Final Report
Capstone Project – The Battle of Neighbourhoods
Analysis of Luxembourgish Population Behaviour

Introduction

Luxembourg is one of the founding countries of the European Union and counts approximately 600.000 habitants. Renowned for its high-income market economy featuring moderate growth, low inflation and high level of innovation, Luxembourg has always been an attractive immigration target over the course of the last decades. Luxembourg is divided in 12 cantons as areas of local government, which in turn are subdivided into 102 communes / municipalities.

In the following report, demographic analysis of the Luxembourgish cantons and municipalities will be performed to gain further understanding of the overall development within the country at a local view. The demographic analysis is augmented by utilization of venue descriptors, characteristic for each region.

Ultimately the aim of this report is to provide an overview on the Luxembourgish State, finding differences and similarities within municipalities and cantons, and providing guideline for potential immigrants to choose the optimal residence place based on the analysed metrics.

Data Section

In order to perform the demographic analyses, several data have to be collected on a variety of topics.

This website belongs to a Luxembourgish statistics portal that performs regular surveys on local and countrywide scale. It allows to collect data from several time periods and everything is free to download in xls-format.

For this study, following data sets were downloaded:

- Population per canton and municipality (1995-2020)
- Population Density per canton and municipality (1995-2020)
- Unemployment rates per canton and municipality (per yearly document 2010-2018)
- Net Migration within cantons and municipalities (2010-2018)
- Victims in accidents per canton and municipality (per yearly document 2010-2018)

Not all data sets are available for the same time period, but there are overlaps in the time periods! Analysis of the subsets will be performed and, given certain time periods, correlations between subsets will be analysed.

The population and population density data are available for the longest time periods, allowing to analyse if the population is increasing/decreasing and in which fashion it does so (linear, exponential, etc…)

The unemployment rates allow to identify certain regions in Luxembourg where unemployment is more present. As rates are always given with respect to the potentially working population (not children or pensioners), they can be easily compared between cantons/municipalities.

The net migration is defined as the proportion of people moving in to and out of the canton/municipality, giving a good estimate on how subsets of the population are changing there.

Finally, the changes in the amount of accident victims per canton and municipality allow to identify the cantons with the highest fatality risk and allow to determine overall trends to extrapolate beyond 2018.

Foursquare Location data will be used to complement the demographic data and try to assess similarities between municipalities and cantons.

Methodology

First, all xls-files need to be imported and processed (*Section A* in script). This includes separating the cantonal from the municipal data and for the unemployment and accident victim data extracting and merging the data in all annual files.

This leaves us with several data sets in hands:

1) Cantonal Population + Municipal Population (1995-2020)
2) Cantonal Population Density + Municipal Population Density (1995-2020)
3) Cantonal Unemployment Rate + Municipal Unemployment Rate (2010-2018)
4) Cantonal Net Migration + Municipal Net Migration (2010-2018)
5) Cantonal Accident Victims + Municipal Accident Victims (2010-2018)
6) Cantonal 2018 dataset and Municipal 2018 dataset (as latest common year)

Then, using the *geolocator* library in python, each longitude and latitude for Luxembourg as a country, the cantons and each municipality is generated and merged with the respective 2018 datasets (*Section B* in script). Using *folium*, a map of Luxembourg is generated and every canton and municipality is plotted with their name and respective population density (see Figure 1)

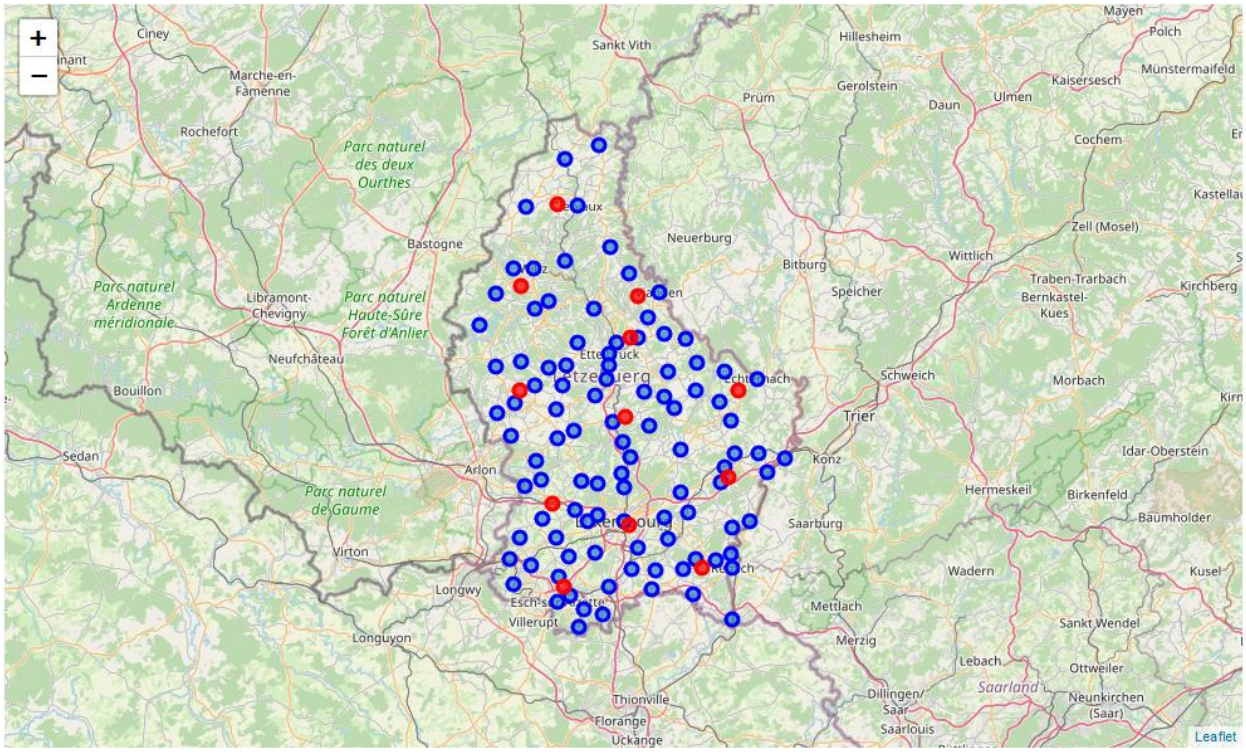

Figure 1 folium interactive map of Luxembourg containing all 12 cantons (red dots) and 102 municipalites (blue dots).

In a next step (*Section C* in the script), the focus was set on the cantonal datasets as they intrinsically include the information of each municipality already and are handy for quick assessment of the countrywide demographic situation.

As depicted in Figure 2, all cantons show a net increase of population over the course of the last 25 years. If we assume a linear increase over the course of the last 25 years, we can define a growth rate (see Table 1).

Surprisingly, this growth rate is relatively slow for most cantons. Esch and Luxembourg show the steepest increase; these cantons are by far (~100.000) the most populated amongst the 12 cantons. Capellen is the third most populated canton, reaching >50000 people in 2020.

Even though, Grevenmacher currently has less people than Diekirch (30887 vs 33782), the growth rate is higher (450 vs 388 people/year), hinting at a preferential population (from net migration / natural balance) of the first canton over the latter.

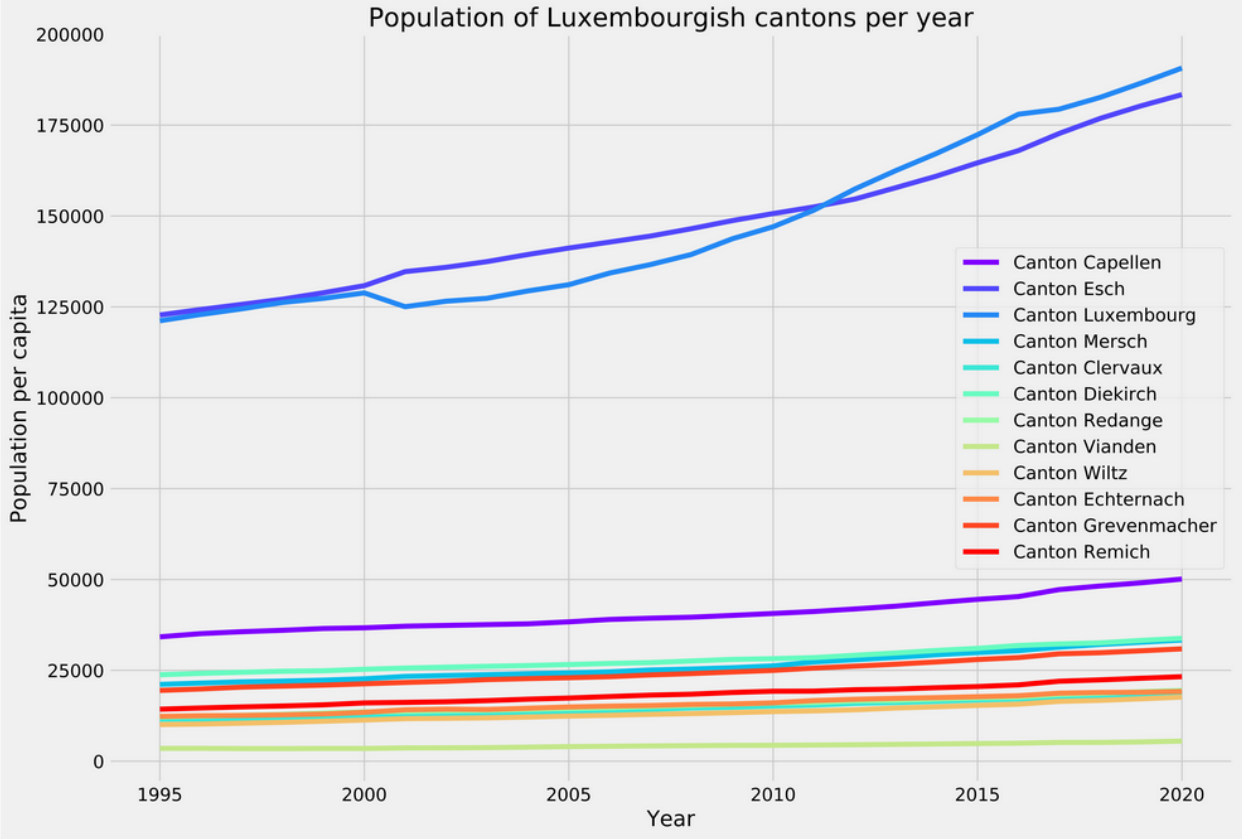

Figure 2 Population per capita of Luxembourgish cantons per year from 1995-2020.

The evolution of net migration in Luxembourgish cantons from 2010-2018 is depicted in Figure 3 and further analysis is based on values in Table 1.

The highest changes in migration during the period of 2010-2018 unsurprisingly occur for the most populated cantons, namely Esch and Luxembourg.

The other 10 cantons have a low standard deviation (<150 people), indicating that there is little fluctuation of incoming/outgoing people with respect to the mean over the years. The mean is close to the respective predicted growth rate (<6%). This highlights that net migration is probably the most contributing factor to population growth (R^2 for those 10 cantons = 0.92 as seen in Figure 4)! If we trace a trendline throughout 2010-2018, we observe negative slopes corresponding to a situation in favour of emigration: this is the case for Luxembourg, Mersch, Echternach and Grevenmacher. Vianden seems to stagnate with a slope close to 0.

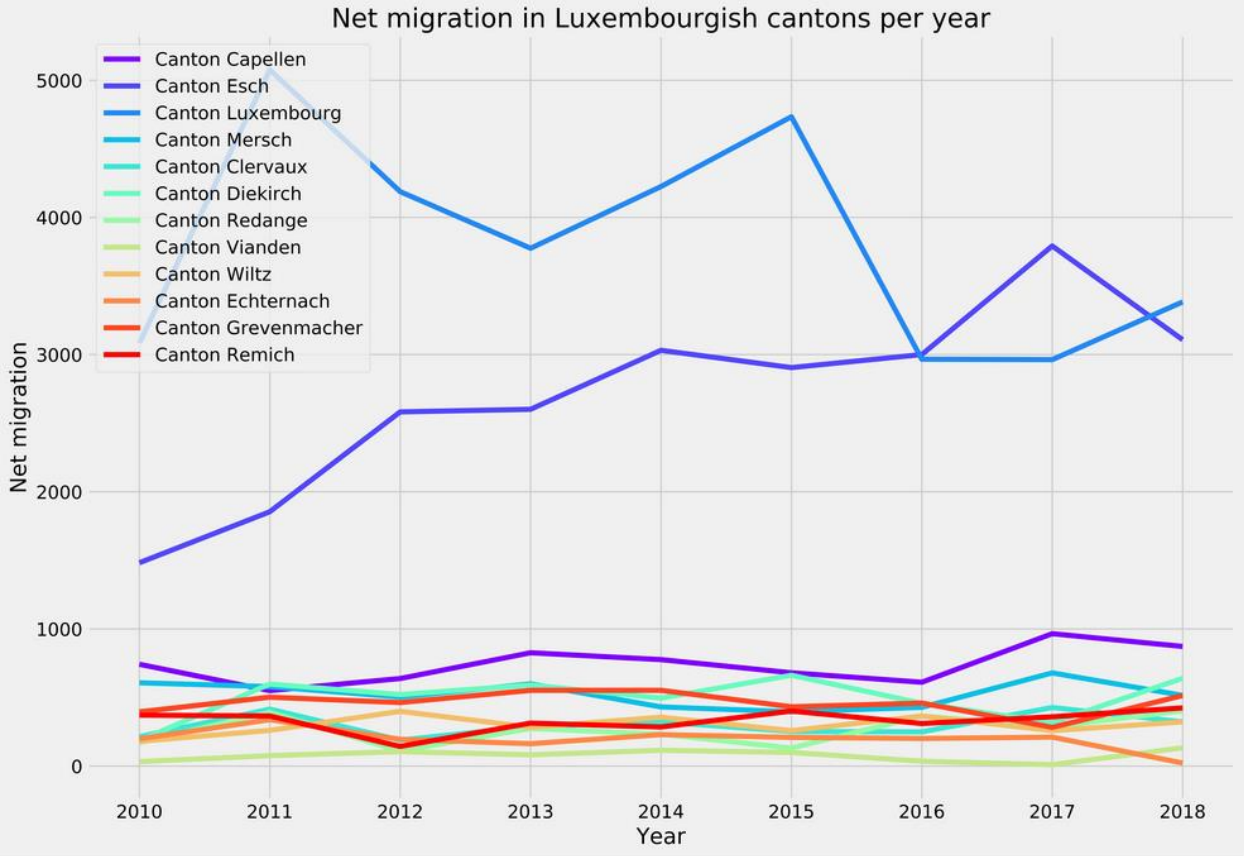

Figure 3 Evolution of net migration in Luxembourgish cantons from 2010-2018.

Cantons	pred. Growth Rates	Std Deviation of Net Migration	Mean of Net Migration	Percent rate (Mean of Net Migration vs Growth rate)	Trendline Slope
Canton Capellen	575	126	740	5.873016	26
Canton Esch	2329	651	2706	4.156682	224
Canton Luxembourg	2881	738	3821	5.177507	-110
Canton Mersch	481	90	526	5.844444	-7
Canton Clervaux	297	78	297	3.807692	9
Canton Diekirch	388	148	496	3.351351	15
Canton Redange	282	104	262	2.519231	16
Canton Vianden	82	39	76	1.948718	1
Canton Wiltz	288	64	297	4.640625	7
Canton Echternach	287	76	196	2.578947	-17
Canton Grevenmacher	450	80	461	5.762500	-5
Canton Remich	343	78	330	4.230769	10

Table 1 Overview of selected indicators extracted from the cantonal population and net migration datasets.

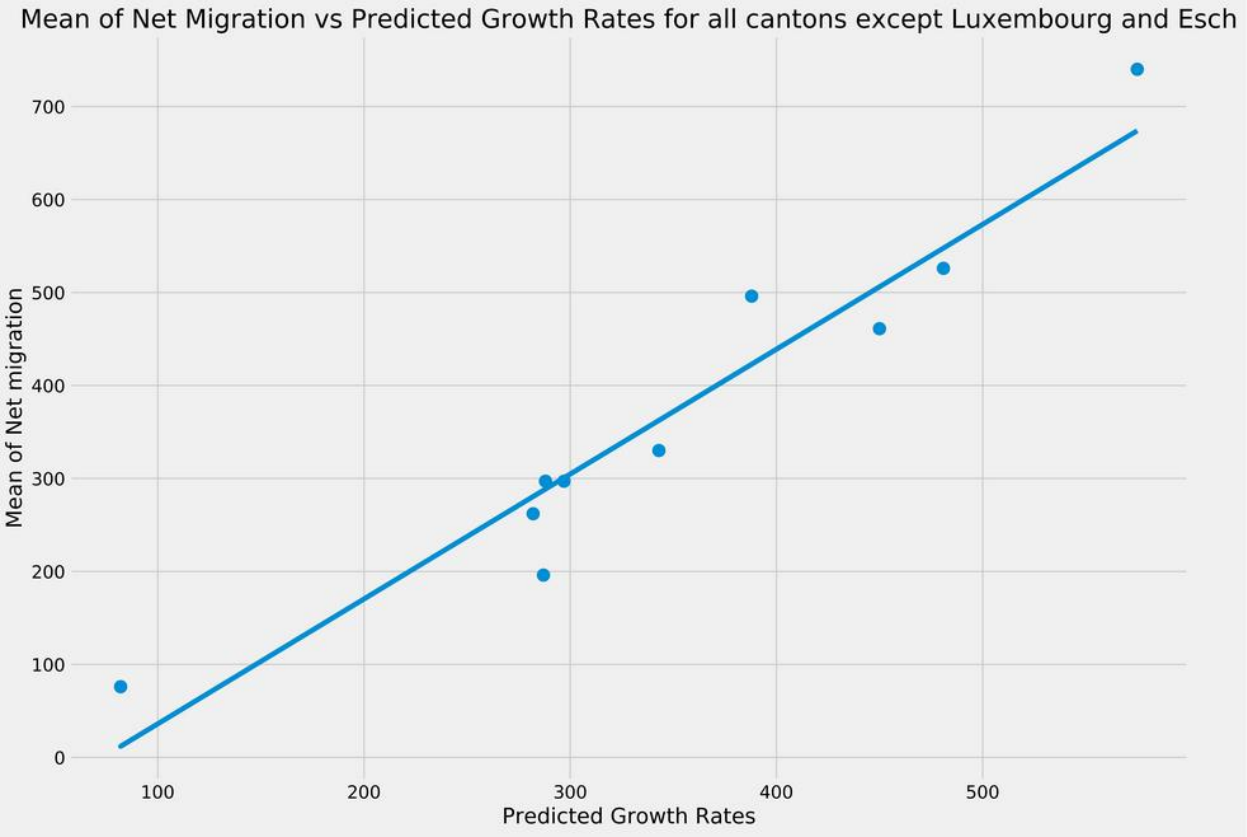

Figure 4 Correlation between the mean of net migration and predicted growth rates in the period of 2010-2018 for all cantons except Luxembourg and Esch.

Next, the amount of victims resulting from accidents in each canton from 2010-2018 was analysed (see *Figure 5*).

The mean of accident victims in 2010-2018 fortunately follows a steady trend, as can be seen through the slopes that are around 0 and the small standard deviations (see Table 2).

The highest number of victims can be found in cantons having the largest population. However, Esch, Grevenmacher, Clervaux and Wiltz fall out of this trend, having higher victim means compared to other cantons that are more populated. This indicates that the number of accident victims does not solely depend on the population in the canton, but also on other factors (likely number of vehicles, number of roads, how much the roads are used, etc..)

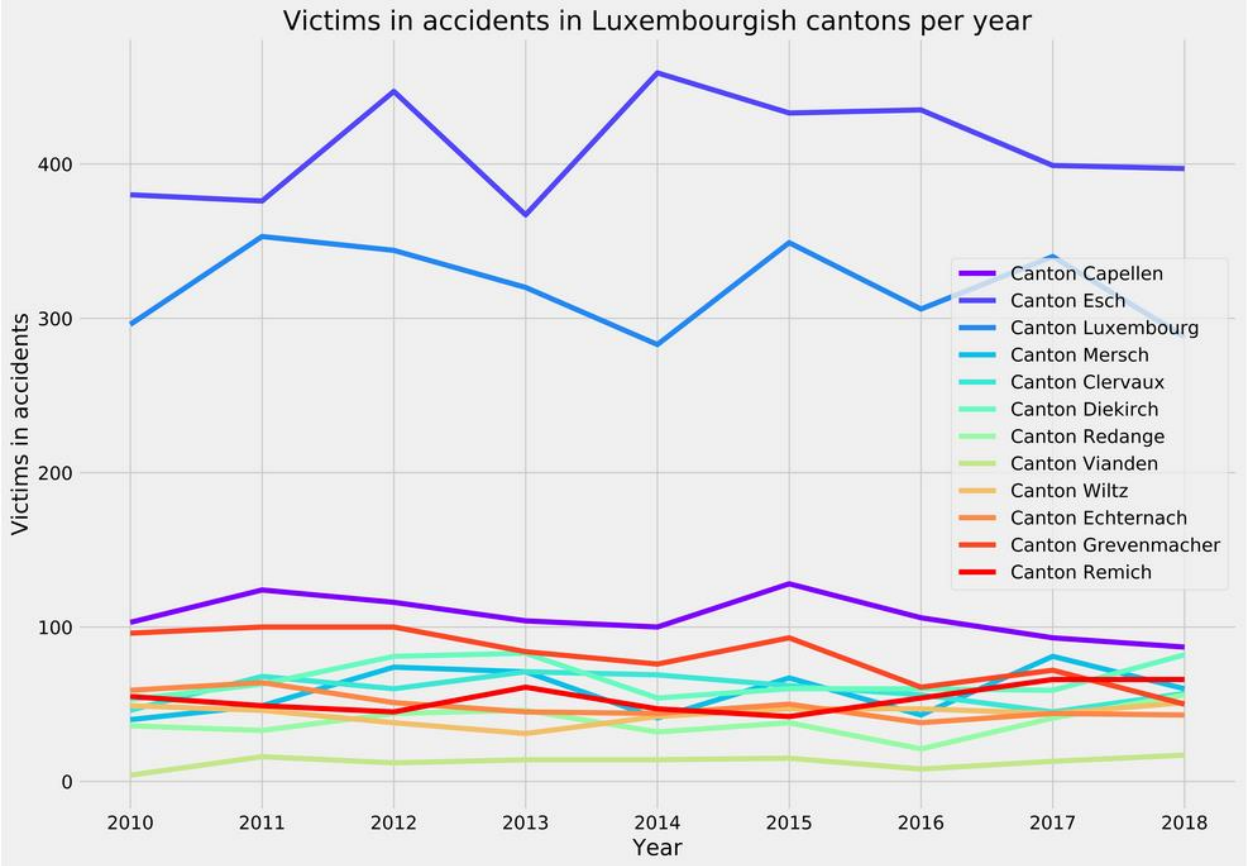

Figure 5 Number of accident victims in Luxembourgish canton from 2010-2018.

Cantons	Population 2018	Std Deviation of Accident Victims	Mean of Accident Victims	Trendline Slope
Canton Vianden	5163.0	3	12	0
Canton Wiltz	16735.0	5	43	0
Canton Clervaux	18081.0	8	59	0
Canton Redange	18664.0	9	38	0
Canton Echternach	18899.0	7	48	-2
Canton Remich	22366.0	8	53	1
Canton Grevenmacher	29828.0	16	81	-5
Canton Mersch	32112.0	14	58	1
Canton Diekirch	32543.0	11	66	0
Canton Capellen	48187.0	12	106	-2
Canton Esch	176820.0	31	410	2
Canton Luxembourg	182607.0	25	319	-1

Table 2 Selected indicator extracted from the Accident victims cantonal dataset.

As a last preliminary demographic analysis, the unemployment rate was checked (see *Figure 6*).

Compared to the other analyses, the unemployment rate in the 12 cantons is not strictly related to the highest population. Less populated cantons like Wiltz or Vianden show higher unemployment rates than the most populated canton, Luxembourg.

We can observe 2 spikes in 2013 and 2016 for most cantons, probably related to a major economic event in the previous year.

Overall however, the unemployment rate seems to decrease in every canton during the time period of 2010-2018.

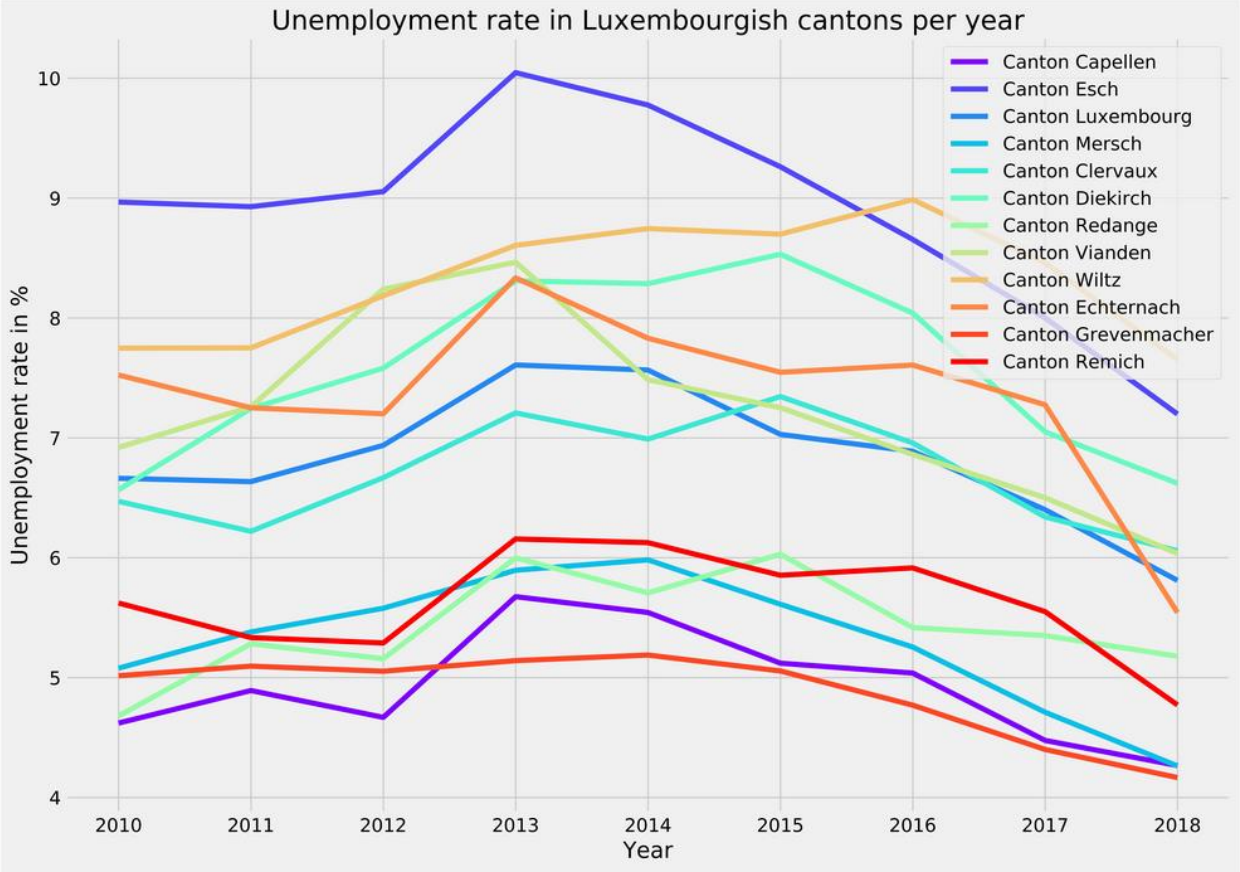

Figure 6 Annual unemployment rates in Luxembourgish cantons from 2010-2018.

To conclude the demographic analysis, Luxembourg is in a steady growth phase with population increasing steadily over the course of the last 25 years. Luxembourg, Esch and Capellen are hereby the largest cantons with regard to population and net migration.

The amount of victims in accidents is steady over the years and does not solely depend on the population. The unemployment rate is on a decrease since 2010.

In order to gain more insight on a local level, Foursquare location data was added to get more information about the similarity between the municipalities (*Section D* in script).

The 2018 municipality dataset was used (as venues should not have changed dramatically in the last two years) and the top venues were searched in a radius of 500 m. After one-hot encoding, the venue scores were added to the demographic data of 2018. As the venue data could not be generated for every municipality, municipality lacking the venue data were eliminated.

After performing a feature scaling, the elbow method was used to identify the optimal amount of clusters for KMeans (*Figure 7*). The optimal amount of clusters is not obvious to determine from the plot, indicating that the clusters might not be distinct enough. We proceeded to use 7 clusters as the curve first flattened at k=7.

The resulting clusters were however difficult to interpret as the municipalities belonging to one cluster sometimes spread over the whole country (see *Figure 8*).

- Cluster 1: Top venues are Asian/Italian restaurants and other food places
- Cluster 2: Biggest and most diverse cluster containing shops, yoga studios and real estate agencies
- Cluster 3: Only Remich as municipality close to the river and with lots of restaurants
- Cluster 4: Diverse cluster with restaurants and playgrounds
- Cluster 5: marked by hotels and historical sites
- Cluster 6: Luxembourg as most densely populated municipality
- Cluster 7: Schuttrange with a cosmetic shop, low population density and in close proximity to a small river

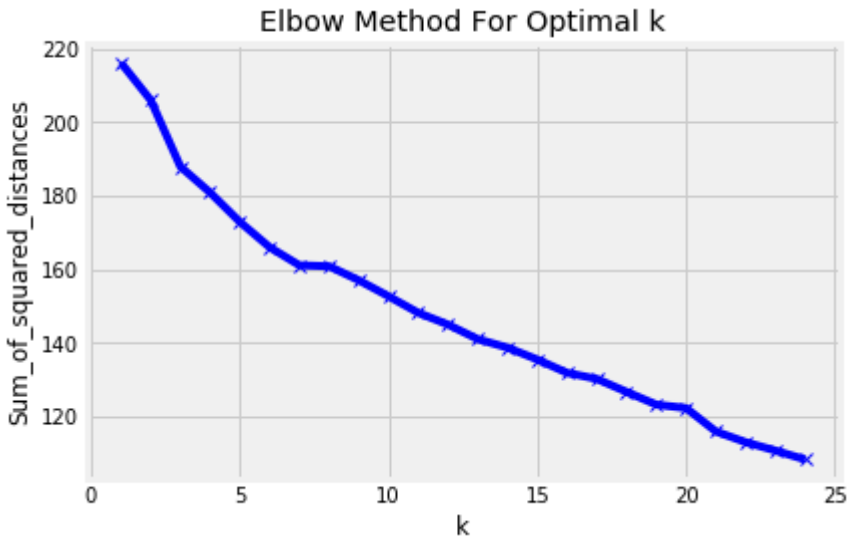

Figure 7 Elbow Method to determine optimal amount of clusters for KMeans in the municipality dataset.

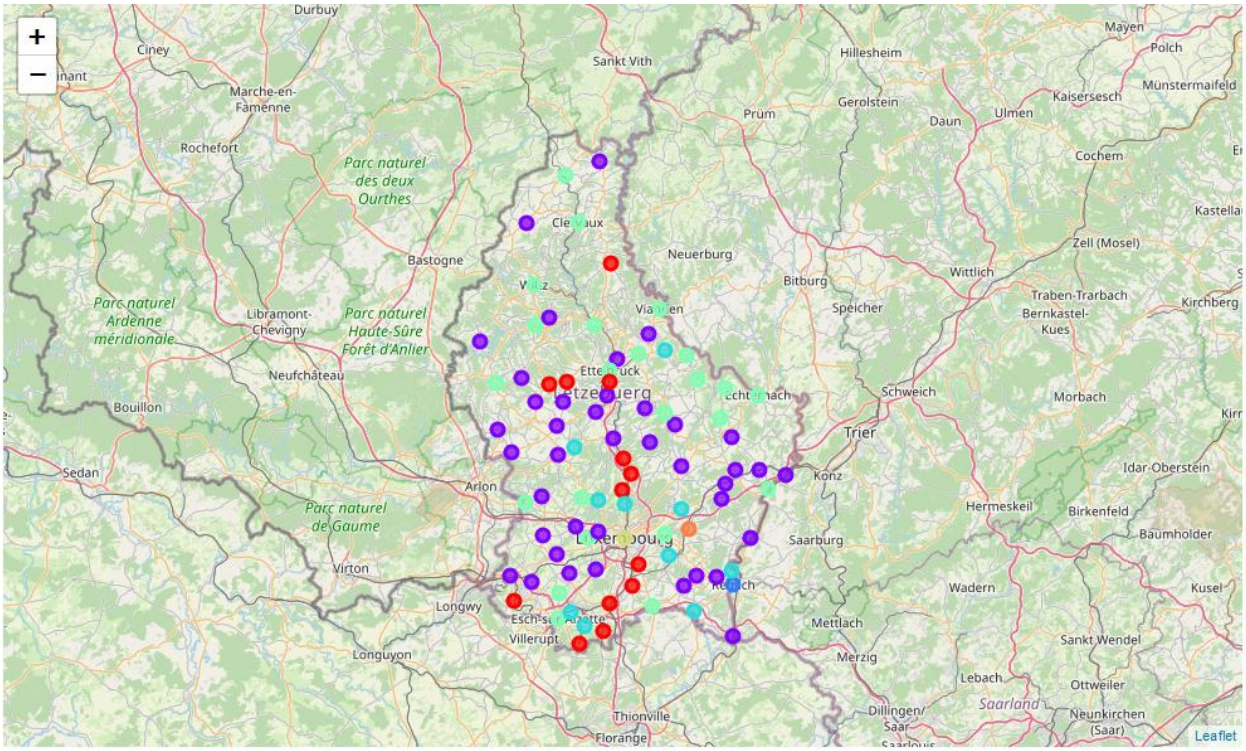

Figure 8 Result of the KMeans clustering (k=7) of the municipalities using venue data and demographic data.

As the municipalities seemed to be to akin to be effectively clustered, the same analysis was repeated with the cantonal 2018 data.

Following the same steps, the radius was chosen to be 2 km here, to cover most area of the canton! The elbow method was more indicative here, suggesting 5 clusters to be the optimal amount (see Figure 9)

Consequent analysis of the clusters shows that 4 cantons stick out of the 12, whereas the rest is very similar (see Figure 10).

- Cluster 1: Contains the smallest cantons by population density, has low amounts of accident victims and are located away from main highways
- Cluster 2: Luxembourg as the most densely populated canton
- Cluster 3: Capellen with a low unemployment rate and a variety of restaurants and shops due to several shopping malls present in this canton
- Cluster 4: Esch as second largest canton by population density, having a high unemployment rate and the highest number of accident victims.
- Cluster 5: Diekirch with high net migration (leaving Esch and Luxembourg apart), having a well-known History museum as most common venue

This clustering is in line with the demographic analysis, where Luxembourg, Esch and Capellen always stuck out from the rest of the cantons in almost every category.

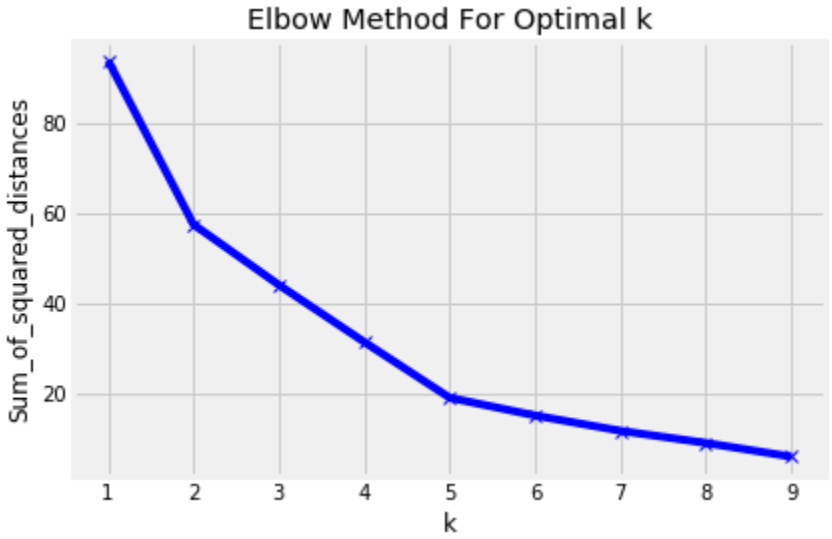

Figure 9 Elbow Method to determine optimal amount of clusters for KMeans in the canton dataset.

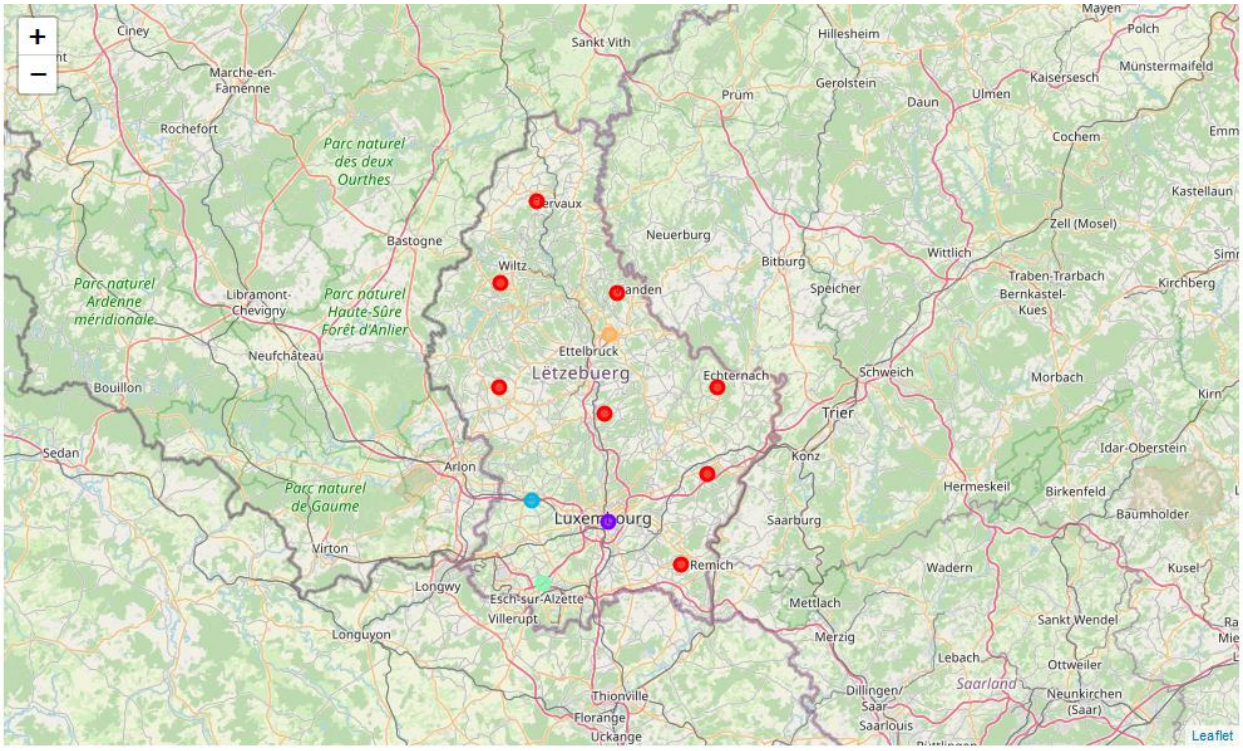

Figure 10 Result of the KMeans clustering (k=7) of the cantons using venue data and demographic data.

Final Thoughts

Analysis of several data sets not only allowed to identify trends in Luxembourg as a country, but specificly in the 12 cantons.

Luxembourg is too small to consider big variations within the municipalities, as demonstrated by a rather unsuccessful attempt of clustering with venues. Clustering the canton however highlighted what was previously observed in the line plots and clarified that at most 5 distinct clusters of cantons exist: Luxembourg, Esch, Capellen, Diekirch and the rest.

The data analysis could serve as a guideline to choose a residence. Higher number of accidents is related to traffic, whereas low net migration together with population density indicates that the areas are not overcrowding.

In the end, the top 10 venues give a good overview of what to expect when chosing a new residence.