Production Project

by Aditya shah

Submission date: 12-Jul-2021 01:02PM (UTC+0545)
Submission ID: 1615597100
File name: Final_Report_77202324.docx (12.97M)
Word count: 13098
Character count: 70680

LEEDS BECKETT UNIVERSITY

The British College
KATHMANDU

The British College

Coursework Submission Coversheet
(individual coursework only)

Faculty of Arts, Environment and Technology

LBU Student Id: 77202324

FOR CHECKING BY THE STUDENT:

Please ensure all information is complete and correct and attach this form securely to the front of your work before posting it in a coursework collection box.

Award name: BSc (Hons) Computing

Module code: 14806

Module name: Production Project

Module run: 2021

Coursework title: Covid-19 Data analysis and prediction using machine learning

Due Date: 2021-07-12

Module leader: (In LBU) Patrick Ingham

Module Supervisor: (In TBC) Mr. Shiva Prasad Nepal

TURNITIN Checked: (YES) NO *(please circle)*

Submission date & time: Date: 2021-07-12 Time: 11:59 PM

Total Word Count: 9224 **Total Number of Pages (including this front sheet):** 113

In submitting this form with your assignment, you make the following declaration:

I declare, that the coursework submitted is my own work and has not (either in whole or part) been submitted towards the award of any other qualification either at LBU or elsewhere. I have fully attributed/referenced all sources of information used during the completion of my assignment, and I am aware that failure to do so constitutes an assessment offence.

Signed:  Date: 2021-07-12

FOR COMPLETION BY THE FACULTY:

Teacher's Signature: _____ **Date:** _____

2021

Covid-19 Data Analysis and Prediction Using Machine Learning

STUDENT ID: C7202324
SUBMITTED BY: ADITYA SHAH
PRODUCTION PROJECT

Abstract

The Covid-19 virus has grown highly causing a global pandemic affecting the country's economic sector, people's health as well as businesses. However, to resolve problems such as poor Covid-19 decision-making, a lack of information about Covid-19 cases, and the inability to visualize and understand data, this report proposes a platform that will allow people to understand and visualize Covid-19 cases globally, make predictions using machine learning, and perform data visualization for effective business decision making. For the SDLC, the agile methodology was used to achieve the end product's goal and vision. As a result of product evaluation, it allowed a user to predict Covid-19 cases, visualize live worldwide Covid-19 data, and perform visualization on the user's data. This paper demonstrates the importance of data analysis in extracting useful information from raw Covid-19 data, the usage of business intelligence tools for data visualization, and the use of machine learning for Covid-19 prediction using ML algorithms, all these elements combined to build a potent product.

Declaration

I, therefore, certify that the project report “COVID-19 DATA ANALYSIS AND PREDICTION USING MACHINE LEARNING” is based solely on my completed work during my studies under the supervision of Mr. Shiva Prasad Nepal. I claim that my research is the result of the remarks and conclusions I made.

I. I also affirm that the paper is original and has been completed by me under my supervisor's general supervision.
II. The paper has not been published at any other university in Nepal or other parts of the country.
III. While composing the report, I adhered to the criteria as determined by the university.

IV. Wherever I applied resources (data, theoretical analysis, and text) from several places, credit was given in the report's text and provided details about them in the references.

Aditya Shah

ID: 77202324

BSc. (Hons) Computing

Acknowledgment

I'd want to convey my heartfelt admiration for the excellent and constant direction and supervision of my supervisor Mr. Shiva Prasad Nepal whose competency, patience, and understanding have amplified my experience in project capstones. Knowledge of such initiatives is precious in the workplace and represents a big advantage in my career.

It is also an opportunity to appreciate my family which encourages me to leverage my experience with love and encouragement. Without their excellent support, this would not have been achievable. I'd also want to acknowledge all my fellow students and colleagues who always supported me all the way.

Contents

List of Figures

List of Tables

1 Chapter-1. Introduction

1.1 Introduction

The goal of the project was to analyze the Covid-19 dataset as well as perform visualization and prediction to provide meaningful insight and additionally, create an analytical report, dashboard, and web app for interactive user experience. In today's trends and scenarios, the Covid-19 disease has become extremely dangerous, resulting in a global pandemic. It has hampered the country's economic sector as well as people's businesses. Many people are unaware of machine learning and cases of Covid-19 all over the world, and proper decision-making is still a long way off. By not utilizing the full potential of machine learning people are struggling to predict the Covid-19 cases to protect themselves and face future problems. With the increase in the size of data and information, it is becoming harder to summarize and visualize the relationship between variables in the dataset to extract useful information and identify patterns. There are numerous factors causing damage as a result of the Covid-19, but people in higher positions are still unable to provide better health care to residents of certain areas and are lacking in the management of various sectors. Citizens are losing lives because of less understanding of coronavirus, bad decision-making, and a poor management system of circumstances. Similarly, many business stakeholders are at a loss and therefore unable to conduct their business effectively due to a lack of understanding of future attacks, poor business intelligence, and a lack of research into Covid-19 data.

To analyze the problem and produce a product that meets the aim and objectives of the project, firstly exploratory data analysis is performed in Google's collaborative IDE using the Python programming language on the Covid-19 dataset. So there will be a pdf document summarizing data analysis and prediction model in detail. Similarly, there are business intelligence tools like

Power BI, Tableau, and many more but for the cost-free and effective option Covid-19 dashboard is created in Google Data Studio for interactive visualization, which will cover the major KPIs of Covid-19 cases. Likewise, instead of limiting the product only to Covid-19 data, there will be a data visualization web app built in the Python Streamlit library that will allow users to predict the Covid-19 cases, visualize live Covid-19 data and also visualize their dataset and generate an overview report of their dataset using the pandas-profiling library. The linear regression model and polynomial features are used for the prediction of covid-19 data. Other libraries like Matplotlib, NumPy, Pandas, Sklearn, etc. will be covered more in-depth in the theoretical framework section of the report.

Data plays a crucial role in the data analysis and prediction part of the project. Data are categorized into quantitative and qualitative forms. So to collect a quantitative time-series Covid-19 dataset which was reliable and accurate the two primary open-source GitHub repositories were chosen. One is the GitHub repository of John Hopkins University, the original link is (https://github.com/CSSEGISandData/COVID-19.git) where global data of Covid-19 is maintained and updated daily which will be used for data analysis and visualization purposes. Similarly, another is laxmimerit repository (https://github.com/laxmimerit/Covid-19-Preprocessed-Dataset.git) while the original data source of this repository is also the John Hopkins University repository which helped to get preprocessed covid-19 data through which a Covid-19 dashboard is developed and also with the assist of this data live Covid-19 data can is fetched into Data Visualization web app. And also after preprocessing these datasets desired data was obtained to apply for machine learning model to get an accurate outcome.

Ultimately, this research will provide meaningful information and resolve real-world questions about covid-19 data by performing data analysis and prediction using machine learning. Similarly, it will give an idea to users for visualizing and interacting with live Covid-19 cases all around the world. This research will also encourage higher authorities and business stakeholders in proper decision making and a good management system using machine learning and business intelligence techniques. And also provide brief information regarding how users can upload and visualize their dataset and get a summary report of their dataset and also predict the covid-19 future outcome. Finally, it will portray the importance of data analysis, business intelligence, and machine learning in this new era of technology.

The first section of the report begins with an introduction that includes the aim and problem description, as well as the formal research and concluding. The second section, following the introduction, is the Review of Literature, which describes the research and findings with the existing system and the new planned system. The third section is a review of technology, which involves a study on tools and technologies that are utilized and rejected throughout product development. The fourth section of the report is Methodology and Design, which discusses the methodology and design approach utilized in the product's creation. Methodology and design is the fourth section of the report which explains the methodology and design approach used in the development of the product. The fifth section of the report is Implementation and testing, which discusses how to implement and test the product to verify its quality. Product evaluation and project evaluation are the sixth and seventh parts, which describe the overall success of the product and project, respectively. The eighth section is the summary and conclusion, which includes a description of all activities undertaken throughout the project, discovered limits of the finished product, and future work or recommendations to upgrade the product.

2 Chapter-2. Review of Literature

2.1 Data analysis and visualization

For many individuals, data analysis is a relatively new word. In essence, it relates to all procedures and instruments needed for data processing and interpretation. Analytics is a more general phrase that incorporates the different data analysis tools and procedures. Also as the function of technology multiples in each area, it produces enormous quantities of data that can provide useful insights into the industry. In the last 10 years, this has led to a boom in the data sector. However, to get decision-making insights, data gathering must be followed by its analysis. Data analytics helps companies and industries understand the huge amounts of subsequent growth and development information. The difference between successful and unsuccessful firms in the current and the coming years is investing in analytics (TechnologyHQ, 2021). With so much of the information acquired in the industry – through data analysis, we need a technique to construct a picture of this information so that we can analyze it. Visualization of data provides us an overview of what information means by providing visual contexts via maps or graphs. This makes the data more familiar to the human mind and consequently facilitates the identification of trends, patterns, and outliers in vast volumes of data (Analytiks, 2020).

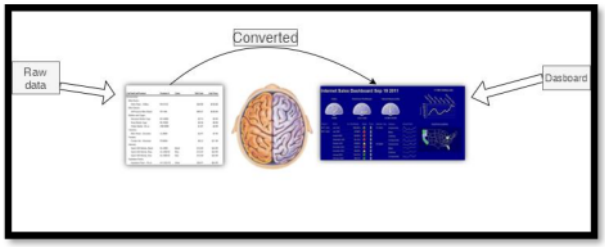


Figure 1. Raw data and dashboard

Figure 1 illustrates how the raw data looks and when it is converted to something like a dashboard it starts to please visually and gives meaning and useful information.

2.2 Business Intelligence

According to (Vercellis, 2010, p. 5) Business intelligence systems are primarily designed to offer knowledge workers tools and techniques to make efficient and timely choices. If verdict can depend on a business intelligence system to help customers, the general quality of the decision-making approach will be substantially improved. It's indeed feasible to examine a significant proportion of potential initatives, obtain furthur precise findings, and make convenient and relevant judgments with the aid of mathematical models and algorithms. Therefore, it can be concluded that the enhanced efficacy of the decision-making process is a significant benefit of using a system of business intelligence.

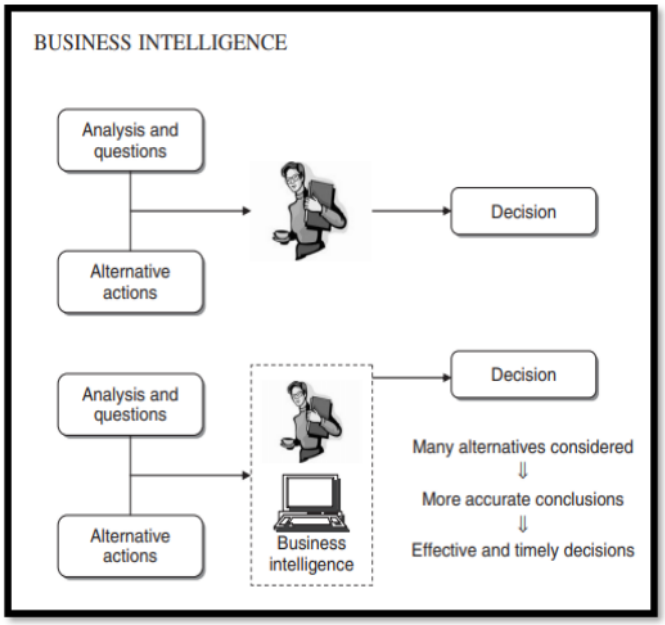


Figure 2. Benefits of Business Intelligence

Figure 2 shows the difference between the uses of business intelligence in decision-making. It represents with the business intelligence how we can consider many alternatives and get more accurate conclusions which help in effective and timely decisions.

2.3 Machine Learning and Algorithms

As per (Burkov, 2019) Machine learning may also be described as obtaining a dataset and constructing a statistical model on that dataset as the process of solving a practice issue. However, if we look in broader aspect (NetApp) says Machine learning (ML), a component of artificial intelligence (AI), is a branch of computational science that involves studying and interpreting data shapes to enable comprehension, reasoning, and decision making without the involvement of the people. Simply said, machine learning enables the user to provide an enormous amount of data to a computer algorithm and instruct the computer to assess and draw data-driven conclusions based only on the supplied data. If any adjustments are found, the algorithm can use that knowledge to make better decisions in the future. (Guido, 2016) Tells machine learning algorithms are at the heart of many modern websites and gadgets, from automated suggestions of which movies to watch, to what meals to order or which items to purchase, to personalized online radio and identifying your pals in your photographs. Each section of the site probably has many machine-learning models when you look at a complicated website like Facebook, Amazon, or Netflix.

Talking about machine learning techniques (Brownlee, 2016, p. 16) claims that In the great majority of current machine learning applications, supervised learning is employed. Supervised learning is a strategy where we utilize input (X) and output variables (Y) and the algorithm to learn how to map from the input to the output. Y = f(X) The purpose is to analyze the mapping function well enough that we can determine the output variables (Y) given fresh input data (X). Because the technique may be compared to a teacher overseeing the learning process, the procedures of an algorithm learning from a training dataset are referred to as supervised learning. Speaking of the machine learning model (Brownlee, 2016, p. 35) says that Linear regression is a linear model, that posits a linear connection between the input variables (x) and a single output variable (y). That is, y may be determined using a linear combination of the input

variables (x). When only one input variable (x) is present, the approach is known as simple linear regression. When there are many input variables, the approach is commonly referred to as multiple linear regression in the statistical literature. The most popular approach for preparing or training the linear regression equation from data is termed Ordinary Least Squares. Each input value or column is assigned one scale factor, known as a coefficient, which is typically symbolized by the Greek letter Beta (β). Yet another coefficient is added, which provides the curve an additional degree of freedom (e.g., moving vertically and horizontally on a two-dimensional graph) and is commonly known as the intercept or bias coefficient. In a basic regression case (a single x and a single y), for example, the model would be:

$$y = B0 + B1 \times x$$

2.4 Coronavirus (Covid-19)

As research is done on Coronavirus (Singhal, 2020) says with the development and spread of the 2019 novel coronavirus (2019-nCoV) or the severe acute respiratory syndrome coronavirus 2, the globe is facing a new public health crisis (SARS-CoV-2). The virus was transmitted to humans in December 2019 in Wuhan, Hubei Province, China, via an unknown intermediate species. To far (05/03/2020), there have been about 96,000 recorded cases of coronavirus disease in 2019 (COVID-2019) and 3300 recorded deaths. Inhalation or contact with contaminated droplets transmits the disease, and the incubation period spans from 2 to 14 days. Among the symptoms include fever, cough, sore throat, shortness of breath, tiredness, and malaise. Most people have a moderate case of the disease; but, in certain cases (typically the elderly and those with comorbidities), it can develop to pneumonia, acute respiratory distress syndrome (ARDS), and multi-organ failure. Many people have no symptoms. The case fatality rate is expected to be between 2 and 3%. The virus is detected in respiratory secretions using specific molecular testing. Normal/low white cell counts with high C-reactive protein are common test results (CRP). Even in people with no symptoms or moderate illness, the computed tomographic chest scan is frequently abnormal. The treatment is mostly supportive; the function of antiviral medicines has yet to be determined. Home isolation of suspected cases and those with mild illnesses, as well as stringent infection control measures in hospitals, including contact and droplet precautions, are all part of the prevention

strategy. Although less deadly, the virus spreads faster than its two predecessors, SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). The global implications of this new pandemic are yet unknown.

2.5 Existing Research and Proposed system

In the case of the use of machine learning algorithm accuracy in Covid-19 data (Muhammad, 2021) express that COVID-19, also known as 2019-nCoV, is no longer pandemic, but rather endemic, with over 651,247 individuals worldwide dying as a result of the disease. There is presently no specific treatment or cure for COVID-19, living with the condition and its symptoms is unavoidable. This fact has placed huge pressure on the world's inadequate healthcare systems, particularly in underdeveloped countries. These approaches might help 2019-nCoV pandemic patients with diagnosis and prognosis. For positive and negative COVID-19 cases in Mexico, this study used an epidemiology labeled dataset to develop supervised machine learning models for COVID-19 infection using learning algorithms such as logistic regression, decision tree, support vector machine, naive Bayes, and artificial neural network. Before building the models, a correlation-coefficient study was conducted for the strength of the relationship between each dependent and independent data-set feature to determine the correlation-binding feature strength. 80% of the training dataset was utilized to train the models, while the remaining 20% percent was used to test the models. According to the results of the model performance evaluation, the decision tree model acquired the greatest precision of 94.99 %, the sensitivity of the Support Vector Machine Model was 93.34%, and the specificity of the Naive Bayes Model was 94.30%.

But according to Corona Tracker Research Group (Binti Hamzah FA et al., 2020), it is said that COVID-19 remains unclear, so after an outbreak, only precise SEIR prediction can be achieved. The spread of the outbreak is heavily influenced by each country's regulation and civic responsibility. In the method, part of the report is claimed that relevant data queries are performed and visualized on their official site and requested data is then used for predictive modeling with sensitivities-exposed-infectious-recovered (SEIR). Based on daily observations, they used SEIR modeling to predict COVID-19 outbreaks both in and out of China. They also examined the queried news and classify the news into negative and positive feelings to understand both politically and economically the impact of the news on human behavior.

(Benjamin D Wissel et al., 2020) Claims that their team created a mechanism for aggregating Covid-19 data at the county level into metropolitan areas and shown on a real-time customizable dashboard. The website's objective was to make this information more accessible to the general public while also allowing for a more comprehensive evaluation of virus transmission and effect. Every hour, the COVID-19 app checks the NYT and COVID Tracking Project data and updates the watch application. Data are immediately transferred to the server and incorporated into the webpage when data updates are made accessible. A quality control review was done to ensure the intended size and format of updated data files. The COVID-19 Watcher dashboard is accessible over the internet (https://covid19watcher.research.cchmc.org/). The database covers all counties in the United States, as well as 188 metropolitan regions with a population of over 277 million people (83.3 percent of the population).

The COVID-19 pandemic has emphasized themes such as reliability or interactive posting of data for real-time event monitoring, two of today's research strongholds in which most models provided in various scientific papers are not widely obtainable, understandable, or reproducible. Proper evaluation of

incidents, such as the current pandemic affected by COVID-19, and also visualization of the negative effects of its spread, has highlighted the importance of bringing together forces in fields that have been well in collaborating, such as medicine, biology, and information technology. The dashboard has been written in R and uses the Shiny framework to build an attractive visualization tool. COVID-19 Spain continuously produces daily updates on cases, deaths, recovered, ICU hospitalizations, and aggregated daily incidence from official sources (Carlos III Research Institute and Ministry of Health, Consumer Affairs, and Welfare). Furthermore, It displays the progression of active, new, and accumulated cases by the autonomous community on a georeferenced map, allowing users to journey in time from the beginning to the end, allowing them to observe disease transmission and serve as a visual aid for epidemiological investigations (Fernandez-Lozano, 2020).

(Hector Florez, 2020) Declares that disease is known as coronavirus disease 2019 is accountable for the 2019-2020 worldwide pandemic (COVID-19). The Severe Acute Respiratory Syndrome coronavirus-2 is to blame for this illness (SARS-CoV-2). The World Health Organization recorded 3,096,626 cases and 217,896 fatalities by April 30, 2020, implying an exponential increase in infection and mortality globally. There are now several computer-based techniques that show COVID-19 data through various types of charts, which is highly beneficial for recognizing its behavior and patterns. However, such techniques do not allow for the observation of any projections on confirmed cases and fatalities, which would be important in understanding COVID-19 trends. In this project, they created and built an online dashboard that displays real-time information on COVID-19. Furthermore, based on this data, they also developed a statistical model to forecast the progression of cases and fatalities worldwide and by nation.

However, the above-mentioned existing research and methods are confined to data from a single country. All of the systems are not fully dynamic and lack a variety of functions. Prediction algorithms are not implemented in web apps anywhere in the research done in the above systems. All of these systems, in addition, do not allow users to visualize their data. So, to provide all of these functionalities and a better user experience in a single product and system, an analytical report of Covid-19 was created to better understand the Covid-19. Likewise, the Covid-19 dashboard was built that includes the main KPIs of Covid-19 cases and allows users to see data based on user input to provide a more dynamic rather than static user experience. Similarly, to avoid limiting the product to Covid-19 data, the web app was developed with three navigation menus: one for deploying a machine learning model where the user can predict future Covid-19 cases, another for visualizing live Covid-19 data, and a third for letting the user visualize their dataset and create a summary. I anticipate that by using this overall structure, users will have a sufficient grasp of Covid-19, data visualization, and Covid-19 prediction. Likewise, to predict covid-19 data Nonlinear Regression model with polynomial features was chosen in comparison to linear regression because the nature of data was nonlinear and to best fit the model and curve path nonlinear regression model was best to predict the data as Nonlinear regression links two variables in a nonlinear (curved) connection, whereas a straight line ($y = mx + b$) connects two variables (X and Y) in simple linear regression.

3 Chapter-3. Review of Technology

3.1 Software and Libraries

3.1.1 Google colaboratory

Colaboratory, or rather "Colab," helps to write and execute Python in your web, with free GPU power, convenient sharing, and no setups involved. Colab can make your job simpler, no matter if you are a student, a data scientist, or an AI researcher. Colab notebooks allow you, along with pictures, HTML, LaTeX, and more, to mix executable and rich text into a single document. They are kept in your Google Drive account when you create your own Colab notebooks. You may simply share your Colab notebooks with colleagues or friends and allow them to remark or update your notebooks. You may use Colab to fully analyze and view data from popular Python libraries (Google Colaboratory, n.d.).

3.1.2 Pycharm

PyCharm is a blended architecture that JetBrains created as Python's IDE. It is often used in the creation of Python applications. Twitter, Facebook, Amazon, and Pinterest are among the firms that use PyCharm as their Python IDE!. PyCharm is a Python IDE featuring a range of tools from Python. The IDE also provides expert Django web development abilities. The sophisticated and versatile editor supports faster and quicker syntax highlighting, snippets, folding and divided displays. Python is quite demanded nowadays. It is frequently used in the creation of software. For this, there are several 'n' causes. We need IDE to code python, An IDE is made up of an editor and a compiler for the compilation of programs. It has a mix of characteristics necessary for software development. The existence of an IDE makes it much easier to create and program. It understands what users type and proposes to add the appropriate keyword. We can distinguish between a class and a method as the IDE assigns various colors. The IDE also provides the proper and the wrong keywords in various colors (Intellipat, 2020).

3.1.3 Google Data Studio

Data Studio is a free application that makes your data visualizations and reports interesting, simple to read, easy to distribute, and configurable. You can quickly report data from a large range of sources using Data Studio without programming. It is easy to tell people, teams, or the globe about your thoughts. Invite people or give them links in scheduled emails to read or amend your reports. You can include your reports on other sites, including Google Sites, blog posts, marketing pieces, and yearly reports, to convey your data stories as broadly as possible. Most of the capabilities of Data Studio are straightforward to use, making it easy to share and schedule reporting. We utilize Data Studio to track customer critical KPIs, display trends, and compare overtime performance (Data Studio Help, n.d.).

3.1.4 Streamlit

(KARANJAGOTA, 2021) Data science today is one of the trendiest Google search topics. Increasing numbers of developers within the community are building new frameworks and libraries to support data scientists or researchers in their regular activities because of their high demand and extensive use in real-world applications. One of them is streamlit. You can develop data applications in no time using streamlit. It integrated smoothly with other python library systems, such as NumPy, Pandas, and Matplotlib. Streamlit is a python library open-source which is useful for creating and sharing web applications. In the data science community, it is gradually gaining traction. Because a data science web app is easy to create, many developers utilize it in their everyday workflow. There are more than 14.1k stars and 1.2k forks in the streamlined repository GitHub. Under the cap, React utilizes it to render the data on the screen as a front-end framework. Through minimal modifications in the code, React developers can simply alter the UI.

3.1.5 Numpy

NumPy is a simple version of Numerical Python that is utilized specially for the numbers for scientific programming in Python. It includes multi-dimensional objects in arrays and the Python implementation integration tools package. It is essentially a blend of C and Python used for MATLAB programming, where data are handled as numbers as arrays for multi-dimensional functions and rearrangement operations (Pedamkar, 2020).

3.1.6 Plotly express

The new Python visualization package Plotly Express is a wrapper for Plotly.py which shows an uncomplicated syntax for complex charts. It was specially intended to have a smooth, consistent, and easy-to-understand API, inspired by Seaborn and Ggplot2: with one import you can produce interactive charts, including faceting, maps, animations, and trend lines, in a single function. The Plotly Express is fully free, exactly as Plotly.py: with a permissible MIT open-source license, you may use it as you like (even for the commercial goods!) This product includes onboard data sets, color scales, and themes (Plotly, 2019).

3.1.7 Pandas

The most widely used open-source library for data analysis and machine teaching is pandas. It is designed to handle multi-dimensional arrays in addition to a Numpy module. This is one of the most often used data pooling packages in the Python environment and generally included in every Python distribution, from your operations to distributions in the commercial market such as ActiveState's ActivePython (ActiveState, n.d.).

3.1.8 Scikit-learn

Scikit-learn is without question the most essential package for machine learning in Python. The Sklearn package offers several efficient machines and statistical

modeling methods like classification, regression, clustering, and reduction of dimensionality.

3.1.9 Pandas- profiling

The Python Pandas profiling package has a ProfileReport() function, which generates a fundamental report for the DataFrame input. The report is DataFrame overview, DataFrames's correlation between characteristics, each attribute for which DataFrame is defined, and DataFrame samples (deepanshu_rustagi, 2020).

3.2 Similar Product Review

3.2.1 JHU Covid-19 Dashboard

Figure 3. Covid-19 Dashboard

Figure 3 represents the Covid-19 dashboard built and handled by the Center for System Science and Engineering (CSSE) at Johns Hopkins University (JHU). Ensheng Dong, a Johns Hopkins University doctorate student, has built the world's famous visualization panel to evaluate the 2019 pandemic of the coronavirus illness (COVID-19). This Covid-19 dashboard was built in the ArcGIS application. Basically, the ArcGIS application has been written in python and maintained by the US business Esri as a geographical information system (GIS).

The overall application looks clean and the User interface is also good enough. The only complaint I have is that it does not offer much of a user interaction and user input is also not available. The red dots in the map in the middle section of the app does not provide any information it could have been better if we hover over the red dot and it displays information like country name, cases, etc. Similarly, It can be improved if the user had the accessibility to give input and receive required data from the application. The red colors accent and numbers view in the application can be minimal to give a minimal and brighter appearance.

3.2.2 Kathmandu Post Covid-19 dashboard

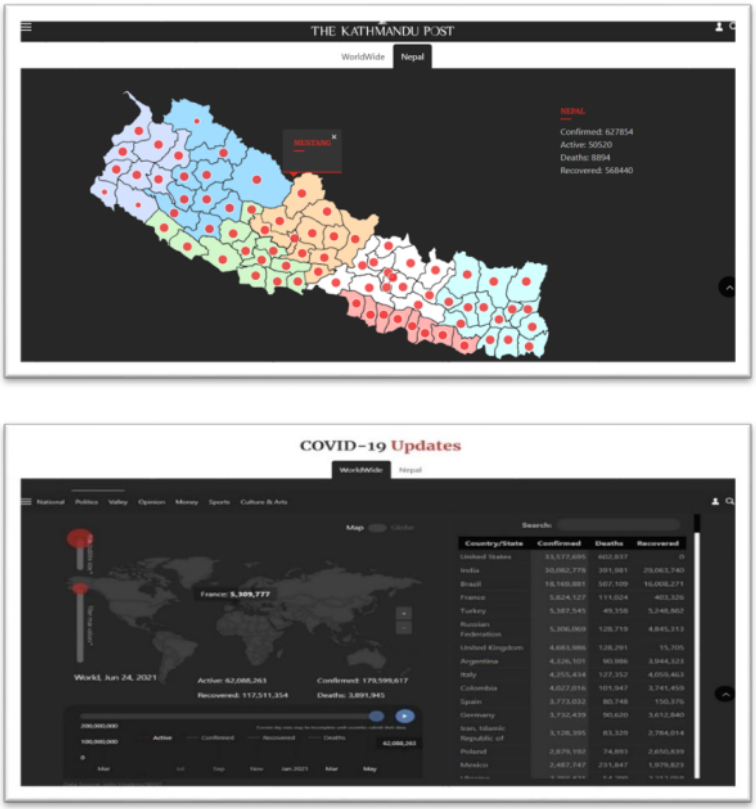


Figure 4. Kathmandu Post Covid-19 web page

Figure 4 shows the web page of The Kathmandu Post news portal and it also contains covid-19 updates. Two sections on the page are Nepal and worldwide. In the Nepal section, it is beautifully shown the number of Covid-19 cases on the right side while big Nepal's map with the red dot representing the presence of Covid-19 in a particular district. If we hover over the map then it displays the name of the district. Similarly, in the case of a worldwide section, it shows the map of the whole world and on the right side the all the country and their cases. There is also a search button to search for specific country data at top of the table. Down below the map, there is a timeline available so whenever the play button is clicked it starts to show Covid-19 cases from the beginning of covid-19 cases to till now. This web page is built-in Html, CSS, and JS by The Kathmandu Post. The UI/UX is clean which makes the user stick to the application.

However, in the Nepal map area, it would have been preferable if when we hover over data, it displayed the district name as well as particular district cases rather than just the district name. It just shows the overall number of cases rather than particular district-level cases, therefore it could be enhanced by include district-level cases. In addition, the color range can be used to show the number of instances in each area, making it aesthetically appealing. The worldwide part might be enhanced if the map was more colorful and visually informative.

3.2.3 Interactive Covid-19 Forecast and Correlation Explorer

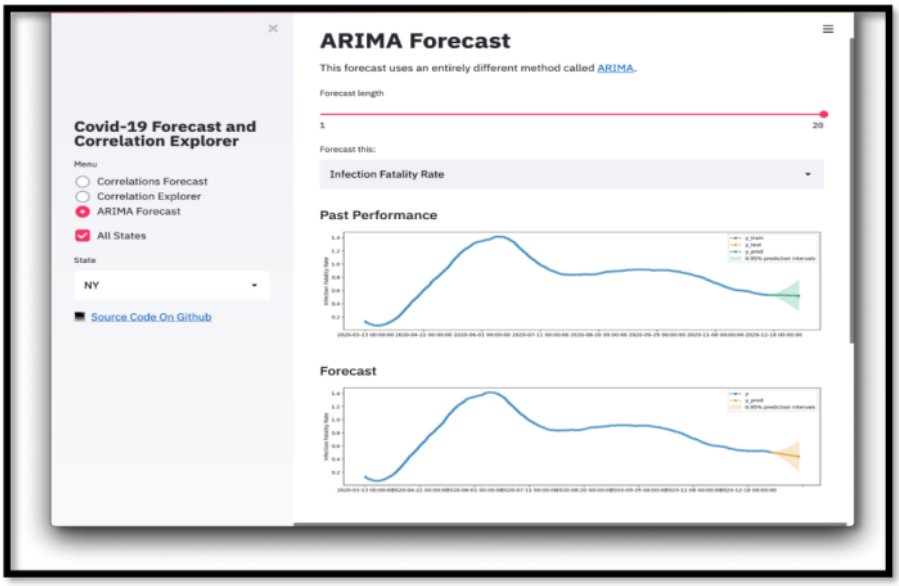


Figure 5. Covid-19 ARIMA forecast and Correlation Explorer

Michael Remington created this Interactive Covid-19 Forecast and Correlation Explorer (see figure 5). This program is written in Python using the Streamlit module. This software is divided into three sections, each with its own set of features. When it comes to the ARIMA Forecaster section, there is a slider called Forecast length that allows users to choose the number of days to be forecasted. Similarly, a dropdown box allows you to pick Covid-19 cases or another characteristic for which data would be predicted. There is a Past performance component that shows a chart of previous data and anticipated data, and we can see the actual data route and expanded forecasted path in yellow in the Forecast part.

This overall website is minimal and very straightforward. The only problem was it is only limited to the US and its state's data. The chart can be more interactive and dynamic to make the user experience much better rather than showing static images.

3.3 Alternative technologies

Ultimately, there are lots of alternative tools and technologies to develop the product as mentioned above. The tools and technologies play a crucial role in creating the product. There is an enormous amount of programming language present today like Php, JavaScript, C#, Java, and many more but Python was chosen because it is the programming language of the future and the benefit of Python's simplicity helps engineers to save time by writing fewer lines of code to perform tasks. Python's speed lends itself nicely to data analysis. And this is owing to widespread support; the availability of a plethora of open-source libraries for a variety of applications, this includes, but is not restricted to, scientific computing. As a result, it's no surprise that it's touted as the ideal programming language for data analysis. Python has a wide range of unique characteristics that make it the best alternative for data analysis. Similarly, an IDE (Integrated Development Environment) is needed to execute Python code, so for data analysis visualization there are options like Jupyter Notebook, Kaggle, Azure Notebook, and many more, but for the development of my product, Google Colaboratory was preferred because it is free open-source and we can execute Python code on a web browser, and it also has cloud sharing functionality. Google Colab has built-in libraries and modules, so we don't have to manually install every module. It also includes free storage space and RAM for analysis.

There are several business intelligence products available, including Tableau, Power BI, Oracle Analytics Cloud, and others. The reasons for not selecting these applications were that they were premium and required payment to utilize their full functionality. They were primarily utilized in the enterprise, and creating a fully deployable dashboard required a lot of money. As a result, the best option was Google Data Studio, which is also a business intelligence tool for creating interactive dashboards and robust data analytics reports. It offers a lot of capability and an intuitive user interface, making it one of the top business intelligence tools. Although Python programming language was used, there are

several frameworks and modules available to help me create a website, such as Django, Flask, Dash, and many others. However, for the development of a web app, Streamlit library was picked as it allows for the rapid development of data-driven apps that are simple to deploy. Unlike Django and Flask, the Streamlit library does not require front-end coding; instead, it contains widgets and components that allow the creation of robust and fast web apps, can deploy machine learning models, and so on. Also, using Streamlit sharing, we can easily launch our web apps for free.

4 Chapter-4. Methodology

The selection of the proper methodology for developing software leads to quick implementation, cost savings, and risk avoidance. It also offers developing companies the finest practices for software applications that satisfy genuine consumer demands. Generally, the process comprises five phases of software development; namely: initiating requirements, developing them, accepting the final release, and supporting them including training, maintenance, and upgrading. Control and management of the software process is a vital responsibility to maintain a high level of integrity, quality, and user satisfaction in the generated software application. Similarly, for a data analysis project, there are steps such as business comprehension, information gathering, preprocessing, exploratory data analysis, modeling, model assessment, and model deployment.

4.1 Alternative methodologies

4.1.1 Waterfall methodologies

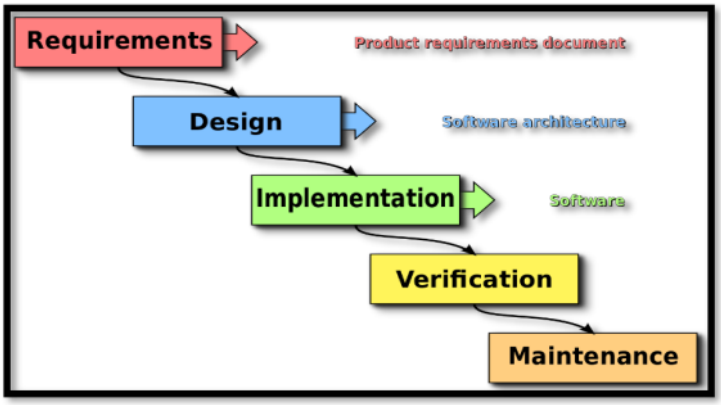


Figure 6. Waterfall model

The waterfall model is a linear project management approach in which the demands of both employees and shareholders are gathered at the initiation of the project. The waterfall model is so-called because every step of the project flows down like a waterfall. The model waterfall contains, in a precise linear order, 5-7 phases in which the stage cannot begin until the preceding stage is finished. The phase names vary, but they were specified by Winston W. Royce, the creator of the waterfall model. These five stages of the waterfall model are requirements, Design, Implementation, verification, and maintenance (Tutorialspoint, n.d.).

The major reason for not choosing the waterfall methodology is its high level of risk and uncertainty. This methodology is also not appropriate for projects with moderate to high risk of changing requirements. As the project had a lot of chances to change the requirement so it was not a suitable methodology. This process approach is hence highly risky and unpredictable. Changing the scope of a project during its life cycle might lead to its failure. Similarly measuring progression within phases is challenging and cannot adapt to changing needs. So this methodology was not best suitable for product development.

4.1.2 Rapid Application Development

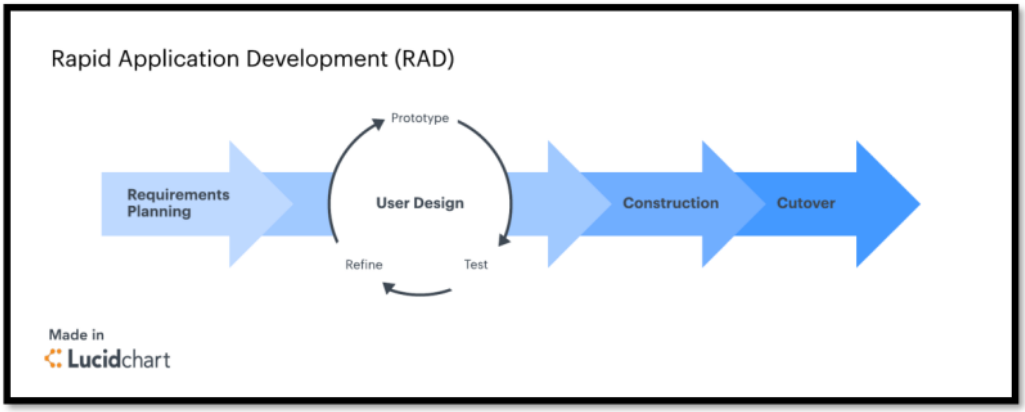


Figure 7. RAD Model

Rapid application development has been designed to utilize little preparation for fast prototyping software development approach. Iterative development approaches are used in the RAD projects paradigm and include a short group of devs, system engineers, customers, as well as other IT personnel, who are functioning on their area or model progressively. It focuses on the implementation of smaller tasks; the larger projects are split into several smaller ones. The reuse of themes, technologies, procedures, and code as one important part of the RAD paradigm has been underlined. (Tutorialspoint, n.d.).

The disadvantage of using Rapid Application Development(RAD) methodology is it depends on professional team members that are skilled to recognize business goals. But in the case of my product development, there is one person to develop the product. This methodology also requires highly qualified designers and developers. It is also suitable for component-based, scalable systems. Due to the expense of modeled and automatic code development, this technique is not suitable for low-cost applications. Hence it was not the best methodology for the development of my product.

4.2 Proposed Methodology

The agile development methodology is a blend of incremental and progressive models that focuses on process adaptability and customer satisfaction by providing functioning software rapidly. Agile techniques break the product down into small incremental structures. These architectures are presented in iterations. According to the agile method, almost every project should be handled differently, and current approaches must adapt to the project's demands. In Agile, the tasks are split into timescales (small timescales) to provide certain release features. The agile approach is a concept of software development that attempts to offer consumers greater value by adopting shorter development cycles and continuous changes (Tutorialspoint, n.d.).

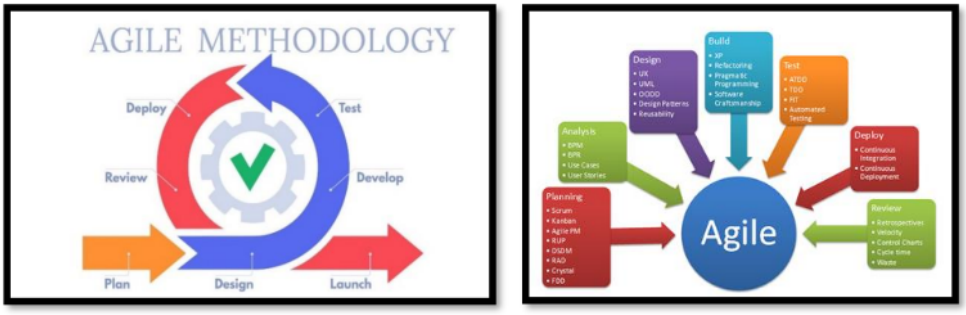


Figure 8. Agile Model

If we compare it with other methodologies it was best for the development of my product. It is also suitable for fixed or changing requirements during the development of the product. While in the development phase product is also easy to manage and this gives flexibility to developers. This methodology requires fewer resources. So Agile methodology is the best fit for my product development as compared to other methodologies.

The following phases were carried out during the product development lifecycle:

i. Planning and requirement analysis

At the beginning of the project, requirements were identified and analyzed. Subsequently, formal theoretical research was done, existing products were researched and a suitable project was identified and the end product was chosen along with its scope and features. The initial project specification was created and submitted for approval. After approval, a review of the Literature and Technical review was written through research. A methodology was established and the work breakdown structure was developed to follow throughout the life cycle of the project.

ii. Design

Following the requirement analysis and planning phase, the design of the end project was created. For designing part of the product, a use case diagram, wireframe, and development process design were designed for the end product.

iii. Implementation and Development

The implementation and development phase refers to the overall development of a product. After designing the whole product was coded and written to develop the end product. The product installation is outlined in this report in the implementation part of this report. In addition, accompanying product development the documentation was produced.

iv. Testing and maintenance

During this phase, the developed final product was tested utilizing the black box testing method to guarantee the product's quality and functioning. Any problems or errors discovered during the testing process were identified and resolved**.**

v. Closure and Evaluation

The product was initially shown/discussed with the customer/supervisor at this phase and based on their comments, necessary product adjustments were performed. Once the product was completed, the product was evaluated and checked if its original objectives have been met and whether the expected outcomes have been achieved. The limitations have been recognized and further work/recommendations were indicated. In addition, the product was shown, final documentation was developed and submitted

5 Chapter-5. Design

5.1 Use Case Diagram

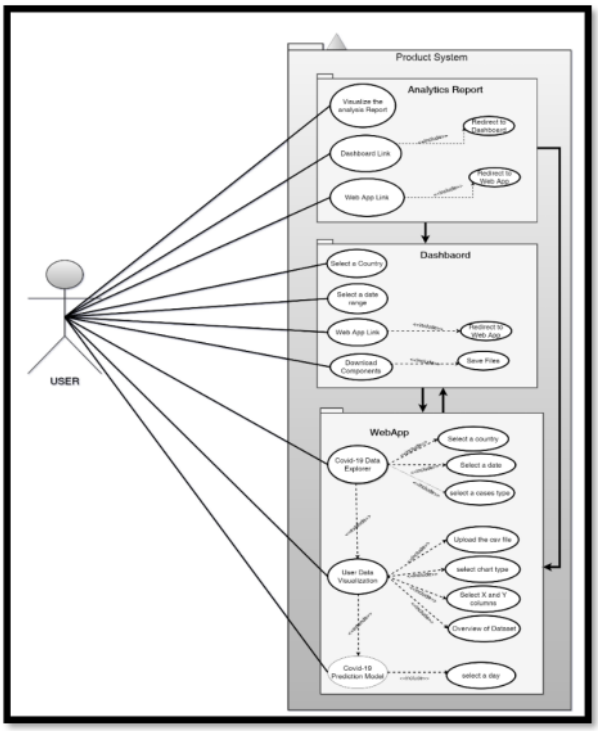


Figure 9. Use Cases Diagram

It is a visual representation of details about a system and its users which can be referred to as a use case diagram. It's also frequently portrayed as a graphical depiction of the interactions between various system components. Use case diagrams to explain the events in a system and how they flow, but they do not describe how those events are implemented.

Since the whole product is separated into three components, such as an analytics report, a dashboard, and a web app, the following picture represents a use case diagram for the entire system. So, to better comprehend the system, each component and its use cases are linked to one another and stored in a single system.

5.2 Flow Chart

A detailed analysis with its steps is a flow chart. It was developed in computer science as an algorithm and programming logic tool but has been expanded to all other operations. It enables us to envision intricate processes or clarify problems and task hierarchies. A process or project to be implemented may also be defined using a flowchart.

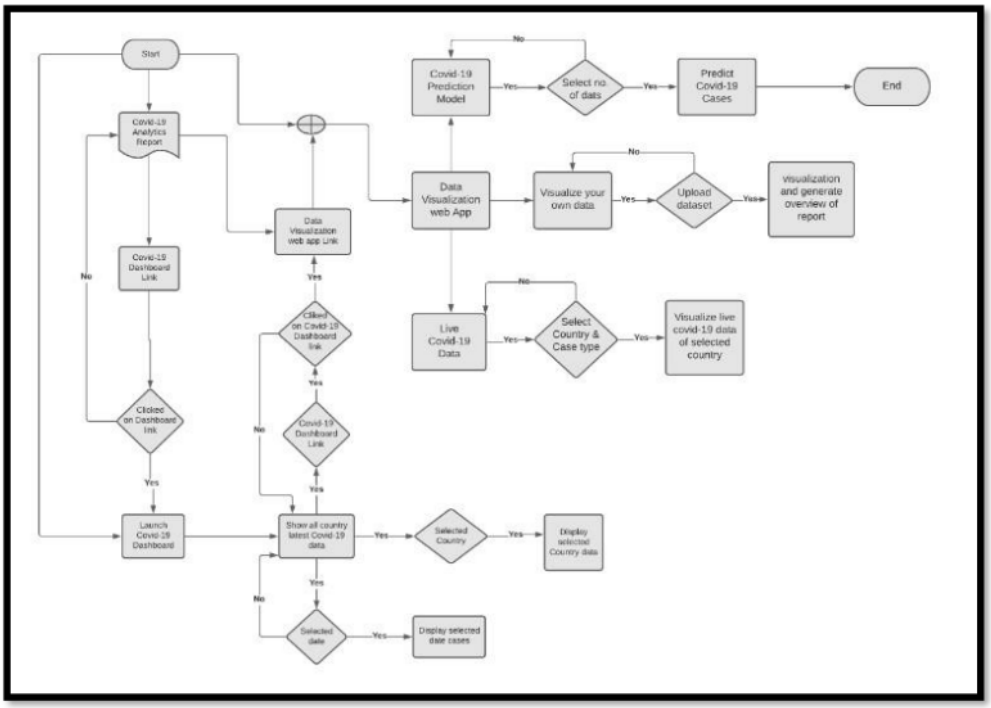


Figure 10. Flow Chart

Figure 10 shows the flowchart of the whole system. It displays the sequential phases of the process and the necessary decisions to implement the procedure.

5.3 The design of the development process

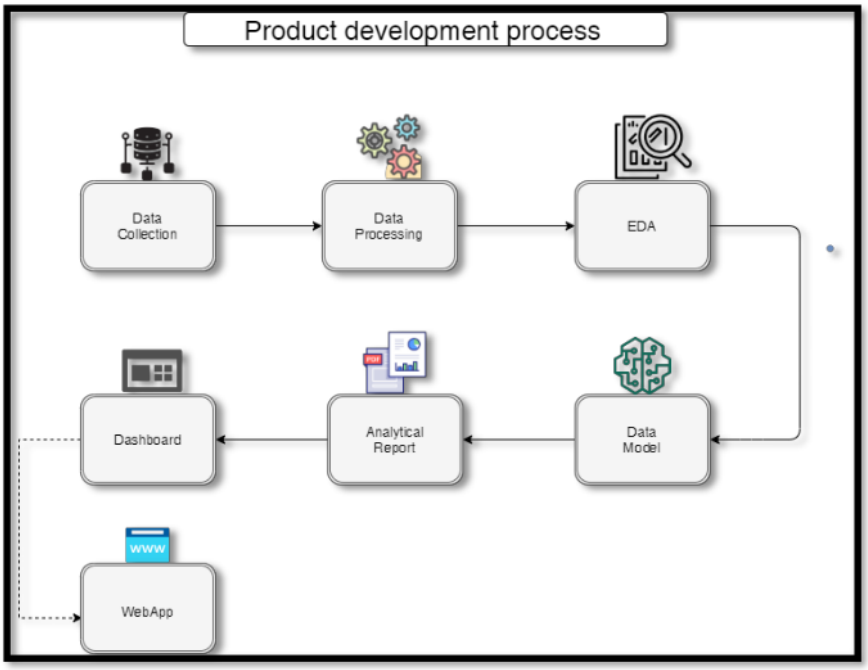


Figure 11. Development Process

Figure 11 is a graphical depiction of the procedure that took place throughout the creation of the entire project system. First, relevant data was collected, followed by data processing to purify the data, accompanied by exploratory data analysis (EDA) to uncover hidden patterns and valuable insight, and then, a prediction machine learning model was developed to forecast Covid-19 data. Similarly, an analytical report for Covid-19 is prepared to present the data analysis work. The Covid-19 dashboard is then created to make the user experience more engaging. At last to deploy a machine learning model, visualize live covid-19 data and let the user visualize their data visualization is developed.

5.3.1 Wireframe

A wireframe illustrates the interface of a two-dimensional page, focusing in particular on the allotment of space and priority of contents, features accessible, and behavior. Wireframes generally don't feature styles, colors or images, for these reasons. Wireframes also assist build links between different layouts of a website. Before the development of any framework begins, a wireframe is necessary. It causes the designer to compose and reorganize his work by visually presenting the prior development of the system.

Covid-19 Dashboard

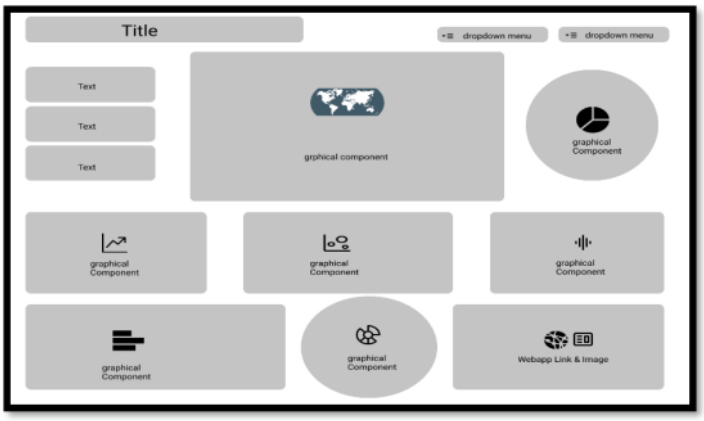


Figure 12. Covid-19 Dashboard wireframe

Figure 12 shows the homepage, or the first page, of a Covid-19 Dashboard. It will have a title on the top left side of the page and two drop-down options for selecting a country and date range in the top right area. Just below the title, there will be three boxes depicting Covid-19 cases, with a globe map displaying the infected country in the middle section. Similarly, a pie chart with the top confirmed cases will be presented in the dropdown option down below. The three columns in the middle part will indicate Covid-19 cases by date, 1-week change/percentage, and deaths & recovered per 100 cases respectively. The last row will contain cases per million, top deaths, and an image link to the web app

Covid-19 data explorer section

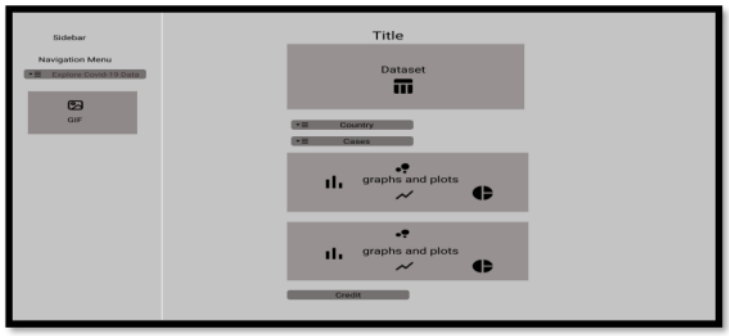


Figure 13. Covid-19 Data explorer section wireframe

Figure 13 depicts the web's initial page, which displays the live covid-19 dataset section in the sidebar navigation menu and data at the top of the page, as well as two dropdown options below for selecting a country and case type to produce graphs and visualize data based on user input.

Covid-19 prediction model

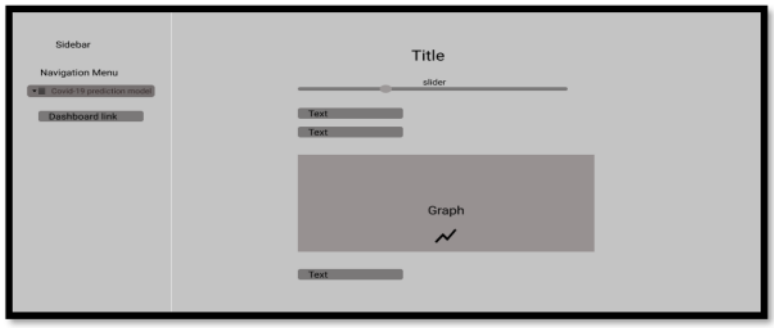


Figure 14. Covid-19 Prediction Model Section wireframe

By selecting the 'Covid-19 prediction model' from the sidebar menu, the user is routed to this page, which includes its title, a slider for prediction data and text labels, and a chart of predicted and actual data. The Covid-19 dashboard link is also available in the sidebar.

Visualize your data section

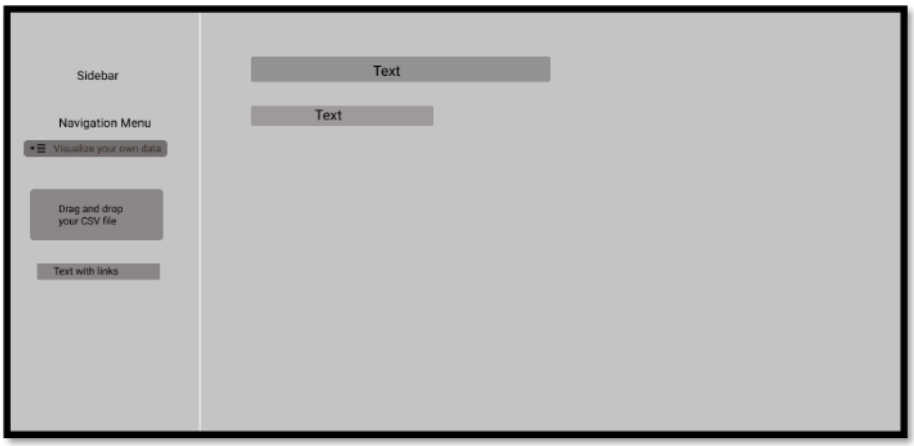


Figure 15. Visualize your data section wireframe

Figure 15 represents the initial page of the final menu ' Visualize your data' where users can upload CSV files in the sidebar and visualize as well as generate an overview of their dataset and also down below there is a text link for the Covid-19 dashboard.

5.4 System Requirement

The functional and non-functional requirements of the project are attached in the appendix A section of the report in Project Specification.

6 Chapter-6 Implementation and Testing

During the Implementation phase of the SDLC, the project starts taking form. We started building here actual products based on the designs we created during performing the design of a system. Every bit of product's requirements are examined. We begin designing the user-friendly interface design by implementing the product using wireframes. Coding is done at this phase of the project's development. As an agile approach has been utilized it is more convenient to apply modifications so that during the coding phase it also improves the implementation of the new feature in the project. Python programming language is used for data analysis, visualization, and developing a web app while business intelligence tool Google data studio has been used for creating a covid-19 dashboard.

6.1 Implementation of Data analysis and visualization

Importing library and loading data

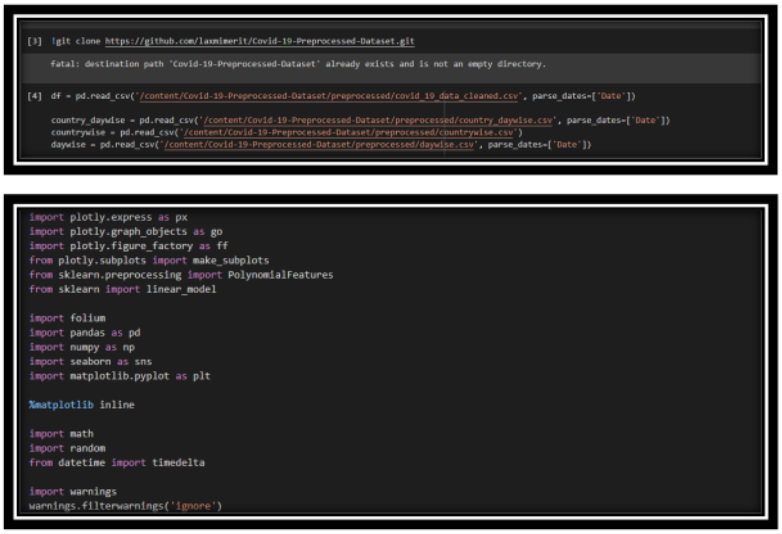


Figure 16. Libraries and data

Figure 16 depicts that all the necessary packages of python are imported required for data analysis and visualization. Similarly, data is loaded and read from the GitHub repository and prepared for analysis.

Worldwide Covid-19 cases on folium map

```
temp = df[df['Date']==max(df['Date'])]

m= folium.Map(location=[0,0], tiles='cartodbpositron', min_zoom=1, max_zoom=4, zoom_start=1)

for i in range(0, len(temp)):
  folium.Circle(location=[temp.iloc[i]['Lat'], temp.iloc[i]['Long']], color='crimson', fill='crimson',
                tooltip='<li><bold> Country: ' + str(temp.iloc[i]['Country'])+
                '<li><bold> Province: ' + str(temp.iloc[i]['Province/State'])+
                '<li><bold> Confirmed: ' + str(temp.iloc[i]['Confirmed'])+
                '<li><bold> Deaths: ' + str(temp.iloc[i]['Deaths']),
                radius =int(temp.iloc[i]['Confirmed'])**0.5).add_to(m)
m
```

Figure 17. Worldwide Covid-19 cases

Figure 17 shows the code and output for generating worldwide covid-19 cases on the folium map.

Worldwide Covid-19 cases on folium map

```
fig= px.choropleth(country_daywise, locations='Country', locationmode='country names', color=np.log(country_daywise['Confirmed']),
                   hover_name = 'Country', animation_frame=country_daywise['Date'].dt.strftime('%Y-%m-%d'), title='Cases over time', color_continuous_scale=px.colors.sequential.Oryel
                   )
fig.update(layout_coloraxis_showscale=True)
fig.show()
```

Figure 18. Cases over time

Figure 18 displays the code that generates a choropleth with a timeframe along with cases over time.

Deaths/100 cases, Recovered/100 cases, Deaths/100 Recovered

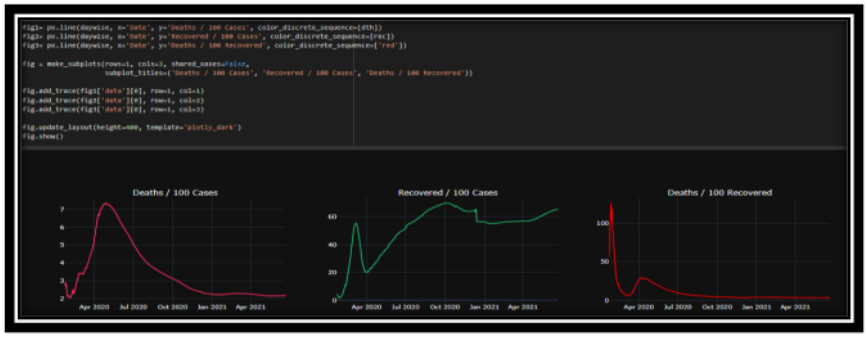


Figure 19. Deaths/100 cases, Recovered/100 cases, Deaths/100 Recovered

Figure 19 represents the code that creates the graph of deaths/100 cases, recovered/100, and deaths/100 cases

Covid-19 Compared with other epidemics

```
temp = epidemics.melt(id_vars='epidemic', value_vars=['confirmed', 'deaths', 'mortality'],
                      var_name='Case', value_name='Value')
temp

fig = px.bar(temp, x = 'epidemic', y ='Value', color='epidemic', text='Value',
             facet_col='Case', color_discrete_sequence=px.colors.qualitative.Bold, template='plotly_dark')

fig.update_traces(textposition='outside')

fig.update_layout(uniformtext_minsize=8, uniformtext_mode='hide')
fig.update_yaxes(showticklabels = False)
fig.layout.yaxis2.update(matches= None)
fig.layout.yaxis3.update( matches = None)
fig.show()
```

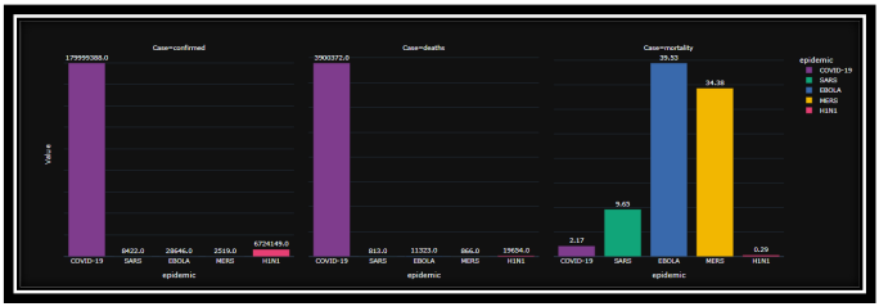


Figure 20. Covid-19 compared with other epidemics

Figure 20 represents the code that generates the bar graph which shows the covid-19 cases and mortality rate compared with other epidemics.

Machine learning model

```
x= np.array(data['id']).reshape(-1,1)
y= np.array(data['cases']).reshape(-1,1)
plt.plot(y, '-m')

polyFeat = PolynomialFeatures(degree=5)
x = polyFeat.fit_transform(x)

model = linear_model.LinearRegression()
model.fit(x,y)
accuracy = model.score(x,y)
print(f'Accuracy:{round(accuracy*100,3)} %')
y0 = model.predict(x)
plt.plot(y0, '--b')
plt.show()

#prediction
days=3
print(f'Prediction - Cass after {days} days:', end='')
print(round(int(model.predict(polyFeat.fit_transform([[584+days]])))/1000000,2),'Million')
```

Figure 21. prediction machine learning model

Figure 21 shows the code that is used to create a machine learning model.

6.2 Implementation of Covid-19 Dashboard

Setting data sources

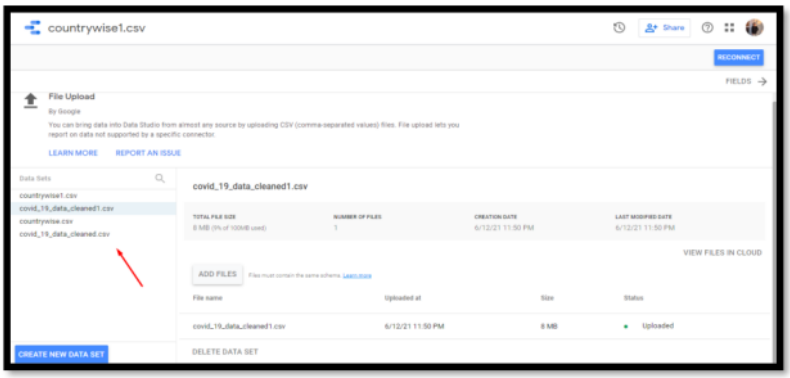


Figure 22. Data Sources

Different data sources are managed and evaluated for creating the Covid-19 dashboard.

Creating Covid-19 Dashboard in Google Data Studio

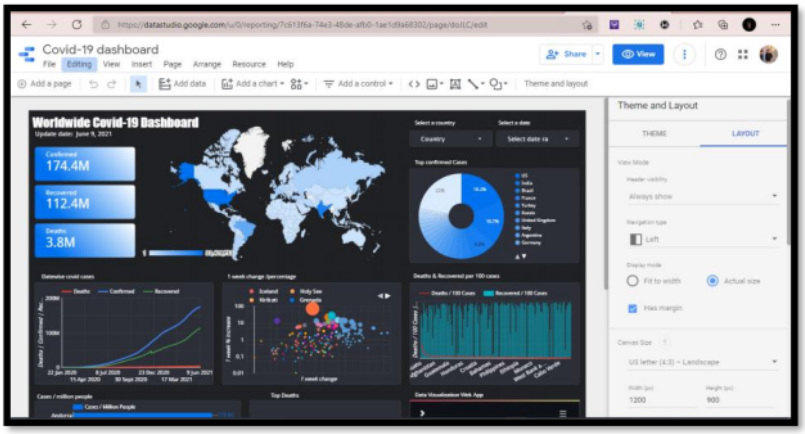


Figure 23. Covid-19 Dashboard

Figure 23 displays the making of the Covid-19 dashboard in Google's data business intelligence tool.

6.3 Implementation of a Web App

Installing python

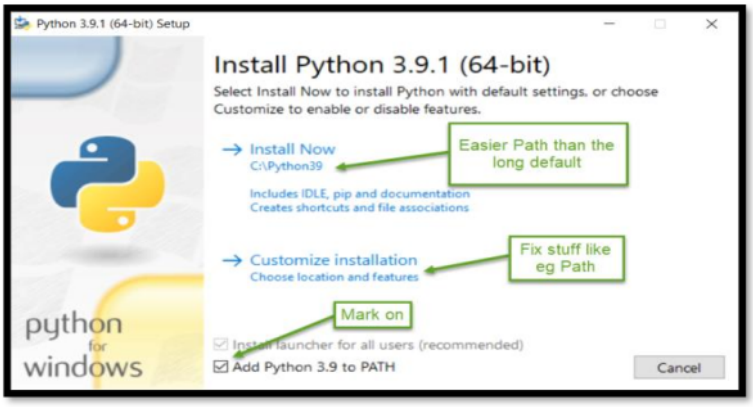


Figure 24. Installing python

Python programming language stable version was downloaded and installed on a laptop.

Importing libraries and packages in PyCharm

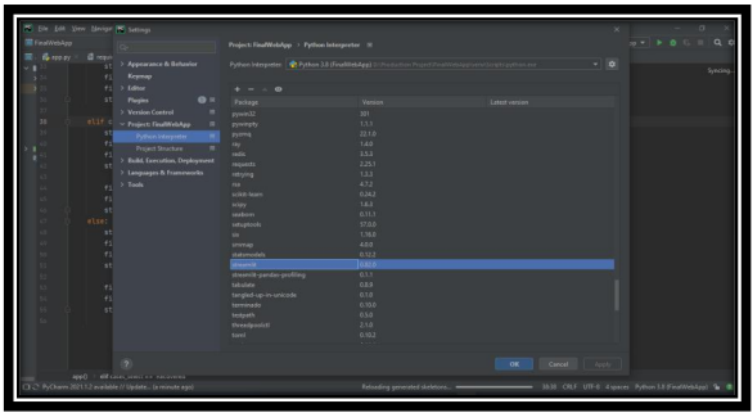


Figure 25. Installing libraries

Essential packages and libraries were installed in PyCharm IDE for the development of the product.

Fetching Live Covid-19 data

```
import pandas as pd
import plotly_express as px
import numpy as np
import streamlit as st


def app():
    st.info('It takes a while ⌛ to fetch data. You can see the data 👁 once its loaded.')
    st.title('Covid-19 Data Explorer')
    st.sidebar.image("https://media.giphy.com/media/d7ksD7AarGUDz0NmxU/giphy.gif", width=300)
    covid = pd.read_csv('https://raw.githubusercontent.com/Laxmiserit/Covid-19-Preprocessed-Dataset/master/preprocessed/covid_19_data_cleaned.csv')
    covid = covid.drop(["Province/State", "Lat", "Long"], axis=1)
    covid.columns = ['Date', 'Country', 'Confirmed', 'Recovered', 'Deaths', 'Active']
    covid['Date'] = pd.to_datetime(covid['Date']).dt.strftime('%Y-%m-%d')
    st.write(covid)

    country_options = covid['Country'].unique().tolist()
    date_options = covid['Date'].unique().tolist()
    date = st.selectbox('Which date would you like to see?', date_options, 100)
    country = st.multiselect('Which country 🌎 would you like to see?', country_options, ['US', 'India', 'Italy'])
    covid = covid[covid['Country'].isin(country)]
    # covid = covid[covid['Date'] == date]
    cases_select = st.selectbox(
        label="Select the Case type",
        options=['Confirmed', 'Recovered', 'Deaths'])
```

Figure 26. Code to fetch live data

Figure 26 represents the code for fetching the raw live data and preprocessing from Github repositories to render on a web app. Similarly, it also has code for a dropdown menu for selecting country and type of cases to visualize graphically.

Generating Graphs and Charts

```
if cases_select == 'Confirmed':
    st.subheader("Confirmed Cases")
    fig = px.bar(covid, x="Country", y="Confirmed", color="Country")
    fig.update_layout(height=600, width=800)
    st.write(fig)

    st.subheader("Date-wise Area plot")
    fig1 = px.area(covid, x="Date", y="Confirmed", color="Country")
    fig1.update_layout(height=600, width=800)
    st.write(fig1)

elif cases_select == 'Recovered':
    st.subheader("Recovered Cases")
    fig = px.bar(covid, x="Country", y="Recovered", color="Country")
    fig.update_layout(height=600, width=800)
    st.write(fig)

    fig2 = px.line(covid, x="Date", y="Recovered", color="Country")
    fig2.update_layout(height=600, width=800)
    st.write(fig2)
else:
    st.subheader("Deaths Cases")
    fig = px.bar(covid, x="Country", y="Deaths", color="Country")
    fig.update_layout(height=600, width=800)
    st.write(fig)
```

Figure 27. Code to generate charts in web app

Figure 27 shows the code that plots the charts and graphs as users input. Whenever the user selects a particular country and type of case in the dropdown menu this code gets executed and plots the chart.

Upload CSV file

```
def app():
    with st.sidebar.header('1. Upload your CSV data'):
        uploaded_file = st.sidebar.file_uploader("Upload your input CSV 📄 file", type=["csv"])
        st.sidebar.markdown("""
[Example CSV input file](https://raw.githubusercontent.com/dataprofessor/data/master/delaney_solubility_with_descriptors.csv)
""")

    # Pandas Profiling Report
    if uploaded_file is not None:
        @st.cache
        def load_csv():
            csv = pd.read_csv(uploaded_file)
            return csv

        df = load_csv()
        pr = ProfileReport(df, explorative=True)
        st.header('**Input DataFrame**')
        st.write(df)
        global x_columns, y_columns
        try:
            x_columns = list(df.select_dtypes(['datetime64', 'float']))
            y_columns = list(df.select_dtypes(['integer', 'float']))

        except Exception as e:
            print(e)
            st.write("Please upload dataset in the sidebar > of the application\n"
                     "to visualize your own data")
```

Figure 28. Code to upload file in web app

Figure 28 represents the code that lets the user upload the CSV file and defines the data types that should be in X and Y columns.

Chart selection

```
if chart_select == 'Scatterplots':
    st.sidebar.subheader("Scatterplot settings")
    try:
        x_values = st.sidebar.selectbox('X axis', options=x_columns)
        y_values = st.sidebar.selectbox('Y axis', options=y_columns)
        plot = px.scatter(data_frame=df, width=750, size_max=9, size=y_values, color=y_values,
                          color_continuous_scale=px.colors.cyclical.mrybm, x=x_values, y=y_values, log_x=True)
        st.plotly_chart(plot)
    except Exception as e:
        print(e)

# Line plots

if chart_select == 'Lineplots':
    st.sidebar.subheader("Lineplots settings")
    try:
        x_values = st.sidebar.selectbox('X axis', options=x_columns)
        y_values = st.sidebar.selectbox('Y axis', options=y_columns)
        plot = px.line(data_frame=df, x=x_values, y=y_values, color=y_values)
        st.plotly_chart(plot)
    except Exception as e:
        print(e)
```

Figure 29. Code for selection of chart

Figure 29 shows the code where the user can select the type of chart and can define what variables should be in x columns and y columns to visualize the data.

Report generation

```
df = load_data()
pr = ProfileReport(df, explorative=True)
st.header('**Input DataFrame**')
st.write(df)
st.write('---')
st.header('**Report of Your Dataset**')
st_profile_report(pr)
```

Figure 30. . Code to generate a report

Figure 30 represents the code that helps to generate an overview of a report using the pandas-profiling library.

Deploying Machine Learning model

```
def app():
    st.title(" Covid-19 Prediction  Model")
    a = st.slider('select a day range from 0 to 5', min_value=0, max_value=5, value=2)
    data = pd.read_csv('total_Cases.csv')
    nan_value = float("NaN")
    data.replace("", nan_value, inplace=True)
    data.dropna(subset=['id', 'cases'], inplace=True)
    m = (data.dtypes == 'float')
    data.loc[:, m] = data.loc[:, m].astype(int)

    x = np.array(data['id']).reshape(-1, 1)
    y = np.array(data['cases']).reshape(-1, 1)
    plt.plot(y, '-m')

    polyFeat = PolynomialFeatures(degree=6)
    x = polyFeat.fit_transform(x)
    model = linear_model.LinearRegression()
    model.fit(x, y)
    accuracy = model.score(x, y)
    st.write(f'Accuracy:{round(accuracy * 100, 6)} %')
    y0 = model.predict(x)
    plt.plot(y0, '--b')

    # prediction
    st.write(f'Prediction - Cass after {a} days:', end='')
    st.write(round(int(model.predict(polyFeat.fit_transform([[335 + a]]))) / 1000000, 2), 'Million Cases')
    st.pyplot(plt)
```

Figure 31. Code to deploy machine learning model

Figure 31 displays the code for the machine learning model which is deployed in the web app to predict the Covid-19 cases. It was build using a nonlinear regression model.

6.4 Testing

6.4.1 Black box testing

The "black box testing" implies a software testing approach that analyzes program functioning without peeking into its core structure or code. The major source of black-box testing is a customer-specified requirement specification. In essence, a tester does not have any link to the code involved. The testing device only examines interface problems, product performance, and data structure. We should not employ the developer of software to verify the system because of this testing as we do not test internal programming faults. We need only a person who can inspect the system in the future.

Test cases

The tests listed in the tables below are performed to ensure that there are no bugs or errors in the system. All tests are completed with 0 mistakes successfully. A picture relating to testing is provided in the appendix section of the report below.

Table 1. Tests

ID	Tests	Details	Expected output	Actual output	Test pass(Yes/No)
1	Null value	Removing null value	Null free data	Null value removed successfully	Yes
2	Accuracy	Prediction model accuracy	Above 90 % accuracy	Above 99% model achieved successfully	Yes

3	Select one or multiple countries	Selecting multiple countries	Give data of the selected country	Showed the selected country data successfully	Yes
4	Select date	Selecting a date range	Display data of particular data	Covid-19 data of specific date was given successfully	yes
5	Select a case type	Selecting Covid-19 cases	Display the chart of particular cases	Shows the live Covid-19 data of specific cases successfully	Yes
6	Upload file	uploading the CSV file	Show the detailed overview of the dataset	Provides the overview of the dataset successfully	yes
7	Chart selection	Able to select the specific chart	Present the chart as per the selected chart type	Displays the selected chart successfully	Yes
8	X column & Y Column	Selecting X variable and Y variable	Show the chart as per X and Y variable	Displayed the chart of X and Y successfully	Yes
9	Report generation	Overview of uploaded dataset	Generates the overview of the uploaded dataset	Displays the overview of report successfully	Yes

10	Prediction	Predict the Covid-19 cases	Display the cases as per the day selected	Provides the no. of cases as per the number of days successfully	Yes
11	Links	links redirecting from one page to another	Should be able to move from dashboard to web app and vice versa.	Successfully moving from dashboard to web app and vice versa	yes

In-depth testing is mentioned in the Appendix C section of this report.

7 Chapter-7. Product Evaluation

Typically, when the project is completed, a product evaluation is performed to determine whether or not the project objectives were met. The final product evaluation is reliable or not; it is based on the estimated results and fulfills or does not meet the system's requirements. The software development process begins with a project plan and outline, which includes design elements such as ERD, Use Case, Gantt chart, and agile approach to ensure that the project stays on track. After each meeting for this project, several modifications were made based on the supervisor's suggestions. Overall, the final product was finished with all of the requirements. The product evaluation is performed done below.

Table 2. Analytics report evaluation

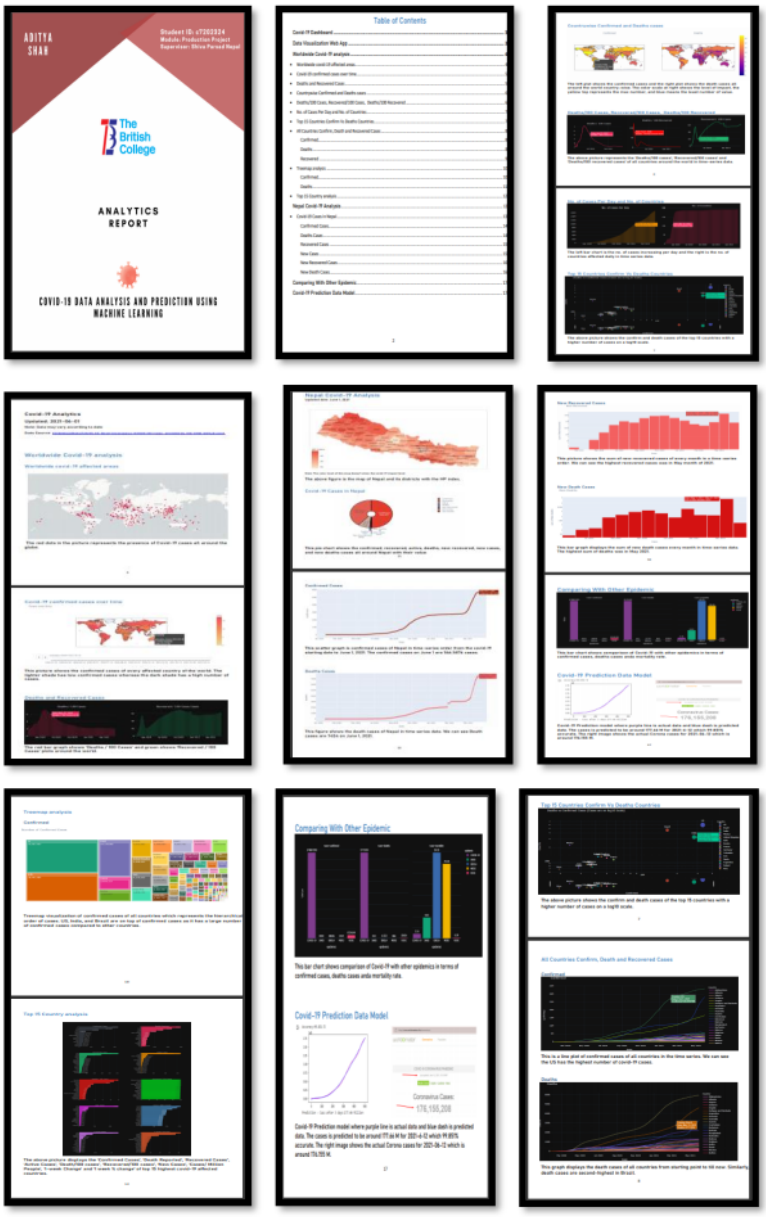
Requirement	Covid-19 Analytics Report
Expected	Result of data analysis and visualization in the analytics report
Outcome	
Evaluation	Success

Table 3. Covid-19 dashboard evaluation

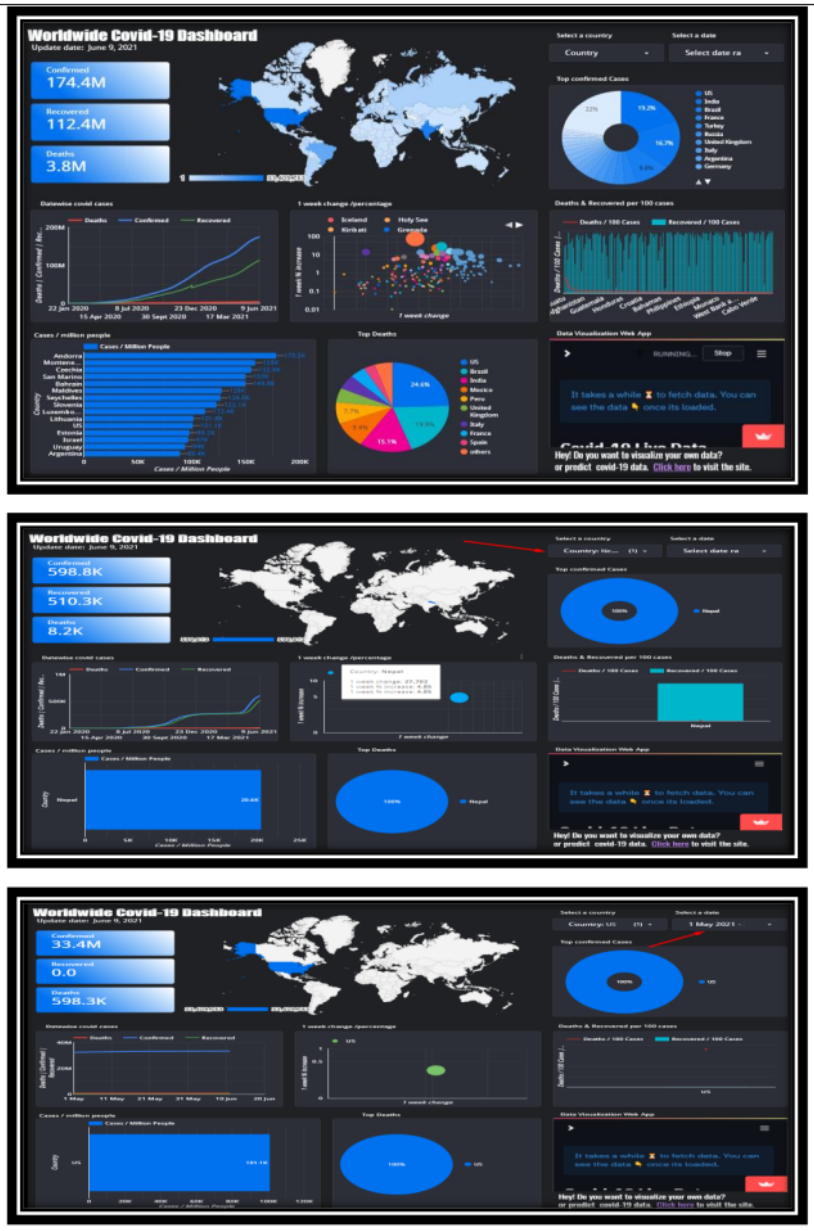
Requirement	Covid-19 Dashboard
Expected	Interactive Covid-19 Dashboard for visualization
Outcome	
Evaluation	Success

Table 4. Web app evaluation

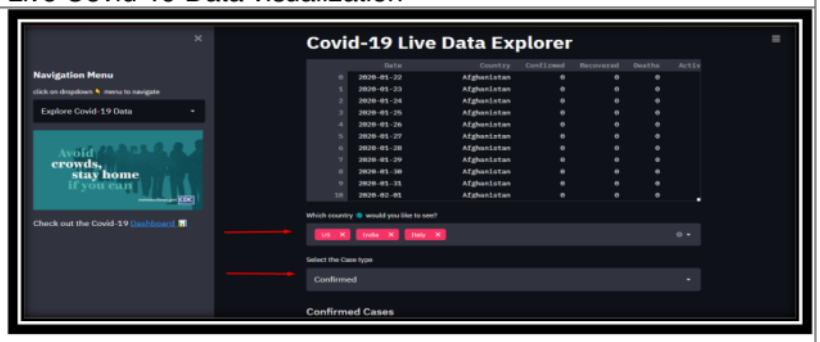
Requirement	Data Visualization Web App
Expected	Live Covid-19 Data visualization
Outcome	  
Evaluation	Success

Table 5. Upload file evaluation

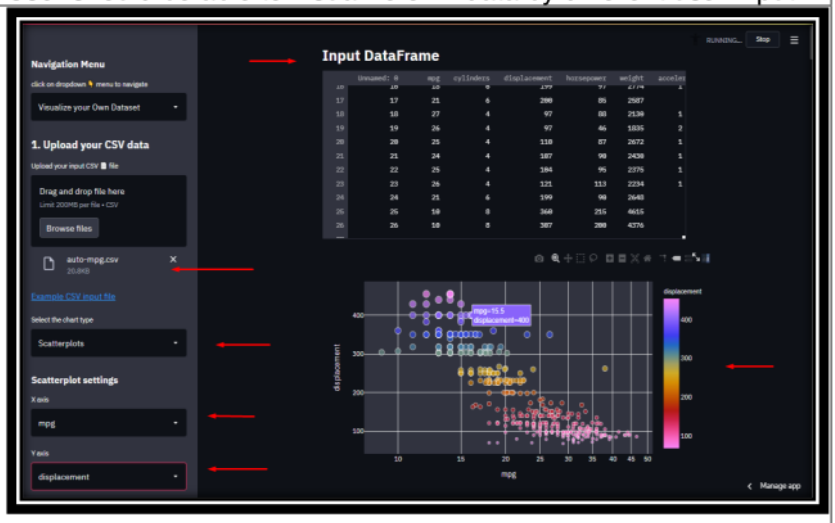
Requirement	Upload own data set and select the chart type
Expected	User should be able to visualize own data by different user input
Outcome	
Evaluation	Success

Table 6. Report evaluation

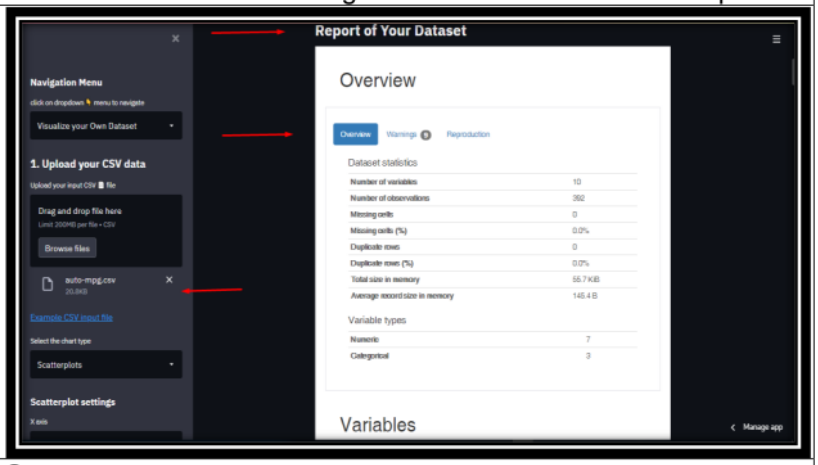
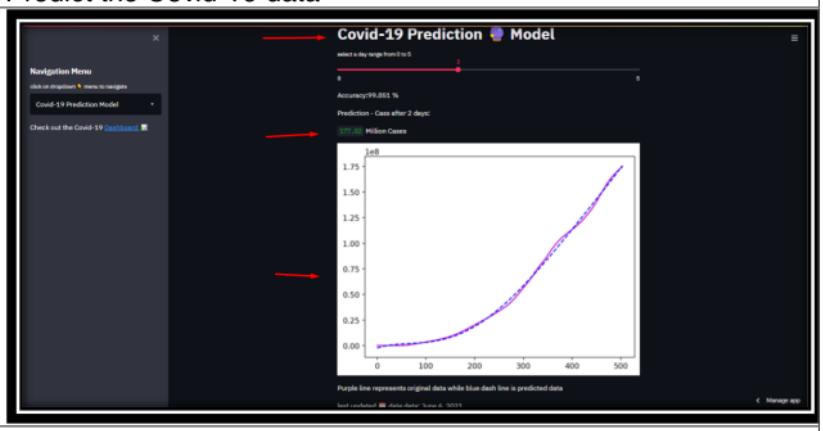
Requirement	Generate an overview of users dataset
Expected	The user should be able to generate an overview of the report
Outcome	
Evaluation	Success

Table 7. Prediction evaluation

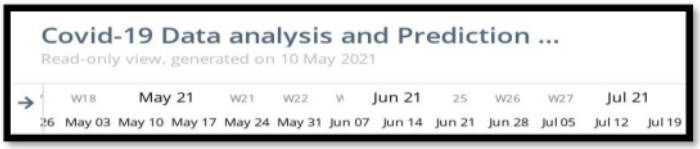
Requirement	Covid-19 Prediction
Expected	Predict the Covid-19 data
Outcome	
Evaluation	Success

8 Chapter-8. Project Evaluation

The project evaluation provides a clear indication of how the project is performing in comparison to the original evaluations. Similarly, it provides an opportunity to consider how the developers alter their technique of working by examining how they operate and changing the items that must be done properly. However, all phases necessary during the product enhancement lifecycle will be evaluated.

Project Timeline

To achieve the aim, objectives, and milestone project was planned, breakdown and managed from starting of the project.



Gantt chart

The breakdown structure design was created to establish the steps to be followed at various phases of the project. Following figure 32 displays the overview of the project.

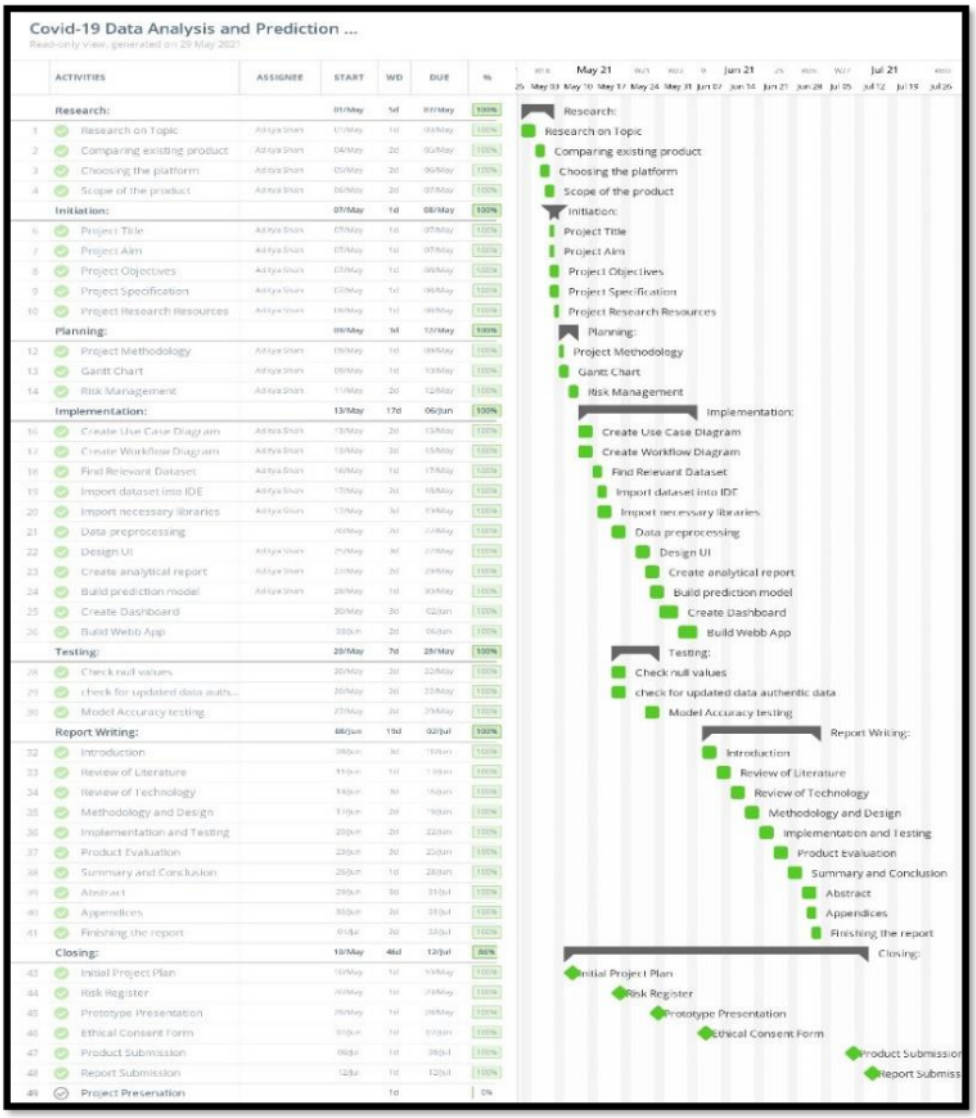


Figure 32.Gantt Chart

Resource sheet

The resource sheet is the list of all the resources that have been used during the development of the product.

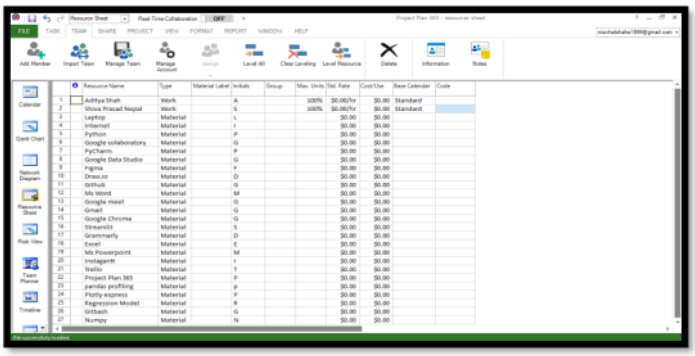


Figure 33. Resource sheet

Task management

Since this project was developed using agile methodology, the Kanban framework was utilized to manage the tasks. Kanban is a management method that is used to visualize work, manage tasks, and enhance overall efficiency in product development. Kanban boards, as illustrated below, uses cards to represent work items and columns to represent each phase of the process to visually depict work at various stages.

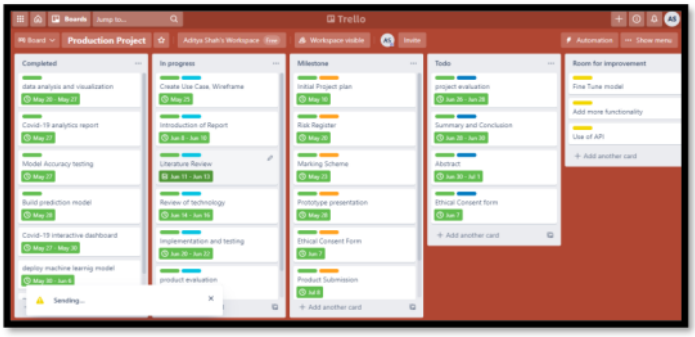


Figure 34. Trello board

GitHub

GitHub was used for the version control as well as deployment of web app since Streamlit sharing requires the whole code in GitHub for the deployment of the web app.

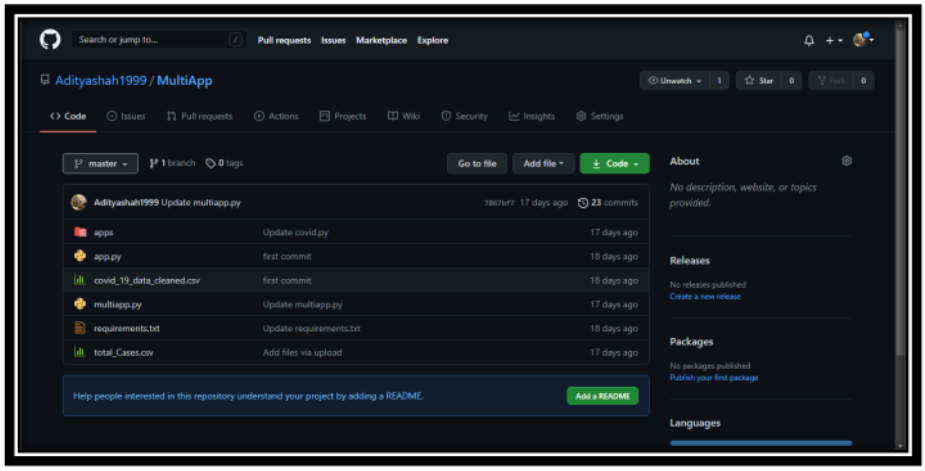


Figure 35. GitHub

Moreover, the project was challenging during the implementation and development phases because the project topic and programming language were intimately familiar to me. Given the challenges, it also helped me learn and acquire expertise after a lot of research and experimenting. Ultimately, the project was beneficial since it allowed me to hone my skills. Despite the occurrence of the problem, planning the whole project in Instagannt, monitoring tasks using Agile Kanban methodology, and utilizing Github for version control resulted in the project's aim and objectives being accomplished within the time limit stated.

9 Chapter-9. Summary and Conclusion

9.1 Summary

Covid-19 Analytical report, Covid-19 dashboard, and data visualization web app give a comprehensive dynamic experience for users in understanding Covid-19 data, Covid-19 prediction using machine learning, and data visualization. The project's goal is to do data analysis and show how a huge amount of Covid-19 raw data is analyzed and visualized to gain useful insight and discover trends and anomalies. It also aimed to show how, using a machine learning model, we can predict Covid-19 cases. Not only did the project aim to render and visualize live worldwide Covid-19 data, but it was also designed to allow the user to visualize and generate an overview of the user's uploaded dataset in a web app. To develop an end product that achieves the project's goal, research was performed in the fields of machine learning, data analysis, and business intelligence techniques. Following that, for the development of the product, research on methodology was undertaken and chosen that followed throughout the project. The ultimate product's conceptual and logical designs were produced, followed by its implementation on the basis of prototyping. The solution was further evaluated to see if it was as successful as it was anticipated. In summary, the finished product/system enabled users to obtain immersive knowledge on Covid-19; access live Covid-19 data and visualize it based on user input; predict Covid-19 cases; visualize their dataset; generate a report about their dataset; aid in effective decision making.

9.2 Recommendation

Even though the features developed were as or even more than what was originally envisioned, several unanticipated limits were detected. Because the data is updated late, there was a data limitation for the Covid-19 data analysis and visualization. Also, because the Google Data Studio business intelligence tool was used for the dashboard, the real-time Covid-19 database was difficult to find and maintain. As a result, the data source in the backend had to be manually

updated for display in the dashboard. However, if we can purchase an API or obtain a backend real-time server, we will be able to fix the problem. Similarly, Covid-19 prediction data is manually updated since it follows a specific model. To address this issue, appropriate data can be retrieved and updated on the server regularly. As data are extremely complicated due to their structure, the user cannot visualize the date-time column variable of their dataset for the time being, and it only handles a maximum 200MB CSV file. The user is limited to predicting Covid-19 cases for the next 5 days since the data changes quickly and its accuracy may fail for a longer number of days. However, we can use various machine learning techniques to fine-tune the model and make it more exact and accurate.

9.3 Future Work

Considering new features are always in existence for every project, new features could be added to further enhance this project. The addition of a real-time time database or API, as indicated in the limitation above, is a highly suggested feature to add. In fact, the ability to view and explore live Covid-19 data in the dashboard and predict case outcomes is a useful feature to provide. Another suggested feature would be the addition of a user profile in the web app, as well as increasing the uploaded size of the dataset file to more than 200MB and supporting formats such as excel. Adding a feature like supporting every variable column and creating more graphs and charts than in this product could make the product stand out. Deploying and implementing different advanced machine learning algorithms for prediction might also advance this project forward. Thus, this project can be extended in an infinite number of ways; the ultimate decision on what features to implement will undoubtedly be determined by the user and developers.

10 Bibliography

10.1 Books

Brownlee, J., 2016. Supervised Machine Learning. In: *Master Machine Learning Algorithms: discover how they work and implement them from scratch.* Melbourne, Australia: Jason Brownlee, p. 16.

Burkov, A., 2019. What is Machine Learning. In: *The hundred-page machine learning book.* s.l.:Andriy Burkov, p. 3.

Guido, A. C. M. a. S., 2016. Introduction . In: D. Schanafelt, ed. *Introduction to Machine Learning with Python.* s.l.:O'Reilly Media, p. 1.

Vercellis, C., 2010. *Business intelligence: data mining and optimization for decision making.* 1 ed. s.l.:John Wiley & Sons.

10.2 Journals

Benjamin D Wissel, P. J. V. C. M. K. C. W. T. A. G. P. S. W. I. S. K. J. W. D., 2020. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *Journal of the American Medical Informatics Association,* 27(7), p. 1121–1125.

Binti Hamzah FA, L. C. N. H. L. D. L. G. T. C., 2020. *CoronaTracker: World-wide COVID-19 Outbreak Data Analysis and Prediction,* s.l.: Bull World Health Organ.

Fernandez-Lozano, C. C. F., 2020. Shiny Dashboard for Monitoring the COVID-19 Pandemic in Spain. *Proceedings,* 54(1), p. 23.

Hector Florez, S. S., 2020. Online dashboard and data analysis approach for assessing COVID-19 case and death data. *F1000Research,* Volume 9, p. 570.

Muhammad, L. J. A. E. A. U. S. S. A. A. C. C. M. I. A., 2021. Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. *SN Computer Science,* 2(1).

Singhal, T., 2020. A Review of Coronavirus Disease-2019 (COVID-19). *The Indian Journal of Pediatrics,* 84(4).

10.3 Website

ActiveState, n.d. *ActiveState.* [Online]
Available at: https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/
[Accessed 16 June 2021].

Analytiks, 2020. *Analytiks.* [Online]
Available at: https://analytiks.co/importance-of-data-visualization/
[Accessed 12 June 2021].

Data Studio Help, n.d. *Data Studio Help.* [Online]
Available at: https://support.google.com/datastudio/answer/6283323?hl=en
[Accessed 21 June 2021].

deepanshu_rustagi, 2020. *GeeksforGeeks.* [Online]
Available at: https://www.geeksforgeeks.org/pandas-profiling-in-python/
[Accessed 14 June 2021].

Google Colaboratory, n.d. *Google Colaboratory.* [Online]
Available at: https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index#scrollTo=UdRyKR44dcNI
[Accessed 20 June 2021].

Intellipat, 2020. *intellipat.* [Online]
Available at: https://intellipaat.com/blog/what-is-pycharm/
[Accessed 20 June 2021].

KARANJAGOTA, 2021. *Analytics Vidhya.* [Online]
Available at: https://www.analyticsvidhya.com/blog/2021/04/developing-data-

web-streamlit-app/
[Accessed 21 June 2021].

NetApp, n.d. *NetApp.* [Online]
Available at: https://www.netapp.com/artificial-intelligence/what-is-machine-learning/
[Accessed 15 June 2021].

Pedamkar, P., 2020. *EDUCBA*. [Online]
Available at: https://www.educba.com/what-is-numpy/
[Accessed 21 June 2021].

Plotly, 2019. *Medium.* [Online]
Available at: https://medium.com/plotly/introducing-plotly-express-808df010143d
[Accessed 19 June 2021].

TechnologyHQ, 2021. *TechnologyHQ.* [Online]
Available at: https://www.technologyhq.org/importance-data-analytics-modern-world/
[Accessed 10 June 2021].

Tutorialspoint, n.d. *tutorialspoint.* [Online]
Available at: https://www.tutorialspoint.com/sdlc/sdlc_rad_model.htm
[Accessed 18 June 2021].

Appendix A

Project specification

Sc (Hons) Computing Course 2020/21

Level 6 Production Project

Name: Aditya Shah	**Student I.D.:** C7202324
Course: BSc (Hons) Computing	**Supervisor's Name:** Shiva Parsad Nepal

Final Project Individual Aim & Objectives

Title of my Project:

Covid-19 Data analysis and Prediction using Machine-Learning

Aim of my Project:

This project aims to analyze the Covid-19 dataset as well as perform visualization and prediction to provide meaningful insight and additionally, create an analytical report, dashboard, and web app for interactive user experience.

Objectives of my Project:

The primary objective of this product is:

1. To create a Covid-19 analytics report of data analysis and prediction.
2. To produce Covid-19 Dashboard for interactive user experience.
3. To develop a data visualization web app for prediction model deployment, live covid-19 data visualization, and user's data visualization.

The secondary objective of this product is:

1. To support higher authorities in effective decision-making and proper management of the system.
2. To predict the Covid-19 cases via machine learning.
3. To assist people in discovering live Covid-19 cases and situations.
4. To analyze and summarize the user's dataset.

Figure 36. Title, aim, and objective

Specification of my Product:

The product specification is divided into functional and non-functional specifications, and the MoSCoW approach is defined below:

Functional Requirements	MoSCoW
Do the data analysis and visualization to bring meaningful information	M
Create a model to predict the data	M
Create an analytical report	M
Create a Dashboard	M
Create a web app for visualization	S
Segregation of data	S
Apply multiple machine learning algorithm	C
Can train or fine-tune the model	W

Non-functional Requirements	MoSCoW
The products and results are easy to visualize.	M
Comparison and validation of the product	M
The graph and charts results should give detailed knowledge about data.	S
The product should be platform-free.	C
Determines the best prediction model	w

Figure 37. System requirements

Research:

It's an artificial intelligence subfield centered around the idea that machines should learn from results, understand patterns, and make decisions with little or no human intervention. (SAS Institute Inc., n.d.) The process of manipulating, transforming, and visualizing data to obtain useful information from the results is known as data analysis. Individuals, corporations, and even governments often base their decisions on these observations. Regardless of how popular these machine-learning models are, humans are still required to conclude the outcomes of data analysis. It is up to us humans to make sense of the findings and decide how to clean the data. (Udacity Team, 2020)

Evaluation:

The evaluation of the final product result will be based on the achievement of the objectives, as mentioned in the product specification and objectives. For evaluation, I will process and analyze data to obtain valuable knowledge concerning covid-19. I will then compare and visualize the data to get a better understanding of the dataset. Then, in the model, I will use a relevant and suitable algorithm to predict the outcome, followed by time series analysis to obtain precise outcomes, and then create an analytical report, dashboard, and a web app for better visualization.

Figure 38. Research and Evaluation

Project Planning & Methodology

Project Timeline:

Fig: Above figure represents timeline of the project

Gantt Chart:

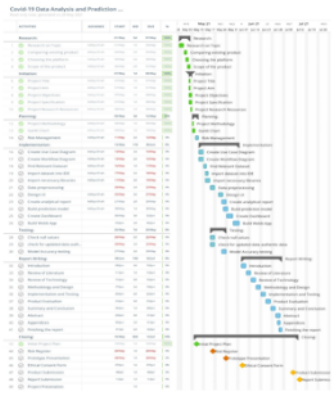


Fig: The Gantt chart above portrays project preparation as well as key dates.

Figure 39. Project Planning

Methodology:

After importing the Dataset, I will do data processing and exploratory data analysis to visualize data. Later, I will then train the machine-learning model with various algorithms, and later perform series analysis to predict the data. I will later create an analytical report, dashboard, and web app for interactive visualization. The Agile methodology will be used for the project, and the planning is further classified using the Project Timeline and Gantt chart.

Resources

The hardware and software I require to complete my Project successfully:

List of Software and Libraries:

- Python(PyCharm)
- Google Colaboratory
- Pandas-profiling
- Matplotlib
- Plotly
- GitHub
- Streamlit
- NumPy
- Scikit-learn
- Figma
- Google data studio

List of Hardware:

- Acer Aspire 5 Laptop
- NVidia Mx250 graphics

Figure 40. Methodology and Resources

Human Resource

I am working on my Project with the following people

Name: Aditya Shah	**Role:**
	Module Leader
	Supervisor: Shiva Prasad Nepal

Initial Bibliography

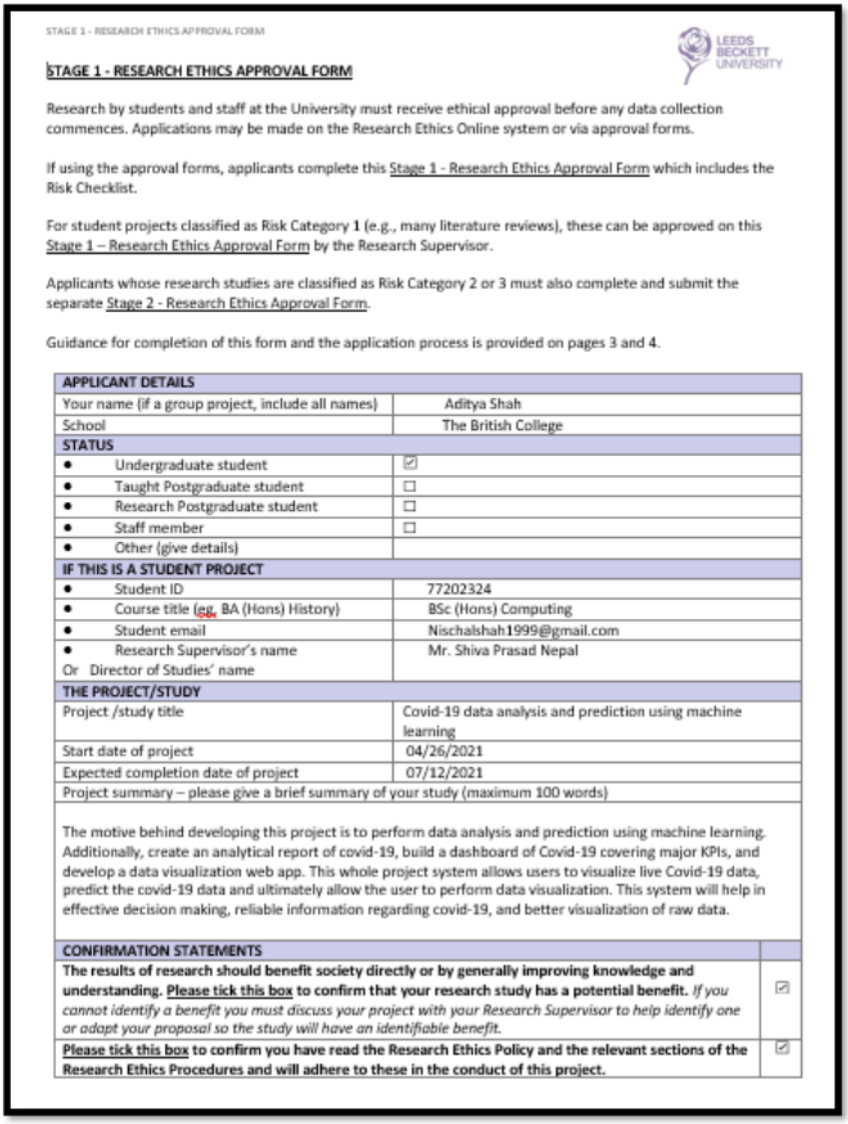
Bibliography

- SAS Institute Inc., n.d. *SAS.* [Online]
  Available at: https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.
  [Accessed 01 05 2021].
- Udacity Team, 2020. *UDACITY.* [Online]
  Available at: https://www.udacity.com/blog/2020/08/machine-learning-for-data-analysis.html
  [Accessed 1 5 2021].
- E. Gambhir, R. Jain, A. Gupta and U. Tomer, "Regression Analysis of COVID-19 using Machine Learning Algorithms," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 65-71, doi: 10.1109/ICOSEC49089.2020.9215356.
- Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. Chaos Solitons Fractals. 2021 Jan;142:110512. doi: 10.1016/j.chaos.2020.110512. Epub 2020 Nov 28. PMID: 33281306; PMCID: PMC7698672.
- A. Abdullha and S. Abujar, "COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective," 2020 15th International Joint

Figure 41. Human resource & Initial bibliography

Appendix B

Ethical Approval



STAGE 1 - RESEARCH ETHICS APPROVAL FORM

LEEDS BECKETT UNIVERSITY

STAGE 1 - RESEARCH ETHICS APPROVAL FORM

Research by students and staff at the University must receive ethical approval before any data collection commences. Applications may be made on the Research Ethics Online system or via approval forms.

If using the approval forms, applicants complete this Stage 1 - Research Ethics Approval Form which includes the Risk Checklist.

For student projects classified as Risk Category 1 (e.g., many literature reviews), these can be approved on this Stage 1 – Research Ethics Approval Form by the Research Supervisor.

Applicants whose research studies are classified as Risk Category 2 or 3 must also complete and submit the separate Stage 2 - Research Ethics Approval Form.

Guidance for completion of this form and the application process is provided on pages 3 and 4.

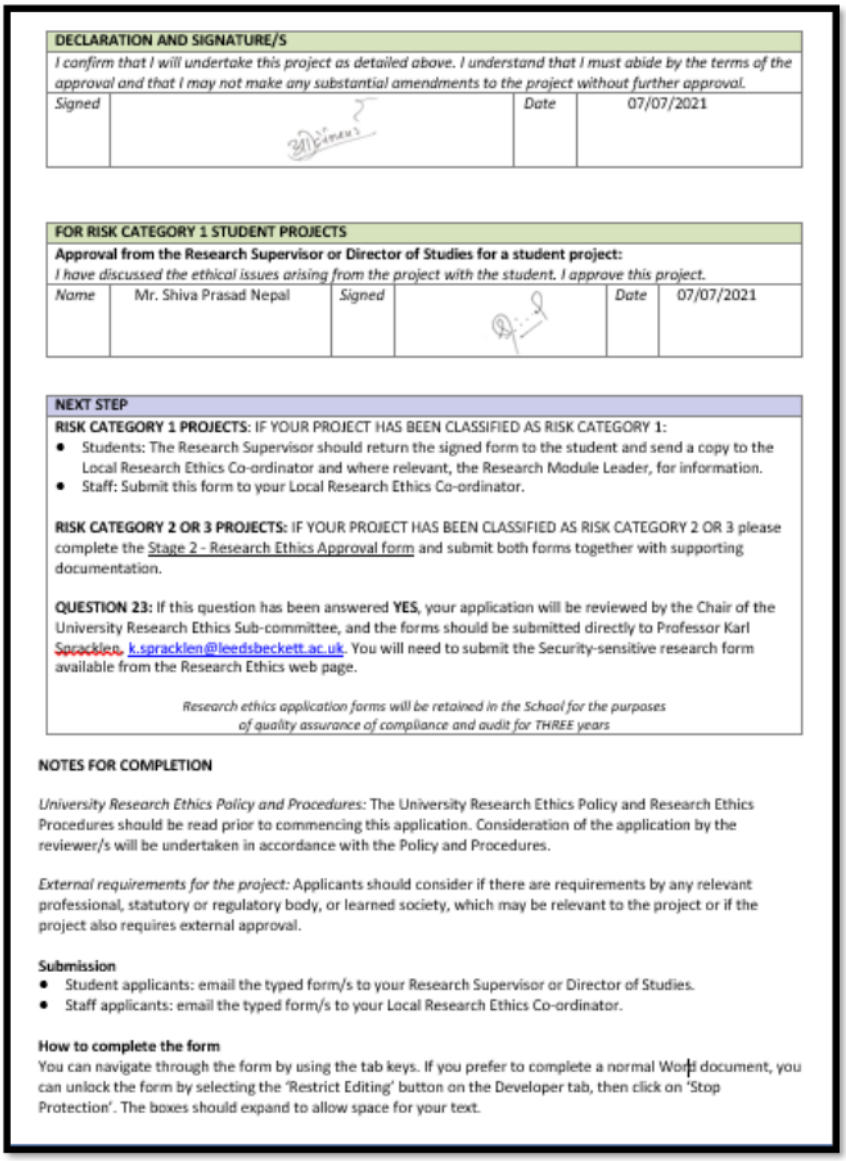
APPLICANT DETAILS	
Your name (if a group project, include all names)	Aditya Shah
School	The British College
STATUS	
• Undergraduate student	☑
• Taught Postgraduate student	☐
• Research Postgraduate student	☐
• Staff member	☐
• Other (give details)	
IF THIS IS A STUDENT PROJECT	
• Student ID	77202324
• Course title (eg. BA (Hons) History)	BSc (Hons) Computing
• Student email	Nischalshah1999@gmail.com
• Research Supervisor's name Or Director of Studies' name	Mr. Shiva Prasad Nepal
THE PROJECT/STUDY	
Project /study title	Covid-19 data analysis and prediction using machine learning
Start date of project	04/26/2021
Expected completion date of project	07/12/2021
Project summary – please give a brief summary of your study (maximum 100 words)	
The motive behind developing this project is to perform data analysis and prediction using machine learning. Additionally, create an analytical report of covid-19, build a dashboard of Covid-19 covering major KPIs, and develop a data visualization web app. This whole project system allows users to visualize live Covid-19 data, predict the covid-19 data and ultimately allow the user to perform data visualization. This system will help in effective decision making, reliable information regarding covid-19, and better visualization of raw data.	
CONFIRMATION STATEMENTS	
The results of research should benefit society directly or by generally improving knowledge and understanding. Please tick this box to confirm that your research study has a potential benefit. *If you cannot identify a benefit you must discuss your project with your Research Supervisor to help identify one or adapt your proposal so the study will have an identifiable benefit.*	☑
Please tick this box to confirm you have read the Research Ethics Policy and the relevant sections of the Research Ethics Procedures and will adhere to these in the conduct of this project.	☑

Figure 42. Applicant details and confirmation statement

RISK CHECKLIST - Please answer ALL the questions in each of the sections below – tick YES or NO WILL YOUR RESEARCH STUDY.....?		YES	NO
1	Involve direct and/or indirect contact with human participants?	☐	☑
2	Involve analysis of pre-existing data which contains personal or sensitive information not in the public domain?	☐	☑
3	Require permission or consent to conduct?	☐	☑
4	Require permission or consent to publish?	☐	☑
5	Have a risk of compromising confidentiality?	☐	☑
6	Have a risk of compromising anonymity?	☐	☑
7	Collect / contain sensitive personal data?	☐	☑
8	Contain elements which you OR your supervisor are NOT trained to conduct?	☐	☑
9	Use any information OTHER than that which is freely available in the public domain?	☐	☑
10	Involve respondents to the internet or other visual/vocal methods where participants may be identified?	☐	☑
11	Include a financial incentive to participate in the research?	☐	☑
12	Involve your own students, colleagues or employees?	☐	☑
13	Take place outside of the country where you are enrolled as a student, or for staff, outside of the UK?	☐	☑
14	Involve participants who are particularly vulnerable or at risk?	☐	☑
15	Involve any participants who are unable to give informed consent?	☐	☑
16	Involve data collection taking place BEFORE informed consent is given?	☐	☑
17	Involve any deliberate deception or covert data collection?	☐	☑
18	Involve a risk to the researcher or participants beyond that experienced in everyday life?	☐	☑
19	Cause (or could cause) physical or psychological harm or negative consequences?	☐	☑
20	Use intrusive or invasive procedures?	☐	☑
21	Involve a clinical trial?	☐	☑
22	Involve the possibility of incidental findings related to health status?	☐	☑
23	Fit into any of the following security-sensitive categories: concerns terrorist or extreme groups; commissioned by the military; commissioned under an EU security call; involves the acquisition of security clearances? If yes, see the guidance.	☐	☑

CLASSIFICATION **The following guidance will help classify the risk level of your study**	Tick the box which applies to your project
If you answered **NO** to all the above questions, your study is provisionally classified as **Risk Category 1** (literature reviews will be Risk Category 1).	☑
If you answered **YES** to any question from 1-13 and **NO** to all questions 14-22, your study is provisionally classified as **Risk Category 2**.	☐
If you answered **YES** to any question from 14-22, your study is provisionally classified as **Risk Category 3**.	☐
If question 23 has been answered **YES**, your application will be reviewed by the Chair of the University Research Ethics Sub-committee	☐

Figure 43. Risk Checklist



DECLARATION AND SIGNATURE/S

I confirm that I will undertake this project as detailed above. I understand that I must abide by the terms of the approval and that I may not make any substantial amendments to the project without further approval.

Signed		*Date*	07/07/2021

FOR RISK CATEGORY 1 STUDENT PROJECTS

Approval from the Research Supervisor or Director of Studies for a student project:

I have discussed the ethical issues arising from the project with the student. I approve this project.

Name	Mr. Shiva Prasad Nepal	*Signed*		*Date*	07/07/2021

NEXT STEP

RISK CATEGORY 1 PROJECTS: IF YOUR PROJECT HAS BEEN CLASSIFIED AS RISK CATEGORY 1:

- Students: The Research Supervisor should return the signed form to the student and send a copy to the Local Research Ethics Co-ordinator and where relevant, the Research Module Leader, for information.
- Staff: Submit this form to your Local Research Ethics Co-ordinator.

RISK CATEGORY 2 OR 3 PROJECTS: IF YOUR PROJECT HAS BEEN CLASSIFIED AS RISK CATEGORY 2 OR 3 please complete the Stage 2 - Research Ethics Approval form and submit both forms together with supporting documentation.

QUESTION 23: If this question has been answered **YES**, your application will be reviewed by the Chair of the University Research Ethics Sub-committee, and the forms should be submitted directly to Professor Karl Spracklen, k.spracklen@leedsbeckett.ac.uk. You will need to submit the Security-sensitive research form available from the Research Ethics web page.

Research ethics application forms will be retained in the School for the purposes of quality assurance of compliance and audit for THREE years

NOTES FOR COMPLETION

University Research Ethics Policy and Procedures: The University Research Ethics Policy and Research Ethics Procedures should be read prior to commencing this application. Consideration of the application by the reviewer/s will be undertaken in accordance with the Policy and Procedures.

External requirements for the project: Applicants should consider if there are requirements by any relevant professional, statutory or regulatory body, or learned society, which may be relevant to the project or if the project also requires external approval.

Submission

- Student applicants: email the typed form/s to your Research Supervisor or Director of Studies.
- Staff applicants: email the typed form/s to your Local Research Ethics Co-ordinator.

How to complete the form

You can navigate through the form by using the tab keys. If you prefer to complete a normal Word document, you can unlock the form by selecting the 'Restrict Editing' button on the Developer tab, then click on 'Stop Protection'. The boxes should expand to allow space for your text.

Figure 44. Declaration and signatures

Signatures
Electronic/typed signatures are acceptable for emailed forms, as the emails provide the audit trail for all parties' agreement and approval of the forms (e.g., student applicant →Research Supervisor →Local Research Ethics Co-ordinator).

Outcome
Applicants will be advised of the outcome of the application by receipt of the signed form from:

- The Research Supervisor or Director of Studies for Risk Category 1 student projects;
- The Local Research Ethics Co-ordinator or the School level group for Risk Category 2 and 3 projects.

YOU MAY ONLY BEGIN ANY DATA COLLECTION ONCE YOU RECEIVE NOTIFICATION THAT THE PROJECT HAS ETHICAL APPROVAL. If the circumstances of your research study change after approval it is your responsibility to revisit the Risk Checklist and complete a further application.

Advice
When completing the Stage 1 - Research Ethics Approval Form, if you are uncertain about the answer to any question, read the relevant section of the Research Ethics Procedures document, and if you are still unsure:

- if you are student, seek guidance from your Research Supervisor or Director of Studies;
- if you are a staff member, contact your Local Research Ethics Co-ordinator.

APPROVAL PROCESS

- Local Research Ethics Co-ordinator = LREC
- School level group (if your School uses a different review process, please follow your School guidance)
- University Research Ethics Sub-Committee = URESC

Category	Student applicants	Staff applicants
Risk Category 1	If your study has been provisionally classified as Risk Category 1, your Research Supervisor (or Director of Studies) can normally give approval for the project. You must complete this form and submit it to your Research Supervisor for consideration. A copy of the signed form if approved must be given or emailed to the LREC and, where relevant, the Research Module Leader, for information.	If your study has been classified as Risk Category 1, you do not need ethical approval for the project. You must complete the remainder of this form so that your research project is registered with the University. Please submit this form to your LREC.
Risk Category 2	If your study has been provisionally classified as Risk Category 2, your Supervisor (or Director of Studies) can recommend approval for your study by the LREC. You must complete this application form and also the separate Stage 2 - Research Ethics Approval form. Once you have completed the forms please submit both forms and supporting documentation to your Research Supervisor for consideration. Your Supervisor may disagree with your assessment and ask you to make revisions or reject your application. When the Research Supervisor is happy to recommend the application for approval, they will send the forms to the LREC. The LREC will review your project and then decide to approve it, ask for revisions, reject it or pass it on for review by the School level group.	If your study has been provisionally classified as Risk Category 2, your project will be considered for ethical approval by the LREC. You must complete this application form and also the separate Stage 2 - Research Ethics Approval form. Please submit both forms and supporting documentation to your LREC for consideration. The LREC will review your project and then decide to approve it, ask for revisions or pass it on for review by the School level group.
Risk Category 3	**Postgraduate Research Students** If your study has been provisionally classified as Risk Category 3, your Supervisor or Director of Studies can recommend approval for your study by the LREC. You must complete this application form and also the separate Stage 2 - Research Ethics Approval form and submit both forms to your Director of Studies. If your Director of Studies recommends approval of your project they will refer it to the LREC who will review your project and decide whether to grant ethical approval, request revisions, reject the application or refer it to the School level group for review. **Undergraduate and Taught Postgraduate Students** If your study has been provisionally classified as Risk Category 3, you should consult with your Research Supervisor immediately as it is unlikely you will be able to proceed and you should negotiate a project that is of lower risk. However, if you have already discussed the project with your Supervisor and they have agreed that a case for approval is warranted, proceed in line with the details above for Research Students.	If your study has been provisionally classified as Risk Category 3, your project will be considered for ethical approval by an appropriate LREC. You must complete this application form and also the separate Stage 2 - Research Ethics Approval form and submit both forms with supporting documentation to your LREC. The LREC will review your project and then decide to approve it, ask for revisions or pass it on for review by the School level group.
Q23	If question 23 has been answered 'yes', your application will be reviewed by the Chair of the University Research Ethics Sub-committee. The answer does not affect the Risk Category.	

Figure 45. Approval Process

Appendix C

Product Testing

Table 8. Null value testing

ID	1
Test	Null value
objectives	To check the data if it contains a null value
precondition	Data should be preprocessed
Expected output	The dataset should not contain any null value
output	
Result	Test pass

Table 9. Accuracy testing

ID	2
Test	Accuracy
objectives	To check the accuracy of the machine learning model
precondition	Train and apply a machine learning algorithm
Expected output	Achieve accuracy above 90%
output	
Result	Test pass

Table 10. Countries testing

ID	3
Test	Select one or multiple countries
objectives	To check if it displays selected country data
precondition	Users should select a country from the dropdown menu
Expected output	Specific data of a particular country
output	
Result	Test pass

Table 11. Date testing

ID	4
Test	Select a date
objectives	To check whether it provides selected date Covid-19 data
precondition	Users should select a date from a dropdown menu
Expected output	Covid-19 data of selected date
output	
Result	Test pass

Table 12. Case testing

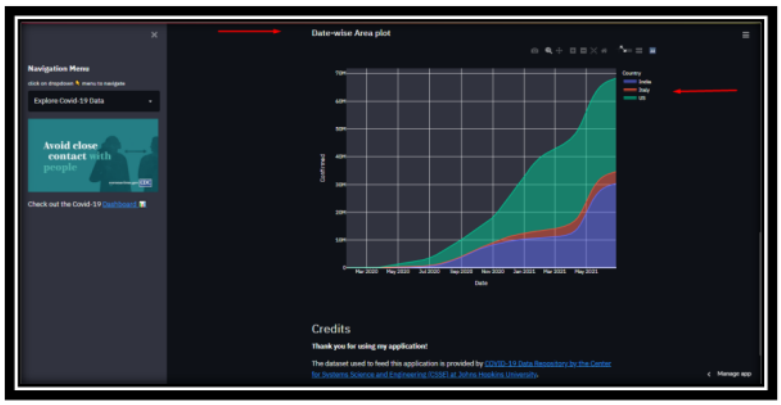
ID	5
Test	Select a country's case type
objectives	To check whether it displays a chart of selected country's case
precondition	Users should select a case type from a dropdown menu
Expected output	Graphs and plots of selected country's case
output	 

Table 13. File testing

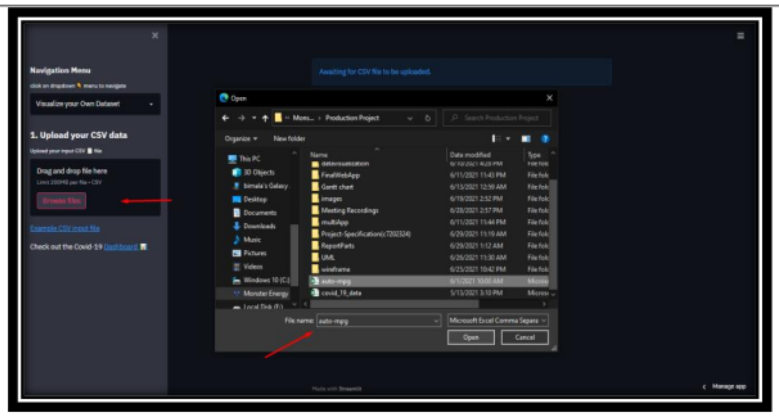
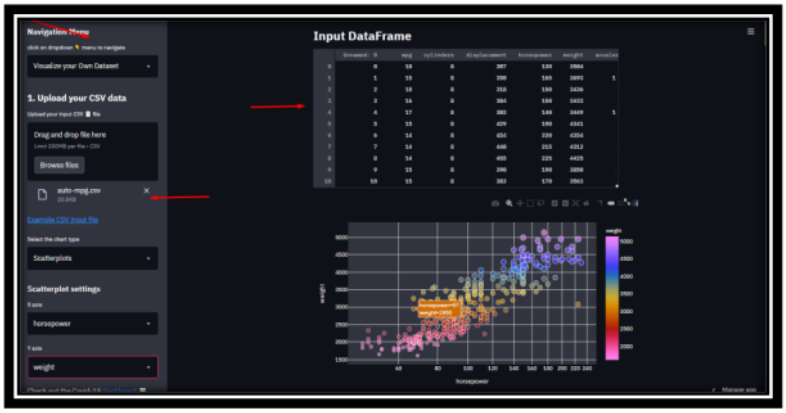
ID	6
Test	Upload a CSV file
objectives	To check whether it uploads the user CSV file
precondition	Users should drag or drop CSV files from a computer
Expected output	Dataframe of the uploaded file and its information
output	 
Result	Test pass

Table 14. Chart testing

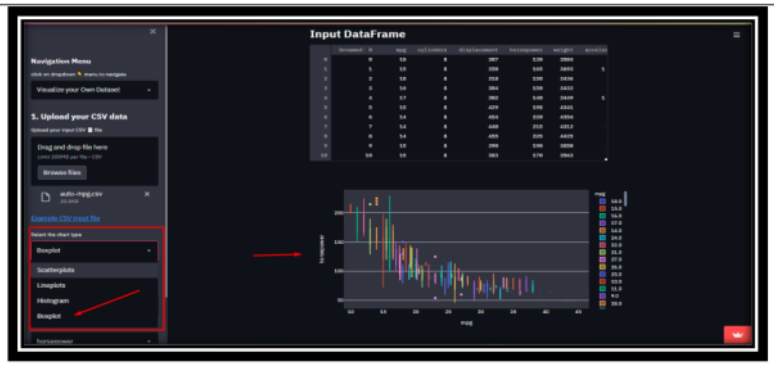
ID	7
Test	Chart selection
objectives	To check whether it renders the data along with a selected chart
precondition	Users should define the chart type from the chart selection dropdown
Expected output	Display data of x and y columns in charts or graphs
output	
Result	Test pass

Table 15. X and Y column testing

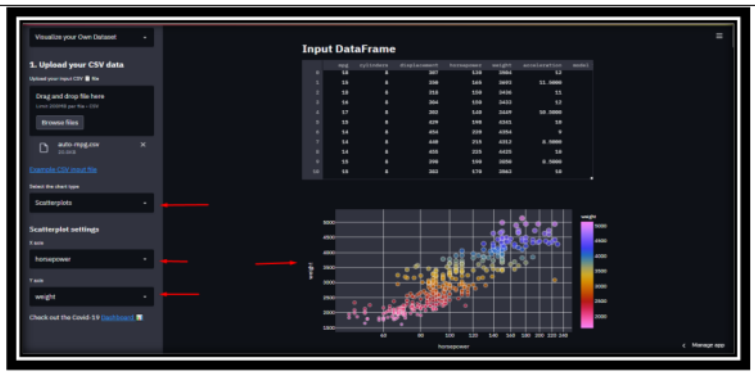
ID	8
Test	X and Y axis columns
objectives	To check whether it displays chart as per x and y columns
precondition	Users should define the variable for x and y columns
Expected output	Display data of x and y columns in charts or graphs
output	
Result	Test pass

Table 16. Report testing

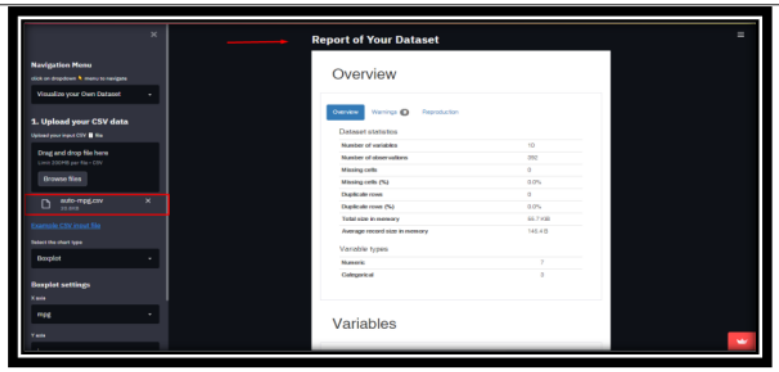
ID	9
Test	Overview of report
objectives	To check whether it generates the report of users uploaded dataset
precondition	Users should upload a CSV file
Expected output	Display the report of an uploaded dataset
output	
Result	Test pass

Table 17. Prediction testing

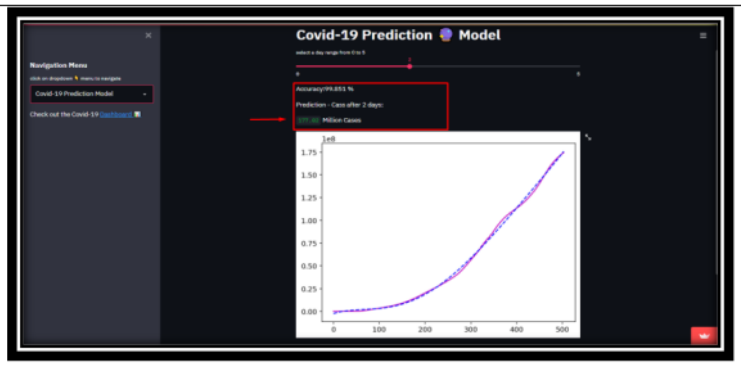
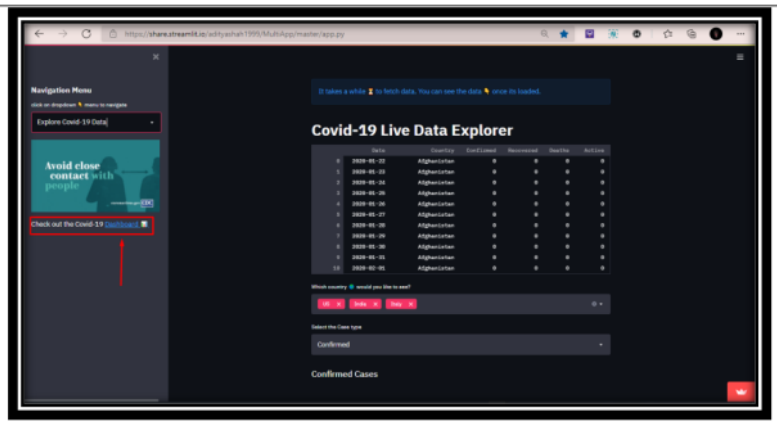
ID	10
Test	Prediction
objectives	To check whether it predicts the covid-19 cases
precondition	Users should select the number of days for prediction
Expected output	Displays the predicted outcome with a graph
output	
Result	Test pass

Table 18. Links testing

ID	11
Test	Working links
objectives	To check whether it redirects from one site to another
precondition	Users should select click on the links and picture to redirect
Expected output	Move from one site to another site efficiently
output	 
Result	Test pass

Appendix D

Meeting Records sheet

School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 1
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-05-16		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Discussion about my topic	Completed
2	Discussion about project-related issues	Completed
3	Discussion about requirements	Completed
4	Discussion about submissions date	Completed
5		
6		
Comments of student (if any):		
***ABOVE here** – student to complete before Meeting with supervisor. **BELOW here** – complete at the Meeting.*		
Next meeting (date/time)**:** 2021-05-20		
Agreed Actions to complete before next meeting:		
1	Risk register	
2	Project Initial Plan	
3		
4		
5		
6		
Comments of supervisor (if any):		

Figure 46. Meeting record 1

School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 2
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-05-20		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Risk Register	Completed
2	Project Initial Plan	Completed
3		
4		
5		
6		
Comments of student (if any):		

***ABOVE here** – student to complete before Meeting with supervisor. **BELOW here** – complete at the Meeting.*

Next meeting (date/time): 2021-05-27		
Agreed Actions to complete before next meeting:		
1	Prototype Presentation	
2	Design of product(use case diagram, wireframe)	
3	Marking scheme	
4		
5		
6		
Comments of supervisor (if any):		

Figure 47. Meeting record 2

School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 3
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-05-27		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Prototype presentation	Completed
2	Marking scheme	Completed
3	Design of product(use case diagram, wireframe)	Completed
4		
5		
6		
Comments of student (if any):		
***ABOVE here** = student to complete before Meeting with supervisor. **BELOW here** = complete at the Meeting.*		
Next meeting (date/time): 2021-06-04		
Agreed Actions to complete before next meeting:		
1	Introduction of Report	
2	Report Structure	
3		
4		
5		
6		
Comments of supervisor (if any):		

Figure 48. Meeting record 3

School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 4
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-06-04		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Introduction of report	Completed
2	Report Structure	Completed
3		
4		
5		
6		
Comments of student (if any):		
***ABOVE here** – student to complete before Meeting with supervisor. **BELOW here** – complete at the Meeting.*		
Next meeting (date/time)**:** 2021-06-15		
Agreed Actions to complete before next meeting:		
1	Literature Review	
2	Improvement in Introduction	
3	Review of technology	
4	Finalizing product	
5		
6		
Comments of supervisor (if any):		

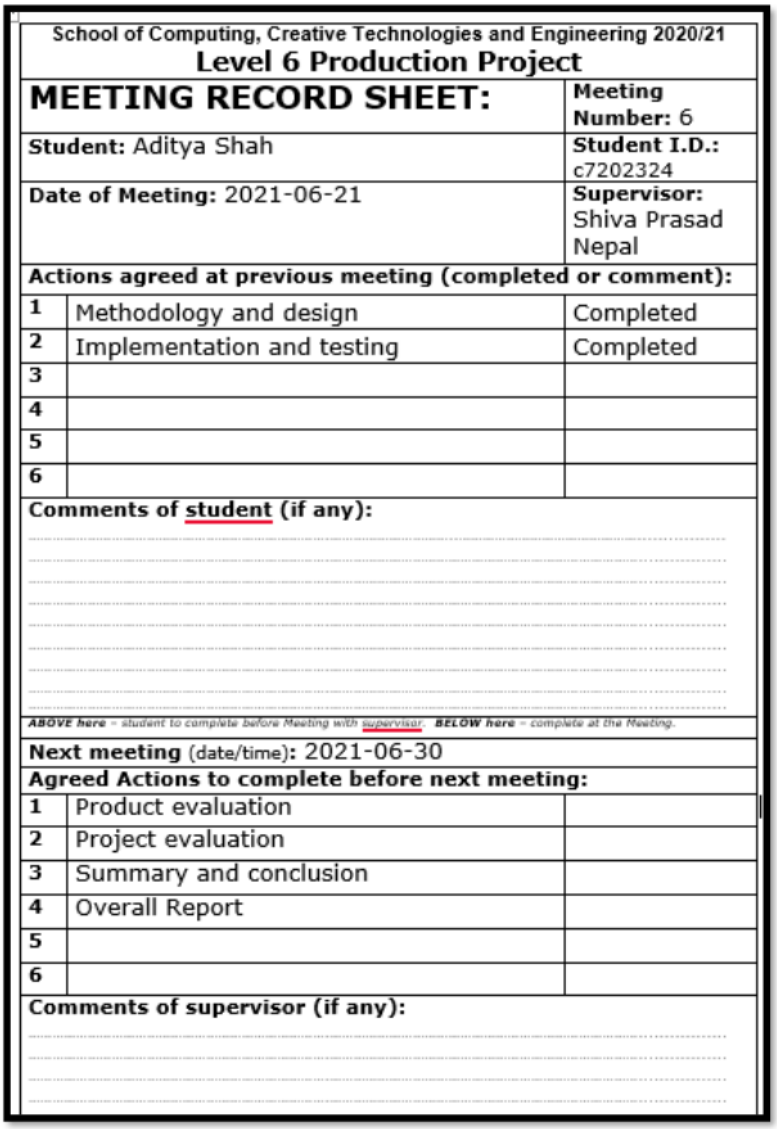
Figure 49. Meeting record 4

School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 5
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-06-15		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Literature Review	Completed
2	Improvement in Introduction	Completed
3	Review of Technology	Completed
4	Finalizing product	Completed
5		
6		
Comments of student (if any):		
***ABOVE here** = student to complete before Meeting with supervisor. **BELOW here** = complete at the Meeting.*		
Next meeting (date/time)**:** 2021-06-21		
Agreed Actions to complete before next meeting:		
1	Methodology and Design	
2	Implementation and Testing	
3		
4		
5		
6		
Comments of supervisor (if any):		

Figure 50. Meeting record 5



School of Computing, Creative Technologies and Engineering 2020/21

Level 6 Production Project

MEETING RECORD SHEET:		**Meeting Number:** 6
Student: Aditya Shah		**Student I.D.:** c7202324
Date of Meeting: 2021-06-21		**Supervisor:** Shiva Prasad Nepal
Actions agreed at previous meeting (completed or comment):		
1	Methodology and design	Completed
2	Implementation and testing	Completed
3		
4		
5		
6		
Comments of student (if any):		
***ABOVE here** – student to complete before Meeting with supervisor. **BELOW here** – complete at the Meeting.*		
Next meeting (date/time)**:** 2021-06-30		
Agreed Actions to complete before next meeting:		
1	Product evaluation	
2	Project evaluation	
3	Summary and conclusion	
4	Overall Report	
5		
6		
Comments of supervisor (if any):		

Figure 51. Meeting record 6

Appendix E

Risk Register

ID	Risk Description	Likelihood	Impact	Severity	Owner	Mitigation	Status
1	Inaccurate model due to various external factors	Medium	High	High	Aditya Shah	Training and fine-tuning the model	Open
2	Exposure to new IDE and programming language	Low	High	Medium	Aditya Shah	Learn the basics of the language and be able to adapt	closed
3	Unable to build a web-based app for visualization	Medium	Low	Low	Aditya Shah	Time management	open
4	Covid-19 and pandemic	High	Medium	High	Aditya Shah	Rely on online resources and tools and take preventive measures	open
5	Hardware and Network issue	Medium	High	High	Aditya Shah	Proper use of hardware tools, internet, and laptops	open
6	Unreliable Data source	Medium	Medium	High	Aditya Shah	Find the relevant dataset to work on	open
7	Health issues	Medium	Medium	Medium	Aditya Shah	Stay healthy and complete the project before the deadline	Open
8	Product Life	Medium	Medium	High	Aditya Shah	Let user visualize their own dataset and not only limited to Covid-19 data.	open

Figure 52. Risk register

Appendix F

User Guidelines For Data Visualization web app

1. Users can visit the web app from the Covid-19 analytics report, Covid-19 dashboard, or by pasting the link https://share.streamlit.io/adityashah1999/MultiApp/master/app.py. on the web browser.

Figure 53. Link on the address bar

2. Whenever users first enter the homepage of looks like as shown in figure (see figure 2). The initial page is the 'Covid-19 data explorer' section.

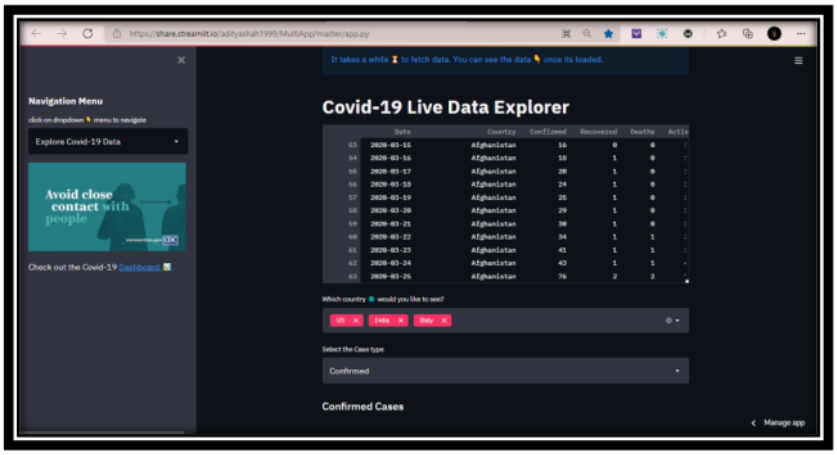


Figure 54. Homepage of web app

3. Figure 3 is the data frame of live covid-19 data. It is being fetched from the GitHub library. Users can view every detail of the data.

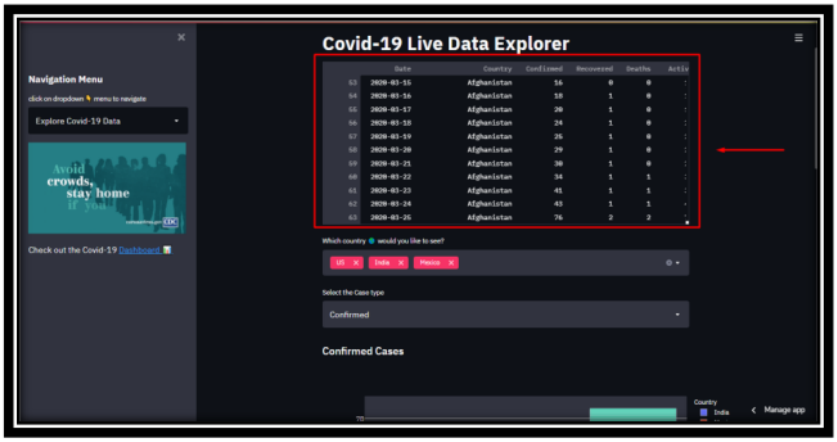


Figure 55. Raw data of Covid-19

4. Users can select one or multiple countries from the dropdown menu and for the case type users also can select the case from the dropdown menu for which they want to visualize the data.

Figure 56. Dropdown menu

5. After selecting the country and their case type from the drop-down menu for visualization. It shows the data in a bar chart and for time series data in an area plot as shown in figure(see figure 5)

Figure 57. Graphs and plots

6. Users can visit back to the Covid-19 Dashboard from the link given below the navigation menu.

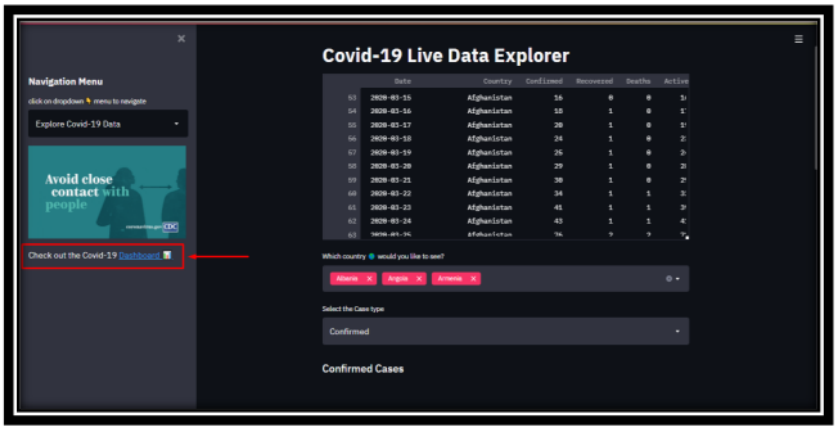


Figure 58. Dashboard link

7. Users can click on the navigation menu select the ‘Visualize your own Dataset’ section of the web app as shown in figure(see image 7).

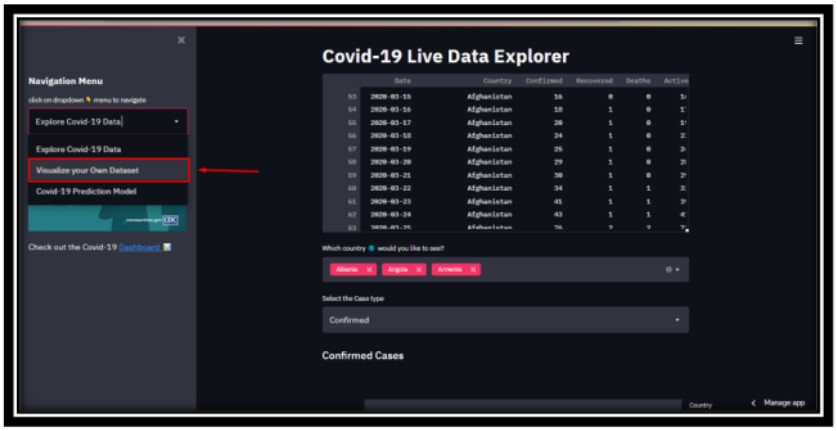


Figure 59. Visualize your data section

8. After clicking on the navigation dropdown option and selecting the ‘Visualize your own Dataset’ menu user is redirected to another page as shown in figure 8.

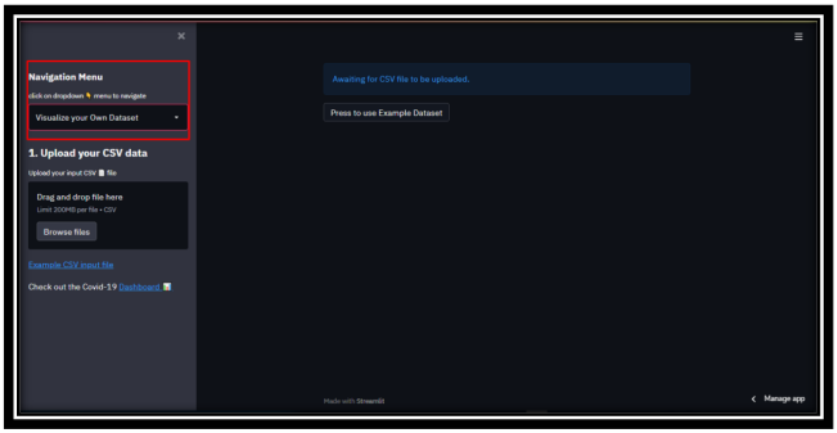


Figure 60. User-data visualization section

9. In the sidebar, a box is provided where the user can drag and drop any CSV files or just browse CSV files from the computer.

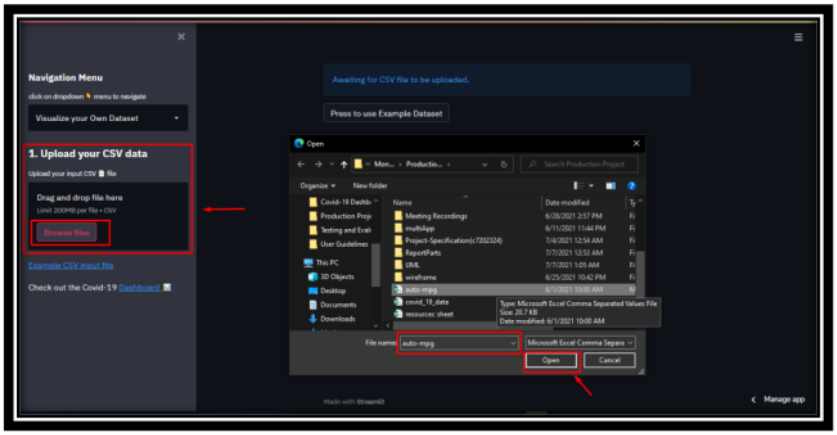


Figure 61. Upload CSV file

10. After uploading the CSV file user is presented with this screen(see figure 10). First at the top input data frame is presented and in the sidebar uploaded file name is given. All the dropdown menus are by default set and presented with the scatterplot.

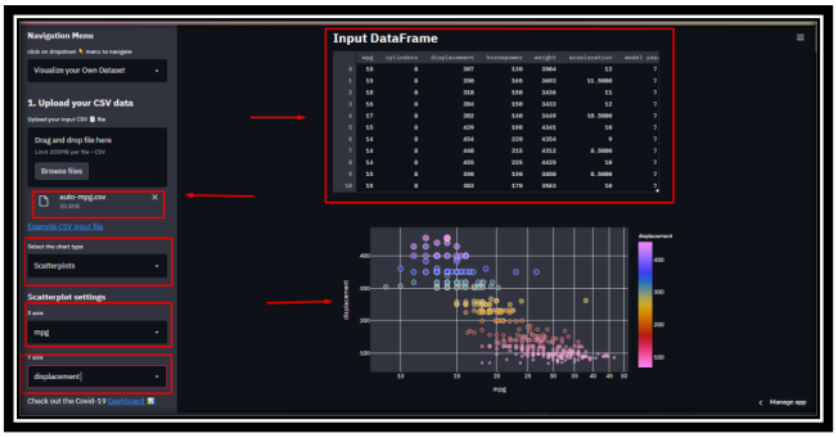


Figure 62. After uploading a CSV file

11. In the sidebar, the user can select the chart type in which they want to visualize the data. The four chart plot option given are scatterplot, line plot, histogram, and boxplot(see figure 11).

Figure 63. Chart options

12. After selecting the chart type. The user now has the option to set the value for the x and y column of graphs. The value of columns is determined by the variable available in the dataset file or input data frame. Here, the x value is chosen for horsepower(see figure 12) from the scatterplot settings.

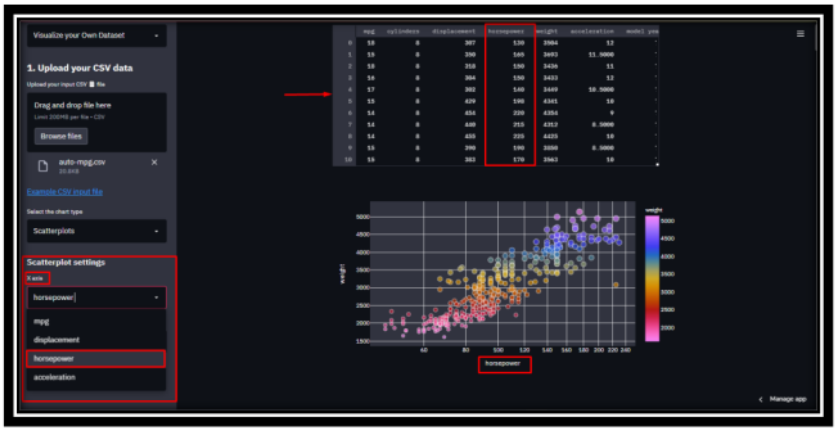


Figure 64. X columns

13. Similarly, for Y-axis the value is chosen from the sidebar ‘y-axis’ menu. Here the value of y is set to be the ‘weight’ column from the input data frame or CSV file and the scatterplot is represented with the weight variable in the y column(see figure 13).

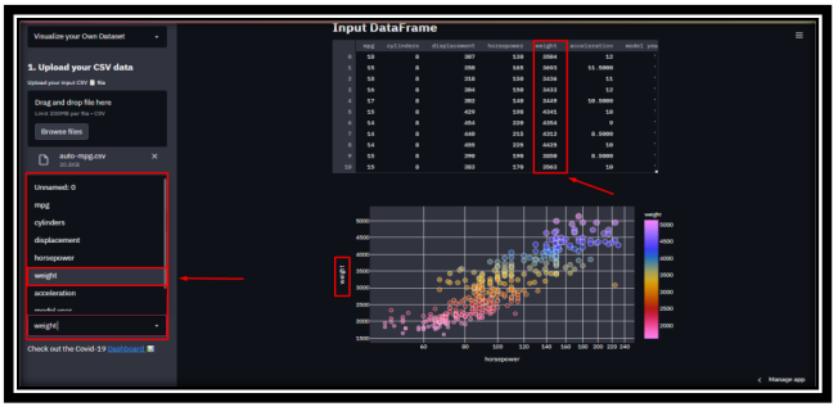


Figure 65. Y column

14. Whenever the user uploads the CSV dataset it automatically generates the overview of the report and the report is fully interactive as shown below(see figure 14).

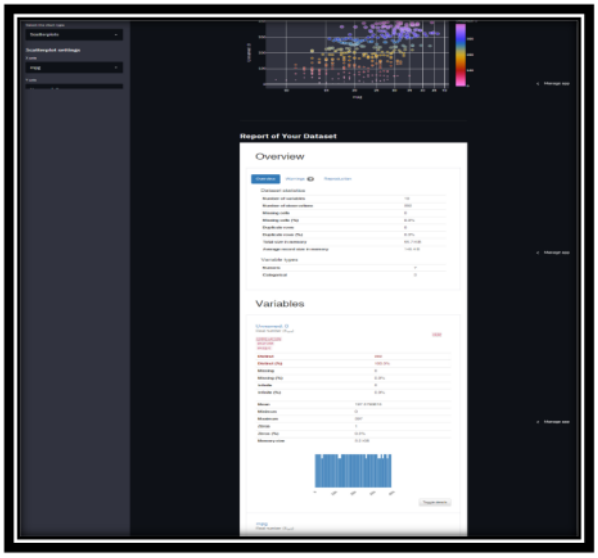


Figure 66. Overview of report

15. Users can also make the interaction and find correlations between variables in the report as shown in the figure (see figure 15).

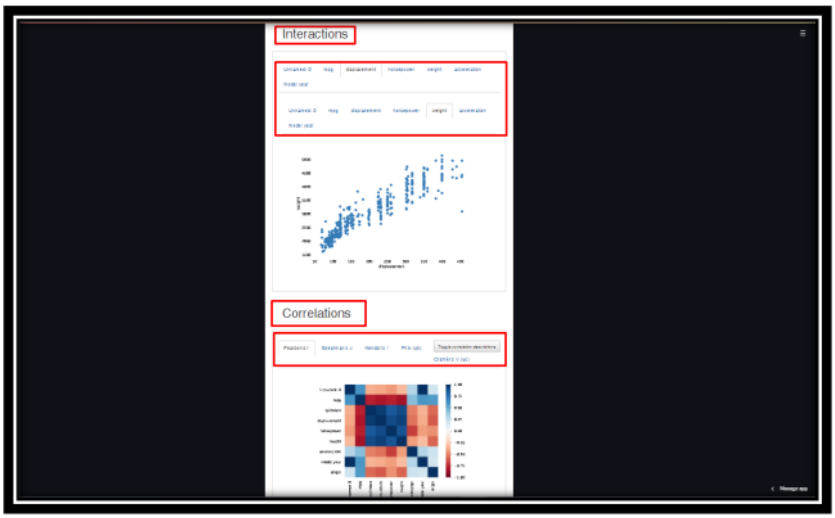


Figure 67. Interaction & correlation

16. By clicking on the link of the example dataset users can also use an example dataset that is built in the app to visualize the dataset. Similarly, clicking on the dashboard link below the example dataset will redirect the user to the covid-19 dashboard.

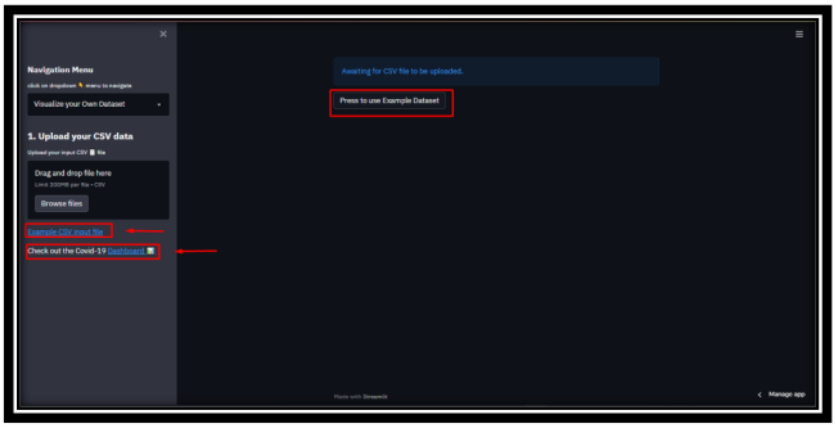


Figure 68. Example data & dashboard

17. Now to move the prediction section of the web user can simply navigate to the navigation menu and click the ‘Covid-19 Prediction Model’ section.

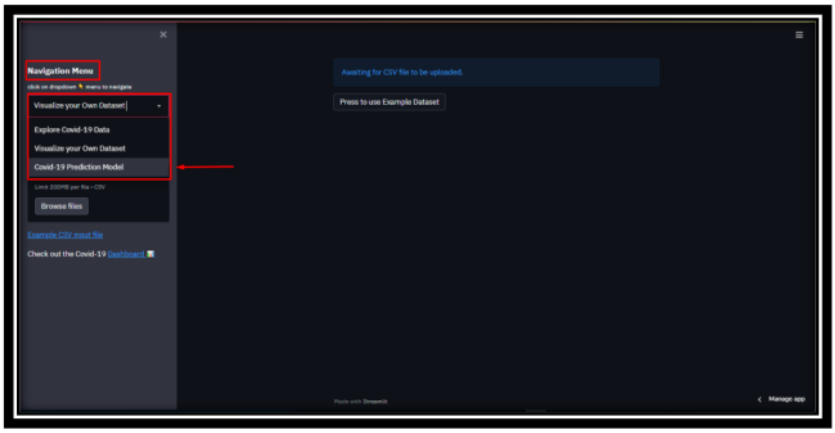


Figure 69. Prediction section

18. After clicking on the ‘Covid-19 Prediction Model’ menu from the navigation menu user is redirected to the prediction page as shown in figure (see figure 18).

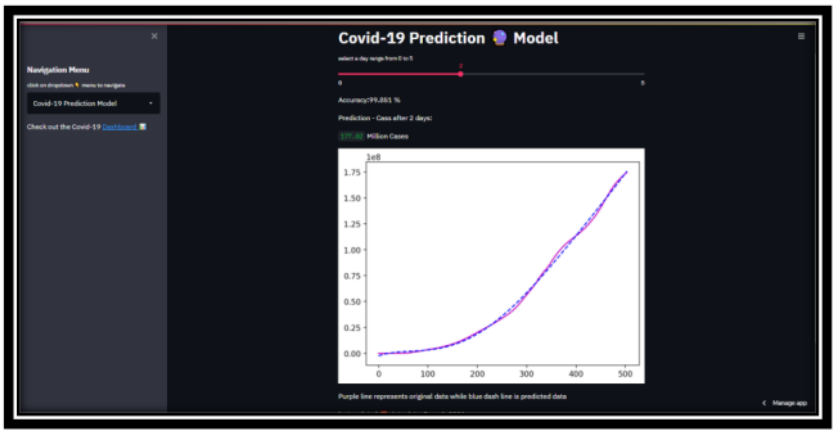


Figure 70.Prediction page

19. Users can select the number of days for the prediction of covid-19 cases from the slider. By default, data is predicted for 2 days. Right down below the slider, the accuracy and the number of cases are given along with the chart representing original data and predicted data.

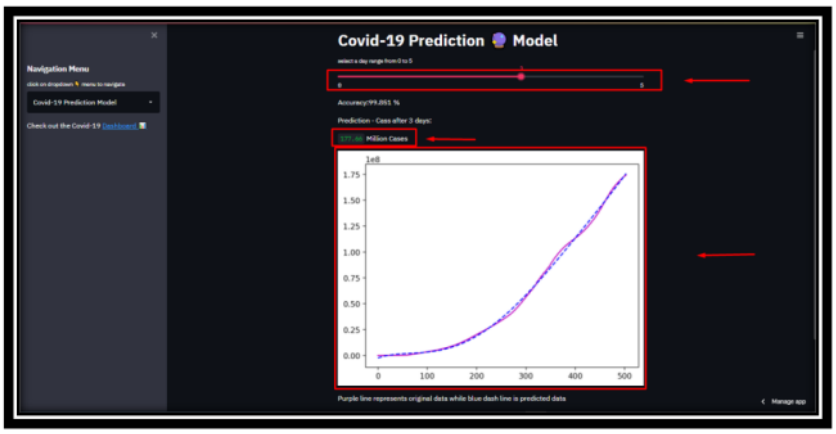


Figure 71. Covid-19 Prediction

20. Users can similarly redirect to the Covid-19 dashboard from the link given in the sidebar for a more interactive dashboard experience.

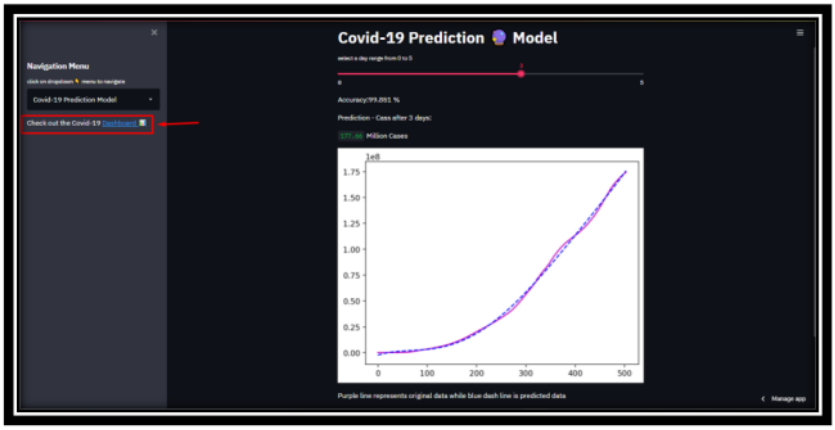


Figure 72. Link for Covid-19 dashboard

Appendix G

User Guidelines For Covid-19 Dashboard

1. Users can visit the Covid-19 Dashboard by visiting the https://datastudio.google.com/embed/reporting/84e28589-6d5a-443c-8a6f-1949473df87e/page/doJLC link or by entering it manually on the browsers address bar.

Figure 73. Dashboard URL

2. After entering on browsers address bar user will be redirected to the homepage of Covid-19 Dashboard.

Figure 74. Homepage of the dashboard

3. This section updates when there is an increase or decrease in the total number of cases worldwide or per selected country. The first scorecard covers total confirmed cases, the second covers recovered cases and the third contains death cases.

Figure 75. Confirmed, recovered, and death cases

4. Users can select a specific country or multiple countries to visualize the data. As per the user's input, all the graphical components change their data.

Figure 76. Select country

5. Whenever users select a specific country all graph and charts changes as pes user inputs as shown in figure(See figure 5). As we have selected India, Nepal, and the US it only provides the data of these countries.

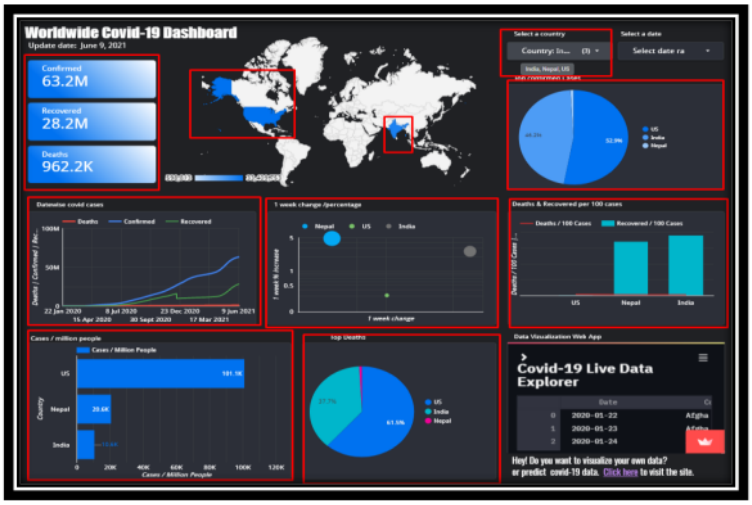


Figure 77. Country output

6. Users can select a specific date for which all graphical component tries to display the data of the selected date only.

Figure 78. Select date

7. Whenever a user selects a date the graphical components and their data or value change as shown in figure

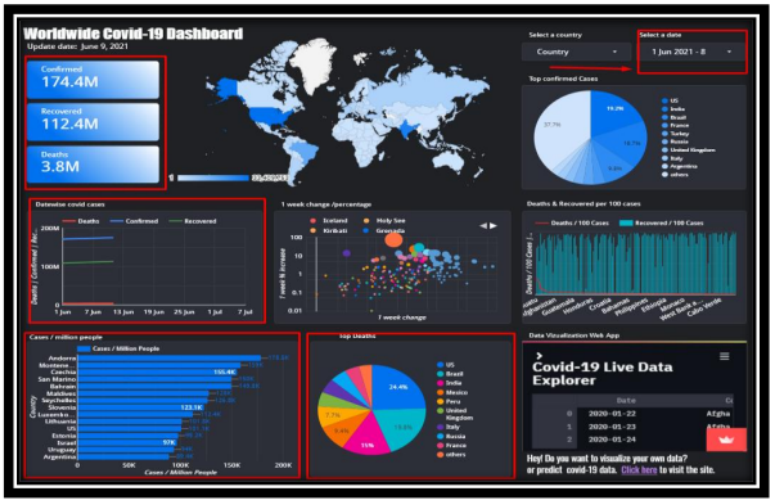


Figure 79. Date output

8. Users can download or export data of graphical components in various file formats as shown in the figure(see figure 7).

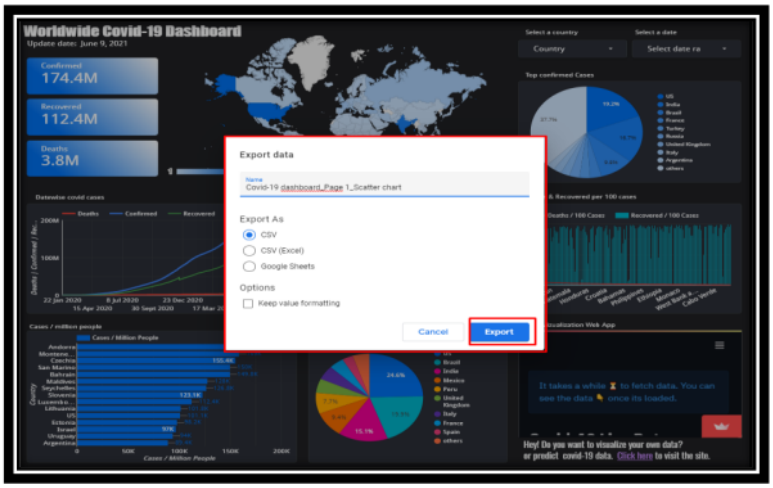


Figure 80. Export data

9. Users can redirect to the data visualization web app by clicking on the link ‘Click here’ under the dashboard. A graphical representation is shown for the web app. Users can take a peek at the app and can visit the web app by clicking on it(see figure 7).

Figure 81. Link for the web app

Appendix H

User Guidelines For Analytics report

1. Users can visit the site to view the analytics report of Covid-19 using the https://adityashah595728501.files.wordpress.com/2021/06/analyticsreport-1.pdf link.

Figure 82. Analytics report URL

2. After clicking on the link or visiting the site user is represented with a Covid-19 analytics report.

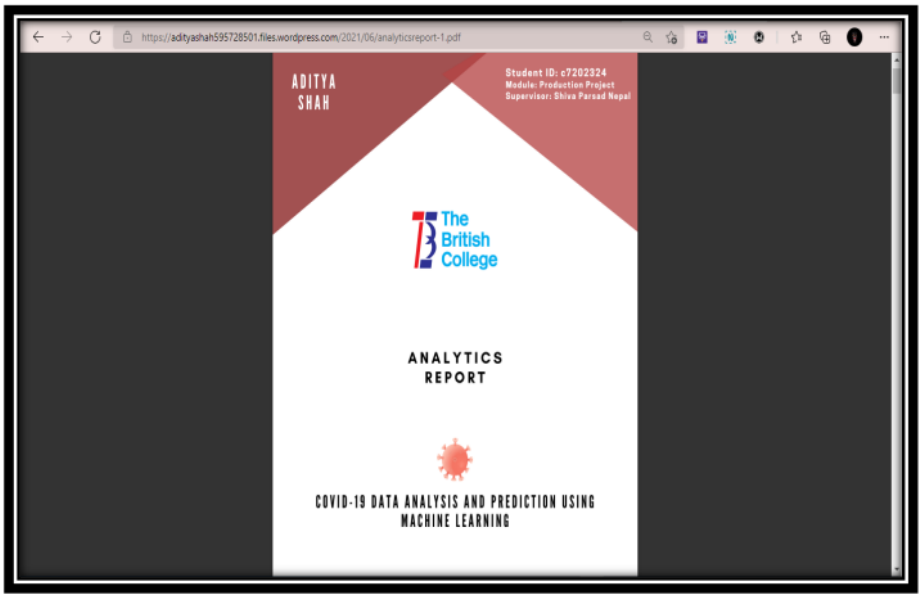


Figure 83. The first page of a report

3. Users can navigate through the table of contents and search for the desired topic to get information from.

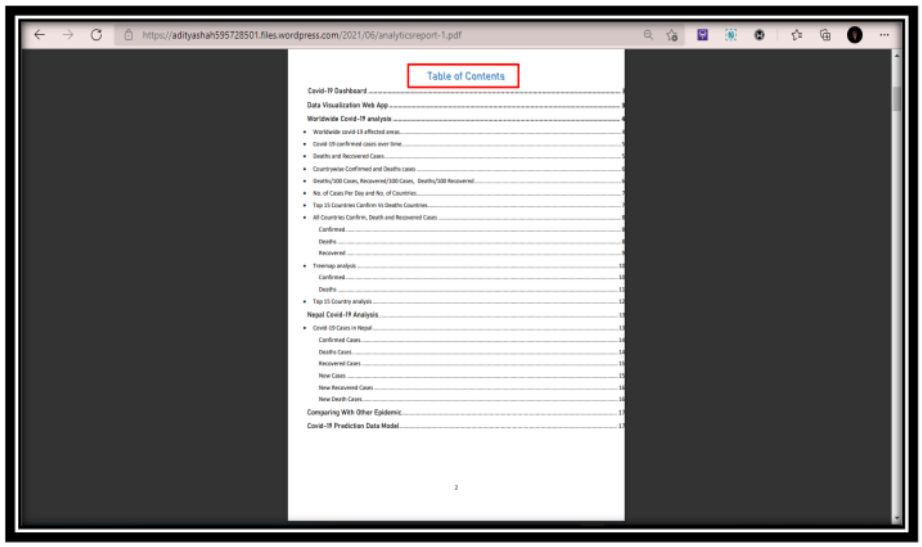


Figure 84. Table of contents

4. Users can download the Covid-19 analytics report from the save menu icon marked in the image.

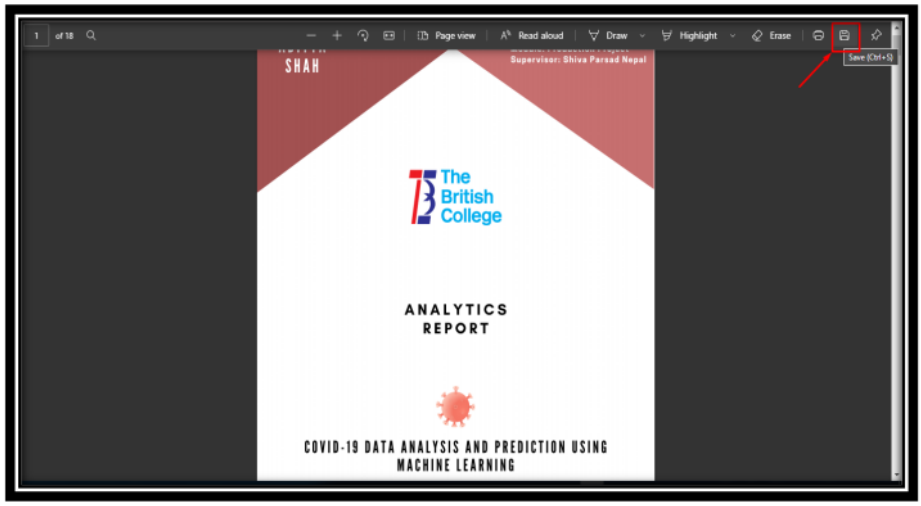


Figure 85. Save icon

5. Users can also print out the report of the Covid-19 analytics report to make hard copies by clicking on the print icon as shown in the figure.

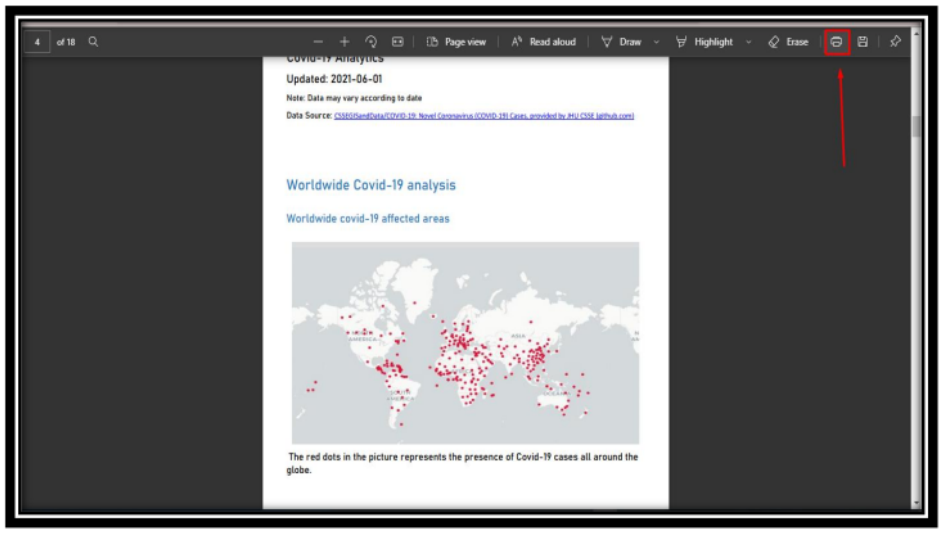


Figure 86. Print icon

6. Users can visit the Covid-19 dashboard and data visualization web app by clicking on the image or click as shown below.

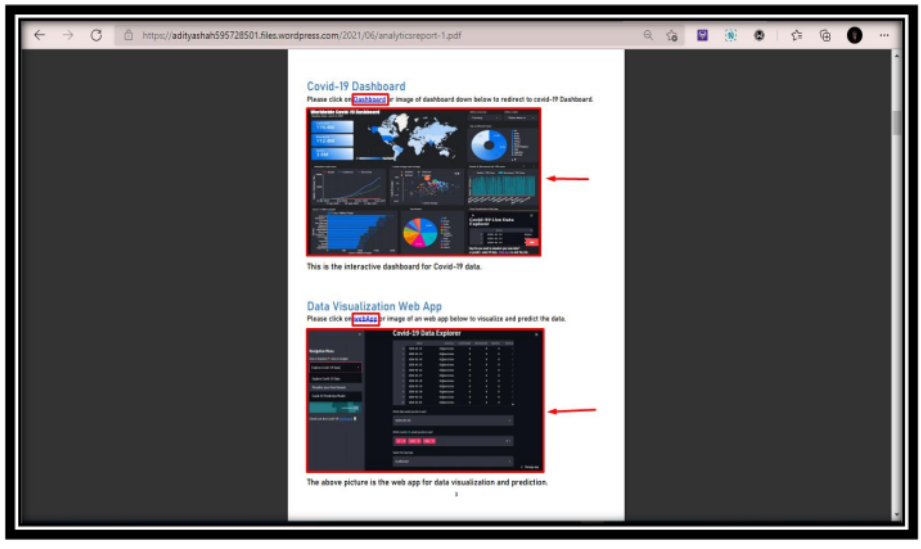
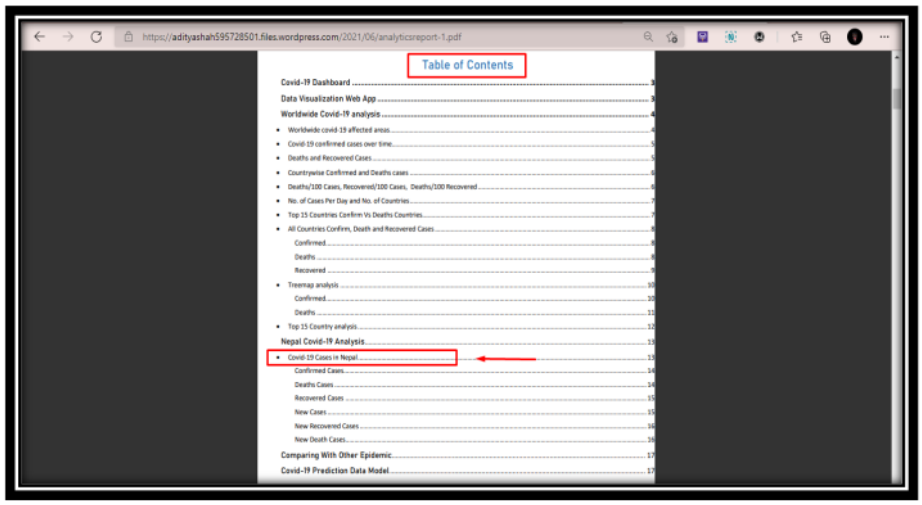


Figure 87. Links

7. To get information regarding Covid-19 cases of Nepal users can go to the table of contents and click ‘Ctrl+click’ on Covid-19 Cases in Nepal.



After clicking the user will be redirected to the Nepal Covid-19 section of the analytics report.

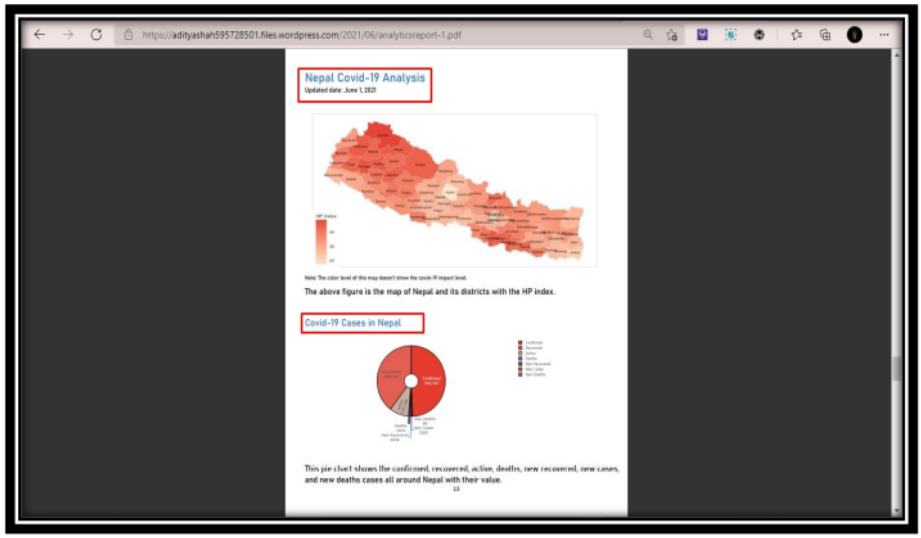


Figure 88. Nepal Covid-19 cases

8. Similarly, users can navigate the table of contents and visit the data source from where it was taken to perform data analysis and visualization.

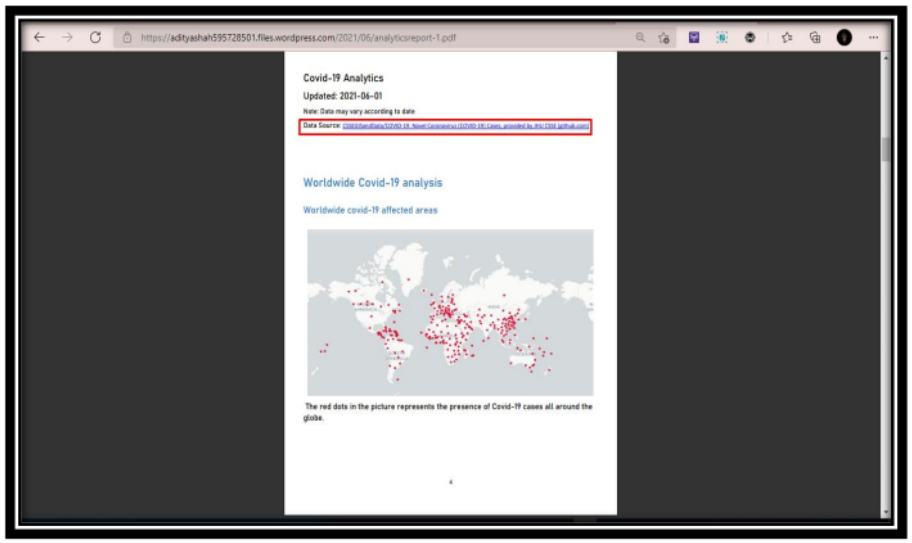


Figure 89. Data source

9. Users can see the chart and visualize information from any topics mentioned in the table contents of the report.

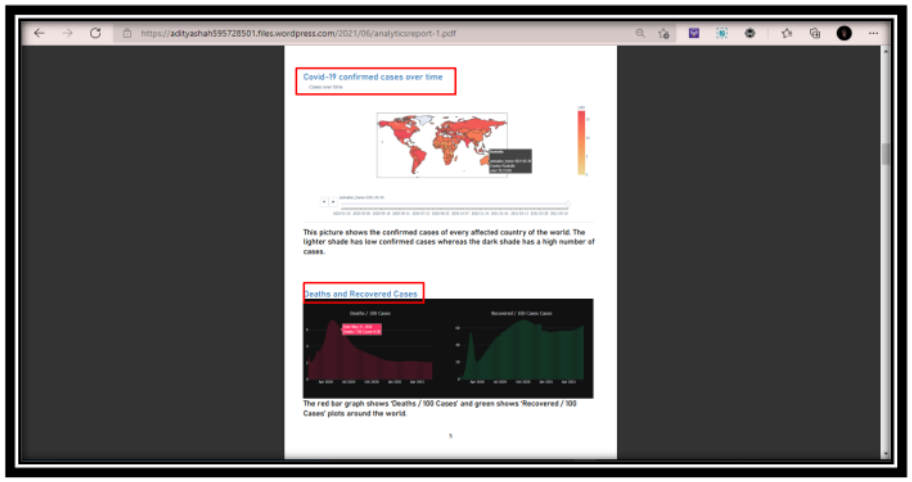
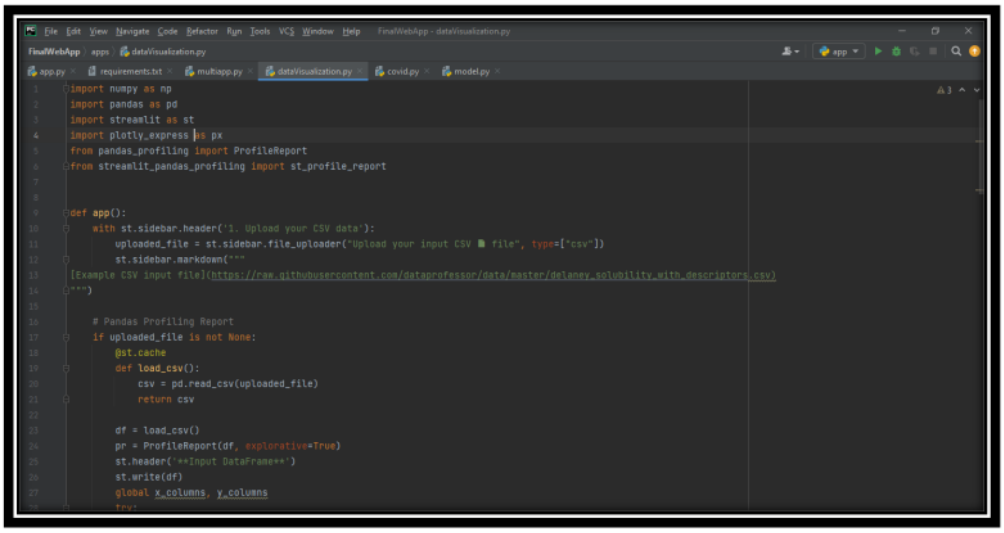


Figure 90. Covid-19 cases

Appendix I

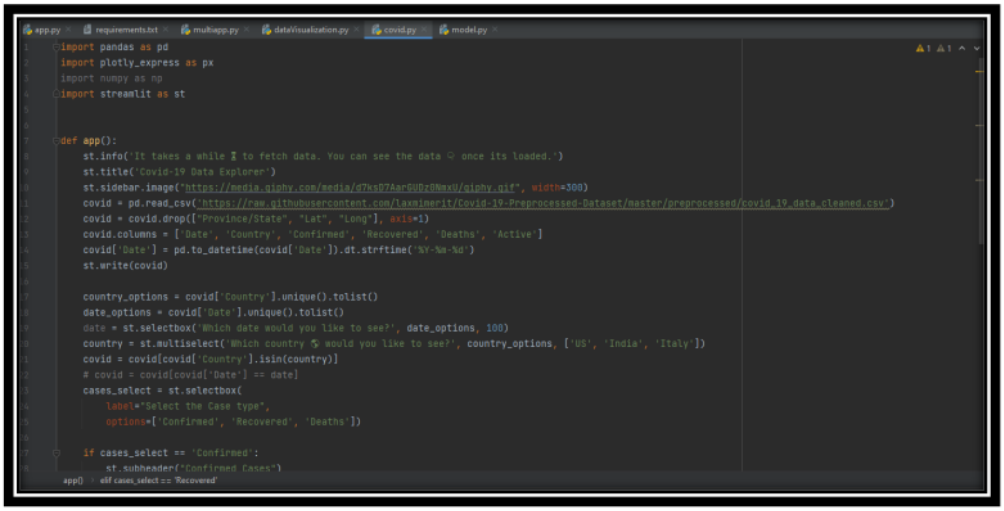
Source Code

Web App Source Code



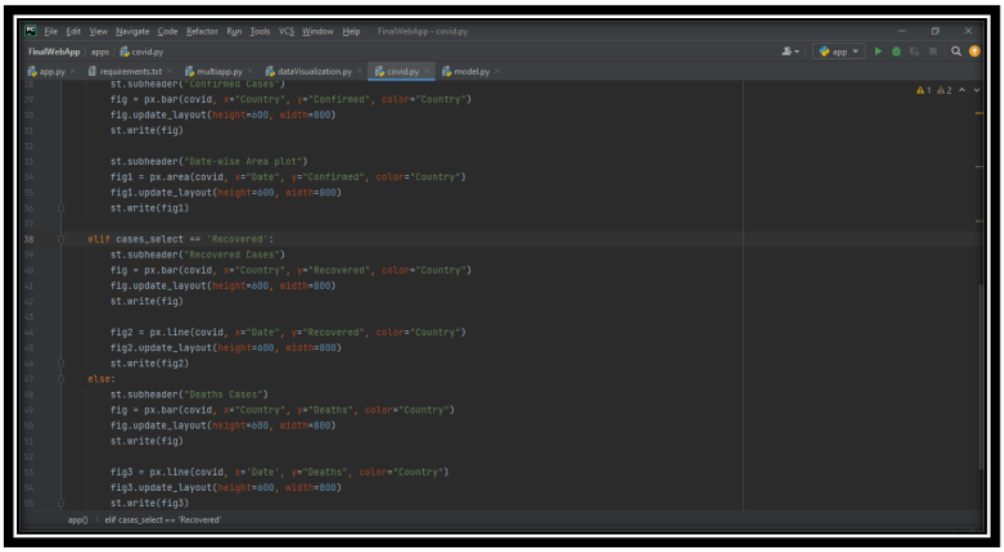
```
import numpy as np
import pandas as pd
import streamlit as st
import plotly_express as px
from pandas_profiling import ProfileReport
from streamlit_pandas_profiling import st_profile_report


def app():
    with st.sidebar.header('1. Upload your CSV data'):
        uploaded_file = st.sidebar.file_uploader("Upload your input CSV file", type=["csv"])
        st.sidebar.markdown("""
[Example CSV input file](https://raw.githubusercontent.com/dataprofessor/data/master/delaney_solubility_with_descriptors.csv)
""")

    # Pandas Profiling Report
    if uploaded_file is not None:
        @st.cache
        def load_csv():
            csv = pd.read_csv(uploaded_file)
            return csv

        df = load_csv()
        pr = ProfileReport(df, explorative=True)
        st.header('**Input DataFrame**')
        st.write(df)
        global x_columns, y_columns
        try:
```

Figure 91. User Data visualization code



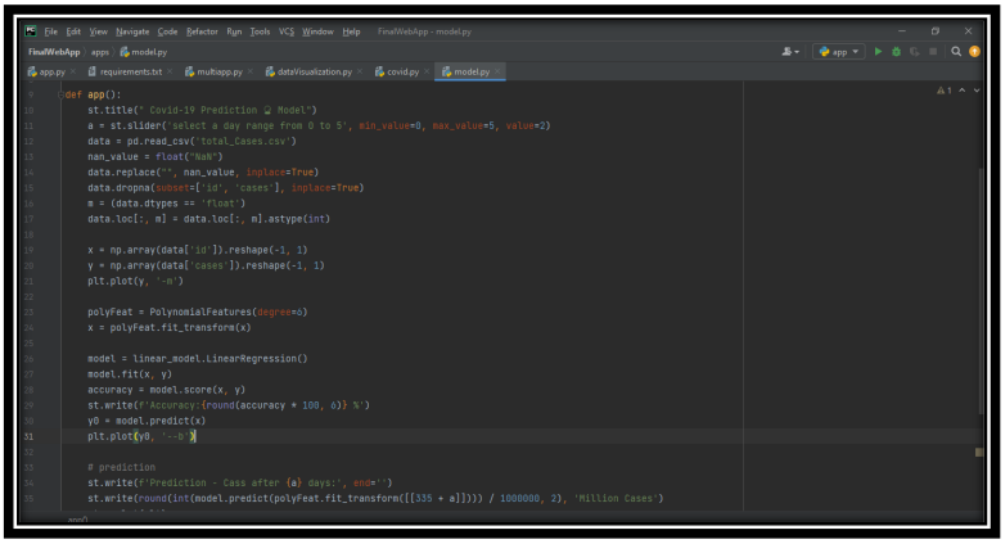
```
import pandas as pd
import plotly_express as px
import numpy as np
import streamlit as st


def app():
    st.info('It takes a while ⏳ to fetch data. You can see the data 👁 once its loaded.')
    st.title('Covid-19 Data Explorer')
    st.sidebar.image("https://media.giphy.com/media/d7ksD7AarGUDz0NmxU/giphy.gif", width=300)
    covid = pd.read_csv('https://raw.githubusercontent.com/laxminerit/Covid-19-Preprocessed-Dataset/master/preprocessed/covid_19_data_cleaned.csv')
    covid = covid.drop(['Province/State', "Lat", "Long"], axis=1)
    covid.columns = ['Date', 'Country', 'Confirmed', 'Recovered', 'Deaths', 'Active']
    covid['Date'] = pd.to_datetime(covid['Date']).dt.strftime('%Y-%m-%d')
    st.write(covid)

    country_options = covid['Country'].unique().tolist()
    date_options = covid['Date'].unique().tolist()
    date = st.selectbox('Which date would you like to see?', date_options, 100)
    country = st.multiselect('Which country 🌎 would you like to see?', country_options, ['US', 'India', 'Italy'])
    covid = covid[covid['Country'].isin(country)]
    # covid = covid[covid['Date'] == date]
    cases_select = st.selectbox(
        label='Select the Case type',
        options=['Confirmed', 'Recovered', 'Deaths'])

    if cases_select == 'Confirmed':
        st.subheader('Confirmed Cases')
app() › elif cases_select == 'Recovered'
```

Figure 92. Covid-19 Live data Code



```
        st.subheader('Confirmed Cases')
        fig = px.bar(covid, x="Country", y="Confirmed", color="Country")
        fig.update_layout(height=600, width=800)
        st.write(fig)

        st.subheader("Date-wise Area plot")
        fig1 = px.area(covid, x="Date", y="Confirmed", color="Country")
        fig1.update_layout(height=600, width=800)
        st.write(fig1)

    elif cases_select == 'Recovered':
        st.subheader("Recovered Cases")
        fig = px.bar(covid, x="Country", y="Recovered", color="Country")
        fig.update_layout(height=600, width=800)
        st.write(fig)

        fig2 = px.line(covid, x="Date", y="Recovered", color="Country")
        fig2.update_layout(height=600, width=800)
        st.write(fig2)
    else:
        st.subheader("Deaths Cases")
        fig = px.bar(covid, x="Country", y="Deaths", color="Country")
        fig.update_layout(height=600, width=800)
        st.write(fig)

        fig3 = px.line(covid, x='Date', y="Deaths", color="Country")
        fig3.update_layout(height=600, width=800)
        st.write(fig3)
```

Figure 93. Cases and Chart Code



```
def app():
    st.title(" Covid-19 Prediction  Model")
    a = st.slider('select a day range from 0 to 5', min_value=0, max_value=5, value=2)
    data = pd.read_csv('total_Cases.csv')
    nan_value = float("NaN")
    data.replace("", nan_value, inplace=True)
    data.dropna(subset=['id', 'cases'], inplace=True)
    m = (data.dtypes == 'float')
    data.loc[:, m] = data.loc[:, m].astype(int)

    x = np.array(data['id']).reshape(-1, 1)
    y = np.array(data['cases']).reshape(-1, 1)
    plt.plot(y, '-m')

    polyFeat = PolynomialFeatures(degree=6)
    x = polyFeat.fit_transform(x)

    model = linear_model.LinearRegression()
    model.fit(x, y)
    accuracy = model.score(x, y)
    st.write(f'Accuracy:{round(accuracy * 100, 6)} %')
    y0 = model.predict(x)
    plt.plot(y0, '--b')

    # prediction
    st.write(f'Prediction - Cass after {a} days:', end='')
    st.write(round(int(model.predict(polyFeat.fit_transform([[335 + a]]))) / 1000000, 2), 'Million Cases')
```

Figure 94. Machine learning model code

Data analysis and visualization code

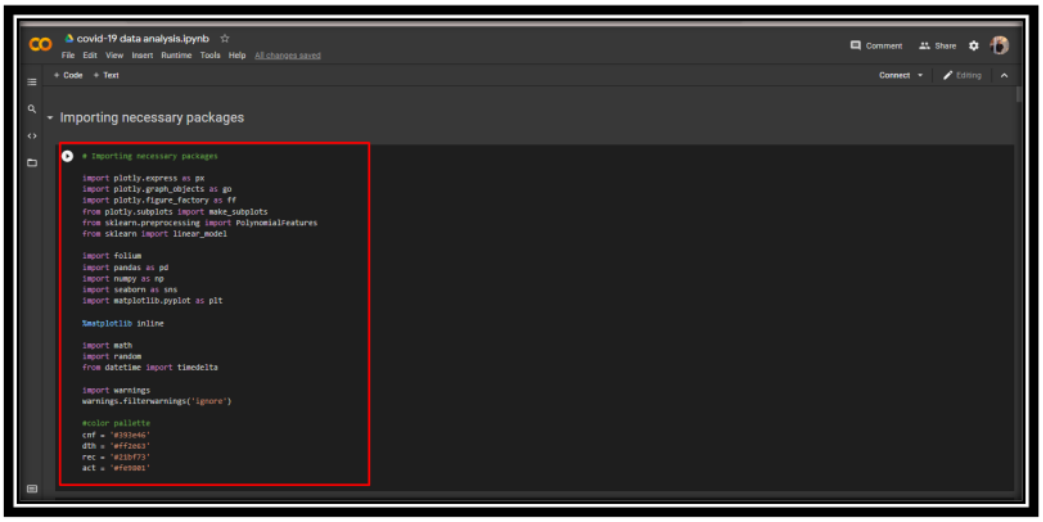


Figure 95. Importing libraries

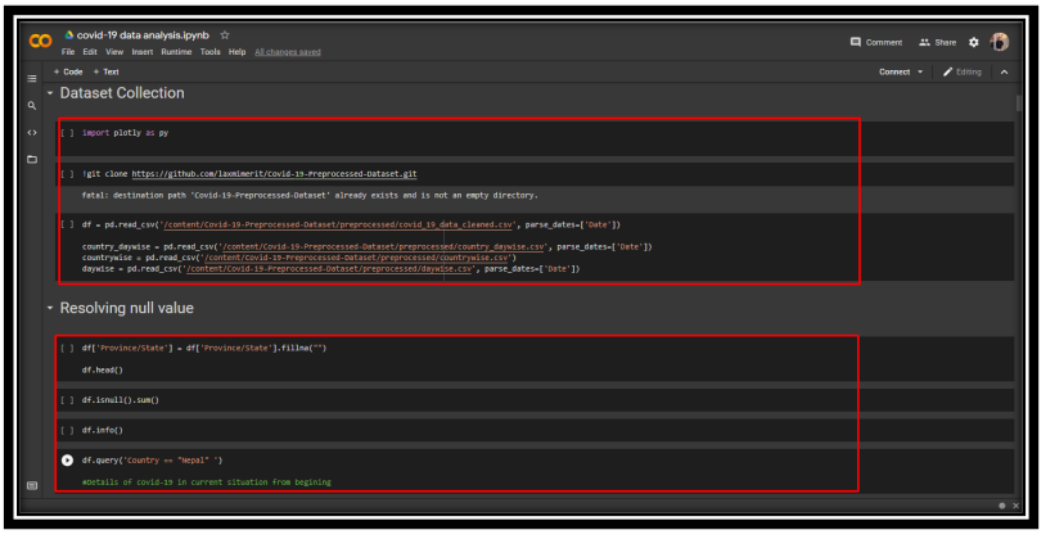


Figure 96. Data collection and loading

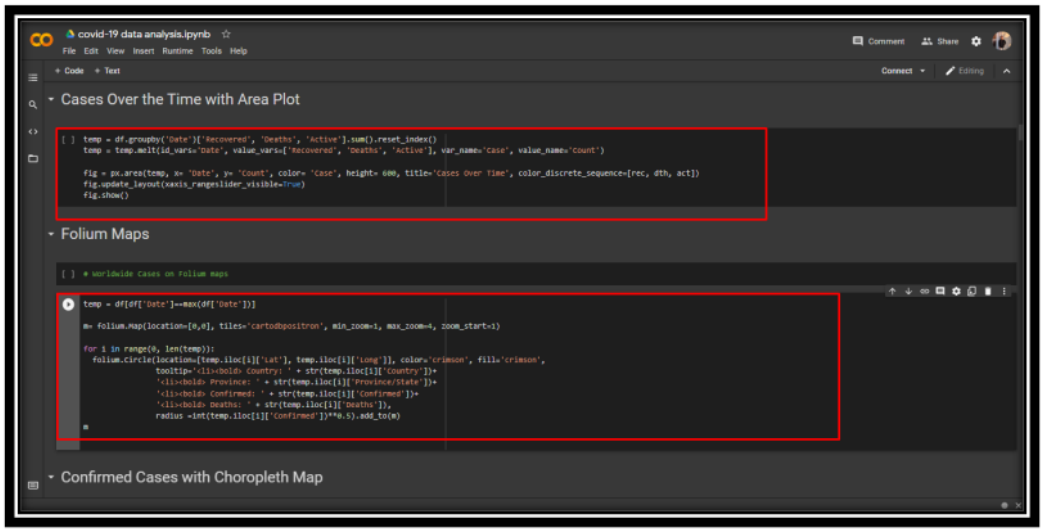


Figure 97. Cases over time and folium map

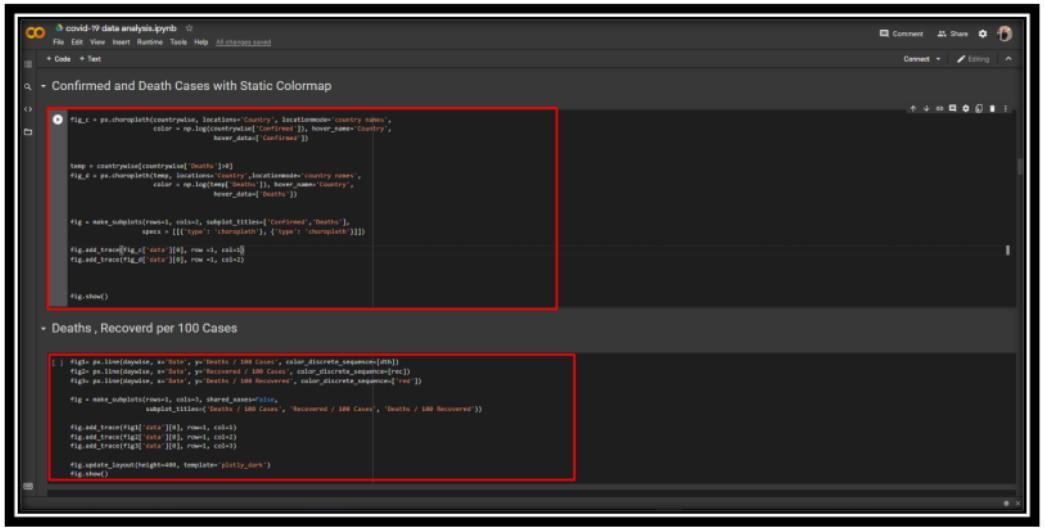


Figure 98. Confirmed and death cases

Figure 99. Training machine learning model

Figure 100. Comparing Covid-19 with others

Production Project

ORIGINALITY REPORT

9%	6%	4%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to The British College Student Paper	1%
2	www.mdpi.com Internet Source	<1%
3	Submitted to Liverpool John Moores University Student Paper	<1%
4	Submitted to Coventry University Student Paper	<1%
5	f1000research.com Internet Source	<1%
6	ijircce.com Internet Source	<1%
7	medium.com Internet Source	<1%
8	Submitted to Lovely Professional University Student Paper	<1%
9	doctorpenguin.com Internet Source	<1%

10	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	<1 %
11	www.researchgate.net Internet Source	<1 %
12	academic.oup.com Internet Source	<1 %
13	Submitted to Middlesex University Student Paper	<1 %
14	Submitted to Arab Open University Student Paper	<1 %
15	Submitted to Ain Shams University Student Paper	<1 %
16	Submitted to King's College Student Paper	<1 %
17	Submitted to University of Keele Student Paper	<1 %
18	www.who.int Internet Source	<1 %
19	Anand Jose Paul, Sahimol Eldhose. "Chapter 67 Development of Frame Work for Residential Building Construction Using Agile Management", Springer Science and Business Media LLC, 2020 Publication	<1 %

20	Submitted to University of Salford Student Paper	<1 %
21	www.ecf.toronto.edu Internet Source	<1 %
22	Submitted to Letterkenny Institute of Technology Student Paper	<1 %
23	Submitted to University of Technology, Sydney Student Paper	<1 %
24	Carlos Fernandez-Lozano, Francisco Cedron. "Shiny Dashboard for Monitoring the COVID-19 Pandemic in Spain", Proceedings, 2020 Publication	<1 %
25	Submitted to University of Ulster Student Paper	<1 %
26	Submitted to Bournemouth University Student Paper	<1 %
27	G. Sreedhar. "chapter 21 A Framework to Improve Performance of E-Commerce Websites", IGI Global, 2021 Publication	<1 %
28	Submitted to Federation University Student Paper	<1 %

29	Oren Miron, Kun-Hsing Yu, Rachel Wilf-Miron, Nadav Davidovitch. "COVID-19 infections following outdoor mass gatherings in low incidence areas: retrospective cohort study", Cold Spring Harbor Laboratory, 2020 Publication	<1 %
30	Reed F. Beall, Aidan Hollis. "Global clinical trial mobilization for COVID-19: higher, faster, stronger", Drug Discovery Today, 2020 Publication	<1 %
31	Submitted to Royal Holloway and Bedford New College Student Paper	<1 %
32	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
33	expertprogrammanagement.com Internet Source	<1 %
34	Submitted to RDI Distance Learning Student Paper	<1 %
35	orca.cf.ac.uk Internet Source	<1 %
36	analyticsindiamag.com Internet Source	<1 %
37	www.ncl.ac.uk Internet Source	<1 %

38	L. J. Muhammad, Ebrahem A. Algehyne, Sani Sharif Usman, Abdulkadir Ahmad, Chinmay Chakraborty, I. A. Mohammed. "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset", SN Computer Science, 2020 Publication	<1 %
39	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
40	ethesisarchive.library.tu.ac.th Internet Source	<1 %
41	worldwidescience.org Internet Source	<1 %
42	www.soi.city.ac.uk Internet Source	<1 %
43	Jianfeng Hua, Xinfan Lin, Liangfei Xu, Ping Guo, Jianqiu Li, Ouyang Minggao. "Application and study of novel electronic technologies on vehicle control system of fuel cell bus", 2008 IEEE Vehicle Power and Propulsion Conference, 2008 Publication	<1 %
44	plus.mcmaster.ca Internet Source	<1 %
45	Hai Lan, Dexuan Sha, Anusha Srirenganathan Malarvizhi, Yi Liu et al. "COVID-Scraper: An	<1 %

Open-Source Toolset for Automatically Scraping and Processing Global Multi-Scale Spatiotemporal COVID-19 Records", IEEE Access, 2021

Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		