Analysis of communes in Denmark: finding a new home

Contents

1. Introduction: Business Problem

In these unprecedented times, many people are creating new ways to work. Some people are frustrated about not being able to go to office and small-talk with colleagues. At the same time, other people find working from home advantageous. It is good for the climate (no need to drive), small-talk can be avoided.

In Denmark, most of companies are located in and around Copenhagen. Not surprisingly, house prices are the highest in and around Copenhagen as well, ranging from 3 mio DKK to 10-15 mio DKK.

Now, what if you are employed in a Copenhagen based company, but work from home and thus live in more affordable commune?

Our purpose is to provide guidelines to employees who would like to move out of Copenhagen without changing their lifestyle dramatically. Our target audience is the people, which have found a new degree of freedom in the time of the crisis: freedom to move.

With this in mind, we try to find a commune (municipality) in Denmark that is similar to Copenhagen, and is not Copenhagen. The new commune should combine both, active business/entertainment and culture life. We choose to focus on the number of private companies and the number of museums.

With all the information described, we can create a map and information chart with the relevant information for each commune; moreover, communes are clustered according to the venues.

2. Data description

Based on the definition of our problem, here is a list of data used in this analysis:

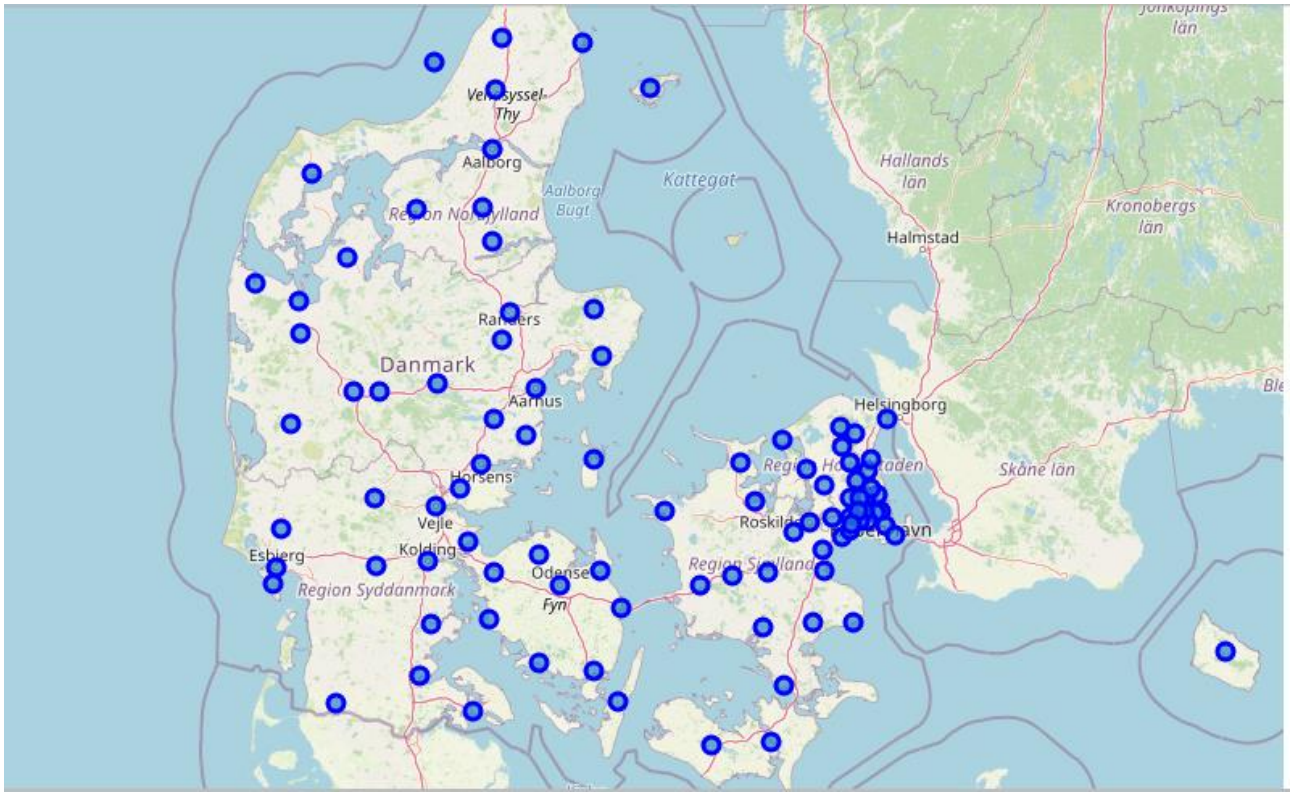
1. There are 98 communes in Denmark. We use the list of communes (municipalities) in Denmark in **Wikipedia** [1].
2. We use **Foursquare API** [2] to get most common venues in each commune. The data is used to cluster all communes and find those similar to Copenhagen.
3. For each commune, we find geographical coordinates using **Geocoder** [3]. Then, using a JSON file of Denmark [4], we create a Cloropleth map with the data for all communes,
4. From statistical database **Statistikbanken**, we find number of museums in each commune [5].

5. From statistical database **Statistikbanken**, we find number of private companies in each commune [5].

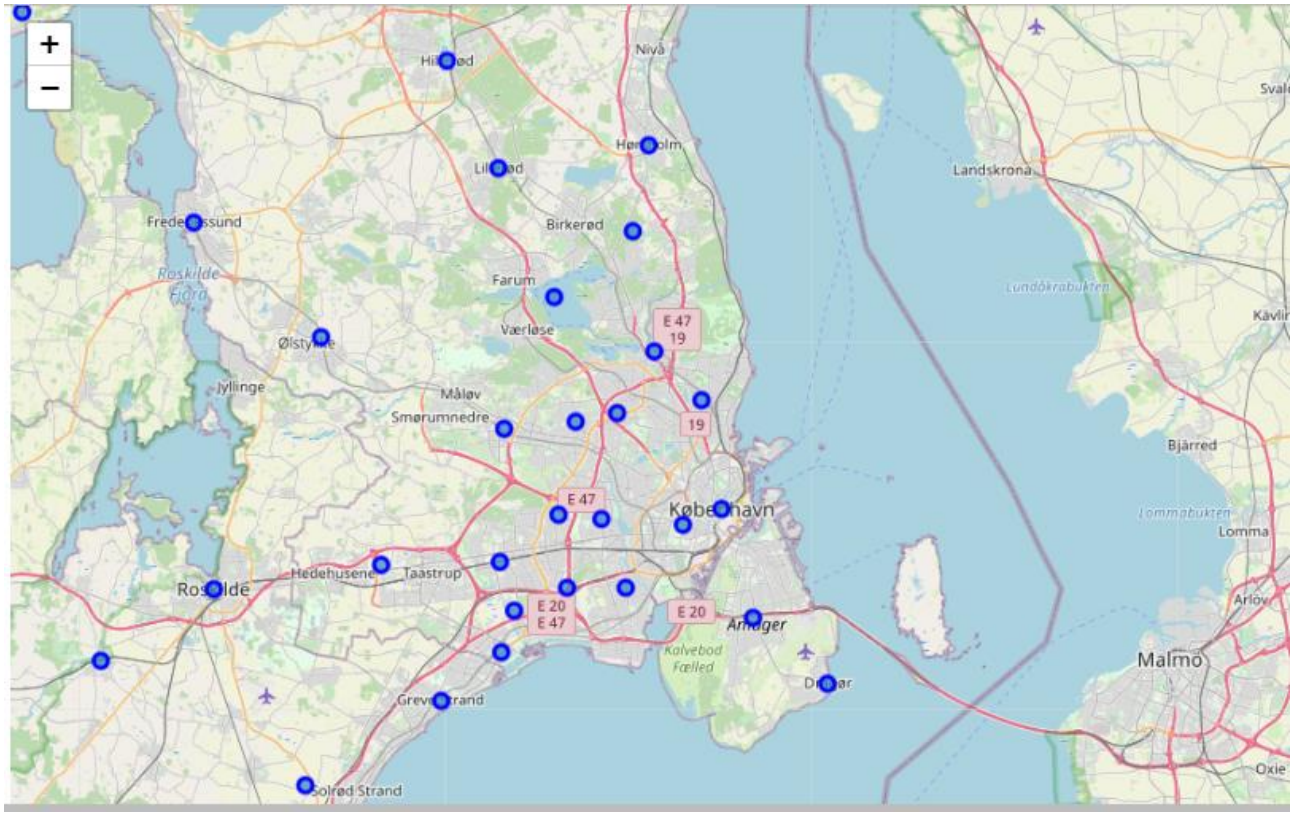
We consider the following factors when recommending a commune:

- Similarity to Copenhagen: clustering based on most common venues in each commune
- Number of museums
- Number of private companies.

Below is a map of Denmark with all communes marked:



The map zoomed in to the Copenhagen area:



Here is a snapshot of datatable with the number of private companies for each commune:

	Commune	PrivateCompanies
0	København	32415
1	Frederiksberg	4842
2	Dragør	621
3	Tårnby	1758
4	Albertslund	1465

Here is a snapshot of datatable with the number of museums for each commune:

	Commune	Museums
0	København	36
1	Frederiksberg	6
2	Dragør	4
3	Tårnby	1
4	Albertslund	0

3. Methodology

There are 98 communes in Denmark. The resulting dataframe, which contains for each commune: name, area, population, population density, and region is below:

	Nr.	Navn	Km2	2019	indb/km2	Region
0	1.0	København	86.6	623404.0	7082.0	Hovedstaden
1	2.0	Aarhus	467.9	345332.0	728000.0	Midtjylland
2	3.0	Aalborg	1137.4	215312.0	188000.0	Nordjylland
3	4.0	Odense	305.6	204182.0	662000.0	Syddanmark
4	5.0	Esbjerg	795.1	115652.0	146000.0	Syddanmark

We use Geocoder to find geographical coordinates of each commune. The resulting dataframe looks as follows:

	Nr.	Navn	Km2	2019	indb/km2	Region	location	point	latitude	longitude	altitude
0	1.0	København	86.6	623404.0	7082.0	Hovedstaden	(København, Københavns Kommune, Region Hovedst...	(55.6867243, 12.5700724, 0.0)	55.686724	12.570072	0.0
1	2.0	Aarhus	467.9	345332.0	728000.0	Midtjylland	(Aarhus, Aarhus Kommune, Region Midtjylland, 8...	(56.1496278, 10.2134046, 0.0)	56.149628	10.213405	0.0
2	3.0	Aalborg	1137.4	215312.0	188000.0	Nordjylland	(Aalborg, Aalborg Kommune, Region Nordjylland,...	(57.0482206, 9.9193939, 0.0)	57.048221	9.919394	0.0
3	4.0	Odense	305.6	204182.0	662000.0	Syddanmark	(Odense, Odense Kommune, Region Syddanmark, 50...	(55.3997225, 10.3852104, 0.0)	55.399723	10.385210	0.0
4	5.0	Esbjerg	795.1	115652.0	146000.0	Syddanmark	(Esbjerg, Esbjerg Kommune, Region Syddanmark, ...	(55.4664892, 8.4520751, 0.0)	55.466489	8.452075	0.0

We use Foursquare API to fetch venues in each commune in the radius of 10 km and with a limit of 100. The venues are stored in a dataframe Denmark_venues.

There are 23 communes that have reached the limit of 100: Tårnby, Gentofte, Lyngby-Taarbæk, Frederiksberg, Hvidovre, Gladsaxe, Odense, Rødovre, Glostrup, København, Aarhus, Brøndby, Furesø, Ishøj, Albertslund, Ballerup, Dragør, Allerød, Vallensbæk, Rudersdal, Herlev, Hørsholm, Helsingør.

Other communes have not reached the limit of 100. For one commune there were no venues found: Morsø. Morsø municipality is a municipality in Region Nordjylland in northern Denmark. The municipality is located on the island of Morsø or Mors. It has a total population of 22,091 (2008). The main town and the site of its municipal council is the town of Nykøbing Mors.

In total, there are 4425 venues found for the 97 communes. Below is a snapshot of a table, which contains all found venues:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	København	55.686724	12.570072	Botanisk Have	55.686169	12.573246	Botanical Garden
1	København	55.686724	12.570072	Torvehallerne	55.683934	12.569244	Food Court
2	København	55.686724	12.570072	The Coffee Collective	55.683695	12.569790	Coffee Shop
3	København	55.686724	12.570072	Rosenborg Slot	55.685683	12.577482	Palace
4	København	55.686724	12.570072	Kongens Have	55.684361	12.580099	Park

We see that some communes have less than 10 venues found. We drop these 16 communes: Brønderslev, Jammerbugt, Lolland, Nordfyn, Kalundborg, Lemvig, Læsø, Langeland, Vesthimmerland, Ringkøbing-Skjern, Skive, Thisted, Tønder, Vejen, Mariagerfjord, Vordingborg.

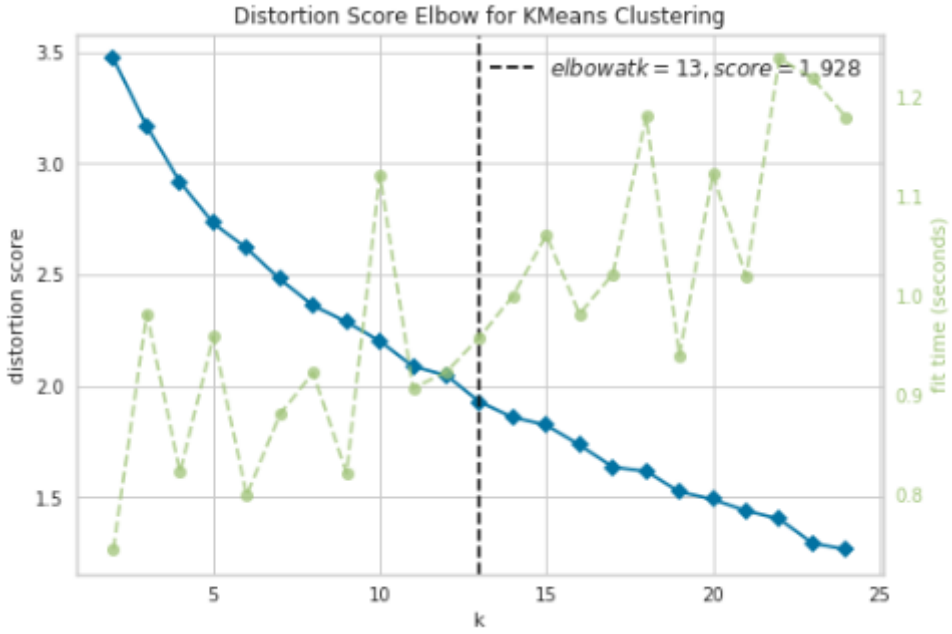
We now consider 81 commune now in our analysis, out of 98. There are 250 unique categories.

Next, we find 10 most common venues in each commune. Below is a table with 10 most common venues for each commune:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aabenraa	Fast Food Restaurant	Gas Station	Café	Harbor / Marina	Convenience Store	Gym / Fitness Center	Hotel	Pub	Bakery	Restaurant
1	Aalborg	Café	Grocery Store	Discount Store	Gym / Fitness Center	Italian Restaurant	Park	Scandinavian Restaurant	Pub	Bakery	Gas Station
2	Aarhus	Grocery Store	Café	Park	Scandinavian Restaurant	Gym / Fitness Center	Italian Restaurant	Discount Store	Hotel	Beach	Sandwich Place
3	Albertslund	Grocery Store	Gym / Fitness Center	Fast Food Restaurant	Furniture / Home Store	Market	Café	Supermarket	Restaurant	Music Venue	Hotel
4	Allerød	Grocery Store	Golf Course	Fast Food Restaurant	Café	Sushi Restaurant	Supermarket	Steakhouse	Music Venue	Park	Trail

In the following, we use unsupervised machine learning algorithm k-means clustering. For this, we transform the categorical variable ‘Venue’ using one hot encoding.

We use the elbow method [6] to find the optimum number of clusters, k. The curve generated by the algorithm is below:

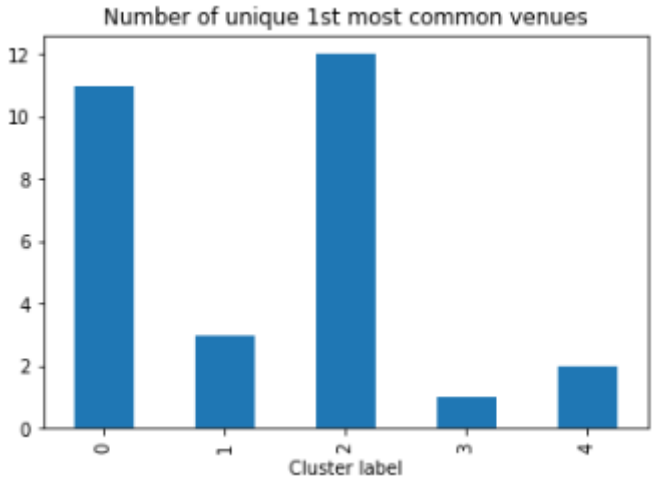


However, we see that the curve does not have an elbow and is a smooth curve. This can be an indication that our data is not very clustered. Thus, the algorithm did not return an optimum k value.

By experimenting with the k-means clustering algorithm and using the domain knowledge, we find that **five clusters** seem to fit the data. Here is a table with the cluster labels:

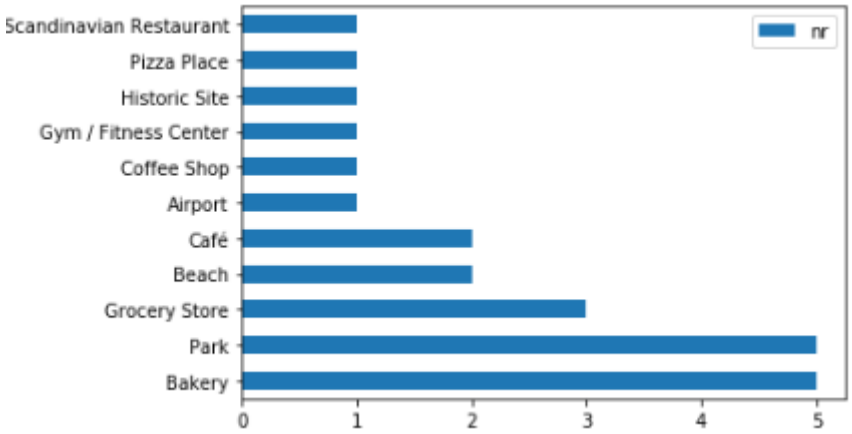
Navn	Region	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
København	Hovedstaden	55.686724	12.570072	0	Coffee Shop	Beer Bar	Park	Bakery	Movie Theater	Café	Ice Cream Shop	Theater	Burger Joint	Playground
Aarhus	Midtjylland	56.149628	10.213405	0	Beach	Café	Grocery Store	Park	Scandinavian Restaurant	Coffee Shop	Discount Store	Athletics & Sports	History Museum	Bar
Aalborg	Nordjylland	57.048221	9.919394	1	Grocery Store	Café	Italian Restaurant	Discount Store	Park	Tapas Restaurant	Sushi Restaurant	Gym / Fitness Center	French Restaurant	Cultural Center
Odense	Syddanmark	55.399723	10.385210	0	Café	Bar	Coffee Shop	Grocery Store	Supermarket	History Museum	Scandinavian Restaurant	Bakery	Concert Hall	Restaurant
Esbjerg	Syddanmark	55.466489	8.452075	1	Grocery Store	Café	Hotel	Beach	Restaurant	Steakhouse	Zoo	Bowling Alley	Sports Bar	Sporting Goods Shop

The figure below shows how many unique venues there are in each cluster:

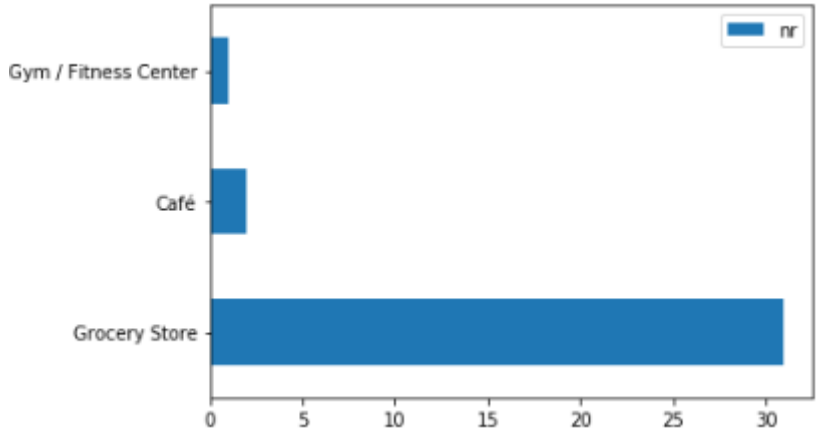


So, the algorithm has created the following five clusters. For each cluster, we show the number of 1st most common venues:

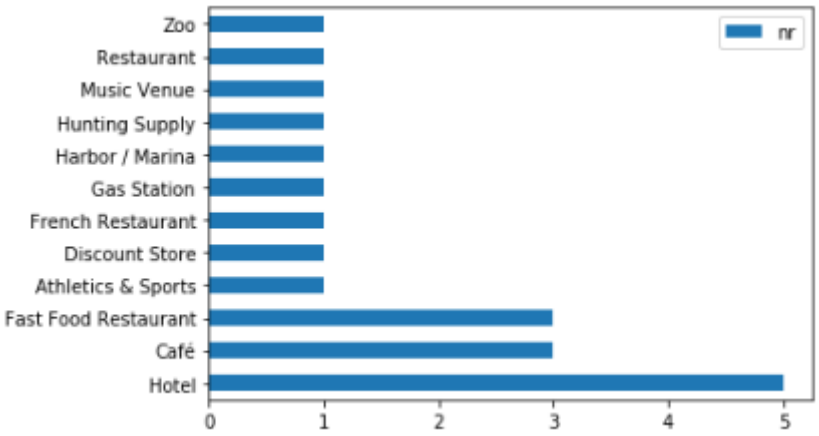
Cluster 1



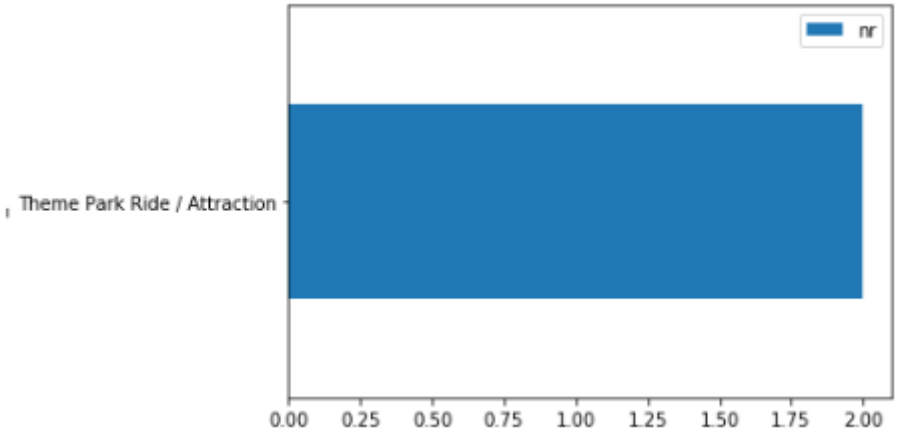
Cluster 2



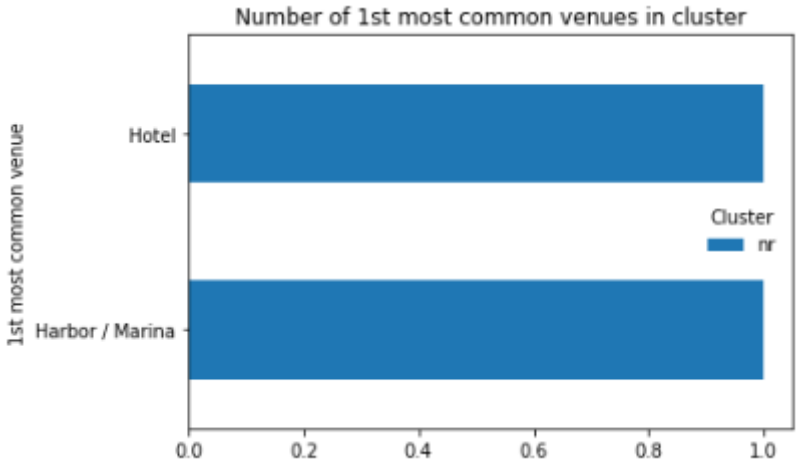
Cluster 3



Cluster 4



Cluster 5



Based on the analysis of the most common venues in each cluster, we can label the clusters:

- Cluster 1. Parks, bakeries, coffee shops
- Cluster 2. Primarily grocery stores
- Cluster 3. Hotel, cafe
- Cluster 4. Theme park
- Cluster 5. Harbor/marina

Let us examine the clusters in the Results section.

4. Results

So, which communes belong to each cluster?

Cluster 1: 23 communes. København, Aarhus, Odense, Frederiksberg, Viborg, Gentofte, Gladsaxe, Helsingør, Rudersdal, Lyngby-Taarbæk, Hvidovre, Ballerup, Tårnby, Syddjurs, Furesø, Bornholm, Rødovre, Brøndby, Herlev, Albertslund, Glostrup, Vallensbæk, Dragør.

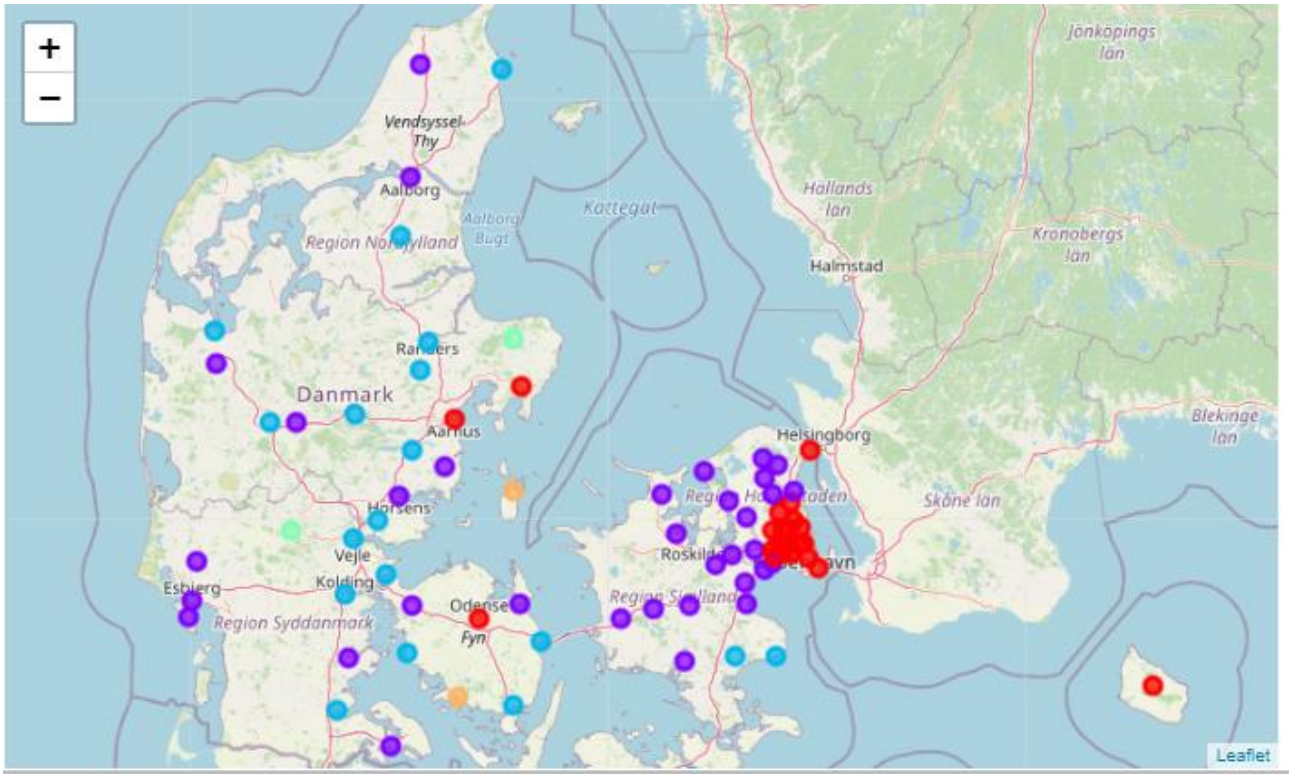
Cluster 2: 34 communes. Aalborg, Esbjerg, Horsens, Roskilde, Næstved, Slagelse, Sønderborg, Holbæk, Hjørring, Køge, Holstebro, Haderslev, Hillerød, Høje-Taastrup, Varde, Greve, Frederikssund, Egedal, Gribskov, Ikast-Brande, Fredensborg, Middelfart, Ringsted, Odsherred, Halsnæs, Sorø, Lejre, Allerød, Hørsholm, Kerteminde, Ishøj, Odder, Solrød, Fanø.

Cluster 3: 20 communes. Vejle, Randers, Kolding, Silkeborg, Herning, Guldborgsund, Skandeborg, Frederikshavn, Aabenraa, Svendborg, Fredericia, Favrskov, Hedensted, Assens, Faxe, Nyborg, Rebild, Stevns, Struer, Ærø.

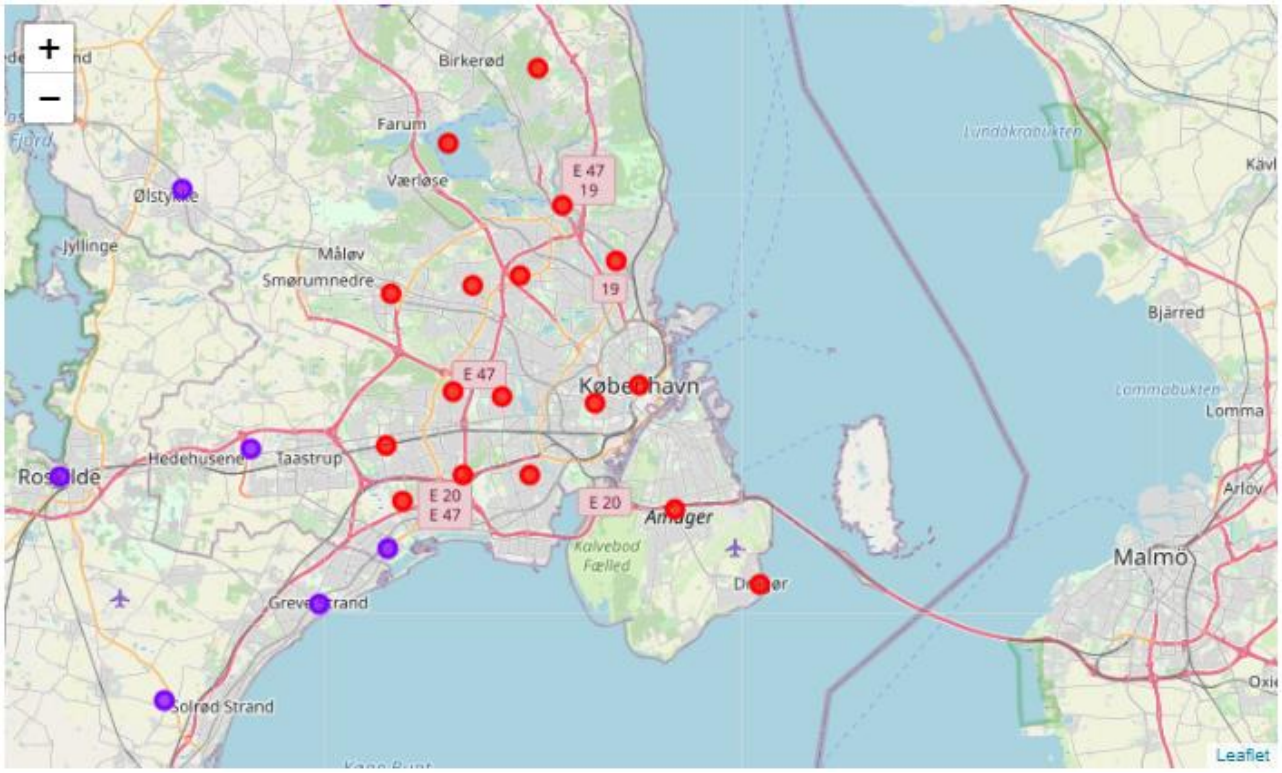
Cluster 4: Norddjurs, Billund.

Cluster 5: Faaborg-Midtfyn, Samsø.

The clusters are visualized below using Folium package:



In addition, zoom into the capital area:



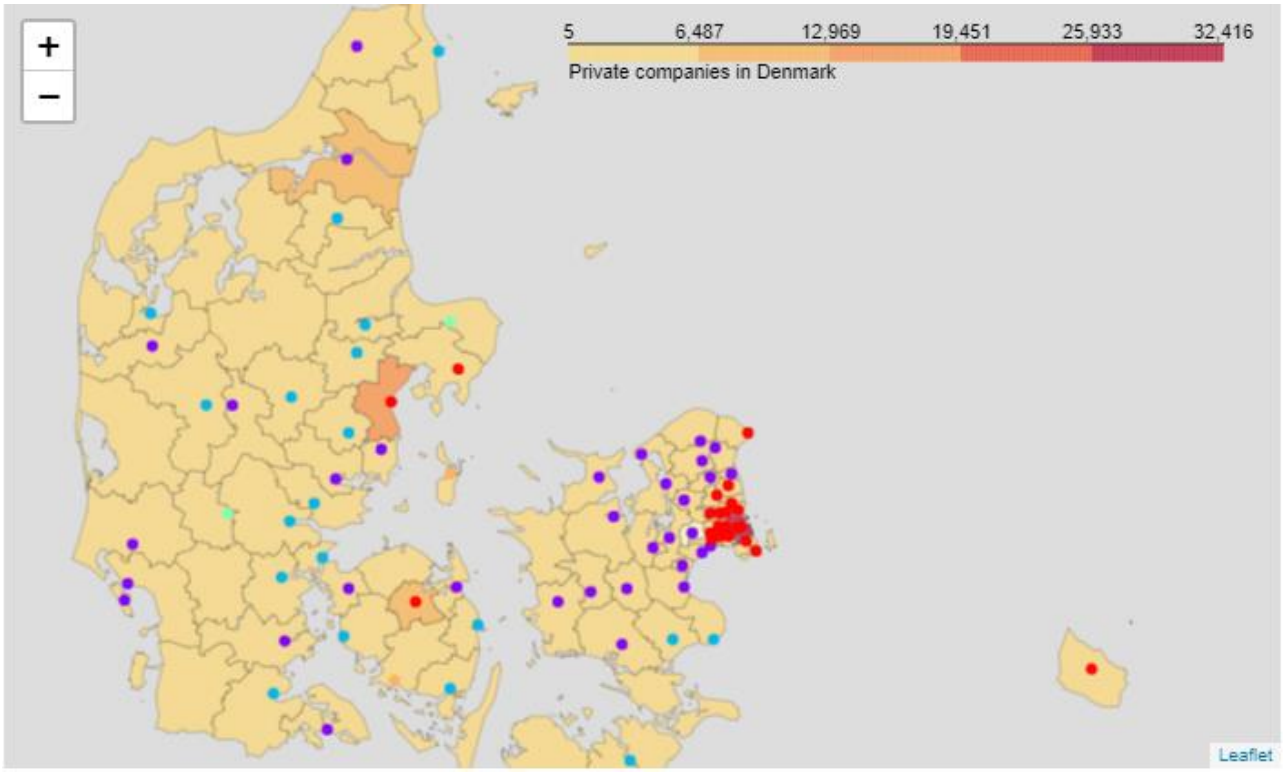
Moreover, we use data from Statistikbanken to compare communes by the:

1. Number of private companies
2. Number of museums
3. Income per family
4. Percentage of women that have a job.

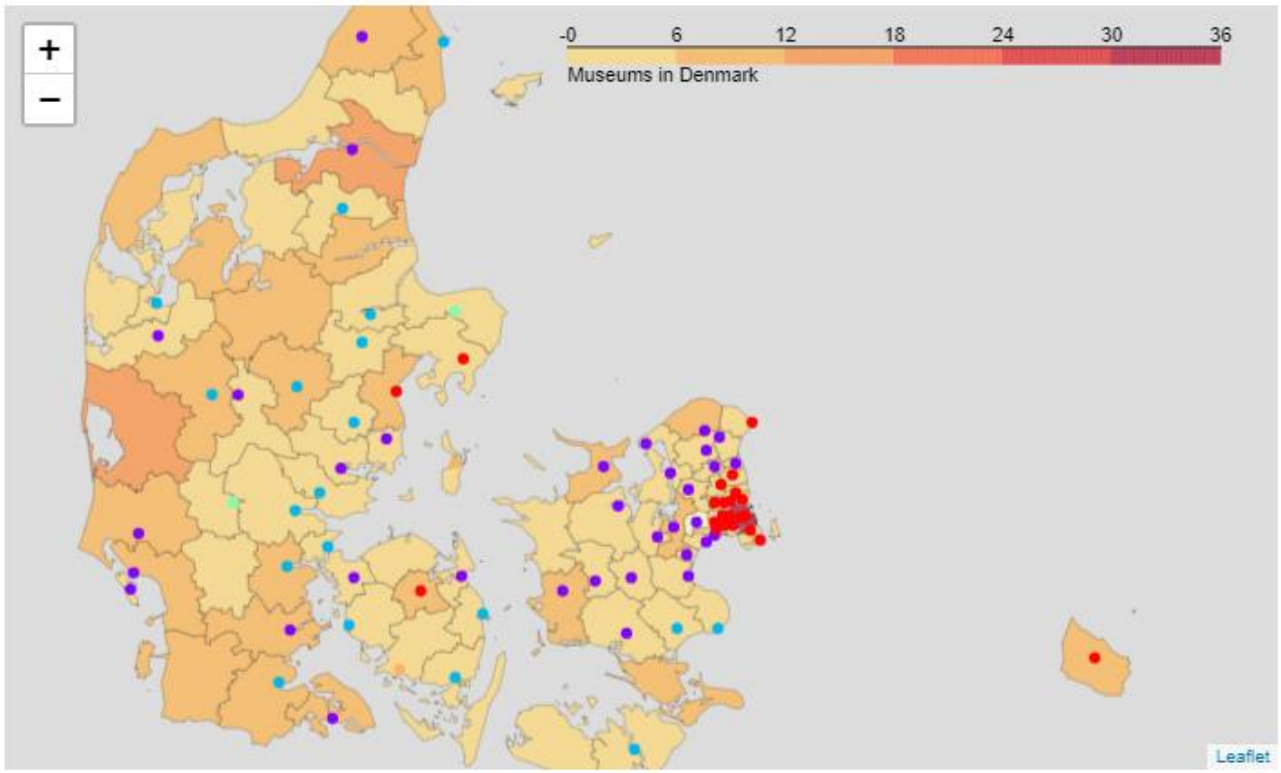
Below is a table that summarizes the collected data.

Navn	PrivateCompanies	Museums	IncomeFamilies	IncomeCouples	IncomeSingles	Divorces	FemaleWork	Km2	2019	indb/km2	Region	latitude	longitude
København	32415.0	36.0	459997.0	807874.0	298018.0	927.0	72.5	86.6	623404.0	7082.0	Hovedstaden	55.686724	12.570072
Frederiksberg	4842.0	6.0	567083.0	971731.0	350841.0	149.0	75.5	8.7	103940.0	12001.0	Hovedstaden	55.678016	12.532619
Dragør	621.0	4.0	728628.0	1004900.0	377713.0	20.0	79.5	18.3	14270.0	780000.0	Hovedstaden	55.592487	12.671995
Tårnby	1758.0	1.0	540538.0	794094.0	312576.0	72.0	76.9	66.1	42984.0	651000.0	Hovedstaden	55.628184	12.599763
Albertslund	1465.0	0.0	465264.0	704773.0	287515.0	47.0	68.8	23.2	27877.0	1196.0	Hovedstaden	55.658602	12.356285

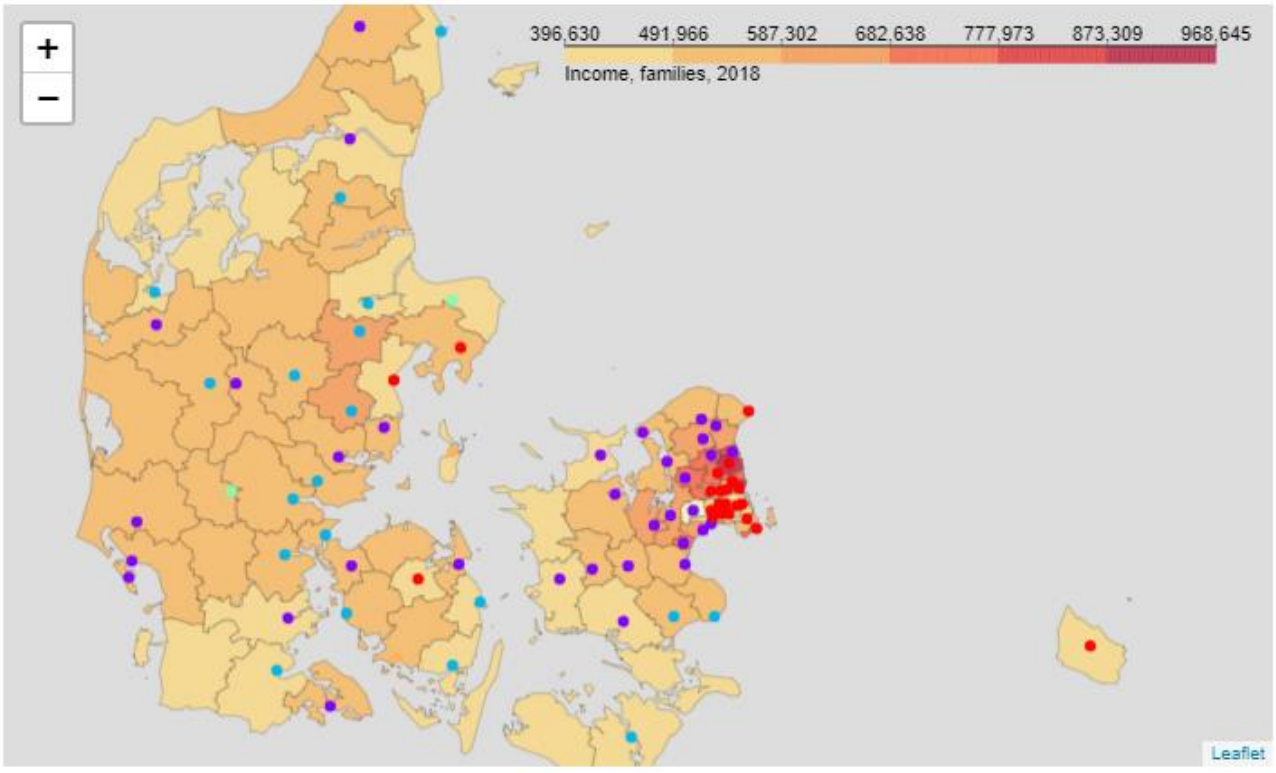
The data is visualized below using Cloropleth maps.



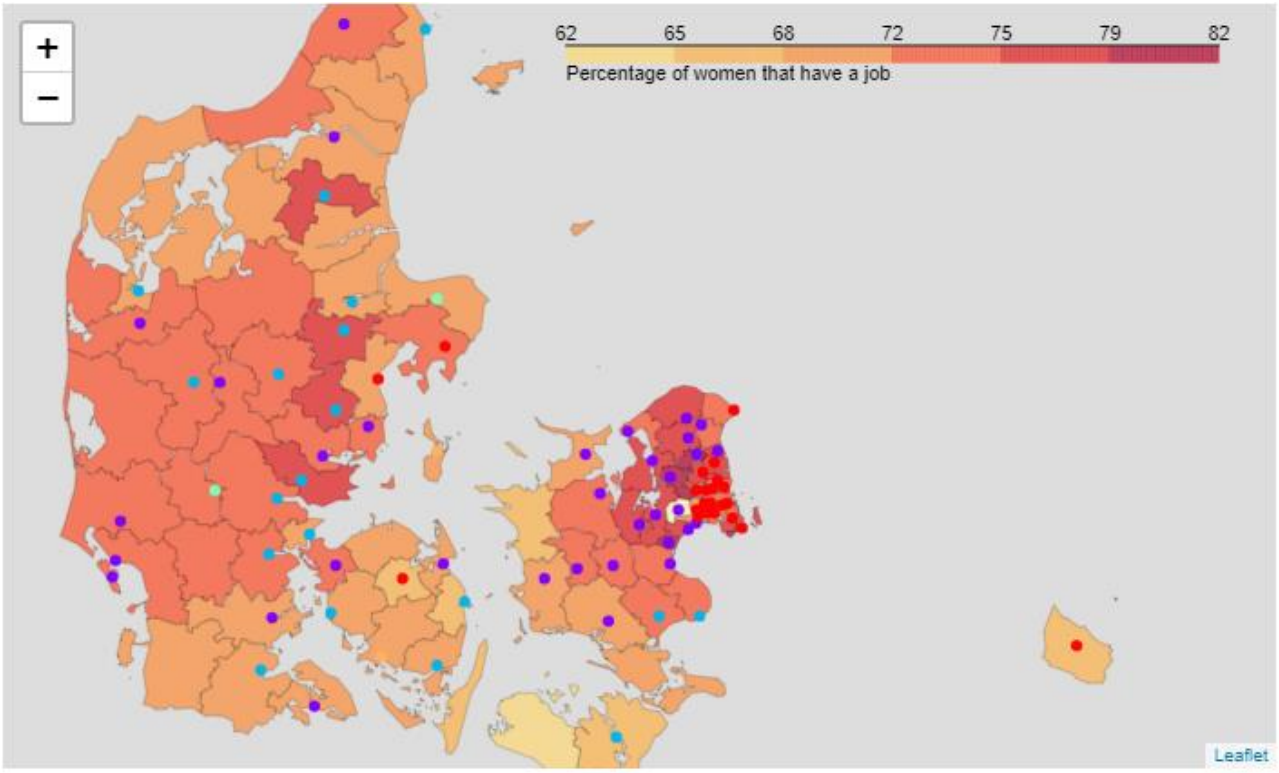
Figur 1. Number of private companies in Denmark, 2019, combined with the cluster labels



Figur 2. Number of museums in Denmark, 2019, combined with the cluster labels



Figur 3. Income, families, danish kroner DKK, 2018, combined with the cluster labels



Figur 4. Percentage of women that have a job, 2019, combined with the cluster labels

Let us discuss the obtained results in the following Discussion section.

5. Discussion

Denmark is a small country with surprisingly many communes. However, the population is rather non-uniformly distributed: many people live in the Copenhagen area. It's not possible to find a city/commune very close to Copenhagen, because Copenhagen Is the only of its kind in Denmark.

Still, according to Cluster 1, 22 communes are similar to Copenhagen. The similarity is expressed in number the diversity of found venues and in the number of parks, bakeries, coffee shops. In the first cluster, most of the communes are on Sealand, the same island where Copenhagen is located. There are only five communes that are located elsewhere: Aarhus, Odense, Viborg, Syddjurs, Bornholm.

Below, we only consider the communes in Cluster 1 that are not on Sealand.

Aarhus and Odense are the biggest cities of Denmark after Copenhagen are somewhat an obvious alternative to the capital. However, Viborg and Syddjurs are communes in Jutland that are similar to Copenhagen according to the algorithm. Bornholm is an island, which can be accessed by ferry and plane. So, for now we have five candidates to the new home:

1. Aarhus
2. Odense
3. Viborg
4. Syddjurs
5. Bornholm

Let us examine the Cloropleth maps which help us to decide.

Both Viborg and Syddjurs are not so far from Aarhus, where many private companies are located.

The same goes for the number of museums. In Aarhus, Odense, and Syddjurs the number of museums is lower than in Copenhagen area but not too low.

Interestingly, the map of family income suggests a rather high-income level in Syddjurs, compared to Aarhus, Odense.

Moreover, for the sake of curiosity, we have plotted a percentage of women that have a job in each commune. Again, the numbers are in favor of the Syddjurs commune!

To conclude, we recommend to all people that are considering moving from the capital area, to investigate the following communes: Aarhus, Odense, Syddjurs, Viborg.

6. Conclusion

The purpose of this study was to recommend an alternative commune to Copenhagen, for all people who have discovered the opportunity of working from home. Thus, all communes in Denmark are under the magnifying glass!

We used Wikipedia, Geocoder, and Foursquare to gather data about communes.

Then, we have used the Kmeans algorithm to cluster the communes of Denmark based on the venues data from Foursquare. The communes have been divided into five clusters. The first cluster contained 23 communes including Copenhagen.

Finally, we used the data from Statistikbanken to investigate the communes by the number of private companies and museums.

The data and the clusters have been visualized using Folium and Cloropleth maps.

7. References

1. Wikipedia page: https://da.wikipedia.org/wiki/Kommuner_i_Danmark_efter_indbyggertal
2. Foursquare API https://developer.foursquare.com/
3. Geocoder https://geocoder.readthedocs.io/
4. JSON file https://github.com/Neogeografen/dagi/blob/master/geojson/kommuner.geojson
5. Danmarks statistik https://www.dst.dk/da/Statistik/statistikbanken
6. Elbow method https://www.scikit-yb.org/en/latest/api/cluster/elbow.html