Understanding Global Unemployment Patterns: A 1991-2021 Regional Analysis

Tianyu Xia*

Shenzhen Foreign Languages School, Shenzhen, Guangdong, 518083, China

*hackerdsgx@gmail.com

Abstract. As one of the most important macroeconomic factors, the unemployment rate has played a significant role in individuals' decision-making and policy-making processes. This paper primarily analyzes unemployment rates and is dedicated to unveiling the connections between regions, years, and unemployment rates. Specifically, I classified all the countries/regions into three clusters according to their unemployment rates spanning from 1991 to 2021. Through Analysis of Variance (ANOVA), I found that cluster 1 represents a relatively low unemployment rate, cluster 2 represents a middle unemployment rate, and cluster 3 represents a relatively high unemployment rate. Moreover, I carried out bivariate correlation tests and identified the general trend of the average unemployment rate of each cluster. Furthermore, I classified all the countries/regions into six continents and performed ANOVA on the unemployment rates of the six continents over the years. This enabled me to discover the differences and similarities between the continents. Additionally, after dividing these 31 years into three time periods, I conducted bivariate correlation tests and determined the general trend of the average unemployment rate of each continent during specific time periods. I also constructed a cross table and discovered the distribution of the three clusters within the six continents. This research serves as guidance for individuals' decision-making processes and is helpful for people who want to study or work abroad.

Keywords: ANOVA, Bivariate Correlations, Hierarchical classification, Unemployment Rate

1 Introduction

Regarding a country's economic success, employment is one of the most crucial macroeconomic factors and an indispensable part of the economic policies of many countries [1]. The unemployment rate, by definition, is the percentage of people who are actively seeking employment but are unable to find a job within the total labor force [2]. High unemployment rates strain economic growth and limit opportunities for citizens, while low unemployment rates promote consumer spending and signal economic vitality [3]. Studying unemployment rates not only helps policymakers and economists make appropriate policies and interventions but also provides sophisticated guidance to

Z. Wang et al. (eds.), *Proceedings of the 4th International Conference on Economic Development and Business Culture (ICEDBC 2024)*, Advances in Economics, Business and Management Research 299,
https://doi.org/10.2991/978-94-6463-538-6_11

people about whether a region is suitable for them to work or study [4]. Therefore, analyzing unemployment rates has a positive effect on people's lives.

As a result, there have been a lot of related studies in this field. Soylu, Ö. B. et al. investigated unemployment in Eastern European countries for the period of 1992-2014 and found that there is a co-integration between unemployment and economic growth [1]. Moreover, Latif, E. examined the impact of permanent international immigration on the unemployment rate in Canada and discovered that the adverse impact on employment is eliminated in the long run [5]. Furthermore, Lin, S. J. used two panel data sets and employed fixed-effects models to identify the relationship between unemployment rates and suicide death rates [6]. Additionally, Proietti, T. applied different mathematical models that deal with hidden patterns in data, and checked how well they predict future outcomes. He conducted this experiment by repeatedly testing these models on data from the past two decades to see which one performs better in terms of making accurate predictions [7]. Amato, P. R., & Beattie, B. analyzed data from 50 states and the District of Columbia from 1960 to 2005 and determined the connections between unemployment rates and divorce rates [8]. Johnston, M. F., & Huimin, L. described the political procedures that influence how China's urban unemployment is defined and measured, and they also corrected some controversial data in the past [9]. Undoubtedly, those researchers have done pretty good jobs and have discovered a large amount of valuable information. However, few of these studies have investigated the differences and similarities between continents and years with respect to unemployment rates.

To fill this gap, this paper focuses on analyzing the panel data of unemployment rates all over the world for a 31-year period, from 1991-2021 [10], and attempts to find the general trend of unemployment rates in some regions during a specific period. People, especially students and employees who plan to go abroad in the future, can take the discoveries in this research paper as guidance to help them choose a prospective continent to live in. The first column of the dataset is country names, comprising 235 countries/regions. The second column is country codes, and the following 31 columns are the unemployment rates of the corresponding countries/regions during 1991-2021. Based on this panel data, I discovered a considerable amount of valuable information through ANOVA, Bivariate Correlation Tests, Cross Tables, and other analytical methods.

The rest of the paper is organized as follows: In Section 2, I conducted hierarchical classification that classified all countries into three groups according to their unemployment rates over the 1991-2021 period. After conducting the ANOVA, I discovered that these three clusters have significant differences between each other. Moreover, I built line charts that visually show the average unemployment rates of the corresponding clusters, and then I identified the trends of these average unemployment rates during a specific time. In Section 3, I conducted ANOVA with respect to the years and continents to find the connections between them. After reorganizing the dataset, I got several line charts and investigated the relationship between the years and the continents with bivariate correlations tests. In Section 4, I created a cross table of continents and clusters which was beneficial for recognizing the distribution of the three clusters within the six continents. Section 5 concludes the paper, clarifies the limitations of this study, and offers some prospects for this field.

2 Cluster Analysis

2.1 Hierarchical Clustering

First, I conducted hierarchical clustering of countries/regions based on their unemployment rates over the 1991-2021 period. Then I appended a new column to the original dataset to indicate the specific cluster to which each country/region belongs. This clustering analysis yielded seven countries/regions in cluster 3, eighteen in cluster 2, and a vast majority of two hundred and ten in cluster 1. This classification provides a fundamental framework for further examination of the disparities in global unemployment rates.

2.2 One-Way ANOVA

2.2.1 Assessment of the Effectiveness of the Clustering.
Second, I carried out an ANOVA on the unemployment rates of the clusters to assess whether the differences between them are significant. The ANOVA Table shows that the p-values between groups are all smaller than 0.05 for 1991-2021, indicating that there are significant differences among the three clusters, demonstrating the effectiveness of the clustering process in distinguishing among them.

2.2.2 Differentiating the Clusters.
One-Way ANOVA employs 31 mean plots to depict the average unemployment rates of each cluster over the 31-year period. By analyzing these mean plots, I have preliminarily identified that cluster 1 represents low unemployment rates, cluster 2 stands for medium unemployment rates, and cluster 3 signifies high unemployment rates. This identification provides a fundamental basis for conducting in-depth analyses within each cluster.

2.2.3 Preliminary Insights of the Variances of the Average Unemployment Rates Among the Clusters.
As demonstrated in Figure 1 and 2, the average unemployment rates exhibit variations over the 31-year span: Cluster 1 shows a slight increase, cluster 2 experiences an increment, while cluster 3 shifts from above 30 percent to below 30 percent. However, while figures like Figure 1 and 2 provide valuable insights into the disparities among clusters between two specific years, they are unable to illustrate the year-to-year changes directly.

2.3 Line Charts Analysis

2.3.1 Dataset Transformation.
To delve into the year-to-year fluctuations in average unemployment rates across the three clusters and examine their differences more comprehensively, I conducted calculations of yearly averages from 1991 to 2021 and created a new dataset. In this dataset,

the first column shows the years from 1991 to 2021, and the subsequent columns comprise the average unemployment rates for each cluster throughout this time. Next, I built three line-charts that are helpful for the in-depth examination of these year-to-year variations and have a beneficial impact on distinguishing these three clusters.

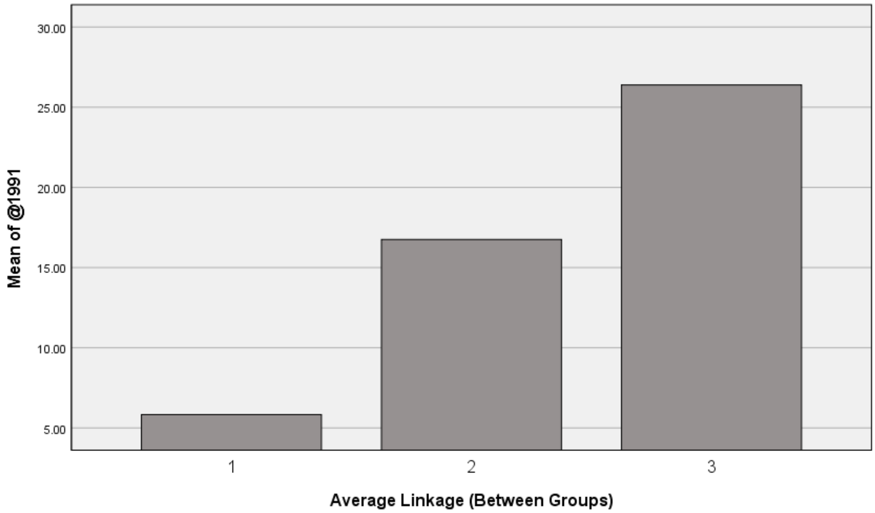

Fig. 1. The average unemployment rates of the three clusters in 1991.

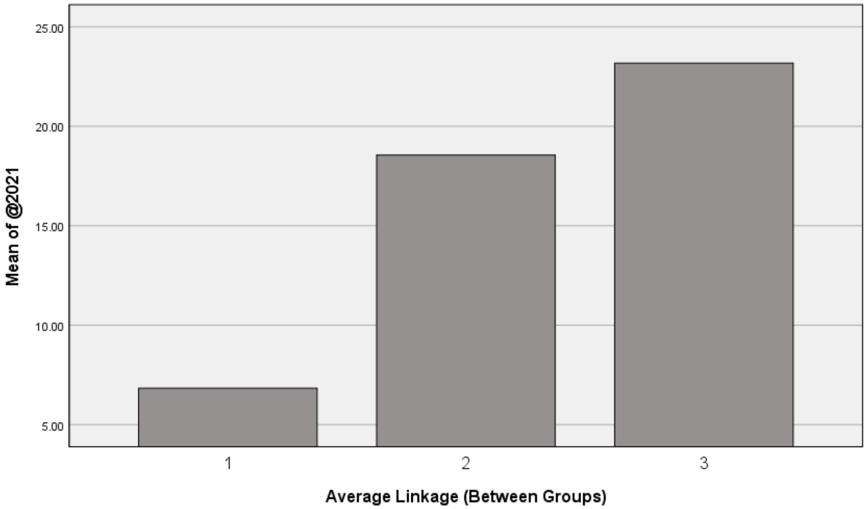

Fig. 2. The average unemployment rates of the three clusters in 2021.

2.3.2 Analysis of Changes in Average Unemployment Rates of Clusters Over Years.

As shown in Figures 3, 4, and 5, the average unemployment rates within the three clusters fluctuate over the years. Initially, an upward trend is observed in these rates, followed by a significant decline that reaches a relatively low level in 2008. After this low point, these average unemployment rates demonstrate a gradual recovery, eventually returning to reasonable levels after a few years. Notably, in 2019, all three clusters experience a convergence toward relatively lower levels of unemployment rates once more. Finally, the average unemployment rates recover rapidly. The analysis of temporal changes in the average unemployment rates within the three clusters provides a basis for further analysis. It is worth acknowledging that the synchronized occurrence of a low point in the unemployment rates across all three clusters may be indicative of significant global events. A comprehensive understanding of these events and their impact on employment is of great importance for policymakers and economists to consider.

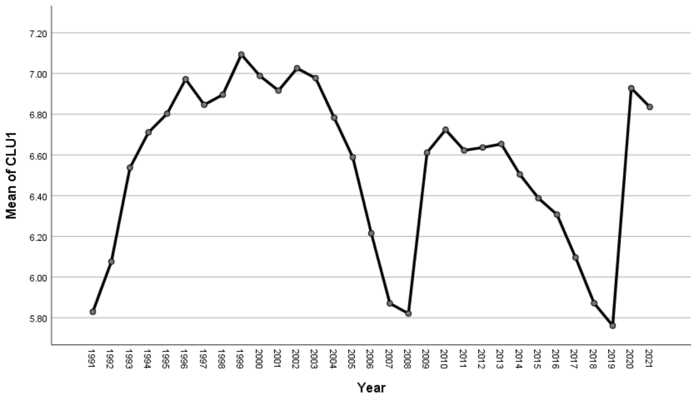

Fig. 3. The line chart of the average unemployment rates of cluster 1 from 1991 to 2021.

Fig. 4. The line chart of the average unemployment rates of cluster 2 from 1991 to 2021.

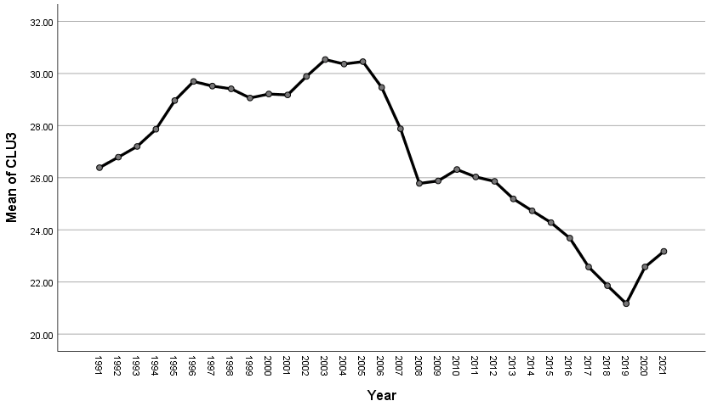

Fig. 5. The line chart of the average unemployment rates of cluster 3 from 1991 to 2021.

2.3.3 Further Examination of Cluster Classification Validity.
To thoroughly assess the validity of the clustering approach and achieve a clearer distinction among the three clusters, I first calculated the mean values of these clusters over the 31-year span as shown in Table 1. Next, I conducted a bivariate correlation analysis to investigate the relationships between the years and the average unemployment rates within each cluster.

The results of this analysis indicate that the p-values between the average unemployment rates of cluster 1 and 2 throughout the years from 1991 to 2021 are greater than 0.05. This suggests that there is no significant correlation between the years and the average unemployment rates for these two clusters. In contrast, for cluster 3, the p-value between the average unemployment rates and the years is much smaller than 0.05. This finding indicates a highly significant correlation between the average unemployment rate of cluster 3 and the passage of time, with a Pearson Correlation coefficient of -0.755.

Furthermore, the mean unemployment rates observed over the past 31 years amount to 6.545, 17.745, and 26.805 for clusters 1, 2, and 3, respectively. The standard deviation values for the unemployment rates within these clusters are 0.411, 0.694, and 2.796, respectively. The substantial disparity in means between cluster 2 and 3 is significant enough that even with the relatively higher standard deviation and Pearson correlation coefficient observed in cluster 3, it is highly unlikely that cluster 2 would exhibit a higher average unemployment rate in any given year compared to cluster 3. The fact that these three clusters have a very clear distinction between high and low levels over the past 31 years is also verified in the 31 mean plots mentioned in 2.2.2. The values of 6.545, 17.745, and 26.805 respectively signify that cluster 1 corresponds to low unemployment rates, cluster 2 represents moderate unemployment rates, and cluster 3 signifies high unemployment rates. This analysis successfully reveals meaningful distinctions in the behavior of unemployment rates among the three identified clusters over the 31-year period.

3 Continental Analysis

3.1 Dataset Transformation

To commence the continental analysis, I added a column to the dataset, designating the respective continents represented by each country. In this newly created column, I employed a numeric coding system: 1 signifies Asia, 2 designates Africa, 3 represents Europe, 4 denotes Oceania, 5 indicates North America, and 6 stands for South America. This numeric coding system simplifies the subsequent analyses.

3.2 One-Way ANOVA

Next, I conducted ANOVA to examine the relationship between the continents and the progression of years. The results indicate that there are significant differences among the continents. However, the years 2002, 2008, 2017, and 2018 stand out as exceptions, where the differences are not significant, which may indicate there were important events that happened globally. Further investigations are needed to explore this disparity.

3.3 Line Charts Analysis

3.3.1 Mean Plots for the Initial and Final Years.
As shown in Figure 6 and Figure 7, the examination of the initial and final years indicates a discernible variation in average unemployment rates spanning the 31-year period. The mean unemployment rate of Africa, for instance, exhibits a notable increase, surpassing 10 percent, while that for North America declines to a level below 10 percent.

3.3.2 Dataset Transformation.
Similarly, Figure 6 and Figure 7 do not illustrate the year-to-year changes in unemployment rates for all six continents visually. Therefore, I calculated the average unemployment rates for each continent and restructured the dataset. This approach allows me to create more incisive line charts, providing visual representations of how the unemployment rates evolved over time. After transforming the dataset, the first column of the transformed dataset represents the 31 years, and the following six columns represent the six continents coded from 1 to 6, corresponding to Asia through South America. Given the observed insignificance in differences between continents during the years 2002, 2008, 2017, and 2018, I split the past 31 years into 1991-2002, 2003-2008, 2009-2021 these specific time spans for conducting bivariate tests pertaining to the six continents.

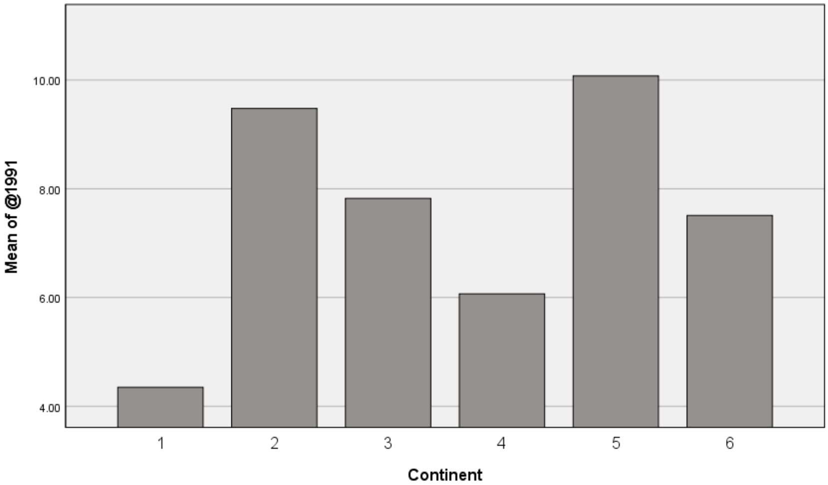

Fig. 6. The average unemployment rates of the six continents in 1991.

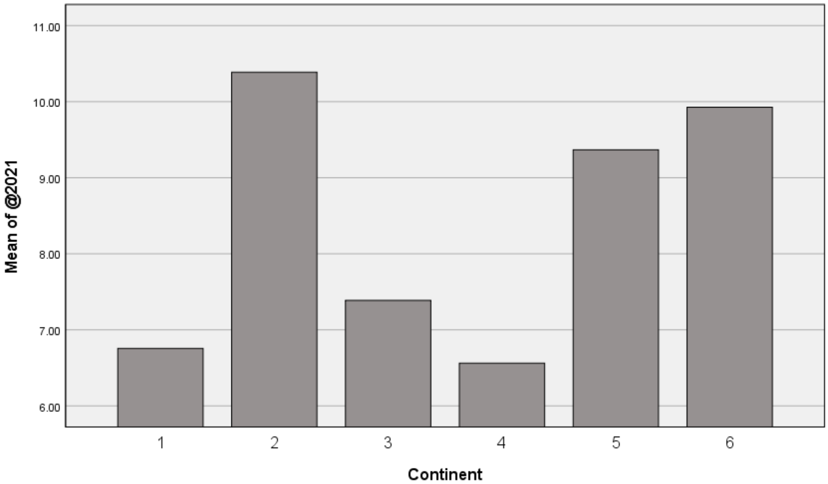

Fig. 7. The average unemployment rates of the six continents in 2021.

3.3.3 Further Analysis of the Continents.

3.3.3.1 Analysis from 1991 to 2002.

Examining the period from 1991 to 2002, Asia exhibits a remarkably robust positive correlation, with a Pearson Correlation Coefficient of 0.985 (see Figure 8). Africa, during this time frame, demonstrates a strong positive correlation with a coefficient of 0.596 (see Figure 9). Oceania also displays a highly significant positive correlation, with a coefficient of 0.782 (see Figure 11). In contrast, North America reveals a very strong negative correlation, boasting a coefficient of -0.935 (see Figure 12). Lastly, South America exhibits a strong positive correlation, with a coefficient of 0.936 (see Figure 13). During this time, the trend of average unemployment rates of the six continents is increasing except for North America, which showcases a very strong negative correlation, illustrates the recovery of the economics in North America.

3.3.3.2 Analysis from 2003 to 2008.
During the period spanning from 2003 to 2008, a comprehensive examination of unemployment rate trends across continents reveals noteworthy correlations. Asia exhibited a notably strong negative correlation, as indicated by a Pearson Correlation Coefficient of -0.926 (see Figure 8). Similarly, Africa demonstrated an exceptionally significant negative correlation during this timeframe, with a coefficient of -0.986 (see Figure 9). Europe also displayed a highly substantial negative correlation, with a Pearson Correlation Coefficient of -0.948 (see Figure 10). Oceania exhibited a significant negative correlation, with a coefficient of -0.826 (see Figure 11). North America and South America likewise showcased remarkably strong negative correlations, with Pearson Correlation Coefficients of -0.960 and -0.982, respectively (see Figure 12 & 13). This analysis highlights the prevalence of negative correlations during this specific period and underscores the importance of further investigation into the factors influencing these synchronous trends in unemployment rates across continents.

A comprehensive global analysis of the average unemployment rates across six major continents during the period from 2003 to 2008 has yielded statistically significant findings. All six continents exhibited highly significant negative correlations with p-values below 0.05, indicating a pronounced negative relationship between average unemployment rates and this specific time frame. Notably, the global average unemployment rate displayed an astonishing Pearson Correlation Coefficient of -0.991 during this period. This pattern underscores a synchronized decline in unemployment rates worldwide, coinciding with an improvement in economic conditions. These findings provide robust evidence for the subsequent exploration of the factors contributing to this global reduction in unemployment rates.

3.3.3.3 Analysis from 2009 to 2021.
During the period from 2009 to 2021, a detailed examination of unemployment rates across continents reveals distinctive patterns. Notably, Africa displayed a p-value of much less than 0.05 for its variation of the average unemployment rates during this timeframe, signifying a significant positive correlation with a Pearson Correlation Coefficient of 0.663. Conversely, Europe exhibited a notably significant negative correlation, marked by a coefficient of -0.847 throughout this period. Furthermore, South America's average unemployment rate also demonstrated a p-value below 0.05, showcasing a robust positive correlation, as indicated by a Pearson Correlation Coefficient of 0.707 concerning the variations within this timespan. Unlike the previous period from 2003 to 2008, these findings illuminate the absence of a common global trend and therefore should be further examined with each individual continent to discover some valuable information.

3.3.3.4 Analysis from 1991 to 2021

Spanning from the period 1991-2021, Asia displayed a p-value of much less than 0.05 for its variation of the average unemployment rates during this 31-year period, indicating a highly significant positive correlation with a Pearson Correlation Coefficient of 0.551 (see Figure 8). Africa exhibited a significant negative correlation, marked by a coefficient of -0.383 throughout this period (see Figure 9). Furthermore, Europe's average unemployment rate also demonstrated a p-value below 0.05 and showcased a notably negative correlation, as indicated by a Pearson Correlation Coefficient of -0.435 (see Figure 10). And for North America, its average unemployment rate has a significantly strong negative correlation with respect to the years, showcasing with a Pearson Correlation Coefficient of -0.556 (see Figure 12). The average unemployment rate of the whole world indicating a strong negative correlation regarding this 31-year time span. The Pearson Correlation Coefficient is -0.367 (see Figure 14).

Fig. 8. The line chart of the average unemployment rates in Asia from 1991 to 2021.

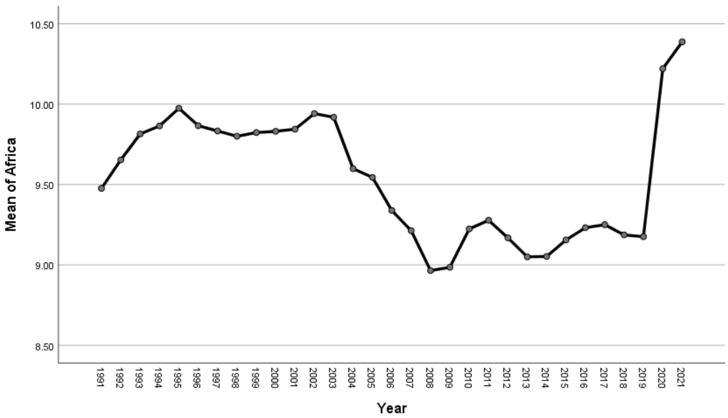

Fig. 9. The line chart of the average unemployment rates in Africa from 1991 to 2021.

Fig. 10. The line chart of the average unemployment rates in Europe from 1991 to 2021.

Fig. 11. The line chart of the average unemployment rates in Oceania from 1991 to 2021.

Fig. 12. The line chart of the average unemployment rates in North America from 1991 to 2021.

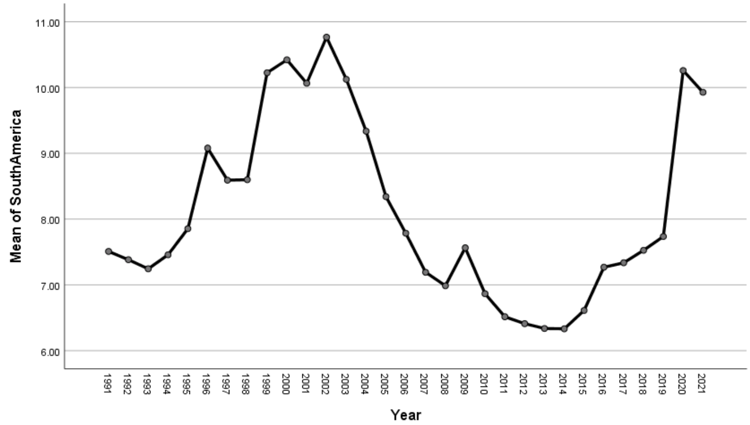

Fig. 13. The line chart of the average unemployment rates in South America from 1991 to 2021.

Fig. 14. The line chart of the average unemployment rates in the world from 1991 to 2021.

4 Cross Table Analysis

Finally, I conducted a cross table analysis to explore the distribution of the three clusters among the six continents. The results, as shown in Table 1 below, provide valuable insights into the distribution of countries across these categories. Notably, all countries in South America exclusively belong to cluster 1, while most countries in Asia, Oceania, and North America fall into cluster 1, with a few categorized under cluster 2. Cluster 3, on the other hand, is primarily represented in Africa and Europe. This analysis sheds light on the distinct clustering patterns within continents, offering a clearer understanding of how different regions are categorized based on their unemployment rates. The Chi-square test shows that the p-value of the six continents and the clusters is greater than 0.05, indicating that there is no specific distribution between the clusters and the continents.

Table 1. Cross Table of the clusters and the continents.

		1	2	3	Total
Continent	1	49	2	0	51
	2	45	10	4	59
	3	35	3	3	41
	4	10	1	0	11
	5	19	2	0	21
	6	13	0	0	13
Total		171	18	7	196

5 Conclusion

All in all, this analysis of global unemployment rates spanning the 31 years from 1991 to 2021 has explored the interconnections among various geographical regions, different time periods, and the fluctuations in unemployment rates through different analysis methods. In the following paragraphs, I'll provide a brief recap, identify some shortcomings of this research paper, and offer prospects for in-depth research in this field.

First, the hierarchical clustering categorized countries/regions into three distinct clusters based on their unemployment rates. The ANOVA confirmed significant differences among these clusters, further substantiating the effectiveness of the clustering process. Through histograms of the mean unemployment rates of these clusters, I identified that cluster 1, 2 and 3 represent relatively low, medium, and high unemployment rates, respectively. Moreover, through bivariate correlation tests, I discerned the trend of the average unemployment rates as time elapses. The results illustrate the varying degrees of correlation between clusters during specific time intervals and provide.

Secondly, the continental analysis unveiled intriguing insights into the relationships between continents, years, and unemployment rates. One-Way ANOVA revealed significant differences among continents, while 2002, 2008, 2017, and 2018 emerge as notable exceptions. Therefore, I divided the 31 years into three specific time spans according to the ANOVA and conducted bivariate correlation analyses in order to get more valuable information. Between 1991 and 2002, distinct regional correlations were evident, with Asia, Africa, and Oceania showing positive relationships, while North America exhibited a significant negative correlation. From 2003 to 2008, all continents displayed synchronized negative correlations, indicative of a global decline in unemployment rates during those periods. However, the years from 2009 to 2021 saw diverse continental correlations, suggesting unique regional influences on unemployment rates.

Furthermore, the Cross Table analysis in this research paper has revealed distinctive clustering patterns within continents based on countries' unemployment rates. Notably, all countries in South America exclusively belong to cluster 1, while Asia, Oceania, and North America mainly align with cluster 1, with some countries represented in cluster 2. Conversely, cluster 3 prevails in Africa and Europe.

These insightful discoveries can be valuable for policymakers and researchers, guiding their understanding of and response to regional unemployment challenges. However, it is essential to acknowledge the limitations of this study. First and foremost, the clustering approach employed in this analysis was a broad categorization into three clusters, and the division of countries was largely based on their continental affiliations. This level of aggregation may obscure important nuances within each continent and country-specific variations. Future research should aim to capture more granular distinctions among nations. Furthermore, this study primarily serves as an initial exploration of the relationships between unemployment rates regarding continents, clusters, and time periods. It has not delved deeply into the underlying factors driving these variations, such as political factors, natural disasters, and so on. Additionally, the correlation analysis conducted in this report provides a rudimentary examination of how average unemployment rates evolved across continents during specific time intervals; however, this study does not possess the capacity to predict future events and therefore cannot precisely forecast how unemployment rates will evolve.

To address the limitations highlighted in this research paper, future scholars could adopt a more granular approach by concentrating on individual countries. This approach would involve a comprehensive investigation into the impact of policy-making decisions and the influence of natural factors on the fluctuations observed in unemployment rates within each nation. Additionally, the development of a more precise and sophisticated model for analyzing this dataset holds substantial promise. Such a model could offer predictive capabilities that illuminate the cyclical patterns of economic shifts, providing policymakers, businesses, and individuals with a more accurate guide to better-informed decisions and strategies for addressing unemployment challenges and fostering economic stability.

References

1. Soylu, Ö. B., Çakmak, İ., & Okur, F. (2018) Economic growth and unemployment issue: Panel data analysis in Eastern European Countries. Journal of International Studies, 11: 93-107.
2. Mankiw, N. G., & Rabasco, E. (2007) Principios de economía. Ediciones Paraninfo, SA, Madrid.
3. Layard, R., Nickell, S. J., & Jackman, R. (2005) Unemployment: macroeconomic performance and the labour market. Oxford University Press, USA.
4. Borjas, G. J., & Van Ours, J. C. (2010) Labor economics. McGraw-Hill/Irwin, Boston.
5. Latif, E. (2015) The relationship between immigration and unemployment: Panel data evidence from Canada. Economic Modelling, 50: 162-167.
6. Lin, S. J. (2006) Unemployment and suicide: panel data analyses. The Social Science Journal, 43: 727-732.
7. Proietti, T. (2003) Forecasting the US unemployment rate. Computational Statistics & Data Analysis, 42: 451-476.
8. Amato, P. R., & Beattie, B. (2011) Does the unemployment rate affect the divorce rate? An analysis of state data 1960–2005. Social Science Research, 40: 705-715.

9. Johnston, M. F., & Huimin, L. (2002) Estimating China's Urban Unemployment Rate: background, mechanics and an alternative. Journal of Contemporary China, 11(31), 189-207.
10. Anjali Pant. (2022) Country's unemployment rate from past 31 years. https://www.kaggle.com/datasets/pantanjali/unemployment-dataset/.