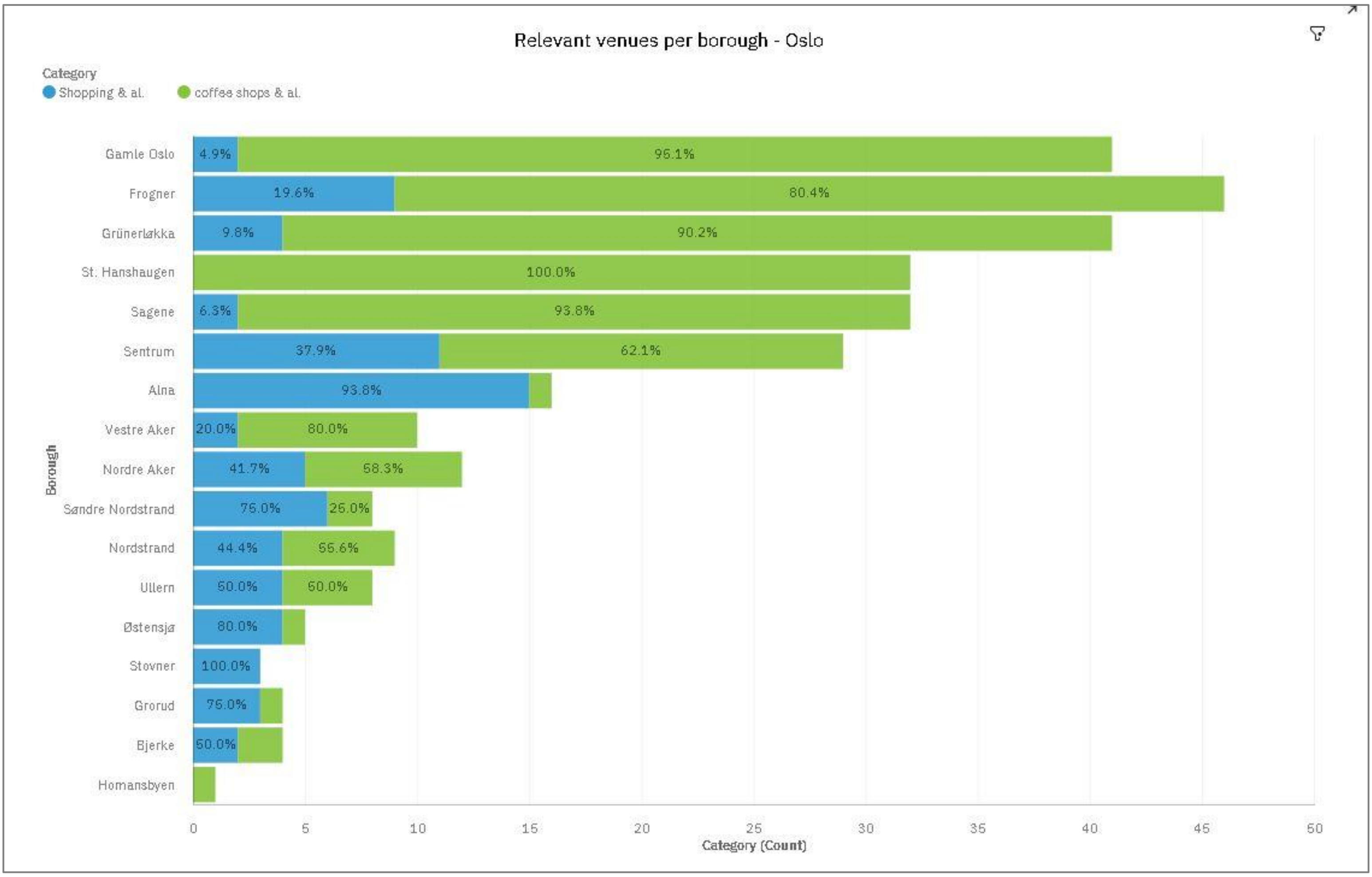
We also have superimposed the boundaries of Oslo's boroughs, to visualize if some places are out of Oslo (outliers) but we see there are not: all our venues are there, inside Oslo.

The east part of Oslo seems to be full of shopping centers. And downtown, as expected, full of coffee shops. We are going to see if clustering and deeper data analysis can provide us some interesting neighborhoods.

We also group our data to find the most common venues: because there are a large number of neighborhoods (163) it is difficult to see meaningful insights, and that is here that we **need to make a clustering analysis** to see **global trends.**

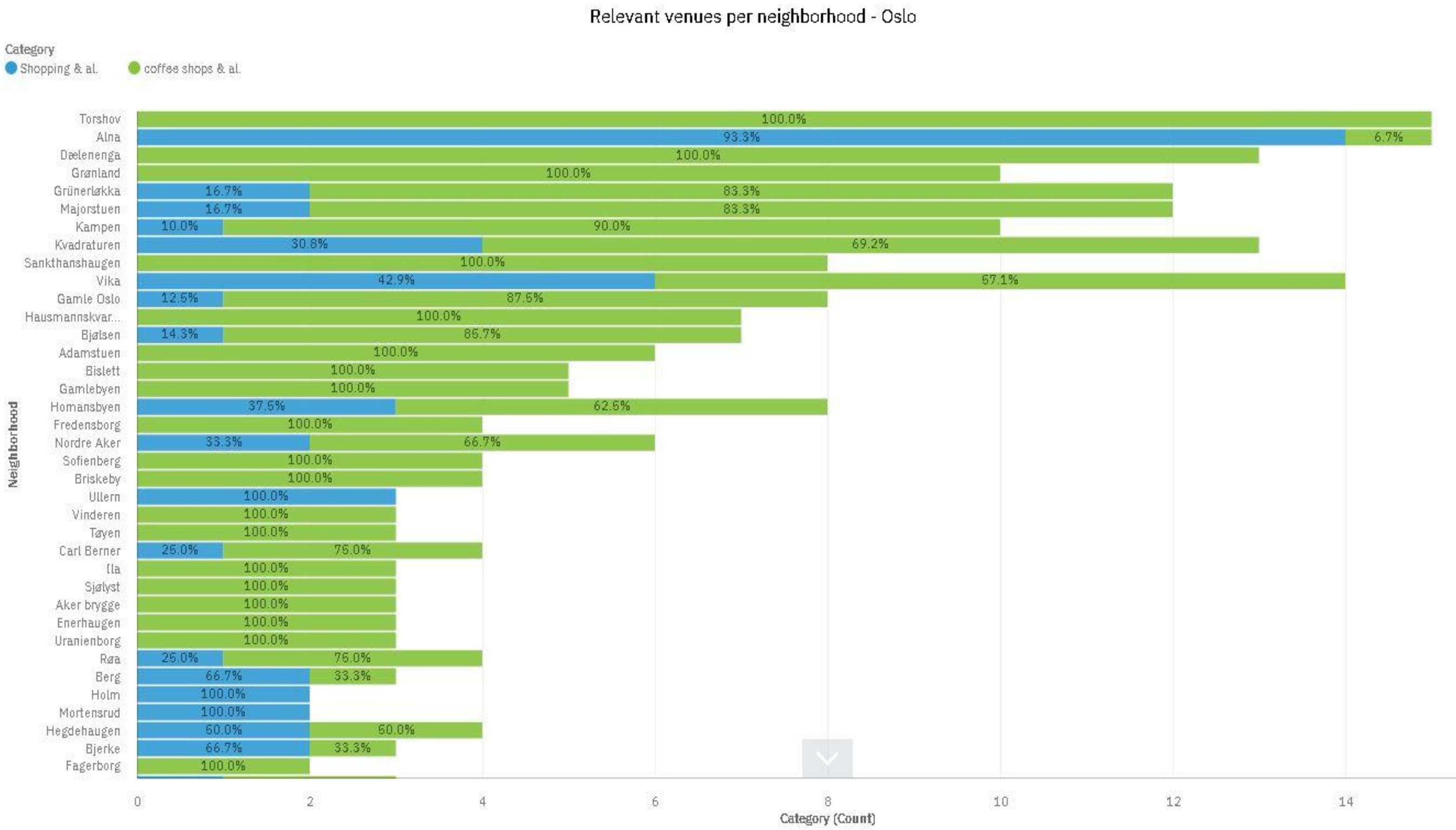
3.3 Data Visualization (first looking for global trends)

These graphics are imported from Cognos Analytics (see link here), which is a tool suitable for data viz, producing nice dashboard in a mouse click without programming. We just must put our data into a csv file, then upload into Cognos. And here are two interesting graphics:



Comments:

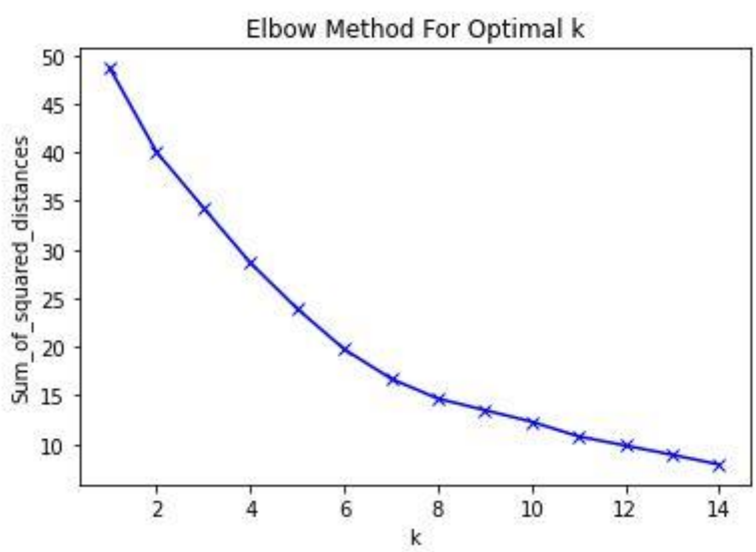
We see that Alna and Søndre Nørdstrand are good candidates, as they have few coffee shops and relatively many shopping areas. We will then go to a deeper level, by looking at the distribution per neighborhood, but this subdivision does not tell much more, as the number of venues for some neighborhoods are very few.

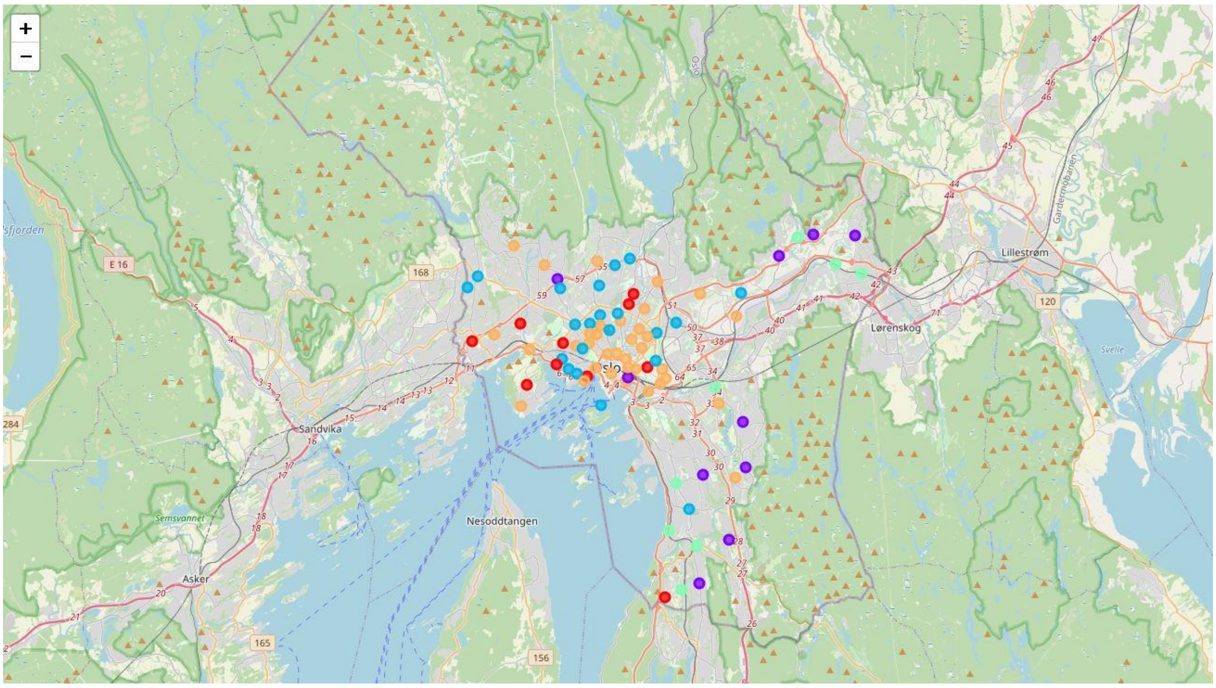


3.4 Clustering Neighborhoods: the power of unsupervised machine learning

Now we have found some interesting areas by Data visualization, we are going to use a popular method named K-means to discover some useful insights according to the density, number, and type of venues in each neighborhood, which will confirm our previous analysis with more accuracy.

Remember that a cluster is a group of objects (=neighborhood) that are like other objects in cluster, and dissimilar to data points in other clusters. We use the elbow method to determine the optimal number of clusters, it seems that 5 or 6 clusters are the ideal choice. We decide to keep 5 clusters as the results are easier to interpret with K =5.





There is a huge cluster (orange one number 4, with 35 points), which is essentially concentrated on downtown. Then, at this level, our previous analysis of most common venues per neighborhood is very useful, as well as the corresponding tab (dataframe, see below). The remaining 4 clusters are more “pure”:

Cluster	**Number of neighborhoods**	**Most proeminent venue**	**Comments**
0 (Red)	10	Café	Downtown neighborhoods, already full of coffee shops !
1 (Purple)	10	Shopping Mall	Seems to be the most interesting
2 (cyan)	21	Bakery	Need to be examined neighborhood by neighborhood
3 (green)	8	Train station	Look like interesting
4 (orange)	35	Coffee Shops	Downtown neighborhoods, already full of coffee shops! but small pockets exists

Cluster 0 is a good example of crowded neighborhoods most located in downtown.

Clusters 1, and 3 satisfy to our criteria of low density of coffee shops and many shopping venues, with a preference for cluster 1, distributed in the East of Oslo (except Sjøtomta which is in Sentrum)

Cluster 2 and 4 need to be examined neighborhood by neighborhood (and even street by street for relevant stakeholders), as there exists some interesting pockets which satisfy our criteria.

```python
# Cluster 0 : concentrated on dense areas, with already lot of coffee shops
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 0].head(20)
```

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Aker brygge	59.909928	10.725042	0	Café	Coffee Shop	Train Station	Theater	Shopping Mall
6	Bjølsen	59.940508	10.759192	0	Café	Electronics Store	Dessert Shop	Bar	Bakery
11	Bygdøy	59.906995	10.679746	0	Café	Train Station	Theater	Shopping Mall	Pub
14	Enerhaugen	59.913294	10.769124	0	Café	Coffee Shop	Train Station	Theater	Shopping Mall
18	Frogner	59.922224	10.706649	0	Café	Furniture / Home Store	Train Station	Theater	Shopping Mall
31	Hoff	59.929658	10.675400	0	Café	Art Gallery	Train Station	Theater	Shopping Mall
45	Lilleaker	59.923150	10.639532	0	Café	Train Station	Theater	Shopping Mall	Pub
61	Sagene	59.936887	10.755306	0	Café	Bar	Train Station	Theater	Shopping Mall
65	Skarpsno	59.914348	10.702332	0	Café	Train Station	Theater	Shopping Mall	Pub
83	Åsbråten	59.828008	10.782272	0	Café	Train Station	Theater	Shopping Mall	Pub

```python
# Cluster 1 : look the most interesting
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 1].head(10)
```

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Bogerud	59.876219	10.842116	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
39	Kalbakken	59.954778	10.866958	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
43	Lambertseter	59.873364	10.810371	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
52	Mortensrud	59.849287	10.829698	1	Shopping Mall	Bakery	Train Station	Theater	Pub
54	Oppsal	59.892992	10.839837	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
55	Prinsdal	59.833067	10.807914	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
56	Ris	59.946205	10.702969	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
59	Romsås	59.962615	10.892461	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
64	Sjøtomta	59.909619	10.754703	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store
71	Stovner	59.962140	10.922823	1	Shopping Mall	Train Station	Theater	Pub	Furniture / Home Store

```python
# Cluster 2 : look like interesting especially some neighborhoods (Gimle, Holtet, Linderud,...)
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 2].head(21)
```

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adamstuen	59.932728	10.734403	2	Bakery	Coffee Shop	Café	Train Station	Theater
3	Berg	59.951247	10.745366	2	Clothing Store	Bakery	Train Station	Theater	Shopping Mall
12	Carl Berner	59.926126	10.775939	2	Furniture / Home Store	Coffee Shop	Bakery	Train Station	Theater
15	Fagerborg	59.929502	10.726581	2	Café	Bakery	Train Station	Theater	Shopping Mall
16	Filipstad	59.910825	10.716861	2	Cocktail Bar	Bakery	Train Station	Theater	Shopping Mall
20	Gamle Oslo	59.899237	10.734767	2	Coffee Shop	Bakery	Electronics Store	Bar	Train Station
22	Gimle	59.916347	10.706452	2	Bakery	Train Station	Theater	Shopping Mall	Pub
33	Holtet	59.943085	10.636071	2	Bakery	Train Station	Theater	Shopping Mall	Pub
46	Linderud	59.940963	10.838420	2	Bakery	Train Station	Theater	Shopping Mall	Pub
48	Lovisenberg	59.933379	10.747348	2	Bakery	Train Station	Theater	Shopping Mall	Pub
49	Løren	59.929834	10.790311	2	Coffee Shop	Bakery	Train Station	Theater	Shopping Mall
50	Majorstuen	59.929271	10.715569	2	Bakery	Pub	Theater	Creperie	Coffee Shop
53	Nordre Aker	59.953638	10.756412	2	Coffee Shop	Café	Bakery	Train Station	Shopping Mall
60	Røa	59.947157	10.643631	2	Shopping Mall	Coffee Shop	Café	Bakery	Train Station
62	Sankthanshaugen	59.927264	10.741108	2	Coffee Shop	Café	Bakery	Cocktail Bar	Train Station
66	Skillebekk	59.912799	10.711163	2	Bakery	Train Station	Theater	Shopping Mall	Pub
72	Sæter	59.860529	10.800535	2	Coffee Shop	Bakery	Train Station	Theater	Shopping Mall
75	Tøyen	59.915757	10.775795	2	Coffee Shop	Café	Bakery	Train Station	Theater
77	Ullevål hageby	59.943508	10.733755	2	Café	Bakery	Train Station	Theater	Shopping Mall
78	Uranienborg	59.920436	10.721020	2	Dessert Shop	Café	Bakery	Train Station	Theater
82	Vinderen	59.942802	10.704988	2	Bakery	Coffee Shop	Café	Train Station	Theater

```python
# Cluster 3 : looks like interesting
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 3].head(10)
```

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
23	Grorud	59.961424	10.880549	3	Train Station	Brewery	Theater	Shopping Mall	Pub
27	Haugenstua	59.951474	10.908540	3	Train Station	Theater	Shopping Mall	Pub	Furniture / Home Store
28	Hauketo	59.847015	10.806027	3	Train Station	Theater	Shopping Mall	Pub	Furniture / Home Store
32	Holm	59.830839	10.794767	3	Train Station	Shopping Mall	Theater	Pub	Furniture / Home Store
35	Høybråten	59.947995	10.927620	3	Train Station	Theater	Shopping Mall	Pub	Furniture / Home Store
36	Høyenhall	59.905908	10.819796	3	Train Station	Bar	Theater	Shopping Mall	Pub
41	Kastellet	59.870296	10.790310	3	Train Station	Theater	Shopping Mall	Pub	Furniture / Home Store
47	Ljan	59.852780	10.784992	3	Train Station	Theater	Shopping Mall	Pub	Furniture / Home Store

```
# Cluster 4 : already crowded by coffee-shops, essentially downtown, except Alna neighborhood
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 4].head(35)
```

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Alna	59.932417	10.835276	4	Furniture / Home Store	Shopping Mall	Electronics Store	Café	Train Station
4	Bislett	59.925553	10.731894	4	Coffee Shop	Pub	Café	Bakery	Train Station
5	Bjerke	59.940668	10.808725	4	Train Station	Shopping Mall	Furniture / Home Store	Café	Theater
7	Bjørvika	59.908382	10.754553	4	Coffee Shop	Train Station	Theater	Shopping Mall	Pub
9	Bolteløkka	59.928825	10.734031	4	Pub	Train Station	Theater	Shopping Mall	Furniture / Home Store
10	Briskeby	59.921150	10.716572	4	Coffee Shop	Café	Brewery	Bar	Train Station
13	Dælenenga	59.927642	10.762951	4	Café	Brewery	Coffee Shop	Bar	Pub
17	Fredensborg	59.898758	10.676228	4	Bar	Cocktail Bar	Train Station	Theater	Shopping Mall
19	Galgeberg	59.907229	10.778955	4	Pub	Train Station	Theater	Shopping Mall	Furniture / Home Store
21	Gamlebyen	59.904354	10.769846	4	Coffee Shop	Café	Cocktail Bar	Bakery	Train Station
24	Grønland	59.912795	10.760161	4	Bar	Coffee Shop	Pub	Café	Train Station
25	Grünerløkka	59.923856	10.757889	4	Coffee Shop	Bar	Shopping Mall	Furniture / Home Store	Creperie
26	Hammersborg	59.918449	10.746358	4	Bar	Train Station	Theater	Shopping Mall	Pub
29	Hausmannskvartalene	59.916638	10.753795	4	Cocktail Bar	Bar	Coffee Shop	Café	Brewery
30	Hegdehaugen	59.925823	10.726439	4	Pub	Clothing Store	Café	Train Station	Theater
34	Homansbyen	59.922189	10.726256	4	Clothing Store	Coffee Shop	Bar	Cocktail Bar	Train Station
37	Ila	59.930672	10.749731	4	Coffee Shop	Brewery	Train Station	Theater	Shopping Mall
38	Jordal	59.909168	10.783130	4	Bar	Train Station	Theater	Shopping Mall	Pub
40	Kampen	59.912731	10.780247	4	Bar	Café	Shopping Mall	Coffee Shop	Bakery
42	Kvadraturen	59.910873	10.742479	4	Coffee Shop	Clothing Store	Bar	Theater	Shopping Mall
44	Langerud	59.872242	10.834396	4	Furniture / Home Store	Train Station	Theater	Shopping Mall	Pub
51	Meyerløkka	59.918314	10.739677	4	Coffee Shop	Train Station	Theater	Shopping Mall	Pub
57	Rodeløkka	59.924670	10.769644	4	Coffee Shop	Theater	Cocktail Bar	Bar	Train Station
58	Rognerud	59.899810	10.822482	4	Clothing Store	Train Station	Theater	Shopping Mall	Pub
63	Sjølyst	59.919962	10.681874	4	Coffee Shop	Train Station	Theater	Shopping Mall	Pub
67	Slemdal	59.951067	10.693321	4	Coffee Shop	Train Station	Theater	Shopping Mall	Pub
68	Sofienberg	59.920539	10.765193	4	Coffee Shop	Cocktail Bar	Brewery	Bar	Train Station

4. Results

We have finally obtained 5 clusters, one of them seems particularly interesting (cluster 1), and some neighborhoods belonging to cluster 0 that we found in our previous graphical analysis. Our choice presented here consists of neighborhoods belonging in cluster 1 and Alna, identified with help of data visualization as a promising place.

They are:

Cluster	Neighborhood	Borough	Geographical situation
1	Bogerud	Østensjø	South
1	Kalbakken	Grorud	East
1	Lambertseter	Nordstrand	South
1	Mortensrud	Søndre Nordstrand	South
1	Oppsal	Østensjø	South
1	Prinsdal	Søndre Nordstrand	South
1	Ris	Vestre Aker	West
1	Romsås	Grorud	East
1	Sjøtomta	Sentrum	Downtown
1	Stovner	Stovner	Stovner
0	Alna	Alna	East

It is interesting to remark that the one looking the most promising (Alna) was more evident with data visualization than clustering. Indeed, clustering is a mean to find global trends, not a particular data point with specific characteristics.

As we could have guessed from the beginning, most of the places of interest according to our criteria are quite far from downtown, which is already full a coffee-shop.

Nevertheless, one of them, Sjøtomta, is located in downtown (Sentrum). Sjøtomta is in the heart of Oslo downtown, and then could be the best choice if one wants to be closest as possible to downtown.

The final choice depends on much more parameters than we choose. Indeed, other neighborhoods could be interesting in cluster 3 for example (where the competitors are Pub, but they rank only in 4th top place)

Globally, the East part of Oslo is the one with less coffee shops but however lot of shopping places.

5. Discussion

Our analysis shows that although there is a great number of coffee shops and related in Oslo, there are still areas with relative few of them and high number of other places of interest (like stores and shops). Lowest concentration of coffee shops was detected in downtown, north and Oslo center, so we focused our attention to areas south and east, corresponding to especially to boroughs Alna and Søndre Nørdstrand and Østensjø to name only those.

Nevertheless, it is possible to find interesting places in downtown, by looking at the interactive map where we put all the locations of coffee shops.

Stakeholders willing to open a coffee shop in Sentrum (downtown) could indeed use our analysis to find small pockets with relative few numbers of coffee shops, as they have a geolocalization of the coffee shops and relatives on a detailed map of Oslo. They can check street by street their ideal location, knowing exactly where their competitors are, as well as potential customers doing shopping.

6.Conclusion

Purpose of this project was to identify Oslo areas with low number of coffee shops to help stakeholders in narrowing down the search for optimal location for a new trendy coffee shop. By calculating coffee shop density distribution from Foursquare data, we have first identified general boroughs that justify further analysis (Alna). Clustering of coffee-shops family (bar, café, coffee-shops) and shopping family (stores, mall...) then confirm our intuition and visualization and create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal coffee shop location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads or railroads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

Thanks for reading!