Cumhuriyet Sci. J., 43(1) (2022) 146-164
DOI: https://doi.org/10.17776/csj.969445

Cumhuriyet Science Journal

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

# Investigating the COVID19 Characteristics of the Countries Based on Time Series Clustering

***Muhammet Oguzhan Yalçın [1,a,*], Nevin Güler Dincer [1,b], Öznur İşçi Güneri [1,c]***

[1]*Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Muğla, Turkey*
**Corresponding author*

**Research Article**

**History**
*Received: 10/07/2021*
*Accepted: 05/03/2022*



**ABSTRACT**

The objective of this study is to reveal the COVID19 characteristics of the countries by using time series clustering. Up to now, various studies have been conducted for similar objectives. But, it has been observed that these studies belong to early time of pandemic and are involved limited number of countries. To analyze the characteristic of COVID19 more, this study has considered 111 countries and time period between the 4th of April 2020 and the 1st of January 2021. Fuzzy K-Medoid (FKM) is preferred as clustering method due to its three abilities: i) FKM enables to determine the similarities and differences between the countries in more detail by utilizing the membership degrees, ii) In FKM, cluster centers are selected among from objects in the data set. Thus, it has the ability of detecting the countries which represent the behavior of all countries, iii) FKM is a robust method against to outliers. Thanks to this ability, FKM prevents that the countries exhibiting abnormal behavior negatively affect to the clustering results. At the results of the analyses, it is observed that 111 countries have three different behaviors in terms of confirmed cases and five different behaviors in terms of deaths.

***Keywords:*** Fuzzy K-medois, Cluster validity, Time series clustering, COVID19

[a] moguzhanyalcin@mu.edu.tr https://orcid.org/0000-0003-4017-5588
[b] nguler@mu.edu.tr https://orcid.org/0000-0003-0361-1803
[c] oznur.isci@mu.edu.tr https://orcid.org/0000-0003-3677-7121

## Introduction

Coronavirus which emerged in China's Hubei province in December 2019 and spread to the all over the world in a short time has been declared as pandemic by World Health Organization (WHO) on the March 11, 2020. As of 5 March 2021, 115,598,160 confirmed COVID19 cases and 2,569,011 deaths have been reported in all of the world. Therefore, the countries have developed various strategies to fight COVID19 such as lockdown, obligation of mask, closing down of cafés and restaurants, restricting international air traffic etc. But, these kinds of restrictions affect the countries quite negatively in terms of economy. Determining correct strategies having the least impact on the economy depends on a good understanding of the COVID19 behavior of countries. In this study, some statistical properties of COVID19 are investigated for each country separately and the countries having similar COVID19 behavior are determined by using time series clustering. Thus, it is hoped to detect the countries that need to take more serious precautions individually and the countries that can develop common strategies.

So far, several studies have been conducted to determine the countries having similar behavior of COVID19. Some of these studies can be summarized as follows: Imtyaz et al.[1] have clustered thirty countries in terms of percentage of their elderly population and, COVID19 mortality rate. They have applied k-means clustering algorithm to the data set of the thirty countries for the time period between 22 January 2020 and 01 June 2020. At the result of the analyses, they have found that mortality rates in countries in Western Europe are high while the mortality rates in countries in South Asia and Middle East are low. Zarikas et al.[2] have used the hierarchical clustering methods to divide the countries into the homogenous groups with respect to active cases, active cases per population and active cases per population and per area for the time period between 22 January 2020 and 4 April 2020. Mahmoudi et al.[3] have clustered the high-risk countries including United States America, Spain, Italy, Germany, United Kingdom, France and Iran with respect to the number of confirmed cases, the number of death cases, cumulative number of confirmed cases and cumulative number of deaths using fuzzy clustering. The data sets they used involve the time period from 22 February 2020 up to 18 April 2020. They also have investigated the correlation between the population size and spread of COVID19. Alvarez et al. [4] have used non-parametric techniques based on correlation distance and Minimal Spanning tree in order to cluster 191 countries in terms of COVID19 dynamics. Hutagalung et al.[5] have focused on the grouping of the 11 countries located in Southeast Asia in terms of the number of confirmed cases and the number of deaths observed on the date of April 2020. They have used the k-means clustering algorithm. Virgantari and Faridhan[6] have applied k-means clustering algorithm to the data set covering COVID19 cases in 34 provinces of Indonesia. Rojas et al.[7] have used the hierarchical clustering algorithm based on dynamic time warping distance measure to determine behavioral relationships between different states of the US with respect to COVID19.

Azarafza et al.[8] have investigated the spatiotemporal distribution and spread pattern of COVID19 in Iran. For this objective, they have utilized k-means clustering algorithm. Crnogorac et al.[9] have carried out a study based on clustering the cumulative relative number of the European countries and territories. They have used three clustering algorithms, including K-Means, agglomerative and BIRCH. Sadeghi et al. [10] used hierarchical clustering algorithm to rank and score 180 countries in terms of COVID19 cases and fatality in 2020. Putra ve Kadyanan[11] have clustered 9 provinces in Bali by using K-Means clustering algorithm. They have used four clustering variables, consisting of number of cases, dead rate, the number of recovered and the number of isolated people. Utomo[12] has applied k-means and k-medoids clustering algorithms to data set, consisting of confirmed and death cases for grouping 34 provinces in Indınesia. Abdullah et al.[13] also clustered provinces in Indonesia of the risk of the COVID19. They used the K-Means clustering algorithm and three clustering variable, including confirmed, death and recovered cases.

The most of these studies are based on the classical clustering approach. But, the clustering approaches based on classical logic have some disadvantages: i) Classical clustering approaches force that each object to be clustered in such a way as to belong to only one cluster. In case an object is approximately equidistant from more than one cluster, the object is assigned to the cluster that is the closest one. Thus, the fact that the object has also the characteristics of other clusters with certain degree is ignored. ii) In classical clustering, there is no difference between the objects within the same cluster. Whereas some objects carry the characteristics of the cluster more, some less. Besides, in most of the studies summarized above, short periods at the beginning of the pandemic and limited number of countries have been considered. In the current study, FKM clustering algorithm is applied to cumulative number of confirmed cases (CCOP) and deaths (DOP) per one million persons of 111 countries. The time period studied covers the period between 1 April 2020 and 22 January 2021. The contributions of this study can be sorted as follows:

•This study uses FKM clustering algorithm based on fuzzy logic. The fuzzy clustering approaches allow assigning a country to more than one cluster with different membership degrees. Thus, it is possible to identify the countries having the characteristics of more than one cluster and the differences between the countries within the same cluster.

•FKM clustering algorithm selects the cluster centers among from the countries. This ability of FKM allows determining a representative country for each cluster separately in order to form an opinion about the COVID19 behavior of the other countries which are assigned to the same cluster with high membership degrees.

•FKM clustering algorithm is robust to outliers. It's this ability considerably decreases the negative effect on the clustering of the countries having abnormal COVID19 pattern.

•Five cluster validity indexes have been considered simultaneously to reveal the number of different COVID19 behavior.

This study is organized as follows. Section 2 gives brief information about the data set and methods used. Section 3 includes the experimental results and in the last section, the study is concluded.

## Materials and Methods

### *Data Set*

The raw data sets used in this study are downloaded from the web site of https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset. Data sets consist of the cumulative number of confirmed cases and cumulative number of deaths. 111 countries and the time period between the 4th of April 2020 and the 19th of January 2021 are considered. Before the clustering process, the raw data are standardized as follows:

$$CCOP_{ij} = \frac{Cumulative\ number\ of\ Confirmed\ Cases_{ij}}{Population\ Size_i} * 1000000 \quad \mathrm{i} = 1,2,\dots,\mathrm{N} \quad \mathrm{j} = 1,2,\dots,\mathrm{n} \tag{1}$$

$$DOP_{ij} = \frac{Cumulative\ number\ of\ deaths_{ij}}{Population\ Size_i} * 1000000 \quad \mathrm{i} = 1,2,\dots,\mathrm{N} \quad \mathrm{j} = 1,2,\dots,\mathrm{n} \tag{2}$$

Where n (294) is the length of time series, N (111) is the number of countries.

### *Time Series Clustering*

Clustering analysis is a data mining technique used for dividing the objects into homogenous groups according to their characteristic properties. According to this method, while objects with the same properties are in the same group, objects with a large difference from each other are placed in different groups. While increasing the difference between groups to the maximum is aimed by the method, the difference between groups is minimum.

Clustering methods are generally divided into two groups as hierarchical and partitioning methods. Hierarchical clustering techniques are the process of combining clusters gradually. In order to perform hierarchical cluster analysis, researchers have to decide how to define similarity or distance and how to merge or separate clusters[14][15].

Partitioning clustering algorithms take c input parameters and divide N objects into c clusters. These techniques perform operations that find single-level clusters instead of working on a nested clustering structure like a dendrogram[16]. All techniques are based on the cluster center representing the cluster. To improve cluster quality, the algorithm is run multiple times with different starting points and the best configuration from total runs is used as output clustering. Partitioning clustering algorithms are widely used due to their easy applicability and efficient results. Partitioning methods are divided into classical clustering and fuzzy clustering [17]. In classical clustering, each object of the data set is assigned to one and only one cluster. In fuzzy clustering, it allows objects to belong to two or more clusters. According to the fuzzy logic principle, each object belongs to each of the clusters with a membership value varying between [0,1].

Time series clustering (TSC) is a special type of clustering in which the objects to be clustered correspond to the time series. Time series can be defined as a set consisting of the observations measured at the successive time points. TSC methods can be collected under three main headings I) distance-based, ii) feature-based and iii) model-based[18]. Distance-based TSC methods directly works with time series themselves without any transforming or preprocessing on them. Therefore, this kind of TSC methods provide the best clustering performance since they do not lead to information lost.

In feature-based TSC methods, time series are converted into feature space with lower dimension which represents its behavior. Clustering algorithm is applied to the features extracted. Lastly, in the model-based TSC methods, a model is predicted for each time series by using statistical or other modeling techniques. To determine similar time series, model parameters are used as a clustering variable.

Distance-based time series clustering is preferred in this study due to its advantage mentioned above. In these methods, data set is organized as follows:

$$\boldsymbol{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N1} & y_{N2} & \cdots & Y_{Nn} \end{bmatrix} \tag{3}$$

Where N is the number of time series, n is the length of time series. In matrix Y, each row corresponds to a time series. In this study, each row is the time series of a country consisting of CCOP or DOP values.

### *Fuzzy Clustering and Fuzzy K-Medoids Algorithm*

Fuzzy clustering algorithms have the ability of assigning an time series to more than one clusters via membership degrees. In fuzzy clustering, the membership degrees must satisfy following conditions:

$$\begin{aligned} & \text{v}0 \le u_{ij} \le 1 \quad \text{i=1,2,…,N, j=1,2,…c} && \text{(a)} \\ & 0 < \textstyle\sum_{i=1}^{N} u_{ij} < N \quad \text{j=1,2,..,c(} && \text{(b)} \\ & \textstyle\sum_{j=1}^{c} u_{ij} = 1 \quad \text{i=1,2,..,N} && \text{(c)} \end{aligned} \tag{4}$$

Where $u_{ij}$ is the membership degree of i^th^ time series to j^th^ cluster, c is the number of clusters and N is the number of time series. 4(a) indicates that membership degrees are between 0 and 1, 4(b) states that the sum of membership degrees of the time series in j^th^ cluster must be between 0 and N. Lastly, 4(c) states that sum of membership degrees of i^th^ time series to all clusters must be equal to one.

Fuzzy clustering methods are based on minimizing the following objective function:

$$J_{F=} \sum_{j=1}^{c} \sum_{i=1}^{N} u_{ij}^{m} \quad d^2(y_i, v_j) \tag{5}$$

m: fuzziness index, $v_j$ : j^th^ cluster center, $d^2(y_i, v_j)$ : distance between j^th^ cluster and i^th^ time series.

When the given objective function is tried to be minimized, in other words, when the derivatives are taken for $u_{ij}^{m}$ and set to 0, the following update equation is obtained.

$$u_{ij} = \left[\sum_{s=1}^{c} \left(\frac{d^2(y_i, v_j)}{d^2(y_i, v_s)}\right)^{\frac{1}{m-1}}\right]^{-1} \quad j = 1,2,\ldots,c, \ i = 1,2,\ldots,N \tag{6}$$

Fuzzy clustering algorithms differ according to form of the cluster centers and distance measure used. Fuzzy C-Means (FCM)[19], Gustafson-Kessel (GK)[20], Fuzzy C-Regression Model (FCRM) [21] and Fuzzy K-Medoids (FKM)[22] are most popular fuzzy clustering algorithms. The form of cluster centers and the distance measure used are given in Table 1.

Table 1. Properties of Widely Used Fuzzy Clustering Algorithms

| Clustering algorithm | Distance Measure | Cluster Centers |
|---|---|---|
| FCM | $d^2(y_i, v_j) = \sqrt{\sum_{k=1}^{n} (y_{ik} - v_{jk})^2}$ | $v_j = \frac{\Sigma_{i=1}^{N} u_{ij}^{m} y_i}{\Sigma_{i=1}^{N} u_{ij}^{m}}$ |
| FCRM | $d^2(y_i, v_j) = \sqrt{(y_{in} - v_{ij})^2}$ | $v_{ij} = \beta_{0j} + \beta_{1j} y_{i1} + \cdots + \beta_{j(n-1)} y_{i(n-1)}$ |
| FKM | $d^2(y_i, v_j) = \sqrt{\sum_{k=1}^{n} (y_{ik} - v_{jk})^2}$ | $q = \underset{1 \le k \le N}{argmin} \sum_{j=1}^{N} u_{ij}^{m} \, d^2(y_k, y_j) \quad v_j = y_q$ |
| GK | $d^2(y_i, v_j) = \sqrt{(y_i - v_j)^T \Sigma_j^{-1} (y_i - v_j)}$ | $v_j = \frac{\Sigma_{i=1}^{N} u_{ij}^{m} y_i}{\Sigma_{i=1}^{n} u_{ij}^{m}}$ |

According to Table 1, FCM, FCRM and FKM use the Euclidian distance measure while GK use the Mahalanobis distance measure. In Mahalanobis distance, $\Sigma_j$ is the variance-covariance matrix of j$^{th}$ cluster. The cluster centers in FCM and GK correspond to the arithmetic means weighted by membership degrees. Thus, FCM and GK algorithms are negatively affected from the outliers because these points pull cluster centers towards to themselves. In FCRM, the cluster center is a hyperplane. This algorithm is generally used in the modeling the data set generating by more than one stochastic process.

Lastly in FKM, the cluster centers are called as medoid. In fact, the medoids correspond to time series in the data set that minimizes its distance from all time series in the datasets depending on the cluster membership[23]. The most important property of the FKM algorithm is to more robust to the outliers when comparing with FCM, GK and FCRM. Besides, the correspondence of cluster centers to time series in the data set in FKM allows to select a representative time series for the time series within the same cluster. In this study, FKM algorithm is used to cluster CCOP or DOP values due to these properties of it. The pseudo code of the FKM is given in Table 2.

Table 2. Pseudo code of FKM[22]

Fix the number of clusters c; Randomly select to initial values of the medoids,
$\boldsymbol{V = \{v_1, v_2, \dots, v_c\}\ from\ Y\ (defined\ in\ Eq.3)}$
Iter =1;
Repeat
    for i=1:N
        for j=1:c
            calculate $\boldsymbol{u_{ij}}$ by using Eq. (6)
        end;
    end;
    $\boldsymbol{V^{iter} = V}$
    for j=1:c
        $\boldsymbol{q = \underset{1\le k\le N}{\operatorname{argmin}} \sum_{j=1}^{N} u_{ij}^{m}\, d^2(y_k, y_j)}$
        $\boldsymbol{v_j = y_q}$
    end;
    iter = iter+1;
    $\boldsymbol{V^{iter} = V}$
Until $\boldsymbol{V^{iter} = V^{iter-1}}$

The most important problem in partitioning clustering methods is to determine the number of clusters. Many algorithms have been proposed to determine the number of clusters. Next subsection gives the cluster validity indexes used in this study.

### *Cluster Validity Indexes*

Cluster validity are techniques used to find the optimal number of clusters without any prior knowledge.

#### Fuzzy silhouette index (FS)

The Silhoutte index technique was first proposed by Peter J. Rousseeuw in 1987[24]. It provides a graphical representation of how well each time series is in its own set. In this technique, a silhouette score is calculated for each number of clusters, and silhouette scores above the average determine the number of clusters. The silhouette score is the score that calculates how well the data is clustered and is calculated to evaluate the quality of clustering algorithms. This score is calculated separately for each time series of the different clusters. For this, the silhouette score (S) is calculated as follows:

$$S = \frac{b_i - a_i}{\max(a_i - b_i)} \tag{7}$$

Where **a** is average of the distances between the time series and other cluster elements (intra-cluster distance) and **b** is average distances of the distances between the time series and the data of the other closest cluster (average closest cluster distance).

Silhouette score ranges from -1 to 1. If the score is 1, it is said that the cluster is a dense cluster and is better separated from other clusters. If the score is close to 0, we can say that it is very close to neighboring clusters. When negative values are seen, it can be said that there is a wrong clustering. The overall average of the entire data set is found, and the largest overall mean silhouette shows the best cluster. The number of clusters with the maximum silhouette width is determined as the optimum number of clusters.

The fuzzy version of the silhouette index is calculated as follows:

$$I_{FS}\,(X;V,U = \frac{\sum_{i=1}^{n}\left(u_{ij} - u_{ij'}\right)^{\alpha}\left(\frac{b_i - a_i}{\max(b_i, a_i)}\right)}{\sum_{i=1}^{n}\left(u_{ij} - u_{ij'}\right)^{\alpha}} \tag{8}$$

b, a: weighting coefficients of fuzzy

The fuzzy silhouette index is designed in such a way that the optimal number of clusters (c) takes the maximum value.

#### Xie-beni index (XB)

The Xie-Beni (XB)[25] index is a popular measure of fuzzy set validity. It is an index that truly measures compactness and separation. This proposed index generally focuses on two features. These are the closeness of object to each other and the difference of clusters from each other. The numerator part in the formula shows the density and the denominator part shows the strength of the separation. The value that makes the index minimum is selected.

$$I_{XB}(Y;V,U) = \frac{\sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}||y_i - v_j||^2}{N\left(\min_{1 \le j,k \le c, j \ne k}\{||v_j - v_k||^2\}\right)} \quad (9)$$

### Partition coefficient (PC)

Bezdek[28] proposed the partition coefficient, which measures the amount of overlap between clusters. A performance measure based on minimizing the fuzzy intercept is defined.

$$I_{PC}(U) = \frac{1}{N}\left(\sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}\right) \quad (10)$$

The range of values for PC is (1/c, 1). The PC index has two disadvantages: it tends to decrease as the number of clusters increases and is sensitive to fuzzier; it prevents the data set from correctly determining the underlying cluster number[27]. Accordingly, the best performance is the value at which the function takes the maximum value.

### Partition entropy (PE)

$$I_{PE}(U) = \frac{1}{N}\left(\sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}\ log_b(u_{ij})\right) \quad (11)$$

The PE[28] index is a scalar measure of the amount of fuzziness in a given U. The best performance in the index is found when it takes the minimum value.

### Modified partition coefficient (MPC)[29]

$$I_{MPC}(U) = (c * I_{PC}(U) - 1)/(c-1) \quad (12)$$

The number of clusters which MPC value is maximum corresponds to optimal number of clusters.

## Experimental Results

Clustering process is performed at two steps. In the first step, optimal number of clusters is determined by using five cluster indexes defined in Section 2.4. In the second step, clustering is performed with the optimal number of clusters and the lengths of clusters (CL) are calculated to determine the risk levels of clusters. The following equation is used for CL:

$$CL_j = \sqrt{\sum_{i=1}^{n} {v_{ij}}^2}\ j = 1,2,..,c \quad (13)$$

Where n is the length of the cluster center (is equal to the length of the time series). The CL values are sorted in descending order. The cluster having maximum CL values is labelled as high risk and the cluster having minimum CL value is labelled as low risk.

### *The Clustering Results for CCOP*

In order to determine the optimal number of clusters, FKM clustering algorithm is executed for all numbers of clusters between 3 and 10. Table 3 gives the values of cluster validity indexes.

According to Table 3, the optimal number of clusters is found as 3. When clustering process is repeated for the optimal number of clusters and CL values are calculated, low risk countries are obtained as seen in Table 4

The results given in Table 4 can be summarized as follows:

- Low risk cluster consists of 52 countries.
- The cluster center corresponds to Uzbekistan. This states that information about COVID19 characteristics of the other countries can be obtained by only monitoring Uzbekistan.
- Greece, Mexico, Kazakhstan, Saudi Arabia, Singapore, and Tunisia also belong to the cluster of middle risk countries with approximately 0.3 membership degrees.
- The membership degrees of the other countries to low risk countries are generally bigger than 0.8.

Descriptive statistics for low risk countries are given in Figure 1

According to Figure 1,

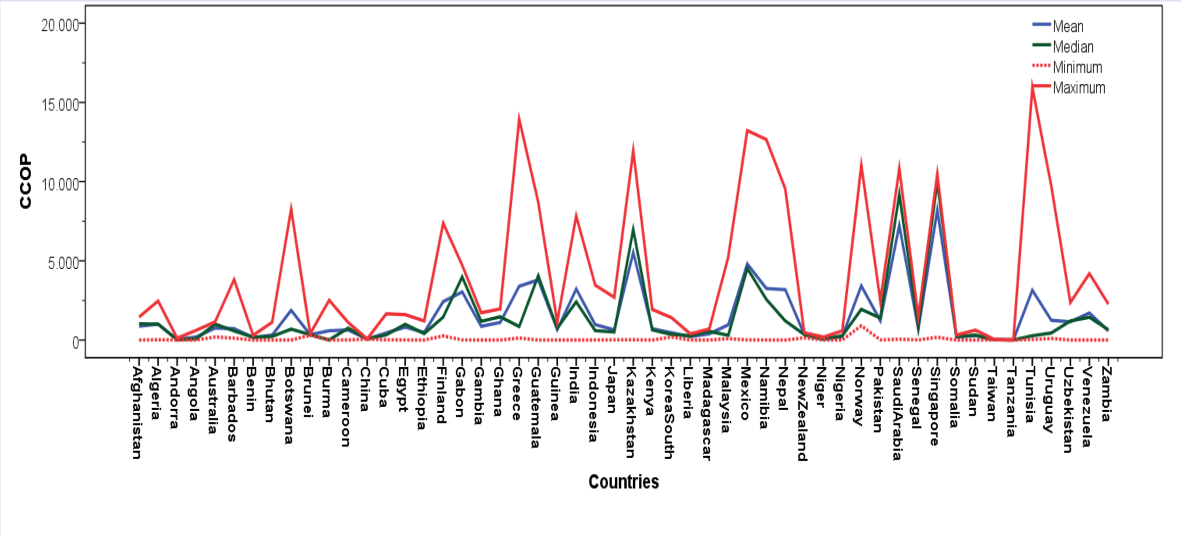
- The countries which have the highest mean and median in terms of the CCOP in the low risk cluster are Kazakhstan, Saudi Arabia and Singapore. Tanzania, Taiwan and Andorra have the smallest mean and median CCOP values. Besides, the maximum value (maximum of maximum CCOP values of the countries) of CCOP value is observed in Tunisia while minimum value is observed in Sudan.
- Variation coefficients given in Figure 1(b) indicate the variability of CCOP values observed in the date of 1 April 2020 - 19 January 2021. Accordingly, the countries having the highest variability are Greece and Tunisia. The smallest variabilities are observed in CCOP values of China, Brunei, New Zealand, Taiwan and Tanzania.
- In Figure 1(c), lines in the middle of boxes show the median values of the countries. Based on this, it can be said that the CCOP values in the countries of Angola, Bhutan, Cuba, Ethiopia, Finland, Greece, India, South Korea, Malaysia, Namibia, Nepal, Norway, Tunisia and Uruguay have increased considerably since 26 August 2020 (the mid of the time period considered) since the median values are at the bottom of the boxes.
- The CCOP values in the countries of Afghanistan, Algeria, Australia, Benin, Cameron, Egypt, Gabon, Gambia, Ghana, Guinea, Kazakhstan, Mexico, Pakistan, Saudi Arabia, Senegal and Singapore have started to increase at the beginning of the time period considered.

Table 3. Cluster Validity Indexes for CCOP

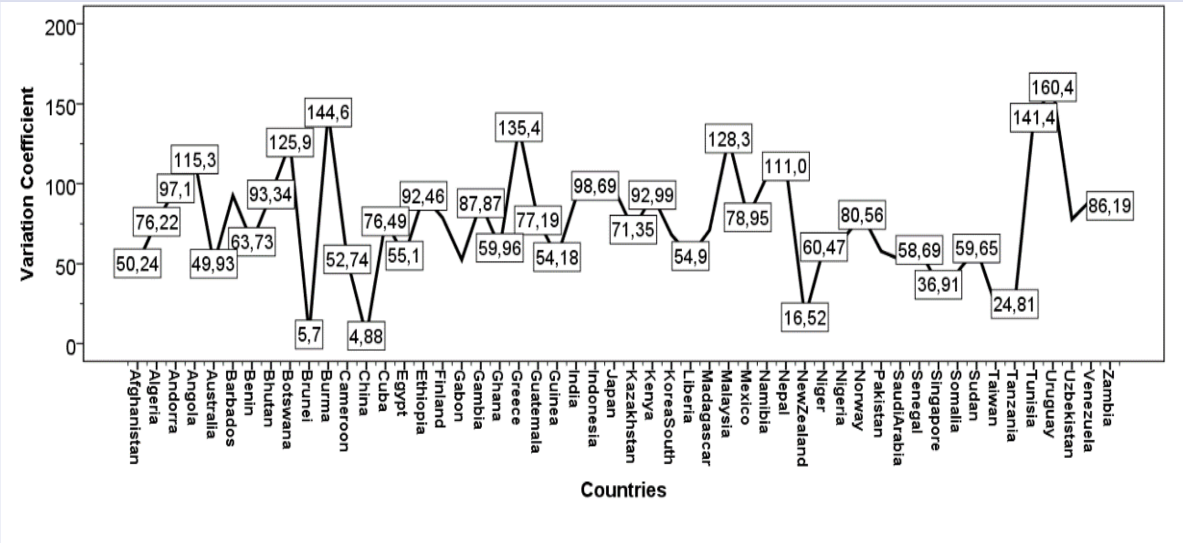
| Index/NC | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| **FS** | **0.75** | 0.75 | 0.73 | 0.72 | 0.71 | 0.65 | 0.65 | 0.67 |
| **XB** | **0.17** | 0.83 | 0.59 | 5.95 | 3.74 | 4.83 | 4.47 | 3.99 |
| **PC** | **0.73** | 0.65 | 0.61 | 0.59 | 0.54 | 0.53 | 0.52 | 0.51 |
| **PE** | **0.46** | 0.66 | 0.83 | 0.90 | 0.96 | 1.06 | 1.13 | 1.19 |
| **MPC** | **0.60** | 0.56 | 0.52 | 0.49 | 0.49 | 0.49 | 0.46 | 0.46 |

Table 4. Low Risk Countries in terms of CCOP

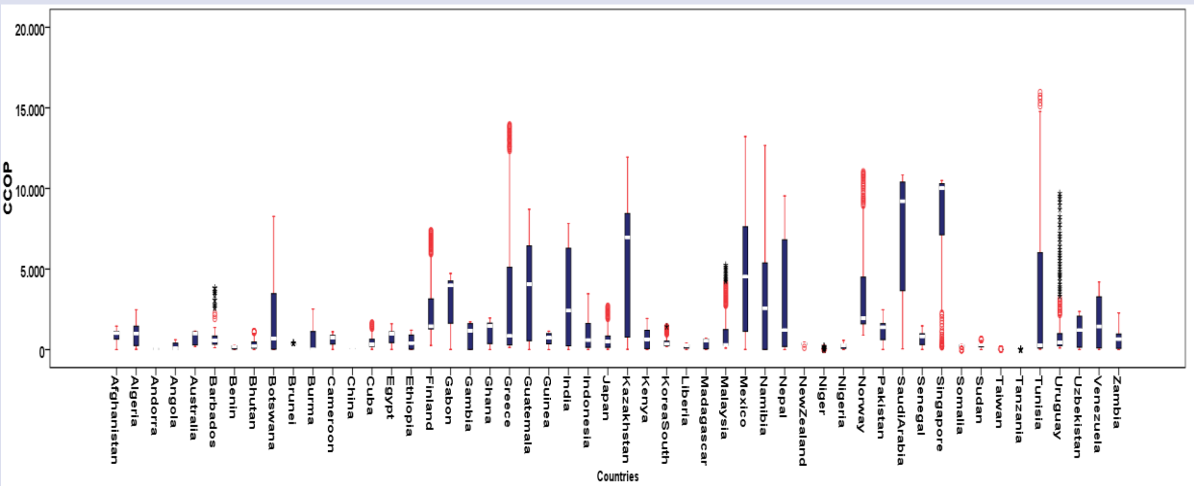
| **CL** | | | **25001.3** | | | | **Low Risk** | | |
|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **No** | **Countries** | **U1** | **U2** | **U3** |
| **1** | Afghanistan | 0.99 | 0.00 | 0.00 | **27** | Kazakhstan | 0.53 | 0.39 | 0.06 |
| **2** | Algeria | 0.99 | 0.00 | 0.00 | **28** | Kenya | 0.99 | 0.00 | 0.00 |
| **3** | Andorra | 0.98 | 0.01 | 0.00 | **29** | South Korea | 0.99 | 0.00 | 0.00 |
| **4** | Angola | 0.98 | 0.01 | 0.00 | **30** | Liberia | 0.98 | 0.01 | 0.00 |
| **5** | Australia | 0.99 | 0.00 | 0.00 | **31** | Madagascar | 0.99 | 0.00 | 0.00 |
| **6** | Barbados | 0.99 | 0.00 | 0.00 | **32** | Malaysia | 0.99 | 0.00 | 0.00 |
| **7** | Benin | 0.98 | 0.01 | 0.00 | **33** | Mexico | 0.58 | 0.36 | 0.05 |
| **8** | Bhutan | 0.98 | 0.00 | 0.00 | **34** | Namibia | 0.78 | 0.18 | 0.03 |
| **9** | Botswana | 0.94 | 0.04 | 0.00 | **35** | Nepal | 0.77 | 0.19 | 0.03 |
| **10** | Brunei | 0.98 | 0.01 | 0.00 | **36** | New Zealand | 0.98 | 0.01 | 0.00 |
| **11** | Burma | 0.99 | 0.00 | 0.00 | **37** | Niger | 0.98 | 0.01 | 0.00 |
| **12** | Cameroon | 0.99 | 0.00 | 0.00 | **38** | Nigeria | 0.98 | 0.01 | 0.00 |
| **13** | China | 0.98 | 0.01 | 0.00 | **39** | Norway | 0.82 | 0.15 | 0.02 |
| **14** | Cuba | 0.9 | 0.00 | 0.00 | **40** | Pakistan | 0.99 | 0.00 | 0.00 |
| **15** | Egypt | 0.99 | 0.00 | 0.00 | **41** | Saudi Arabia | 0.46 | 0.44 | 0.09 |
| **16** | Ethiopia | 0.99 | 0.00 | 0.00 | **42** | Senegal | 0.99 | 0.00 | 0.00 |
| **17** | Finland | 0.95 | 0.04 | 0.00 | **43** | Singapore | 0.45 | 0.42 | 0.11 |
| **18** | Gabon | 0.94 | 0.05 | 0.01 | **44** | Somalia | 0.98 | 0.01 | 0.00 |
| **19** | Gambia | 0.99 | 0.00 | 0.00 | **45** | Sudan | 0.98 | 0.01 | 0.00 |
| **20** | Ghana | 0.99 | 0.00 | 0.00 | **46** | Taiwan | 0.98 | 0.01 | 0.00 |
| **21** | Greece | 0.58 | 0.36 | 0.04 | **47** | Tanzania | 0.98 | 0.01 | 0.00 |
| **22** | Guatemala | 0.80 | 0.16 | 0.03 | **48** | Tunisia | 0.63 | 0.31 | 0.04 |
| **23** | Guinea | 0.99 | 0.00 | 0.00 | **49** | Uruguay | 0.96 | 0.02 | 0.00 |
| **24** | India | 0.84 | 0.13 | 0.02 | **50** | **Uzbekistan** | **1** | **0** | **0** |
| **25** | Indonesia | 0.99 | 0.00 | 0.00 | **51** | Venezuela | 0.99 | 0.00 | 0.00 |
| **26** | Japan | 0.99 | 0.00 | 0.00 | **52** | Zambia | 0.99 | 0.00 | 0.00 |




(a) Mean, Median, Minimum and Maximum Values For Low Risk Countries




(b) Variation Coefficients for Low Risk Countries




(c) Box-Plot for Low Risk Countries

Figure 1. Descriptive Statistics for Low Risk Countries

According to CL value, middle risk countries are obtained as in Table 5.

As can be seen in Table 5,

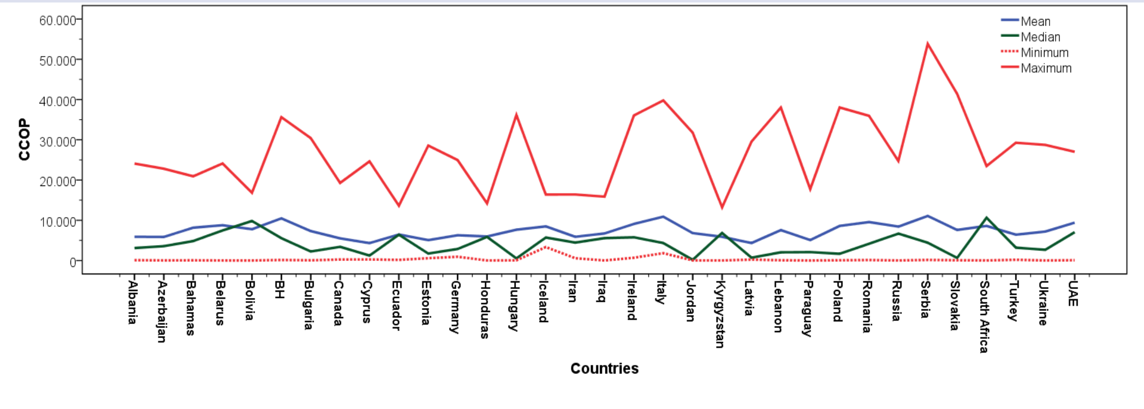
- This cluster includes 33 countries.
- The cluster center of middle risk countries is Ukraine. The behavior of Ukraine in terms of CCOP can be used to gain insight about the CCOP behavior of the countries which belong to this cluster with especially high membership degrees.
- Many countries such as Bolivia, Canada, Ecuador, Kyrgyzstan etc. have the characteristics of low risk countries at the same time since the membership degrees of these countries to low risk clusters are bigger than 0.2.
- Bosnia and Herzegovina (BH) and Serbia also belong to cluster of high risk with membership values of 0.304 and 0.441 respectively.
- The other countries belong with high membership degrees to middle risk cluster.
- Fig 2. illustrates the descriptive statistics of the middle risk countries
- When examined Figure 2, it can be seen that
- The country having the highest CCOP value in mean is Serbia.
- According to median values, the country having the highest CCOP value is South Africa.
- The highest CCOP value is observed in Serbia while the smallest CCOP value is in Paraguay.
- The countries having the highest variation coefficients are Latvia, Jordan, Slovakia and Hungary. This states that CCOP values of these countries show the most variation over time. The minimum variations are observed in Iceland, Ecuador and Bolivia.
- In Albania, Bahamas, BH, Bulgaria, Canada, Cyprus, Estonia, Germany, Hungary, Iceland, Ireland, Italy, Jordan, Latvia, Lebanon, Paraguay, Poland, Romania, Slovakia, Ukraine and UAE, CCOP values have increased much since 26 August 2020.
- In Bolivia, Kyrgyzstan and South Africa, high CCOP values have been observed at the beginning of the time period considered.

In the other countries assigned to this cluster, CCOP values have showed a more homogenous distribution.

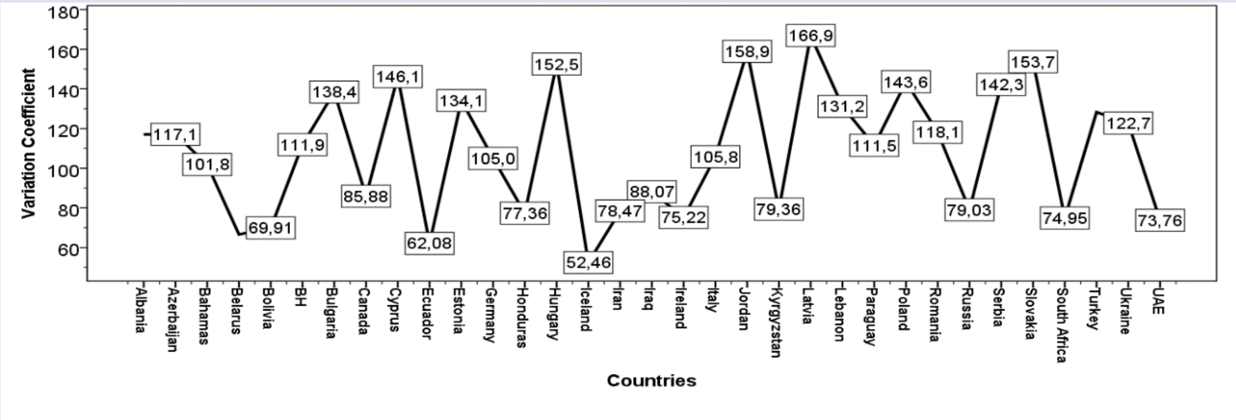
High risk countries for CCOP are given in Table 6

**Table 5.** Middle Risk Countries in terms of CCOP

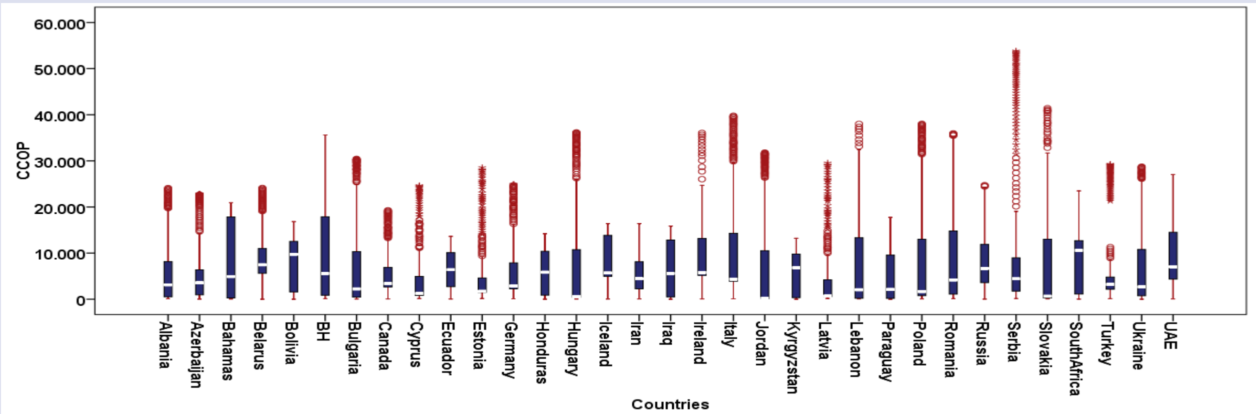
| CL | | | 195048.6 | | | | Middle Risk | | |
|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **No** | **Countries** | **U1** | **U2** | **U3** |
| **1** | Albania | 0.08 | 0.88 | 0.03 | **18** | Ireland | 0.11 | 0.78 | 0.10 |
| **2** | Azerbaijan | 0.10 | 0.85 | 0.03 | **19** | Italy | 0.06 | 0.63 | 0.29 |
| **3** | Bahamas | 0.11 | 0.78 | 0.10 | **20** | Jordan | 0.04 | 0.90 | 0.05 |
| **4** | Belarus | 0.14 | 0.76 | 0.09 | **21** | Kyrgyzstan | 0.41 | 0.51 | 0.07 |
| **5** | Bolivia | 0.30 | 0.58 | 0.11 | **22** | Latvia | 0.22 | 0.71 | 0.05 |
| **6** | BH | 0.06 | 0.63 | 0.30 | **23** | Lebanon | 0.02 | 0.95 | 0.02 |
| **7** | Bulgaria | 0.02 | 0.95 | 0.02 | **24** | Paraguay | 0.28 | 0.65 | 0.05 |
| **8** | Canada | 0.34 | 0.59 | 0.05 | **25** | Poland | 0.06 | 0.76 | 0.16 |
| **9** | Cyprus | 0.28 | 0.65 | 0.05 | **26** | Romania | 0.05 | 0.78 | 0.16 |
| **10** | Ecuador | 0.41 | 0.50 | 0.07 | **27** | Russia | 0.08 | 0.85 | 0.06 |
| **11** | Estonia | 0.20 | 0.74 | 0.05 | **28** | Serbia | 0.10 | 0.45 | 0.44 |
| **12** | Germany | 0.10 | 0.86 | 0.03 | **29** | Slovakia | 0.05 | 0.84 | 0.09 |
| **13** | Honduras | 0.40 | 0.52 | 0.07 | **30** | South Africa | 0.19 | 0.68 | 0.12 |
| **14** | Hungary | 0.05 | 0.85 | 0.09 | **31** | Turkey | 0.11 | 0.82 | 0.05 |
| **15** | Iceland | 0.24 | 0.65 | 0.10 | **32** | **Ukraine** | **0** | **1** | **0** |
| **16** | Iran | 0.33 | 0.60 | 0.06 | **33** | UAE (United Arab Emirates) | 0.08 | 0.88 | 0.03 |
| **17** | Iraq | 0.23 | 0.68 | 0.08 | | | | | |




(a) Mean, Median, Minimum and Maximum Values For Middle Risk Countries




(b) Variation Coefficients for Middle Risk Countries

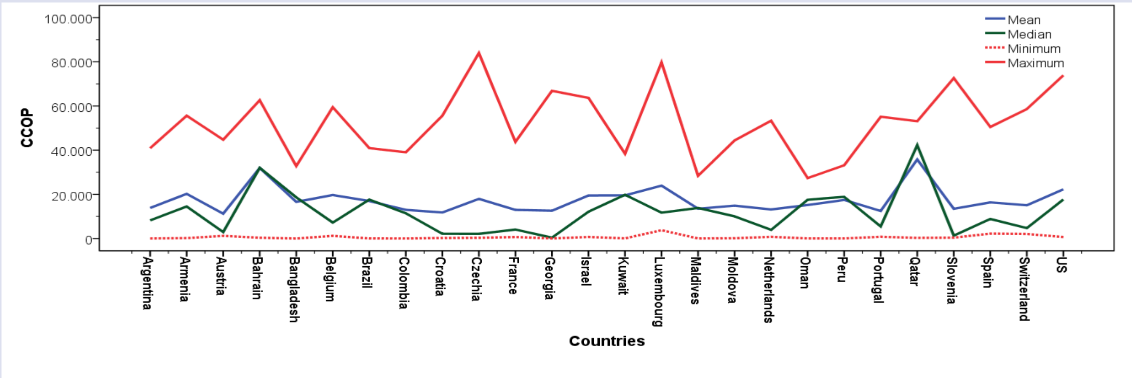



(c) Box-Plot for Middle Risk Countries

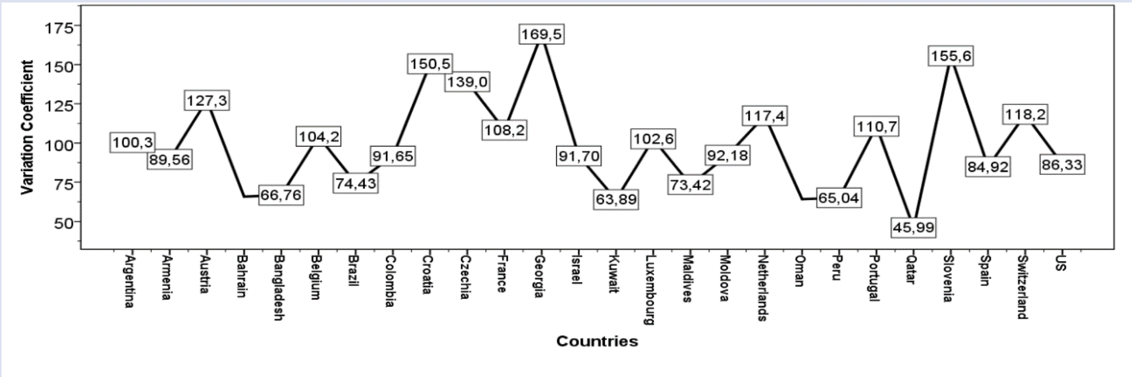
Figure 2. Descriptive Statistics for Middle Risk Countries

**Table 6.** High Risk Countries in terms of CCOP

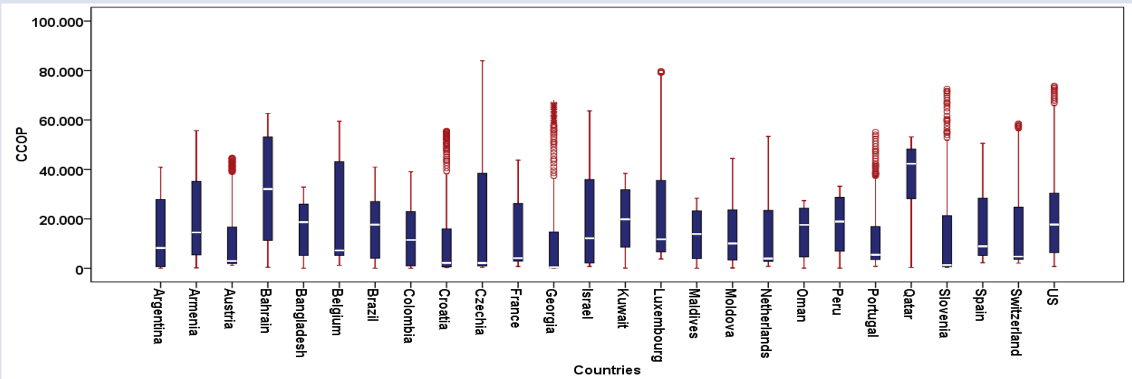
| CL | | 195048.6 | | | | High Risk | | | |
|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **No** | **Countries** | **U1** | **U2** | **U3** |
| **1** | Argentina | 0.030 | 0.118 | 0.853 | **14** | Kuwait | 0.072 | 0.171 | 0.757 |
| **2** | Armenia | 0.052 | 0.130 | 0.819 | **15** | Luxembourg | 0.115 | 0.236 | 0.649 |
| **3** | Austria | 0.073 | 0.427 | 0.501 | **16** | Maldives | 0.107 | 0.370 | 0.523 |
| **4** | Bahrain | 0.155 | 0.255 | 0.589 | **17** | Moldova | 0.016 | 0.066 | 0.918 |
| **5** | Bangladesh | 0.087 | 0.236 | 0.677 | **18** | Netherlands | 0.047 | 0.208 | 0.745 |
| **6** | Belgium | 0.068 | 0.165 | 0.766 | **19** | Oman | 0.125 | 0.333 | 0.542 |
| **7** | Brazil | 0.055 | 0.157 | 0.789 | **20** | Peru | 0.085 | 0.213 | 0.702 |
| **8** | Colombia | 0.061 | 0.287 | 0.652 | **21** | Portugal | 0.061 | 0.334 | 0.605 |
| **9** | Croatia | 0.089 | 0.345 | 0.567 | **22** | Qatar | 0.190 | 0.281 | 0.529 |
| **10** | Czechia | 0.105 | 0.240 | 0.655 | **23** | Slovenia | 0.098 | 0.288 | 0.614 |
| **11** | France | 0.038 | 0.185 | 0.777 | **24** | **Spain** | **0.000** | **0.000** | **1.000** |
| **12** | Georgia | 0.113 | 0.328 | 0.559 | **25** | Switzerland | 0.055 | 0.185 | 0.759 |
| **13** | Israel | 0.052 | 0.128 | 0.820 | **26** | US | 0.075 | 0.172 | 0.753 |




(a) Mean, Median, Minimum and Maximum Values For High Risk Countries




(b) Variation Coefficients for High Risk Countries




(c) Box-Plot for High Risk Countries

Figure 3. Descriptive Statistics for High Risk Countries

According to Table 6,

- High risk cluster contains 26 countries.
- The cluster center of high risk cluster is Spain. The CCOP characteristics of Argentina, Armenia, Belgium, Brazil, France, Moldova, Israel, Kuwait, Netherlands, Peru, Switzerland and US show high similarity with those of Spain since the membership degrees of these countries are bigger than 0.7.
- Argentina, Armenia, Austria, Brazil, Colombia, Georgia and US are assigned to high risk cluster with high membership values. Therefore, it can be said that these countries are the most risk countries in terms of CCOP.
- Bahrain, Bangladesh, Maldives, Oman, Peru, Croatia, Czechia, Portugal, Qatar, and Slovenia also belong to the Middle Risk cluster with membership values which are bigger than approximately 0.2

- Descriptive statistics for high risk cluster are given in Figure 3.
- Figure 3 can be summarized as follows:
- According to mean and median values of CCOPs in the time period monitored, the highest CCOP values are observed in Qatar and Bahrain.
- The countries seen in the maximum CCOP values are Czechia, Luxembourg and US.
- The countries whose CCOP values show the most variation are Georgia, Slovenia and Croatia while the smallest variations are obtained from Qatar, Kuwait and Oman.
- According to Fig 3(c), CCOP values observed in the countries of Argentina, Armenia, Austria, Belgium, Croatia, Czechia, France, Georgia, Israel, Luxembourg, Moldova, Netherlands, Portugal, Slovenia, Spain and Switzerland have increased much after the second half

of the time period considered. In the Bangladesh, Brazil, Oman, Peru and Qatar, the increase in the CCOP values has started in the first half of the time period. In the other countries, increase in the CCOP values is more regular.

### The Results for Cumulative Number of Deaths

Cluster validity indexes for DOP are given in Table 7. According to Table 7, it is decided that the optimal number of clusters is equal to 5 when all cluster validity indexes are evaluated simultaneously. Low risk countries in terms of DOP are given in the Table 8.

When looking at the Table 8, it can be seen that

- Low risk cluster in terms of DOP consists of 56 countries.
- The cluster center of low risk cluster is Sudan.
- Azerbaijan, Belarus, Estonia, Kazakhstan and Lebanon are also element of clusters of middle risk1 and middle risk 2 with different membership degrees. All countries except these countries show the highest similarity with Sudan in terms of DOP values.

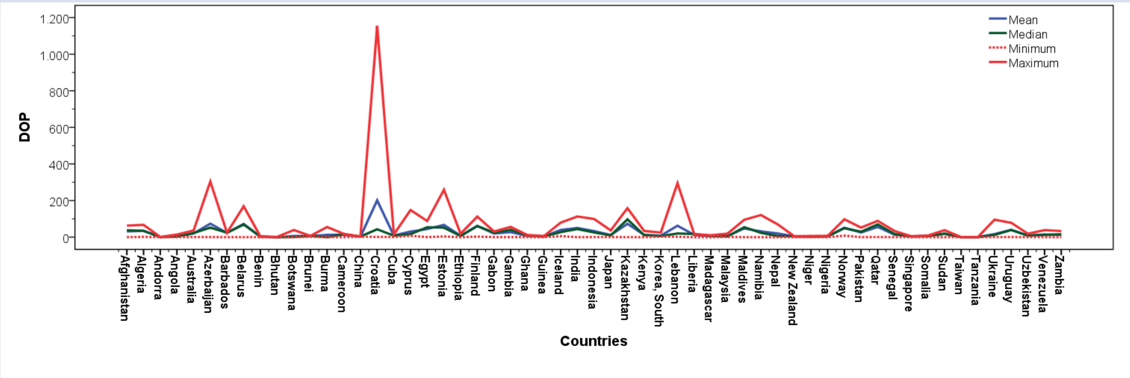
Figure 4 shows the descriptive statistics for low risk cluster.

**Table 7.** Cluster Validity Indexes for DOP

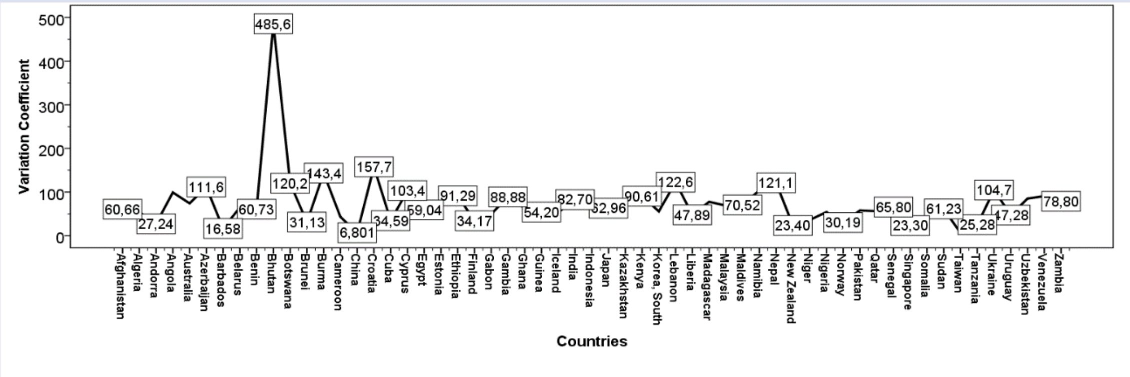
| Index/NC | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| **FS** | 0.76 | 0.77 | **0.77** | 0.73 | 0.72 | 0.70 | 0.70 | 0.67 |
| **XB** | 29.26 | 2.31 | **0.75** | 0.62 | 6.96 | 1.04 | 5.63 | 5.61 |
| **PC** | 0.39 | 0.70 | **0.68** | 0.61 | 0.58 | 0.57 | 0.54 | 0.56 |
| **PE** | 0.42 | 0.61 | **0.71** | 0.86 | 0.95 | 0.99 | 1.07 | 1.13 |
| **MPC** | 0.65 | 0.57 | **0.59** | 0.55 | 0.51 | 0.54 | 0.48 | 0.48 |

**Table 8.** Low Risk Countries in terms of DOP

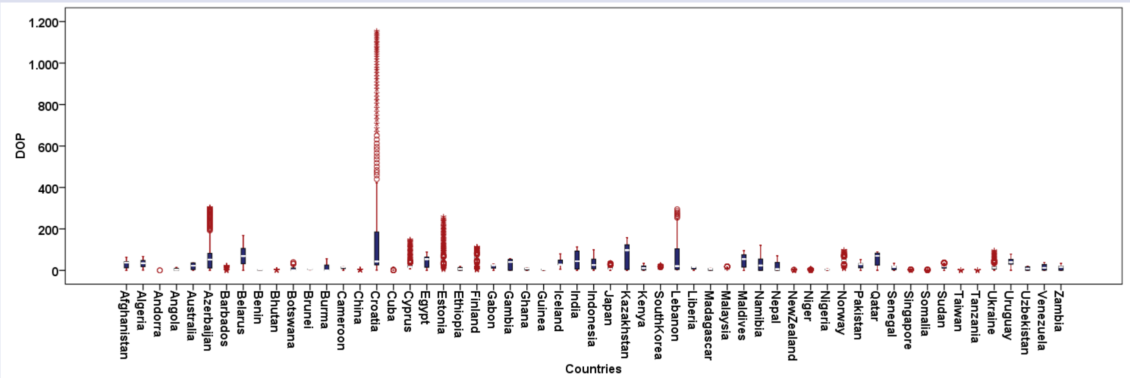
| CL | | 356.82 | | | | | Low Risk | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** | **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | U5 |
| **1** | Afghanistan | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 | **29** | Japan | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | Algeria | 0.97 | 0.01 | 0.01 | 0.00 | 0.00 | **30** | Kazakhstan | 0.53 | 0.19 | 0.22 | 0.04 | 0.02 |
| **3** | Andorra | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 | **31** | Kenya | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | Angola | 0.99 | 0.01 | 0.01 | 0.00 | 0.00 | **32** | South Korea | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | Australia | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **33** | Lebanon | 0.47 | 0.28 | 0.19 | 0.04 | 0.02 |
| **6** | Azerbaijan | 0.36 | 0.35 | 0.22 | 0.04 | 0.03 | **34** | Liberia | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **7** | Barbados | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **35** | Madagascar | 0.99 | 0.01 | 0.01 | 0.00 | 0.00 |
| **8** | Belarus | 0.61 | 0.16 | 0.17 | 0.03 | 0.02 | **36** | Malaysia | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **9** | Benin | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 | **37** | Maldives | 0.87 | 0.05 | 0.06 | 0.01 | 0.01 |
| **10** | Bhutan | 0.97 | 0.01 | 0.01 | 0.00 | 0.00 | **38** | Namibia | 0.94 | 0.03 | 0.03 | 0.01 | 0.00 |
| **11** | Botswana | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **39** | Nepal | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **12** | Brunei | 0.99 | 0.01 | 0.01 | 0.00 | 0.00 | **40** | N. Zealand | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **13** | Burma | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **41** | Niger | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **14** | Cameroon | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **42** | Nigeria | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **15** | China | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 | **43** | Norway | 0.90 | 0.04 | 0.04 | 0.01 | 0.01 |
| **16** | Cuba | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **44** | Pakistan | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **17** | Cyprus | 0.94 | 0.03 | 0.02 | 0.01 | 0.00 | **45** | Qatar | 0.83 | 0.07 | 0.08 | 0.02 | 0.01 |
| **18** | Egypt | 0.91 | 0.04 | 0.04 | 0.01 | 0.01 | **46** | Senegal | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **19** | Estonia | 0.66 | 0.16 | 0.13 | 0.03 | 0.02 | **47** | Singapore | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **20** | Ethiopia | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **48** | Somalia | 0.99 | 0.01 | 0.01 | 0.00 | 0.00 |
| **21** | Finland | 0.82 | 0.07 | 0.08 | 0.02 | 0.01 | **49** | Sudan | **1.00** | **0.00** | **0.00** | **0.00** | **0.00** |
| **22** | Gabon | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **50** | Taiwan | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **23** | Gambia | 0.97 | 0.01 | 0.01 | 0.00 | 0.00 | **51** | Tanzania | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 |
| **24** | Ghana | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | **52** | Uruguay | 0.95 | 0.02 | 0.02 | 0.01 | 0.00 |
| **25** | Guinea | 0.98 | 0.01 | 0.01 | 0.00 | 0.00 | **53** | Ukraine | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **26** | Iceland | 0.95 | 0.02 | 0.02 | 0.01 | 0.00 | **54** | Uzbekistan | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **27** | India | 0.81 | 0.08 | 0.08 | 0.02 | 0.01 | **55** | Venezuela | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |
| **28** | Indonesia | 0.95 | 0.02 | 0.02 | 0.01 | 0.00 | **56** | Zambia | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 |




(a) Mean, Median, Minimum and Maximum Values For Low Risk Countries




(b) Variation Coefficients for Low Risk Countries




(c) Box-Plot for Low Risk Countries

Figure 4. Descriptive Statistics for Low Risk Countries

According to Figure 4.

- Although Croatia has the highest DOP value (mean=200.86, median=43.12) according to mean values, the country having the highest DOP value (mean =73.12, median =97.43) according to median values is Kazakhstan.
- According to variation coefficients, the highest variations are observed in the countries of Bhutan and Croatia. The reason for this high variation in Bhutan is that no DOP value is reported at the beginning of the time period. The countries whose DOP values show the least variability are China and Taiwan.
- In the countries of Angola, Croatia, Lebanon, Nepal, Uruguay and Venezuela, DO values have increased since 26 August 2020. The DOP values of Afghanistan, Egypt, Gambia, Kazakhstan, Maldives, Qatar, Senegal and Zambia started to increase before August 2020

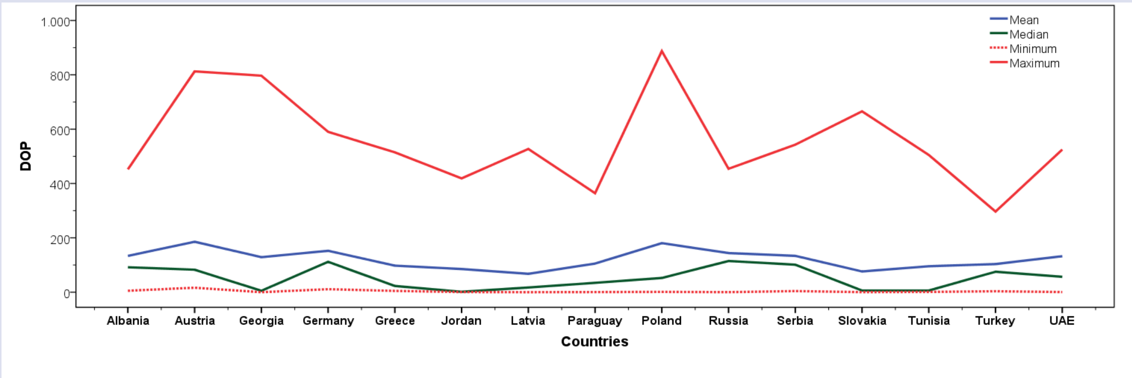
Middle risk1 countries are given in Table 9. As shown in Table 9,

- The cluster of middle risk1 consists of 16 countries and its cluster center is UAE. The countries that are the closest to UAE in terms of DOP are Albania, Greece, Jordan, Serbia and Tunisia. The other countries also have the characteristics of the other clusters with different membership degrees.

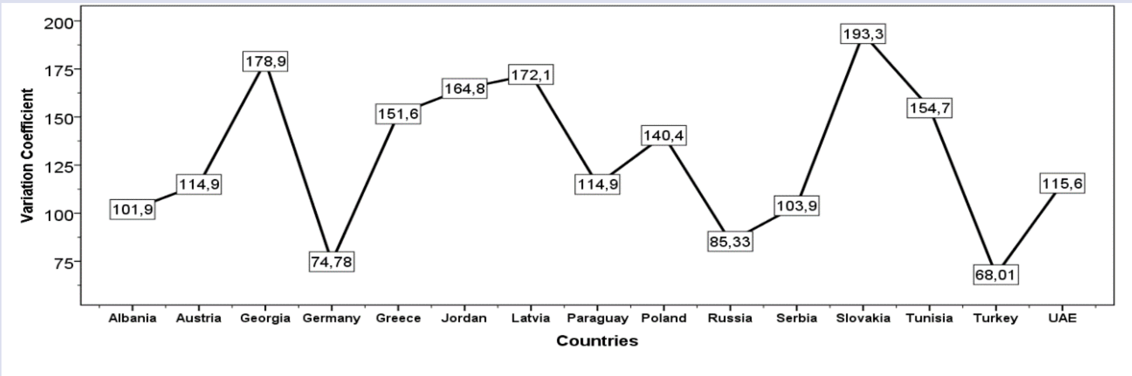
Descriptive statistics for the cluster of middle risk1 are given in Figure 5

**Table 9.** Middle Risk 1 Countries in terms of DOP

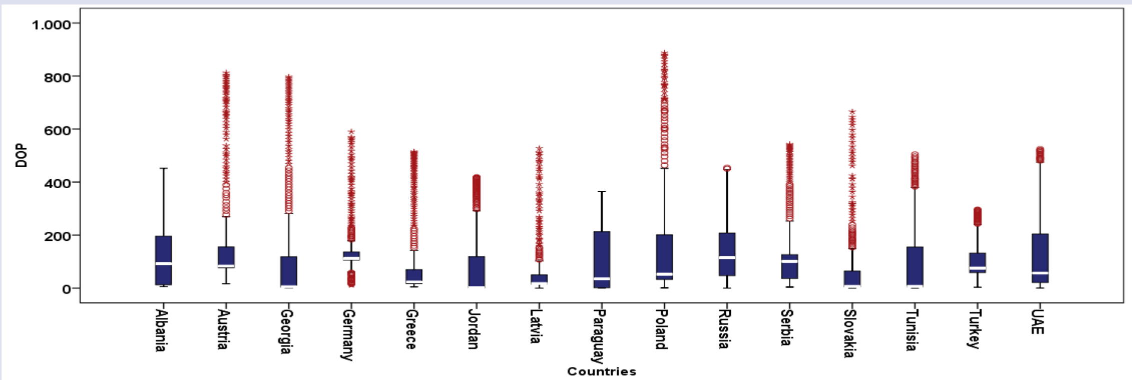
| CL | | 3464.92 | | | | | Middle Risk 1 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** | **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** |
| **1** | Albania | 0.02 | 0.85 | 0.11 | 0.02 | 0.01 | **9** | Paraguay | 0.09 | 0.60 | 0.26 | 0.04 | 0.02 |
| **2** | Austria | 0.06 | 0.49 | 0.18 | 0.17 | 0.10 | **10** | Poland | 0.07 | 0.41 | 0.17 | 0.21 | 0.14 |
| **3** | Croatia | 0.08 | 0.29 | 0.16 | 0.25 | 0.23 | **11** | Russia | 0.04 | 0.61 | 0.31 | 0.03 | 0.02 |
| **4** | Georgia | 0.08 | 0.58 | 0.16 | 0.11 | 0.07 | **12** | Serbia | 0.04 | 0.79 | 0.13 | 0.03 | 0.02 |
| **5** | Germany | 0.09 | 0.55 | 0.27 | 0.06 | 0.03 | **13** | Slovakia | 0.17 | 0.57 | 0.17 | 0.06 | 0.04 |
| **6** | Greece | 0.08 | 0.73 | 0.12 | 0.04 | 0.02 | **14** | Tunisia | 0.06 | 0.79 | 0.10 | 0.03 | 0.02 |
| **7** | Jordan | 0.10 | 0.70 | 0.14 | 0.04 | 0.02 | **15** | Turkey | 0.26 | 0.36 | 0.31 | 0.05 | 0.03 |
| **8** | Latvia | 0.28 | 0.45 | 0.19 | 0.05 | 0.03 | **16** | **UAE** | **0.00** | **1.00** | **0.00** | **0.00** | **0.00** |




(a) Mean, Median, Minimum and Maximum Values For Middle Risk Countries




(b) Variation Coefficients for Middle Risk Countries




(c) Box-Plot for Middle Risk Countries

Figure 5. Descriptive Statistics for Middle Risk Countries

Based on Figure 5, it can be said that

- The highest DOP values are observed from the countries of Poland and Austria.
- According to mean values, the countries having the highest mean of DOP values are also Austria and Poland
- While the countries that variation coefficients are the highest are Slovakia and Georgia, the smallest are Turkey and Germany.
- DOPs in the all countries except for Russia and Serbia have increased in the second time period (after 26 August 2020).
- Russia has the most regular behavior in terms of DOP. In the Serbia, the increase of DOPs has generally occurred in the first time period.

The countries in the cluster of middle risk2 are given Table 10. According to Table 10,

- Middle risk2 cluster includes 13 countries.
- The cluster center of middle risk2 cluster is Honduras. The countries belonging to this cluster with high membership degrees are Guatemala, Iraq and Oman. Thus, it can be said that information about DOP values of these countries can be obtained by monitoring Honduras.
- All countries also belong to the cluster of middle risk1 with different membership degrees except the countries of Guatemala, Iraq and Oman.
- In Figure6, the countries of middle risk 2 cluster are shown. When looking at the Figure 6, it can be seen that
- According to mean and median values, while the highest DOP values are observed in the countries of Bangladesh and Canada, Saudi Arabia is the country which has the smallest DOP values.
- Maximum DOP value in the time period monitored is observed in the country of South Africa.
- The country having maximum variation coefficient is Bahamas. Canada exhibits more stable behavior in terms of DOP values. Therefore, minimum variation coefficient is obtained for Canada.
- In the countries of Bahamas and Israel, the increase of DOPs is higher in the second time period. Bahrain, Canada, Honduras, Iraq, Kuwait, Oman and Saudi Arabia have more stable behavior in terms of increase of DOP values. In the other countries, high DOP values have been observed in the first time period.

Table 11 shows the countries which belong to cluster of high risk1

As shown in Table 11,

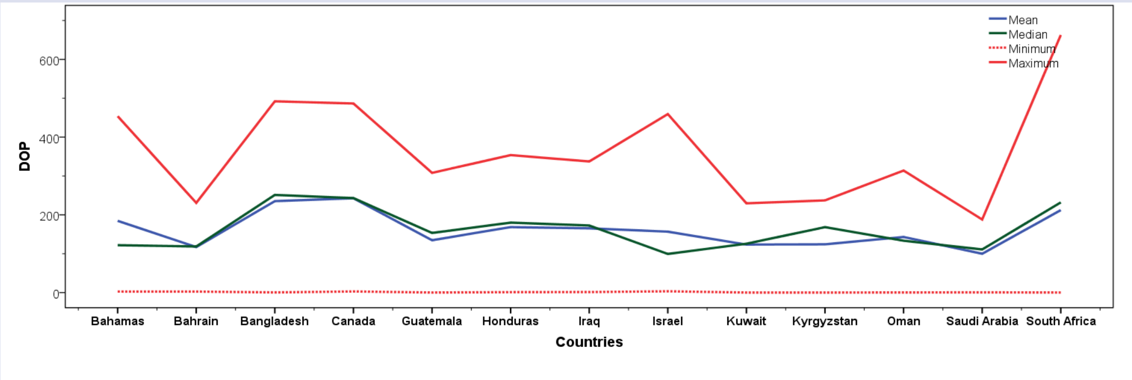
- The cluster of high risk1 contains 8 countries and the cluster center of this cluster is Iran.
- Hungary, Luxembourg, Netherlands, Romania and Switzerland also have the characteristics of the cluster of high risk 2.
- Ireland also belongs to the clusters of middle risk 2 and high risk with approximately 0.2 membership degrees.
- Portugal exhibits a more unstable behavior in terms of being assigned to the clusters.

Figure 7 demonstrates the descriptive statistics of this cluster

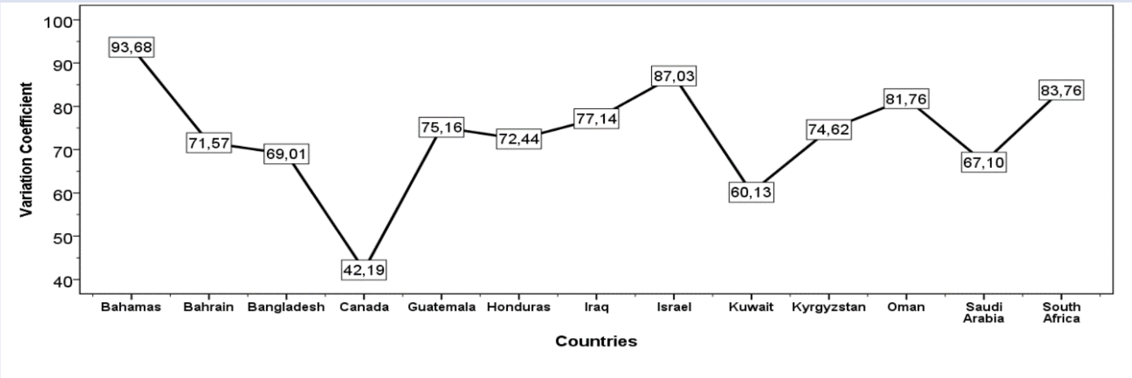
**Table 10.** Middle Risk2 Countries in terms of DOP

| | CL | | 3563.41585 | | | | | | Middle Risk 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** | **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** |
| **1** | Bahamas | 0.05 | 0.32 | 0.44 | 0.14 | 0.06 | **8** | Israel | 0.03 | 0.39 | 0.51 | 0.04 | 0.02 |
| **2** | Bahrain | 0.15 | 0.25 | 0.54 | 0.04 | 0.02 | **9** | Kuwait | 0.16 | 0.25 | 0.51 | 0.05 | 0.03 |
| **3** | Bangladesh | 0.04 | 0.15 | 0.43 | 0.28 | 0.09 | **10** | Kyrgyzstan | 0.16 | 0.21 | 0.54 | 0.05 | 0.03 |
| **4** | Canada | 0.07 | 0.21 | 0.43 | 0.20 | 0.09 | **11** | Oman | 0.03 | 0.14 | 0.79 | 0.02 | 0.01 |
| **5** | Guatemala | 0.06 | 0.17 | 0.73 | 0.03 | 0.02 | **12** | Saudi Arabia | 0.30 | 0.24 | 0.38 | 0.05 | 0.03 |
| **6** | **Honduras** | **0.00** | **0.00** | **1.00** | **0.00** | **0.00** | **13** | South Africa | 0.04 | 0.22 | 0.43 | 0.22 | 0.08 |
| **7** | Iraq | 0.00 | 0.01 | 0.98 | 0.00 | 0.00 | | | | | | | |

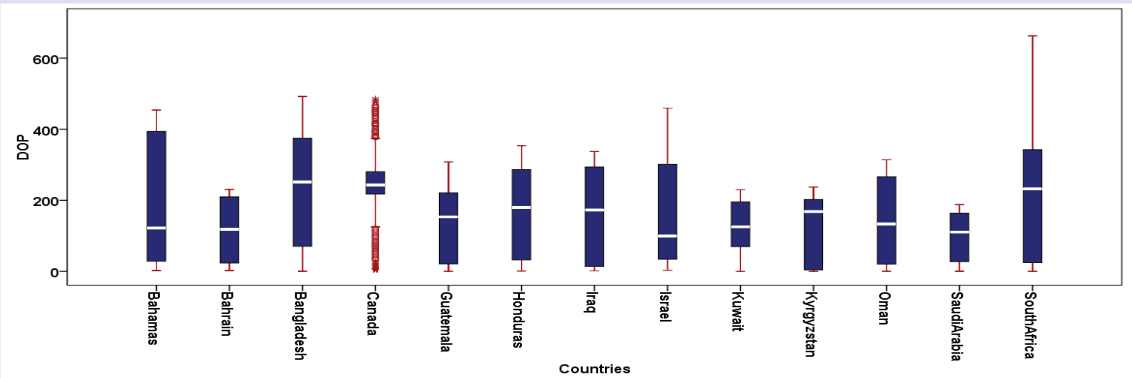
.




(a) Mean, Median, Minimum and Maximum Values For Middle Risk2 Countries




(b) Variation Coefficients for Middle Risk2 Countries

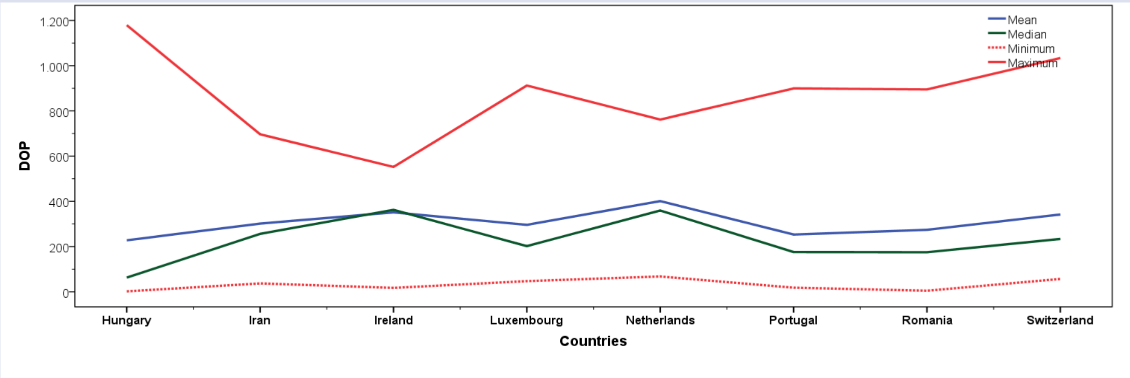



(c) Box-Plot for Middle Risk2 Countries

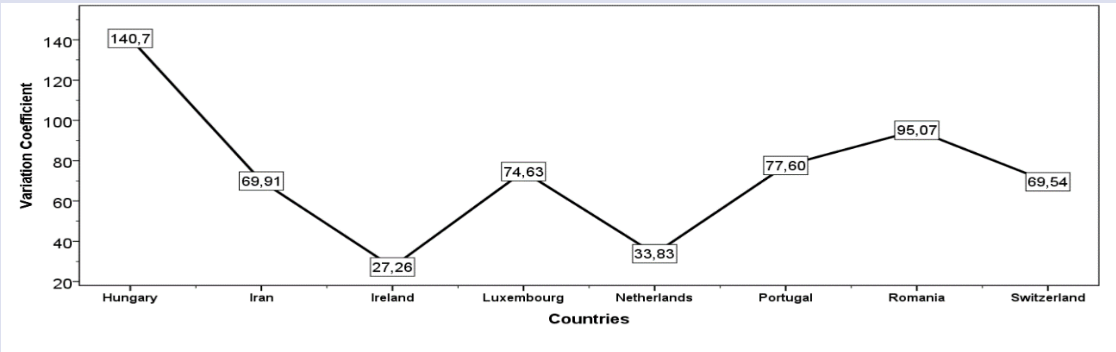
Figure 6. Descriptive Statistics for Middle Risk2 Countries

**Table 11.** High Risk1 Countries in terms of DOP

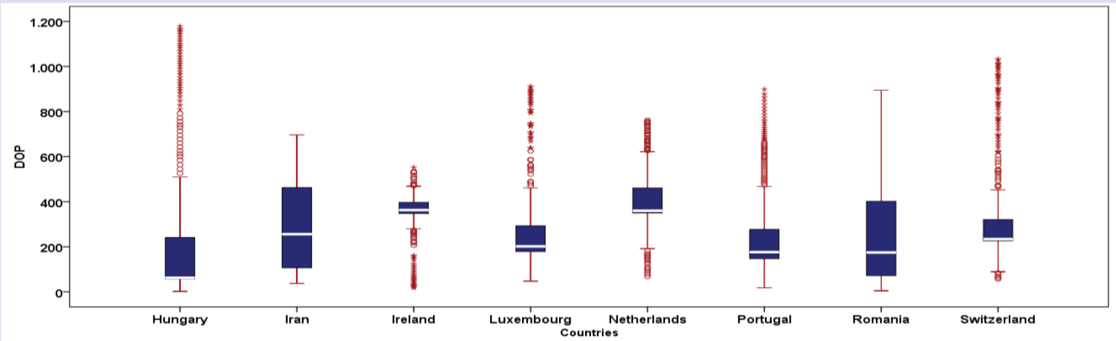
| CL | 6307.885832 | | | | | | High Risk 1 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** | **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | U5 |
| **1** | Hungary | 0.07 | 0.24 | 0.14 | 0.28 | 0.27 | **5** | Netherlands | 0.05 | 0.11 | 0.13 | 0.41 | 0.30 |
| **2** | **Iran** | **0.00** | **0.00** | **0.00** | **1.00** | **0.00** | **6** | Portugal | 0.04 | 0.22 | 0.16 | 0.41 | 0.17 |
| **3** | Ireland | 0.08 | 0.15 | 0.23 | 0.35 | 0.20 | **7** | Romania | 0.02 | 0.08 | 0.07 | 0.50 | 0.33 |
| **4** | Luxembourg | 0.04 | 0.13 | 0.11 | 0.46 | 0.26 | **8** | Switzerland | 0.04 | 0.11 | 0.09 | 0.39 | 0.37 |




(a) Mean, Median, Minimum and Maximum Values For High Risk1 Countries




Variation Coefficients for High Risk1 Countries




(c) Box-Plot for Middle Risk2 Countries

Figure 7. Descriptive Statistics for High Risk1 Countries

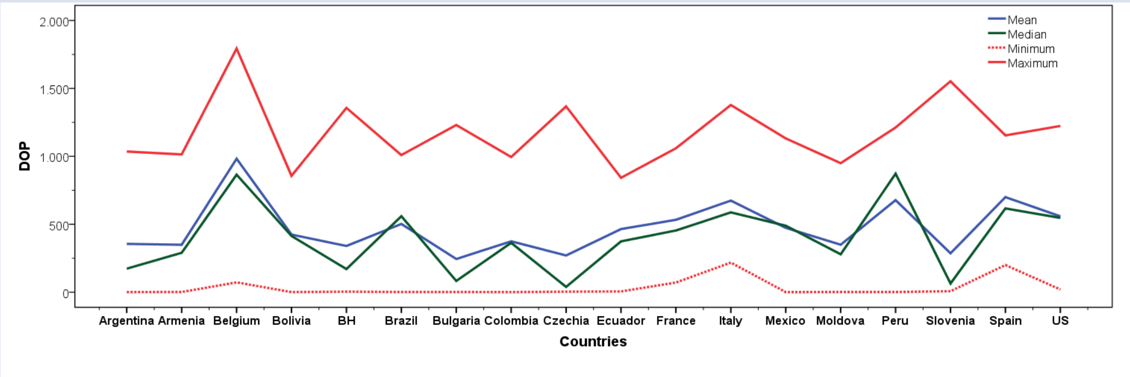
**Table 12.** High Risk2 Countries in terms of DOP

| CL | 7623.690332 | | | | | | High Risk 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** | **No** | **Countries** | **U1** | **U2** | **U3** | **U4** | **U5** |
| **1** | Argentina | 0.03 | 0.08 | 0.08 | 0.26 | 0.55 | **10** | Ecuador | 0.04 | 0.08 | 0.10 | 0.29 | 0.49 |
| **2** | Armenia | 0.01 | 0.03 | 0.03 | 0.15 | 0.79 | **11** | France | 0.06 | 0.11 | 0.12 | 0.31 | 0.41 |
| **3** | Belgium | 0.12 | 0.16 | 0.17 | 0.25 | 0.30 | **12** | Italy | 0.08 | 0.14 | 0.14 | 0.28 | 0.36 |
| **4** | Bolivia | 0.04 | 0.09 | 0.11 | 0.29 | 0.48 | **13** | Mexico | 0.04 | 0.08 | 0.09 | 0.26 | 0.53 |
| **5** | BH | 0.05 | 0.12 | 0.10 | 0.26 | 0.48 | **14** | **Moldova** | **0.00** | **0.00** | **0.00** | **0.00** | **1.00** |
| **6** | Brazil | 0.05 | 0.09 | 0.11 | 0.28 | 0.47 | **15** | Peru | 0.09 | 0.14 | 0.16 | 0.26 | 0.35 |
| **7** | Bulgaria | 0.07 | 0.20 | 0.13 | 0.28 | 0.32 | **16** | Slovenia | 0.08 | 0.18 | 0.14 | 0.26 | 0.34 |
| **8** | Colombia | 0.02 | 0.05 | 0.05 | 0.22 | 0.66 | **17** | Spain | 0.09 | 0.14 | 0.15 | 0.28 | 0.34 |
| **9** | Czechia | 0.06 | 0.17 | 0.13 | 0.28 | 0.36 | **18** | US | 0.06 | 0.10 | 0.11 | 0.28 | 0.45 |

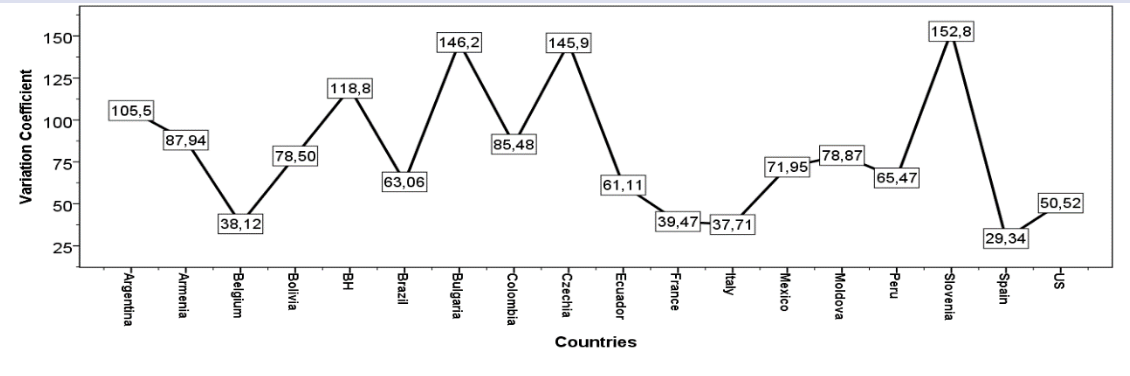
As can be seen from Figure 7,

- When mean and median values of the countries are examined, it is seen that Netherlands has the highest DOP values; Hungary has the smallest DOP values.
- The country whose DOP values change the most is Hungary. When the reason for this is investigated, it is observed that the DOP values of Hungary have increased dramatically since 2 October 2020.
- In the all countries except Iran, DOP values have increased in the second time period.
- Lastly, high risk2 countries are given in Table 12
- As shown in Table 12,
- The cluster of high risk2 consists of 18 countries and the cluster center is Moldova. Armenia and Colombia have high membership degrees. Thus, information about the DOP values of Armenia and Colombia can be obtained by monitoring the behavior of Moldova.
- All the countries except Armenia and Colombia belong to the cluster of high risk2 with small membership degrees. These countries also are members of the cluster of high risk 1.

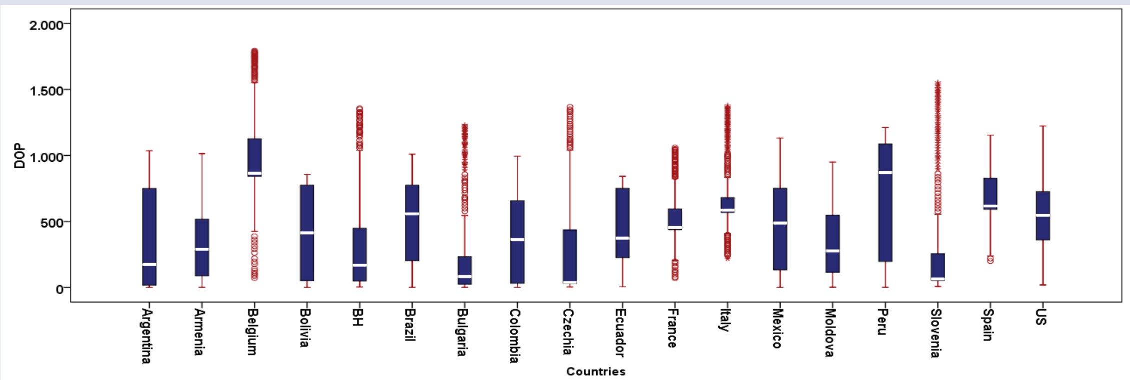
Fig 8. shows the descriptive statistics for this cluster.




(a) Mean, Median, Minimum and Maximum Values For High Risk2 Countries




(b) Variation Coefficients for High Risk2 Countries




(c) Box-Plot for Middle Risk2 Countries

Figure 8. Descriptive Statistics for High Risk2 Countries

As seen in Figure 8,

- According to mean and median values, the highest DOP values are reported from Belgium.
- Czechia and Bulgaria have the smallest DOP values in mean.
- The highest DOP value is observed in the Belgium.
- The highest variation coefficients are obtained from the countries of Bulgaria, Czechia and Slovenia. The smallest variations are observed in Belgium, France, Italy and Spain.
- According to Figure 8(c), Armenia, Bolivia, Colombia, Mexico and US have exhibited a more regular behavior in terms of DOP in the time period considered.
- In Argentina, Belgium, BH, Bulgaria, Czechia, Ecuador, France, Italy, Moldova, Slovenia and Spain, the increases in DOP values have been observed in the second time period.

## Conclusions

In this study, it is aimed to determine the countries exhibiting similar and different behavior in terms of spread of COVID19. For this objective, the data set consisting of CCOP and DOP values of 111 countries are used. Firstly, the optimal number of clusters is found by using five cluster validity indexes for each variable (CCOP and DOP). The number of clusters is determined as 3 and 5 for CCOP and DOP respectively. FKM clustering algorithm is executed with the optimal number of clusters and CL value is calculated for each cluster separately. CL values are used to reveal the risk levels of the countries with respect to COVID19. The results obtained for CCOP values are as follows:

- The cluster of low risk includes 52 countries. The cluster center of this cluster is Uzbekistan. All the countries except Greece, Mexico, Kazakhstan, Saudi Arabia, Singapore, and Tunisia have been assigned to this cluster with high membership degrees. From here, it can be said that the information about the spread of COVID19 spread in 45 countries can be obtained by following the spread of COVID19 in Uzbekistan.
- The cluster of middle risk contains 33 countries. The cluster center of this cluster is Ukraine. The COVID19 behavior of the countries of Albania, Azerbaijan, Bahamas, Belarus, Bulgaria, Estonia, Germany, Hungary, Ireland, Jordan, Latvia, Lebanon, Poland, Romania, Russia, Slovakia, Turkey and UAE show high similarity with that of Ukraine.
- High risk cluster consists of 26 countries and its cluster center is Spain. Argentina, Armenia, Belgium, Brazil, France, Israel, Kuwait, Moldova, Netherlands, Peru, Switzerland and US are assigned to this cluster with membership degrees which are bigger than 0.7 and thus, it can be said that these countries have similar COVID19 behavior with Spain.

According to DOP values, following results are obtained:

- Low risk cluster consists of 56 countries and the cluster center of this cluster is Sudan. All the countries except Azerbaijan, Belarus, Estonia, Kazakhstan and Lebanon belong to this cluster with high membership degrees. In other words, 51 countries have similar DOP behavior with Sudan.
- The cluster of middle risk 1 includes 16 countries. The cluster center is found as UAE. Albania, Greece, Jordan, Serbia and Tunisia have high similarity with UAE in terms of DOP behavior.
- Middle risk2 cluster contains 13 countries and the cluster center is Honduras. All the countries except Guatemala, Iraq and Oman also belong to the other clusters with different membership degrees.
- High risk1 cluster contains 8 countries. The cluster center is Iran. All the countries in this cluster are also assigned to the other clusters. Thus, it can be said that the countries in this cluster exhibit unstable behavior in terms of DOP.
- Lastly, the cluster of high risk2 consists of 18 countries and the cluster center is Moldova. All the countries except Armenia and Colombia are the element of the other clusters with different membership degrees.

When the clusters of CCOP and DOP are compared, the following results are obtained:

- Although the countries of Greece, Guatemala, Saudi Arabia and Tunisia are element of low risk cluster in terms of CCOP, they belong to middle risk cluster in terms of DOP values.
- Mexico belongs to low risk cluster according to CCOP values while it is the element of high risk cluster according to DOP values.
- Azerbaijan, Belarus, Cyprus, Estonia, Iceland, Lebanon and Ukraine are assigned to low risk cluster with respect to DOP. But, these countries are the element of middle risk cluster with respect to CCOP.
- Lastly, while Maldives and Qatar belong to low risk cluster with respect to DOP, they are assigned to high risk cluster with respect to CCOP.

In the future work, we planned that countries are clustered by considering three COVID19 behavior, including the number of active cases, the number of deaths and the number of recovered cases, simultaneously

## Conflicts of Interest

Sample sentences if there is no conflict of interest: The authors state that did not have conflict of interests.

## References


[1] Imtyaz A, Haleem A, Javaid M, Analysing Governmental Response to The COVID19 Pandemic, *Journal of Oral Biology and Craniofacial Research*, 10 (2020) 504-513.

[2] Zarikas V, Poulopous SG, Gareiou Z, Zervas E, Clustering Analysis of the Countries COVID19 Data Sets, *Data in Brief*, (2020) 31.

[3] Mahmoudi MR, Baleanu D, Mansor Z, Tuan BA, Pho K, Fuzzy Clustering Method to Compare The Spread Rate of COVID19 in The High-risk Countries, *Chaos, Solitions and Fractals*, (2020) 140.

[4] Alvarez E, Brida JG, Limas E, Comparisions of COVID19 dynamics in the different countries of the World using time series clustering, (2020), medRxiv

[5] Hutagalung J, Ginantra NLWSR, Bhawika GW, Parwita WGS, Wanto A, Panjaitan PD. COVID19 Cases and Deaths in Southeast Asia Clustering Using K-means Clustering, Annual Conference on Science and Technology Research, Journal of Physics: Conference Series, (2021) 1783.

[6] Virgantari F, Faridhan YE. K-means clustering of COVID19 cases in Indonesia's provinces, Proceedings of the International Conference on Global Optimization and Its Applications, Jakarta, Indonesia, November 21-22, (2020).

[7] Rojas F, Valenzuela O, Rojas I,. Estimation of COVID19 dynamics in the different states of the United States using time series clustering, (2020), medRxiv.

[8] Azarafza M, Azarafza M, Akgün H, Clustering Method for Spread Pattern Analysis of Corona-virus (COVID19) Infection in Iran, *Journal of Applied Science, Engineering, Technology, and Education,* 3(1) (2021).

[9] Crnogorac V, Grbic M, Dukanovic M, Matic D, Clustering of European countries and territories based on cumulative relative number of COVID19 patients in 2020, 20th International Symposium INFOTEH, (2021).

[10] Sadeghi B, Cheung RCY, Hanbury M, using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020, *BMJ Open*, (2021).

[11] Putra PMA, Kadyanan GAGA, Implementation of K-Means clustering algorithm in determining classification of the spread of the COVID19 virus in Bali, *Jurnal Elektronik Ilmu Komputer Udayanan*, 10(1) (2021).

[12] Utomo W, The comparision of k-means and k-medoids algorithms for clustering the spread of the COVID19 outbreak in Indonesia, *ILKOM Jurnal Ilmiah*, 13(1) (2021).

[13] Abdullah D, Susilo S, Ahmar AS, Rusli R, Hidayat R, The application of K-Means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data, *Quality and Quantity*, (2021).

[14] Everitt B, Landau S, Leese M, Cluster analysis, 4th ed. London: Arnold, (2001).

[15] Hair J, Black W, Babin B, Anderson R, Tatham R, Multivariate data analysis. 6th ed. Uppersaddle River, N.J.: Pearson Prentice Hall, (2006).

[16] Jain A, Murty M, Flynn P, Data Clustering: A Review, *ACM Computing Surveys*, 31(3) (1999) 264-323.

[17] De Carvalho FAT, De Melo FM, Lechevallier Y, A Multi-view Relational Fuzzy C-medoid Vectors Clustering Algorithm, *Neurocomputing*, 163 (2015) 115-123.

[18] Rani S, Sikka G, Recent Techniques of Clustering of Time Series Data: A Survey, *International Journal of Computer Applications*, 52 (2012).

[19] Bezdek JC, Pattern Recognition with Fuzzy Objective Function Algoritms. New York: Plenum Press , (1981)

[20] Gustafson DE, Kessel WC, Fuzzy Clustering with a Fuzzy Covariance Matrix. IEEE CDC San Diego, (1979) 761-766.

[21] Hathaway RJ, Bezdek JC, Switching regression models and fuzzy clustering, *IEEE Transactions On Fuzzy Systems*, 1(3) (1993) 195–204.

[22] Krishnapuram R, Joshi A, Yi L, A fuzzy relative of the k-medoids algorithm with application to web document snippet clustering, IEEE International Fuzzy Systems, Conference Proceedings, (1999).

[23] Labroche N, New incremental fuzzy C medoids clustering algorithms, Annual Conference of the North American Fuzzy Information Processing Society—NAFIPS; Toronto, ON, Canada, (2010).

[24] Rousseeuw PJ, Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Computational and Applied Mathematics*, 20 (1987) 53–65.

[25] Xie X, Beni G, A Validity Measure for Fuzzy Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(8) (1991) 841-846.

[26] Bezdek JC, Numerical taxonomy with fuzzy sets, *Journal of Mathematical Biology*, 1(1) (1974) 57-71.

[27] Li CS, The Improved Partition Coefficient, *Procedia Engineering*, 24 (2011) 534-538.

[28] Bezdek, JC, Cluster validity with fuzzy sets. Journal of cybernetics, 3(3) (1974) 58-72.

[29] Dave RN, Validating fuzzy partitions obtained through c-shells clustering, *Pattern Recognition Letters*, 17(6) (1996) 613-623.