Clustering of Danish cities

“Where should we live?”

Introduction

- Denmark is a small country in Northern Europe with a very mobile workforce
- The associated moving of families are eased by being able to identify cities in new regions with similar characteristics
- We develop a tool for this task
- Businesses setting up a new branch are stakeholders as well

Data acquisition and cleaning

- Location data on Danish cities from:
http://www.tageo.com/index-e-da-cities-DK.htm

- Geographical data on Danish citites from:
https://raw.githubusercontent.com/Neogeografen/dagi/master/geojson/regioner.geojson

- Venue data for the 30 most common venues in a 10km radius around each city was obtained from Foursquare.

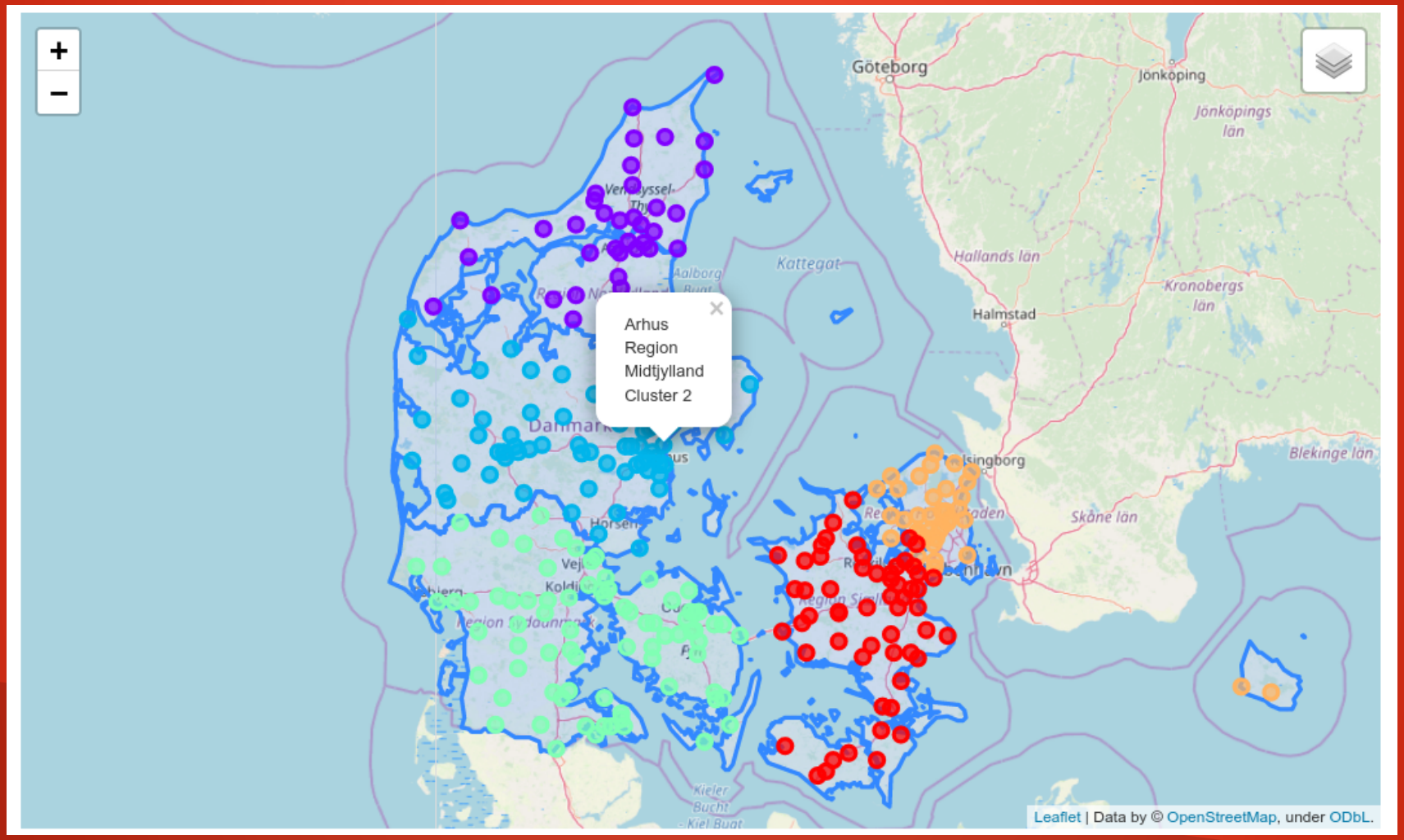
- The cleaned and combined data consisted of 298 cities with the following features
    - Population
    - Coordinates
    - Region
    - Frequencies of 30 most common venue categories

Exploratory data analysis

	Population	Latitude	Longitude
count	2.980000e+02	298.000000	298.000000
mean	1.330973e+04	55.857162	10.506980
std	6.551004e+04	0.654665	1.311002
min	1.900000e+03	54.650000	8.130000
25%	2.700000e+03	55.412500	9.555000
50%	3.850000e+03	55.750000	10.150000
75%	8.500000e+03	56.200000	11.787500
max	1.089700e+06	57.730000	15.150000

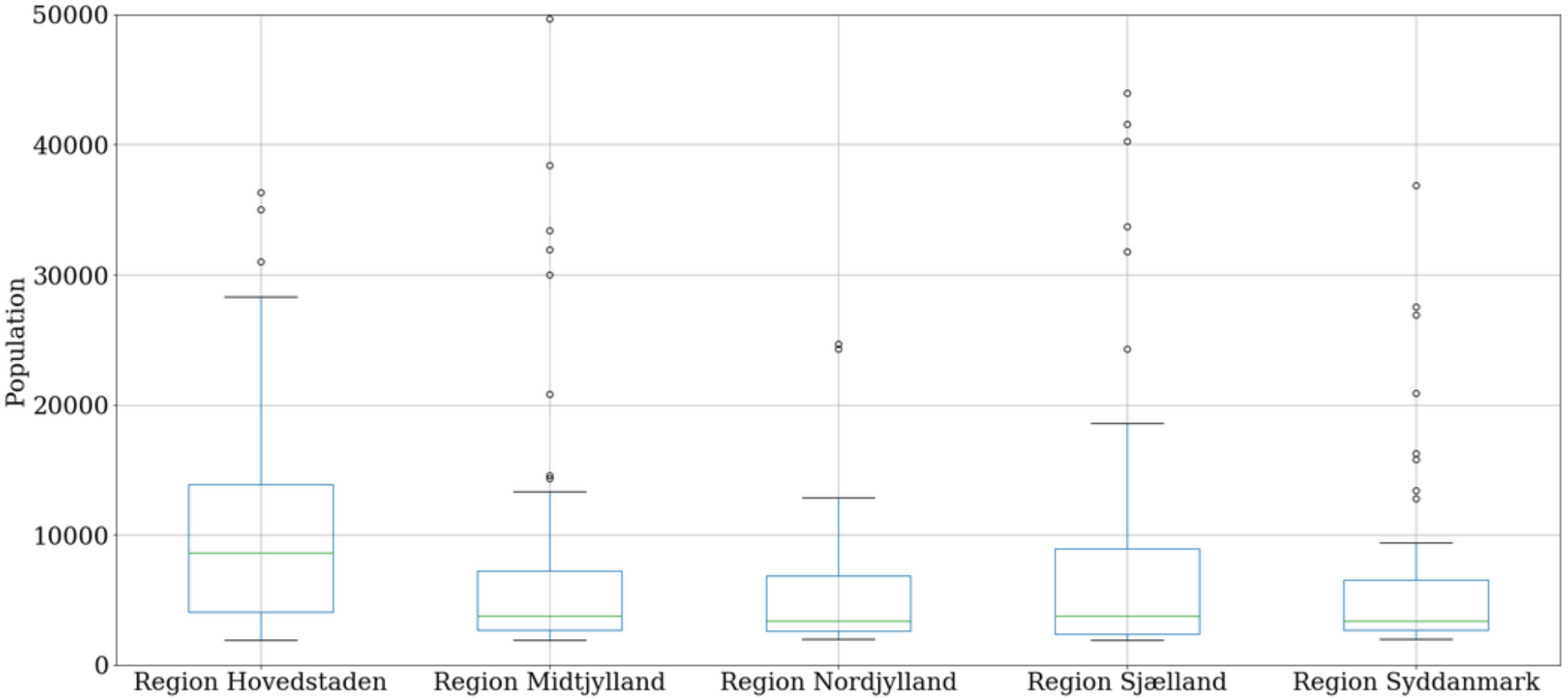
Summary statistics for the population data

Exploratory data analysis



Cities marked on a map according to their region (colors)

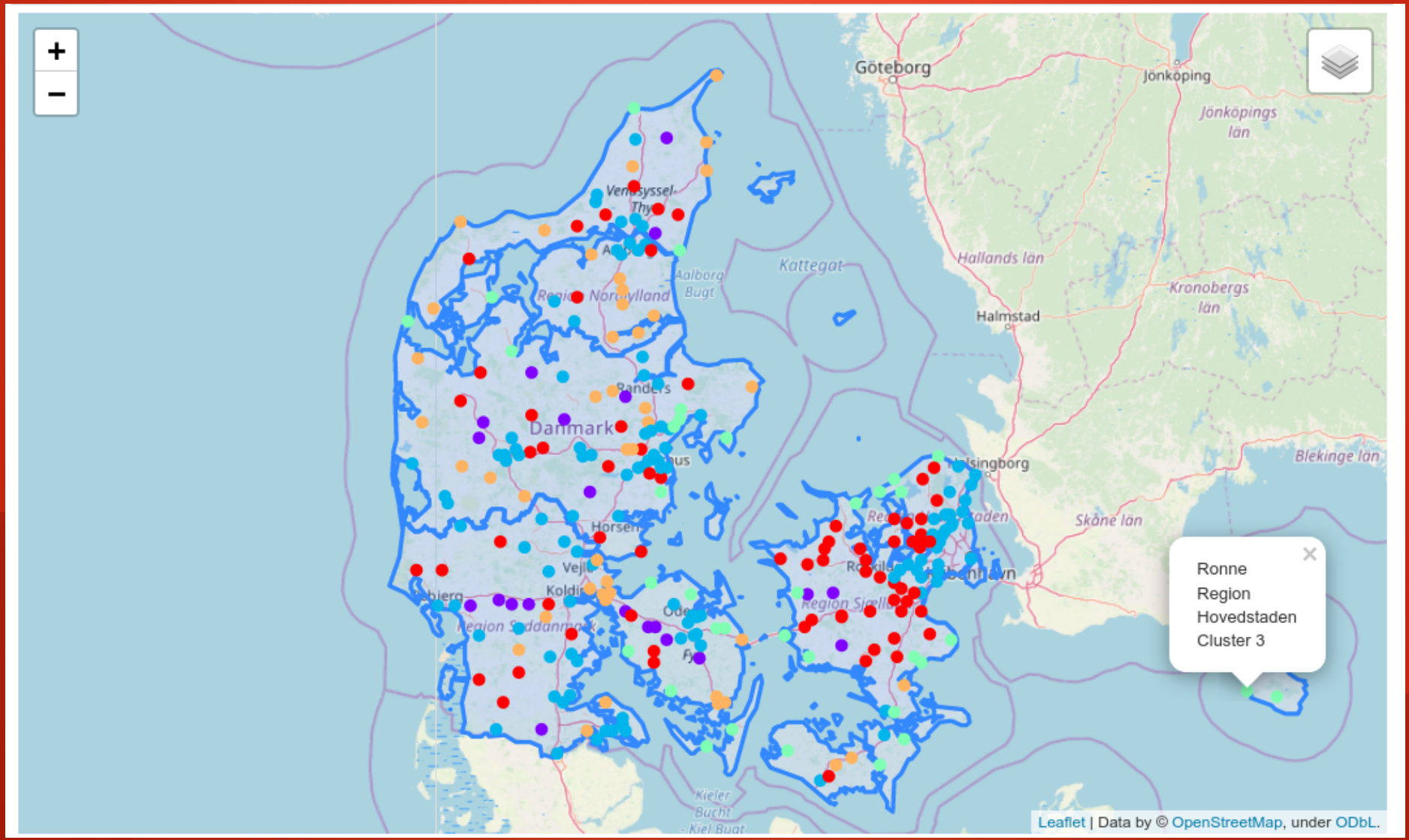
Exploratory data analysis



Distribution of cities in each of the five Danish regions.

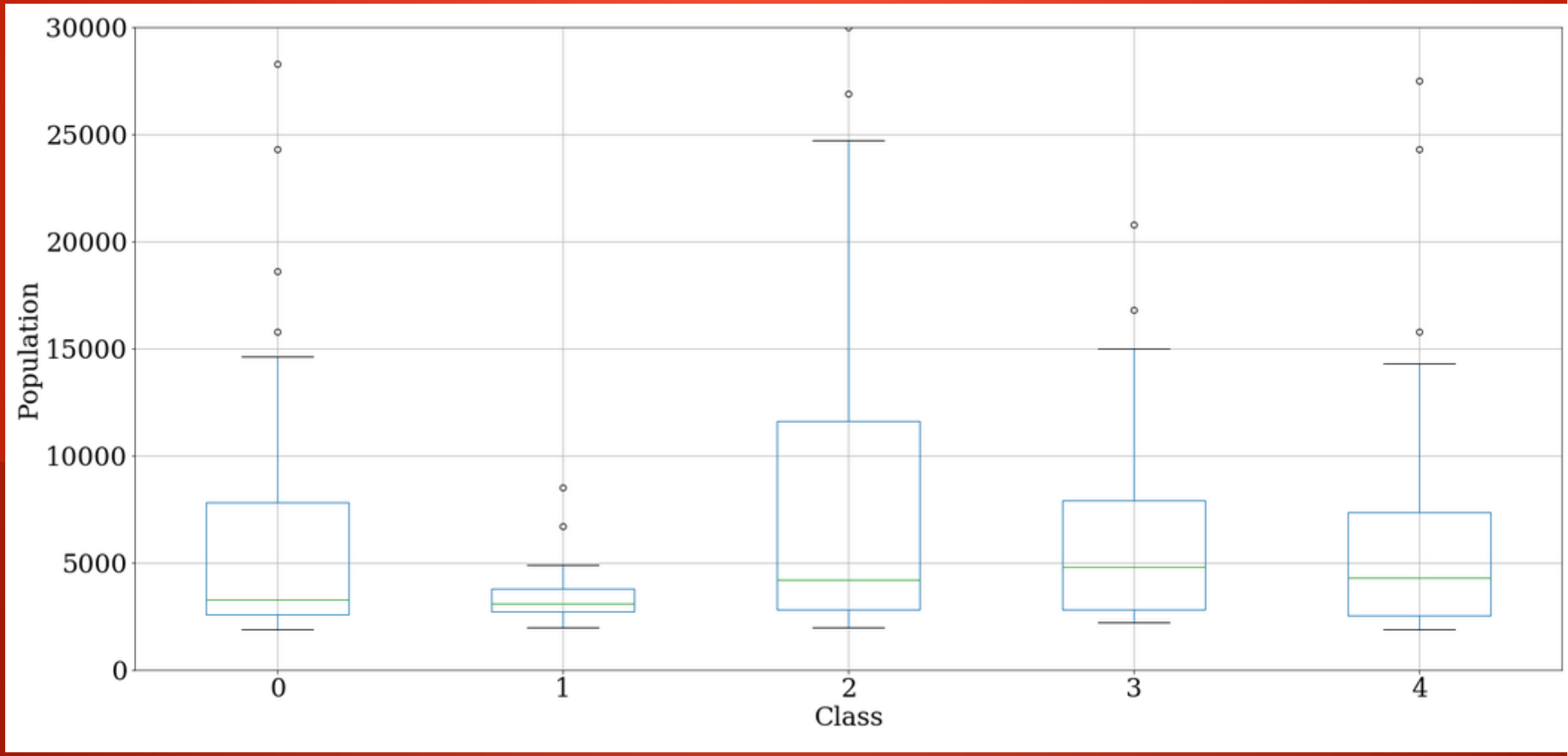
Clustering

Method: K-means with five clusters (marker colors on map)



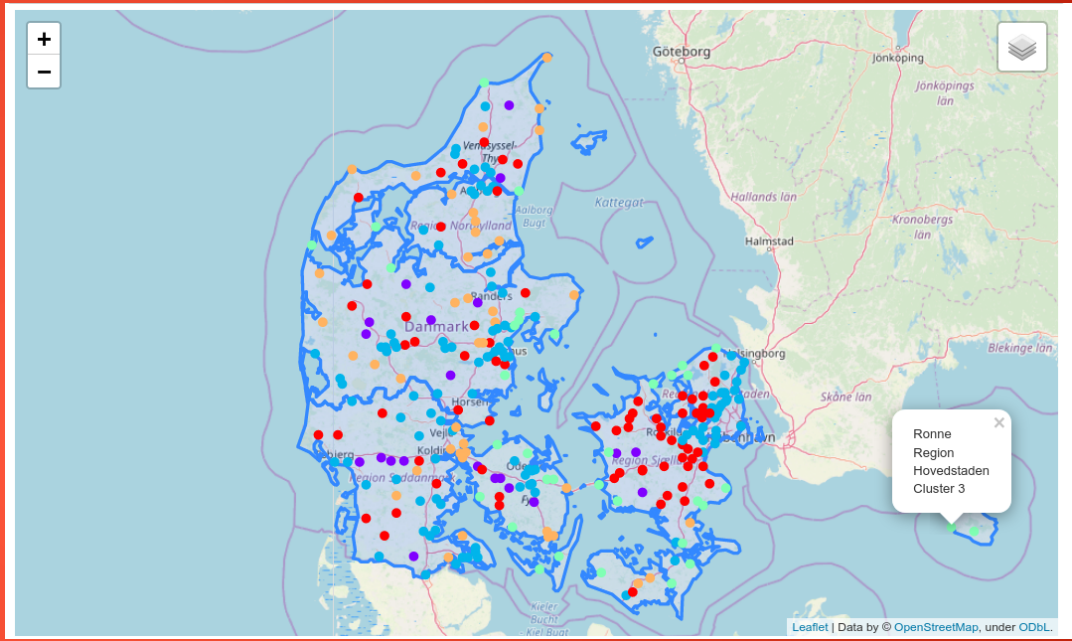
Clustering

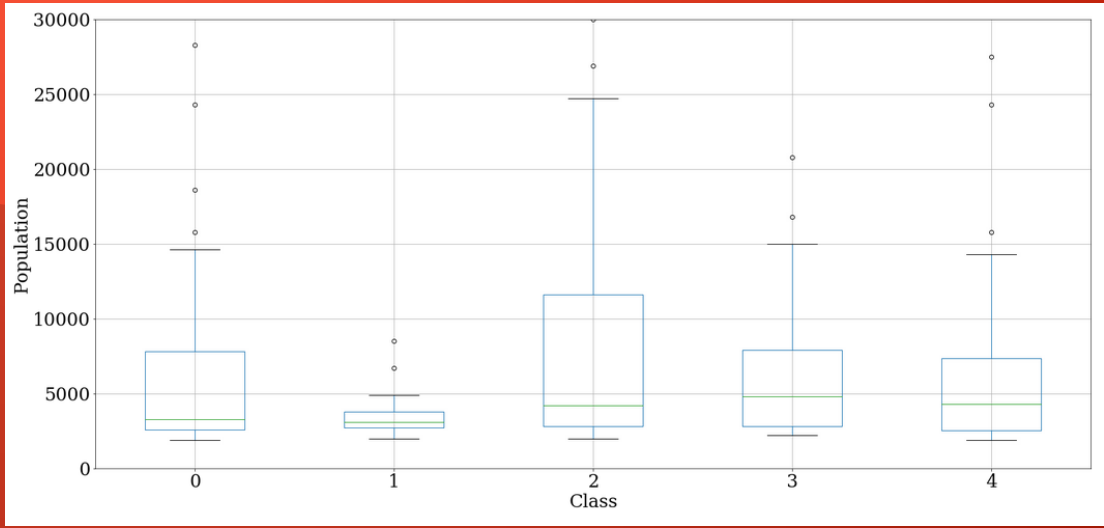
Population distributions within each of the five classes



Catagorization of clusters

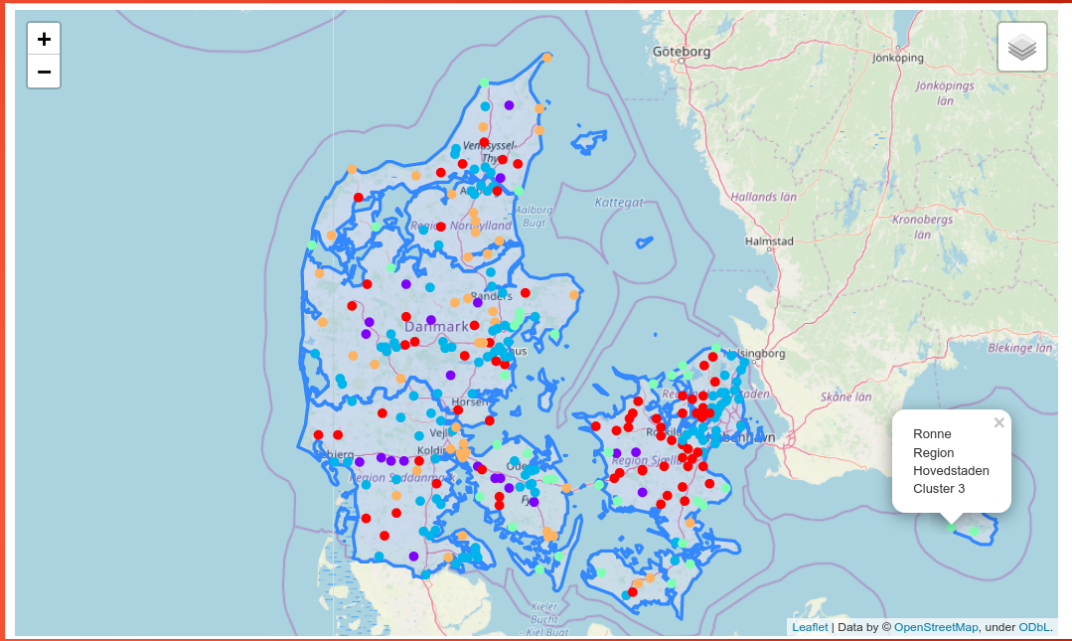
Cluster 0 (red markers):
72 towns with a mean population of 7272. The by far dominating venues are grocery stores. This cluster represents smaller towns from where people mainly commute to a larger city whenever they want to eat out of attend cultural or other activities. The towns themselves are mainly used for living and shopping of everyday goods and food. From Fig. 4, we note that these towns are predominately located in rings around larger cities (Cluster 2, blue markers), which offer a rich variety of venue offerings in close driving distance.

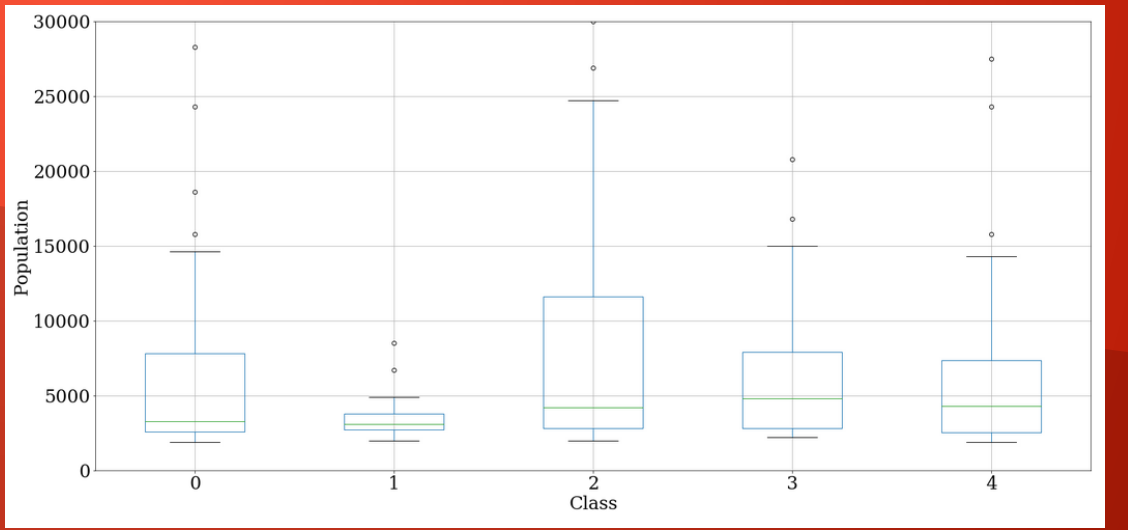




Catagorization of clusters

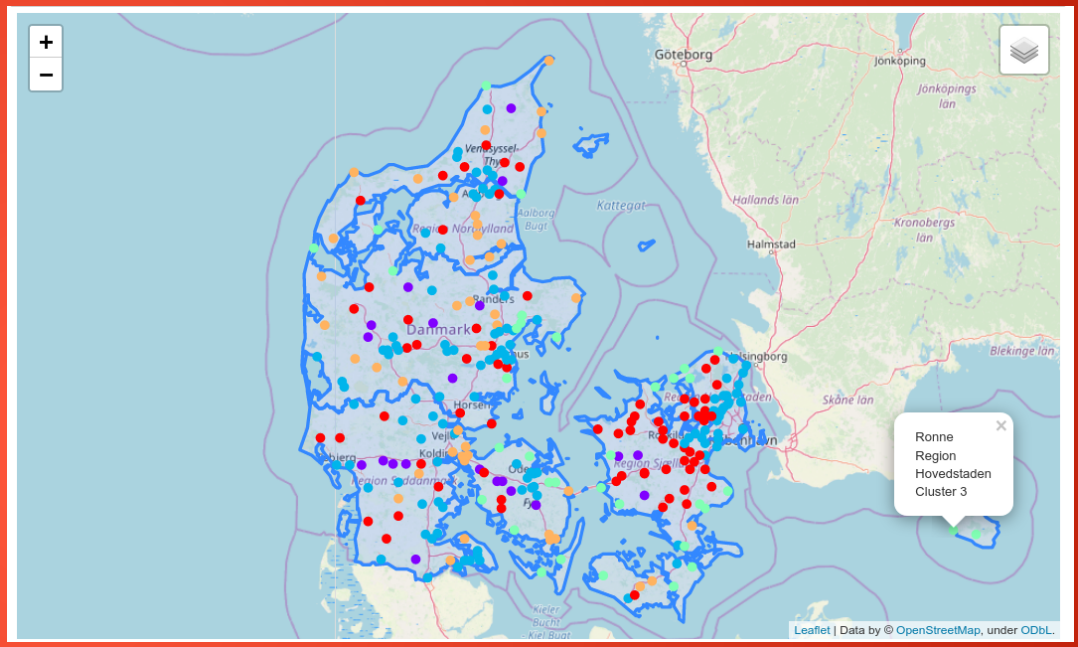
Cluster 1 (purple markers):
21 towns with a mean population of 3542. These are small towns which all have train stations. The cities in this cluster are characterized as commuter towns where people in general need to commute whenever they want to go to work, eat out or enjoy any cultural activities.

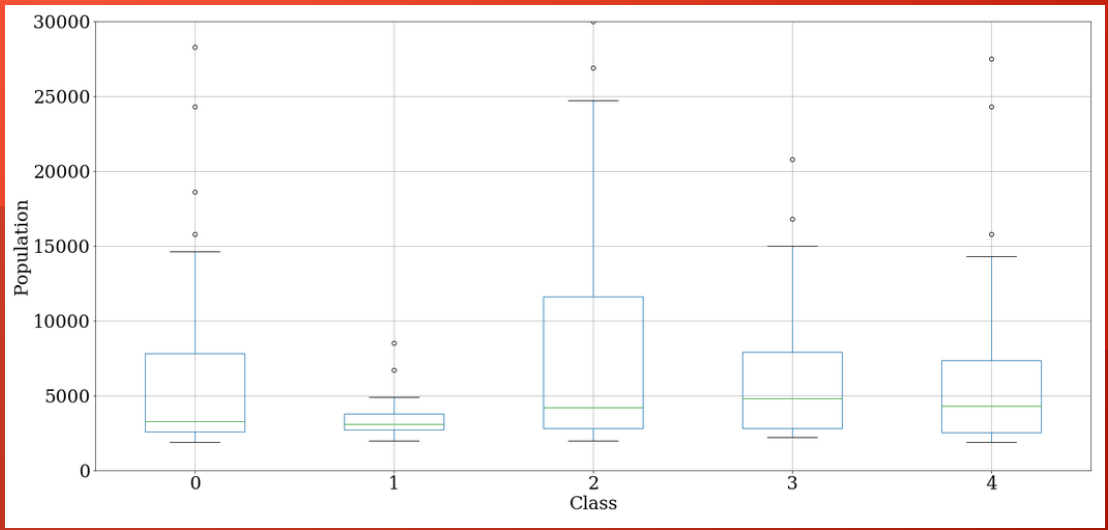




Catagorization of clusters

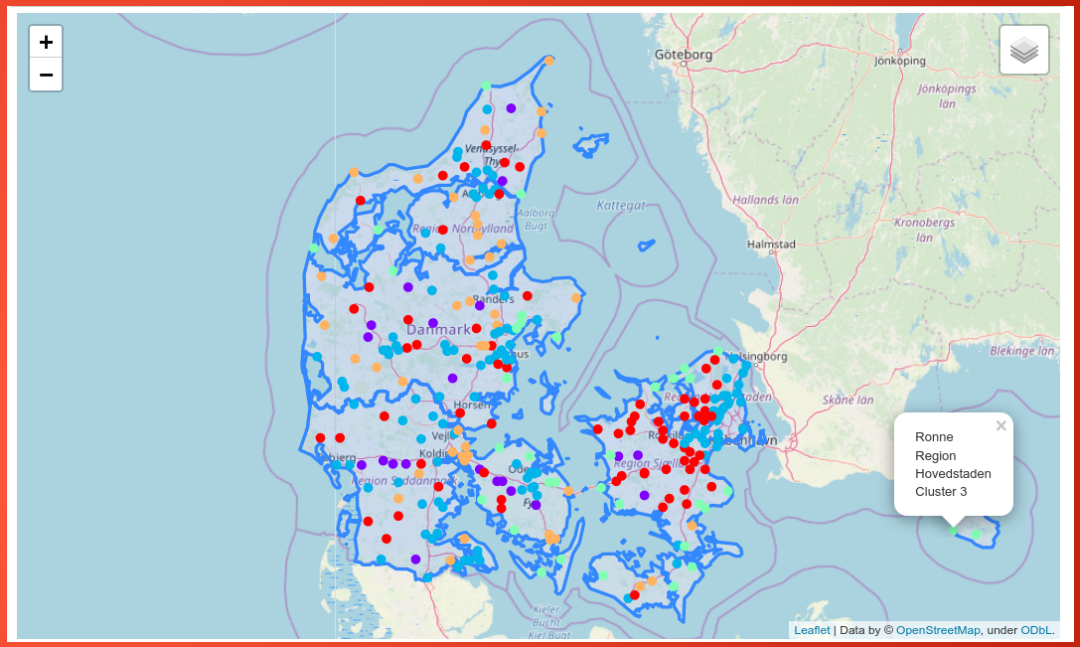
Cluster 2 (blue markers):
110 cities with a mean population of 24338. These are larger cities with a wide variety of offering for activities such as fitness centers, parks, golf courses, pools as well as many different cafés and restaurants.

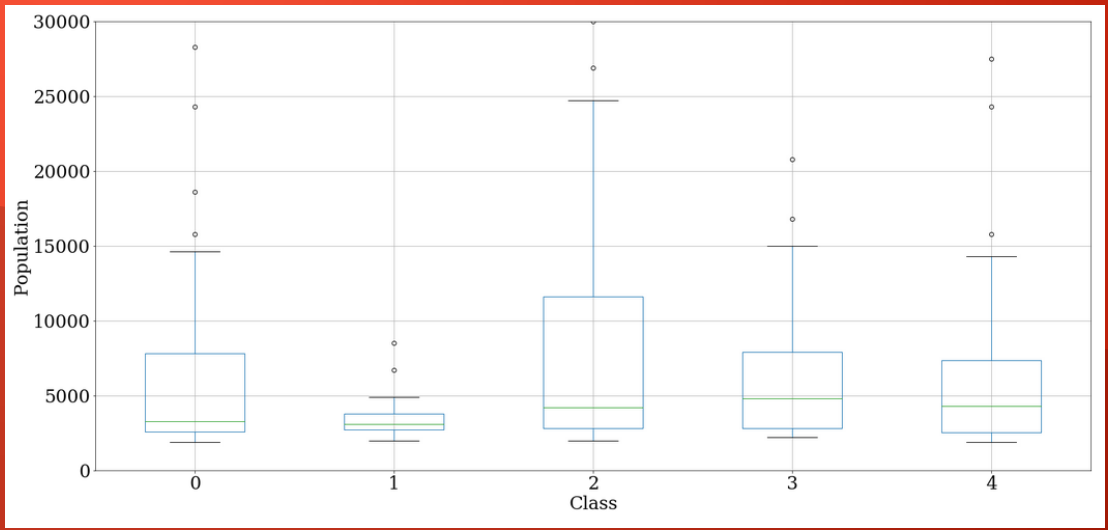




Catagorization of clusters

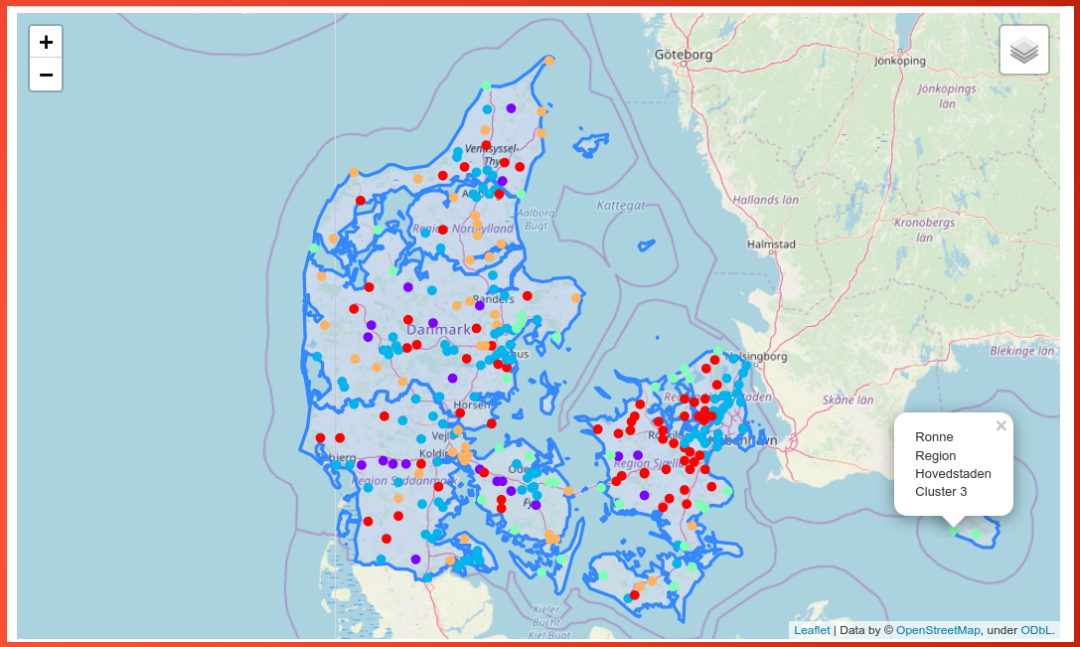
Cluster 3 (green markers):
35 towns with a mean population of 6440. These are smaller towns, very similar to those in Cluster 0 with grocery stores as the dominating venues. The difference is that the cities in Cluster 3 all have harbors or marinas, and many of them feature beaches. Looking af Fig. 4 we note that these cities are all located at the seaside.

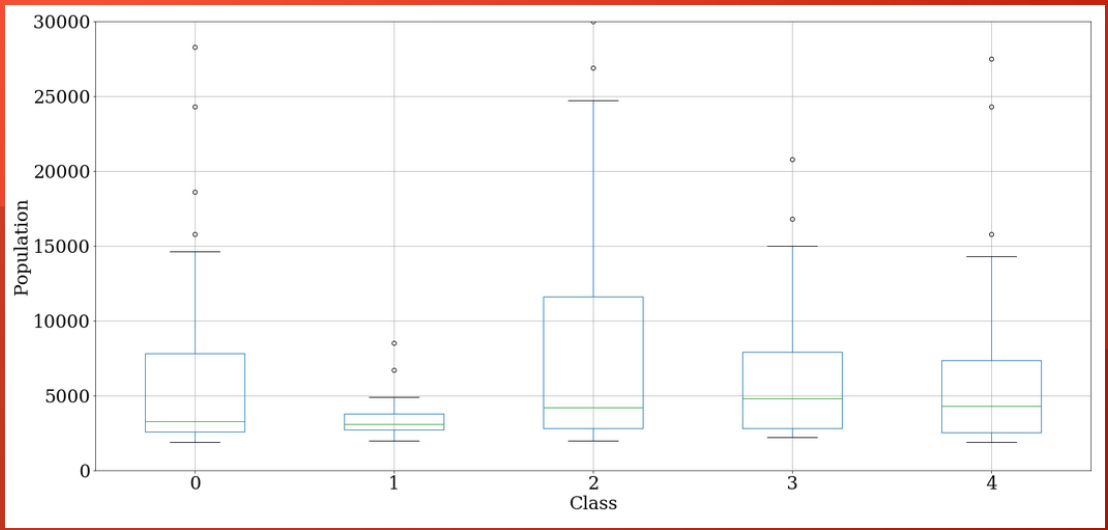




Catagorization of clusters

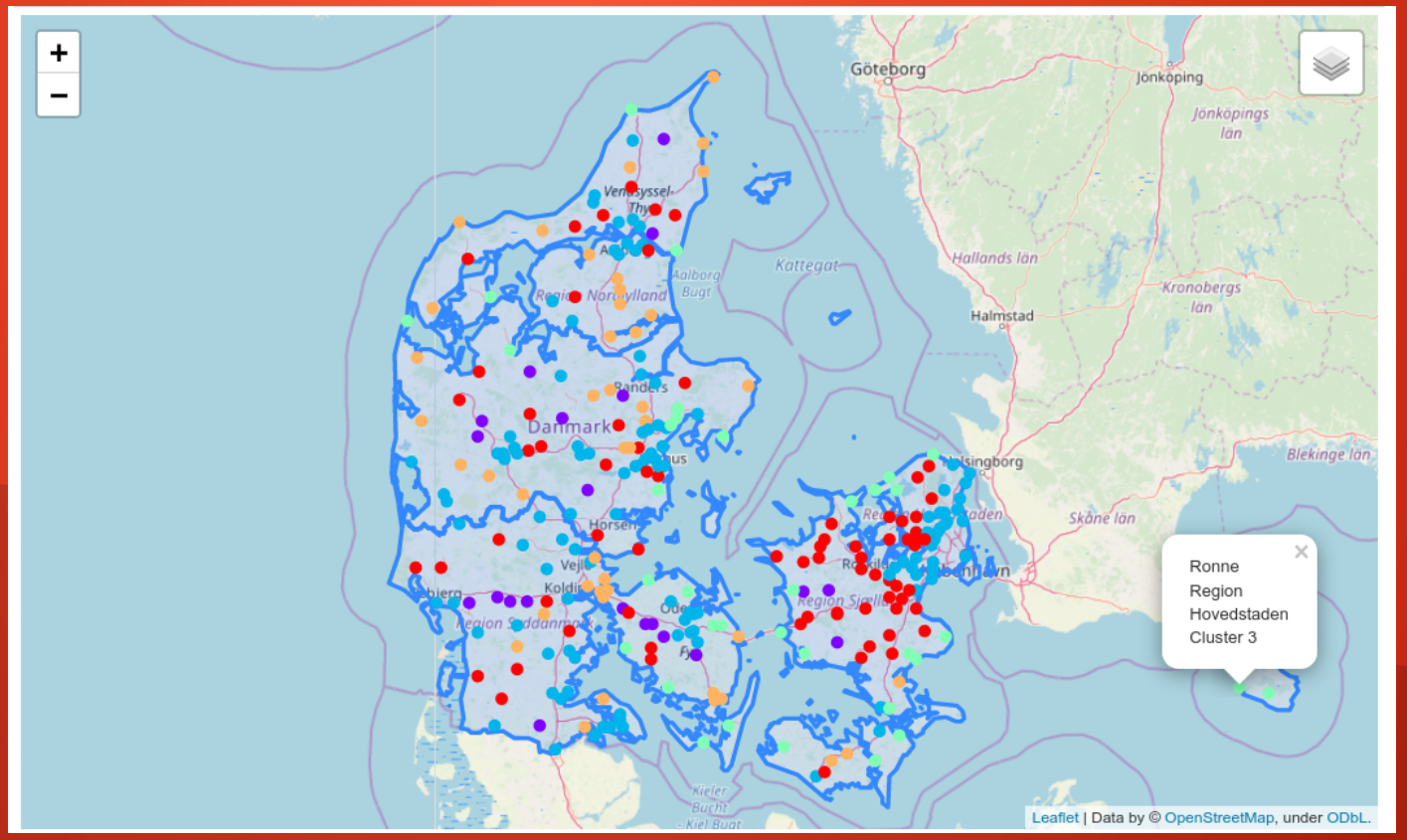
Cluster 4 (orange markers):
43 towns with a mean population of 6923 . These are smaller towns with some variety amongst them. One common thing is that they almost all have hotels as their most common venue and many of them feature gas stations. At the same time we note that there are many attractions (aquariums, historic sites, zoos, theme parks, etc.) among their common venues. From this, we infer that this cluster represents (mainly) cities where tourism and visitors play a large role.





Conclusion

- We have produced an interactive map which allows cities of similar type to be identified across the Danish regions
- The cities are divided into 5 classes, each of which have a distinct description



Outlook

- Interesting include different data (occupation, educational background, mean income and sports activities)
- Extend this kind of analysis and tool beyond Denmark