COVID DATASET

A Project Report submitted in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by

MEDISETTY VENKATA NIKHIL	**121910312016**
MAREMALLA SANDEEP KUMAR	**121910312023**
SAGI LAKSHMI PASYANTHI	**121910312035**
VIVEK CHADARAM	**121910312046**

Under the esteemed guidance of

Dr. S. VENKATA LAKSHMI
Assistant Professor, GST

DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING

GITAM

(Deemed to be University)

VISAKHAPATNAM

October 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

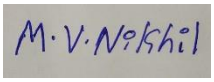
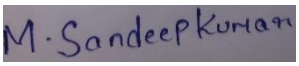
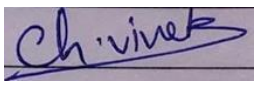
GITAM

(Deemed to be University)

DECLARATION

I/We, hereby declare that the project report entitled "**POST COVID UPI EXPENDITURE**" is an original work done in the Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University) submitted in partial fulfilment of the requirements for the award of the degree of B.Tech. in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree or diploma.

Date:28-10-22

Registration No(s).	**Name(s)**	**Signature(s)**
121910312016	MEDISETTY VENKATA NIKHIL	
121910312023	MAREMALLA SANDEEP KUMAR	
121910312035	SAGI LAKSHMI PASYANTHI	
121910312046	VIVEK CHADARAM	

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM

(Deemed to be University)

BONAFIDE CERTIFICATE

This is to certify that the project report entitled "**POST COVID UPI EXPENDITURE**" is a Bonafide record of work carried out **by MEDISETTY VENKATA NIKHIL (121910312016), MAREMALLA SANDEEP KUMAR (121910312023), SAGI LAKSHMI PASYANTHI (121910312035), VIVEK CHADARAM (121910312046)** students submitted in partial fulfilment of requirements for the award of the degree of Bachelors of Technology in Computer Science and Engineering.

PROJECT GUIDE	HEAD OF THE DEPARTMENT
Dr. S.Venkata Lakshmi	**Dr. R.Sireesha**
(Assistant Professor)	**(Professor, HOD)**
CSE, GST	**CSE, GST**
GITAM	**GITAM**

ACKNOWLEDGMENT

We would like to thank our project guide **Dr. S.VENKATA LAKSHMI**, Assistant Professor, Department of CSE for her stimulating guidance and profuse assistance. We shall always cherish our association for her guidance, encouragement and valuable suggestions throughout the progress of this work. We consider it a great privilege to work under her guidance and constant support.

We also express our gratitude to the project reviewers **Mrs K.NAGA SOUJANYA KETHA**, Assistant Professor and **Prof.**, Associate Professor, Department of CSE, GITAM (Deemed to be University) for their valuable suggestions and guidance in the completion of our project.

We consider it a privilege to express our deepest gratitude to **Dr R.SIREESHA**, Head of the Department of Computer Science and engineering for her valuable suggestions and constant motivation that greatly helped us to complete this project.

Our sincere thanks to **Prof.Ch. VIJAY SHEKAR**, Principal, GITAM Institute of Technology, GITAM (Deemed to be University) for inspiring us to learn new technologies and tools.

Finally, we deem it a great pleasure to thank one and all that helped us directly and indirectly throughout this project.

MEDISETTY VENKATA NIKHIL **121910312016**

MAREMALLA SANDEEP KUMAR **121910312023**

SAGI LAKSHMI PASYANTHI **121910312035**

VIVEK CHADARAM **121910312046**

TABLE OF CONTENTS

ABSTRACT

Since Covid-19's emergence in India, most states were affected dreadfully. The reasons for wide spread of this infectious virus had many reasons and factors which varied from state-to-state. We used data extracted automatically from daily health bulletins published by state governments to a) breakdown the behaviour of covid waves in each state and b) analyse the reasons for surges in covid cases c) to analyse and develop real life machine learning models for prediction. We learned that major factors for surge in cases were elections, festivals, restrictions, relaxations set by the state governments and poor initial vaccination drives. But in later phases less fatality rates were achieved by huge vaccinations drives and herd immunity.

INTRODUCTION

A) PROBLEM DEFINITION:

The COVID-19 India Dataset is one of the most comprehensive datasets on the pandemic in India. It aggregates data from health bulletins published online daily by governments of major Indian states. The main idea of this research is to come up with models for analysis, prediction, and insights on the evolution of the pandemic in India.

B) OBJECTIVES:

The main objectives of the analysis are listed below: -

1. To Identify Cause and Effect of different covid-waves.
    a. To identify different phases in the dataset.
    b. To discuss cause and effects in brief.
    c. Conclusion.
2. To identify the merits of vaccine and lockdown to reduce hospitalization and fatality rate.
    a. To justify the advantages of vaccine on general terms.
    b. Merits of Lockdown.
    c. Conclusion.
3. To identify relation, association to develop machine learning models to predict future values.
    a. To identify best attributes to predict cumulative positive cases and cumulative deaths.
    b. To identify and choose best models for prediction.
    c. To identify the factors which affects cumulative cases and deaths among different states.
    d. Conclusion.

PROBLEM IDENTIFICATION AND OBJECTIVES

The whole world was affected with a pandemic Covid-19. It impacted the ways of the world very deeply. India also subdued a lot damage and changes. The whole period of 2020-2022 was very chaotic. There was an outrageous surge in the amount of positive covid cases and deaths, sadly.

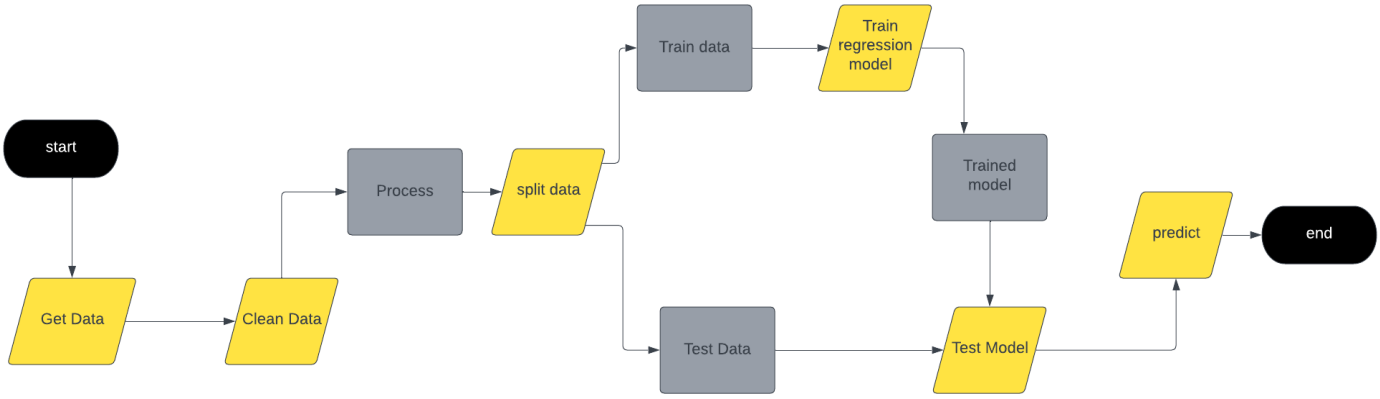
Over those 2 years, we accumulated a lot of data, which when studied properly will give a lot of useful insights. Analyzing those 2 years of pandemic can help us get ready for any such disasters in the future.

The main idea of this research is to study, analyze and predict the trends of covid, positive cases, and deaths in Telangana and Tamil Nadu.

Following are the objectives of this research:

1. To Identify Cause and Effect of different covid-waves.
2. To identify the merits of vaccine and lockdown to reduce hospitalization and fatality rate.
3. To identify relation, association to develop machine learning models to predict future values.

RESEARCH METHODOLOGY



- The study is based on primary sources of data/information. Indian central Govt. and state Govt. covid websites have been consulted to make the study an effective one.
- This fetched data from various sources is cleaned first, ie. duplicate or irrelevant observations are removed. Unwanted observations from the dataset, including duplicate observations or irrelevant observations, are also removed. Along with that filtering unwanted outliers and handling missing data are also done.
- Next, the data is split. Data visualization is done, Data visualization is a way to represent information graphically, highlighting patterns and trends in data and helping the reader to achieve quick insights. ie. The analysis of data is done using pie charts, bar graphs and scatterplots on the data.
- Then the observation of the analysis takes place and then Multivariate linear regression is performed on the observations. Linear regression helps create models to make predictions, such as predicting the cumulative no. of positive cases and deaths

OVERVIEW OF TECHNOLOGIES

Various technologies are used for achieving results in this research work i.e. Data must be sorted, structured, and visually displayed in a way that makes sense because it is difficult to grasp and make sense of data in its raw form. Data visualization is useful in situations like this. The practice of drawing insights from data through the use of graphs, sparklines, infographics, heat maps, or other visual representations is termed as data visualization. This makes it easier to understand, and use data. Tools for transforming data into pictorial representations are known as data visualization tools. Jupyter Notebook is the one utilized here to code as it is easier to write and maintain code for data visualization and machine learning. One of the most popular web-based applications for data visualization, JupyteR, allow users to produce and share documents with visualizations, equations, and live code. Statistical modelling, numerical simulation, interactive computing, and machine learning are all excellent uses for JupyteR. Data visualizations using scatterplots, bar charts, pie charts, histograms, etc. are the most popular. The popular library that is used for visualization is Matplotlib. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib is a popular Python 2-D charting library known for its incredibly high level of customization. With the help of matplotlib, interactive 2D graphs may be created, including line, scatter, and bar graphs. The language used is Python. A high-level, all-purpose programming language is Python. Code readability is prioritized in its design philosophy, which makes heavy use of indentation. Using dots to depict the link between variables, scatter plots are used to observe the relationship between them. To create a scatter plot, use the matplotlib library's scatter() method. A

circular statistical graphic called a pie chart can only show one series of data at a time. The overall 8 percentage of the provided data is represented by the chart's area. The percentage of the data pieces is represented by the area of the pie slices. A bar plot, often known as a bar chart, is a graph that uses rectangular bars with lengths and heights that are proportional to the values to represent a category of data. Both horizontal and vertical graphs of the bars are possible. The comparisons between the distinct categories are shown in a bar chart. One of the simplest and most widely used Machine Learning techniques is linear regression. It is a statistical technique for performing predictive analysis. For real, numerical, or continuous data, linear regression makes predictions. It executes a regression operation.

IMPLEMENTATION

CONCLUSION AND FUTURE SCOPE

Festivals, elections, mass gathering were some of the major reasons contributed to massive widespread virus from April-2021 to June-2021 in most states. But it was also seen that on the same reasons even some states like Kerala witnessed increase in new cases till December-2021.

Poor medical infrastructure and less vaccination added to the misery of common people.

1. In Tamil Nadu number of covid cases decreased from July-2020 to December-2020 as seen in **Fig-12**. One primary reason the government could achieve that because the lockdown was implemented strictly till December-2020.

To see full sized image, Click Here

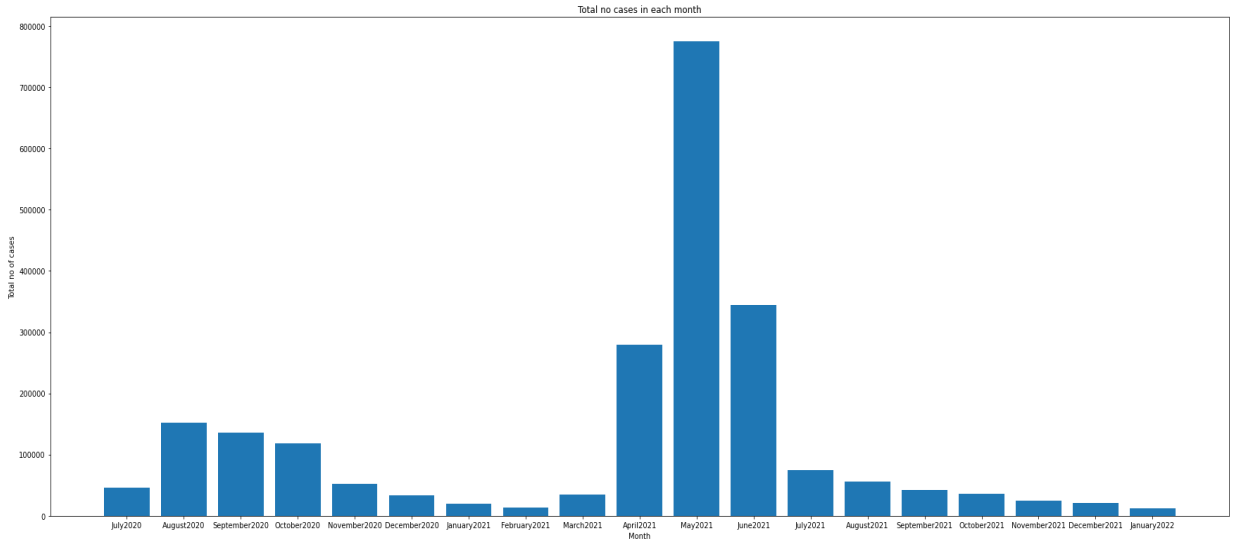


Fig-12: The graph plots total no of positives cases per month in Tamil Nadu.

Vaccination indeed worked as an antidote to reduce hospitalisation and thus fatality rate in many states. Strict implementation of lockdown also helped in Tamil Nadu to control the positive cases for much longer period whereas in West-Bengal as soon as lockdown was lifted, we can see rise in positive cases.

To see full sized image, **Click Here**

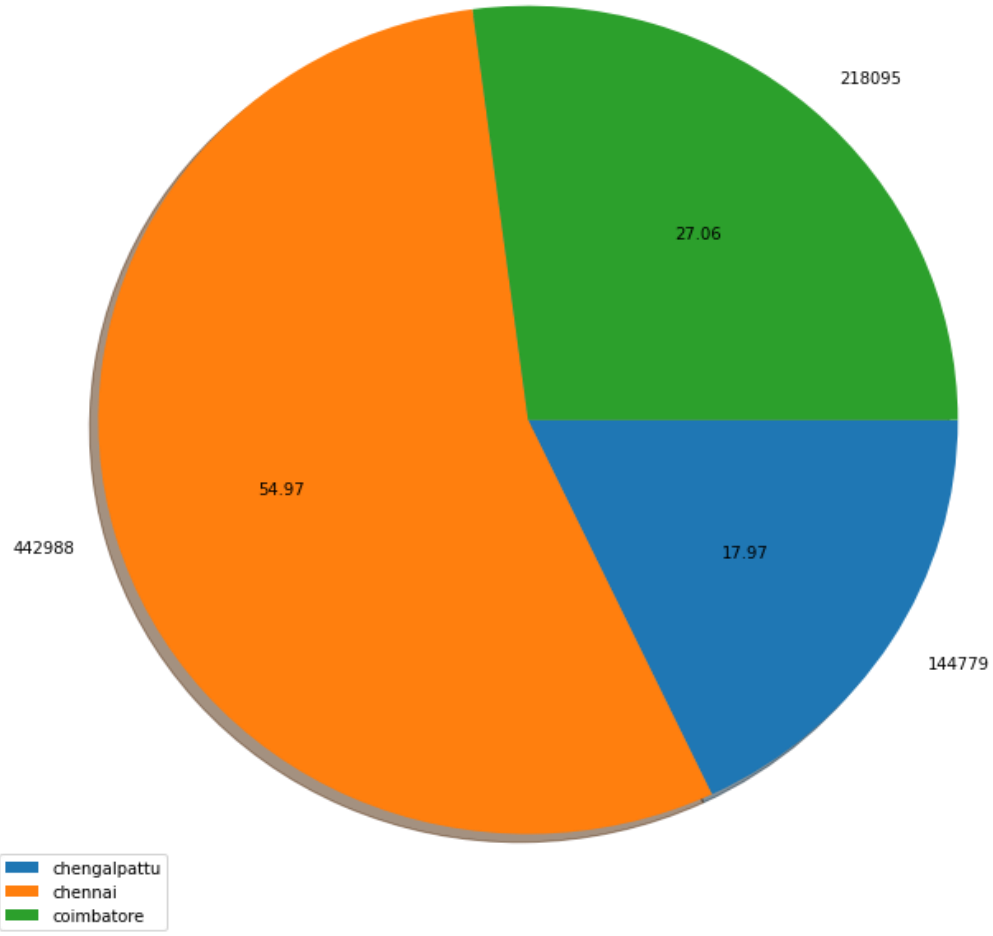


Fig-15: The pie chart **shows major cities contribution to total cases** in Tamil Nadu.

Objective-2: Trend in Fatalities and different comorbidities involved:

In Tamil Nadu it was found that death rate is very high in patients of age 40 above compared to deaths in patients below 40 age group. Again, in patients of age 40 above huge number of deaths can be seen in patients with comorbidities than without patients without comorbidities. (**Check table 5.2.1**)

Table 5.2.1 shows **no. of deaths in 40+ age group with vs without comorbidities.**

Fatalities category	**Number of deaths**
With comorbidity	Nearly 8000
Without comorbidity	Nearly 19000

Table 5.2.1

- **To view bar graph** Click Here

To see full sized image, Click Here

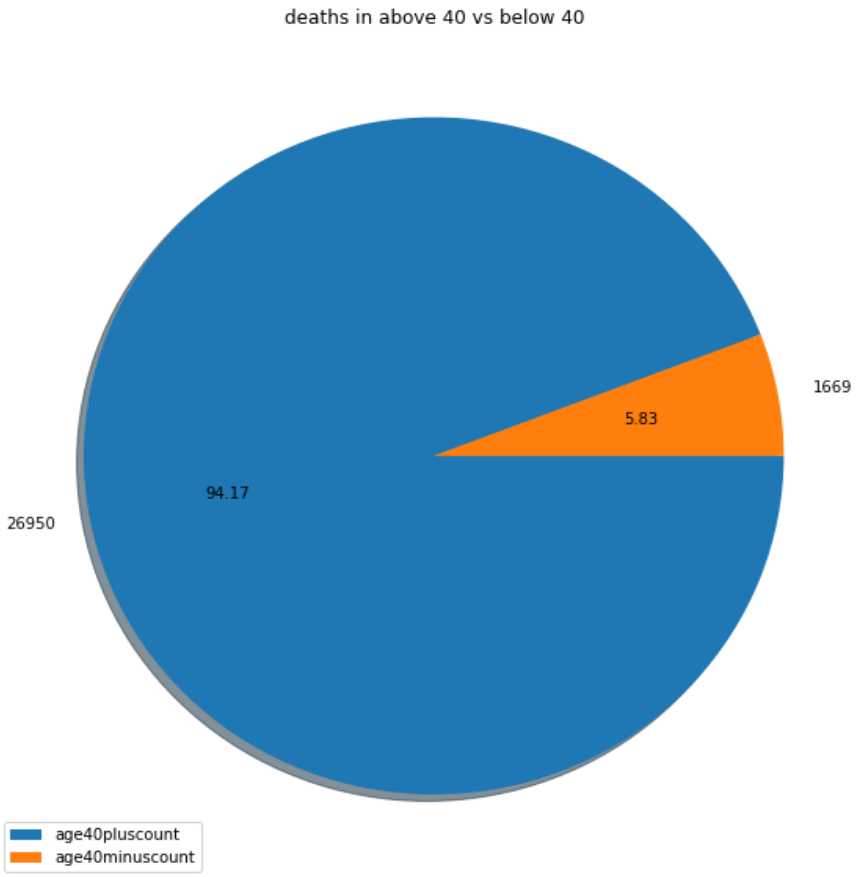


Fig-16: The pie chart depicts **the no. of deaths in 40+ age group vs 40+ age group (with comorbidity)** in Tamil Nadu**.**

FINAL CONCLUSION

1. Festivals, mass gathering in election rallies, not following strict lockdown in states were some of the major reasons for high positive case, hospitalisation and fatality.
2. Poor medical infrastructure and poor rate of vaccination, in the country has caused a lot of discomfort to people and to the medical fraternity in this pandemic.
3. Vaccination played a major role to reduce hospitalisation and fatality.
4. It was identified that all perfect models truly displayed the actual factors behind “total cases” and “total death”. At the same time practical model gave us an idea on how to develop machine learning model which are feasible in real world.
5. We also identified that cities and district with big population were major contributors in positive cases and fatalities in their respective states.
6. With available data in some states, it was found that fatalities with comorbidities were more than fatalities without comorbidities.
7. With available data in some states, fatality was more in the age group 45 and above.

REFERENCES

1. Data sourced from - https://india-covid-19-data.mybluemix.net/