

## DATA MANAGEMENT PROJECT REPORT

(November-December 2022)

Covid-19 in India

Submitted by

Vishnudeep

Shukla

**12002534**

B.Tech (Computer Science) KM073

INT-217

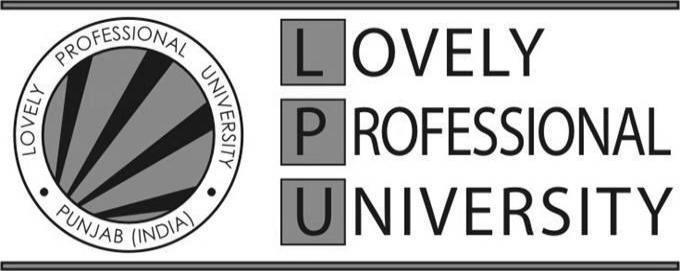
Under the Guidance of

## Maneet Kaur

## 15709

(Faculty)

## Discipline of CSE/IT

**Lovely School of Computer Science and Engineering Lovely Professional University, Phagwara**

**DECLARATION**

I, **Vishnudeep Shukla**, student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 15 November 2022

Registration No. **12002534** Name of the student: **Vishnudeep Shukla**

# ACKNOWLEDGEMENT

I would like to take this opportunity to show my sincere gratitude towards, Ms**. Maneet Kaur**, our subject teacher for course code INT217 as well as our university: Lovely Professional University whose constant help and guidance allowed me to complete this project in due time. I would also like to thank my classmates whose support was crucial in helping me finish this project.

# TABLE OF CONTENTS

|  |  |
| --- | --- |
| **Introduction** | **5-6** |
| **Objectives** | **7** |
| **Source of dataset** | **8-9** |
| **ETL process** | **10** |
| **Dashboard and Screenshot** | **11-13** |
| **Bibliography** | **14** |

**INTRODUCTION**

The outbreak of novel pathogenic Coronavirus (2019-nCoV) [[1](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r1)-[6](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r6)] was first identified in Wuhan city of Hubei Province of South China on 31st December, 2019. It was recognized as a Pandemic by the World Health Organization on 11th March, 2020. The Coronavirus disease has been officially named as COVID-19.

COVID-19

is an infectious disease caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2). World

Health Organization (WHO) had declared the outbreak a Public Health Emergency of International Concern

on 30th January, 2020. There are several reports regarding the timeline of the Coronavirus outbreak.

According to media reports on unpublished Chinese government data, the first case of COVID-19 emerged

on 17th November, 2019 [[7](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r7)-[10](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r10)]. According to a study by Chinese researchers published in the Lancet

medical journal, the first patient with the symptoms of COVID-19 was identified on 1st December, 2019

[[11](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r11)-[15](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r15)]. Like Middle East Respiratory Syndrome (MERS-CoV) [[16](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r16)-[20](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r20)] and Severe Acute Respiratory

Syndrome (SARS-CoV)

[[21](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r21)-[24](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r24)]. Coronavirus is a large family of viruses that causes illness ranging from the common cold to severe disease. COVID-19 is a disease that mainly affects the respiratory system and most infected people recover without requiring specific treatment. People above the age of 60 years and who have underlying medical conditions

are at higher risk. According to World Health Organization (WHO), the symptoms of COVID-19 include:

* Fever, tiredness, dry cough as the most common symptoms.
* Sore throat, diarrhea, pains and aches, conjunctivitis, loss of taste or smell, headache,
* discoloration of toes or fingers or rashes on skin as less common symptoms.
* Shortness of breath or difficulty breathing, loss of movement or speech, chest pain as some
* serious symptoms.

The Coronavirus disease first emerged in China in December 2019 and spread globally rapidly. When a

person with COVID-19 [[5](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r5)-[8](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r8)] exhales or coughs, small droplets spread out in the surrounding environment

and land on surfaces and objects around the person. Other people then catch COVID-19 directly by inhaling

these small droplets or indirectly by touching contaminated surfaces or objects then touching their nose,

mouth or eyes. The survival of Coronavirus depends on various factors such as humidity or temperature of

the environment and type of surface. The risk of catching COVID-19 from the faeces of an infected person

and someone with no symptoms at all appears to be very low. In India, the first case of coronavirus was

reported in Kerala on 30th January, 2020 [[25](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r25)]. As of 5th December, 2020, the recovery rate of COVID-19 in

India is 98.48%, a total of 9499710 cases and 140216 deaths have been confirmed by worldometer website

in India. The number of infections could be a substantial underestimate due to low testing rates of India

[[26](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r26), [27](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r27)]. Following are some protective measures against COVID-19, advised by the World Health

Organization (WHO) [[28](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r28)].

1. Maintain social distancing: you should maintain at least 1-meter distance from a person who is

sneezing or coughing.

1. Regular hand washing: one should thoroughly and regularly clean his/her hands with soap and water

or alcohol-based hand wash.

(3) Seek medical care early, if you have cough, fever and shortness of breath.

(4) Avoid unnecessary travel

(5) Maintain respiratory hygiene: you should not sneeze in the bare hand, always make sure that you cover

your nose and mouth with a tissue or with your bent elbow.

(6) Avoid touching nose, mouth and eyes: if you touch your eyes, mouth and nose with contaminated hands

, then the virus can enter your body.

On 28th March 2020, there were 25,778 numbers of government hospitals and 7.13 lakh number of beds in government hospitals and in government hospitals, the availability of beds was 0.55 beds per 1000

population [[28](https://opennursingjournal.com/VOLUME/15/PAGE/62/#r28)]. Therefore, India had taken several measures to stop coronavirus spread. Such measures

are as follows:

1. Lockdown the entire nation for 21 days from 24th March 2020 to 14th April 2020 to break Coronavirus
2. chain.

(2) India is trying to quickly ramp up its fragile healthcare facilities.

(3) Conversion of railway coaches and medical colleges into isolation wards, etc.

In this study, we tried to present an early prediction of the epidemic of COVID- 19 in India based on infection

rate and suspected cases and our mathematical model forecasts the minimum and the maximum number

of deaths, due to COVID-19 in India, for the different number of infected cases. Our results are supposed to

provide important information regarding the upcoming infected number of people, which will help manage COVID-19 crisis.

The contribution of this paper is structured as follows: Section 2 elaborates motivation for research. Section

3 depicts the methodology and section 4 discusses the result of the proposed model and section 5

concludes the research paper by summarizing our work.

# OBJECTIVE

The objectives are carefully selected to analyze the dataset to get useful information. The scope of each analysis is limited to the dataset to get accurate results for every objective. Following analysis are done on the dataset:

**Objective 1 –** Covid-19 Spreading in Asia, Europe Region

The scope of this analysis is to find the Covid-19 Spreading in every state with time. The analysis can be done by using power pivot.

**Objective 2 –** Total Confirmed Cases Reported in Asia, Europe Region

This analysis helps in finding Number of Confirmed Cases Reported from Countries to Countries from Jan 2020 to Aug 2021 helps in analyzing this objective.

**Objective 3 –** Total Deaths Reported due to Covid-19 in Asia, Europe Region

The scope of this analysis is to find out total Deaths Reported From Countries to Countries from Jan 2020 to Aug 2021.

**Objective 4 –** Total Cured From Covid-19 in Asia, Europe Region

This analysis helps to find out Total people Cured from Covid-19 , from Countries to Countries from Jan 2020 to Aug 2021.

**Objective 5 –** Total Comparison of covid outbreak in Asia, Europe Region

The scope of this analysis is to find the Relation between Confirmed ,deaths and Cured

Covid\_19 from Countries to Countries from Jan 2020 to Aug 2021.

# SOURCE OF DATASET

Kaggle is an Airbnb for Data Scientists –A subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore, and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start in 2010 by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. this is where they spend their nights and weekends. It’s a crowd-sourced platform to attract, nurture, train, and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries, and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars, then ultimately been acquired by the Google in March of

2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

Kaggle enables data scientists and other developers to engage in running machine learning contests, write and share code, and to host datasets. The types of data science problems posted on Kaggle can be anything from attempting to predict cancer occurrence by examining patient records to analyzing sentiment to evoke by movie reviews and how this affects audience reaction. Different sources post projects on this trailblazing platform. While some are just for educational purposes and fun brain exercises, others are genuine issues that companies are trying to solve. Kaggle makes the environment competitive by awarding prizes and rankings for winners and participants. The prizes are not only monetary but can also include attractive rewards such as jobs

or free products from the company hosting the competition.

## Why is Kaggle Worth Your Time?

1. Interesting and challenging projects where contributors can learn and practice Kaggle competitions involve solving challenging and interesting problems. Companies post projects to numerous contributors. It especially a great place for beginners who are just trying to break into the data science field. Aside from the competitions that are open to the general public, Kaggle also has private competitions which are only open to top rated participants (Kaggle Masters).
2. Insightful discussions with Industry leaders and learned experts apart from the projects, forums are very interesting, stimulating, and informative. Through these discussions, you can either seek advice from others or offer advice to people who are dealing with issues you understand
3. Kaggle offers its audience a chance to get into the biggest data science community in the world. This platform is trusted by some of the largest data science companies of the world such as Walmart, Facebook, and Winton Capital. On Kaggle, data scientists get exposure and a chance to work on problems faced by big companies in real-time. While it is not a guarantee, there is always the chance that the company will be impressed enough to recruit.

## How Kaggle Works

The host of the competition is in-charge of preparing the data and preparing a detailed description of the problem at hand. To make it more convenient for hosts, Kaggle offers an additional consulting service that can help prepare data and describe the problem in the best possible format.

## Do Kaggle Projects have any Real impact

One of its biggest and most recognized projects is one by Heritage Health which offered a remarkable cash price of $3 million. Competitions hosted on Kaggle have had far-reaching impacts such as enhancing and enabling state of the art HIV/AIDS research and improving traffic forecasting. Essentially, Kaggle has given companies the opportunity to seek solutions from the best data scientist in the world and to have external pairs of eyes to look at the problems they are trying to solve.

# ETL

ETL stands for Extract, Transform and Load and it is the process of extracting data from a data source, transforming it using specialized software and then loading the transformed data into a data warehouse.

The data of Indian General Election from 1977 to 2014, extracted from Kaggle, which is an open- source database website. The data is then transformed (cleaned) using Excel tools and Tableau then saved to the device. The dataset has 11 columns related to the election like, name of candidates and total vote polled etc. Following are the operations performed in Tableau to transform the data:

* First, fetch the dataset of Covid-19 India – 2020 to2021 from Kaggle.
* Analyze the dataset carefully to look for any ambiguity.
* Load the dataset into Microsoft Excel and perform the cleaning operation.
* Take the output of the cleaned dataset.

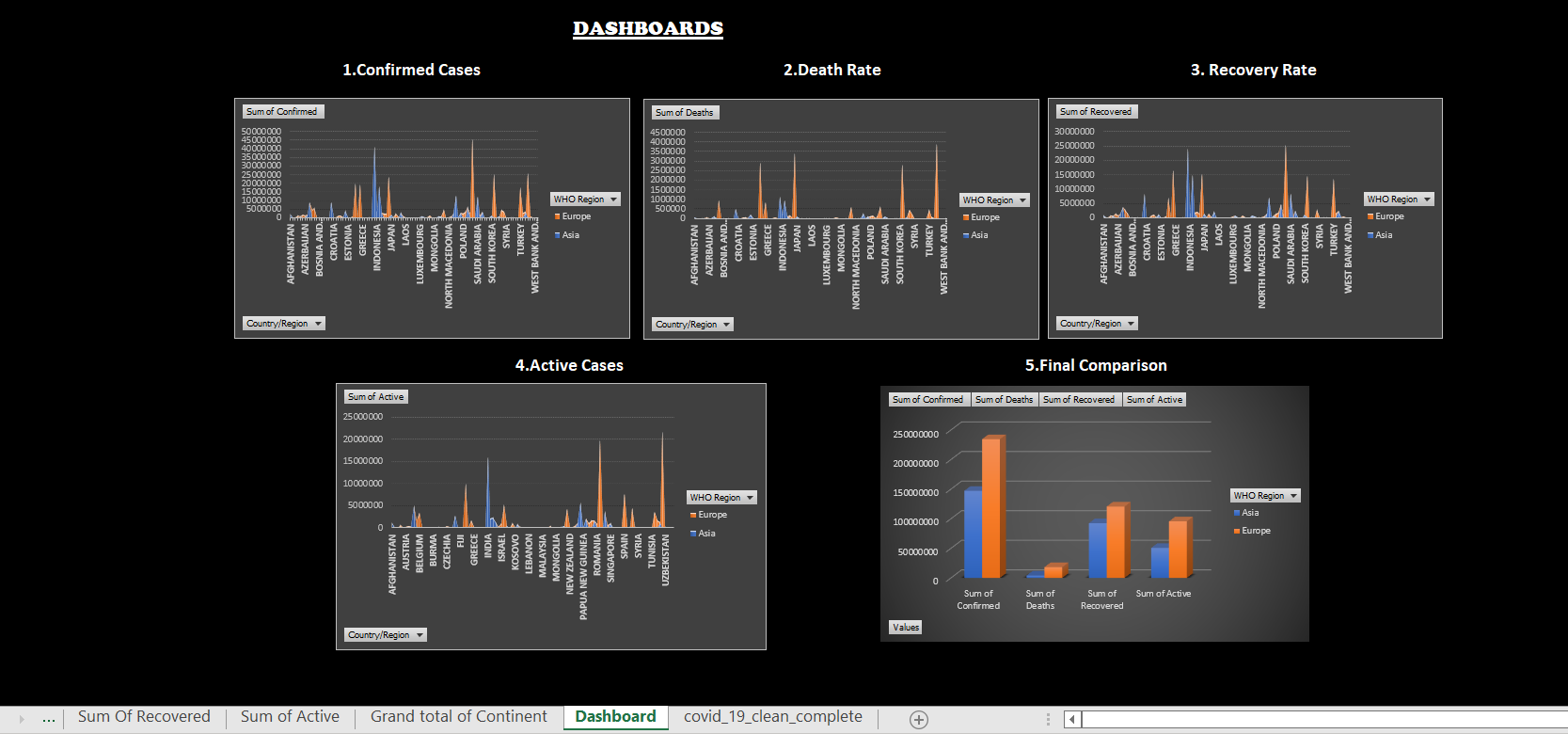
## Why Do We Need ETL?

It is essential to properly format and prepare data to load it in the data storage system of your choice. The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:

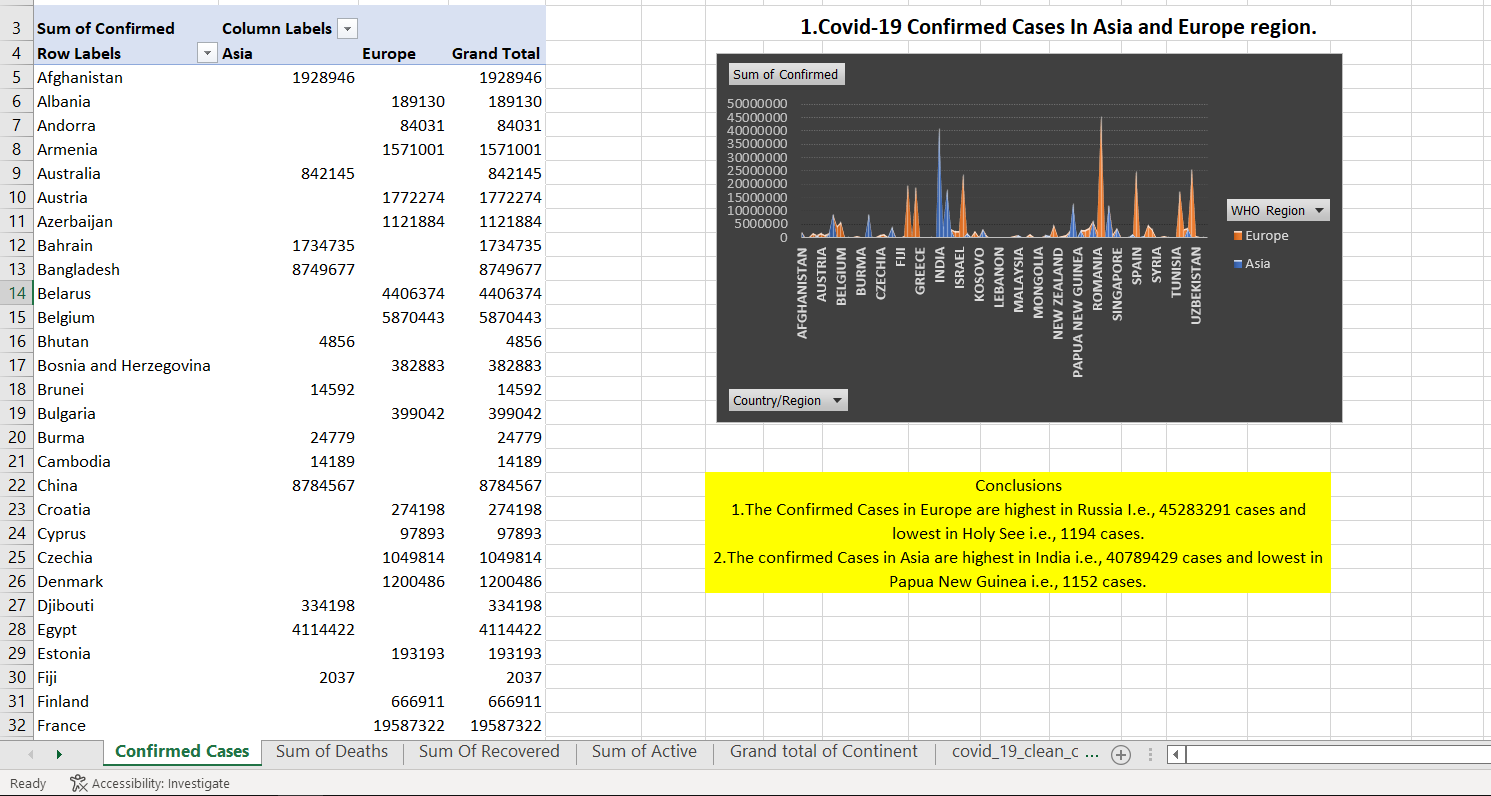
* Offers deep historical context for business.
* Enhances Business Intelligence solutions for decision making.
* Enables context and data aggregations so that business can generate higher revenue and/or save money.
* Enables a common data repository.
* Allows verification of data transformation, aggregation, and calculations rules.

**Screenshot**

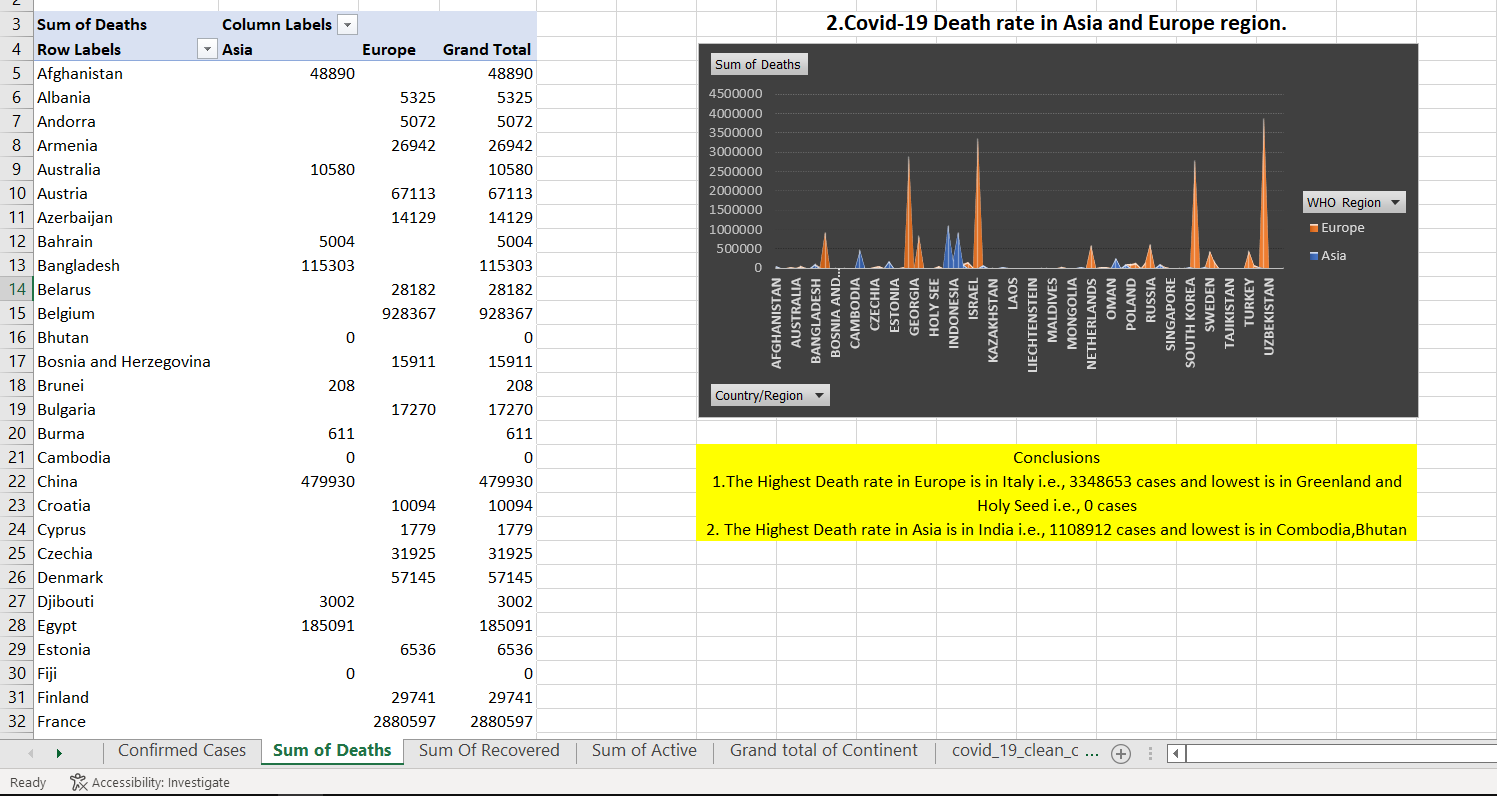
1. **Dashboard**

****

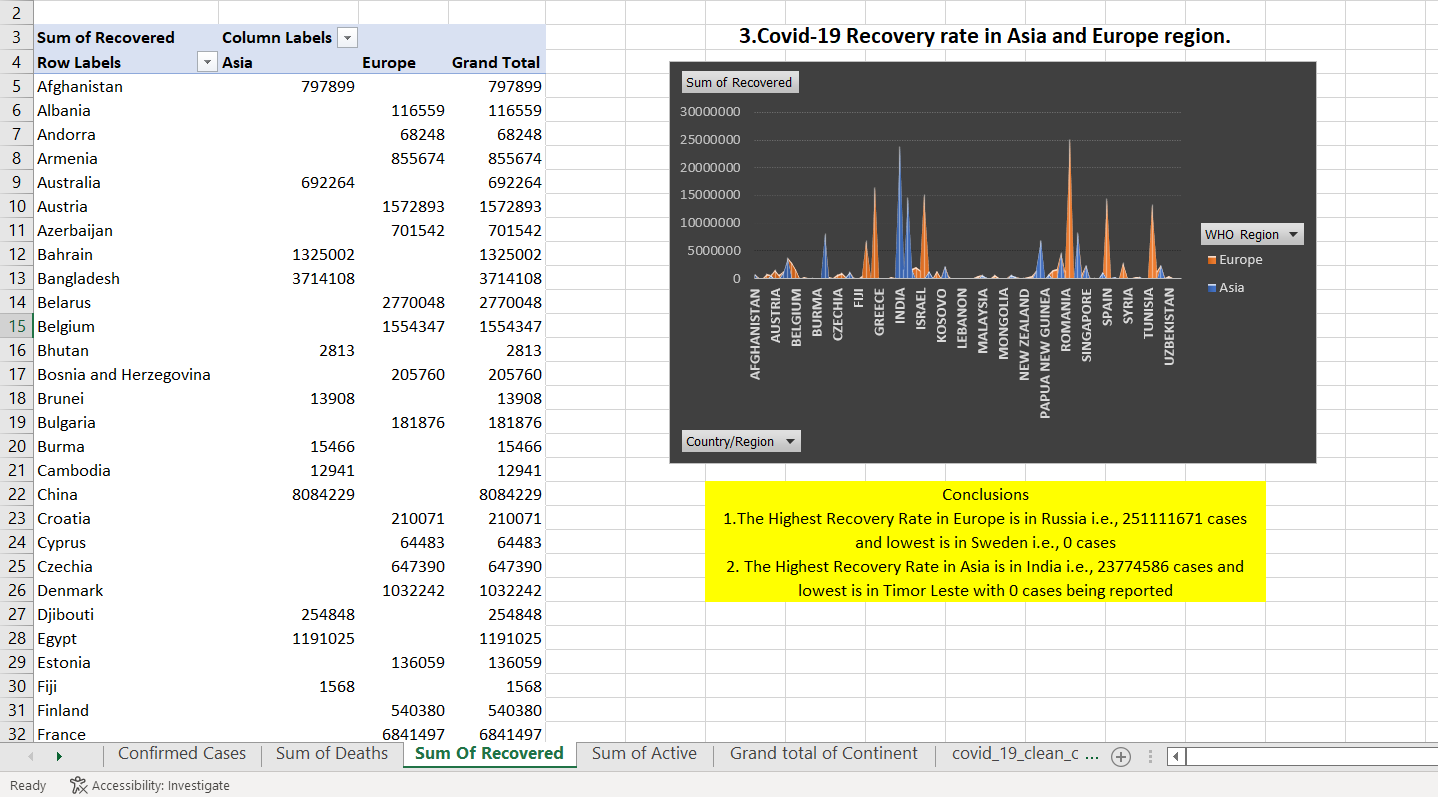
1. **Covid-19 Confirmed Cases In Asia and Europe region.**

****

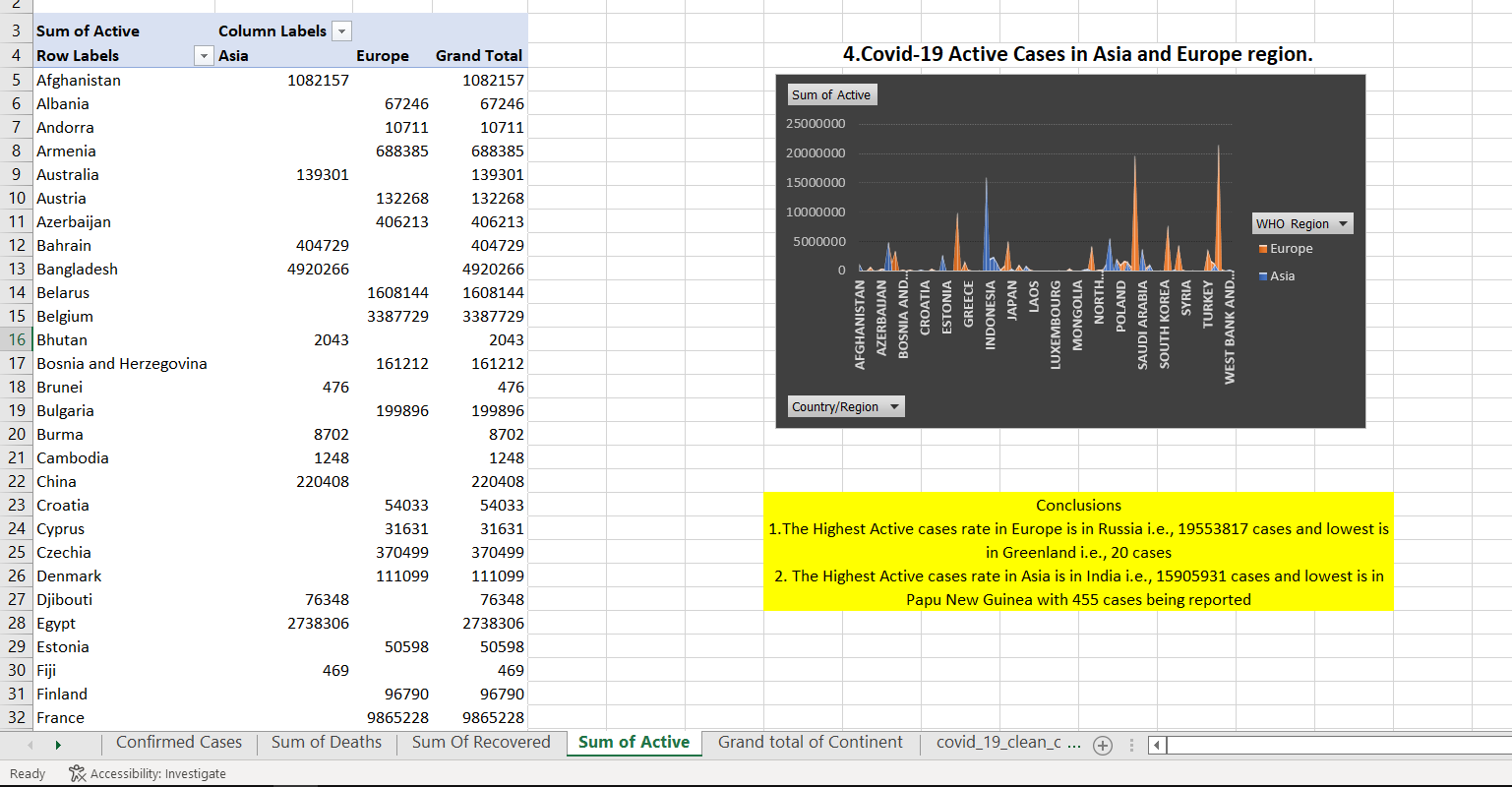
1. **Covid-19 Death rate in Asia and Europe region.**

****

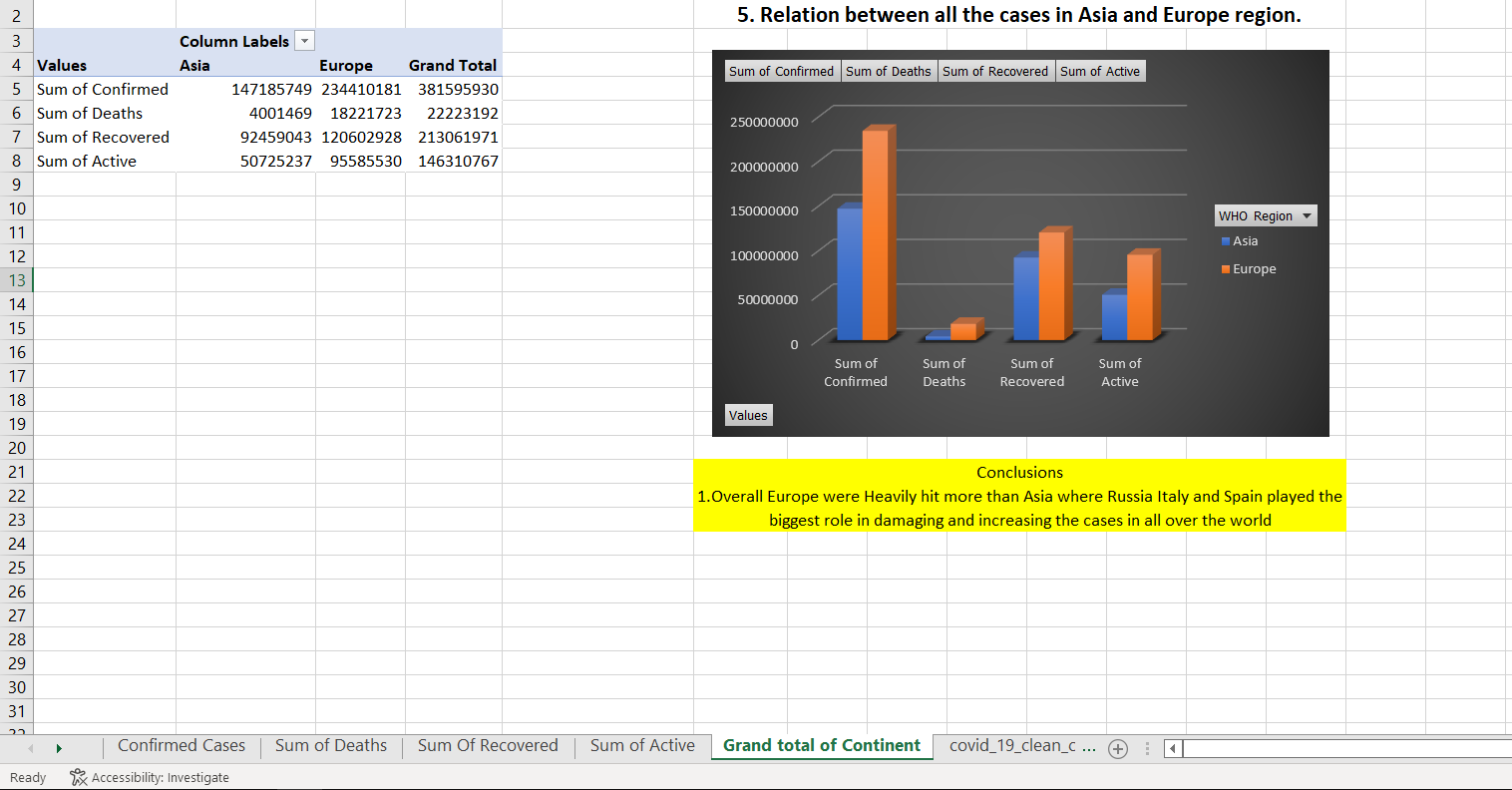
1. **Covid-19 Recovery rate in Asia and Europe region.**

****

1. **Covid-19 Active Cases in Asia and Europe region.**

****

**6.Relation between all the cases in Asia and Europe region.**



# BIBLIOGRAPHY

1) https://www.kaggle.com/datasets/imdevskp/corona-virus-dataset

2) 2020. [online] Available at: <htt[ps://www.quor](http://www.quora.com/What-is-Kaggle-and-)a[.com/What-is-Kaggle-and-](http://www.quora.com/What-is-Kaggle-and-) how-

exactly-should-I-use-it> [Accessed 14 December 2020].

3) Talend Real-Time Open-Source Data Integration Software. 2020. *What Is Extract,*

*Transform, Load? Definition, Process and Tools*. [online] Available at:

<htt[ps://www.talend.com/resourc](http://www.talend.com/resources/what-is-etl/)e[s/what-is-etl/>](http://www.talend.com/resources/what-is-etl/) [Accessed 14 December 2020].

**-------------------------------------------THANK YOU ----------------------------------------------**