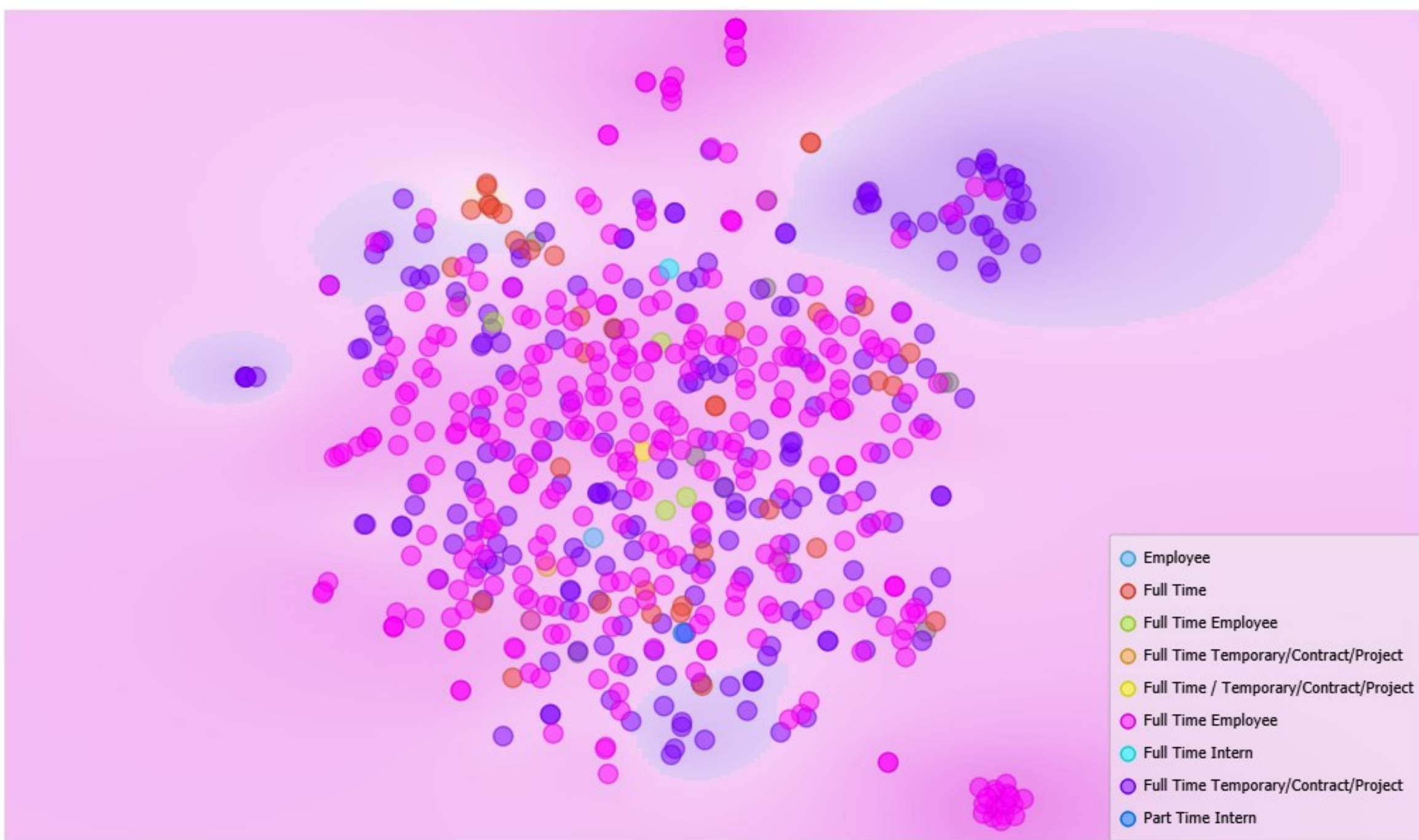
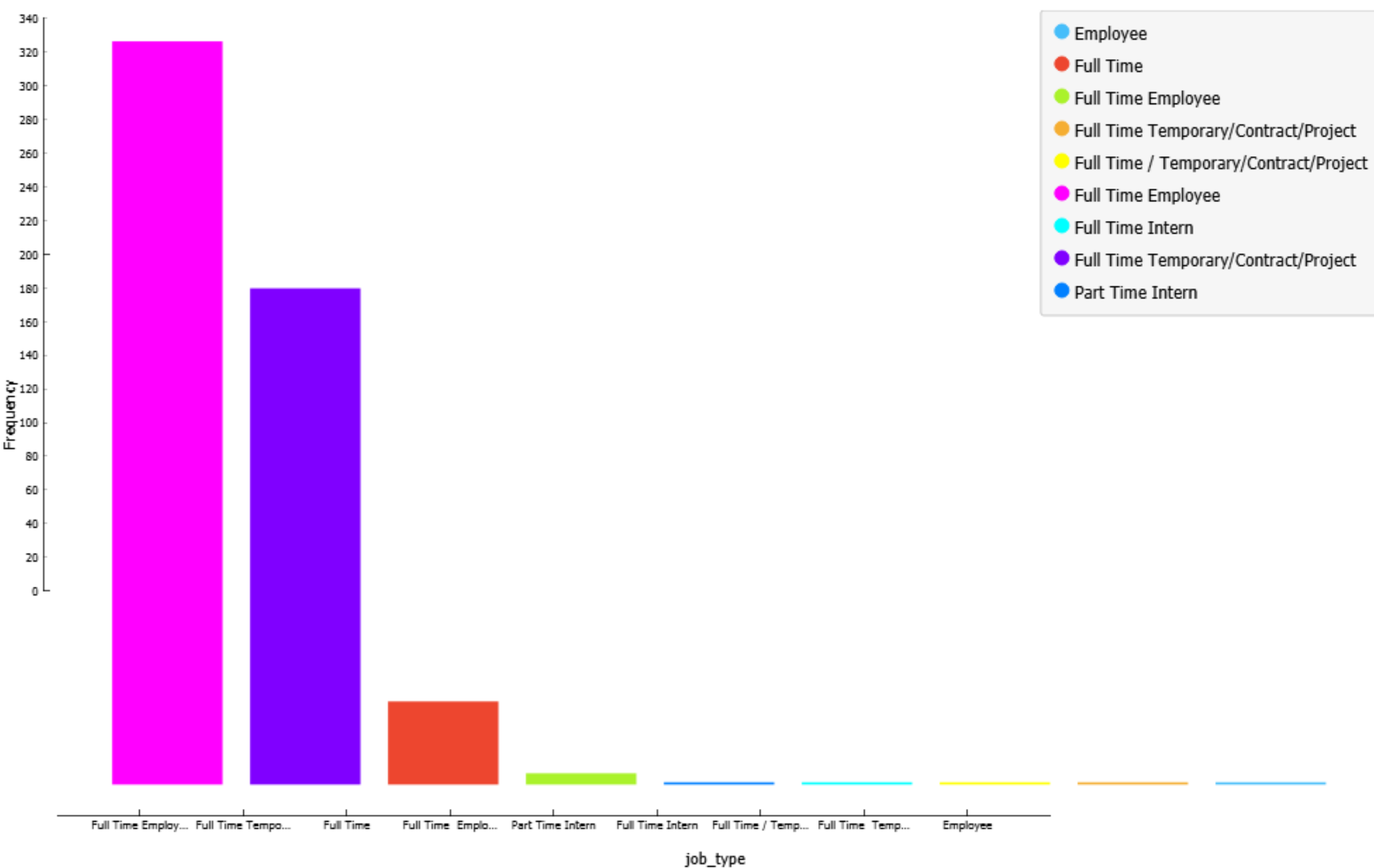


# Job postings

## General overview

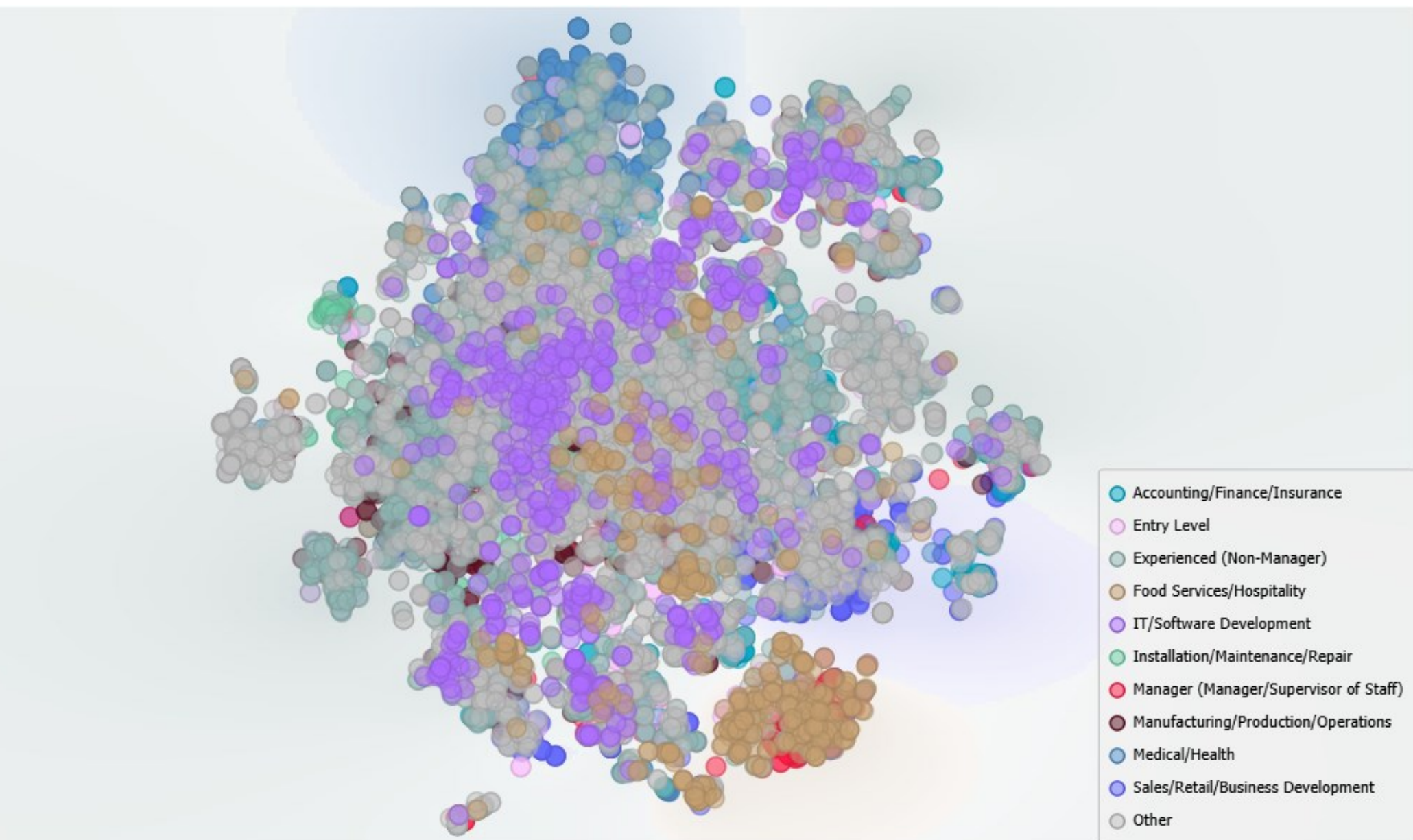
The researched dataset is from „monster.com“ job posting website. The job offers were posted in 2016 and are all from the United States. It consists of 21932 rows, which means the dataset contained the same amount of job postings. In this research, the focus was on getting a general picture about the usual job postings and conduct text mining and unsupervised learning on them with software tools. The research has been conducted using the „Orange“ data mining software.

In the first part we will get information about the job types, their variance, variance of the job sectors, general observations and findings.

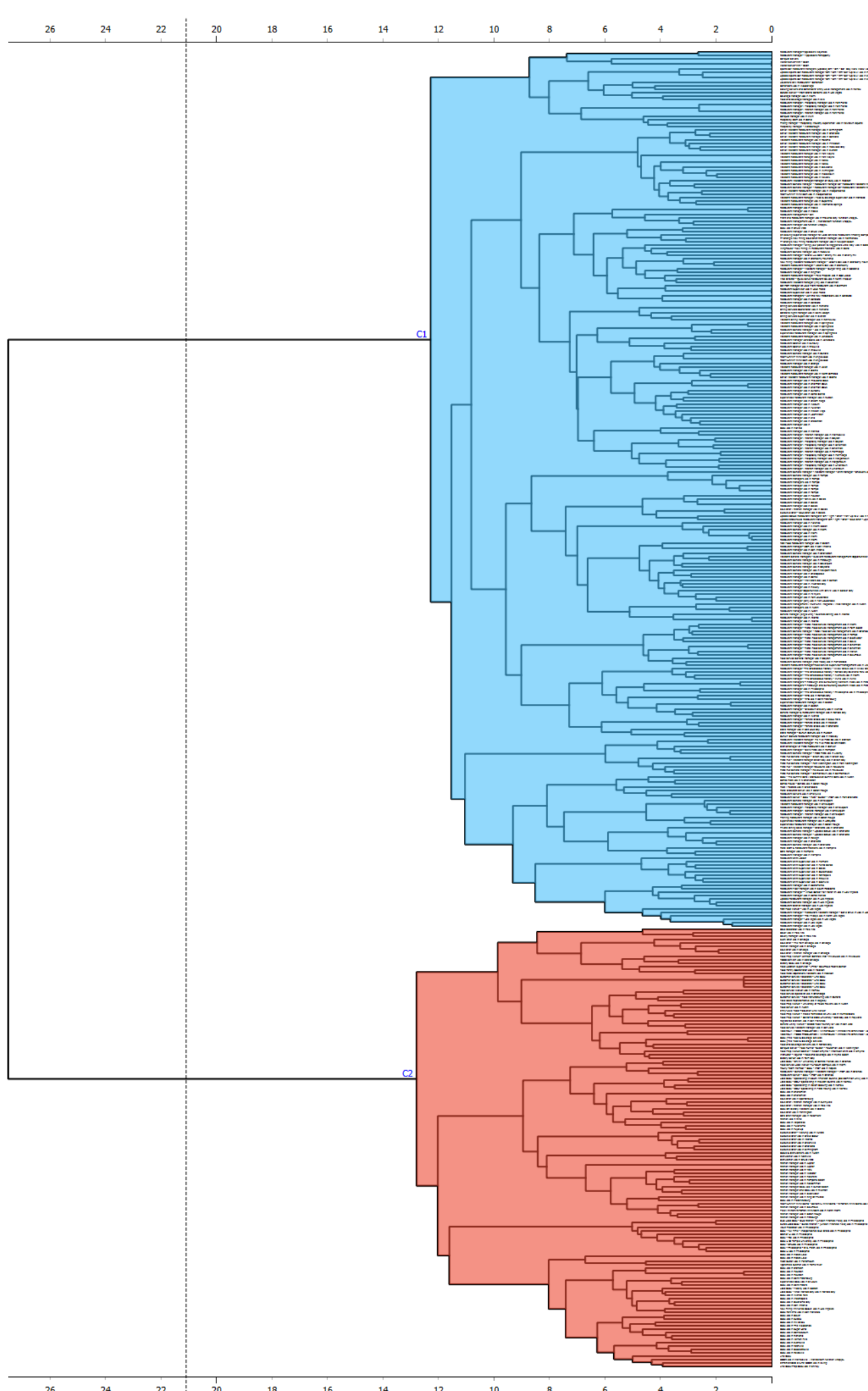


On the bar chart on the **left** we can see the distribution of the types given to the job postings. In the dataset, it is further classified what the exact working order of each jobs are. According to this chart, full-time jobs are hugely overrepresented, there are barely any part-time jobs (even then, there are only internships). Unfortunately, the different ways how the description to job types was typed, created more categories for the same working order.

On the T-SNE plot on the **right**, we can see a T-SNE plot for the job types. Generally, the whole plot shows various distances between the job postings from each job types. However, there is a spot which we can look at. On the upper right corner we can see a group of job postings in **purple** with some **pink**. After further investigation, it has been revealed that most of the purple jobs are web designer jobs, meanwhile the six jobs marked with pink are various types of manager jobs.



This T-SNE plot shows the distance between each job sectors in 2 dimensions. What we see here, is that most of the job sectors' postings show a high level of variance. However, there is an interesting spot on the bottom of the plot marked in **brown**. Even though the job postings of the „Food services/Hospitality“ sector is also scattered throughout the graph, most of them are concentrated on the bottom. Coincidentally, almost all of the job postings of the „Manager (Manager/Supervisor of Staff)“ sector is concentrated on the same spot. Further research shown, that high majority of management jobs with staff-supervising role, are within the Food or Hospitality industry. On the right side, this spot's hierarchical clustering is shown. In this clustering, it clearly shows that the posted jobs can be divided into two parts: manager (**blue**) and worker (**red**) in the food/hospitality industry.

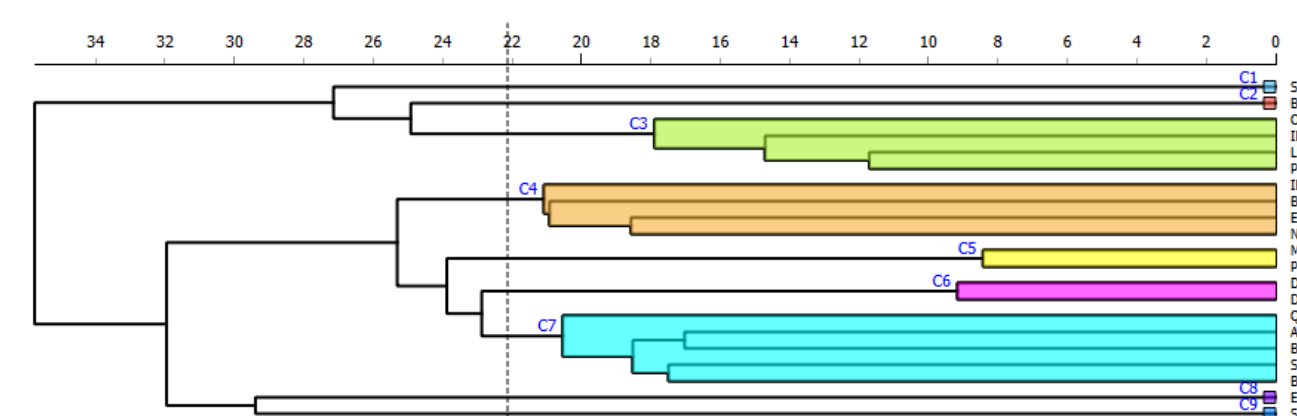


Weight	Word
19477	?
1965	-
1908	\$
1047	/year
743	/hour
167	1000000
159	6000000
148	4000000
139	5000000
130	7500000
128	8000000
117	6500000
110	9000000
106	7000000
104	1500

In the dataset, for many job postings there was no salary given, neither as a fixed sum nor as a range. Out of the 21932 job postings, around 19477 records are detected to have a question mark in the place of the field for salary, meaning there is no available value for the amount of money offered. Unfortunately, with salaries missing from around 89% of job postings, no representative statistics can be made.

## Specialised Sector Research: IT/Software Development

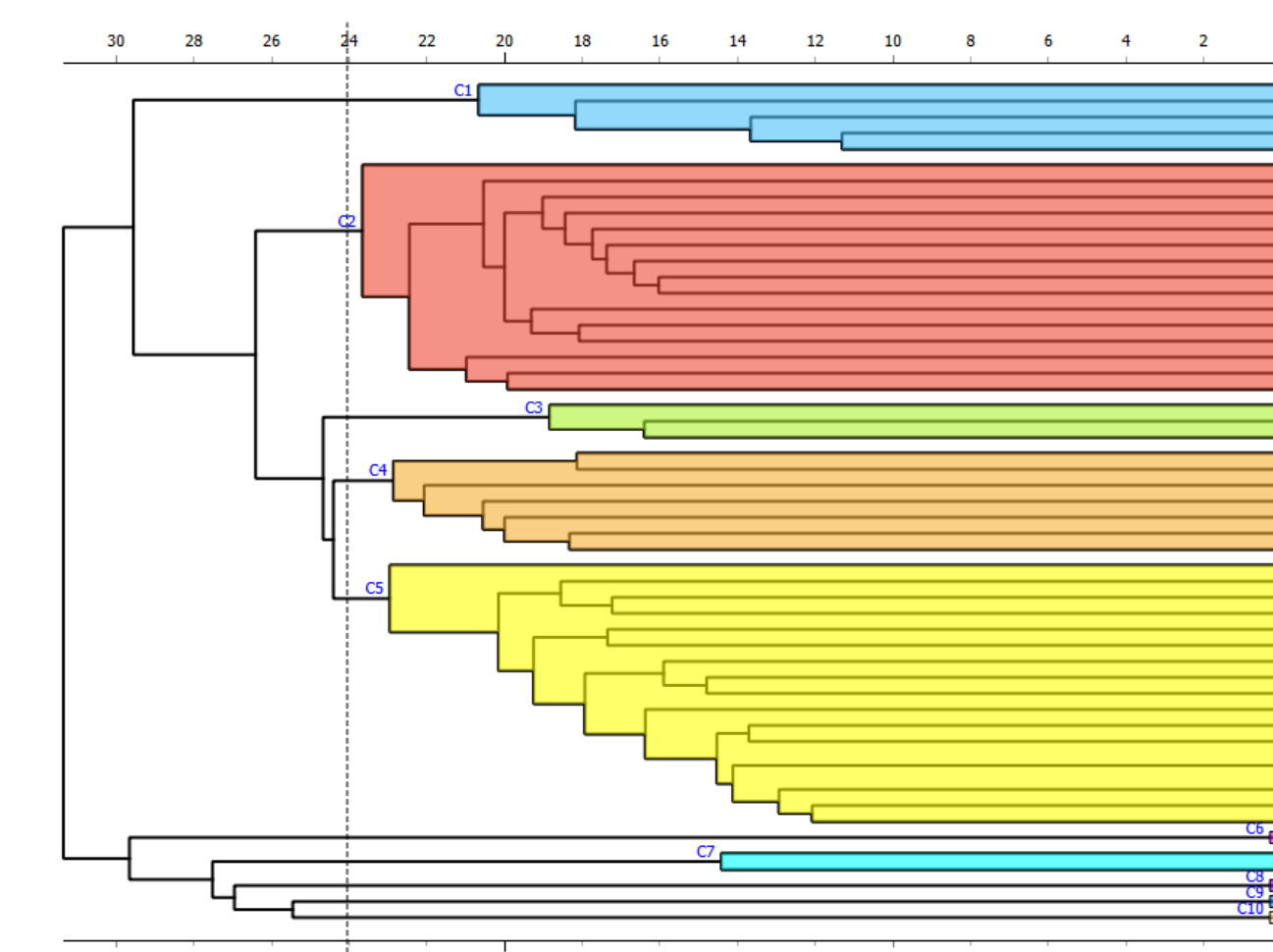
In this section, we will look at a chosen sector of jobs to research on it further. In this case, it is the sector of „IT/Software Development“. Please look at the right side of this poster from the top to the bottom, to see the full clustering generated for IT/Software development. There are four different clusters chosen, on which we will have a further insight. The number of jobs posted in this job sector is 610.



Weight	Word
6.76	data
1.38	management
1.29	services
1.19	analysis
1.00	client
1.00	systems
0.90	excel
0.90	group
0.86	reports
0.86	team
0.86	classroom
0.86	quality
0.81	information
0.76	include
0.76	entry
0.71	design
0.67	big
0.67	consulting
0.62	knowledge

Based on the wordlist and their weights in shaping the **Cluster 14**, most likely, it's a cluster of data analyst/ data management jobs. Among the words with the higher weights, we can find „data“, „management“, „services“, „analysis“, and a little bit further we also find „excel“, which is telling in what type of most jobs are in this cluster.

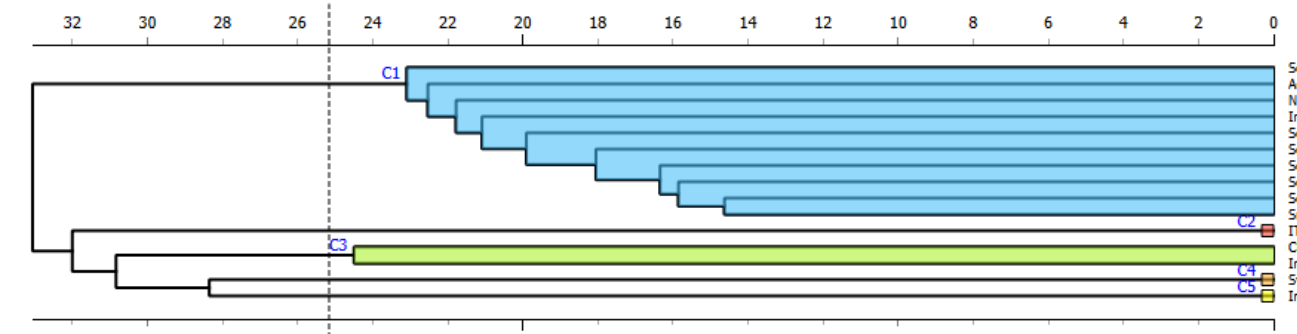
There are nine sub-clusters, out of which four are individual, meaning there are only one job posting belonging there. Overall, this cluster shown a greater variance for its size of only 21 job postings.



Weight	Word
2.81	development
1.81	software
1.32	team
1.13	application
1.06	systems
1.04	design
0.94	web
0.91	technical
0.83	code
0.83	support
0.74	client
0.70	applications
0.68	developer
0.68	services
0.66	testing
0.66	knowledge
0.64	management
0.62	strong
0.62	well

The **Cluster 21** indicates the it is a cluster of software developer job postings, as the biggest weight is on the word „development“, followed by „software“, „team“, „application“, „systems“ and „design“, which are the words with the highest weight in the cluster.

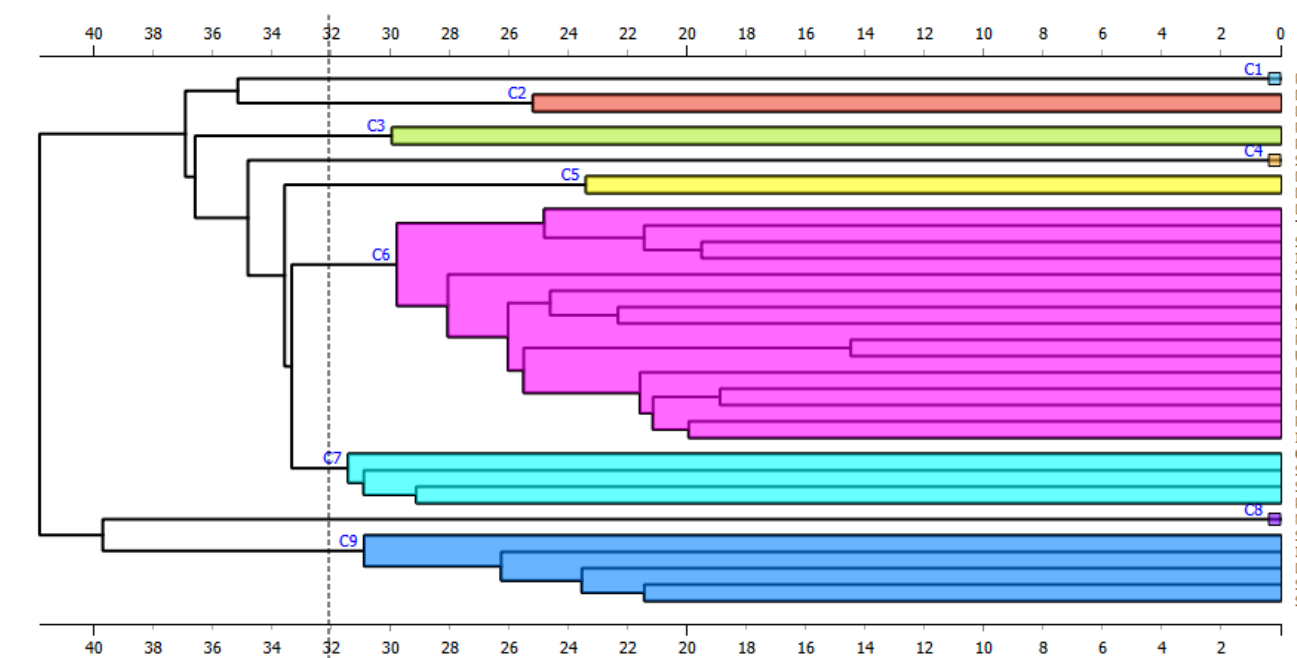
The 53 job posting that build up this cluster, are grouped into 10 sub-clusters, where most of the jobs can be found in 4 bigger sub-cluster. In these sub-clusters, the more specific types of the job postings are recognizable. For example, the sub-cluster marked with C2 (colored in red) implies that this sub-cluster is specialized to development of web applications.



Weight	Word
11.67	security
3.13	information
2.33	systems
2.33	network
1.93	support
1.47	management
1.40	technical
1.33	technologies
1.27	knowledge
1.20	risk
1.20	services
1.13	related
1.07	data
1.00	computer
1.00	technology
0.93	standards
0.93	system
0.87	architecture
0.87	enterprise

In **Cluster 8**, there is a huge weight on the word „security“ compared to the others, then we find words such as „information“, „systems“ and „network“ also with bigger relevance. It is safe to come to a conclusion, that this cluster contains jobs related to IT security.

The 15 job posting are grouped into 5 sub-clusters, where 5 job posting are (mostly) in individual clusters, meanwhile there are 10 job postings which appear to be similar to each other as they are in the same sub-cluster.



Weight	Word
6.55	management
2.21	team
2.15	development
1.97	projects
1.73	client
1.73	knowledge
1.48	responsible
1.39	risk
1.30	strong
1.27	communication
1.27	process
1.24	processes
1.21	support
1.18	manage
1.12	data
1.12	manager
1.06	change
1.00	delivery
1.00	information

The **Cluster 31**, containing 33 job postings, has the word „management“ as its most significant word, after which come „team“, „development“, „projects“, but a little bit further down, there is „responsibility“. These word are associated with IT management, and this is the type of job this cluster mostly contains.

Cluster 31 is divided into 9 sub-clusters. The size of these sub-clusters are different, there is one bigger cluster 2 middle sized, the rest are either dual sub-clusters (with 2 members) or individual sub-clusters.

