Research Lab

Incidence Response Model For Covid-19

Group Report

Submitted By

Name	Email	Matriculation
Irshad Hussain Mohammed	imohammed@uni-koblenz.de	220200902
Manpreeth Vankadara	manu@uni-koblenz.de	221202815
Pavan Kumar Tokachichu	pavankumar@uni-koblenz.de	220200833
Pavan Kumar Reddy Kancharla	pavankancharla@uni-koblenz.de	221202462
Nirav Satani	niravsatani24@uni-koblenz.de	221100483
Peram Navachandu Reddy	peramnavachandu@uni-koblenz.de	220202360
Ramesh Reddy Modulla	modullarameshreddy@uni-koblenz.de	221100615
Rehan Ali Syed	syedali@uni-koblenz.de	221100697
Sudheer Kumar Gandham	sudheerkumargandham@uni-koblenz.de	221202511
Venkata Satya Aditya Jagarlapudi	adityajvs@uni-koblenz.de	221202338

Koblenz,December 12, 2022

Supervisors: Prof. Dr. Maria Wimmer
Dr. Ulf Lotzmann

Abstract

When Covid-19 pandemic hit the world, Governments across the world had no prior Incidence Response Model to deal with the pandemic. As a result, they had to abort all the global commutes, call for Lockdowns and close their borders. These actions could not stop the spread of the Covid-19 virus but in turn, cost the lives of people and financial loss.

With the motive of decreasing the spread of viruses by tracking them from their source and decreasing the economic loss caused by border closures, this research lab builds a model by analyzing Covid-19 data, tracking the evolution of the virus from its source, analyzing the air traffic from the source, and clustering regions into high, medium, and low-risk areas.

The model analyses the outgoing flights from South Africa (The place where the Omicron variant was found), predicts the number of cases and clusters them based on cases predicted and their population.

The clusters are then represented in Tableau as a dashboard where it shows the different clusters for upcoming days helping the government to take better decisions.

This approach would let governments cluster the countries with a high risk of viruses and not close the borders completely. Not closing the borders completely can lessen the economic crisis caused due to pandemic and can prevent loss of employment in sectors like tourism, global supply chain, international transportation industry etc.

In addition to that, this model analyses the vaccinations across different countries in the world and presents the trends of vaccinations and deaths in form of a dashboard. Keywords: Incidence Response Model, Data Analysis of Air Traffic, Predictions, Clustering, Tableau Dashboard, Vaccinations.

Keywords: Incidence Response Model, Data Analysis of Air Traffic, Predictions, Clustering, Tableau Dashboard, Vaccinations.

Contents

List of Figures

List of Tables

1. Introduction

Author: Irshad Hussain Mohammed(60%), Ramesh Reddy Modulla(20%), Rehan Ali Syed(20%)

Corona Virus pandemic created havoc in the world, and it affected the world in a way for which Countries were not prepared for. According to Eleni Smitham and Amanda Glassman [1], the probability of a new pandemic is much higher and sooner than what we can think of.

But this is the time to question ourselves if we are ready for the next pandemic if we have any response model that can be implemented as soon as a new variant comes out if Governments are ready to face another global pandemic [2].

When the world was hit by the Covid-19 virus, it costed many lives, many jobs, shortage of food and medicinal facilities. All policymakers were surprised by the spread.

The focus of the project is to build an Incidence Response Model based on Covid-19 virus data. In this project, we emphasize analyzing the data of air traffic from the source country of variant and finding a relation between air traffic and covid cases. With this relation, we apply algorithms to predict the cases in the destination countries.

The Omicron variant of Covid-19 was chosen to analyze data . The first case of Omicron variant was found in South Africa [3]. So, we collect the data of international flights from South Africa and identify the destinations. Then we collect the data of cases in destination countries, train them with time series algorithms, predict cases in these destinations. The then results are used to make clusters based on the incidence rate calculated. A dashboard is created to visualize the results and for stakeholders to understand patterns and take decisions.

Meanwhile, it is found that Omicron Variant despite of its high spread ability, the recovery rate was very high [2]. Shows that the vaccinations were responsible for less hospitalization. So, as a part of our research, vaccinations data is analyzed and a dashboard with predicted deaths is built.

2. Project Management

Author: Rehan Ali Syed(80%), Ramesh Reddy Modulla(20%)

2.1. Team Organisation

2.1.1. Project Leaders

Team Leaders	From	To
Sudheer Kumar Gandham, Nirav Satani	19/05/2022	06/06/2022
Ramesh Reddy Modulla, Rehan Ali Syed	07/06/2022	13/07/2022
Irshad Hussain Mohammed, Peram Navachandu Reddy	14/07/2022	19/08/2022
Pavankumar Tokachichu, Pavan Kancharla	20/08/2022	06/09/2022
Vankadara Manpreeth, Venkata Satya Aditya Jagariapudi	07/09/2022	24/09/2022
Irshad Hussain Mohammed, Ramesh Reddy Modulla	25/09/2022	27/10/2022

Table 2.1.: Project Leaders

2.1.2. Report

Project Supervisors
Prof. Dr. Maria A. Wimmer
Dr. Ulf Lotzmann

Table 2.2.: Project Supervisors

2.2. Project Meetings

2.2.1. Official Meetings

Table 2.3.: Official Meetings

Nr.	Purpose of Meeting	Date	Feedback
1.	Pitching Idea.	10.05.2022	• The panel agreed upon giving us the opportunity for the research lab and suggested we narrow down our project model to China only and limit it to Covid-19 rather than pandemics.
2.	Architecture and data acquisition result presentation.	25.05.2022	• The panel suggested researching more about Chinese Flight data. • In alternative case search for flights from Omicron-originated Countries. • The panel asked to present a more detailed Architecture and Gantt Chart.

Nr.	Purpose of Meeting	Date	Feedback
3.	Presenting Architecture and delivering data acquired.	21.06.2022	• The panel suggested developing a more detailed Architecture with components. • The panel suggested extending the Gantt chart by adding deliverables, and tasks. • The panel asked to create a GitHub repository and upload all the deliverables and documentation related to the project in the repository. • The panel asked to identify artifacts and test the hypothesis regarding factors affecting Omicron. • Migrate to Confluence or Open Project and invite Professors.
4.	Deliver the Dataset, prepare for analysis and discuss the algorithms.	11.07.2022	• The panel suggested including about Algorithm in the Documentation. • Individual contribution should be placed at the top. • Document about steps followed in Collecting Data, and Analysis part. • Gantt chart should be hierarchical. • Add Reference in the Documentation.

Nr.	Purpose of Meeting	Date	Feedback
5.	Deliver a draft model and discuss results.	24.08.2022	• The panel suggested creating groups in open projects. • Upload Documentation in Github. • Create a Dashboard for Vaccinations. • Gantt chart should be hierarchical. • Derive a conclusion of how vaccinations affected Covid diseases.
6.	Deliver dashboard.	19.09.2022	• The panel suggested in the Map that vaccination data should be changed from 3 to 10 categories. • In Bar chart, the number of vaccinations per day should be changed to per week. • December 3 data came out of bounds. Should be rechecked. • Choose the best algorithm (DEEPAP, LSTM, ARIMA) for vaccination data. • Should mention the limitations of LSTM in the documentation.

Nr.	Purpose of Meeting	Date	Feedback
7.	Deliver vaccination predictions, updated dashboard, and documentation.	27.10.2022	• The panel recommended adding a label cumulative for vaccination data. • Comparison between prediction and actual data. • To add a legend for Predicted data and Analysis of actual death. • In the documentation, the individual contribution should be in table format after the table contents. • The actual number of deaths should be per day and not cumulative data. • The user will provide a date to IR Model and the model should deliver 20 days of prior data and 10 days of prediction data.

2.2.2. Internal Meetings

Table 2.4.: Internal Meetings

Nr.	Purpose of Meeting	Date
1.	Meeting to know the team	27.04.2022
2.	Meeting to prepare for pitching	04.05.2022
3.	Separation of tasks	11.05.2022
4.	Update on Deliverable	18.05.2022
5.	Update on Deliverable	25.05.2022
6.	Preparing Deliverable for Presentation 2	01.06.2022
7.	Separation of Tasks	08.06.2022
8.	Update on Deliverable	15.06.2022
9.	Separation of Tasks	22.06.2022
10.	Update on Deliverable	29.06.2022
11.	Update on Deliverable	06.07.2022
12.	Separation of Tasks	13.07.2022
13.	Discussing Blockers	20.07.2022
14.	Discussing Blockers	27.07.2022
15.	Update on Deliverable	03.08.2022

Nr.	Purpose of Meeting	Date
16.	Update on Deliverable	10.08.2022
17.	Update on Deliverable	17.08.2022
18.	Separation of Tasks	24.08.2022
19.	Discussing Blockers	31.08.2022
20.	Discussing Blockers	07.09.2022
21.	Update on Deliverable	14.09.2022
22.	Separation of Tasks	21.09.2022
23.	Discussing Blockers	28.09.2022
24.	Discussing Blockers	05.10.2022
25.	Update on Deliverable	12.10.2022
26.	Discussing Blockers	19.10.2022
27.	Preparing for Presentation	26.10.2022

2.2.3. Task Allocation

Table 2.5.: Task Allocation

Sprint Nr.	Name	Task
Sprint 1	Irshad Hussain Mohammed	Presentation, Documentation, Minutes of Meeting, Data Analysis Team, Architecture.
	Manpreeth Vankadara	Research about Algorithms
	Pavan Kumar Tokachichu	Research about Algorithms
	Pavan Kumar Reddy Kancharla	Data Acquisition
	Nirav Satani	Management, Research about Algorithms
	Peram Navachandu Reddy	Research about Algorithms
	Ramesh Reddy Modulla	Data Acquisition
	Rehan Ali Syed	Data Acquisition
	Sudheer Kumar Gandham	Management, Data Acquisition
	Venkata Satya Aditya Jagarlapudi	Data Acquisition
Sprint 2	Irshad Hussain Mohammed	Presentation, Algorithms and Artifacts Research
	Manpreeth Vankadara	Research about Algorithms
	Pavan Kumar Tokachichu	Research about Algorithms
	Pavan Kumar Reddy Kancharla	Correlation
	Nirav Satani	Research about Algorithms
	Peram Navachandu Reddy	Research about Algorithms
	Ramesh Reddy Modulla	Management, Data Acquisition
	Rehan Ali Syed	Management, Data Acquisition
	Sudheer Kumar Gandham	Correlation
	Venkata Satya Aditya Jagarlapudi	Data Acquisition
Sprint 3	Irshad Hussain Mohammed	Presentation, Management, Architecture, Algorithms, Artifacts, Documentation.
	Manpreeth Vankadara	Research about Algorithms
	Pavan Kumar Tokachichu	Documentation, Reseach about Algorithms, Architecture
	Pavan Kumar Reddy Kancharla	Correlation
	Nirav Satani	Research about Algorithms
	Peram Navachandu Reddy	Management, Research about Algorithms

Sprint Nr.	Name	Task
	Ramesh Reddy Modulla	Data Acquisition
	Rehan Ali Syed	Data Acquisition
	Sudheer Kumar Gandham	Data Acquisition, Correlation
	Venkata Satya Aditya Jagarlapudi	Data Acquisition
Sprint 4	Irshad Hussain Mohammed	Presentation, Algorithms, Artifacts Research, LSTM Model Training, Documentation.
	Manpreeth Vankadara	Prophet Draft Model
	Pavan Kumar Tokachichu	Management, LSTM Draft Model
	Pavan Kumar Reddy Kancharla	Management, VARIMA Draft Model
	Nirav Satani	DEEPAR
	Peram Navachandu Reddy	ARIMA, VAR, Exponential Smoothing
	Ramesh Reddy Modulla	ARIMA, VAR, Exponential Smoothing
	Rehan Ali Syed	DEEPAR Model, Documentation
	Sudheer Kumar Gandham	Prohet Model
	Venkata Satya Aditya Jagarlapudi	Coding, Troubleshoot, Clustering
Sprint 5	Irshad Hussain Mohammed	Presentation, Dashboard, Extended Model Planning and Implementation, Open Project
	Manpreeth Vankadara	Management, Documentation about Prophet
	Pavan Kumar Tokachichu	Documentation on LSTM
	Pavan Kumar Reddy Kancharla	Documentation on Correlation
	Nirav Satani	Documentation on DEEPAR
	Peram Navachandu Reddy	Documentation on VARIMA
	Ramesh Reddy Modulla	Documentation on Data Acquisition
	Rehan Ali Syed	Overall Documentation
	Sudheer Kumar Gandham	Documentation about Prophet
	Venkata Satya Aditya Jagarlapudi	Management, Coding, and Vaccination Data Acquisition, Vaccinations prediction by LSTM
Sprint 6	Irshad Hussain Mohammed	Presentation, Management, Data Acquisition and Preparation.
	Manpreeth Vankadara	Documentation about Prophet
	Pavan Kumar Tokachichu	Documentation on LSTM
	Pavan Kumar Reddy Kancharla	Documentation on Correlation
	Nirav Satani	Documentation on DEEPAR

Sprint Nr.	Name	Task
	Peram Navachandu Reddy	Documentation on VARIMA
	Ramesh Reddy Modulla	Management, Documentation on Data Acquisition
	Rehan Ali Syed	Overall Documentation
	Sudheer Kumar Gandham	Documentation about Prophet
	Venkata Satya Aditya Jagarlapudi	Training and Predicting Vaccinations and changes to Vaccination Data
Sprint 7	Irshad Hussain Mohammed	Presentation, Management, Data Acquisition, Preparation and final documentation.
	Manpreeth Vankadara	Documentation about Prophet
	Pavan Kumar Tokachichu	Documentation about LSTM
	Pavan Kumar Reddy Kancharla	Documentation about Correlation
	Nirav Satani	Documentation about DEEPAR
	Peram Navachandu Reddy	Documentation about VARIMA
	Ramesh Reddy Modulla	Management, Documentation on Data Acquisition
	Rehan Ali Syed	Overall Documentation
	Sudheer Kumar Gandham	Documentation about Prophet
	Venkata Satya Aditya Jagarlapudi	Coding, Troubleshooting, Adding Flexibility

2.2.4. Project Plan

Gantt Chart

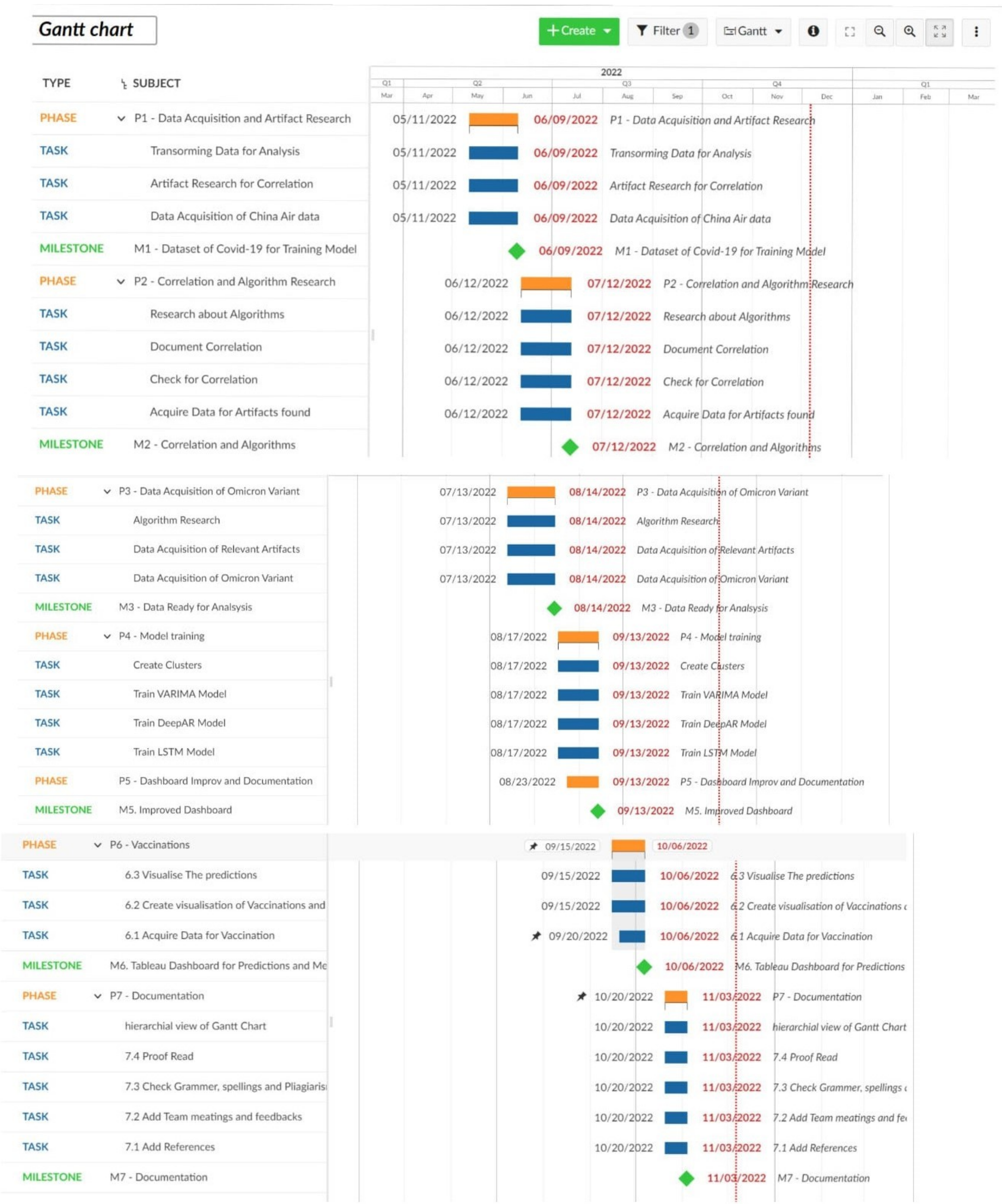

Figure 2.1.: Gantt Chart

3. Research Process

3.1. Former Research

Author: Pavan Kumar Reddy Kancharla(25%), Peram Navachandu Reddy(25%), Nirav Satani(25%), Ramesh Reddy Modulla(25%)

Over the recent few months, a number of researchers have analyzed the impact of various aspects on the spread of COVID-19 around the world.

Different variables have a significant impact on the spread of COVID-19 globally. For instance, Amanda M Y Chu [4], uses the network density method on the COVID-19 confirmed cases of different countries and found out that the higher the network density, the higher the tendency that a country's number of confirmed COVID-19 cases to increase together, thus indicating increasing pandemic risk. Omar Sharif [5], forecasts the number of COVID-19 infection cases based on demographic variables such as population density and literacy rate using ANOVA and Holt's Exponential Smoothing method. Matteo Chinazzi, Jessica T. Davis [6], investigated the effect of travel restrictions on the spread of COVID-19 globally using a stochastic and spatial epidemic model called the global epidemic and mobility model (GLEM) on Chinese flight data and found various results on how effectively COVID-19 transmission can be decreased by banning international flights.

Yusuf Jamal, Mayank Gangwar, and Moiz Usmani [7], identify the various thresholds in population density for understanding COVID-19 transmission using logistic regression functions on USA population data and found out that there is a 95 percent chance that 43380 people will be infected by COVID-19. If the population density is more than 3000 persons per square mile in the US. Sylvia Xiao Wei Gwee, Pearleen Ee Yong Chua, Min Xian Wang, and Junxiong Pang [8], did a comparative study on the impact of travel ban implementation on Singapore, hong kong, and South Korea using the 7-day and 14-day moving average method and found out that the effectiveness of the travel ban can be seen only when it was implemented at the early phase and local surveillance is concurrently implemented. Hien Lau, Veria Khosrawipour [9] abstracts the relationship between international and domestic air traffic and the coronavirus (COVID-19) outbreak. The authors collected flight data from China. The correlation between the frequency of the flight data and the outbreak of the corona cases in each region was found using a linear regression method. With their statistical approach, the research concludes that the total passenger numbers in mainland China directly correlate to reported COVID-19 cases with travel history to China. Jenna K. Pang [10] suggested that the risk of COVID- 19 transmission on an aircraft is low, even with infectious persons onboard. In this research, they used statistical analysis with their defined formula which entirely contradicts the other research works.

Almost each prior research work predicted COVID-19 cases with the help of statistics and related conventional methods. Additionally, these research works forecasted the corona cases with limited features I.e., flight frequency and neglecting the actual number of passengers. The frequency of flights with predicted passengers number, and population density of the country are considered for the research process.Neural networks are used for better forecasting and cutting the limitations of the earlier works.

3.2. Research Question

Author: Pavan Kumar Tokachichu(100%)

What Required Objectives Do the Model Meet?

- Helps in Better Understanding of the pandemic situation.
- Reduces complexity for categorization of regions.
- Able to Analyze and Interpret data reliably, and quickly.
- Helps decision makers with decision recommendations.

What data can be explored from existing data sources?

The data on modes of international transportation are to be obtained from the vast data sources and transformed as required.

What methods for data analysis from the areas of AI, big data/big data analytics, data science, and statistical methods can be applied to the available data?

The data analysis methods are going to be obtained-data dependent, mostly from the analysis methods such as Data Mining, Time Series Analysis.

3.3. Research about algorithms

The data series is stationary, which means that the mean and variance should not vary with time. A series can be made stationary by using log transformation or differencing the series(d).

3.3.1. Vector ARIMA

Author: Peram Navachandu Reddy(80%), Pavan Kumar Reddy Kancharla(20%)

A Multivariate Time Series consists of more than one time-dependent variable and each variable depends only on its past values.

To deal with multi-variate time series, one of the most popular methods is Vector Auto Regressive Integrated Moving Average models (VARIMA) which is a vector form of autoregressive integrated moving average (ARIMA) that can be used to examine the multivariate time series analysis [11].

The Vector ARIMA model is a commonly used machine learning model which is used to forecast time series data. It is the merger of two components which are:

AR (Auto Regressive)

This model attempts to predict future values based on past values. For this model, time series data should be stationary [12].

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t \tag{3.1}$$

Where:

y_t Is the value at time step t, c is a constant, ϕ_1 is a coefficient, and ϵ_t is a white noise error term with $\epsilon_t \sim N\ (0, \sigma^2)$.

From this model, we will find the value p, which represents how many prior time steps to use in the time regression.

MA (Moving Average)

It is the linear regression of the present value of the series against previously observed error terms [12].

$$y_t = c + \theta_1 \varepsilon_{t-1} \tag{3.2}$$

Where:

y_t Is the value at time step t, c is a constant, ϕ_1 is a coefficient, and ϵ_t is a white noise error term with $\epsilon_t \sim N\ (0, \sigma^2)$.

From this model, we will find the value q, which represents how many prior errors we need to consider.

Applications

- VARIMA model uses statistical analyses in combination with collected historical data points to predict future trends.so, data should be more reliable.
- Consists of more components (such as AR, MA), so that the error will be less.

Limitations

- Data should be stationary for the model to work.
- Long time span of data is needed for the model to forecast future values effectively.

3.3.2. Deep AR

Author: Nirav Satani(80%), Ramesh Reddy Modulla(20%)

The DeepAR algorithm is a supervised learning algorithm. It is a forecasting time series algorithm based on recurrent neural networks (RNN). Some conventional models like Autoregression (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and Simple Exponential Smoothing (SES) learn from past observations and predict future values using recent history. Whereas, instead of fitting separate models for each time series, DeepAR aims to create a global model that learns using all-time series in the dataset.

DeepAR outperforms the standard ARIMA and ETS methods when the data has multiple time series. DeepAR can forecast the data of multiple time series by only single model training [13].

The training input for the DeepAR algorithm is one or more target time series data that are generated by the same or similar processes. The input data is given as a JSON file or in Parquet format. Input data should contain the start and target fields. The start field is a string with the format YYYY-MM-DD HH:MM: SS. The target field is an array of floating-point values or integers that represent the time series.

Prepared data should be trained with the DeepAR model. The below parameters are required to define the model.

Required Parameters [14]:

- Context_length: Hyper-parameter controls how far in the past the network can see
- Prediction_length: Hyper-parameter controls how far in the future predictions can be made.
- Epochs: The maximum number of passes over the training data.
- Time_freq: The granularity of the time series in the dataset.

Based on this input, DeepAR trains a model and uses it to forecast the target time series data.

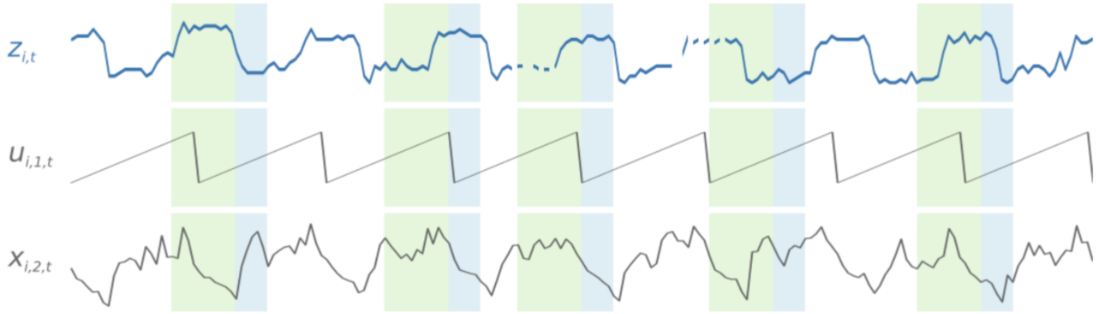

Figure 3.1.: Input and Output

Figure 3.1 depicts the general overview of how the output of the DeepAR algorithm is a lookalike. Here, Z is the target time series, and U and X are the features.

Pros and cons

DeepAR has the advantage of training several hundred or thousands of time series simultaneously, potentially offering significant model scalability. It also has the following technical benefits:

- Minimal Feature Engineering: The model requires minimal feature engineering, as it learns seasonal behavior on given covariates across time series.
- Monte Carlo Sampling: It is also possible to compute consistent quantile estimates for the sub-ranges of the function, as DeepAR implements Monte Carlo sampling. This could, for instance, be useful when deciding on safety stock.
- Built-in item supersession: It can predict on items with little history items by learning from similar items
- Variety of likelihood functions: DeepAR does not assume Gaussian noise, and likelihood functions can be adapted to the statistical properties of the data allowing for data flexibility [15].

3.3.3. Long Term Short Memory (LSTM)

Author: Pavan Kumar Tokachichu(70%), Irshad Hussain Mohammed(30%)

LSTM network is one of the most advanced models used to forecast time series data. LSTM extends the memory of recurrent neural networks (RNN) because RNN suffers from the problem of vanishing gradients [16]. LSTMs are designed to tackle the problem of long-term dependency problem. Unlike the standard feedforward neural networks, LSTM has feedback connections. Because of this characteristic, LSTMs can analyze whole data sequences, such as time series, without having to consider each point separately [16].

Relative insensitivity to gap length is an advantage of LSTM over RNNs and other sequences learning methods [17].

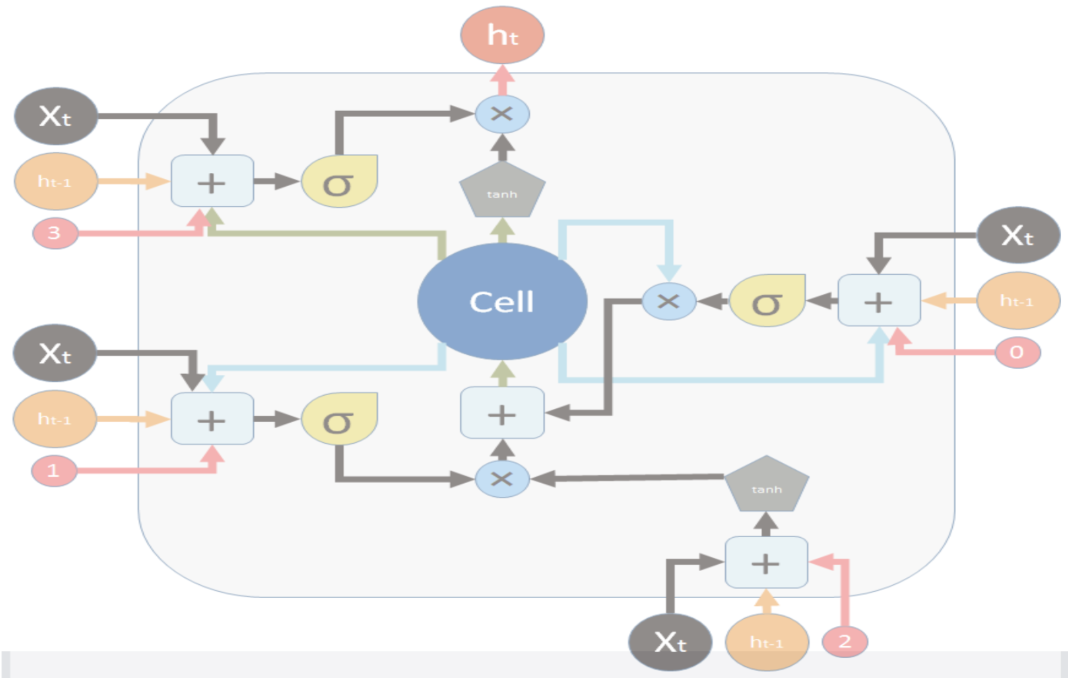

Figure 3.2.: LSTM Architecture

LSTMs use a sequence of gates that regulates the flow of information flowing through the sequence chain. Gates are contained in memory blocks which are contained through layers [18]. Within a unit, there are three important gates Input gate, Output gate, and Forget gate [19].

Forget Gate:

For the given both the hidden state (h_t-1) and new input data(X_t), forget gate decides which bits of cell state are useful.

Input Gate:

Both the input gate and new memory network (tanh) are neural networks and for the given previous hidden state and new input data, it decides what new information should be added to the cell state.

Output Gate:

The output gate is controlled by the new input, previously hidden state, and newly updated cell state. This controls how much memory should be output to the next LSTM unit.

Limitations of LSTM

- LSTM requires more memory to train or to forecast future values effectively [20].
- LSTMs are easy to overfit and get affected by different random weight initialization [20].

4. Implementation

4.1. Architecture

Author: Irshad Hussain Mohammed(50%), Nirav Satani(30%), Rehan Ali Syed(20%)

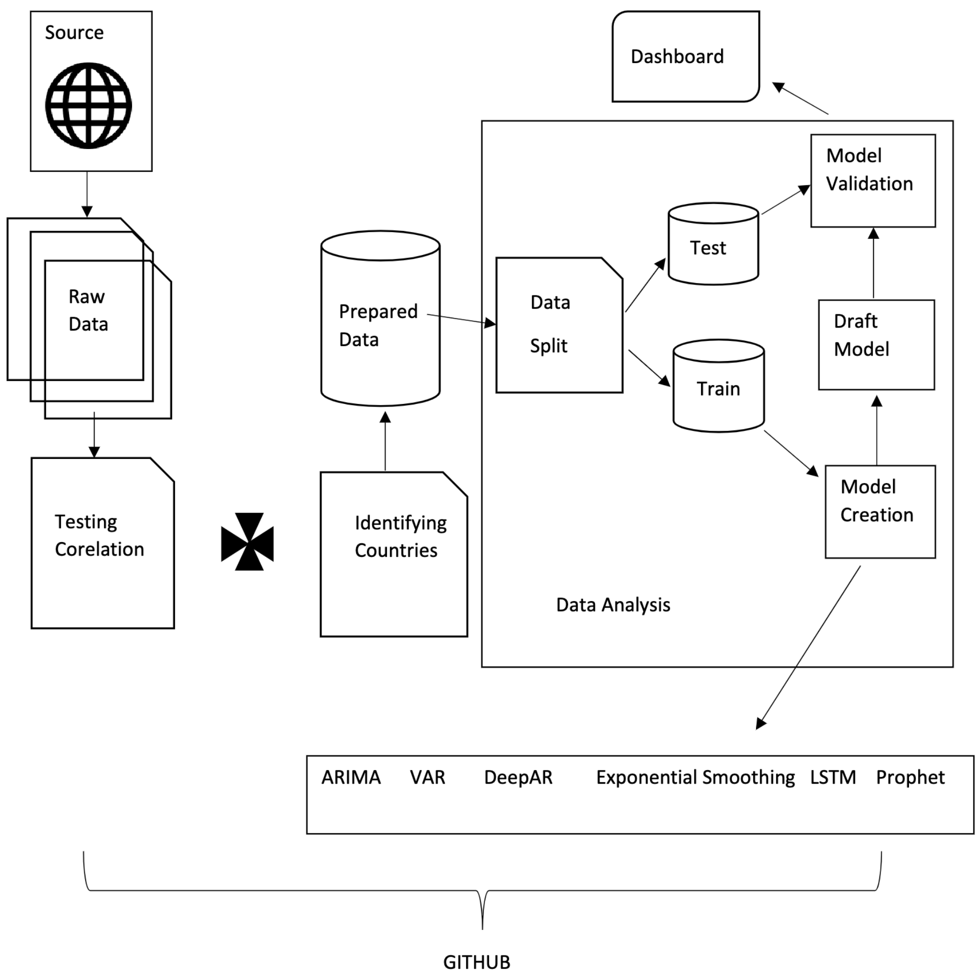

Figure 4.1.: Architecture

4.2. Data Acquisition

Author: Rehan Ali Syed(50%), Ramesh Reddy Modulla(50%)

4.2.1. Acquiring Chinese Data

The idea of the project is to acquire Chinese travel data and analyze the destination countries for covid cases. We tried to get the data from different sources(websites) e.g. Flight aware [21], Aviation Stack [22] but it turned out that Chinese flight data is not available on any of the sources. The reason behind this was that the Chinese Government doesn't allow the ADS-B method. The ADS-B method is a method that is widely used to track flight data by websites. All the sources (including paid sources) we looked for the data, they were either taking too long to respond to or doesn't have the Chinese flight data.

Due to all these complications, we decided to move on instead of spending more time on Chinese flight data, we started focusing on the second variant i.e., the Omicron variant.

4.2.2. Acquiring Data of Omicron Variant

According to WHO, the SARS-CoV-2 Omicron variant(B.1.1.529), was first reported from South Africa in November 2021. As South Africa and the United Kingdom were the first countries to witness the first omicron cases, we decided to look for the data for these two countries.

Unlike Chinese flights, the data for South Africa and the United Kingdom was available in abundance. We were able to acquire data from Zenodo [23] which was providing data from the Open sky network [24].

4.2.3. Data Preparation

After acquiring the data, the data is available from the start of 2019 till march 2022 but as the first Omicron case was hit on 27th November 2021, we decided to filter our data from November 2021 till the end of January 2022. The data is available in many CSV format files which were split periodically, so we transformed those files into one CSV file and applied data cleaning techniques by removing null values and irrelevant fields. Then we again filtered this data into two different CSVs, one for the United Kingdom and the other for South Africa.

From those Unique Destinations, we have provided the file to Data Analysis Team to Identify countries for Analysis.

4.2.4. Identifying Countries

After preparing the first phase of our data we have selected destination countries having high, moderate, and low incoming traffic from South Africa and the United Kingdom. These countries from South Africa are Qatar, Netherlands, Great Britain, Germany, and France having a high amount of traffic, Israel, Morocco, Spain, Algeria and Guatemala have a moderate amount of traffic, and Iceland, Barbados, Norway, Poland, and Bulgaria have a low amount of traffic. Similarly, the destination countries from the United Kingdom are Germany, Spain, Ireland, France, and Italy having a high amount of traffic, Barbados, Malta, Latvia, Luxembourg, and Singapore have a moderate amount of traffic, and Malaysia, Kazakhstan, Philippines, Vietnam, Kenya having a low amount of traffic. So, we filtered our data again with only the above-mentioned destination countries.

4.2.5. Population Density

After identifying the destination countries and filtering the data according to that, we need more attributes in the data which can help to achieve our goal. One important factor in the covid transmission is population density, so we include a column of the population density of the destination countries. We downloaded another dataset i.e., of population density of the countries [25] and mapped the population density to destination countries in the flight data by country names.

4.2.6. Finding the Seating Capacity

After adding the population density, another important attribute is to know how many passengers are traveling on the flight. But the challenge is that no source is providing information about the number of passengers traveling in a flight. Also, the dataset does not have information about the seating capacity of a flight. So, we downloaded another dataset which consists of the aircraft model and its seating capacity. We use this dataset and map the seating capacity of a plane to the flight's dataset by aircraft model.

After mapping the seating capacity of the flights, we still have the problem of the number of passengers traveling on a flight, as this type of data is not available. So, according to the travel restrictions and CNN travel article [26] we considered that about 60 percent of people were traveling on the flights according to the flight capacity.

4.3. Data Analysis

Author: Sudheer Kumar Gandham(100%)

For Data Analysis we have a dedicated team, who are working on finding the artifacts for analysis and finding out the algorithms that could be used for the analysis of our Data

After Research we have decided to Use Time Series Analysis and we will be Using the following Algorithms for our Analysis:

1. VARIMA
2. LSTM
3. Deep AR

4.4. Correlation

Author: Pavan Kumar Reddy Kancharla(80%), Sudheer Kumar Gandham(20%)

The term "correlation" describes how strongly two quantitative variables (such as height and weight) are correlated with one another [27].

If two variables appear to be moving in the same direction, then what? If it does, the relationship between them could be characterized as positively correlated, which often denotes that if one variable rises, the other will follow suit. In the alternative, we might discover that they are inversely connected. These numbers are often between -1 and 1, indicating a strong negative and positive correlation. To put it another way, they may also be described as strong correlation and weak correlation, respectively.

In our instance, we determined the association in two situations:

1. Determine the relationship between the population per density with Covid cases.
2. Determine the connection between the number of occupied airplane seats with Covid instances of the Omicron version.

4.4.1. Covid cases vs the population per density

Why look for a correlation?
In order to investigate the impact of people living close to one another and how it affects the rise in Covid cases, we would like to establish a relationship between Covid cases and population density.

As a result, we gathered information about each country's population density per square kilometer and compared it to the average daily total of cases in the top 21 nations. This data collection was obtained from a number of places, including Kaggle [28] and wiki [29] they were discovered to be exceptionally positively connected throughout a specific time period, such as from 01-11-2021 to 31-01-2022.

Using flight data from South Africa, we chose the top 21 nations, dividing each into three risk categories: high, medium, and low.

4.4.2. Covid Cases vs the number of Flight Seats

Why look for a correlation?
We are interested in learning the relationship between the rise in Covid cases (Omicron) and the occupancy of flight seats so that we may determine whether the number of travelers entering the country is a contributing factor to the Covid cases.

We used the filtered data, which includes information on the overall number of Omicron variant cases in each of the top 21 nations, and we discovered a strong positive correlation between the two.

4.4.3. Results

Case1 :
In our first scenario, we discovered that the relationship between Covid cases and population density is positively connected with a 0.51 or 51% score, indicating that the likelihood of Covid spreading in an area is very high if there are many people there.

In Figure 4.2, the linear increase of two metrics, such as the population per square kilometer and average Covid cases, shows a strong correlation between them. However, in a few nations, like Kenya and Morocco, even though the population is higher, the proportion of average covid cases is lower.

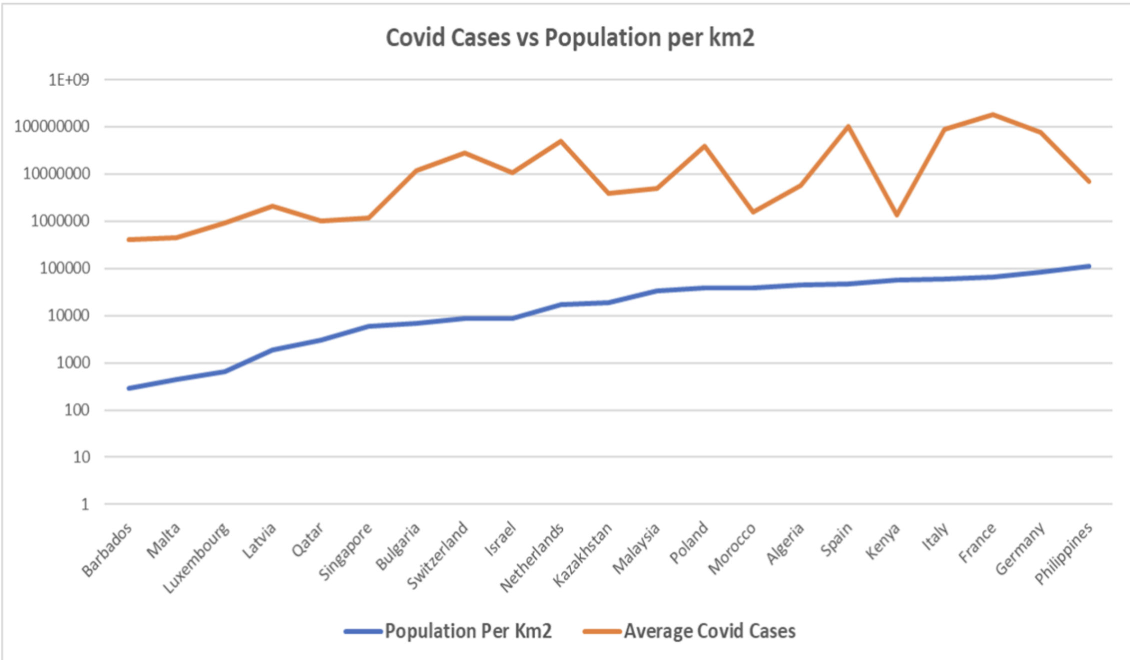

Figure 4.2.: Chart between the correlation of Population per km square vs Avg. Active cases

Case2 :

In our second case, we discovered that there was a strong correlation—a score of 0.701 or 70.1%—between the number of flight tickets or travelers and the number of Covid cases of the Omicron variant. This suggests that if more people travel to these countries, there is a good chance that the number of Covid cases will rise.

Due to the linear increase in both curves in figure ??, it is possible to see the strong association between the omicron covid cases and the average number of seats in use. Despite the fact that fewer people are traveling to these nations, there is a significant incline in the covid cases in the Netherlands and Spain due to the local transportation.

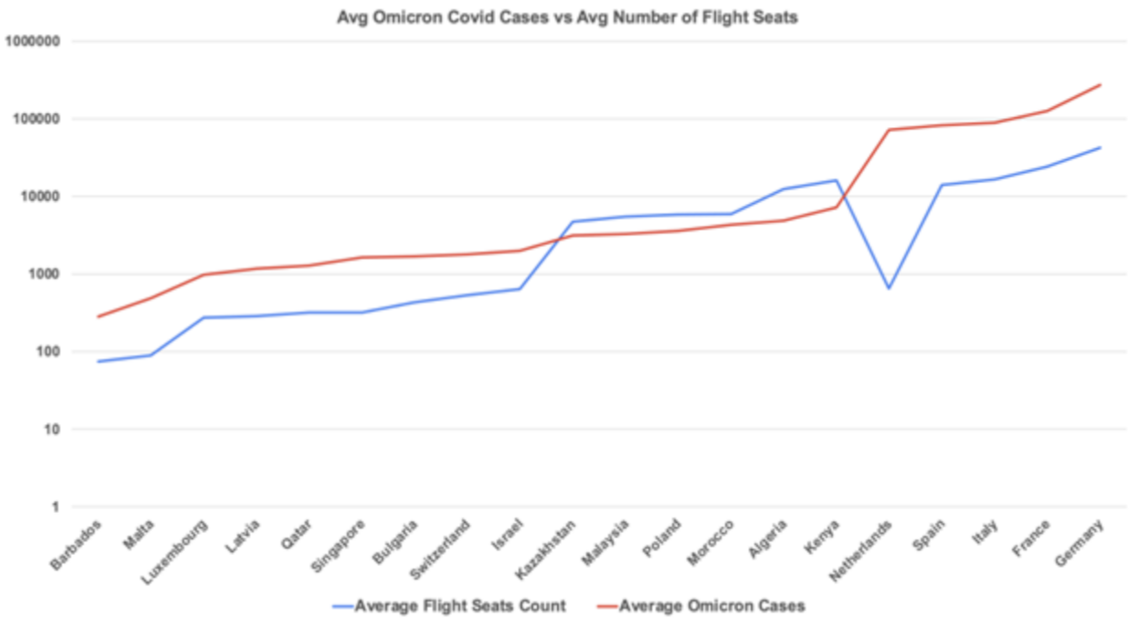

Figure 4.3.: Chart between Avg Omicron cases vs Number of Flight Seats

4.5. Flowchart of Algorithms

Author: Manpreeth Vankadara(100%)

4.5.1. Vector ARIMA

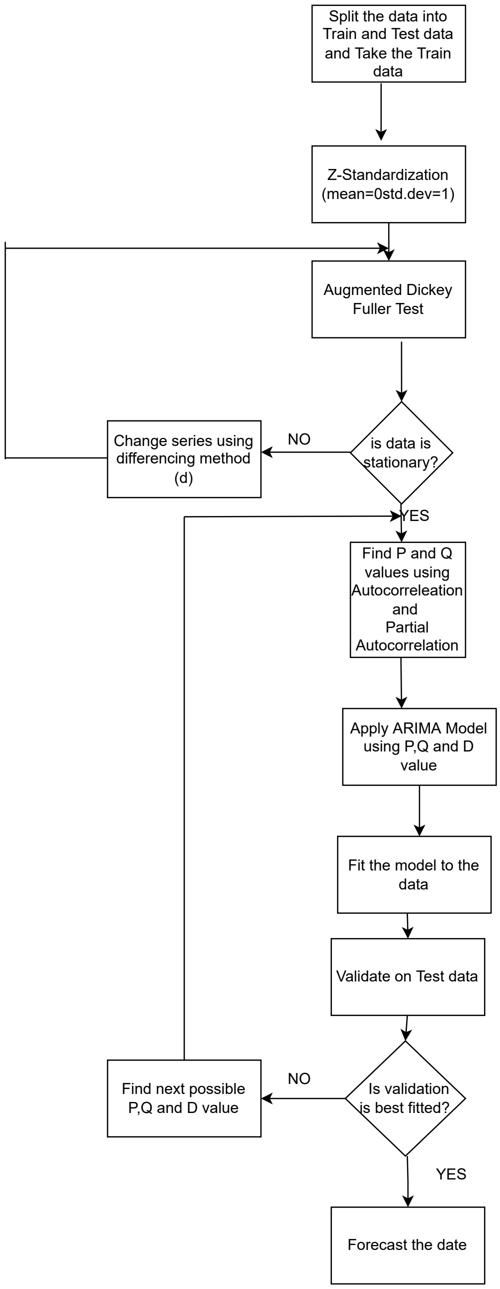

Figure 4.4.: Flow chart of Vector ARIMA model

4.5.2. Deep AR

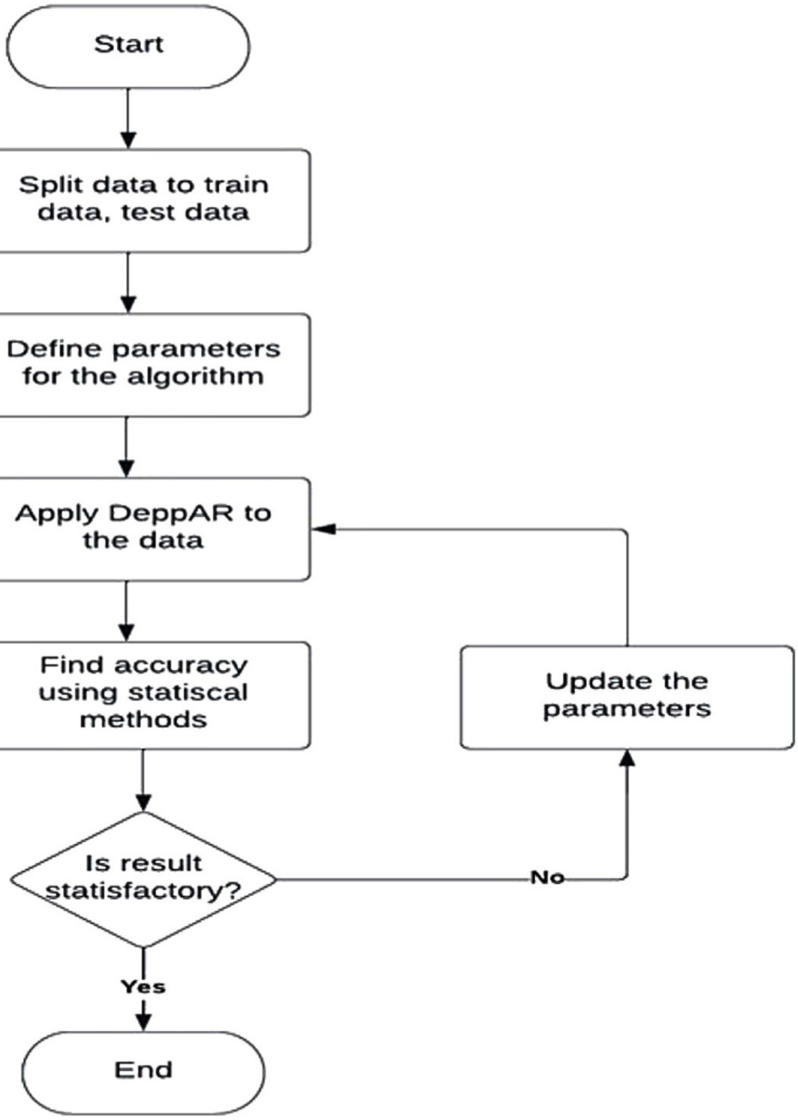

Figure 4.5.: Flow chart of DeepAR model

4.5.3. Long Term Short Memory (LSTM)

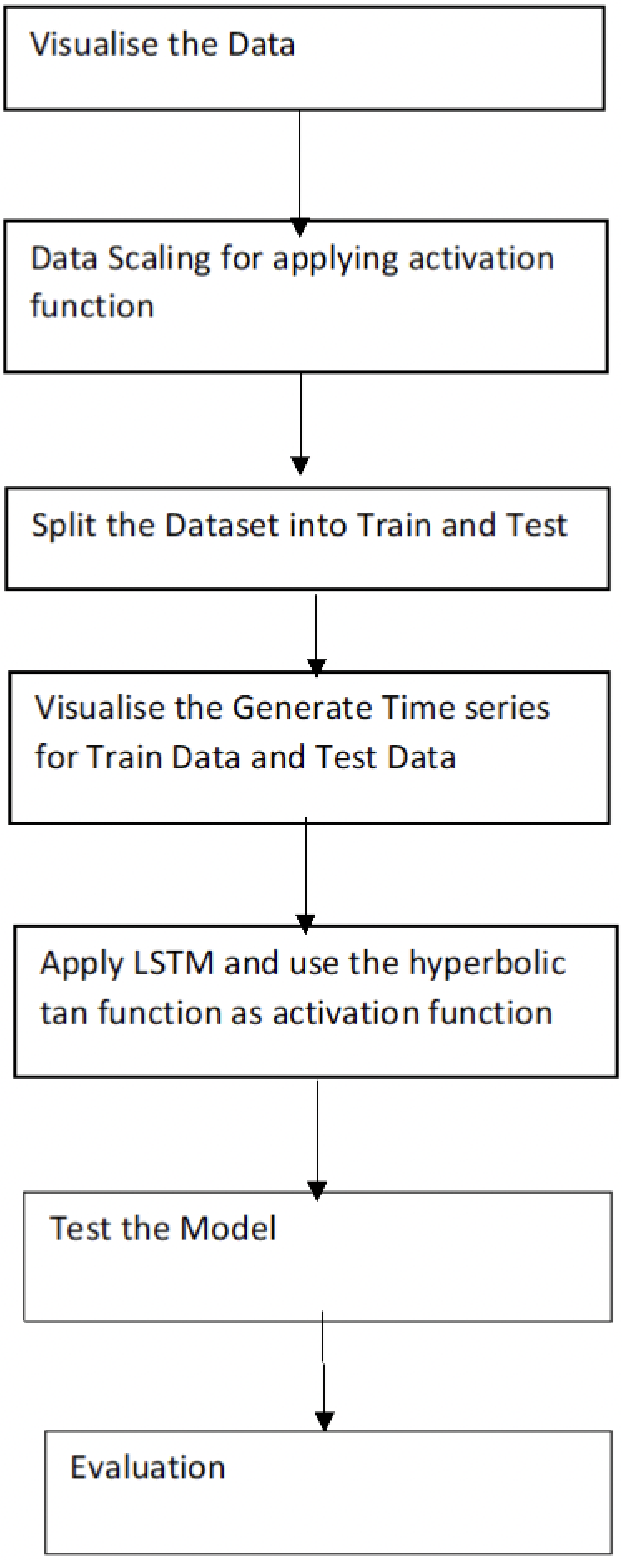

Figure 4.6.: Flow chart of LSTM

4.6. Results of the Model

Author: Venkata Satya Aditya Jagarlapudi(50%), Nirav Satani(25%), Peram Navachandu Reddy(25%)

4.6.1. LSTM Results

When the model was tested, It outperformed DEEPAR and VARIMA by having 38 countries lessthan 20 MAPE range as shown below:

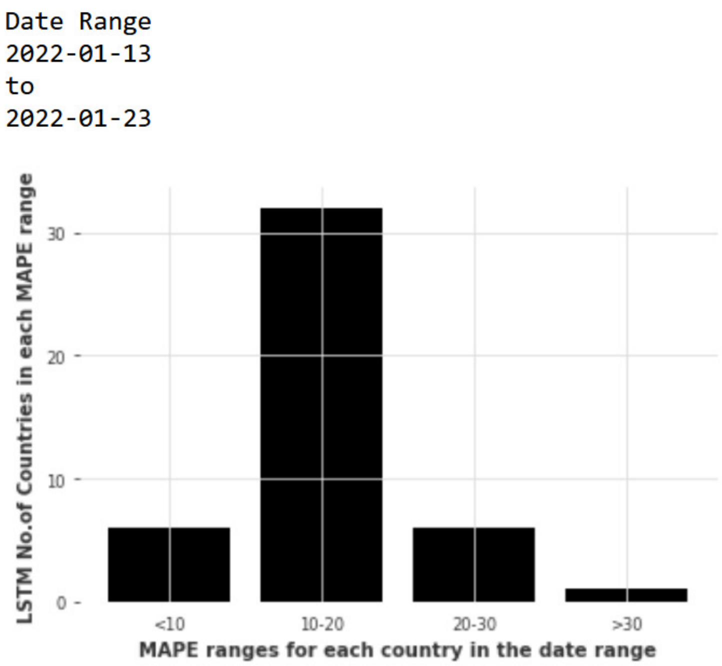

Figure 4.7.: LSTM Results

4.6.2. VARIMA Using Darts Library

In the research, we used the Darts library to perform VARIMA because it is giving more accuracy compared to other libraries.

After implementing the algorithm, we got Auto Correlation and Partial Auto Correlation graphs as shown in Figure 4.8. Darts library automatically calculates P and Q values from the ACF and PACF graphs.

Results

After calculating the MAPE for each country, we conclude that from Figure 4.9

- 10 countries got less than 10% error.
- 12 countries got an error in between 10%-20%.
- 8 countries got an error in between 20%-30%.

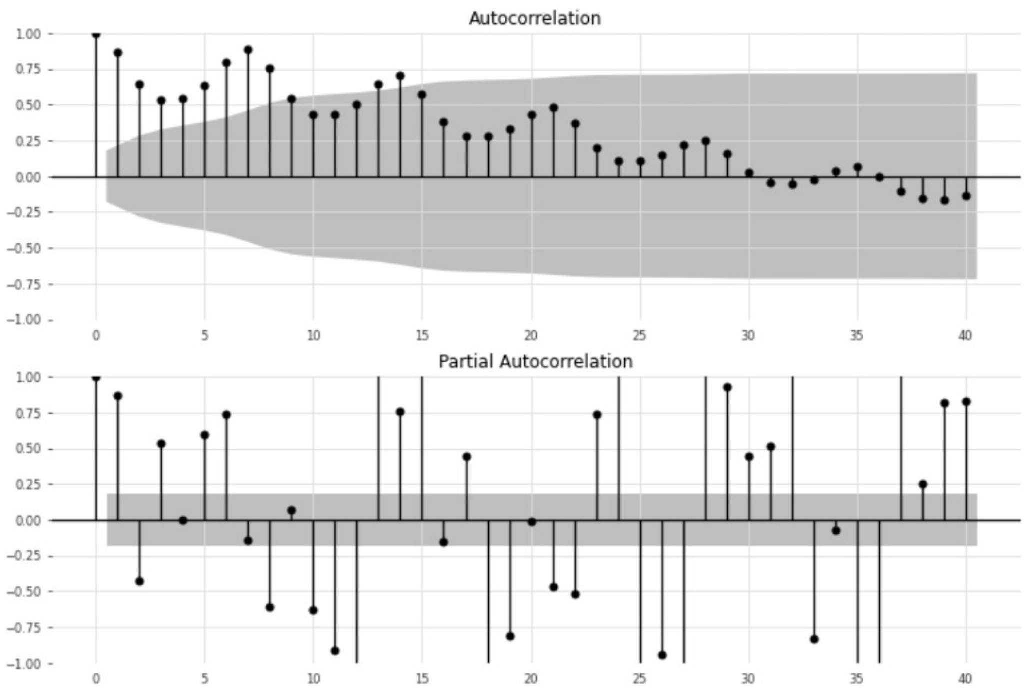

Figure 4.8.: ACF and PACF graphs

- 6 countries got an error greater than 30%.

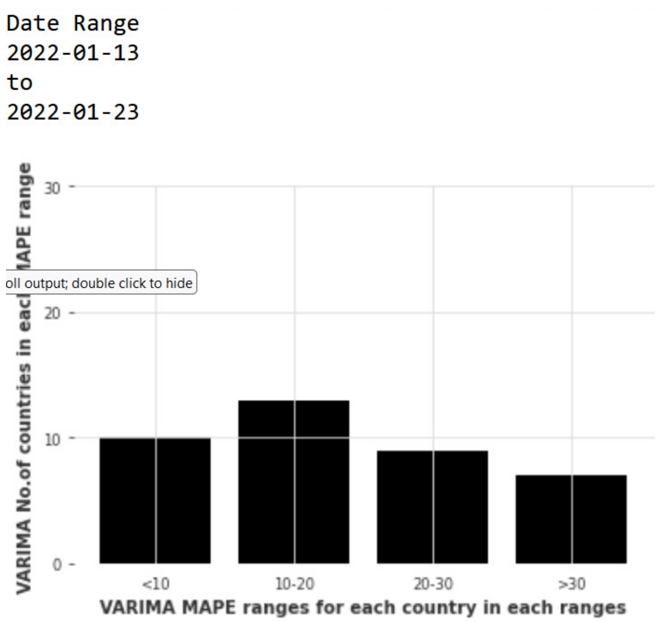

Figure 4.9.: VARIMA Results

We are not considering the algorithm because

- 6 Error is more compared to other algorithms.
- The main drawback of the algorithm is it is not able to find the hidden relations between the data which results in more errors in forecasted data.
- The historical data is very less which leads to an ineffective forecast of future values.

4.6.3. DeepAR

After dividing the data into train and test parts, we applied the DeepAR algorithm using the Dart package to our dataset. The forecast visualization of the test data is given below.

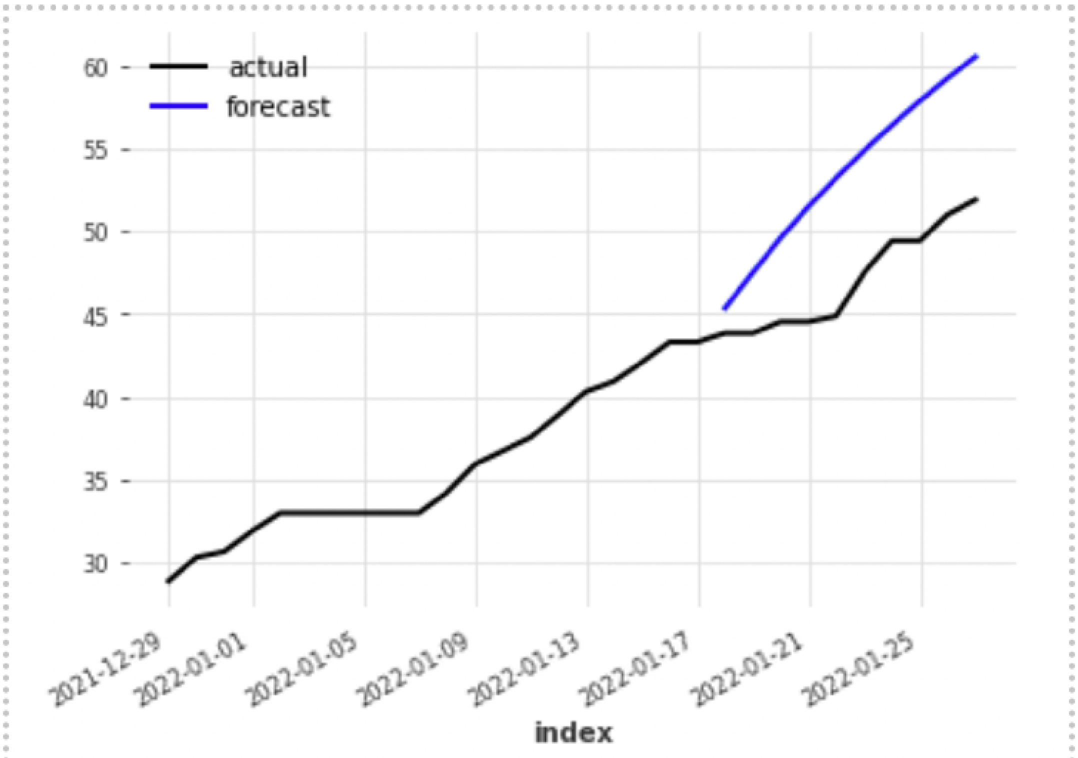

Figure 4.10.: Results

Here, we get the MAPE= 0.32%, which is higher than the LSTM algorithm. The data in our case contains only a single time-series field. However, DeepAR is efficient, it is more suitable and robust for multiple time-series forecasting and chunky data. As the error rate of DeepAR is higher, our data is straightforward (contains only one time series with uniform features), so we choose the LSTM over the DeepAR.

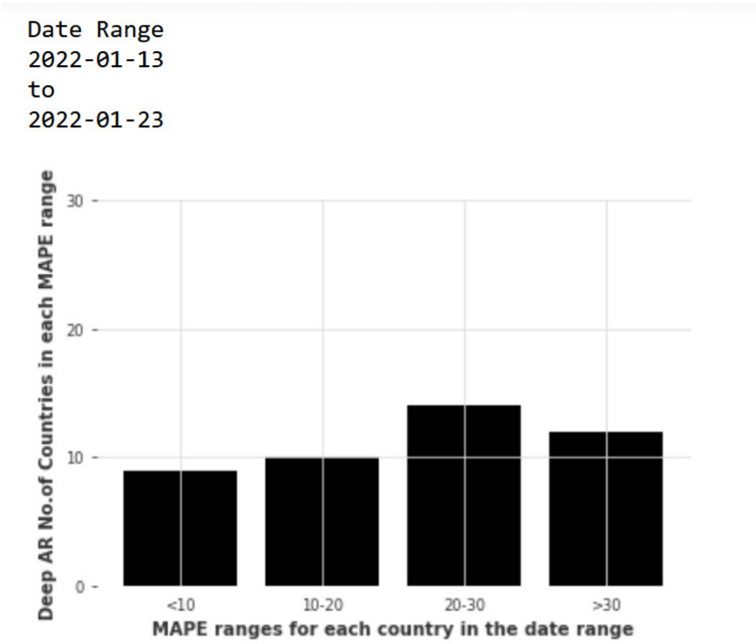

Figure 4.11.: DEEPAR Results

5. Extended Model

Author: Venkata Satya Aditya Jagarlapudi(80%), Irshad Hussain Mohammed(20%)

5.1. Vaccination

5.1.1. Vaccines and Deaths forecasting

After forecasting cases based on seats. We were curious to see the correlation between Cases, Seats, Deaths, and Vaccines. Following is the discussion of the results we achieved from forecasting using all 4 features using 3 different algorithms DeepAR, LSTM and ARIMA.

5.1.2. Metrics

Mean absolute error is on average how different is the prediction from the original value. For example, If there is a mean absolute error of 2000 cases and if the cases range from 0-1000 then 2000 would be a very high error since it's very higher than the range. In this way to MAE is independent of the range of the original values, evaluating it by itself won't give proper output. So instead we used the percentage of the difference between the range and the MAE to evaluate the outputs which is the Mean Absolute percentage error. And, a mean absolute percentage below 10% is very good. And between 10-20 is good. And above 50 is not good.

5.1.3. LSTM Results

After cleaning the data for deaths and vaccines by filling the NA and various other operations. We trained it with LSTM time series forecasting. And, We got the following results.

Below we can see there aren't many countries that have a mean absolute error percentage less than 10. But on the whole, the results are good but the results derived from just cases and seats were better

Below we can see there aren't many countries that have a mean absolute error percentage less than 10. But on the whole, the results are good but the results derived from just cases and seats were better.

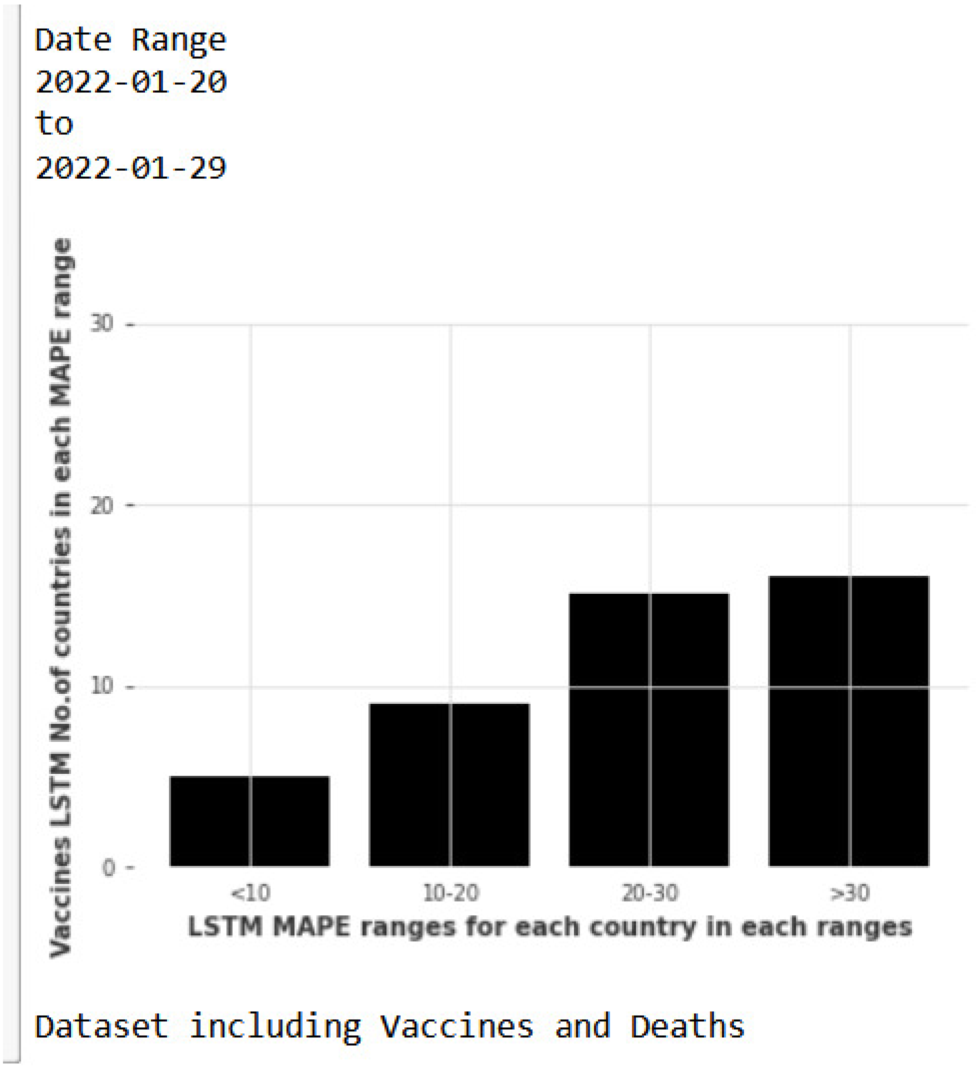

Figure 5.1.: LSTM MAPE ranges of each country

5.1.4. Reasons for Less efficiency

There are some inconsistencies in reporting of vaccines and deaths. We did proper cleaning and preprocessing but still the data on the whole might not be accurate.

And, Deaths per day in most of cases is comparatively a less number in most of the countries compared to cases and vaccines. And, the percentage compared with the population also would be a very small number too. So, even after scaling, it might lead to inconsistent results on the whole.

5.1.5. Deep AR Results

After dividing the data into train and test parts, we applied the DeepAR algorithm using the Dart package to our dataset. The forecast visualization of the test data is given below.

To evaluate the model, we also use the same mean absolute error metric that we used for other algorithms. We find some noticeable trends in the error rate with this algorithm.

- We can generate almost accurate forecasting for seven countries (error rate less than 10). Moreover, with the help of DeepAR, the error rate of nine countries rely on between 10-20%, meanwhile, another six countries are forecasting the results with a high error rate(20-30%).
- Half of the total countries accounted for a greater than 30% error rate. We cannot ignore such an error rate, hence we tried parameter tuning to get better results. In conclusion, this is the best optimal result we could get with our dataset by applying this algorithm.
- Although DeepAr is a neural network-based favorable alternative choice for forecasting, here, our dataset does not contain multiple time series fields, and the data is not too complex (in terms of features). As we can find a more accurate output with another algorithm than this, we do not require DeepAr for this data to construct our forecasting dashboard.

Although DeepAR is a neural network-based favorable alternative choice for forecasting, here, our dataset does not contain multiple time series fields, and the data is not too complex (in terms of features). As we can find a more accurate output with another algorithm than this, we do not require DeepAR for this data to construct our forecasting dashboard.

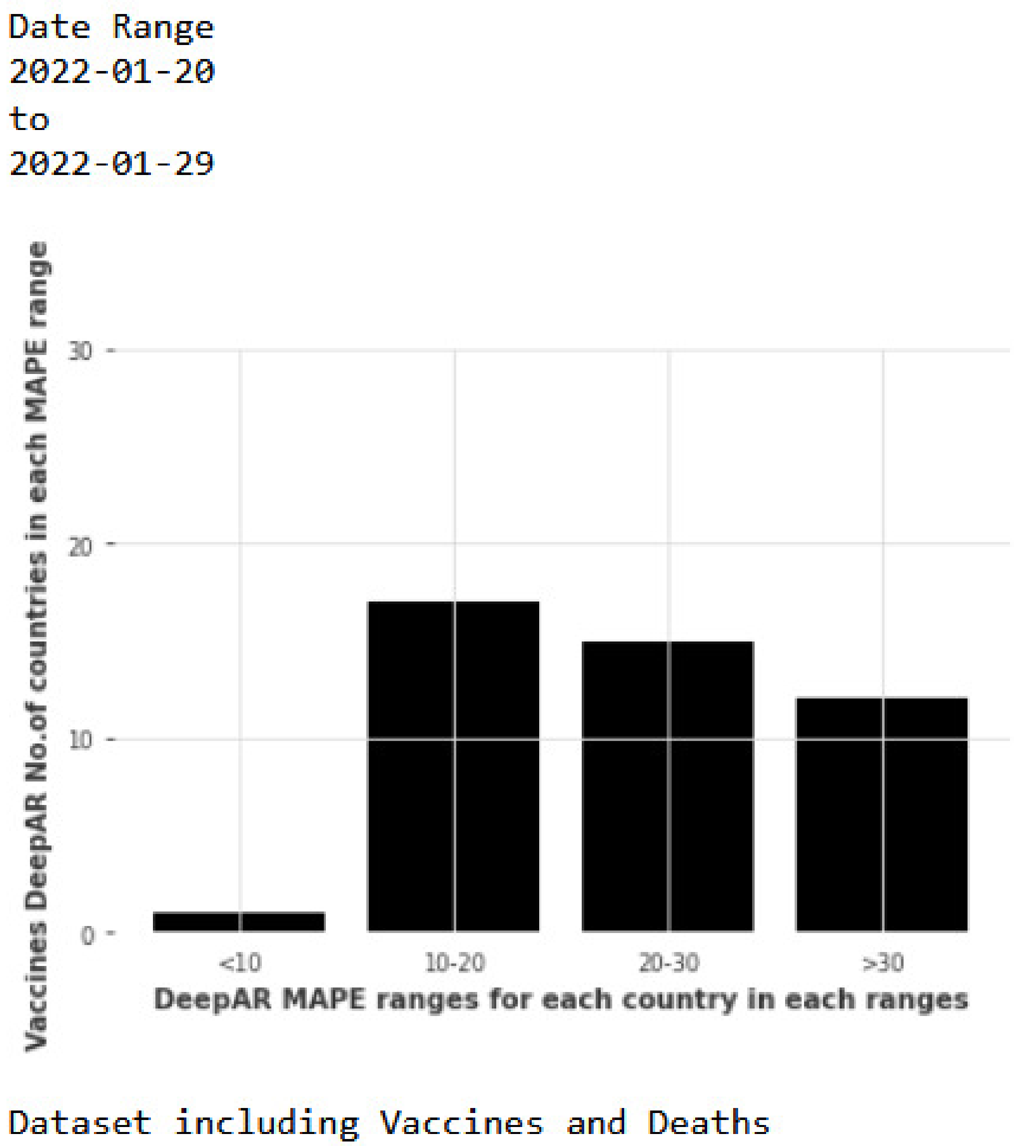

Figure 5.2.: DeepAR MAPE ranges of each country

5.1.6. VARIMA Results

The Varima Algorithm could produce result only for one country as shown in figure. In contrast to algorithms, Varima needs every country data to be trained individually. However, when tested it could only produce result for one country

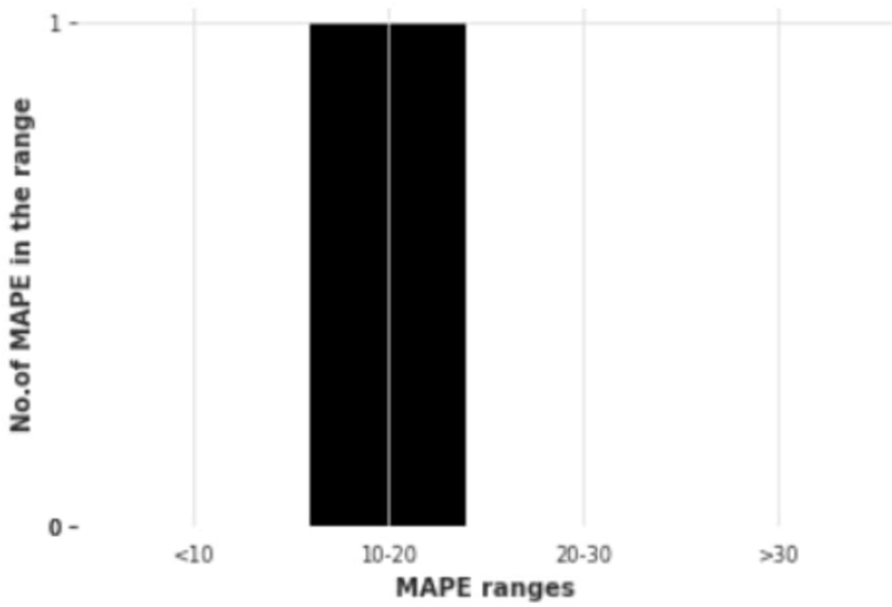

Figure 5.3.: VARIMA Results for Extended Model

6. Clustering

Author: Venkata Satya Aditya Jagarlapudi(70%), Pavan Kumar Reddy Kancharla(30%)

Clustering Using K means

After Forecasting the Values using LSTM. Since LSTM gave the most efficient outputs. The forecasted Cases and Seats are used with the incident rate. The formula for Incident rate is

Incident Rate = Cases*100000/Population

Usually the incident rate is calculated for a population of around 100000 so we multiply with 100000.

After Calculating the incident rate. The new data consists of Cases and Seats or Vaccines and Deaths in the cases of Vaccines forecasting.

K means working

1. Randomly select 'c' cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4. Recalculate the new cluster center.
5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3.

Reasons for using K means

1. It is very efficient for numerical data since we don't have any categorical data.
2. Since the data is distinct or well separated from each other. K means is efficient.
3. And, It's fast and easy to implement.

7. Dashboard

Author: Irshad Hussain Mohammed(50%), Nirav Satani(40%), Rehan Ali Syed(10%)

7.1. Incidence Response Model Dashboard

The predicted cases are used to compute the incidence rate. With the incidence rate, clusters are created using K-means clustering and depicted in a dashboard.

The dashboard has 3 clusters low, medium, and high

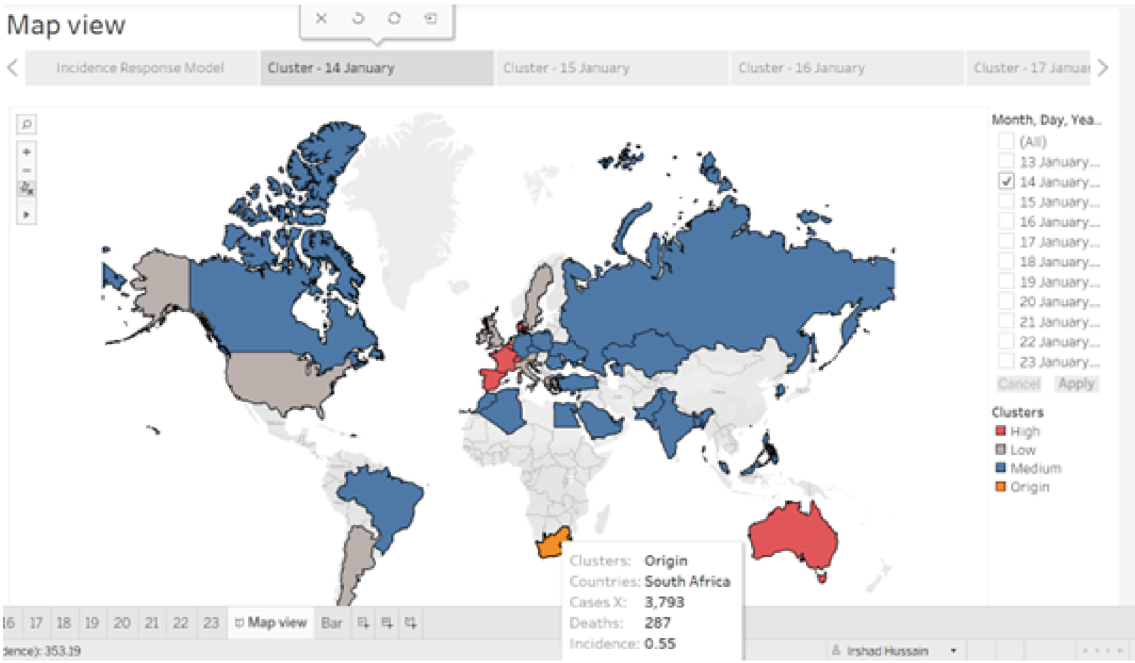

Figure 7.1.: Incidence Response Model Dashboard

The information provided in the dashboard is

- Cluster
- Country Name
- Predicted Cases
- Actual Deaths
- Incidence Rate

7.2. Vaccines vs Deaths Dashboard

In this Dashboard, The First Vaccination data from countries across world is visualized with the deaths due to omicron variant. The dashboard also clusters the countries into 10 parts based on the population vaccinated.

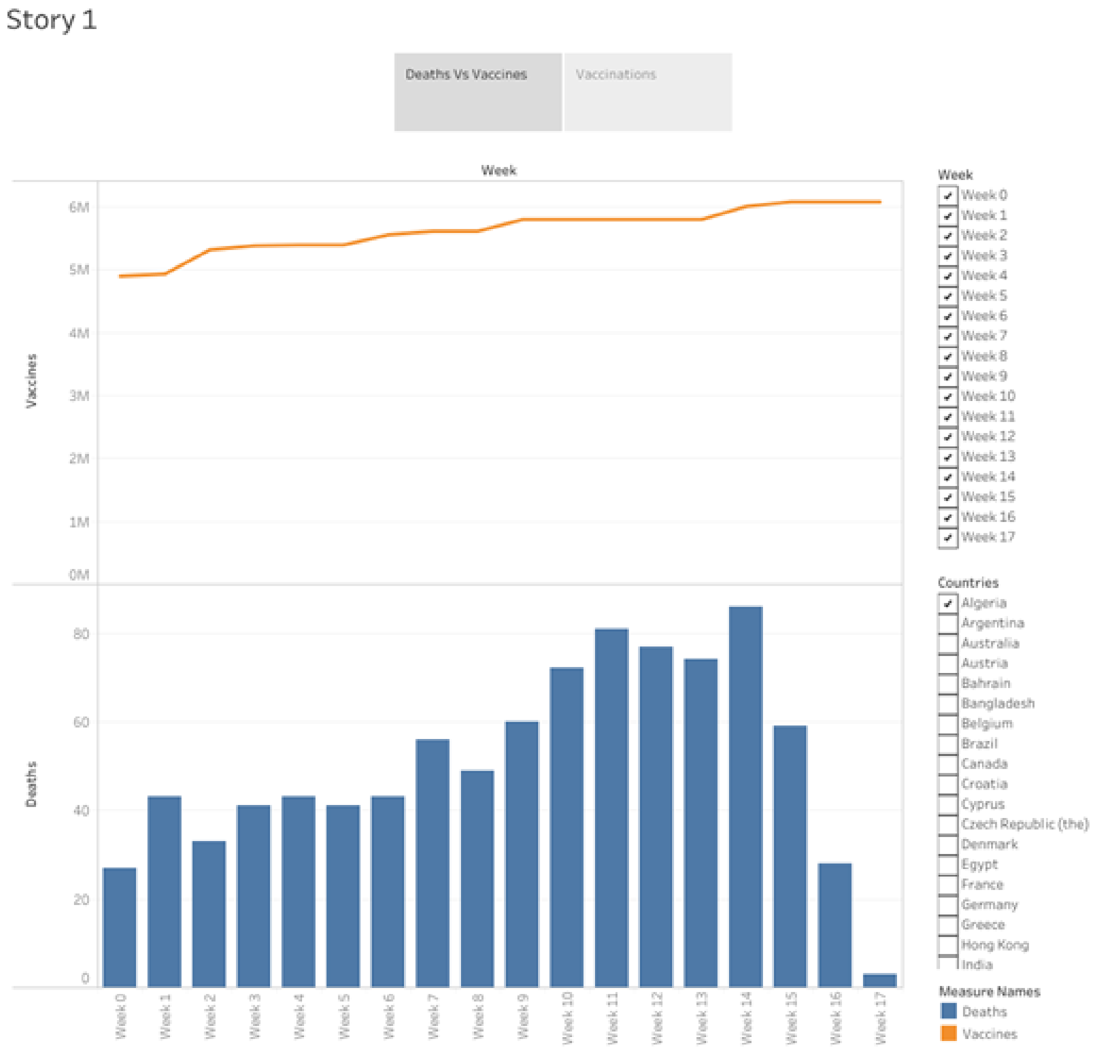

Figure 7.2.: Vaccines vs Deaths

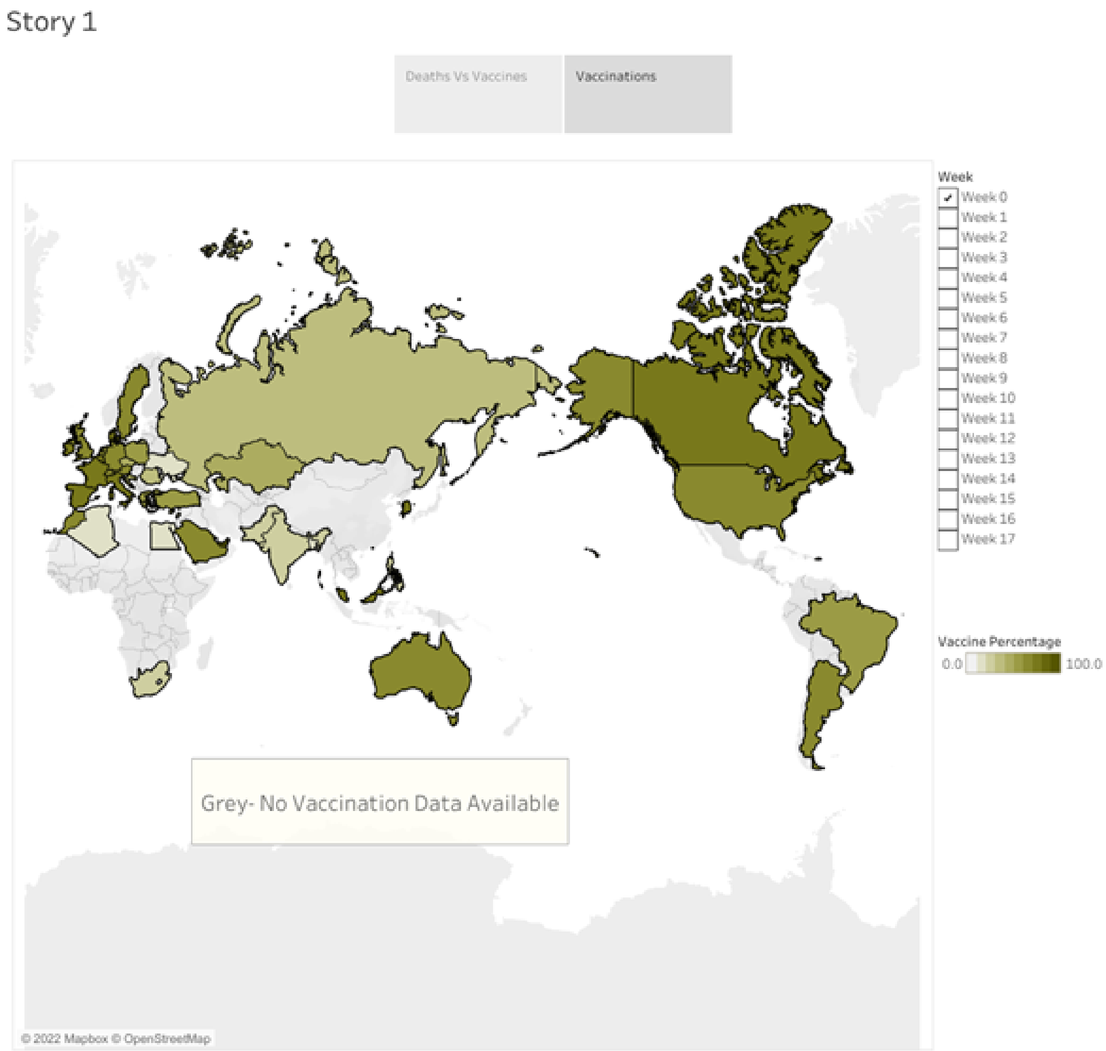

Figure 7.3.: Clusters Based on Actual Vaccinations

7.3. Predicted Vaccinations and Deaths Dashboard

The Vaccinations and Deaths predicted by LSTM model are clustered and illustrated in this dashboard.

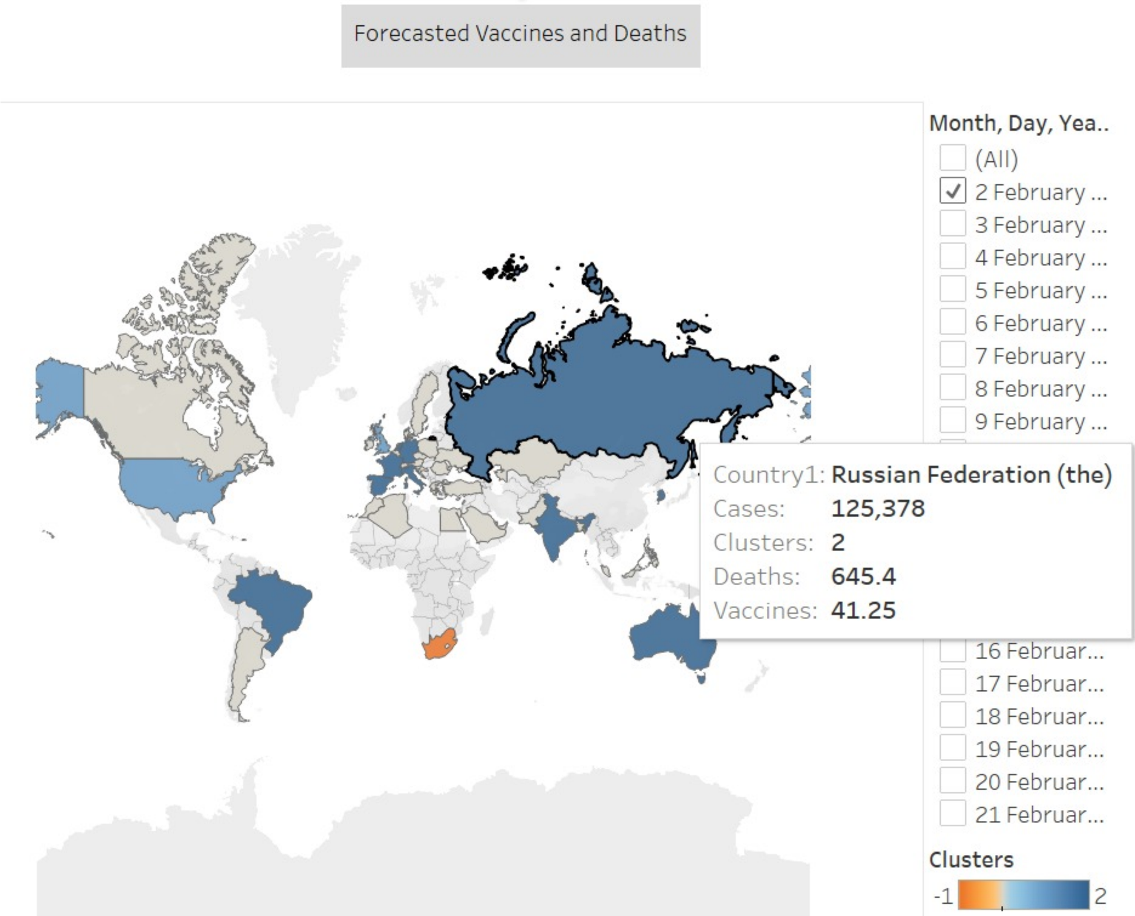

Figure 7.4.: Predicted Vaccinations and Deaths Dashboard

8. Limitations and Recommendations

Author: Peram Navachandu Reddy(50%), Ramesh Reddy Modulla(50%)

The model trained despite of being able to produce expected results and create clusters, has some limitations as the following:

- The Model is based on Third wave of Covid whereas Incidence Response should have been made by taking first wave into account.
- It is only confined to the Flights Data and doesn't consider other modes of transportations.
- It assumes flights were occupied with 60 percentage of it's capacity bot doesn't have access to actual number of passengers.
- It doesn't provide any User Interface to the stakeholders.
- It doesn't provide the flexibility to generate clusters over a period of time and it is specific.

To overcome the limitations, It is recommended to:

- Acquire the transportation data of First Covid Wave.
- Acquire and Analyze the effect of Other modes of transportation and Internal Transportation for a given Company.
- Get Accurate Data of Number of Passengers and analyzing the data by keeping Covid Restrictions in focus.
- Building a GUI to allow users to enter the desired time period and automate the pipeline of Dashboard Creation.

9. Conclusion

Author: Irshad Hussain Mohammed(100%)

In this Research Lab, Analysis of Air Traffic from the origin country of Covid-19 Omicron Variant i.e., South Africa was made to understand the impact of international travel on spread of virus. With this analysis a Model has been built by feeding the data acquired to different Time Series Algorithms and with LSTM producing more accurate results than the others, clusters were made using K-means Clustering Algorithm on the results to divide countries based on the cases predicted. Later on, an Analysis was done on the vaccinations to check how the rate of vaccinations was affecting the spread of cases and death rate. Based on the results, an LSTM Model has been built to predict vaccinations and deaths across countries. Dashboards has been created for the obtained clusters and to visualize the Rate of actual Vaccinations and Deaths. The Model being specifically confined to limited time frame can be expanded to have flexibility of producing results for any given point of time and can be automated to produce dashboards from the results obtained.

10. Technologies Used

Operating System	:	Windows 10 and above
Programming Language	:	Python
IDE	:	Jupyter Notebook
Visualization	:	Tableau Public
Dataset	:	Microsoft Excel
Version Control	:	GitHub
Management	:	Open Project
Documentation	:	Latex
File Storage	:	GitHub

11. Future scope

Author: Sudheer Kumar Gandham(50%), Pavan Kumar Reddy Kancharla(25%), Manpreeth Vankadara(25%)

11.1. Acquiring the Data of China Air Traffic

As the first disease caused by Covid-19 variant was found in China, the Incidence Response Model developed based on the outbound traffic from China can be more useful and reliable. It is recommended to acquire the outbound data from China, feed it to the model and improvise the accuracy.

11.2. Acquiring Actual Number of Passengers Information

The dataset being acquired, only provided the number of flights between South Africa and all other destinations. It is assumed that 60% of the flight capacity is full because of the covid regulations. The reliability of the model can be improved by training the model with actual number of Passengers.

11.3. Investigating Other modes of Transportation

In this model, only flights data was acquired and used to train model. Whereas, there many countries which have other modes of transportations such as road, rail and waterways. Investigating other modes of transportation and acquiring the number of passengers can result in finding the correlation between other modes and Covid-19 cases and to develop more accurate model.

11.4. Adding Flexibility of Dates

The model is restricted to a specific time frame. However, adding flexibility of generating Clusters over entire pandemic time involves considering many factors such as mask mandates, lockdown situations, travel frequency, vaccination rate, number of vaccinations. Considering all these factors if a model is developed to generate clusters at any point of time, it increases the usability.

11.5. Adding Interface to the Model

The model has an option to enter the desired date but in the python code and the clusters then generated are exported as an Excel file and then visualized in Tableau. If this is automated and the dashboard is directly connected to the dashboard, it will improvise the user experience.

Bibliography

[1] A critical moment for a new fund for pandemic preparedness — center for global development — ideas to action. https://www.cgdev.org/blog/critical-moment-new-fund-pandemic-preparedness. (Accessed on 12/08/2022).

[2] Shabir A Madhi, Gaurav Kwatra, Jonathan E Myers, Waasila Jassat, Nisha Dhar, Christian K Mukendi, Amit J Nana, Lucille Blumberg, Richard Welch, Nicoletta Ngorima-Mabhena, et al. Population immunity and covid-19 severity with omicron variant in south africa. New England Journal of Medicine, 386(14):1314–1326, 2022.

[3] How quickly omicron is spreading in south africa? - the mail & guardian. https://mg.co.za/health/2021-12-21-how-quickly-omicron-is-spreading-in-south-africa/#:~:text=The%20Omicron%20variant%20was%20identified%20in%20South%20Africa,wave%2C%20which%20was%20driven%20by%20the%20Delta%20variant. (Accessed on 12/02/2022).

[4] Amanda MY Chu, Agnes Tiwari, and Mike KP So. Detecting early signals of covid-19 global pandemic from network density. Journal of travel medicine, 27(5):taaa084, 2020.

[5] Omar Sharif, Md Rafiqul Islam, Md Zobaer Hasan, Muhammad Ashad Kabir, Md Emran Hasan, Salman A AlQahtani, and Guandong Xu. Analyzing the impact of demographic variables on spreading and forecasting covid-19. Journal of Healthcare Informatics Research, 6(1):72–90, 2022.

[6] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science, 368(6489):395–400, 2020.

[7] Yusuf Jamal, Mayank Gangwar, Moiz Usmani, Alison E Adams, Chang-Yu Wu, Thanh H Nguyen, Rita Colwell, and Antarpreet Jutla. Identification of thresholds on population density for understanding transmission of covid-19. GeoHealth, 6(9):e2021GH000449, 2022.

[8] Sylvia Xiao Wei Gwee, Pearleen Ee Yong Chua, Min Xian Wang, and Junxiong Pang. Impact of travel ban implementation on covid-19 spread in singapore, taiwan, hong kong and south korea during the early phase of the pandemic: a comparative study. BMC infectious diseases, 21(1):1–17, 2021.

[9] Hien Lau, Veria Khosrawipour, Piotr Kocbach, Agata Mikolajczyk, Hirohito Ichii, Maciej Zacharski, Jacek Bania, and Tanja Khosrawipour. The association between

international and domestic air traffic and the coronavirus (covid-19) outbreak. Journal of Microbiology, Immunology and Infection, 53(3):467–472, 2020.

[10] Flightaware - flight tracker / flight status. https://flightaware.com/. (Accessed on 12/02/2022).

[11] A multivariate time series modeling and forecasting guide with python machine learning client for sap hana — sap blogs. https://blogs.sap.com/2021/05/06/a-multivariate-time-series-modeling-and-forecasting-guide-with-python-machine-learning-client-for-sap-hana/. (Accessed on 12/02/2022).

[12] A thorough introduction to arima models — by lleyton ariton — analytics vidhya — medium. https://medium.com/analytics-vidhya/a-thorough-introduction-to-arima-models-987a24e9ff71. (Accessed on 12/04/2022).

[13] Deepar forecasting algorithm - amazon sagemaker. https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html. (Accessed on 12/04/2022).

[14] Deepar hyperparameters - amazon sagemaker. https://docs.aws.amazon.com/sagemaker/latest/dg/deepar_hyperparameters.html. (Accessed on 12/04/2022).

[15] Prophet vs deepar: Forecasting food demand — by sam mourad — towards data science. https://towardsdatascience.com/prophet-vs-deepar-forecasting-food-demand-2fdebfb8d282. (Accessed on 12/04/2022).

[16] Lstm networks — a detailed explanation — towards data science. https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9. (Accessed on 12/08/2022).

[17] Long short-term memory - wikipedia. https://en.wikipedia.org/wiki/Long_short-term_memory. (Accessed on 12/08/2022).

[18] What is lstm? — packt hub. https://hub.packtpub.com/what-is-lstm/. (Accessed on 12/08/2022).

[19] Understanding lstm and its diagrams — by shi yan — ml review. https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714. (Accessed on 12/08/2022).

[20] Understanding of lstm networks - geeksforgeeks. https://www.geeksforgeeks.org/understanding-of-lstm-networks/. (Accessed on 12/08/2022).

[21] Flightaware - flight tracker / flight status. https://flightaware.com/. (Accessed on 12/02/2022).

[22] aviationstack - real-time flight tracker & status api. https://aviationstack.com/. (Accessed on 12/02/2022).

[23] Martin Strohmeier, Xavier Olive, Jannis Lübbe, Matthias Schäfer, and Vincent Lenders. Crowdsourced air traffic data from the opensky network 2019–2020. Earth System Science Data, 13(2):357–366, 2021.

[24] The opensky network - free ads-b and mode s data for research. `https://opensky-network.org/`. (Accessed on 12/02/2022).

[25] Population density (people per sq. km of land area) — data. `https://data.worldbank.org/indicator/EN.POP.DNST`. (Accessed on 12/02/2022).

[26] Though far from the norm, nearly 500 us flights per day depart more than 70% full — cnn travel. `https://edition.cnn.com/travel/article/flight-capacity-united-states-coronavirus/index.html`. (Accessed on 12/02/2022).

[27] What is correlation? — correlation examples — displayr.com. `https://www.displayr.com/what-is-correlation/`. (Accessed on 12/02/2022).

[28] Our world in data - covid-19 — kaggle. `https://www.kaggle.com/datasets/caesarmario/our-world-in-data-covid19-dataset`. (Accessed on 12/02/2022).

[29] List of countries and dependencies by population density - wikipedia. `https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population_density`. (Accessed on 12/02/2022).

[30] Identification of thresholds on population density for understanding transmission of covid-19 - pmc. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9347488/`. (Accessed on 12/02/2022).

[31] Impact of travel ban implementation on covid-19 spread in singapore, taiwan, hong kong and south korea during the early phase of the pandemic: a comparative study — bmc infectious diseases — full text. `https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-021-06449-1`. (Accessed on 12/02/2022).

[32] The association between international and domestic air traffic and the coronavirus (covid-19) outbreak - sciencedirect. `https://www.sciencedirect.com/science/article/pii/S1684118220300864?via%3Dihub`. (Accessed on 12/02/2022).

A. Appendix

A.1. Correlation

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/Correlation/Corelation.ipynb

A.2. Dataset for Predicting Cases

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/PredictingCases/Dataset.csv

A.3. Model Implemented with Algorithms

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/PredictingCases/All%20Algorithms.ipynb

A.4. Clusters created on LSTM Model Output

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/PredictingCases/LSTM%20Model%20with%20Clusters.ipynb

A.5. Extended Model Dataset

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/Extended%20Model/Dataset.csv

A.6. Extended Model Algorithm

https://github.com/lotzmann/ResearchLabST2022-Group2/blob/main/Extended%20Model/Final_Model_Extended.ipynb

A.7. Dashboard for Clusters of risk areas based on cases

https://public.tableau.com/app/profile/irshad.hussain/viz/IRM_166348349724 40/Mapview

A.8. Dashboard for Clusters of Actual Vaccinated Population

https://public.tableau.com/app/profile/irshad.hussain/viz/DeathsvsVaccines /Story1

A.9. Visualisation of Actual Vaccines and Deaths

https://public.tableau.com/app/profile/irshad.hussain/viz/DeathsvsVaccines /Story1

A.10. Dashboard for Predicted Vaccinations and Deaths

https://public.tableau.com/app/profile/irshad.hussain/viz/Forcastedvaccine s/Story1

A.11. Minutes of Meeting

https://github.com/lotzmann/ResearchLabST2022-Group2/tree/main/MinutesOfM eeting

A.12. Presentations

https://github.com/lotzmann/ResearchLabST2022-Group2/tree/main/Presentati ons