CZECH SALARIES IN THE PUBLIC SPHERE: EMPIRICAL EVIDENCE

Diana Bílková[1]

Abstract

The labour market comprises one of the basic economic factors in all developed countries. However, the labour market is not homogeneous; moreover, it can be analysed applying various parameters, i.e. economic, structural, flexibility, mobility, education, and demographic. This paper addresses the construction of clusters of districts in the Czech Republic that are as similar as possible in terms of the gross employee salary, age, education and number of years of employment. The cluster analysis approach was applied in the research supplemented by Ward's method and the Euclidean distance. The various districts of the Czech Republic were divided into fourteen clusters. It was determined that the resulting clusters of districts do not correspond to the country's regional structure, rather the clusters relate to districts that are similar to each other in terms of their populations, areas and industrial and other factors as opposed to their geographical locations.

Keywords

Salaries of Men and Women, Factors that Influence Salaries, Cluster Analysis, Ward's Method, Euclidean Distance

I. Introduction

The labour market is not homogeneous. It can be analysed applying various parameters, i.e. economic (value added, salary/wage, tax burden), structural (the representation of various sectors, the share of low value-added sectors, the public-private employment ratio), flexibility and mobility factors (the composition in terms of working hours, commuting, share of foreigners), educational (the level of completed education, continuing vocational education) and demographic parameters (employment by gender, age and the associated amount of experience). While the analysis of all the above factors is of interest, this paper concentrates on the following labour market components: gross monthly salaries, age, education and length of experience. The paper focuses on the issue of remuneration for work performed in the public sphere of the Czech economy in the years 2019 and 2020, thus taking into consideration the changes brought about by the onset of the coronavirus crisis in 2020.

As Ariffin & Ha (2014), Gimpelson (2019) and Sargeant (2010) have demonstrated, the remuneration of employees for work is related to the employee's age. Moreover, the influence of the length of the working period on remuneration has been clearly shown in a study by Bílková (2016). Similarly, the relationship between the amount of remuneration for work and the level of education attained by the employee has been presented by Bílková (2015), Carraher (2011), Lleras (2008) and Ruiz, Gómez & Narváez (2010). Moreover, a number of Czech and foreign authors have addressed labour market issues in connection with social issues, see for example Beran & Godarová (2017) or Červenka (2021).

The objective of this paper is to construct clusters of districts of the Czech Republic that are as similar as possible in terms of the gross monthly salary of employees, their age, education and number of years with the same employer, and to compare the structure of these clusters with the composition of the country's regions according to the districts thereof. Clusters were constructed separately for men and for women and the data considered covered 2019 and 2020, thus allowing

[1] Research Institute for Labour and Social Affairs, Delnicka Street 213/12, 170 00 Prague 7, Czechia. E-mail: bilkova@vse.cz.

for the identification of changes related to the onset of the coronavirus crisis in 2020. The composition of the various regions of the Czech Republic in terms of districts is shown in Table 1.

Table 1 Distribution of Czech districts according to region

Region	District	Region	District
Prague	Prague	Liberec	Ceska Lipa
Central Bohemia	Benesov		Jablonec nad Nisou
	Beroun		Liberec
	Kladno		Semily
	Kolin	Hradec Kralove	Hradec Kralove
	Kutna Hora		Jicin
	Melnik		Nachod
	Mlada Boleslav		Rychnov nad Kneznou
	Nymburk		Trutnov
	Prague-vychod	Pardubice	Chrudim
	Prague-zapad		Pardubice
	Pribram		Svitavy
	Rakovnik		Usti nad Orlici
South Bohemia	Ceske Budejovice	Highlands	Havlickuv Brod
	Cesky Krumlov		Jihlava
	Jindrichuv Hradec		Pelhrimov
	Pisek		Trebic
	Prachatice		Zdar nad Sazavou
	Strakonice	South Moravia	Blansko
	Tabor		Breclav
Pilsen	Domazlice		Brno-mesto
	Klatovy		Brno-venkov
	Pilsen-jih		Hodonin
	Pilsen-mesto		Vyskov
	Pilsen-sever		Znojmo
	Rokycany	Olomouc	Jesenik
	Tachov		Olomouc
Karlovy Vary	Cheb		Prerov
	Karlovy Vary		Prostejov
	Sokolov		Sumperk
Usti nad Labem	Decin	Zlin	Kromeriz
	Chomutov		Uherske Hradiste
	Litomerice		Vsetin
	Louny		Zlin
	Most	Moravia-Silesia	Bruntal
	Teplice		Frydek-Mistek
	Usti nad Labem		Karvina
			Novy Jicin
			Opava
			Ostrava-mesto

Source: Czech Statistical Office (2022)

The data used in the research was obtained from the Ministry of Labour and Social Affairs of the Czech Republic, the administration of which is entrusted to the Ministry of Finance of the Czech Republic. The data set covers the salaries of all employees in the public sphere, i.e. it is not merely a sample. The sizes of the processed data files are shown in Table 2. The data was processed using the SPSS statistical programming environment.

Table 2 File ranges

	Year	
	2019	2020
Men	203 662	201 678
Women	414 170	397 442

Source: Own research

The cluster analysis method was used in the research supplemented by Ward's method and the Euclidean distance for the analysis of data from 2019 and 2020. The Ward's method approach, a hierarchical clustering method, is not based on the optimisation of the distances between clusters, rather it addresses the minimisation of cluster heterogeneity according to the increment of the intra-cluster sum of the squares of the deviations of objects from the centres of gravity (centroids) of the clusters. Ward's method acts to eliminate overly small clusters, thus forming clusters of approximately the same size, which is advantageous in terms of the clustering of districts in the Czech Republic; hence the application of Ward's method in the cluster analysis in this study. In terms of the measurement of the distance between and the similarity of the objects, the selection was based on how we needed to strengthen the influence of those variables for which extremely large differences were observed on the total. The Euclidean distance was applied in our case since we did not need to strengthen the influence of any of the variables (the points with the same Euclidean distances from the centre lay on the circumference).

Although various methods and recommendations are available in cluster analysis for determining the optimal number of clusters, they do not provide definitive conclusions since cluster analysis basically comprises an exploratory approach, i.e. it is not a statistical test. The explanation and interpretation of the resulting hierarchical structure depends on the context and, in theory, there may be several possible solutions. The number of clusters was chosen according to the number of regions in the Czech Republic, i.e. fourteen.

Cluster analysis was applied separately for the various factors considered: the employee's salary in CZK, age in whole years, highest level of education and period of employment with his/her current employer in years.

II. Methodology: Cluster Analysis

The essence of this multidimensional statistical method has been explained, for example, in (Rencher and Christensen, 2012). We assume that we have a data matrix **X** of type $n \times p$, where n is the number of objects and p is the number of variables. We consider various decompositions of the $S^{(k)}$ set of n objects into k clusters. The purpose is to determine the decomposition that is the most advantageous in certain respects; however, for the purposes of this paper we consider only decompositions with disjoint clusters in order to obtain a state in which the objects within one cluster are as similar as possible, and the objects of the various clusters are as similar as possible. Since the notion of similarity only applies to the p variables selected for investigation, the choice of the set of variables determines the success of the analysis.

An important step in the process comprises the assessment of the extent to which the aim of the cluster analysis has been attained applying the specific algorithm. Several decomposition quality functionals have been proposed for this purpose. The most frequently applied criteria comprise characteristics such as the matrices of intragroup variability

$$\mathbf{E} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)\, (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)^{\mathrm{T}} \tag{1}$$

and the matrices of intergroup variability

$$\mathbf{B} = \sum_{h=1}^{k} n_h (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})\, (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})^{\mathrm{T}}, \tag{2}$$

which, together, provide the matrix of the overall variability

$$\mathbf{T} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}})\, (\mathbf{x}_{hi} - \bar{\mathbf{x}})^{\mathrm{T}}, \qquad \mathbf{T} = \mathbf{E} + \mathbf{B}. \tag{3}$$

In relations (1) – (3) they appear as:

$\mathbf{x}_{hi}$ … observation vector of the *i*-th object in the *h*-th cluster;

$\bar{\mathbf{x}}_h$ … average vector for the *h*-th cluster;

$\bar{\mathbf{x}}$ …vector averages for the whole set.

If we apply the *p* variables, they concern the *p*-member vectors, and **E**, **B** and **T** represent the symmetric square matrices of *p*-th order.

The basic aim is to construct maximally distant compact clusters, which is attained provided that the minimum of the total sum of the squares of the deviations of all the values from the respective cluster averages is achieved[1].

$$G_1 = st\ \mathbf{E} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} \sum_{j=1}^{p} \left(x_{hij} - \bar{x}_{hj}\right)^2. \tag{4}$$

The G_1 criterion is known as the Ward criterion. In addition to this criterion, it is also possible to use its simple monotonic functions. Since st **T** is the same for all the decompositions, the minimising of st **E** is the same as the maximising of st **B**.

Should it be necessary to attain the independence of the applied units of measurement (more generally invariance to linear transformations), it is possible to recommend the minimising of the determinant of the intragroup variability matrix $G_2 = |\mathbf{E}|$ or the maximising of the trace criterion $G_3 = \mathrm{st}\,(\mathbf{B}\mathbf{E}^{-1})$ or $G_4 = \mathrm{st}\,(\mathbf{B}\mathbf{T}^{-1})$.

Following the selection of the variables for the characterisation of the properties of the clustered objects and the determination of their values, it is necessary to decide on the method to be applied for the evaluation of the distance between or similarity of the objects. Most commonly, the initial stage of the implementation of cluster algorithms comprises the calculation of the appropriate measurements for all the pairs of objects. This creates a symmetrical square matrix of type $n \times n$, along the diagonal of which are zeros in the case of the distance measurement matrix **D**, or ones in the case of the similarity measurement matrix **A**.

If the individual variables are roughly at the same level or are, at least, expressed in the same units, it is possible to apply the Hemming distance, which is also referred to in the literature as the Manhattan or city-block distance

[1] In relation (4) the expression "st **E**" is the trace of matrix **E**.

$$D_H(\mathbf{x}_i, \mathbf{x}_{i\backslash}) = \sum_{j=1}^{p} |x_{ij} - x_{i\backslash j}|, \tag{5}$$

the Euclidean distance

$$D_E(\mathbf{x}_i, \mathbf{x}_{i\backslash}) = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i\backslash j})^2}, \tag{6}$$

or the Chebyshev distance

$$D_c(\mathbf{x}_i, \mathbf{x}_{i\backslash}) = \max_j |x_{ij} - x_{i\backslash j}|. \tag{7}$$

We choose between these distances according to the extent to which the influence of the variables for which we observe an extremely large difference on the total sum needs to be strengthened.

The most frequently applied procedures used in cluster analysis include the creation of a hierarchical sequence of decompositions - hierarchical clustering, which can be described as follows:

1. Calculation of the **D** matrix of the appropriate distance measurements.
2. Commencement of the process from the $S^{(n)}$ decomposition, i.e. from the n clusters, each of which contains one object.
3. The searching of the **D** matrix to find the two clusters (the h-th and the $h^{\backslash}$-th), the $D_{hh^{\backslash}}$ distance between which is minimal.
4. The joining of the h-th and the $h^{\backslash}$-th clusters so as to form a new g-th cluster. We remove the h-th and $h^{\backslash}$-th rows and columns in matrix **D** and replace them with rows and columns for the new cluster. The order of matrix **D** has decreased by one.
5. The noting of the order of the v = 1, 2,…, $n - 1$ cycle and the identification of the h and $h^{\backslash}$ connected objects and the level for the $d_1 = D_{hh^{\backslash}}$. connection
6. If the decomposition process has not ended by the merging of all the objects into a single $S^{(1)}$cluster, we proceed to step 3.

The above comprises an <u>agglomerative hierarchical procedure</u> (also referred to as AGNES − AGlomerative NESting), a less frequently applied alternative to which comprises the divisive hierarchical procedure (also referred to as DIANA − DIvisive ANAlysis). This is the opposite procedure, concerning which we commence from a single $S^{(1)}$ cluster; each step then involves the splitting of one of the clusters into two clusters, thus at the end of the process we obtain $S^{(n)}$.

The results of hierarchical clustering procedures can be graphically represented in the form of a tree diagram known as a <u>dendrogram</u>, the horizontal axis of which comprises a scale for the joining of the various levels. The tree starts on the left with n branches; subsequently, two branches join in each step at the point that corresponds to the connection level.

When considering the variables X_1, X_2, …, X_p and given the distances matrix **D**, the results of the implementation of the described algorithm differ in terms of the procedure with which we assess the distances between the clusters.

If the method is given for evaluating the proximity or similarity of the clusters, which also results in a way in which to recalculate the distance matrix within each cycle, it is possible to apply the described algorithm in the considered task to construct a hierarchical sequence of decompositions and thus create a dendrogram.

The distance between the clusters in the described cluster algorithm was derived from the distances between the objects. Such distances for two connected clusters are, in total, $q = n_h\, n_{h^{\backslash}}$, with the smallest of these distances being selected. However, the maximum or average distance could just as well be used, thus resulting in three different procedures. Moreover, a number of other such procedures are also available.

Nearest neighbour method (single link, single join)
This method comprises the oldest method available; in this case, the two clusters we are considering are represented by the objects that are closest to each other. The distance between the h-th and the $h^\backslash$-th cluster $D_{hh^\backslash}$ represents the minimum of all the $q = n_h\, n_{h^\backslash}$ distances between their objects. Thus, the procedure in the third step of the algorithm is specified. The fourth step consists of the replacement of the h-th and $h^\backslash$-th row and column in the distance matrix with the row and column of the distances of the new g-th cluster from the other clusters (denoted $g^\backslash$). The total of the $n - v - 1$ distances is written in the v-th cycle determined on the basis of

$$D_{gg^\backslash} = min\left(D_{g^\backslash h}, D_{g^\backslash h^\backslash}\right) \tag{8}$$

When using this method, the situation often arises that even very distant objects may meet in the same cluster if a large number of other objects establish a bridge between them. This characteristic "chaining" is considered to be a disadvantage, especially if we have a reason to require that the clusters have an elliptical shape with a compact core.

Furthest neighbour method (full links, full joins)
This method is based on the opposite principle; the criterion for joining the clusters is the maximum of the $q = n_h\, n_{h^\backslash}$ possible inter-cluster distances of the objects. As part of the adjustment of the distance matrix, we proceed applying

$$D_{gg^\backslash} = max\left(D_{g^\backslash h}, D_{g^\backslash h^\backslash}\right) \tag{9}$$

The undesirable chaining effect is eliminated in this case; indeed, conversely, there is a tendency towards the formation of compact clusters rather than extraordinarily large clusters.

Average binding method (Sokal-Sneath method)
This method considers the average from $q = n_h n_{h^\backslash}$of the possible inter-cluster distances of the objects as the criterion for the joining of the clusters. We use the following for the recalculation of the distance matrix

$$D_{gg^\backslash} = \frac{n_h D_{g^\backslash h} + n_{h^\backslash} D_{g^\backslash h^\backslash}}{n_h + n_{h^\backslash}}\,. \tag{10}$$

This method often leads to analogous results as with the furthest neighbour method.

Centroid method (Gower method)
Rather than being based on the summarising of information on the inter-cluster distances of the objects, the criterion applied in this method is the Euclidean distance of the centroids

$$D_E\left(\bar{\mathbf{x}}_h, \bar{\mathbf{x}}_{h^\backslash}\right) = \sum_{j=1}^{p} \left(\bar{x}_{hj} - \bar{x}_{h^\backslash j}\right)^2, \tag{11}$$

the recalculation of the distance matrix is performed employing

$$D_{gg^\backslash} = \frac{1}{n_h + n_{h^\backslash}} \left(n_h D_{g^\backslash h} + n_{h^\backslash} D_{g^\backslash h^\backslash} - \frac{n_h\, n_{h^\backslash}}{n_h + n_{h^\backslash}}\, D_{hh^\backslash}\right). \tag{12}$$

Ward method

This method applies the G_1 decomposition quality functional from equation (4); the criterion for joining the clusters comprises the increment of the total intragroup sum of the squares of the deviations observed from the cluster average

$$\Delta G_1 = \sum_{i=1}^{g}\sum_{j=1}^{p}\left(x_{gij} - \overline{x}_{gj}\right)^2 - \sum_{i=1}^{h}\sum_{j=1}^{p}\left(x_{hij} - \overline{x}_{hj}\right)^2 - \sum_{i=1}^{h\backslash}\sum_{j=1}^{p}\left(x_{h\backslash ij} - \overline{x}_{h\backslash j}\right)^2. \tag{13}$$

The increment is expressed as the sum of the squares in the newly-emerging cluster minus the sums of the squares for the two disappearing clusters. Equation (13) can be simplified applying elementary arithmetic adjustments to the following form

$$\Delta G_1 = \frac{n_h\, n_{h\backslash}}{n_h + n_{h\backslash}} \sum_{j=1}^{p}\left(\overline{x}_{hj} - \overline{x}_{h\backslash j}\right)^2, \tag{14}$$

as the product of the Euclidean distance between the centroids of the clusters considered for the connection and the coefficient that presents the dependence on the cluster size, so that the value of a given coefficient increases with the increasing cluster size, and it is maximal for fixed $n_h + n_{h\backslash}$ if we have clusters of the same size $n_h = n_{h\backslash}$. Due to the fact that we form a connection that ensures the minimisation of the $\Delta\, G_1$ criterion, Ward's method often has the welcome feature of a tendency to remove small clusters, and thus to form clusters of approximately the same size. If we apply Ward's method based on the matrix of Euclidean distances between the objects, it is possible to apply the relationship

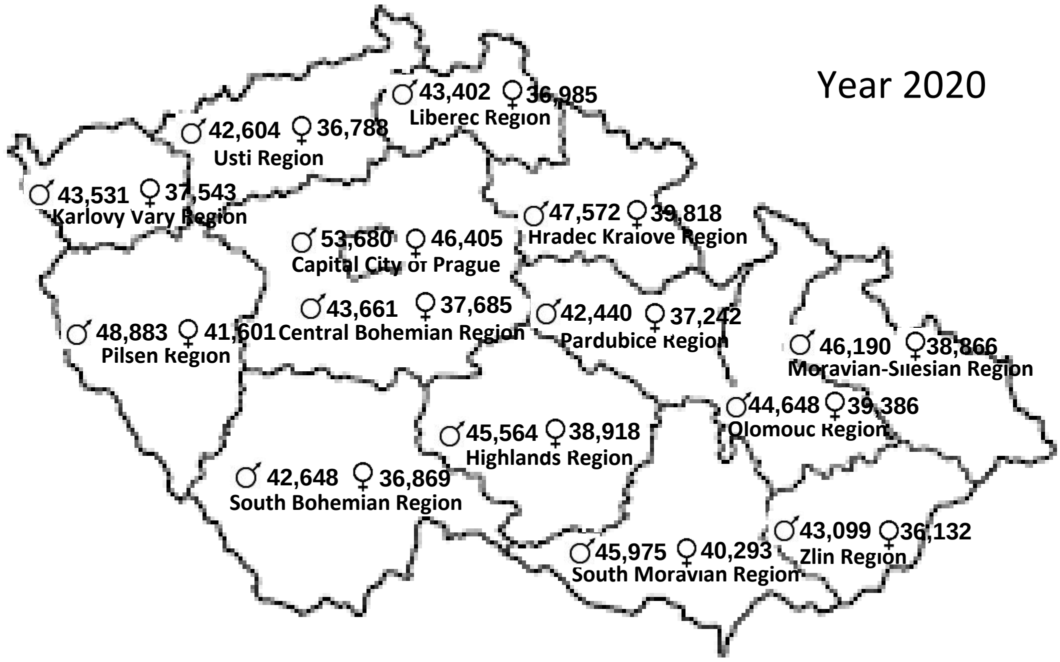
$$D_{gg\backslash} = \frac{1}{n_h + n_{h\backslash} + n_{g\backslash}}\left[\left(n_h + n_{g\backslash}\right)D_{hg\backslash} + \left(n_{h\backslash} + n_{g\backslash}\right)D_{h\backslash g\backslash} - n_{g\backslash}D_{hh\backslash}\right]. \tag{15}$$

III. Results

Figures 1 and 2 present the salaries of employees in 2020 and 2019 according to both regions and gender. A significant gender pay gap is evident for both years under review in all fourteen regions of the Czech Republic. Moreover, it is clear from the two figures that employees in the Prague region consistently earn the highest salaries of all fourteen regions.

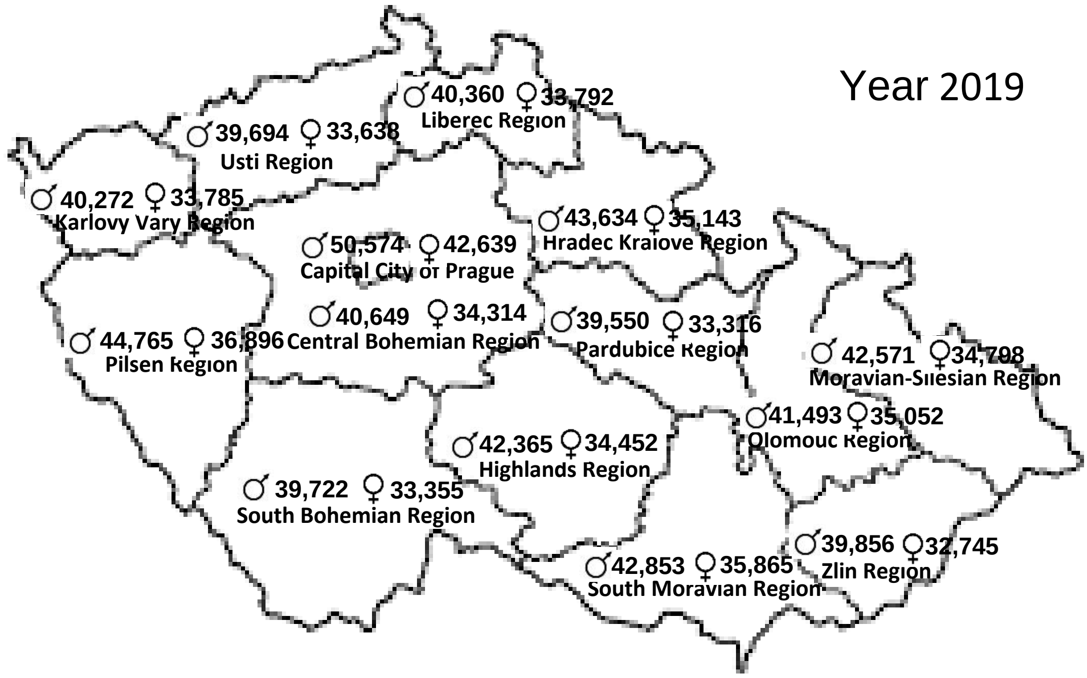
Figures 3 to 6 present the results of the cluster analysis in the form of dendrograms. The clusters are created based on gross monthly salary of employees, their age, education, and number of years with the same employer in the public sector. All four of these variables have identical weights and are therefore equally important.

Figure 1 Average salary by region and gender in 2020



Source: Own research

Figure 2 Average salary by region and gender in 2019

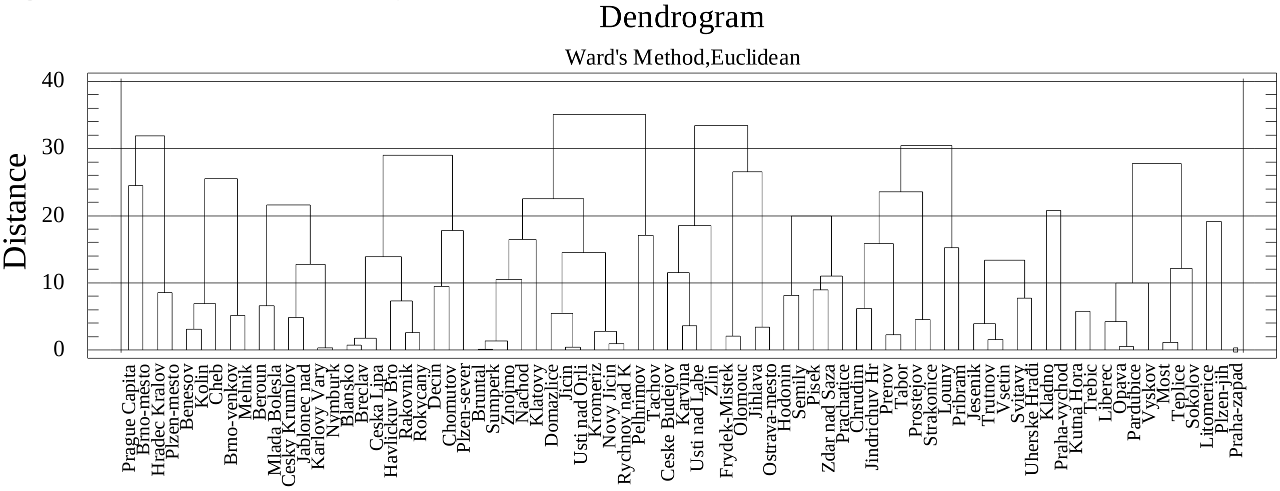


Source: Own research

The classification of the various districts of the Czech Republic into regions is presented in Table A of the Appendix. It is clear from the table that the newly-created clusters of districts do not overlap with the structure of the official geographical regions shown in Table 1. It is clear from Table A of the Appendix that in terms of both data sets for women (2019 and 2020) Prague forms a separate cluster, which clearly results from the significantly above-average salaries earned in the country's capital city compared to the other districts of the country. A further factor most likely comprises the overall higher level of education in Prague compared to the rest of the country.

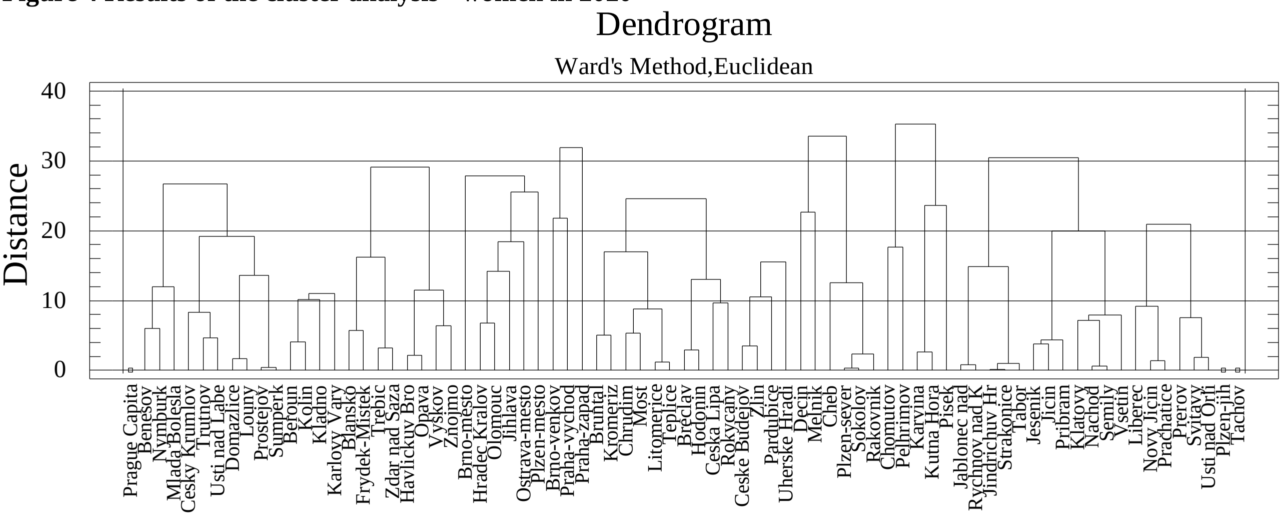
With respect to the two data sets for men (2019 and 2020), in both cases Prague forms a joint cluster with three other districts, the capital cities of which are also the regional capitals of their respective regions, i.e. Brno-mesto, Hradec Kralove and Pilsen-mesto, i.e. three of the largest cities in the Czech Republic.

Figure 3 Results of the cluster analysis - men in 2020



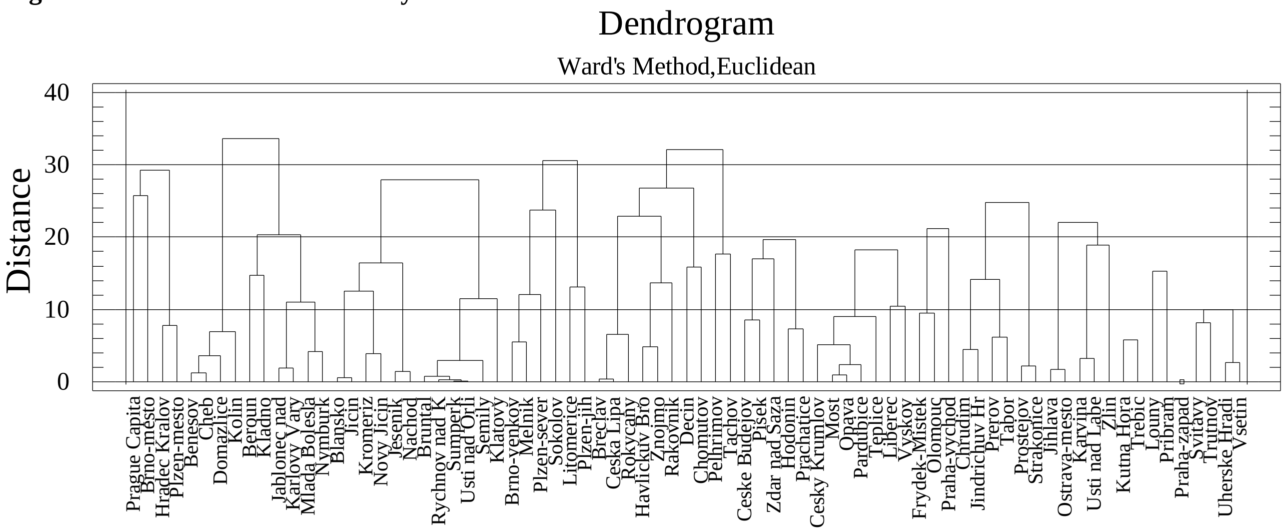
Source: Own research

Figure 4 Results of the cluster analysis - women in 2020



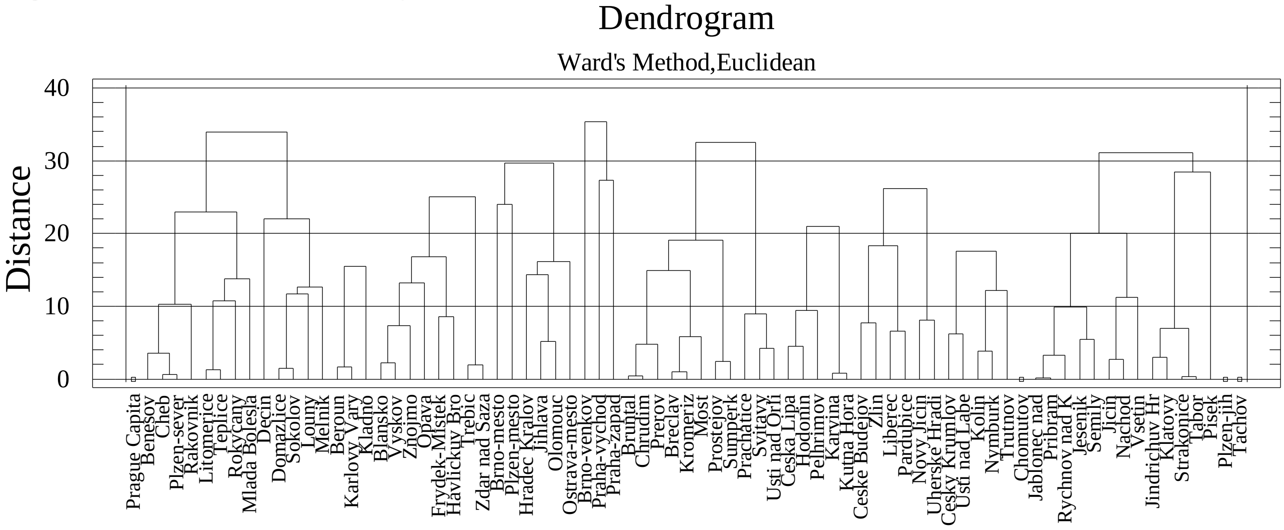
Source: Own research

Figure 5 Results of the cluster analysis - men in 2019



Source: Own research

Figure 4 Results of the cluster analysis - women in 2019



Source: Own research

The comparison of the composition of clusters for the data sets on men in 2019 and 2020 reveals groups of districts that appear in the same clusters in both years:

1. Bruntal, Jicin, Klatovy, Kromeriz, Nachod, Novy Jicin, Rychnov nad Kneznou, Sumperk and Usti nad Orlici;
2. Chrudim, Jindrichuv Hradec, Prerov, Prostejov, Strakonice and Tabor;
3. Liberec, Most, Opava, Pardubice, Teplice and Vyskov;
4. Beroun, Jablonec nad Nisou, Karlovy Vary, Mlada Boleslav and Nymburk;
5. Chomutov, Decin, Havlickuv Brod, Rakovnik and Rokycany;
6. Jihlava, Karvina, Ostrava-mesto, Usti nad Labem and Zlin;
7. Brno-mesto, Hradec Kralove, Pilsen-mesto and Prague Capital;
8. Hodonin, Pisek, Prachatice and Zdar nad Sazavou;
9. Svitavy, Trutnov, Uherske Hradiste and Vsetin;
10. Benesov, Cheb and Kolin;

11. Pelhrimov, Tachov and Znojmo;
12. Breclav and Ceska Lipa;
13. Brno-venkov and Melnik;
14. Frydek-Mistek and Olomouc;
15. Kutna Hora and Trebic;
16. Litomerice and Pilsen-jih;
17. Louny and Pribram;
18. Prague-zapad.

Similarly, the comparison of the composition of the clusters identified for women in 2019 and 2020 also reveals groups of districts that appear in the same clusters in both years:

1. Jablonec nad Nisou, Jesenik, Jicin, Jindrichuv Hradec, Klatovy, Nachod, Pribram, Rychnov nad Kneznou, Semily, Strakonice, Tabor and Vsetin;
2. Blansko, Frydek-Mistek, Havlickuv Brod, Opava, Trebic, Vyskov, Zdar nad Sazavou and Znojmo;
3. Brno-mesto, Hradec Kralove, Jihlava, Olomouc, Ostrava-mesto and Pilsen-mesto;
4. Cheb, Decin, Melnik, Pilsen-sever, Rakovnik and Sokolov;
5. Benesov, Domazlice, Louny and Mlada Boleslav;
6. Bruntal, Breclav, Chrudim, and Kromeriz;
7. Ceske Budejovice, Pardubice, Uherske Hradiste and Zlin;
8. Beroun, Karlovy Vary and Kladno;
9. Brno-venkov, Prague-vychod and Prague-zapad;
10. Karvina, Kutna Hora and Pelhrimov;
11. Nymburk, Trutnov and Usti nad Labem;
12. Prerov, Svitavy and Usti nad Orlici;
13. Ceska Lipa and Hodonin;
14. Liberec and Novy Jicin;
15. Prostejov and Sumperk;
16. Rokycany and Teplice;
17. Chomutov;
18. Pilsen-jih;
19. Tachov.

It is clear from the results of the analysis that only exceptionally do districts that are located in the same region appear in the same cluster. It thus follows that the geographical location of the district does not reflect its similarity with other districts in the same region in terms of the gross monthly salary of employees, their age, education and length of employment with the same employer. Large district cities, which are also regional cities, tend to form groups (see group number three of the women’s data set), as do districts just outside the boundaries of large cities (see group number nine of the women’s data set). It can be said that districts with approximately the same population tend to belong to the same cluster. Other characteristics of districts belonging to the same cluster were not observed. In the case of a dataset of women, Prague always forms a separate cluster in both monitored years. The reason is the specific character of Prague, where the level of salaries is extremely high compared to other areas of the Czech Republic. Young university graduates also often move to Prague, which makes Prague very different from the rest of the Czech Republic, too.

Another relationship between individual clusters cannot be observed from such an analysis. This require a detailed analysis of the distance matrices. Another possibility in this direction is to divide the districts into a smaller number of clusters (using otherwise the same procedures) and to monitor the division of the districts into a smaller number of clusters in comparison with the first fourteen clusters. This topic is a possible issue for further research and a separate publication.

The cluster analysis also revealed a number of changes that took place between 2019 and 2020 in terms of the structures and sizes of certain clusters in both groups; however, with respect to both the men's and the women's data sets for both years, a significant number of districts continued to appear in the same clusters.

IV. Conclusion

This study provides a basic view of clusters of Czech districts that are most similar in terms of the level of salaries of employees in the public sphere, their age, education and length of employment, i.e. the number of years worked for the same employer. In a similar way to Morariu, (2015), this paper focuses only on employees in the public sector. According to Lee & Sabharwal, (2016), significant differences exist between the development of salaries in the private sector and the public sector. As shown by Amanatullah & Morris (2010) and Jena, Olenski & Blumenthal (2016), our research also revealed a significant gender pay gap in the public sector as a result of the separate cluster analysis of men and women.

The results obtained clearly suggest that the clusters of districts do not coincide with the district structures of the fourteen regions of the Czech Republic. Although a number of differences were identified in the structures of district clusters between 2019 and 2020, when the start of the coronavirus crisis affected all the world's economies, the majority of districts appeared in the same clusters in both of the monitored years.

The composition of clusters of districts of the Czech Republic for men clearly differs from the composition of clusters of districts of the Czech Republic for women. The reason is a completely different behaviour of the whole salary distribution, which is characterized by a higher level and variability for men and, conversely, lower skewness and kurtosis compared to the whole salary distribution for women in the same category (education in this case).

It is anticipated that the results of the research will form the basis e.g. for decision-making at the national level or for the development of regional development policy measures.

References

Amanatullah, E. T. & Morris, M. W. (2010). Negotiating Gender Roles: Gender Differences in Assertive Negotiating Are Mediated by Women's Fear of Backlash and Attenuated when Negotiating on Behalf of Others. *Journal of Personality and Social Psychology*, 98(2), 256–267.

Ariffin, H. F. & Ha, N. C. (2014). Examining Malaysian Hotel Employees Organizational Commitment: The Effect of Age, Education Level and Salary. *International Journal of Business and Technopreneurship*, 4(3), 413–438.

Beran, V. & Godarová, J. (2017). The Impact of Population Aging on GDP in the Czech Republic. In: Economic Policy in Global Environment 2017, *Conference Proceeding*, Ostravice, 05.09.2017–07.09.2017, 47–60.

Bílková, D. (2016). Vliv délky praxe na výši mezd zaměstnanců v České republice od počátku hospodářské recese v roce 2009 **(**Influence of Length of Practice on Wages of Employees in the Czech Republic in the Period from the Beginning of the Economic Recession in 2009**).** *Politická ekonomie*, 64(6), 674–694.

Bílková, D. (2015). Financial Position of Czech Employees at the Beginning of the 3rd Millennium According to Educational Attainment. *Prague Economic Papers*, 24(3), 307–331.

Carraher, S. M. (2011). Turnover Prediction Using Attitudes Towards Benefits, Pay, and Pay Satisfaction among Employees and Entrepreneurs in Estonia, Latvia, and Lithuania. *Baltic Journal of Management*, 6(1), 25–52.

Červenka, F. (2021). Employer of Last Resort for the Czech Republic. *Prague Economic Papers*, 30(6), 748–767.

Gimpelson, V. (2019). Age and Wage: Stylized Facts and Russian Evidence. *HSE Economic Journal*, 23(2), 185–237.

Jena, A. B., Olenski, A. R. & Blumenthal, D. M. (2016). Sex Differences in Physician Salary in US Public Medical Schools. *JAMA Internal Medicine*, 176(9), 1 294–1 304.

Lee, Y. & Sabharwal, M. (2016). Education-Job Match, Salary, and Job Satisfaction across the Public, non-Profit, and for-Profit Sectors: Survey of Recent College Graduates. *Public Management Review*, 18(1), 40–64.

Lleras, C. (2008). Do Skills and Behaviors in High School Matter? The Contribution of Noncognitive Factors in Explaining Differences in Educational Attainment and Earnings. *Social Science Research*, 37(3), 888–902.

Morariu, A. (2015). Implications of the Evolution of the Romanian Economy Regarding the Personnel Salary in the Public Sector. *Ecoforum Journal*, 4(1), 226–231.

Ruiz, A. C., Gómez, L. N. & Narváez, M. R. (2010). Endogenous Wage Determinants and Returns to Education in Spain. *International Journal of Manpower*, 31(4), 410–425.

Sargeant, M. (2010). The UK National Minimum Wage and Age Discrimination. *Policy Studies*, 31(3), 351–364.

Appendix

Table A Results of the cluster analysis

Men 2020	Women 2020	Men 2019	Women 2019
1st cluster:	1st cluster:	1st cluster:	1st cluster:
1. Prague Capital	1. Prague Capital	1. Prague Capital	1. Prague Capital
2. Brno-mesto	2nd cluster:	2. Brno-mesto	2nd cluster:
3. Hradec Kralove	1. Benesov	3. Hradec Kralove	1. Benesov
4. Pilsen-mesto	2. Cesky Krumlov	4. Pilsen-mesto	2. Decin
2nd cluster:	3. Domazlice	2nd cluster:	3. Domazlice
1. Benesov	4. Louny	1. Benesov	4. Cheb
2. Brno-venkov	5. Mlada Boleslav	2. Beroun	5. Litomerice
3. Cheb	6. Nymburk	3. Domazlice	6. Louny
4. Kolin	7. Prostejov	4. Cheb	7. Melnik
5. Melnik	8. Sumperk	5. Jablonec nad Nisou	8. Mlada Boleslav
3rd cluster:	9. Trutnov	6. Karlovy Vary	9. Pilsen-sever
1. Beroun	10. Usti nad Labem	7. Kladno	10. Rakovnik
2. Cesky Krumlov	3rd cluster:	8. Kolin	11. Rokycany

3. Jablonec nad Nisou	1. Beroun	9. Mlada Boleslav	12. Sokolov
4. Karlovy Vary	2. Karlovy Vary	10. Nymburk	13. Teplice
5. Mlada Boleslav	3. Kladno	3rd cluster:	3rd cluster:
6. Nymburk	4. Kolin	1. Blansko	1. Beroun
4th cluster:	4th cluster:	2. Bruntal	2. Karlovy Vary
1. Blansko	1. Blansko	3. Jesenik	3. Kladno
2. Breclav	2. Frydek-Mistek	4. Jicin	4th cluster:
3. Ceska Lipa	3. Havlickuv Brod	5. Klatovy	1. Blansko
4. Decin	4. Opava	6. Kromeriz	2. Frydek-Mistek
5. Havlickuv Brod	5. Trebic	7. Nachod	3. Havlickuv Brod
6. Chomutov	6. Vyskov	8. Novy Jicin	4. Opava
7. Pilsen-sever	7. Znojmo	9. Rychnov nad Kneznou	5. Trebic
8. Rakovnik	8. Zdar nad Sazavou	10. Semily	6. Vyskov
9. Rokycany	5th cluster:	11. Sumperk	7. Znojmo
5th cluster:	1. Brno-mesto	12. Usti and Orlici	8. Zdar and Sazavou
1. Bruntal	2. Hradec Kralove	4th cluster:	5th cluster:
2. Domazlice	3. Jihlava	1. Brno-venkov	1. Brno-mesto
3. Jicin	4. Olomouc	2. Litomerice	2. Hradec Kralove
4. Klatovy	5. Ostrava-mesto	3. Melnik	3. Jihlava
5. Kromeriz	6. Pilsen-mesto	4. Pilsen-jih	4. Olomouc
6. Nachod	6th cluster:	5. Pilsen-sever	5. Ostrava-mesto
7. Novy Jicin	1. Brno-venkov	6. Sokolov	6. Pilsen-mesto
8. Pelhrimov	2. Prague-vychod	5th cluster:	6th cluster:
9. Rychnov nad Kneznou	3. Prague-zapad	1. Breclav	1. Brno-venkov
10. Sumperk	7th cluster:	2. Ceska Lipa	2. Prague-vychod
11. Tachov	1. Bruntal	3. Decin	3. Prague-zapad
12. Usti nad Orlici	2. Breclav	4. Havlickuv Brod	7th cluster:
13. Znojmo	3. Ceska Lipa	5. Chomutov	1. Bruntal
6th cluster:	4. Hodonin	6. Pelhrimov	2. Breclav
1. Ceske Budejovice	5. Chrudim	7. Rakovnik	3. Chrudim
2. Frydek-Mistek	6. Kromeriz	8. Rokycany	4. Kromeriz
3. Jihlava	7. Litomerice	9. Tachov	5. Most
4. Karvina	8. Most	10. Znojmo	6. Prachatice
5. Olomouc	9. Rokycany	6th cluster:	7. Prostejov
6. Ostrava-mesto	10. Teplice	1. Ceske Budejovice	8. Prerov
7. Usti nad labem	8th cluster:	2. Hodonin	9. Svitavy
8. Zlin	1. Ceske Budejovice	3. Pisek	10. Sumperk
7th cluster:	2. Pardubice	4. Prachatice	11. Usti nad Orlici
1. Hodonin	3. Uherske Hradiste	5. Zdar nad Sazavou	8th cluster:
2. Pisek	4. Zlin	7th cluster:	1. Ceska Lipa
3. Prachatice	9th cluster:	1. Cesky Krumlov	2. Hodonin
4. Semily	1. Decin	2. Liberec	3. Karvina
5. Zdar nad Sazavou	2. Cheb	3. Most	4. Kutna Hora
8th cluster:	3. Melnik	4. Opava	5. Pelhrimov
1. Chrudim	4. Pilsen-sever	5. Pardubice	9th cluster:
2. Jindrichuv Hradec	5. Rakovnik	6. Teplice	1. Ceske Budejovice

3. Louny	6. Sokolov	7. Vyskov	2. Liberec
4. Prostejov	10th cluster:	8th cluster:	3. Novy Jicin
5. Prerov	1. Chomutov	1. Frydek-Mistek	4. Pardubice
6. Pribram	2. Karvina	2. Olomouc	5. Uherske Hradiste
7. Strakonice	3. Kutna Hora	3. Prague-vychod	6. Zlin
8. Tabor	4. Pelhrimov	9th cluster:	10th cluster:
9th cluster:	5. Pisek	1. Chrudim	1. Cesky Krumlov
1. Jesenik	11th cluster:	2. Jindrichuv Hradec	2. Kolin
2. Svitavy	1. Jablonec nad Nisou	3. Prostejov	3. Nymburk
3. Trutnov	2. Jesenik	4. Prerov	4. Trutnov
4. Uherske Hradiste	3. Jicin	5. Strakonice	5. Usti nad Labem
5. Vsetin	4. Jindrichuv Hradec	6. Tabor	11th cluster:
10th cluster:	5. Klatovy	10th cluster:	1. Chomutov
1. Kladno	6. Nachod	1. Jihlava	12th cluster:
2. Prague-vychod	7. Pribram	2. Karvina	1. Jablonec nad Nisou
11th cluster:	8. Rychnov nad Kneznou	3. Ostrava-mesto	2. Jesenik
1. Kutna Hora	9. Semily	4. Usti nad Labem	3. Jicin
2. Trebic	10. Strakonice	5. Zlin	4. Jindrichuv Hradec
12th cluster	11. Tabor	11th cluster:	5. Klatovy
1. Liberec	12. Vsetin	1. Kutna Hora	6. Nachod
2. Most	12th cluster:	2. Trebic	7. Pisek
3. Opava	1. Liberec	12th cluster:	8. Pribram
4. Pardubice	2. Novy Jicin	1. Louny	9. Rychnov and Kneznou
5. Sokolov	3. Prachatice	2. Pribram	10. Semily
6. Teplice	4. Prerov	13th cluster:	11. Strakonice
7. Vyskov	5. Svitavy	1. Prague-zapad	12. Tabor
13th cluster:	6. Usti nad Orlici	14th cluster:	13. Vsetin
1. Litomerice	13th cluster:	1. Svitavy	13th cluster:
2. Pilsen-jih	1. Pilsen-jih	2. Trutnov	1. Pilsen-jih
14th cluster:	14th cluster:	3. Uherske Hradiste	14th cluster:
1. Prague-zapad	1. Tachov	4. Vsetin	1. Tachov

Source: Own research