**INTRODUCTION TO DATA MANAGEMENT  
 PROJECT REPORT**

(Project Semester August-December 2020)

***COVID 19 DATA ANALYSIS***

Submitted by

PABITRA KUMAR PANDA

Registration No: 11903542

Programme and Section: P192-ND & KM007

Course Code: INT 217

Under the Guidance of

**Sandeep Kaur : 23614 and Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**



# 

# **CERTIFICATE**

This is to certify that PABITRA KUMAR PANDA bearing Registration no. 11903542 has completed INT 217 project titled, **“Covid 19 Data Analysis”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor: Ms. Sandeep Kaur**

**Designation of the Supervisor: Assistant Professor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 10-12-2021

# **DECLARATION**

I, Pabitra Kumar Panda, student of Integrated B.Tech – M.Tech 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: 10-12-21 Signature:  Registration No. : 11903542 Name of the student: Pabitra Kr Panda

# **ACKNOWLEDGEMENT**

A project work is a combination of views, ideas, suggestions and contribution of many people. Thus, I feel obliged as well as pleasure to extend my gratitude towards all those contributors who helped me during my endeavor of making this project.

I would like to thank my Professor as well as supervisor Ms. Sandeep Kaur for providing me with this opportunity to make a minor project as well as implement my practical knowledge by making a dashboard on the project titled, **“Covid 19 Data Analysis”** by providing me constant support as well as constructive criticism which helped me to understand my mistakes and improve my project thoroughly.

I would like to extend my gratitude to all those as well who responded to my queriesin a well-defined manner and helped me acquiring knowledge without whom my project would not have been such a great learning experience.

Regards

PABITRA KUMAR PANDA

Section: KM007

Registration No: 11903542

# **TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| S.NO. | TITLE | PAGE NO. |
| 1. | [Certificate](#_CERTIFICATE_1) | 2 |
| 2. | [Declaration](#_DECLARATION) | 3 |
| 3. | [Acknowledgement](#_ACKNOWLEDGEMENT) | 4 |
| 4. | [Introduction](#_INTRODUCTION) | 6 |
| 5. | [Scope Of The Analysis](#_SCOPE_OF_THE) | 8 |
| 6. | [Source Of The Dataset](#_SOURCE_OF_THE) | 9 |
| 7. | [ETL Process](#_ETL_PROCESS) | 9 |
| 8. | [Objectives](#_OBJECTIVES) | 12 |
| 9. | [Analysis Of The Dataset](#_ANALYSIS_OF_THE) | 13 |
| 10. | [List Of The Analysis With Results](#_LIST_OF_ANALYSIS) | 21 |
| 11. | [References](#_REFERENCES) | 23 |
| 12. | [Bibliography](#_BIBLIOGRAPHY) | 24 |

# **INTRODUCTION**

1.1 *What is Excel?*

Microsoft Excel is a software program included in the Microsoft Office suite. It is used to create spreadsheets, which are documents in which data is laid out in rows and columns — like a big table.

Due to its extreme versatility and power, Excel has become one of the most-used software programs in the business world since its launch in 1985. Indeed, the personal computing renaissance of the 1980s and 1990s was largely driven by the many uses of Excel and other spreadsheet software.

1.2 *Some Important Terms*

* + 1. Dataset: A data set is an ordered collection of data. While [handling the data](https://byjus.com/maths/data-handling/), the data set can be a bunch of tables, schema and other objects. The data are essentially organized to a certain model that helps to process the needed information. The set of data is any permanently saved collection of information that usually contains either case-level, gathered data, or statistical guidance level data.
    2. Spreadsheet: A spreadsheet is a special way of organizing data into rows and columns to make it simpler to read and manipulate.  
       A default spreadsheet is often named as Sheet1, Sheet2, Sheet3 and so on.
    3. Column: A column is a vertical series of cells in a chart, table, or spreadsheet.
    4. Row: A row is a series of data banks laid out horizontally in a table or spreadsheet.
    5. Workbook: A workbook is a collection of one or more spreadsheets also called worksheets in a single file. A default workbook is often named as Book1, Book2, Book3 and son on.
    6. Formula Bar: The formula bar is a section in Microsoft Excel and other spreadsheet applications in which it allows the contents of the current cell and also allows us to create and view formulas.
    7. Pivot Table: Pivot table in excel is used to categorize, sort, filter, and summarize any length of data table which we want to get count, sum, values either in tabular form or in the form of 2 column sets.
    8. Pivot Chart: A pivot chart is especially useful for user when dealing with tremendous amounts of data by quickly reorganizing and visualizing data in an understandable manner and facilitate the entire process.
    9. Slicers: Slicers in Excel are software filters used along with excel tables or pivot tables over a large amount of data. Not just filtering out the data, but slicers also help you with an easy understanding of the information being extracted and displayed on the screen.
    10. Timeline: Timeline in Excel actually represents the time span from the start to end on a bar. So for this, we should have any time frame such as Dates, Months, Minutes, Hours, etc
    11. Hyperlinks: Hyperlinks in Excel allow users to create a shortcut way to reach any certain worksheet, file, folder or webpage. It helps us to reach to any specific folder or link quickly.
    12. Data Analytics: Data analytics is the science of analysing raw data to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.
    13. Data Visualization: Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.
    14. Dashboard: Excel dashboards make it easy to perform quick overviews of data reports rather than going through large volumes of data. Overviews help in making quick and urgent decisions since one can skim through a lot of information at once and within a short time.

# **SCOPE OF THE ANALYSIS**

2.1 *Project Description*

The report gives us an overview on the Covid 19 pandemic of the India over a period of 2years (2020-2021). For the country’s betterment, it is important to know where this virus spread the most and where maximum number of doses were taken. We have chosen this dataset to analyse and generate some insights on the covid. The data collected provides us with the basic information on the covid 19.

*Columns Present In The Dataset*

* Date
* Time
* State/UnionTerritory
* Confirmed Indian National
* Confirmed Foreign National
* Cured
* Deaths
* Confirmed
* Sum of total samples
* Sum of positive
* Sum of negative
* Total Doses Administered
* Sessions
* Sites
* First Dose Administered
* Second Dose Administered
* Male Dose Administered
* Female Dose Administered
* Transgender Dose Administered
* Covaxin
* CoviSheild
* Sputnik V
* Total Individuals Vaccinated
  1. *Scope*

In order to have a better insight it is important to analyze the data so that we can get ready for the future for different waves to come.

It is also important to have made strategies ready on the basis of the report in order to have a better performance.

# **SOURCE OF THE DATASET**

Source Of The Dataset: https://www.kaggle.com/sudalairajkumar/covid19-in-india

Author: SRK

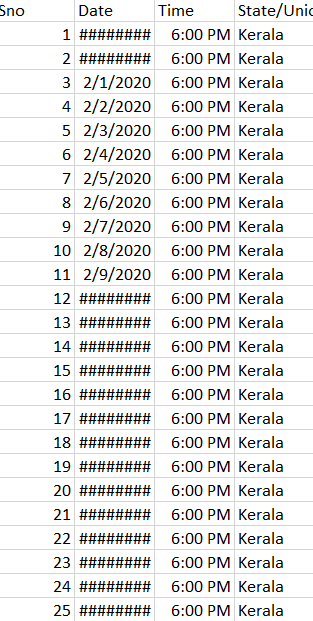
Last Updated: 4 months ago

# **ETL PROCESS**

 ETL is the process of extracting huge volumes of data from a variety of sources and formats and converting it to a single format before putting it into a database or destination file.

4.1 Extraction : Raw data is extracted from various sources like Kaggle, mydata.gov.in, etc. in the format of CSV files. So we are going to extract the file from the raw data and transform it into a target file and load it in the output.

4.2 Transformation: Dataset has been cleaned in order to make it into more readable format in excel and summarized into various tables according to our objectives in the form of pivot tables and charts.



🡨 Original CSV File (Raw Data)

**Table 4.1**



🡨 Formatted Excel File

After cleaning the data and making it into readable format

**Table 4.2**

4.3 Loading: The load stage of the ETL process depends largely on what you intend to do with the data once it’s loaded into the data warehouse. Uses could include:

* Layering a business intelligence or analytics tool on top of the warehouse.
* Analyzing and visualizing the data in the form of Dashboard which is our intended target in this project.
* Creating a tool for site search.
* Building a machine learning algorithm to detect fraud

# **OBJECTIVES**

Objective -1

1. Month Wise Analysis of Covid in India.

Objective – 2

1. Yearly Comparison Between Confirmed, Cure & Death.

Objective -3

1. State Wise Analysis for Positive and Negative Cases.

Objective -4

1. Top 5 most vaccinated State.

Objective – 5

1. Total No. of Doses Administered.

# **ANALYSIS OF THE DATASET**

***6.1. Covid Based On Months***

1. Introduction: The analysis shows state wise spread of covid.
2. Specific Requirements/Functions and Formulas:

* Pivot table of the dataset.
* GETPIVOTDATA function.
* Pivot Chart.

1. Analysis Results: We can easily analyse with the help of this Line graph that in between Jun & Jul had the highest no. of cases.
2. Visualization: Blue line color represents the growth of covid.

**Fig 6.1**

* 1. ***Covid in year***

1. Introduction: This analysis shows sum of confirmed, sum of cured & sum of death in both year.
2. Specific Requirements/Functions and Formulas:

* Pivot table of the dataset.
* Pivot Bar Chart.

1. Analysis Results: We can analyze from this chart that 2021 has maximum no.
2. Visualization:

**Fig 6.2**

* 1. ***Maximum Tested State***

1. Introduction: This analysis shows sum of samples taken in a state.
2. Specific Requirements/Functions and Formulas:.

* Stack chart.

1. Analysis Results: We can easily analyse through this chart that which state has tested the most.
2. Visualization:

**Fig 6.3**

* 1. ***Which company vaccine was used the most***

1. Introduction: This analysis show that which state has taken the maximum of which vaccine.
2. Specific Requirements/Functions and Formulas:

* Pivot table of the dataset.
* Bar chart.

1. Analysis Results: We can analyze from this chart that covishield is the most preferred vaccine.
2. Visualization:

**Fig 6.4**

* 1. ***Top 5 State has taken maximum of vaccine***

1. Introduction: This analysis represents which state has taken more doses.
2. Specific Requirements/Functions and Formulas:

* Pivot table of the dataset.
* Area chart.

1. Analysis Results: We can easily analyze from this chart that himachal and J&K were in the top
2. Visualization:

# **LIST OF ANALYSIS WITH RESULTS**

In this section we can see that we are able to answer so many questions due to this analysis.

* *In which month covid was the most ?*  
  Ans: Oct 2020
* *In which year covid was the most?*   
  Ans: 2021.
* *Which Region had highest covid testing?*  
  Ans: Maharastra
* *Which vaccine was taken the most?*And: Covidshield
* *Which was the state taken maximum vaccine?*Ans: Maharastra

# **REFERENCES**

* [www.google.com](http://www.google.com)
* [www.kaggle.com](http://www.kaggle.com)
* [www.youtube.com](http://www.youtube.com)
* [www.smartsheet.com](http://www.smartsheet.com)
* [www.github.com](http://www.github.com)
* [www.support.microsoft.com](http://www.support.microsoft.com)

# **BIBLIOGRAPHY**

1. Microsoft Excel 2016 Bible: The Comprehensive Tutorial Resource by John Walkenbach, Wiley
2. Fundamentals of Business Analytics by R.N. Prasad, Seema Acharya, Wiley