Capstone Project
Coronavirus: The New Challenge

1. Introduction

The year 2020 was expected to bring in some new hope for the world. The ongoing global slowdown, fueled by many issues such as US-China trade war, Iran-US conflicts had made huge impact on the global economy. Everybody hoped for positive developments. However, two months onwards, we have a challenge from an unexpected corner. The Coronovirus (Covid14), that outbroke in China initially, has become a threat for the whole world. According to the latest updates, more than one hundred thousand people have been infected by coronavirus throughout the world[1], across more than 100 countries and there have already been 6513 deaths reported.[2]

The severity of the CoV infection arises because it can be transmitted through human contacts and through sneeze of infected people. Human movement in crowded places are an absolute problem in this case. Malls, fairs, airports, railway stations, busy markets have the potential to spread this disease. Another problem is that its symptoms are similar to fever and influenza, if not controlled it can cause serious situations like pneumonia and other diseases leading to death. So far there has been no medicines or vaccine available for this viral infection. The only preventive measures are hand wash, covering mouth, etc, [3]

This capstone project is an attempt to look at various trends of spread of the virus across the globe and to look at various patterns of infection and recovery. This project includes data analysis of time series data on infection, recovery, and death cases. The objectives of this project are:
- To understand the trend of change in infection, recovery and death due to CoV
- To identify the clusters (locations) in which the virus has spread
- To understand the recovery efforts and their effectiveness

2. Data Description and Sources

Data on CoV infection, recovery, and death are available day-wise till 14 March 2020. Data are available on the following attributes:
- Date
- Location (including latitude and longitude)
- Country
- Confirmed cases of infection, recovery, and death data from 22 January 2020 onwards

Source of datea: Data has been downloaded from Kaggle website (https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset).

---

[1] https://www.who.int/news-room/detail/07-03-2020-who-statement-on-cases-of-covid-19-surpassing-100-000

[2] https://coronavirus.jhu.edu/map

[3] https://www.who.int/health-topics/coronavirus

3. Methodology

The following methodology has been followed for this project work:

- Downloaded data were put into dataframe and the dataframes were cleaned with formatted using various methods such as drop of columns, group by countries, transpose etc.
- Data Visualisation was done using bar graphs and line graphs for presentation of information.
- Machine learning methods were followed by use of clustering methods. Hierarchical Clustering was done to understand whether

4. Results

It can be observed that China has the highest number of Corona Virus infections. This is followed by Italy, Iran and South Korea (Fig 1).
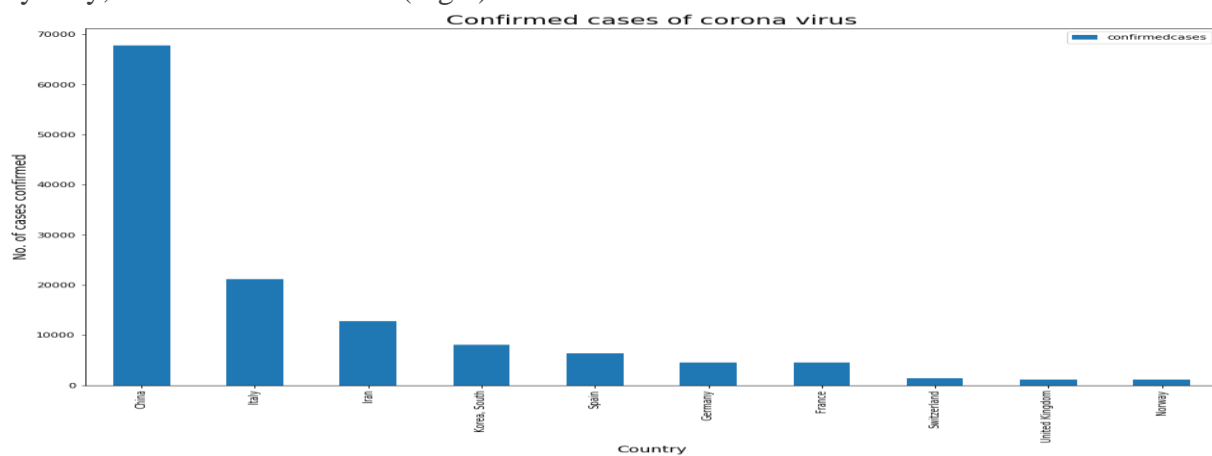
Fig 1: Confirmed cases of corona virus

Overall, there was 128,796 cases out of which 52% was in China. Although the outbreak of the virus is attributed to Wuhan, China, surprisingly Italy has already got over 20,000 confirmed cases followed by about 12,000 in Iran. South Korea has reported about 8,000 cases which is not surprising because of its proximity to China, both geographically and also in terms of trade relationship.

The cases of deaths were about 4.33 per cent of confirmed cases so far. Total 5,578 people have died as per official reports. And as expected majority of deaths have been found in China only. Fig 2 shows the death cases due to corona virus. It can be observed that the deaths in China, the highest, is followed by Italy and Iran.
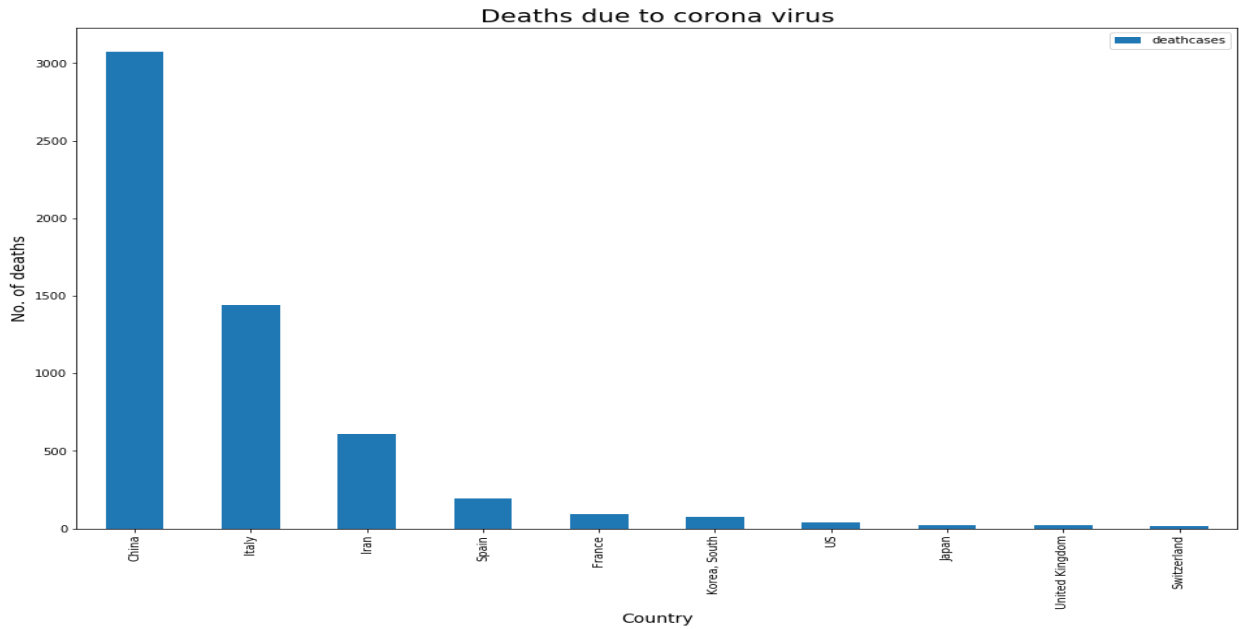
Fig 2: Death Due to corona virus

Surprisingly, the death proportion is very high in Italy (6.81%) and Iran (4.80%), which are higher than China which has reported death of 4.54 per cent. South Korea has been able to contain it remarkably reporting only 0.89 per cent of death.

So far, about 60,000 cases of recovery has been reported, the highest (52,960) reported in China. The high recovery means that China has been resilient in containing the healthcare. However, it does not mean that other countries have very low recoveries. Actually, the outbreak was first observed in China, hence recoveries are also faster.
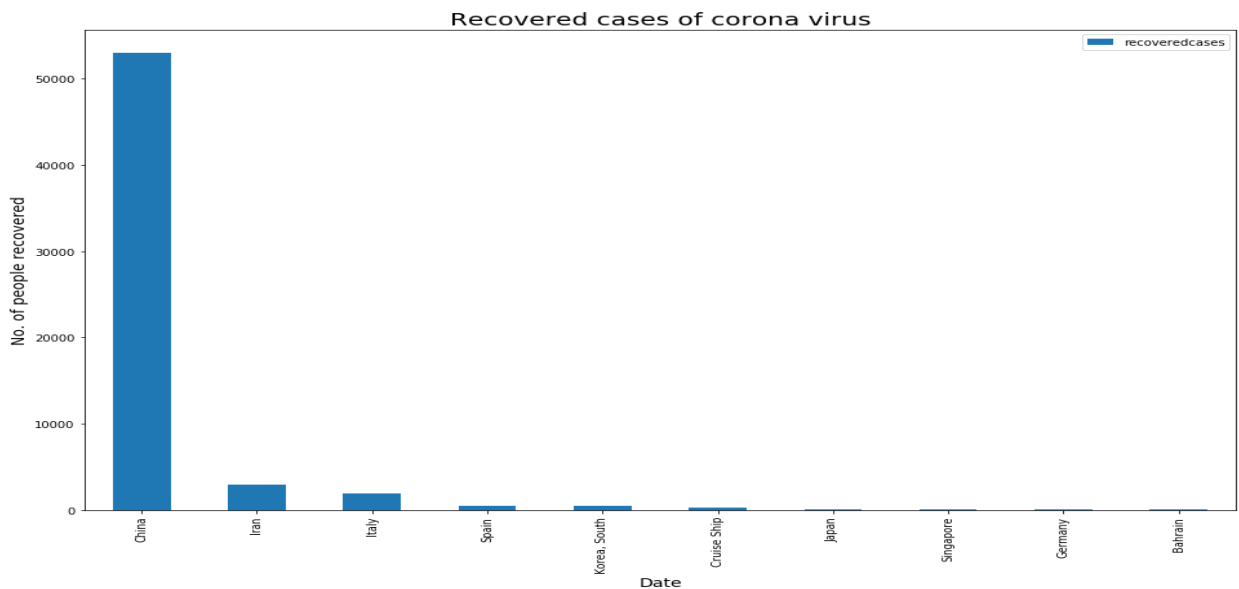

Fig 3: Recovered cases from Corona Virus.

Now let's have a look at the trend of how the outbreak has taken place. Fig 4 provides a glance of how the outbreak took place over time. It can be seen that in China, the outbreak has started plateauing in the initial days of the month of February, 2020.
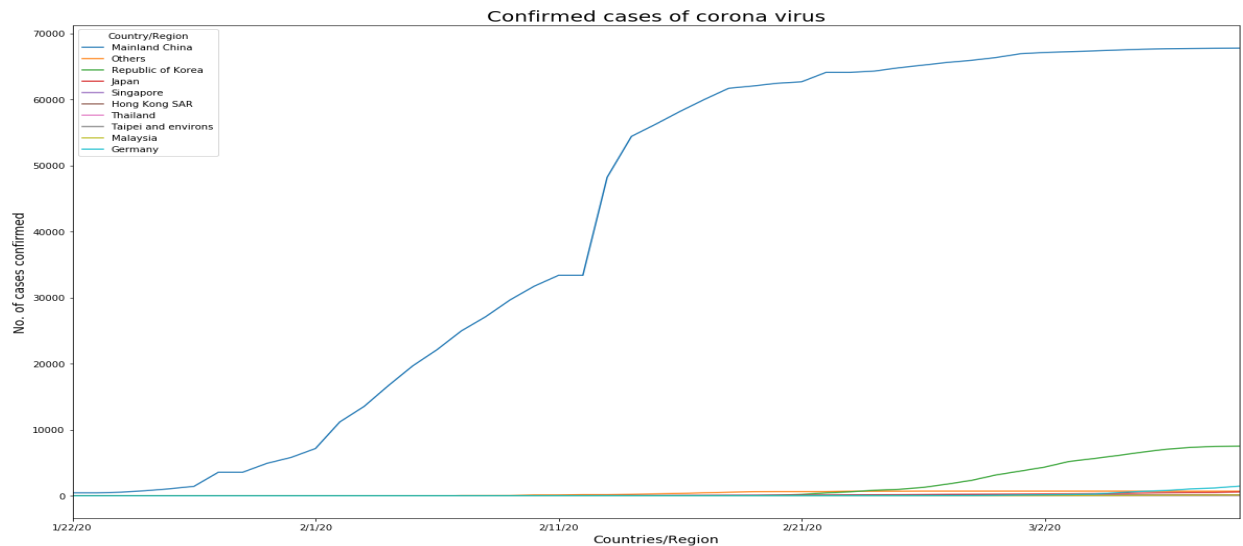
**Fig 4**: Outbreak of Coronavirus over time.

Fig 5 provides a glance of how the outbreak has taken place in countries other than China. It can be observed that in Italy and Iran there has been a major surge towards the end of February.
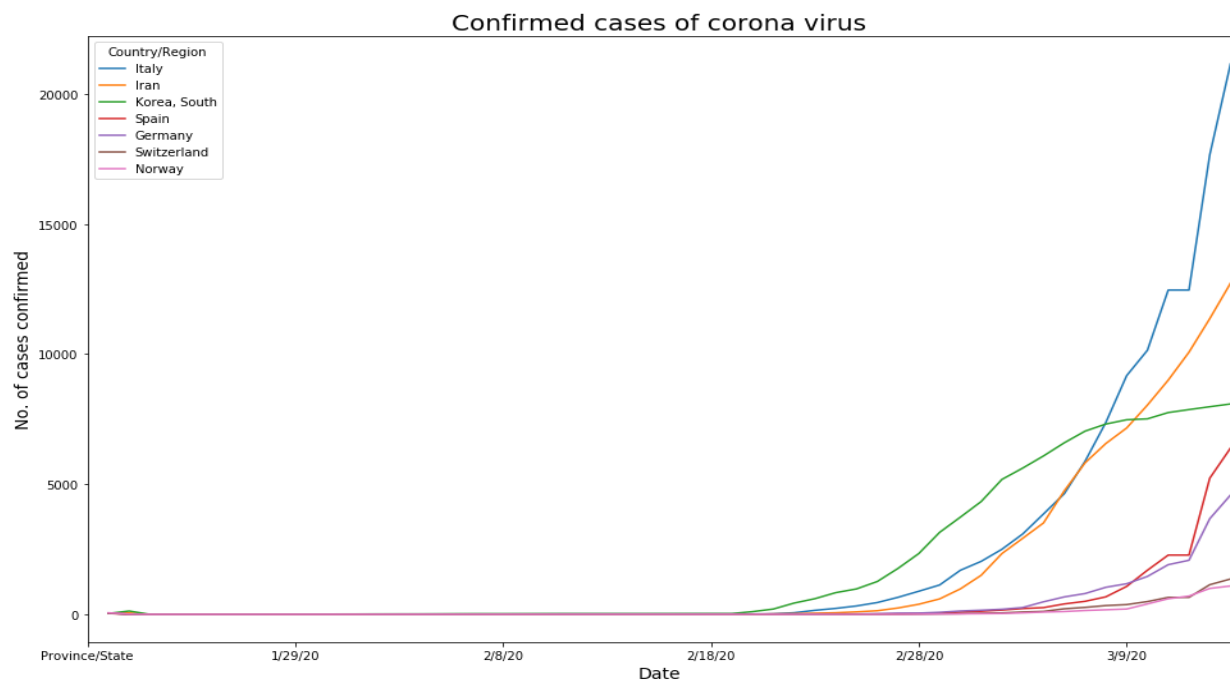
Fig 5: Outbreak of Coronavirus in Countries other than China

Hierarchical Cluster Analysis

Hierarchical Clustering was done using the latitude and langitude data. As can be seen in the following Denndrogram, three clusters can be observed clearly. The Silhoutte coefficient (which ranges from -1 to 1), which denotes the overalap among the cluster was found to be 0.913 whch shows that there is extremely weak overlap.
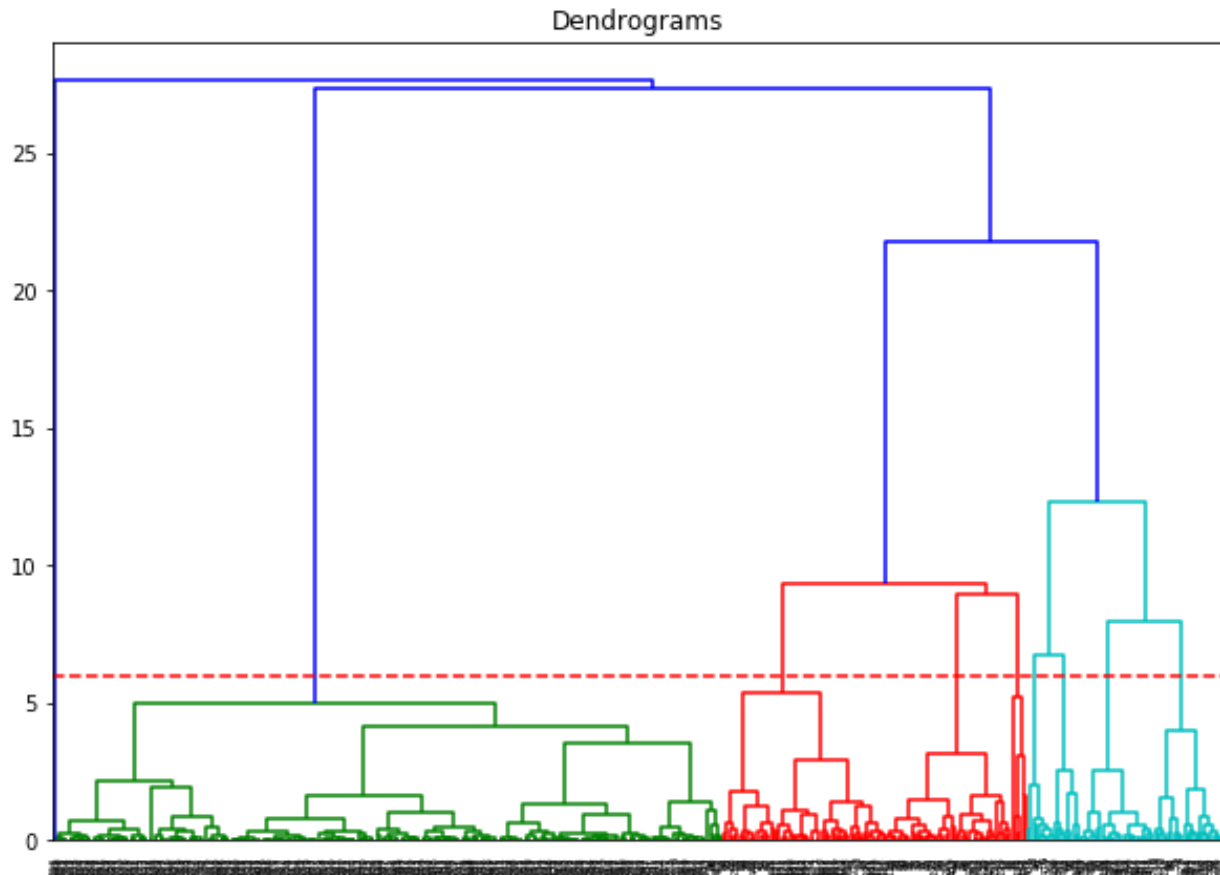

Fig 6: Dendrogram of lat long data of Corona virus observations (confirmed cases)

Let's now see how the clustering has been done. Fig 7 represents the scatter diagram of the latitude and longitude where confirmed cases of corona virus infection has been found. We can confirm that there are about three clusters, which can be categorized into
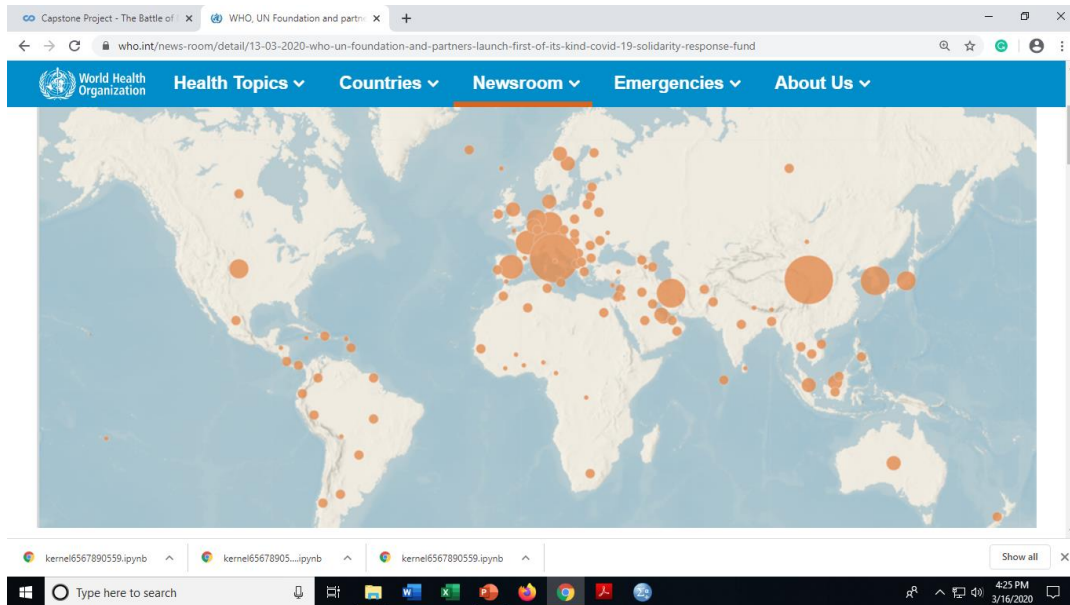- China cluster
- Iran cluster
- Italy Cluster

In other places, the outbreak is not as
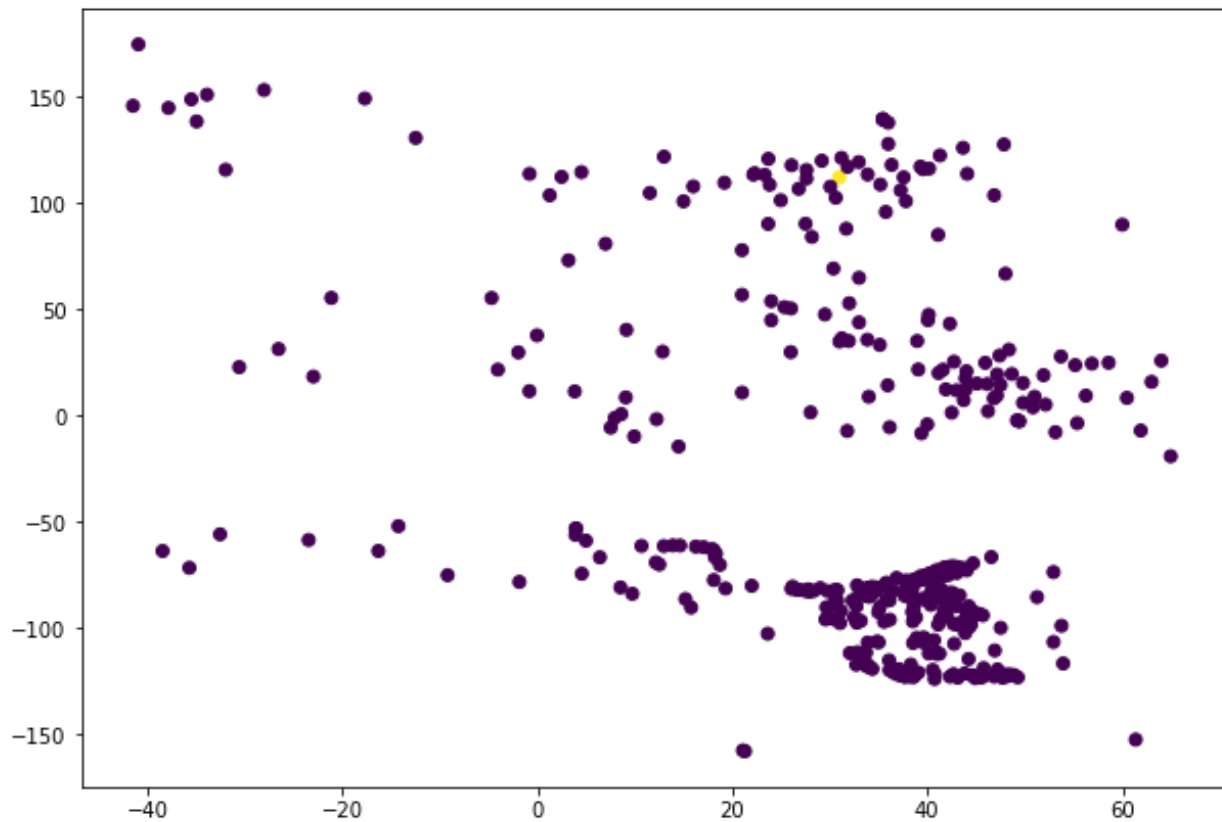
Fig 7: Scatter diagram of clusters of corona virus observations (confirmed cases).

This observation is similar to the following map: [4]

[4] Source: https://www.who.int/news-room/detail/13-03-2020-who-un-foundation-and-partners-launch-first-of-its-kind-covid-19-solidarity-response-fund

Fig 8:
Recovery Efforts and Effectiveness
From the above data analysis, it can be observed that corona virus has spread exponentially over the last two to three months. However, it also has been contained in a brilliant way. In China, the progress is significant. In Italy and Iran, the increasing infection and increasing death rate has been a matter of concern. However, all the countries have started taking emergency measures to contain further spread of the virus. Italy's lockdown is an example how cautions are being taken[5]. South Korea is an excellent example of appropriate measures which could contain the death to less than one percent of confirmed cases.

One interesting, but unfortunate observation is that the spread of the virus has taken place in tourist places like Italy. Also, celebrities have reported being affected by the virus which shows that it is not about normal practices of hygiene[6]. Public places like airports and stations provide a vulnerable context for the spread of the virus.

Basic hygienic practices like hand washing, use of mask have been suggested. The problem is that masks are not available in enough numbers. Exorbitant prices of masks have been a matter of concern.

One problem is that there is not clear information known to people. There are many rumors spreading across the world on how it could be dealt with. Many of them are fake news[7]. This needs to be contained.

5. Conclusion and Recommendations
The pandemic of corona virus has become a global threat. While the initial outbreak was observed in China, it has spread to distant places like Iran and Italy to the extent of a national disasters. It not properly contained; it can spread easily to the other parts of the world. In this context, the following measures may be taken:
- Producing and distributing enough test kits so as to identify the cases
- Taking appropriate policy measures to control gathering so as to contain further spread
- Providing appropriate information so that people know how to prevent the spread

As a result of the shutdown, there is a possibility that businesses would suffer from huge losses. The governments may provide appropriate relaxations like tax cuts to contain the situation.

---

[5] https://edition.cnn.com/2020/03/09/europe/coronavirus-italy-lockdown-intl/index.html
[6] https://nypost.com/2020/03/13/all-the-celebrities-athletes-and-public-figures-diagnosed-with-coronavirus/
[7] https://www.bbc.com/news/entertainment-arts-51858555