

COVID-19 USING COGNOS

ABSTRACT :

Covid-19 is an infectious illness caused by a newly identified form of coronavirus. This is a new virus and illness that was previously unknown before the December 2019 outbreak in Wuhan, China. The number of confirmed cases of Covid-19 and the number of deaths due to this virus in World are increasing and quite alarming. Therefore this study will discuss the grouping of Cases and Deaths of COVID-19 in EU/EEA Countries. The method used is the K-Means Clustering Data Mining and standard deviation using Two path algorithm . By using this method the data that has been obtained can be grouped into several clusters, where K-Means Clustering Process is applied and by following the mean to the standard deviation with the metioned formula. Data used are Country statistics, Area of recorded laboratory confirmed cases of COVID-19, and April 2020 deaths from WHO (World Health Organization). Data is divided into 3 clusters: high (C1), medium (C2) and low (C3). The results obtained are that there are four countries with a high level cluster (C1), one country with a moderate level cluster (C2), and 6 countries with a low level cluster (C3). This can be an input for each country to increase awareness of the transmission of Covid-19.

METHODOLOGY :

1.MEAN USING K-MEANS ALGORITHM :

1.1. RESEARCH DATA :

Data used are Country statistics, Area of recorded laboratory-confirmed cases of COVID-19, and April 2020 deaths from WHO (World Health Organization) .

DATASET LINK : <https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases>

1.2. RESEARCH METHOD :

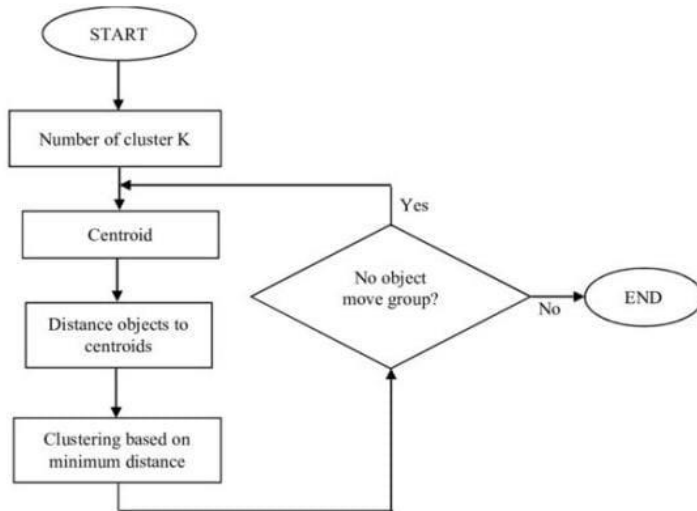
This study uses the K-Means Clustering method. K-Means is one of the clustering algorithms used in the Unsupervised learning group that is used to classify data into several classes with a partition of the system. This algorithm accepts data entries in the form of class labels .

Centroid

A midpoint value, or centroid data, is generated when implementing the K-means algorithm. The method of determining a midpoint value is achieved by

COVID-19 USING COGNOS

following the largest (maximum) for high cluster value (C1), the mean value for a medium cluster (C2), and the lowest Cluster value (C3).

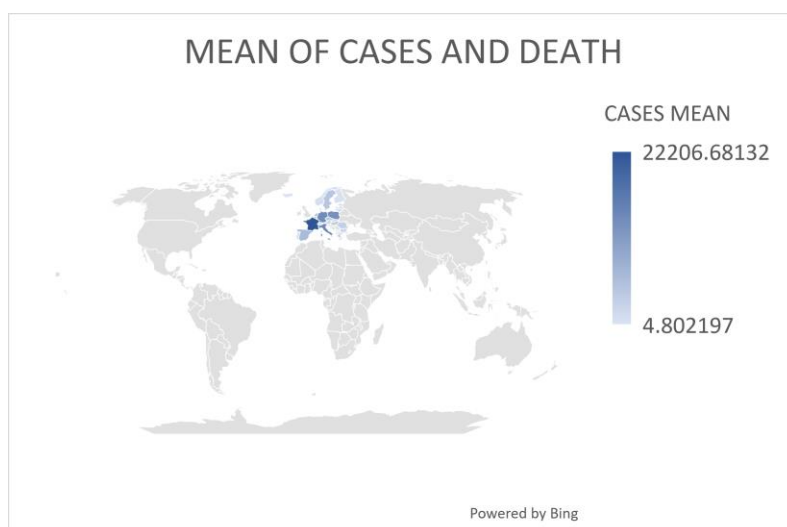


RESEARCH FLOWCHART FOR K-MEANS

ALGORITHM :

Steps to perform clustering using k-means algorithm :

- Determine cluster counts (k) in the data set.
- Determine the center value (Centroid).
- On each record, calculate the closest distance to Centroid
- Distance Group objects to nearest Centroid
- Repeat step a to step b, iterating until Centroid is optimal



COVID-19 USING COGNOS

Let us consider the data segment of country Austria from the dataset, here we taken account of death

[5,6,11,4,19,8,3,3,8,12.....21,23,9,16,24,19,14,17,25

Now we select K random points from the data as centroids

So, No.of clusters into 3 sets named as K1, K2 &K3

K1={3,4,5,6,....13,14,15}

K2={16,17,18,19,....26,27,28}

K3={30,31,32,.....,48,46,51}

Assigning all the points to the closest cluster centroids as

$M1=(3+4+5+6+.....13+14+15)/100$

= 9

$M2=(16+17+19.....+26+27+28)/13$

=22

$M3=(30+31+32+.....+48+46+51)/19$

=37

Compare the first value of k2 with the m values of k1,k2 and k3 then find the difference these three values

If the mean value is same with above step stop orelse repeat the steps

CONCLUSION FOR MEAN :

Clustering for Cases and Deaths caused by COVID-19 EU/EEA Countries. clustering uses 3 clusters, that is: (C1) high, (C2) regular cluster, and (C3) low cluster. From the results of clustering, 30 countries is in (C1) High Cluster that is included in the red zone category. The average of cases were greater than average of deaths. The cases having an average of 3661.011 data and the deaths having an average of 65.29194 data. The country France has a highest mean of cases is 22206.68132. The country Liechtenstein has a lowest mean of cases is 4.802197. The country Poland has a highest mean of death is 329.3296703. The country Iceland has a lowest mean of death is 1.09889.

COVID-19 USING COGNOS

2.STANDARD DEVIATION FROM VARIANCE

Standard deviation measures how far apart numbers are in a data set. Variance, on the other hand, gives an actual value to how much the numbers in a data set vary from the mean. Standard deviation is the square root of the variance and is expressed in the same units as the data set. Thus we use two pass algorithm to find the variance of the data .

2.1. VARIANCE BY TWO PASS ALGORITHM :

A two-pass algorithm is a computational method that processes data or performs a task in two sequential passes or steps. Each pass serves a specific purpose and often involves reading, analyzing, or transforming data. Two-pass algorithms are commonly used in various fields of computer science and data processing.

Step 1: finding a mean by mentioned formula. (death in Sweden)

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n},$$

Where , x- death data of sweden country.

n- time period , 90.

Mean=15.96703

Step 2: finding variance by Two pass algorithm.

$$\text{sample variance} = s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Where,

$$s^2 = ((1-15.96703)^2 + (5-15.96703)^2 + (815.96703)^2 + (215.96703)^2 + \dots + (19-15.96703)^2) / (90-1)$$

variance = 33.21001

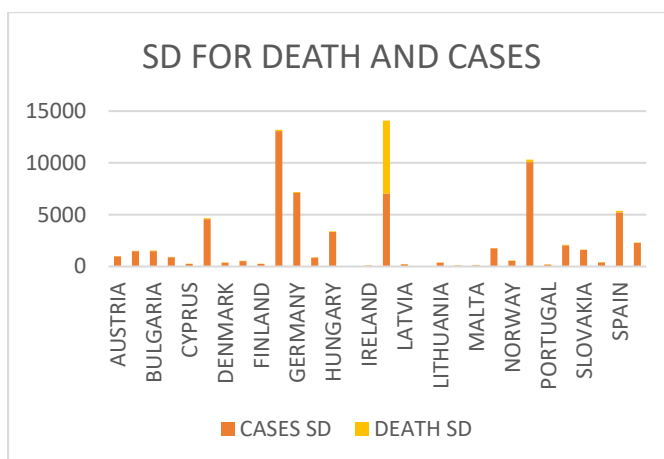
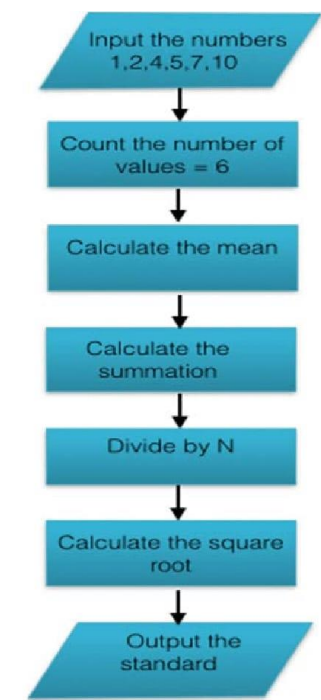
step 3 : standard deviation.

By taking square root for variance we get standard deviation .

SD = 5.762813

COVID-19 USING COGNOS

FLOW CHART FOR TWO PASS ALGORITHM



CONCLUSION FOR SD :

Standard deviation for Cases and Deaths caused by COVID-19 EU/EEA Countries. The cases having an standard deviation of 6490.51 data and the deaths having an standard deviation of 113.9566 data. The country Poland has a highest SD of cases is 10077.11733. The country Iceland has a lowest SD of cases is 8.0349. The country Italy has a highest SD of death is 7041.66. The country Iceland has a lowest SD of death is 1.20483.