Annals of "Dunarea de Jos" University of Galati
Fascicle I. Economics and Applied Informatics
Years XXVII – nº3/2021
ISSN-L 1584-0409 ISSN-Online 2344-441X
www.eia.feaa.ugal.ro

DOI https://doi.org/10.35219/eai15840409220

Cluster Analysis on the Unemployment Rate during Lockdown Period

Ludmila-Daniela MANEA★, Iulian Adrian SORCARU★★

ARTICLE INFO

Article history:
Accepted October 2021
Available online December 2021
JEL Classification
P16, H11, H60

Keywords:
unemployment, lockdown, crisis, Covid-19 , labor force, Cluster analysis on the unemployment rate during lockdown period

ABSTRACT

The problem of unemployment is a permanent one because elements of imbalance appear in the process of macro-stability. In the context of a free market, we can also discuss about the labour market, which can no longer be regulated administratively, so that labour is absorbed, reconverted and unemployment is eliminated as much as possible, leaving supply and demand to settle employment. Unemployment is an element of balance in a certain way, because it regulates the need for labour with the needs and the existence on the labour market of available people, but they should be in the structure by profession, by trade, required by the market. From time to time, when macroeconomic destabilization occurs, unemployment as the valve for the use of the working population makes its presence felt. Under the current circumstances, in which the world economy, of the European countries and consequently also of Romania are under the effect of the coronavirus health crisis, combined with the financial and economic crisis, inflation has become a growing phenomenon, being even worrying. The unemployed population receives for a limited period of time a payment, after which it becomes an unoccupied free population. A paradox is also that as the quotas of the unemployed increase and there are vacancies that are taken out to be filled, the latter are not completed, because the structure of the places taken out to be filled does not correspond to the needs of the economy in terms of structure by trades, by professions. During the current crisis that we have spoken of, it is very clear that unemployment has increased quite a lot compared to January this year, but even more worrying is the fact that a large number, more than 1.3 million people, are in technical unemployment, who will eventually be able to supply the number of people who get unemployed. In order to consider how the labor resource can be used in Romania, we must take into account that over 1.5 million people, who returned from the European states where they worked, not on the basis of contract and precise commitments, returned to the country being part of the number of those unoccupied people.

1. Introduction

Unlike the global financial crisis, which has been caused by developed weaknesses in the financial sector, the current economic shock is affecting the entire economy by disrupting private consumption, trade operations and global supply chains.

EU-27 GDP was contracted at 7,4% in 2020 and may not return to its pre-crisis level in 2021, creating a risk of significant increases in insolvency cases and job cuts.

The causes of the current economic crisis are unique in modern history. While recent crises have been caused by an unsustainable increase in demand and/or accumulation of macro-financial imbalances, the current significant drop in production is caused by the effects of the pandemic and health policy measures on both demand and supply.

Despite the fact that the nature of the current crisis is different from that of the previous crisis, since 2008-2013, there is a close correlation between the decreases in GDP in the Member States, which may create new challenges for economic convergence within the EU. For example, Greece, Italy and Spain seem to be one of the most affected economies in both crises.

Since March 2020, the economic shock caused by measures limiting the movement of people has required large-scale economic policy measures at EU and national level in response to the crisis caused by the COVID-19 pandemic to compensate for income losses for businesses and households.

In terms of unemployment, the countries worst affected in 2020 are: Spain (15.54%), and Greece (16.45%), all of which exceed the EU average (Figure 1). For four countries, the unemployment rate in 2020 is more than double that of 2019 (the Czech Republic, Hungary, Poland and Estonia, with low pre-crisis

★,★★Dunarea de Jos University of Galati, Romania. E-mail addresses: daniela.manea@ugal.ro (Corresponding author – L. D. Manea), iulian.sorcaru@ugal.ro (I. A. Sorcaru).

unemployment rates). Although unemployment is forecast for 2021 in all countries in Europe, unemployment rates remain above 2019 levels.

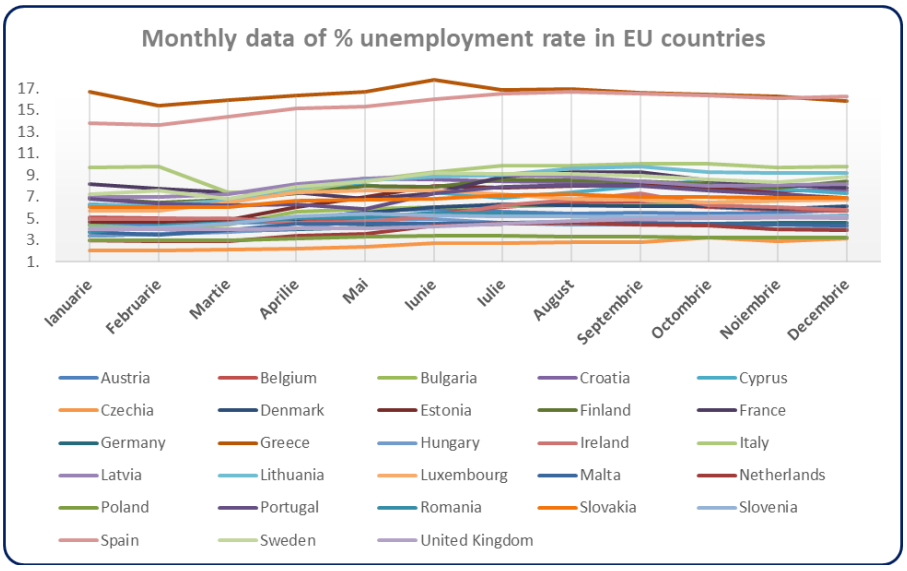

Figure 1. Monthly data of % unemployment rate

In Romania, one of the most serious economic effects of the coronavirus pandemic is the increase in unemployment, as the activity of some economic operators has been affected by closures and restrictions. In fact, the unemployment rate was only 3.9% at the end of 2019, but increased to 5.5% in June 2020.

The European Commission anticipates that the upward trend in the unemployment rate will continue throughout the next year. Thus, the European institution expects an unemployment rate of 6.2% at the end of 2021, probably an even higher value in the middle of the year before a slight decrease in the second part.

The unprecedented growth in the inactive population since February has led to the disappearance of more than 30 million employees from official unemployment statistics in twenty-five OECD (Organization for Economic Co-operation and Development) member countries and emerging markets. If we look beyond the main indicators of the labor market, we find that the true impact of unemployment on household consumption could be underestimated by USD 14 billion. This is just the tip of the iceberg: so far the labor markets of many countries have proved much more resilient than economic activity - at least on the surface. However, given the long-term relationship between GDP growth and changes in the unemployment rate, the true impact of Covid-19 on the labour market in some countries should be much more serious than what the main indicators suggest.

The number of unemployed people who have appeared in official statistics since February 2020 has reached 2.4 million in France, Italy, Spain and Ireland. Apart from the discouraging effect of the pandemic state of work, the sharp increase in inactivity can be mainly explained by the characteristics of the Covid-19 recession: severe restrictions on movement, unavailability of government services as well as no job search due to the responsibility of childcare in the context of school closures.

The increases in the actual unemployment rate by mid-2020 differ considerably from one EU Member State to another. A substantial part of these differences between Member States during the crisis caused by the COVID-19 pandemic can be statistically linked to an index of the degree of stringency of measures to limit the spread or social distance and the share of tourism in the economy.

For example, the closure measures adopted by Germany in March were less stringent than those adopted by Italy, France or Spain, where all non-essential undertakings were obliged to close down, This partly explains the stronger GDP declines in the latter countries (see also Figure 2). Preliminary estimates show that each month of strict isolation leads to an increase in the unemployment rate of up to two percentage points.

Short-time work schemes are public programs that allow companies in economic difficulties to temporarily reduce the number of hours worked, while at the same time providing their employees with income support from the State for the hours not worked. Such programs may involve a partial reduction in the number of hours worked or a complete suspension of the employment contract for a limited period. The aim is to preserve employment during the crisis and to enable companies to retain their skills, expertise and labor.

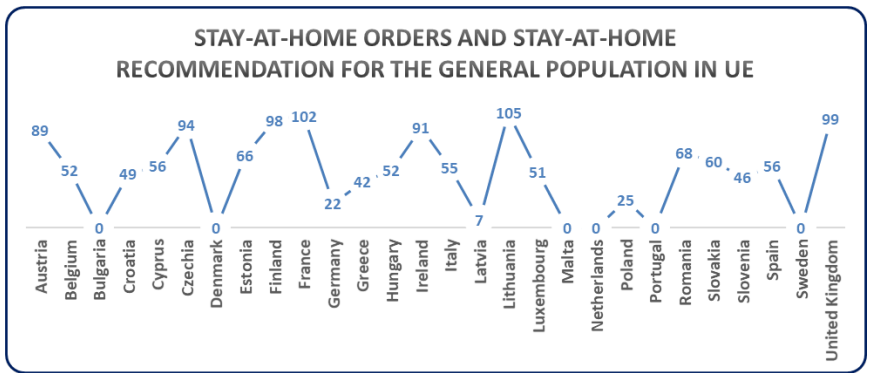

Figure 2. Lockdown in EU contries

On 13 March 2020, the European Commission recommended the use of these schemes to tackle the adverse effects of measures to limit the movement of people on employment. More than 42 million workers were unemployed in the EU-27 in April 2020, at the height of the containments of the movement of people as a result of the pandemic, which accounted for around a quarter of the total workforce. According to the Commission's estimates, the immediate budgetary cost of short-time work schemes is half the cost of unemployment benefits that would be paid without this employment support. The total value of the measures included in the survey aimed at directly supporting employment through short-time work schemes is 184 billion.

For example, spending measures that finance short-time work schemes have a direct positive impact on employment.

So far, the short-time work schemes and State aid schemes made available by governments have covered the liquidity needs of companies and thus mitigated the massive redundancies. Indeed, recent increases in the unemployment rate are low compared to the decline in performance.

The total value of the measures included in the survey aimed at directly supporting employment through short-time work schemes is 184 billion euro, with more than 60 % of the total amount spent by France (32.5 billion euro), Spain (29.4 billion euro), the Netherlands (25.2 billion euro) and Italy (23.7 billion euro) (see Figure 2). The budgetary costs of short-time work depend on several legal characteristics of the schemes (eligibility, replacement rates of salaries, duration, etc.), the nature of the containment measures and the volume of activities that can be carried out by telework in each country.

In Romania, the official unemployment rate reached a peak of 5.5% during July-August and, although it decreased slightly, it still remained above 5% in September. In reality, however, the actual number of those without a job could be higher, given the population who are able to work but who do not meet the conditions of the 'unemployed' category and are not included in the statistics. Closely related to this indicator, consumption, through the evolution of retail sales, remains at 9 months above the level of September 2019 but slows its growth rate to below 4% after reaching a post-crisis high in July (series adjusted for the number of working days and seasonality). However, the encouraging performance of retail (except for cars and motorcycles) in the first 9 months of the year, up 1.7%, suggests a rebound after the second-quarter dive. Despite the spread of the pandemic, consumption and, in particular, construction (and demand-side segments) are expected to remain high throughout the year.

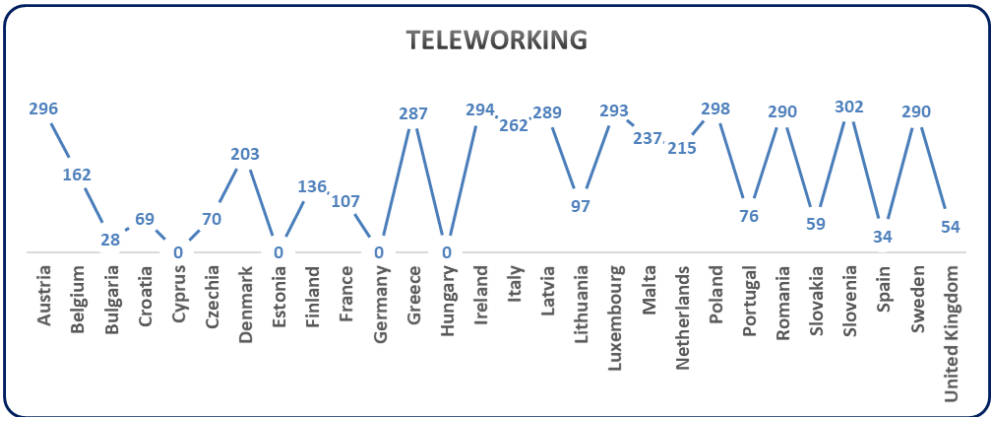

Figure 3. Teleworking in EU contries

Teleworking has proven to be a prominent feature of the pandemic. While teleworking jobs accounted for 33% of employees in 2019, less than 10% of potential teleworking (3.3% of employees) reported working from home either usually or sometimes. This has changed with the pandemic and the

associated lock-ins, during which more than a third of Europeans have started teleworking. In the first months after the outbreak of the pandemic, teleworking supported employment and working hours in some sectors, especially among workers with higher levels of education. Looking ahead, the pandemic is likely to accelerate the continued digital transformation of the euro area economy, with teleworking and the use of digital technology becoming increasingly prominent.

The use of cluster analysis, a concept introduced in the middle of the twentieth century [15], is a research theme that has aroused the interest of numerous researchers in various scientific fields (economists, sociologists, geographers, engineers, mathematicians). Thus, it is found in the profile literature, both in theoretical research, which presents the specificity of this method based on the similarity of the analyzed elements [1][7][2][12], as well as in applied research that tries to identify a series of patterns of the dynamics of some economic indicators, including unemployment, in territorial comparisons [3].

2. Literature Review

The analysis of unemployment, based on clusters, is captured in studies over the last decade for both stable periods of the world economy [4] and especially those of crisis, with a decisive impact on this socio-economic indicator [5].

Thus, k-means methods and spectral clustering are used in some studies that comparatively analyze unemployment in the administrative units of some EU states, NAFTA or ASEAN.

Babucea&Dănăcică, in a study on the comparative analysis of the unemployment rate in the Counties of Romania, applies the Hierarchical Cluster algorithm on the number of unemployed persons registered at the beginning of 2009, as the first effect of the global economic crisis. The identification of the similarity between the counties was based on a set of indicators, considered by the authors to be the most relevant: the unemployment rate registered (the ratio between the number of unemployed and the civilian active population), the total number of registered unemployed, the total number of unemployed persons compensated and unemployed and the number of unemployed in the private sector [3].

Dias, uses a modified version of the k-means-r method to analyse the grouping of euro area countries during the recent sovereign debt crisis, combining monthly data on unemployment and government bond yield rates. The analysis shows that the separation of southern Europe from the other euro area is not necessarily a good characterization of this area before the crisis, but the group of externally assisted countries plus Italy is gaining consistency as the crisis has evolved, although there is no perfect homogeneity in this group, because the problems they have faced, the kind of response requested, the speed of response to the crisis and the lasting effects were not the same for all countries [6].

Tatarczak&Boichuk identified groups of Polish regions on the basis of similarity in youth unemployment. In order to achieve this objective, a set of labour market indicators (number of employees, number of unemployed, structure by age group -15-24 years, 25-54 years, 55-64 years; education, work experience) were first selected, and later the authors compared the situation on the labour market of young people in an analysis of the Polish regions in 2005 and 2014. The methodology for data extraction was based in particular on hierarchical cluster analysis and Hellwig's methods, resulting in the association of the 16 administrative units (voivodeships) in five groups, based on similar labour market characteristics [14].

Balsa&Nunes&Barros, in a study on the 278 municipalities (concelhos) of the continental part of Portugal, concludes on the basis of the socio-demographic characteristics of the unemployed (sex, age, levels of formal education, duration of unemployment) that both methods show the importance of classifying Portuguese concelhos into two well-defined space groups, which could be the subject of distinct public policies and special measures to reduce unemployment. In this way, the economic policies/measures applicable to the labour market would be more effective, the analysis of the clusters allowing the identification of different target regions, with similar characteristics and problems. The patterns obtained based on k-means and spectral clustering analysis could be used to identify and apply customized labour market policies for each region [4].

Rudianto&Budiman, in a recent study, used the k-means algorithm to classify the level of open unemployment in administrative units in the province of Banten, Indonesia. The characteristics/attributes introduced in the k-means algorithm to analyze the level of unemployment were the number of households, the total population, the population structure by sex, age and the number of unemployed. Following the analysis, the authors identified two large categories of clusters: a category specific to administrative units (districts) in which the high unemployment rate correlates with low wages and a poor qualification of the labor force, and another category in which the reduced unemployment rate correlates with higher wages and a higher qualification of human resources [12].

Byers, in a 2013 study, applies cluster analysis and groups U.S. states with similar unemployment rate characteristics based on data from the Bureau of Labor Statistics (BLS) from January 2008 to June 2012. Based on the average unemployment rate, Byers ranks the resulting clusters and finds that the order/hierarchy of the groups at the beginning of the recession and again in 2012, remains unchanged (the cluster of California, Nevada, Michigan, Florida and Rhode Island has the highest levels of unemployment in the US in both 2008 and 2012). At the same time, the author tries to correlate the results obtained with the

economic specificity of the regions grouped in clusters. Thus, the conclusion is that clusters dependent on natural resources are not as strongly affected by periods of recession. In contrast, clusters in which industrial/productive activities predominate were particularly affected during the recession, registering a gradual recovery in 2010. At the same time, clusters with an active population employed in the service sphere, although they initially showed unemployment trends similar to those in which the population is employed in the secondary sector, subsequently recorded relatively high levels of unemployment for a longer period.

Another idea stipulated by Byers would be that cluster analysis could be useful in further research on the relationship between the structure of the region industry and unemployment rates [4].

3. Research methodology

The unemployment rate is analyzed from the perspective of the coronavirus health crisis, combined with the economic and financial crisis that is definitely developing. In presenting the data, it was taken into account both the level recorded by the indicator the number of unemployed persons aged 15-74 years, so who can carry out activity and the concrete results recorded in the evolution of the Gross Domestic Product.

Unemployment is a socio-economic situation, which refers to the way in which the working population is occupied. In this article we have made a presentation of the methodological problems involved in calculating the number of unemployed and the unemployment rate, using the basic elements of the methodology used by the National Institute of Statistics and Eurostat in the calculation of this indicator, whether it is expressed in the form of AMIGO or BIM.

In the current conditions of the coronavirus health crisis, the number of jobs will be reduced. For the moment, more than 1.3 million people are technically unemployed or unemployed.

In order to facilitate the understanding of the opinions and conclusions from the present analysis on the evolution of unemployment, we have taken over and present the main methodological aspects contained in the methodology of the National Institute of Statistics and Eurostat, used in the calculation of statistical indicators on unemployment.

The source of the data is the Household Labour Force Statistical Survey (AMIGO) which is carried out on a quarterly basis, in accordance with Regulation No 577/1998 of the Council and the European Parliament on the organisation of a selective statistical survey on labour force in the European Community.

The unemployed, according to the international definition (BIM), are people aged 15-74 years who simultaneously meet the following 3 conditions: they do not have a job, they are available to start work in the next two weeks and they have been actively looking for a job, at any time during the last four weeks.

The unemployment rate represents the share of unemployed people in the working population, calculated as their ratio.

The economically active population comprises all persons who provide the labour force available for the production of goods and services during the reference period, including the employed population and the unemployed.

The registered unemployed are the persons registered with the National Agency for Employment (ANOFM), who benefit from the provisions of the legislation on social protection of the unemployed.

Cluster analysis is a supervised classification method and aims to divide objects (data) into homogeneous groups on the basis of known characteristics. One cluster or group will contain elements that are similar according to the variables by which the division is desired, and another will contain elements different from the others.

In this work we will use the analysis of the clusters to determine the groups of European Union countries according to the unemployment rate during the economic crisis generated by the lockdown. In order to analyze the groups resulting from clustering, we chose two methods of analysis of clusters: hierarchical analysis and K-means. The software applications used to implement the classification algorithm are Python and PyCharm. The data set used includes information on the monthly values of the unemployment rate and the parameters that define the restrictions imposed during the lockdown period. Thus, the application of the classification algorithm was made according to the parameters of the restrictions imposed for each member country of the European Union.

3. Results

Hierarchical cluster analysis

The ierahic analysis of clusters is also used in statistics and creates cluster ierahii. There are two types of such hierarchical cluesters: agglomerating – when each observation is a cluster and then pairs of clusters are created, climbing the hierarchy and dividing – when all observations start from a cluster and then divide hierarchically.

The basic concept behind this type of cluster is distance. This refers to the interval between pairs of observations and is used to describe the similarity or similarity between points or clusters. The most commonly used method of estimating the distance between points is Euclidean distance.

The objective of our paper is to analyze the evolution of the unemployment rate during the lockdown period at the level of the European Union and to group the countries in clusters.

This grouping method tries to generate clusters that decrease the variation between groups and uses the Ward procedure by which the average of the variations for each cluster is calculated and then the squared Euclidean distance for each object. At each stage, taking into account this distance the clusters with the slightest increase in distance will be combined.

Hierarchical clustering is considered to be a system of uncontrolled recognition, because the classification process starts without the existence of information on the number of classes and the belonging of the forms to these classes. In this case, the classes are built according to the increase in the number of analyzed forms, the number of possible classes being determining at the end of the recognition process.

At the basis of the hierarchical analysis of the clusters is the dendogram and the 'Elbow' method, which illustrate the arrangement of the clusters according to the analysis performed.

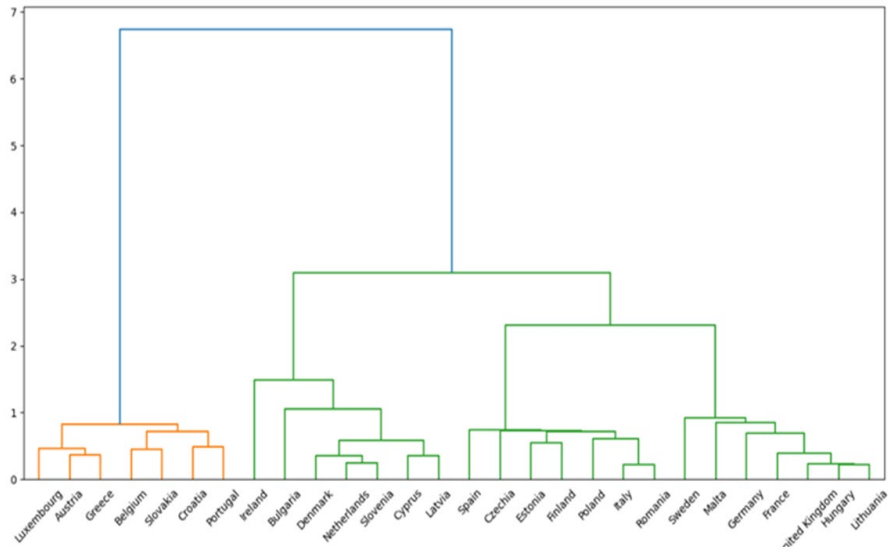

Figure 4. Dendrogram of a hierarchical cluster analysis

The tree diagram in Figure 4 shows the countries that make up the clusters, and the lines that are joined to the variables indicate a group. In the opinion of the hierarchical trees obtained, the optimal number of clusters is 4, since the smaller number of clusters will lead to large distances between countries in the middle of a bunch.

K-means cluster analysis

The second type of grouping procedures, the non-hierarchical method, classifies objects into k groups so that each group contains at least one case and each case belongs to a single group. The number of groups may not be greater than that of the cases.

In market research, the most widely used method of partitioning is the K-means approach. This algorithm starts with a random distribution of cases in k groups. The centroids of the groups are calculated and used in the analysis. Each case is assigned to the group with the nearest centroid. Each time a case is added to a group, the centroids are recalculated. This process continues until all cases are grouped together.

The Ward distance between two clusters measures the cumulative intracluster variability, induced by the agave of two clusters, at the level of the resulting cluster configuration. The merging of two clusters is aimed at achieving maximum homogeneity at the level of all clusters belonging to a given confiscation of objects on clusters. It follows that the distance Ward is the only one that takes into account the minimization of intracluster variability or, in other words, the maximization of intercluster variability, that is, the degree of homogeneity of clusters.

The second way we can group countries into clusters is by K-means cluster. This is a commonly used non-hierarchical grouping method. Through this method, all objects are distributed in k clusters using the k-means algorithm. The algorithm determines all objects that are closer to the mean (the center of the cluster). This algorithm starts with a random distribution of objects in k clusters, and then, through an iterative process, you will search for the objects that are closest to the media of the cluster to which it belongs.

As with the hierarchical analysis of clusters, the number of clusters chosen to group the countries of the European Union according to the criteria given is four. Applying this algorithm the same distribution was obtained on the cluster.

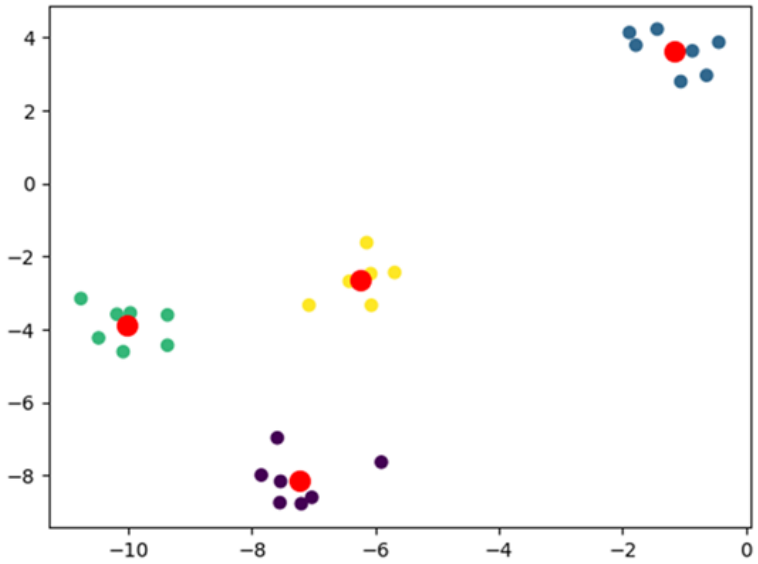

Figure 5. Clusters in scatter plots

Figure 5 shows the four clusters that have been formed with the optimum k value using the K-mean method. You can clearly see four clusters in the image above, each of the groups represented by a different color.

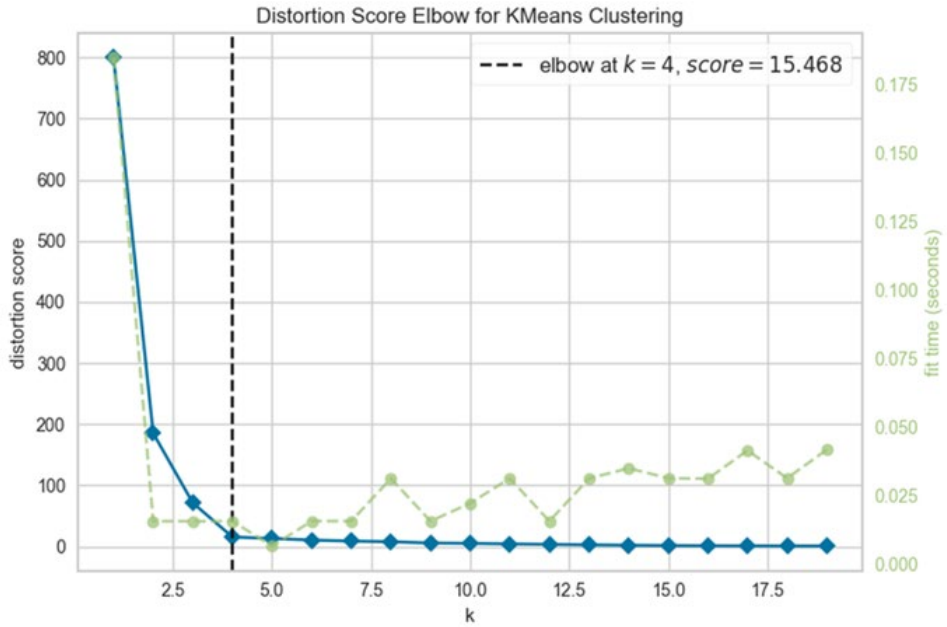

Figure 6. Elbow Method for optimal value of k in KMeans

In the analysis of the cluster (Figure 6), the "elbow" method consists of representing the variation explained as a function of the number of clusters and choosing the bend of the curve as the number of clusters to use. The K-Elbow visualizer implements the "elbow" method of selecting the optimal number of clusters for clustering with K-averages.

To determine the optimal number of clusters, we must select the value k at "elbow", that is, the point after which the distortion / inertia begins to decrease in a linear way. Thus, for the given data, we conclude that the optimal number of clusters for the data is 4.

4. Conclusion

From the study of the article based on a thorough research of the evolution of the unemployment rate during the lockdown period, a series of theoretical and practical conclusions emerge.

One conclusion is that at a time when, at national level or as it is currently wider, in the world, there are crises that lead to the onset of the economic and financial crisis, the number of unemployed and the unemployment rate will increase and this will have a negative effect on economic growth.

Another conclusion is that in the conditions of the economic and financial crisis, imposed by the emergence and accelerated development of the coronavirus health crisis, it is anticipated an increase in the

negative effect of the increase in unemployment. Therefore, economic development programmes must provide as far as possible for the possibility of increasing investment in order to create new jobs, to attract redundant staff who will become unemployed, or even to people who are already unemployed. The unemployment rate is a negative factor, as it affects the results of the creation and evolution of gross domestic product.

A final conclusion is that unemployment also appears as an element of regularisation of macroeconomic imbalances and must be taken into account, especially at the expense of investment, which is the basic element of economic growth, along with the retraining of unemployed or unemployed people.

Countries with a low unemployment rate during the lockdown period have promoted a more flexible labour market in the last 2 decades, part-time work models, temporary work contracts, work from home, teleworking, able to respond quickly and effectively to the new requirements and productivity skills and to facilitate the reconciliation of work life with the responsibilities of private life.

Governments and forward-looking organisations (such as universities) should start thinking about coping with the immediate and long-term consequences of the economic crisis created by COVID-19, especially in the area of unemployment. Creating meaningful interventions to help the new unemployed will be difficult because of the unprecedented number of people and families who are affected and because of the various contextual and personal factors that characterise this new population. Due to this diversity of contextual and personal factors, different interventions will be required for different models of individual /contextual characteristics.

References

1. *Anderberg, M. R., Cluster Analysis for Applications, New York: Academic Press, 1973;*
2. *Babucea, A. G., Analiza Cluster în statistica teritorială, Volumul Simpozionul stiintific international "Universitaria ROPET 2003", Editura Universitas, Petrosani;*
3. *Babucea, A.G., Dănăcică, E.D., 2009, Using Cluster Analysis For Studying The Proximity Of Registered Unemployment At The Level Of Counties In Romania At The Beginning Of The Economic Crisis, Analele Universităţii "Constantin Brâncuşi" din Târgu Jiu, Seria Economie, Nr. 1/2009*
4. *Byers, W.B., 2013, The Great Recession and State Unemployment Trends, Economic Development Quarterly, Vol 27, Issue 2, https://doi.org/10.1177%2F0891242413479653*
5. *[Clatworthy et all, 2005] Clatworthy, J., Buick, D., Hankins, M., Weinman, J., and Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. British Journal of Health Psychology, 10(3):329–358.*
6. *Dias, J., 2017, Unemployment and sovereign debt crisis in the Eurozone: A k-means-r analysis, Physica A: Statistical Mechanics and its Applications, Volume 482/2017, Elsevier*
7. *Furtună, T. F., Algoritmi de clasificare în statistica teritorială, Revista de statistică, nr.2/2002, p. 72-80*
8. *[Guojun G., Ma C., Wu J., 2007] Gan, Guojun, Chaoqun Ma, and Jianhong Wu, Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2007*
9. *[Hodges and Wotring, 2000] Hodges, K. and Wotring, J. (2000). Client typology based on functioning across domains using the CAFAS: Implications for service planning. Journal of Behavioral Health Services and Research, 27(3):257–270.*
10. *Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis.Wiley, New York (1990)*
11. *Tryon, R.C. (1939) Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality. Edwards Brothers, Ann Arbor.*
12. *Řezanková H., Cluster Analysis of Economic Data, Statistika: Statistics and Economy Journal, Czech Statistical Office, 94(1), 2014*
13. *Rudianto, Budiman, R., 2020, The Grouping of Regions are Based on the Unemployment Rate in the Attacking Districts with the K-Means Method, Advances in Social Science, Education and Humanities Research, volume 410, 1st International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019), Serang Banten, Indonesia*
14. *Saunders, J. (1980). Cluster analysis for market segmentation. European Journal of Marketing, 14(7):422–435.*
15. *Tatarczak, A., & Boichuk, O. (2018), The multivariate techniques in evaluation of unemployment analysis of Polish regions, Oeconomia Copernicana, 9(3), 361–380*