

# Student Employment in Finland

## I. Introduction

This research aims to explore the employment rate across different education fields in various regions of Finland. The analysis of employment rates within specific education fields across different areas provides understanding of the patterns and variations in employment opportunities across the country.

Research Questions:

- Can regions be clustered based on their employment rates across different education fields?
- What distinct employment patterns emerge, and how do they vary across regions?

The data is obtained from Tilastokeskus - Statistics Finland. The dataset used in this project comprises employment rate data categorized by education field and areas across Finland in 2022. The data is transformed and formatted so that columns are fields of education, rows are areas, and the value of each instance is the employment rate of corresponding field of education and area. More specific details on fields of education and areas are provided in Figure 1 and Figure 2. The employment rate is calculated by dividing the number of employed students for a specific education field and area by the total number of students studying that corresponding field and area. The employment rate is used because it takes into account the differences in student populations among areas.

"Generic programmes and qualifications"	"Education"
"Arts and humanities"	"Social sciences, journalism and information"
"Business, administration and law"	"Natural sciences, mathematics and statistics"
"Information and Communication Technologies (ICT)"	"Engineering, manufacturing and construction"
"Agriculture, forestry, fisheries and veterinary"	"Health and welfare"
"Services"	"Unknown"

Figure 1: Education fields

"Uusimaa"	"Southwest Finland"	"Kanta-Häme"	"Päijät-Häme"	"Kymenlaakso"
"South Karelia"	"Satakunta"	"Pirkanmaa"	"Central Finland"	"South Ostrobothnia"
"Ostrobothnia"	"South Savo"	"North Savo"	"North Karelia"	"Central Ostrobothnia"
"North Ostrobothnia"	"Kainuu"	"Lapland"	"Åland"	

Figure 2: Areas in Finland

## II. Univariate Analysis

In this part, several univariate analyses are performed to understand the distribution and characteristics of the employment rates within each education field separately. Based on the summary (Figure 3), education fields with high employment rate are 'Education', 'Social sciences, journalism and information', 'Business, administration and law', and 'Health and welfare'. It can be seen that the ranges of education fields vary a lot from each other.

Generic programmes and qualifications	Education	Arts and humanities	Social sciences, journalism and information
Min. :24.50	Min. :47.30	Min. :30.80	Min. :44.00
1st Qu.:31.45	1st Qu.:59.70	1st Qu.:40.85	1st Qu.:57.95
Median :33.50	Median :67.00	Median :43.40	Median :66.70
Mean :34.42	Mean :65.87	Mean :43.17	Mean :65.25
3rd Qu.:37.45	3rd Qu.:71.60	3rd Qu.:46.80	3rd Qu.:73.95
Max. :49.40	Max. :83.90	Max. :54.10	Max. :79.40
Business, administration and law	Natural sciences, mathematics and statistics	Information and Communication Technologies (ICT)	Engineering, manufacturing and construction
Min. :55.00	Min. :40.90	Min. :23.30	Min. :40.20
1st Qu.:62.45	1st Qu.:46.55	1st Qu.:37.30	1st Qu.:50.00
Median :65.80	Median :51.60	Median :42.10	Median :51.80
Mean :64.99	Mean :52.81	Mean :41.28	Mean :51.54
3rd Qu.:67.35	3rd Qu.:59.25	3rd Qu.:44.70	3rd Qu.:54.50
Max. :73.40	Max. :67.80	Max. :57.50	Max. :57.10
Agriculture, forestry, fisheries and veterinary	Health and welfare	Services	Unknown
Min. :47.50	Min. :58.10	Min. :49.10	Min. :15.50
1st Qu.:55.40	1st Qu.:62.85	1st Qu.:50.50	1st Qu.:25.10
Median :59.20	Median :64.50	Median :52.40	Median :27.50
Mean :59.52	Mean :66.02	Mean :54.15	Mean :26.92
3rd Qu.:63.25	3rd Qu.:67.80	3rd Qu.:57.10	3rd Qu.:29.55
Max. :75.50	Max. :84.60	Max. :62.90	Max. :33.70

Figure 3: Summary of employment rate by education field

In histograms in Figure 4, we can take a look at distributions of education fields to understand their shape, central tendency, and variability in the dataset.

Boxplots in Figure 5 give a clearer view on employment rate by areas. The ranges of employment rates across areas vary a lot: Central Finland, North Karelia, Pirkanmaa and Uusimaa have smaller ranges than the other regions. The means of employment rates across areas are within 45%-65%.

Boxplots in Figure 5 provide information on employment rate by education fields. The ranges of employment rates across education fields are small. The means of employment rates vary very much across education fields; this makes sense since each education field has intrinsic aspects that affect students' employment. 'Arts and humanities', 'Generic programmes and qualifications', 'Information and Communication Technologies (ICT)' and 'Unknown' are variables with low mean of employment rate.

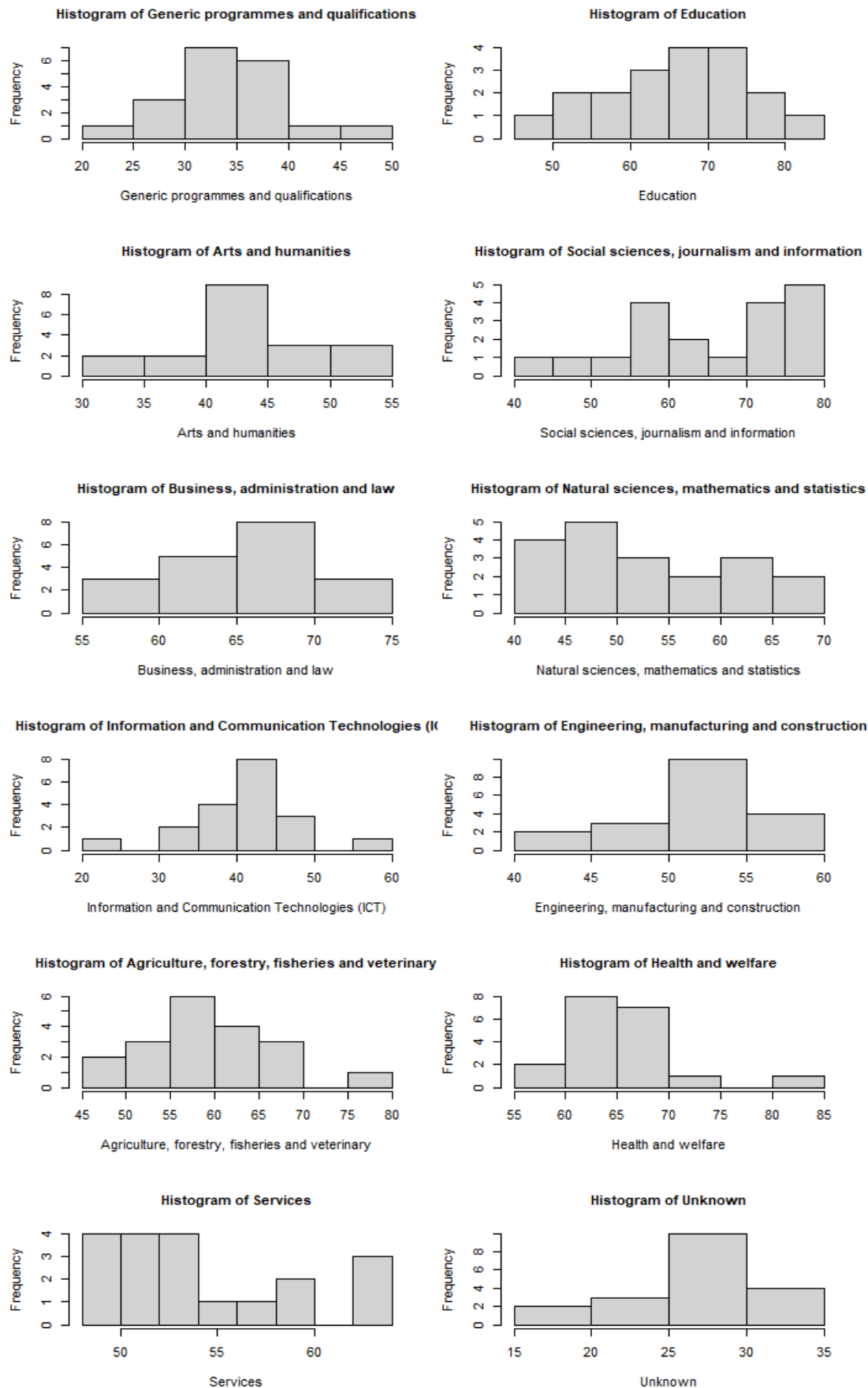


Figure 4: Distributions of education fields variables

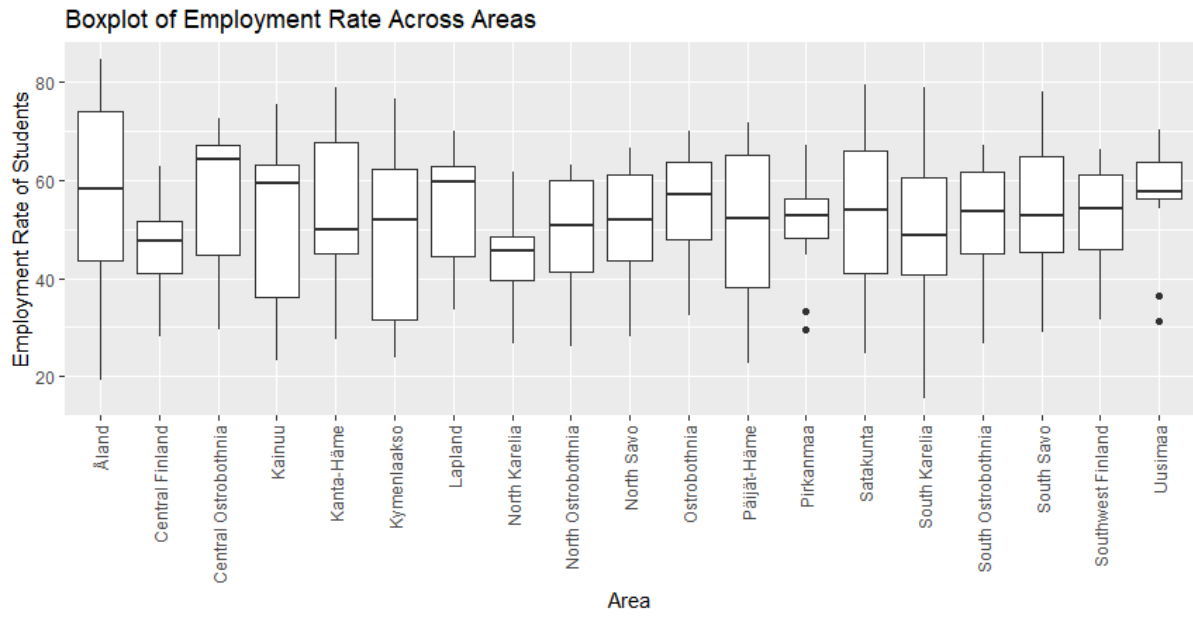


Figure 5: Boxplot of Employment rate across areas

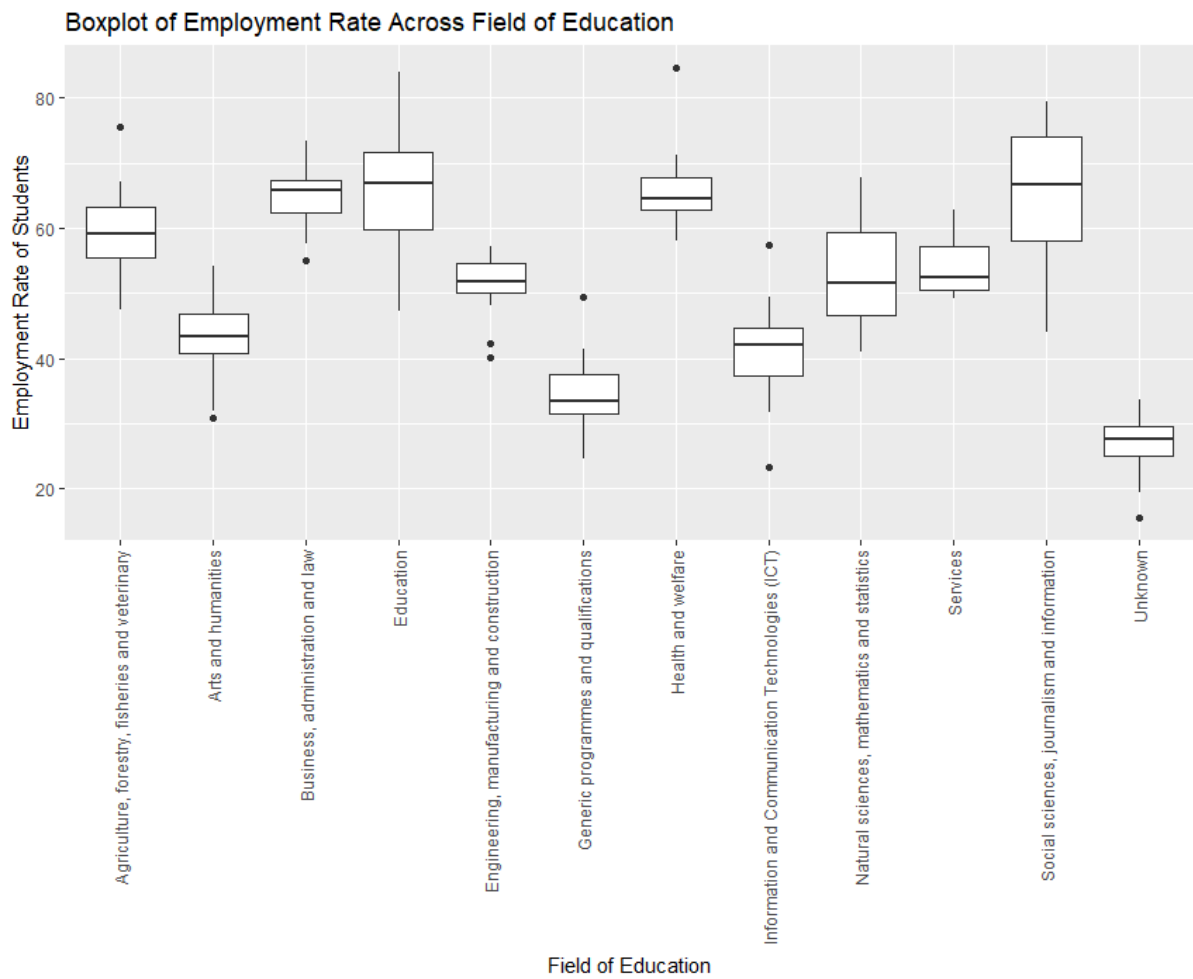


Figure 6: Boxplot of Employment rate across education fields

### III. Bivariate Analysis

For bivariate analysis, pairwise scatter plots of the variables are considered; however, it is quite hard to spot clusters or have any meaningful interpretation from the visual. For that reason, I compare the employment rates of students across different fields of education with different areas as the legend.

Employment rates of 'Health and welfare' are scattered evenly among areas. Åland and South Karelia seem to have outliers for many education fields. For education fields with generally low employment rates such as 'Arts and humanities', and 'Information and Communication Technologies (ICT)', Uusimaa has the highest employment rates.

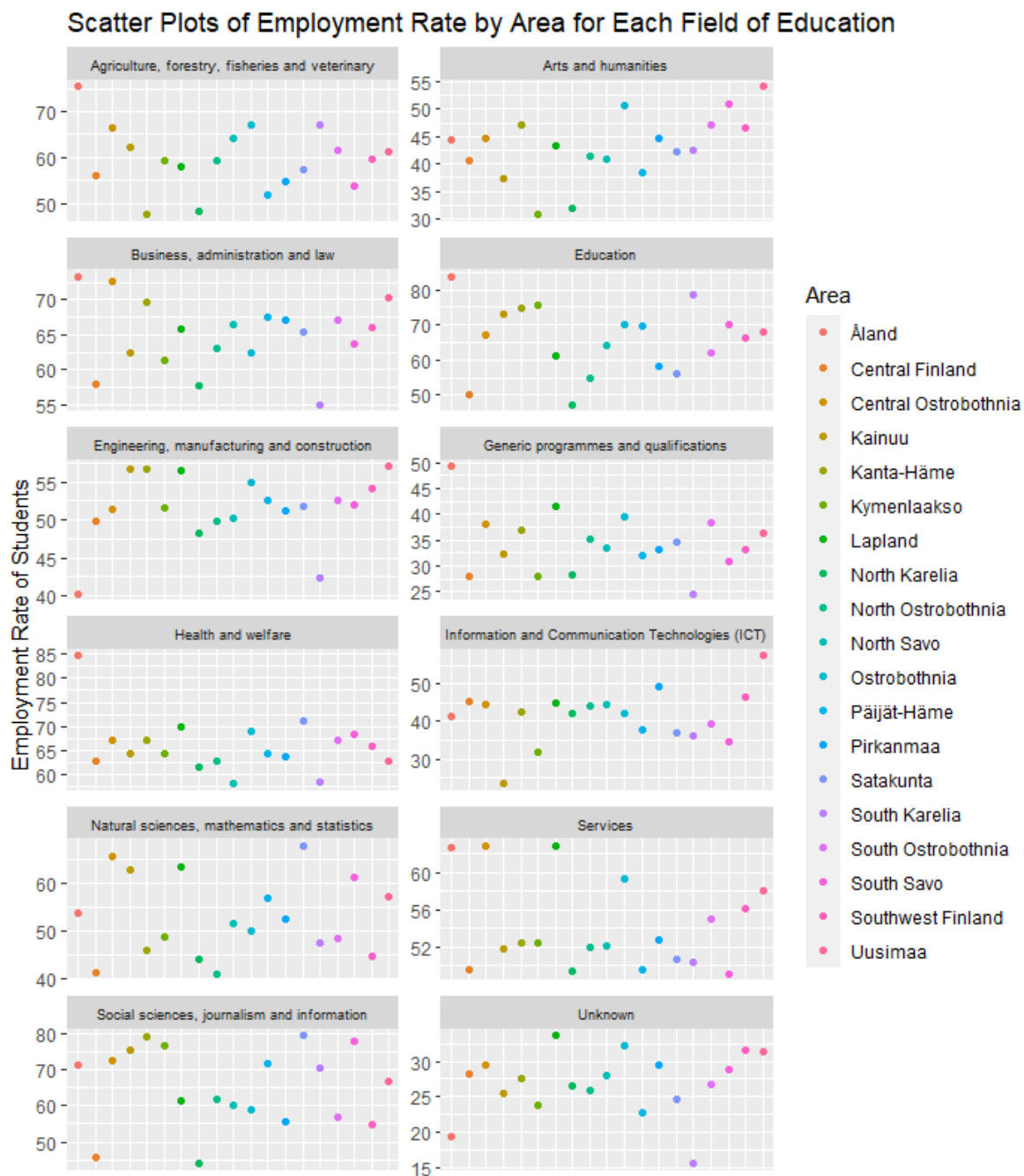


Figure 7: Scatter plots of employment rate by area and education field

## IV. Multivariate Analysis

### A. Method Selection

Agglomerative hierarchical clustering algorithm is a suitable method for analyzing the employment rate data across different education fields and regions because this method does not require knowledge of the number of clusters. Agglomerative hierarchical clustering is a versatile and intuitive method that allows the identification of distinct patterns and variations in employment patterns across regions.

Clustering methods rely on two (separate) issues: the choice of a distance or dissimilarity measure between objects and the choice of a group building algorithm. With quantitative data, the Euclidean distance is a classical choice with straightforward interpretation. As for linkage functions, average linkage provides a safer choice than minimum and maximum linkage. It can avoid chaining in single linkage and sensitivity to outliers in maximum linkage. As a result, agglomerative hierarchical clustering is applied to the data set. The metric used is Euclidean distance, and linkage criteria is the average linkage.

### B. Results & Interpretation

When the cluster dendrogram is obtained, the difficult part is to choose where to cut the tree, i.e., choose the number of clusters. This is typically an expert decision as it requires context knowledge. However, some approaches are applied to choose a threshold to cut the dendrogram. First, I look for points where the branches of the dendrogram become noticeably longer or where there are clear gaps between clusters. Next, I consider a number of clusters that are interpretable and meaningful for the analysis because too few clusters may oversimplify the data, while too many clusters may obscure meaningful patterns. With these considerations, the tree is cut at the height of 30, and 5 clusters are obtained.

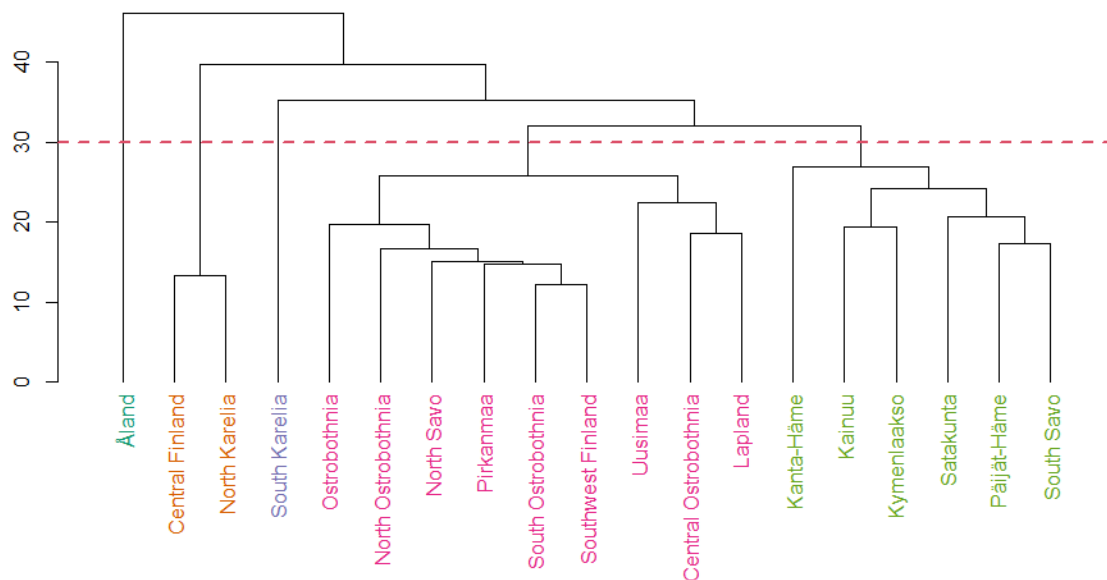


Figure 8: Cluster dendrogram

Although it is difficult to investigate what the underlying reason of the clusters is with respect to the variables used in the analysis, speculative interpretations are made. With close examination of employment rates for each cluster, interpretations are made based on average employment rate, the context of the regions and patterns within each cluster.

Cluster	Area	Population	Average Employment Rate (%)
1	Åland	30,359	58.33
2	Central Finland	272,437	46.29
	North Karelia	162,540	44.11
3	South Karelia	125,353	49.03
4	Central Ostrobothnia	67,805	56.88
	Lapland	175,795	55.18
	North Ostrobothnia	416,543	49.28
	North Savo	247,689	51.15
	Ostrobothnia	176,323	54.72
	Pirkanmaa	532,671	51.08
	South Ostrobothnia	190,774	51.85
	Southwest Finland	485,567	52.09
5	Uusimaa	1,733,033	56.74
	Kainuu	70,521	52.25
	Kanta-Häme	169,537	53.91
	Kymenlaakso	159,488	50.31
	Päijät-Häme	204,528	51.23
	Satakunta	212,556	53.16
	South Savo	130,451	53.46

Table 1: Cluster result combined with area population and average employment rate

There is a significant size difference between clusters, so interpretations may not be representative and need to be validated by experts. There is no clear distinction of clusters based on average employment rate. However, we can see that cluster 2 has two areas with the lowest average employment rates, and cluster 1 has the area with the highest average employment rate.

For the context of the regions, we look at the populations of each region to see if regions' population might explain the employment patterns observed in each cluster. Again there is no clear underlying reasoning of clusters based on regions' population. Cluster 1 has the area with the lowest population. Cluster 4 has many high population areas such as North Ostrobothnia, Pirkanmaa, Southwest Finland and Uusimaa.

For patterns within each cluster, closer investigation is done by looking at the employment rate for each area in a specific cluster by education fields. Cluster 1 has significantly high employment rates in these education fields: Generic programmes and qualifications; Education; Business, administration and law; Agriculture, forestry, fisheries and veterinary and Health and welfare. It seems that cluster 2 has the lowest employment rate for all education fields. Cluster 3 does not seem to have any distinctive difference from other clusters. Cluster 4 has high employment rates in STEM fields (Natural sciences, mathematics and statistics, Information and Communication Technologies (ICT), Engineering, manufacturing and construction), Business and Finance, Arts and humanities and Services. Areas in cluster 5 have significantly high employment rates for Social sciences, journalism and information.

## **V. Evaluation**

After doing analysis, it's time we reflect on research questions. It seems that regions can be clustered based on their employment rates across different education fields: the five clusters obtained seem to have fair interpretability. The patterns of employment rates are not distinctive for each cluster; however, some speculated insights can be gained. Areas in cluster 2 generally have low employment rates for all education fields. Areas in cluster 4 seem to be more developed and sophisticated with high employment rates in STEM fields, Arts and humanities and Business, administration and law.

The main drawback of the analysis is the imbalanced sizes of clusters. It can be challenging to interpret the characteristics and patterns of the smaller clusters, as they may not represent a sufficient number of data points to draw meaningful conclusions. This can lead to biased or misleading interpretations of the clustering results.

Moreover, we should take into account the fact that different clustering methods can produce different solutions, and the different distance methods and linkage may also have an effect on the outcome. In the clustering problem, the labels are unknown. Therefore, it is essential to validate the clustering results using external sources or domain knowledge to ensure that the clusters make sense and are consistent with existing understanding of the data.

The scope of data is quite small as the analysis is only based on data in 2022. Including data from previous years can improve the reliability and robustness of the analysis by providing a larger and more representative dataset.



## References

[https://pxdata.stat.fi/PXWeb/pxweb/en/StatFin/StatFin\\_tyokay/statfin\\_tyokay\\_pxt\\_13g2.px](https://pxdata.stat.fi/PXWeb/pxweb/en/StatFin/StatFin_tyokay/statfin_tyokay_pxt_13g2.px)  
[https://en.wikipedia.org/wiki/Demographics\\_of\\_Finland](https://en.wikipedia.org/wiki/Demographics_of_Finland)