BUSINESS ANALYTICS PROJECT ON
VISUALISATION AND TIMESERIES FORECASTING

A PROJECT REPORT

Submitted by

LOKESH B

18AA21

In partial fulfilment of the requirements for the award of

MASTER OF BUSINESS ADMINISTRATION

MARCH 2020

CERTIFICATE

This is to certify that the project work titled

BUSINESS ANALYTICS PROJECT ON VISUALISATION AND TIMESERIES FORECASTING

is a bonafide work done by

LOKESH B

18AA21

In partial fulfilment of the requirements for the award of

MASTER OF BUSINESS ADMINISTRATION

Dr T G Vijaya	Dr R Chitra
Director	Faculty Guide

Submitted for the Viva- Voce examination held on _____

(Signature of Internal Examiner with date)	(Signature of External Examiner with date)

ACKNOWLEDGEMENT

I thank our Director, **Dr. T G Vijaya** for providing me with an opportunity to work in this project and, for the help that she has rendered throughout the course.

I would like to express my sincere gratitude and dedication to my guide and mentor **Dr. Chitra** and **Dr. A Mansurali**, who has been supportive and a great source of encouragement for this project. I also thank them for their valuable help and knowledge that has been shared with me. I also thank all our Faculty members for their support in the completion of the course and a special thanks to my course coordinator **Dr. J Joshua Selvakumar** who has been supportive throughout the course.

Finally I take this opportunity to thank my parents, my friends and everyone who directly or indirectly provided constant support and enduring encouragement.

Lokesh B

EXECUTIVE SUMMARY

This report is about my MBA project, which has been done on different phases of learning and applying the learning to one master project. The learning projects are project on visualization of website analytics data to generate business intelligence and the other on air passenger travel forecasting using timeseries modelling. The master project is on International trade Forecasting of India on which the learnings from previous projects are applied.

The objective of this project was to study various tools available for data analytics and various machine learning techniques for timeseries modelling and to apply them on a impactful area. In this project I learnt a Business reporting tool "Google datastudio" on which a report on website visitors is created for generating insights from the website analytics data. Also, learning timeseries modelling combined with machine learning techniques which has been applied for the purpose of forecasting the air passengers travel and International trade.

The data used for the project is collected from google analytics tool, R language and the data on the international trade has been scraped from World Trade Organisation website and reports. With these data in hand various timeseries models have been built on these data and one best model has been selected among the models for deployment in out of sample forecasting. The models chosen for this purpose are Simple Exponential Smoothing, Auto Regressive model, Auto Regression Integrated with Moving Average, Artificial Neural networks, Multi-layer Perceptron Neural Networks and TBATS model. These models are then evaluated on the forecast accuracy metrics such as Mean Error, Root Mean Square Error, Mean Absolute Error, Mean Percentage Error, Mean Absolute Percentage error. The models are compared with each other across these accuracy metrics to select a best model.

The result of the model building exercise leaves us with the Multi-Layer perceptron model being performing well for forecasting international trade and ARIMA model for forecasting Air Passengers, on evaluating with forecast accuracy metrics. Also when these models are deployed for out of sample prediction these models are found to be forecasting with least amount of deviation from the actual values. Though these models forecasts well, the volatility nature of the International trade was difficult to observe into the data and further models can be built to observe the volatility in the trade by incorporating timeseries data about other macroeconomic variables. Thus, making this univariate timeseries modelling into a multivariate timeseries modelling for better forecasting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 01

1. INTRODUCTION

1.1 BACKGROUND

1.1.1 Back ground to Visualisation of Web analytics data:

The rise of social media and information sources on the internet has highlighted the need to access digital content from different sources of information related to specific topics. The content may be enhanced and enriched by entities producing derivatives such as documents, comments, renditions, film trailers, music remixes, etc. Analytical devices provide very powerful tools for visualization using various plots on an analytics visualization dashboard. For this purpose, Google data studio is used for Visualization of visitors report. Google Data Studio is a simple to use, customize and distribute dashboard and reporting tool. This helps to transform the data into attractive and informative reports for the audience. It is a great tool for monitoring KPIs, which support business strategies and produce periodic reports.

1.1.2 Back ground to Air passengers forecasting:

Forecasting air travel demand is an integral part of an airport's plan. It reflects the capacity utilization, which is crucial when it comes to making decisions. For the development of infrastructure facilities and reduction in the airport risk, it is important to evaluate and forecast the volume of air passenger demand in the future. The Peak demand in passenger flows at airports, are typically determined by cyclical and seasonal patterns. Thus, it is essential to manage the facilities such as passenger terminal capacity planning and design, runway, and to cover demand during the horizon of planning. Terminal capacity, runway utilisation and the availability of facilities handling arrival and departure of passengers, are the main entities that will affect the required capacity. This project focused on developing a timeseries model to forecast the demand growth in the airline passengers flow.

1.1.3 Background to International Trade Forecasting:

World Trade Organisation (WTO) established in 1995 replacing General Agreement on Tariffs and Trade (GAAT) is the largest international organisation responsible for economic cooperation between the nations of the world. It regulates trade in goods and services, intellectual properties of the countries and provides frameworks and dispute resolution process to make the global trade smooth. In 2008 WTO said foreign trade plays an important role in globalising the economy of the nation [42]. Globalisation of the economy makes the country

to explore growth opportunities outside the nation's border easily accessible and helps in development of the economy of the world as a whole.

Given the context, for every business in the country it is crucial to know which industries are likely to see growth in global trade, when it comes to performance of industries in domestic and foreign markets the interest always lies in the hands of private companies and public authorities. For these entities to make correct decisions, forecasts of key economic variables are very important. In the globalised world, where the dependency of nations on each other is increasing in an exponential levels it is much more important to predict the key economic variables like GDP, Trade, Inflation of a country.

Since the opening of India's border for global trade in the year of 1991, foreign trade plays a crucial role in the GDP of the country. India, sixth largest economy in the world, whose imports accounts for 23.64% and exports account for 19.74% of GDP in the year 2018 [41]. India exports around 7500 commodities and imports 6000 commodities. So, for a country like India which enjoys opportunity of demographic dividend than any other nation in the world, it becomes crucial to forecast the international trade component of GDP, to carry the nation towards the path of growth. The forecasts of this nature serves the purpose of policy making and helps the policy makers to take required actions to put the economy in the path of growth. With trade playing important role in India's GDP it's no surprise that forecasts of exports and imports with various trading blocs, region and countries in the world are very important to determine trade deficit and surplus with other nations and make policies to make the nation an economic power house.

1.2 PROJECT OVERVIEW

1.2.1 Over view to Visualisation of Web analytics data:

Google analytics tool captures many datapoints about a website for analysis of visitors preferences and behaviours on the site. With the idea do automate the analysis of website visitors and making the process of generating insights from the data faster, the project creates a visitor dashboard for the purpose of report generation and insights generation. The dashboard updates itself in the real time, thus providing real time insights on the webpage.

1.2.2 Overview to Air passengers forecasting:

Time series analysis is one of powerful forecasting tools which can predict the future with good accuracy. The classic Box and Jenkins airline wants to forecast the number of passenger who will travel in their air line for the period of 1961 over each month. For

forecasting the number of passengers they have the monthly data of passenger travelled in their airlines from 1949 to 1960 measured in thousands.

- Exploratory data analysis
- Treating the data – Decomposing, De-seasonalising, Differencing.
- Fitting ARIMA Model
- Predicting or forecasting the passengers for the period 1961
- Evaluating the model and finding accuracy measures.

1.2.3 Overview to International Trade forecasting:

Given the importance of the foreign trade, forecasting is very important for every business to strategize them. The two prominent organisations, Organisation for economic Cooperation (OECD) and International Monetary Fund (IMF) are the leading providers of the forecasts, macroeconomic data and performance of every countries under their ambit. To forecast both organisations use large structural economic models and heavily rely on expert judgement in making predictions. Taking a timeseries modelling approach the project tries to find the best timeseries model which can predict the value of international trade. The data for the project is taken from World trade Organisation (WTO) World trade statistics from the year 1948 to 2018 and applies various machine learning and timeseries methods to forecast the value of the trade with least error values.

Since, only the value of trade is considered for forecasting the international trade which is only the variable considered to model only univariate models can be applied to study the data. Such univariate models identified are Simple Exponential Smoothing, Auto regression models, Auto Regressive Moving Averages (ARMA) model, Auto Regressive Integrated Moving Averages (ARIMA) model are applied in addition to these models advanced models such as Artificial Neural Networks, Multilayer perception Neural Networks, TBATS are applied. All these models are built on total value of merchandise goods traded from 1948 to 2017 and results are compared with that of 2018 to predict the value. And the difference between observed value and forecasted value are analysed to select the best model.

1.3 PROBLEM STATEMENT

1.3.1 Problem statement of Visualisation of Web analytics:

The basic basis of web analytics is the collection and analysis of data about website usage. Today web analytics are used for various purposes in many industries, including traffic

monitoring, optimization of e-commerce, marketing / advertising, web development, information architecture, enhancement of website efficiency, web-based campaigns / programs etc. In web analytics the fundamental problem is to analyse what activity drives sales. It could be certain traffic patterns or user paths (and where friction is highest / lowest on those paths — and for predictive / prescriptive, we can dynamically show if friction at any point increases or decreases), attribution models to show how different channels and campaigns factor into different touch points. The need for Visualization in web analytics is to easily visualize various parameters about the users visiting the website, and we have used Google data studio for visualizing the same.

1.3.2 Problem statement of Air passengers forecasting:

Population across the nations of the world are increasing and the globalisation of the world makes the people to travel across the globe in seek of job, better life, business opportunities, etc. so for an airline industry to capitalise this opportunity and grow their business the future projections becomes crucial. Based on these projections in the business only the companies can make strategic decisions. So, forecasting becomes very important for the airline industry. In forecasting many models and methods are available, here ARIMA model is used because of it performance in forecasting out of sample data.

1.3.3 Problem statement of International Trade forecasting data:

International trade is hard to forecast and existing methods and models are very complex in determining the future as trade can be impacted by many external variables. The project tries to address the problem of forecasting the value international trade of India with various nations and world using timeseries approach. By employing time series method we can determine timeseries models can be applied to forecast the trade. For this purpose the project studies from the results of the model, that whether timeseries models can be applied to predict future value of trade since trade depends on many macroeconomic, political and environmental variables. The project applies simple timeseries models to complex machine learning algorithms to determine the best model to forecast the value of merchandise goods traded. For this purpose the univariate models such as from basics simple exponential smoothing, ARIMA, TBATS, ANN and Multi-Layer perception neural networks. From these models the best model which can predict or forecast is selected for deployment.

1.4 SCOPE OF PROJECT

1.4.1 Scope of visualisation of web analytics:

The web analytics data from google analytics tool is fed into the Google data studio software for creating dashboards. The dashboards about the website visitors are created for he purpose of making reports in the real time and generating insights about the performance of the website and visitors to it. The project limits itself only to the analysis of the behaviour of the visitors only. Thus the created report can offer real time intelligence for the website owners to tune the website for the wishes and preferences of the visitors to it.

1.4.2 Scope of Air passengers forecasting:

Forecasting for any industry is crucial to project its revenue and growth opportunities. Here airline industry passengers growth is forecasted using ARIMA model. The limitations and assumptions of the ARIMA model are taken into considerations and then the model is built. Based on the assumptions and limitations, required pre-processing is done on the data and the pre-processed data is split into train and test for building the model. The model is applied for out of sample forecasting and the accuracy measures are derived.

1.4.3 Scope of International trade forecasting:

The international trade has two components to it one is export and other is import. India trades both goods and services to the foreign nations. The project limits itself to determine the best model to predict the value of merchandise goods. So, the models that are above said, are built upon the timeseries data of total merchandise goods trade with world is collected from the WTO and three commonly traded commodity categories in the world that are Agriculture, Fuels and Manufactures are used to model both exports and imports separately and compared to find the best model. If the model forecasts for the above mentioned categories are predicted with a good amount of accuracy then best model among them can selected based on least amount of error and then applied to forecast trade value for other commodities across various countries.

1.5 OBJECTIVE

1.5.1 Objective of visualisation of web analytics:

The objective of the project is

- To create a dashboard to monitor the visitors of the website and generate reports as and when required.

- To generate insights from the generated report, for better decision making in modifications to the website.

1.5.2 Objective of Air passengers forecasting:

The objective of the project is to model an ARIMA Timeseries model for Air passengers data and forecast the passengers for next horizon.

1.5.3 Objective of International trade forecasting:

The objectives of the study are

- To determine whether International Trade can be predicted/forecasted using timeseries modelling.
- To model the trade timeseries data to forecast the future.
- To find the best univariate time series model among the models chosen based on the complexity of each models.

CHAPTER 02

2. LITERATURE REVIEW

Timeseries forecasting involves in predicting future occurrence based on past behaviour. Timeseries models have explanatory power of their coefficients to make prediction and works well for out of sample data [27]. In the macroeconomic models it is often difficult to explain the causal variables and causality determination is a complex work[28]. Thus timeseries approach can be applied to forecast the trade with good amount of accuracy.

Simple exponential smoothing is widely used model of forecasting which smooths original series to forecast the future values[20]. The popularity can be attributed to the simplicity of the model, its efficiency in computation, responsiveness of the model and the reasonable accuracy[29]. This technique is a simple and pragmatic approach to forecasting, as the forecast is constructed from an exponentially weighted average of past observations. The largest weight is assigned to the present observation, less weight to the preceding values (exponential decay) of past data [03].

Relationship between foreign trade, sustainable economic growth and a environmental protection, the impact of terms of trade and terms of trade volatility on economic growth using time series data was examined in article (Wong, 2010) [40] and the results of the generalized forecast error variance decomposition has shown a favourable and less volatile terms of trade are important for economic growth. In study (Rahman & Mamun, 2016) [31] was provided with an evidence of the trade growth hypothesis over the energy-led growth hypothesis for the Australian macro economy. Numerous studies have examined the relationship between trade liberalization policies and economic growth (Manwa & Wijeweera, 2016; Salahuddin & Gow, 2016) [27].

Artificial neural networks are one of the accurate models that are employed in forecasting of economic, social, engineering applications today. They are data driven self-adaptive methods and of non-linear nature[43]. According [26] to neural networks are numeric in nature, which are suitable for processing numeric data such as economic indicators and financial information. They can be applied to short term prediction and it provides good accuracy for such forecasts [44]. [01] applied artificial neural network (ANN) for forecasting government size in Iran. They made comparisons various transfer functions, architectures, and learning algorithms on the operation of network.

Multilayer perceptron (MLP) [02], have been extensively implemented in financial market for price prediction. Among the many MLP training algorithms, genetic algorithms (GA) can be used to determine the initial weights of the MLP [22]. Support vector regression (SVR) has become a popular tool for time series prediction [24].

Robert et al. (2011) [33] covered time series analysis and applications it presents a balanced and comprehensive treatment for both frequency and time domain methods with accompanying theory. In addition to coverage of classical methods of time series regression, ARIMA (autoregressive integrated moving averages) models, spectral analysis and state-space models. In his page, Robert (2014) [34] has presented linear regression and time series forecasting models with focus on ARIMA models. Testing for trends in ARIMA models was also found in Rob (2014)[32]. International trade is most important engine for economic growth and exports are its essential parts (Ali & Talukder, 2009) [05]. Exports have strong a relationship and positive impact on economic growth (Alam, 2011) [04]. The theory of Comparative Advantage and Customs Union provides basis for the formation of regional organization (Ruffin, 2002) [36].

TBATS model may seem peculiar, as given the number of other available models((Box, et al., 2016) [09] ,(Zeliaś, et al., 2016) [43] ,(Armstrong, 2001) [06] ,(Brockwell & Davis, 1996) [10]. The most frequently used models are ARIMA/ARMA, (Box & Jenkins, 1976) [08],(Lee & Ko, 2011)[25], (de Andrade & da Silva, 2009)[14] , (Pappas, et al., 2008) [30], (Chen, et al., 1995)[12] and exponential smoothing(Taylor, 2003)[39],(Hyndman, et al., 2008) [19].Using the ARIMA and exponential smoothing models is perfectly appropriate as long as they are not used for very complicated time series. These models are used in our previous works, as they were sufficient for the investigation of other phenomena (Brożyna, et al., 2016)[11],yet forecasts of more complicated time series require more advanced models that employ Bayesian procedures (Cottet & Smith, 2003)[13],Gaussian processes (Blum & Riedmiller, 2013)[07],ant colony optimization (Dongxiao, et al., 2010)[16], and many other methods (Zhou, et al., 2006) [44],(Taylor, et al., 2006)[38], (Küçükdeniz, 2010) [23] depending on the specifics of the data.

CHAPTER 03

3. METHODOLOGY

3.1 PROJECT METHODOLOGY

3.1.1 Project methodology of visualisation of web analytics:

The study employees the visualisation techniques with the aim of creating a dashboard to track and maintain the website traffic. The tool selected for this purpose is google data studio. The methodology employed for the data visualisation is usage of different types of visualisation to create an interactive dashboard to monitor the website visitors.

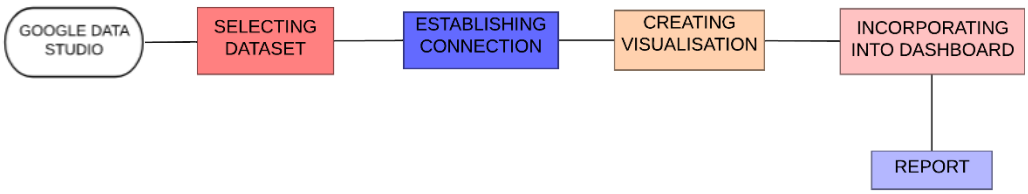

Figure 01 Methodology of visualisation project

3.1.2 Project methodology of Air passengers forecasting:

The study employees use of ARIMA model for forecasting the passengers travel in airlines. The methodology is explained in the image shown below

The data is fed into ARIMA model taking into the considerations of assumptions and constraints of the model as mentioned in figure 02 and required pre-processing is done before feeding the data to the model.

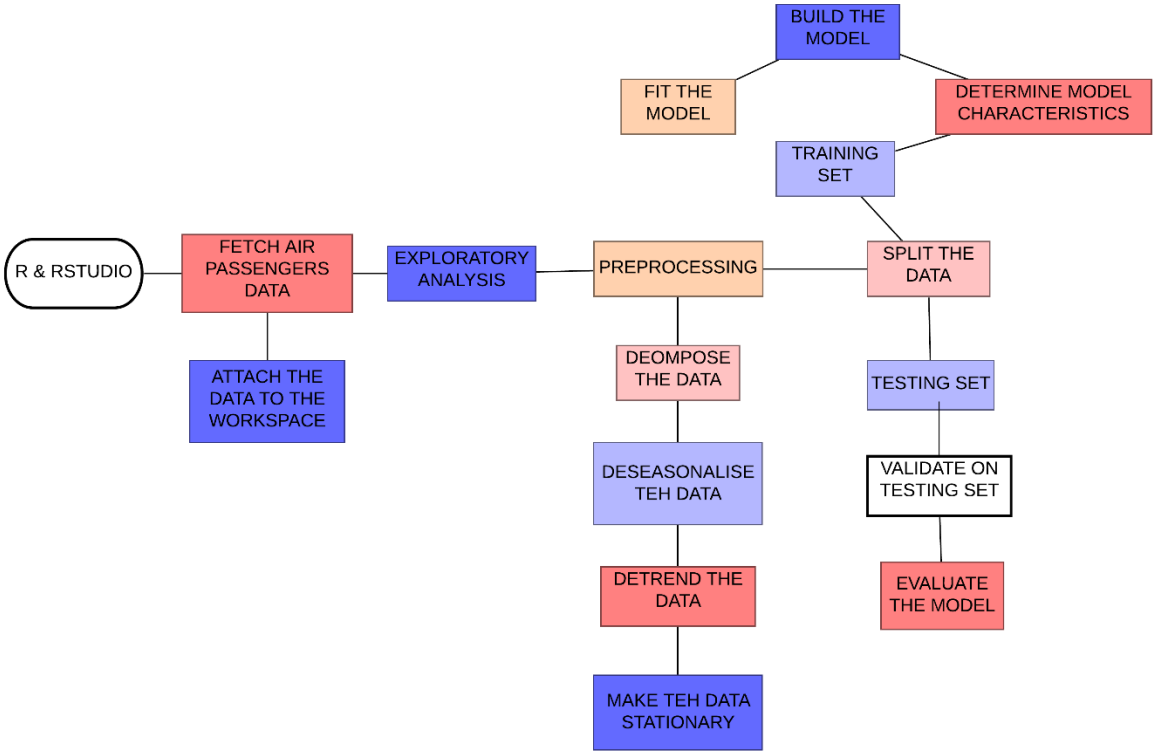

Figure 02 methodology of Air passenger forecasting

3.1.3 Project methodology of International trade forecasting:

The study employees the use of traditional methods and advanced machine learning methods to forecast the value. The timeseries data is fed into Simple exponential smoothing, Auto regression models, Manual Auto Regressive Integrated Moving Averages (Manual-ARIMA) model, Automatic Auto Regressive Integrated Moving Averages (Auto-ARIMA) model, Artificial Neural Networks, Multilayer perception Neural Networks, TBATS models separately and the forecast accuracy measures for each model is taken and compared with all other models to select the best model among them. The data is fed into each model taking into the considerations of assumptions and constraints of each model as mentioned in figure 03 and required pre-processing is done before feeding the data to the model.

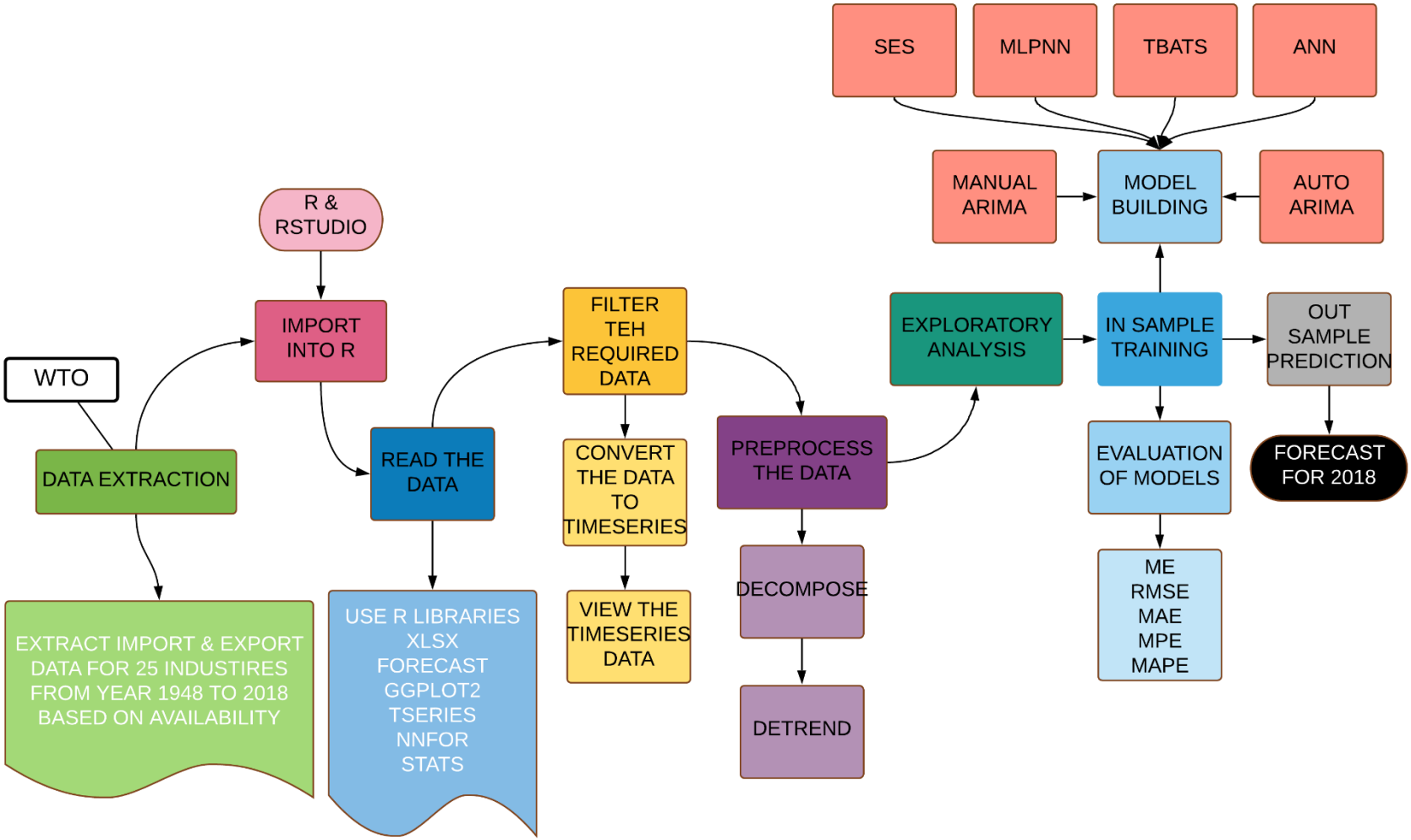

Figure 03: Methodology of International trade forecast.

3.2 DATA COLLECTION

3.2.1 Data collection of visualisation of web analytics:

The data used for the visualisation is a secondary data taken from google open source datasets and the dataset contains many variables about the website to analyse the performance of an website. The data set is an exported data from google analytics tool. Some of the variables used for the purpose of visualisation are

1. Average time on page
2. Page views
3. Bounce rate
4. Organic search
5. Users
6. Device category
7. Channel grouping
8. Date
9. Page
10. User type

These variables are used to create a report about visitors to the webpage.

3.2.2 Data collection of Air passengers forecasting:

The data used for modelling the ARIMA forecast model is available in R software. The data about monthly totals of international airline passengers between 1949 to 1960 is taken to build the model.

3.2.3 Data collection of International trade forecasting:

The time series data required for the analysis is scraped from the world Trade Organisation (WTO) data base from the year 1948 to 2018. Data about the import and export of merchandise goods of India with different countries, regions and trading blocs are taken. The data for around 25 commonly traded commodities namely

1. Automotive products
2. Chemicals
3. Clothing
4. Electronic data processing and office equipment
5. Fish
6. Food
7. Fuels

8. Integrated circuits and electronic components
9. Iron and steel
10. Machinery and transport equipment
11. Miscellaneous manufactures
12. Non-ferrous metals
13. Office and telecom equipment
14. Ores and other minerals
15. Other chemicals
16. Other semi manufactures
17. Other manufactures
18. Personal and household goods
19. Pharmaceuticals
20. Raw materials
21. Scientific and controlling instruments
22. Telecommunications equipment
23. Textiles
24. Total merchandise
25. Transport equipment

Are collected from the WTO database and are categorised to three commonly traded group of items namely

1. Agricultural Products
2. Fuel and mining product
3. Manufactures

Putting this all data together, the dataset contains about 14861 data points.

3.3 STATISTICAL TOOL USED

Statistical tool used for visualisation of web analytics:

Google Data Studio is a dashboard and reporting tool that is customizable and provides feature for sharing the reports and dashboards. With the Google data studio, we can transform our data into an appealing and informative reports for the audience. It is a great tool for monitoring KPIs, which support business strategies and produce periodic reports. Data Studio beta is currently offered free of charge to Google account users and customers of the Google Cloud Platform. Data studio is like Google Analytics dashboard on steroids where there is limitation of 12 widgets per dashboard. But Google data studio has an advantage over this. There is no limitation of 12 widgets and have the ability to add as many report pages. We can connect reports with multiple Google analytics accounts and views. Google data studio does

not provide a wide range of capabilities and features compared to other BI tools like Tableau and Power BI.

- Data Studio has built-in connections to several data sources (listed below) eliminating the need to schedule periodic data refreshes for the reports.
- Another benefit of a live data connection is the ability to toggle between date ranges.
- We can insert dynamic controls so viewers can filter through the content with dimension and date range selectors.
- Also we can include and exclude content from a filter if there's ever a situation where we don't want all the content on the page to change.
- One of the best features of Data Studio is how easy it is to share and collaborate the reports with others. We can share our reports with anyone and even work on the same report at the same time.
- Calculated metrics helps to create custom metrics using formulas. Formula types range from simple to sophisticated, covering a broad range of logical, mathematical, and other typical functions.

Statistical tool used for Air passenger & International trade forecasting:

Statistical tool used for the project is R language. R language is an environment for statistical computing and graphics [20] and it also supports machine learning functions. The advantages of R language are

- It is a free source software and is available on the internet.
- Statistical data analysis and visualizations are possible in R.
- There are a lot of packages available for analysis, and multiple ways of doing them.
- Comprehensive technical documentation and tutorials is supported by R.
- R coding is simple and easy to learn and write.
- Objects of unlimited size and complexity can be handled more easily.
- It support various data formats such as MS Excel, text files, CSV files and it is easy for importing existing datasets, and the results computed in R can be exported.
- R platform supports in writing our own functions, and new packages can be written on inventing some new analysis.

So, it is easy with R language to model complex mathematical models for forecasting foreign trade.

3.4 EVALUATION CRITERIA

Evaluation criteria for Air Passenger and International trade forecasting

With all the models been built on the data it is very important to evaluate the accuracy of each model. To evaluate the forecast accuracy of each model the below mentioned metrics are used

1. **ME – Mean Error**

   Mean error is the mean value of the difference between the actual value and the forecasted value.

   It is calculated using the formula

   ME = Mean (Actual – Forecast)

2. **RMSE – Root Mean Square Error**

   Root mean square error is the square root of squared values of mean error. It is standard deviation of residuals of prediction values. It denotes how the residuals are spread out and how concentrated the data is around the best fit line.

   It is calculated using the formula,

$$RMSErrors = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

3. **MAE – Mean Absolute Error**

   It is the mean of absolute values of difference between the actual value and the forecasted value.

   It is calculated using the formula,

   MAE = Average (ABS (Actual – Forecast))

4. **MPE – Mean Percentage Error**

   It is average of percentage error in each forecast ie it is the average value of percentage deviation between actual and forecasted values.

   It is calculated using the formula

$$\mathrm{MPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{a_t - f_t}{a_t}$$

5. **MAPE – Mean Absolute Percentage Error**

   It is also known as mean absolute percentage deviation, is the average value of percentage of absolute error in the forecast. It is used as loss function in regression techniques, so, it's important in regression based techniques.

   It is calculated using the formula,

$$\mathrm{M} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|,$$

CHAPTER 04

4 ANALYSIS

Before starting analysis of the project there are certain assumptions for each model. Based on the assumptions each model is built and why the model has been used and approach used to model the data is listed below in the table. With understanding of each model it becomes easy for model building and validating.

MODEL	ASSUMPTIONS	WHY THE MODEL	APPROACH USED
SES	1. Time series is locally stationery 2. Non zero constant linear trend can be added to the model	1. Problem is short term forecasting 2. Uses weighted moving averages while exponentially decreasing the weights	1. The model is built on the time series data 2. Model is used to forecast for next year 3. Model is validated across evaluation metrics
ARIMA	1. Timeseries should be stationary 2. Transformation is required in the case of non-stationary data 3. There are no level shifts 4. The error process is homoscedastic (constant) over time 5. The model parameters are constant over time	1. The model is flexible 2. Do not over fits the data 3. Predicts the out of sample with much better accuracy.	1. The data is made stationary 2. ACF and PACF are determined for the data 3. Model parameters p, d, q are determined 4. Model is built based on p, d, q 5. model is evaluated across metrics 6. Forecast is done.

ANN	Artificial neural networks has no assumptions	1. Model has an advantage over other techniques since there is no assumptions to it. 2. Model predicts well for the out of sample data.	1. The data is passed on to neural networks 2. ANN model is built on the data 3. Model is validated across the evaluation metrics 4. Out of sample forecast is done.
MLPNN	No assumptions for neural networks	1. Model uses MSE loss function to optimise the forecast 2. It predicts well for medium term forecast	1. Model Is built on the timeseries data 2. Model is optimised for MSE loss function 3. Models is tuned with lags. 4. Model is evaluated with all metrics 5. Out of sample forecast is done.
TBATS	1. Horizon of the analysis is determined before building the model.	1. Model incorporates multiple seasonality 2. Model uses box-cox transformation.	1. Model Is built on the timeseries data 2. Model is optimises with ARMA errors. 3. model is evaluated across all evaluation metrics. 4. Out of sample forecast is done.

Table 01: Model building considerations.

4.1 ANALYSIS OF VISUALISATION OF WEB ANALYTICS:

To analyse the data and create a reporting dash board following steps are performed sequentially

1. Click Create -> Report to create a new report
2. Then select the data source from the already existing data or create a new data source and click the data source -> Add to report.
3. Now Data source is added to the report and DS displays Layout and Theme dialog box.
4. In layout section of the dialog box, one can change the layout options like Canvas size, Grid setting etc.
5. In Theme section of the dialog box, one can set background colour, font styles etc.
6. Now, click on Insert and select Scorecard. In data pane select the metric to be Users. Select Compare data range -> previous period.
7. Similarly add scorecard for Pageviews, Bounce rate, Average time on page, organic search.
8. Now, click on Add chart in the tool bar, select Pie chart. Change the dimension to Default channel Grouping and metric to be Sessions. Under style select No. of slices to be 6 for better view of chart.
9. Now, move the slider under Manage dimension value colours to make the pie chart into doughnut chart.
10. Select Add chart -> bar chart for viewing device category vs page views.
11. Now, Add a time series chart to see the growth in users. Select compare with data range select previous period to add the previous period line too. (Add a trend line to it if needed.)
12. Now for better understanding insert a table with dimension to Page and Metrics Page views, Unique users and Bounce rate.
13. In style option, Under metrics section change Column 2 type to Bar and column 3 type to heat map.
14. Now add a text box to insert header. And add a date range to it.
15. You can also add filter if needed.

4.2 ANALYSIS OF AIR PASSENGERS FORECASTING:

Data overview

The data is the time series data collected monthly over a period of 1949 to 1960. The snapshot is shown below

```
> AirPassengers
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
```

Exploratory analysis

The below shown is the description and source of the dataset

Monthly Airline Passenger Numbers 1949-1960

Description

The classic Box & Jenkins airline data. Monthly totals of international airline passengers, 1949 to 1960.

Usage

```
AirPassengers
```

Format

A monthly time series, in thousands.

Source

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) *Time Series Analysis, Forecasting and Control*. Third Edition. Holden-Day. Series G.

Dataset Description

Number of rows = 12

Number of columns = 13

Plot of the data over the range of 1949 to 1960

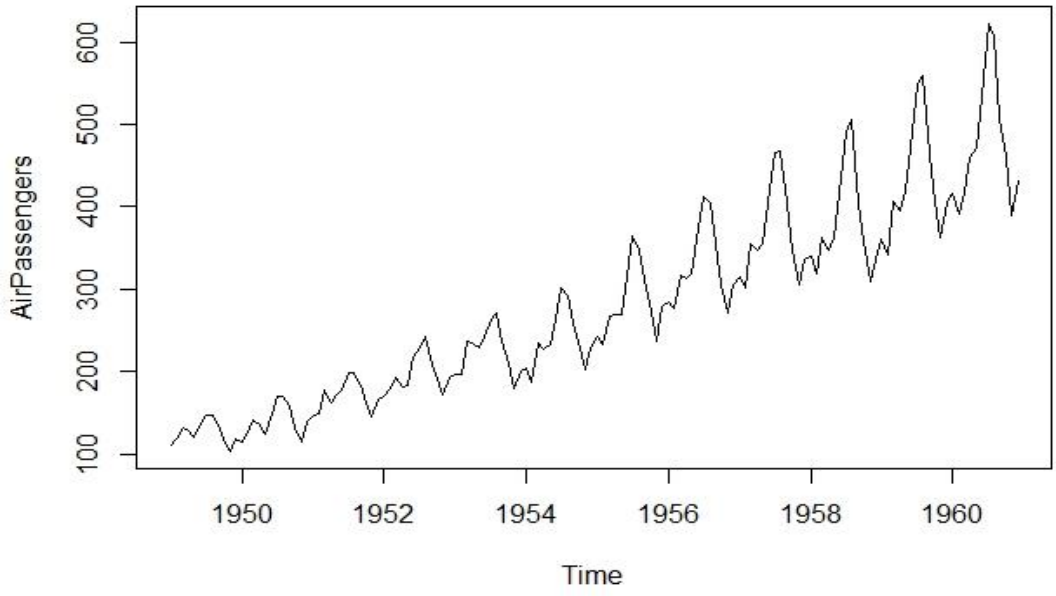

Figure 04: Timeseries plot of the data.

Fitting a regression line in the above graph shows the trend in the dataset. From the below image there is the upward trend in the data, which means the number of passenger travelling in the airlines growing over the period of the time.

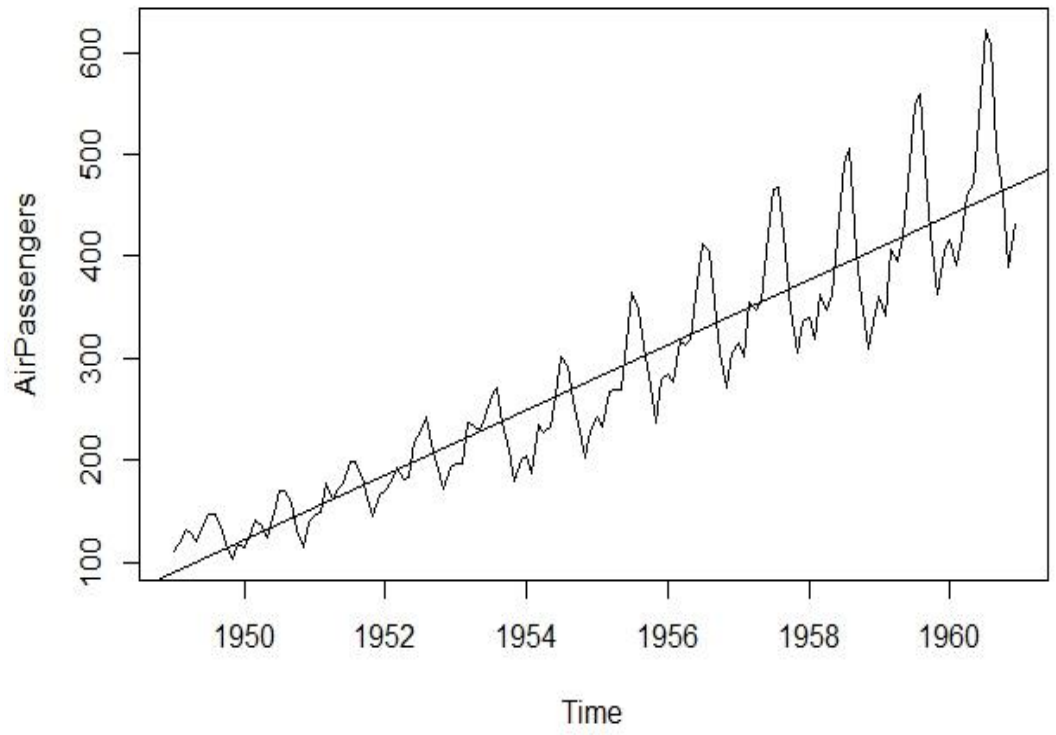

Figure 05: Fitting the regression line in the timeseries plot

Also from the graph we can observe there lies a seasonal pattern in the data. To dig more on the seasonal pattern present in the data a graph plotting the entire data month wise.

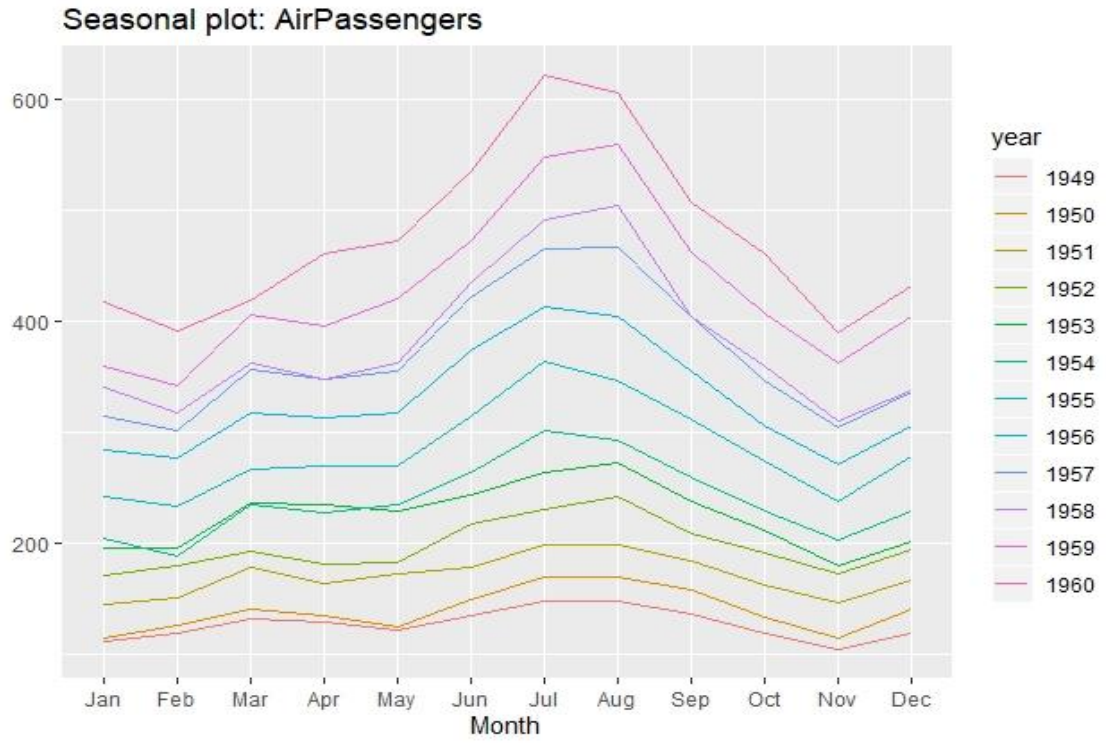

Figure 06: Seasonal plot to check seasonality

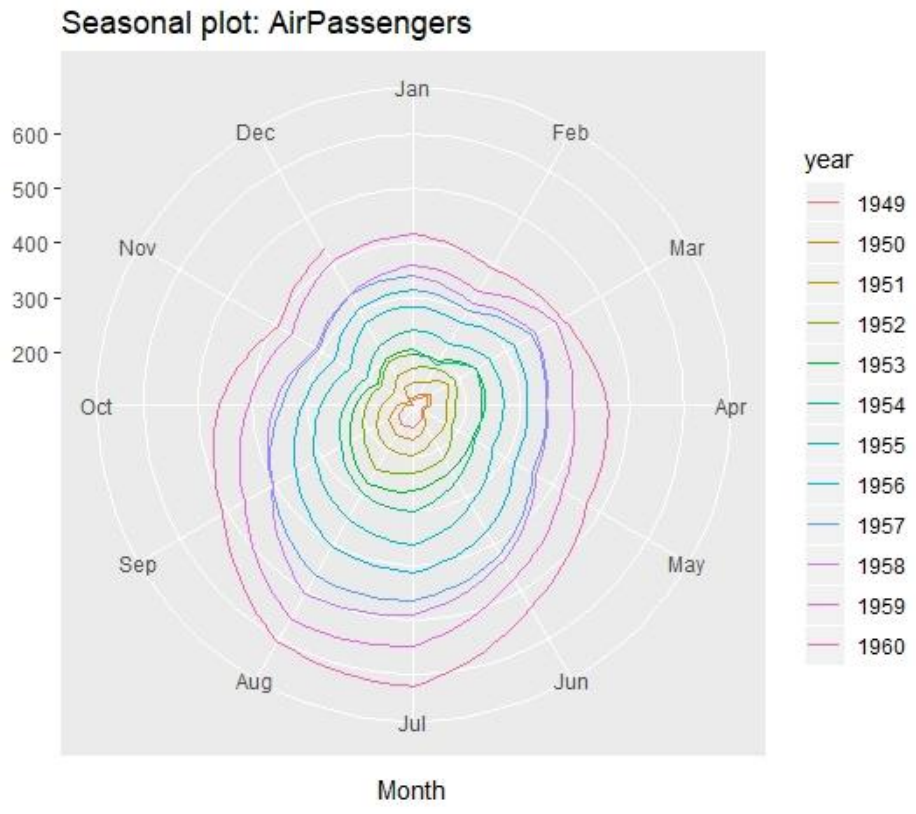

Figure 07: Radial seasonal plot

From the above graphs it can be inferred that passengers travel more in the month of July and august and lesser in the month of November and February.

Month plot

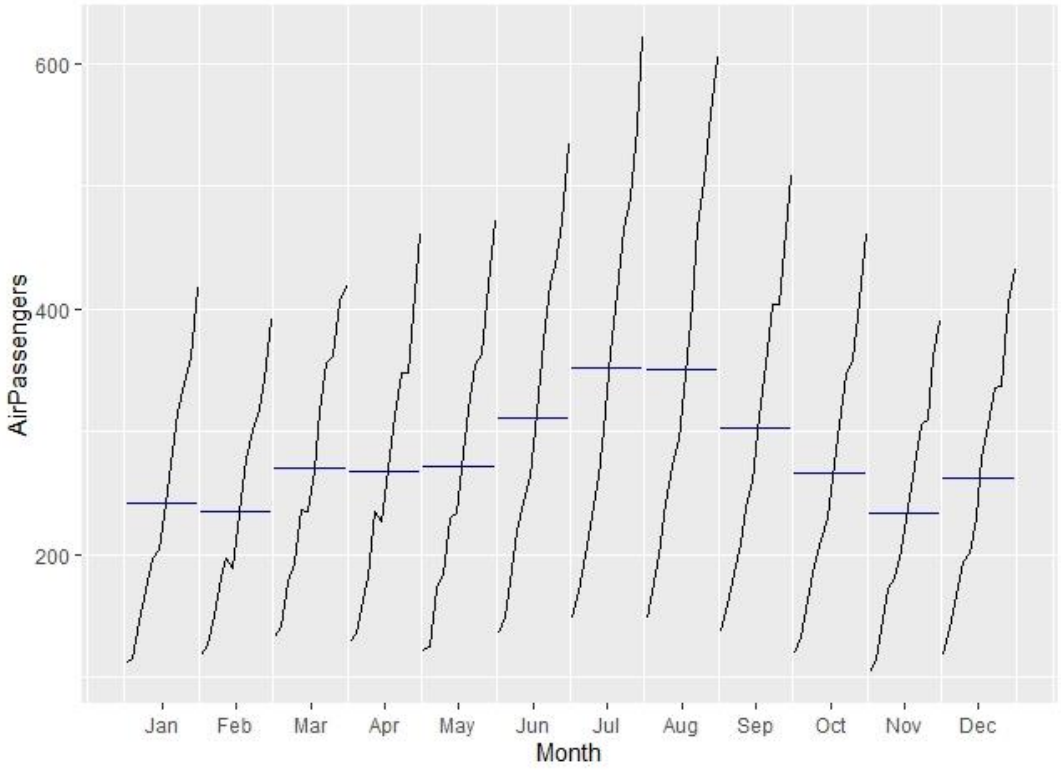

Figure 08: Month plot of the timeseries to check passenger travel

The purpose of month plot is to identify changes across season. From the above image it can be inferred that July witnessed highest passengers and November and February having lowest passenger travel.

Boxplot

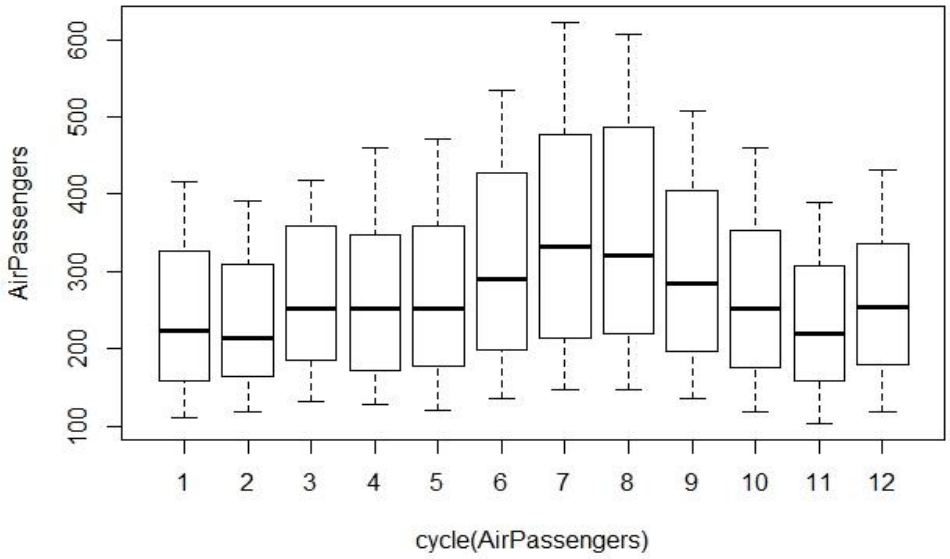

Figure 09: Box plot of the timeseries data

Box plot shows July and August have highest passenger travel and November and February sees least passenger travel.

Lag plot

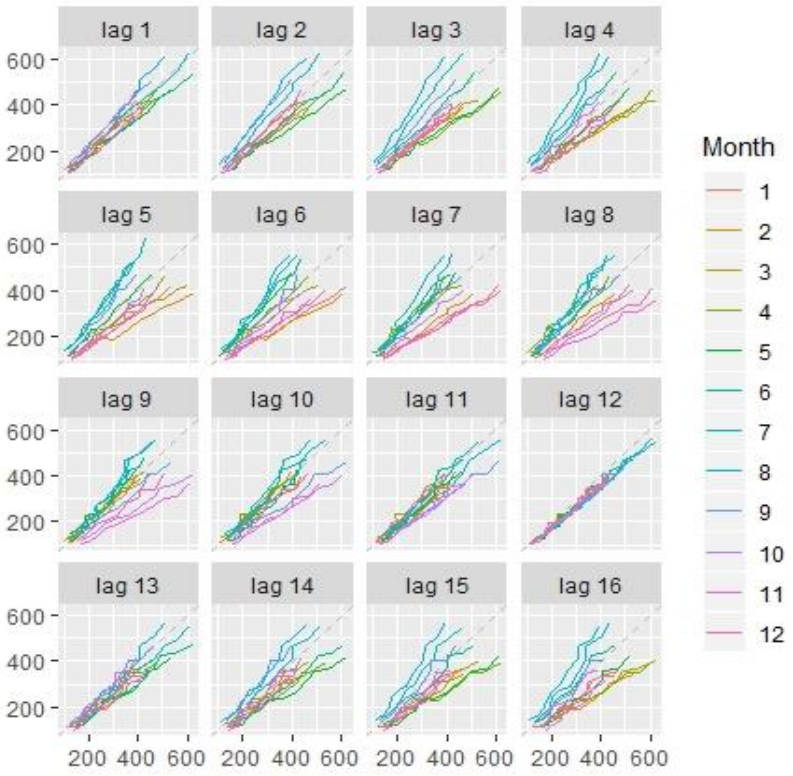

Figure 10: Lag plot of timeseries data

Lag plots are plotted between the current period and the previous period, for example lag 1 plot the plot is made between January and December and goes on for corresponding lag plots.

TS display, ACF and PCF plots

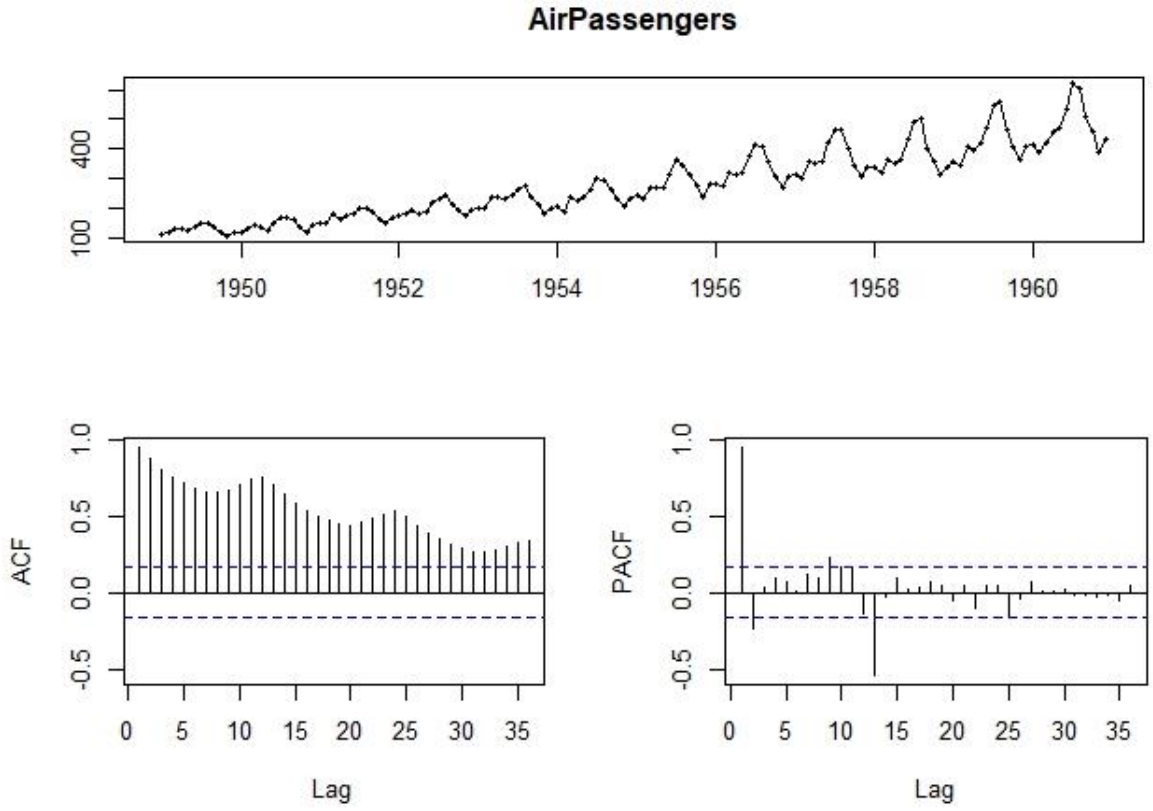

Figure 11: TS plot of the timeseries data

Splitting the dataset

The dataset is split into training and testing dataset. The split ratio is 5:1 for train and test.

Treating the data

As we have seen from the exploratory analysis the data has following components which are need to be treated for better forecast

1. Seasonality
2. Trend
3. Cyclicity
4. Randomness

Decomposing Timeseries

Before treating the data the data has to be decomposed to view the individual components trend seasonality, Cyclicity and Randomness. The code used for decomposing the timeseries is

```
> Air_decompose1 <- stl(AirPassengers, s.window = 9)
> plot(Air_decompose1)
```

The decomposed plot looks like the image shown below

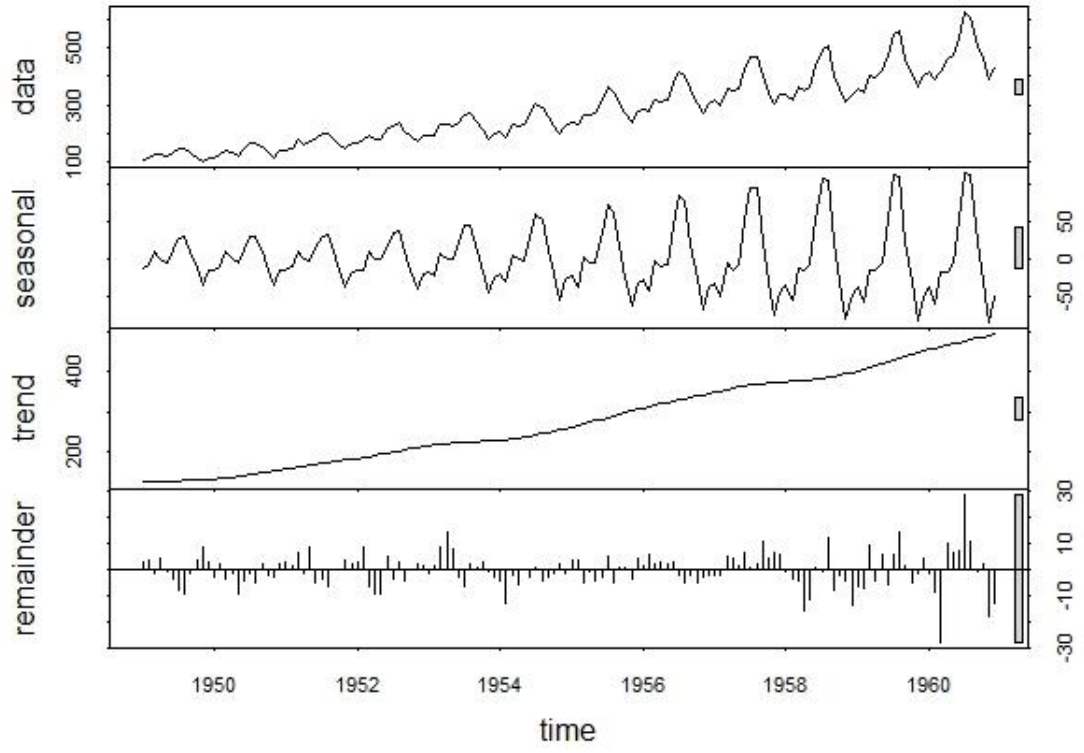

Figure 12: Decomposition of timeseries

De-seasonalising the data

The data is de-seasonalised with the code

```
> deseasoned_air <- seasadj(Air_decompose2)
> plot(deseasoned_air)
```

After removing seasonality from the data the graph looks like the image shown below

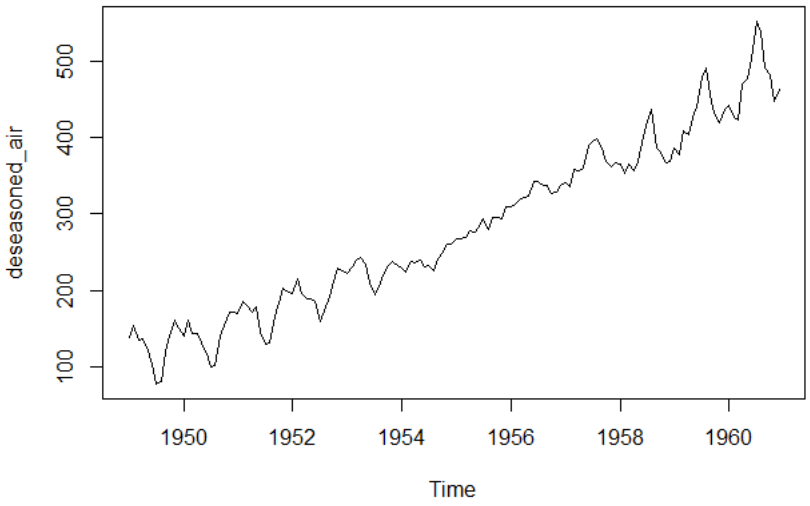

Figure 13: De-seasoned plot of timeseries

Dickey fuller test

The data is then tested for dickey fuller test to check whether the data is stationary or not. The data is passed to augmented dickey fuller test and the results are shown below

```
> adf.test(diff(log(AirPassengers)),alternative = "stationary", k = 0)
ng

        Augmented Dickey-Fuller Test

data:  diff(log(AirPassengers))
Dickey-Fuller = -9.6003, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Thus the data is of stationary nature and timeseries model can be built on it.

ACF plot (decompose)

Auto Correlation Function provides correlation of data with the lagged values. It represents how well the present value in the series represents the past values.

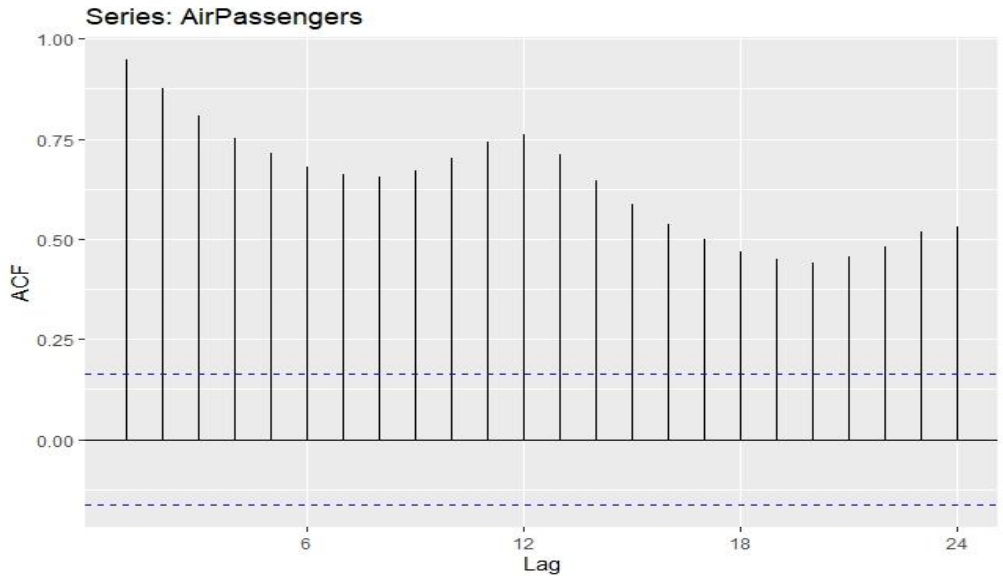

Figure 14: ACF plot of timeseries data

PACF plot (decompose)

Partial Auto correlation Function finds the correlation unlike the ACF by finding Correlation of residuals with the succeeding lag value. So if there is any hidden information in the residual we may get that in the next lag value. The PACF plot is shown below

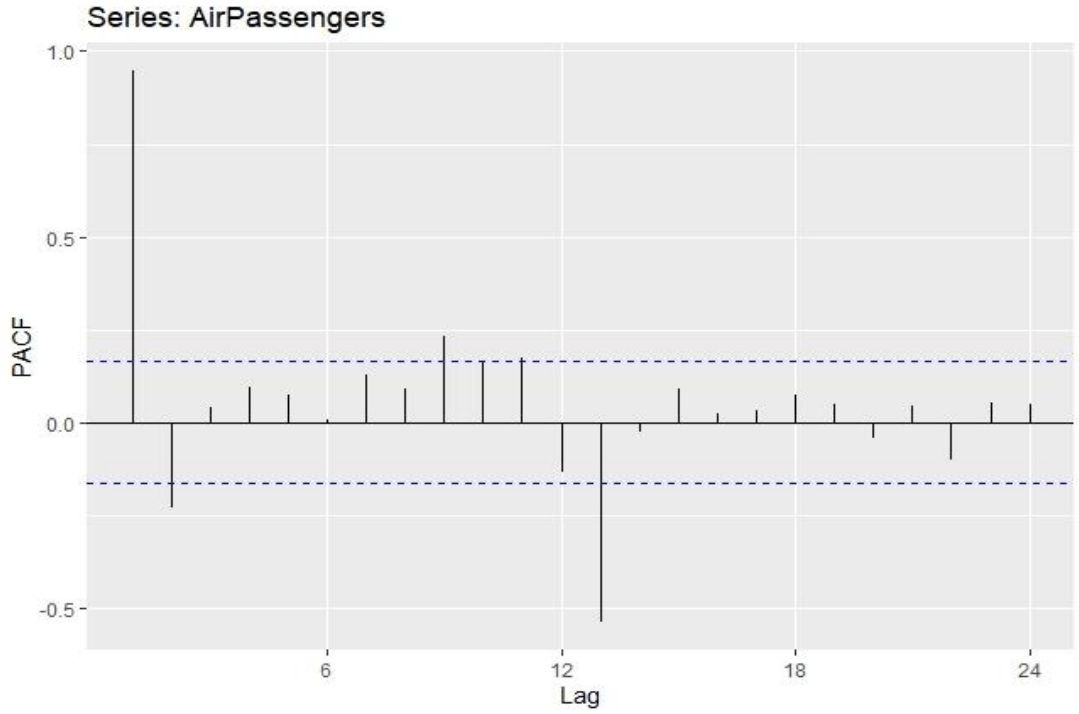

Figure 15: PACF plot of timeseries data

Model building

After treating the dataset the ARIMA model is built with the following code

```
Airfit = arima(log(AirPassengers), c(0, 1, 1),seasonal = list(order = c(0, 1, 1), period = 12))

Airfit

tsdisplay(residuals(Airfit), lag.max = 15, main = "Model Residuals")
```

ARIMA – Auto Regressive Integrated Moving Average model forecasts the timeseries based on the past lag values and lagged forecast errors.

An ARIMA model is characterized by three terms namely p,q,d

P = order of AR term

Q = order of MA term

D = Differencing term which is required to make the series stationary

Here values of (p,q,d) are (0,1,1) as found in previous analysis.

Forecasting the model

The model built is used to forecast the next year passenger volume with the code shown below we get the forecasted values for the year 1961. When forecasting is done it is most important to evalute the forecast model.

```
> fcast_Air = forecast(Airfit, h=12)
> fcast_Air
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jan 1961       6.110186 6.063133 6.157239 6.038224 6.182147
Feb 1961       6.053775 5.998947 6.108604 5.969922 6.137628
Mar 1961       6.171715 6.110084 6.233346 6.077459 6.265971
Apr 1961       6.199300 6.131547 6.267054 6.095681 6.302920
May 1961       6.232556 6.159189 6.305923 6.120351 6.344761
Jun 1961       6.368779 6.290198 6.447359 6.248600 6.488957
Jul 1961       6.507294 6.423825 6.590763 6.379639 6.634949
Aug 1961       6.502906 6.414820 6.590993 6.368190 6.637623
Sep 1961       6.324698 6.232224 6.417172 6.183271 6.466125
Oct 1961       6.209008 6.112346 6.305670 6.061176 6.356840
Nov 1961       6.063487 5.962811 6.164164 5.909516 6.217459
Dec 1961       6.168025 6.063488 6.272562 6.008149 6.327900
```

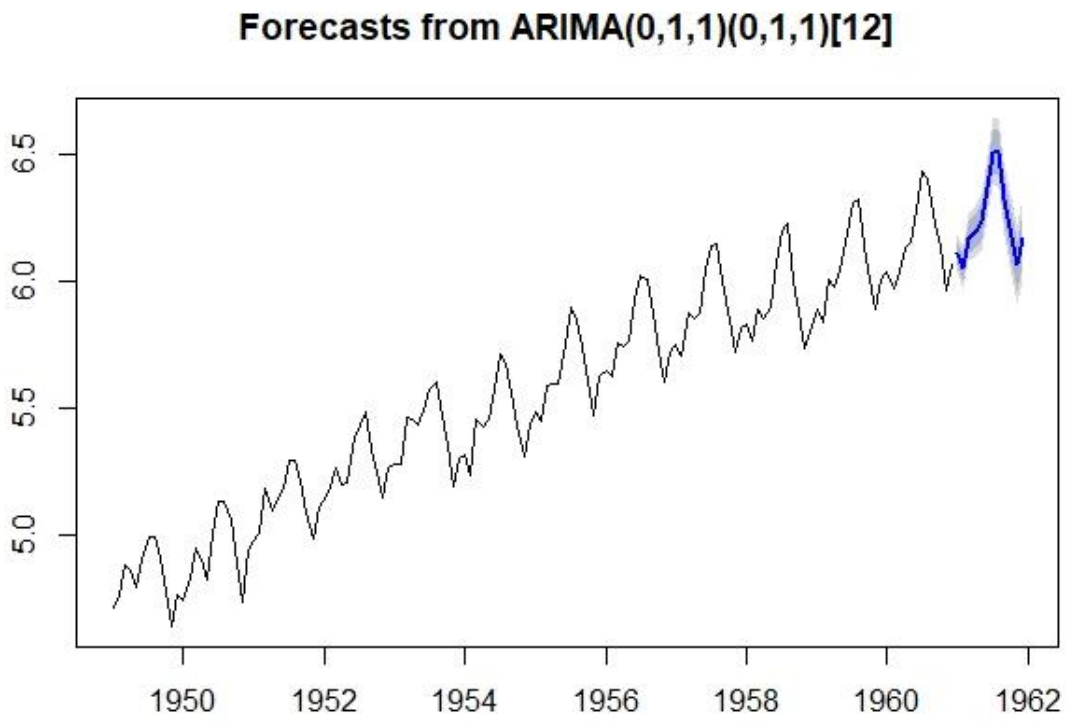

Figure 16: Forecast of air passengers

4.3 ANALYSIS OF INTERNATIONAL TRADE FORECASTING:

The purpose of analysis is to build various machine learning models and compare them across common evaluation criteria and select one best model which can predict the value accurately among the models. For building model total export and import value of merchandise goods traded with the world as a whole from the year 1948 and 2018 are only taken and based on the results from the analysis the best model can be applied to forecast for other categories and to individual countries, blocs, regions.

DATA OVERVIEW

The dataset contains import and export value of merchandise goods traded between the year 1948 and 2018 is extracted and stored in the name of ts_data for the purpose of processing. The following commands has been used to understand the nature of dataset

str(ts_data)

start(ts_data)

end(ts_data)

frequency(ts_data)

summary(ts_data)

EXPLORATORY DATA ANALYSIS

To explore the data various graphs are drawn and various commands are used. The data contains 142 data points with 71 data for exports and 71 data for imports. These data is plotted against year and value as shown in Figure 17.a and 17.b

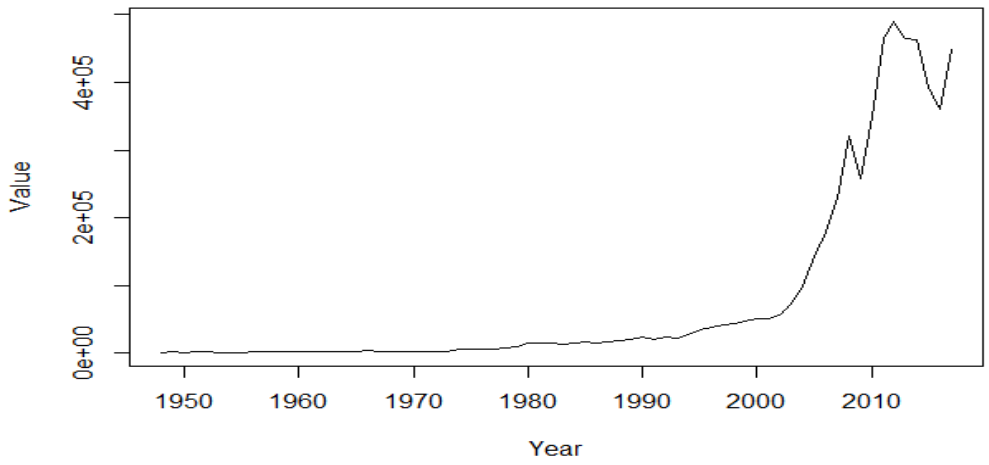

Figure 17.a: Total Imports of India between 1948 and 2018

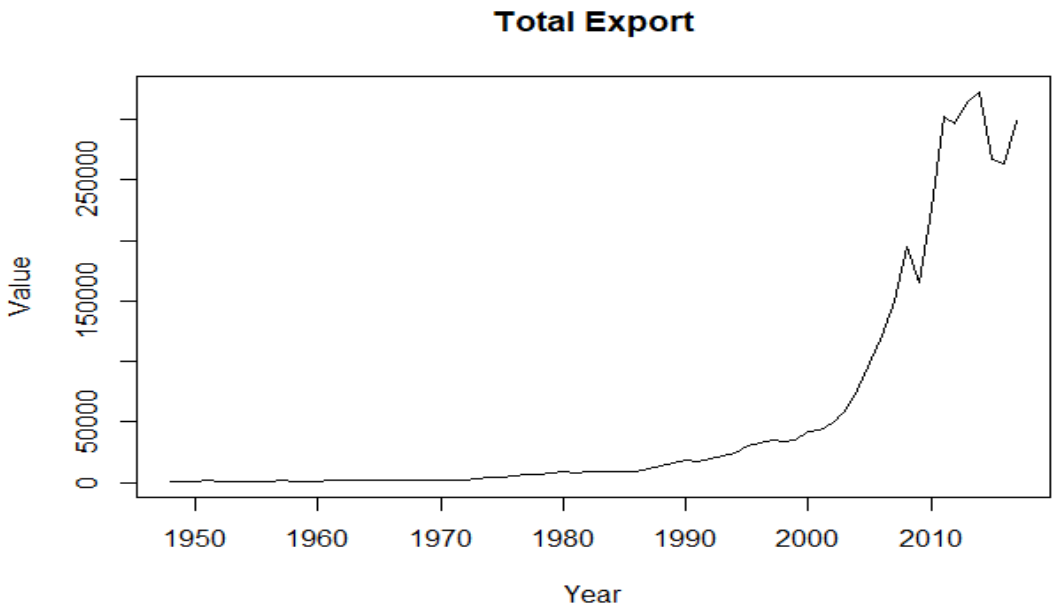

Figure17.b: Total Exports from India between year 1948 and 2018.

From the above figures it is evident that India has grown tremendously in international trade between the time periods. In early 21st century the growth in Total traded value of goods and services from India has grown exponentially. To observe the trend in the data a trend line is drawn as shown in the figure 18.a and 18.b below,

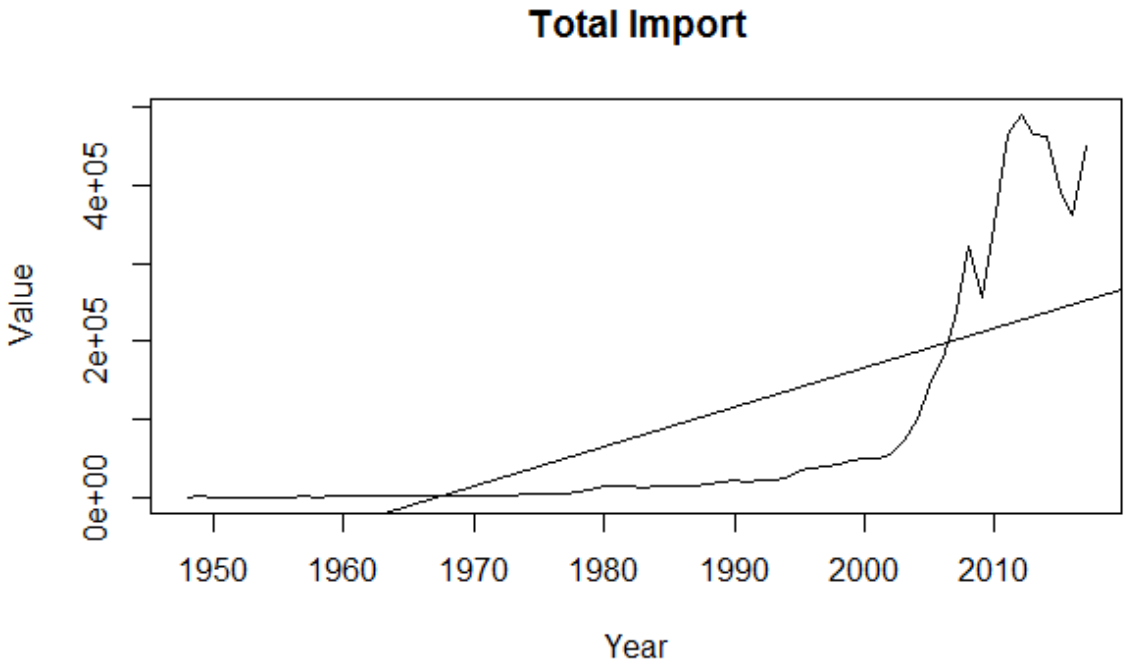

Figure 18.a: Trend in Imports of goods

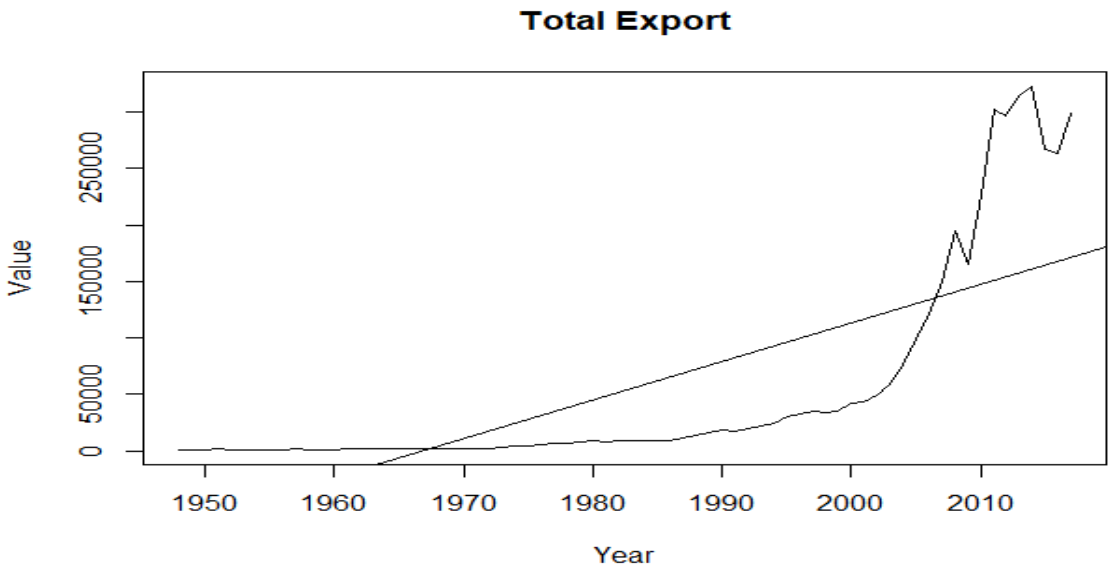

Figure18.b: Trend in Export of goods

From both the figure and it is evident that traded value of goods from India has increased during the time period.

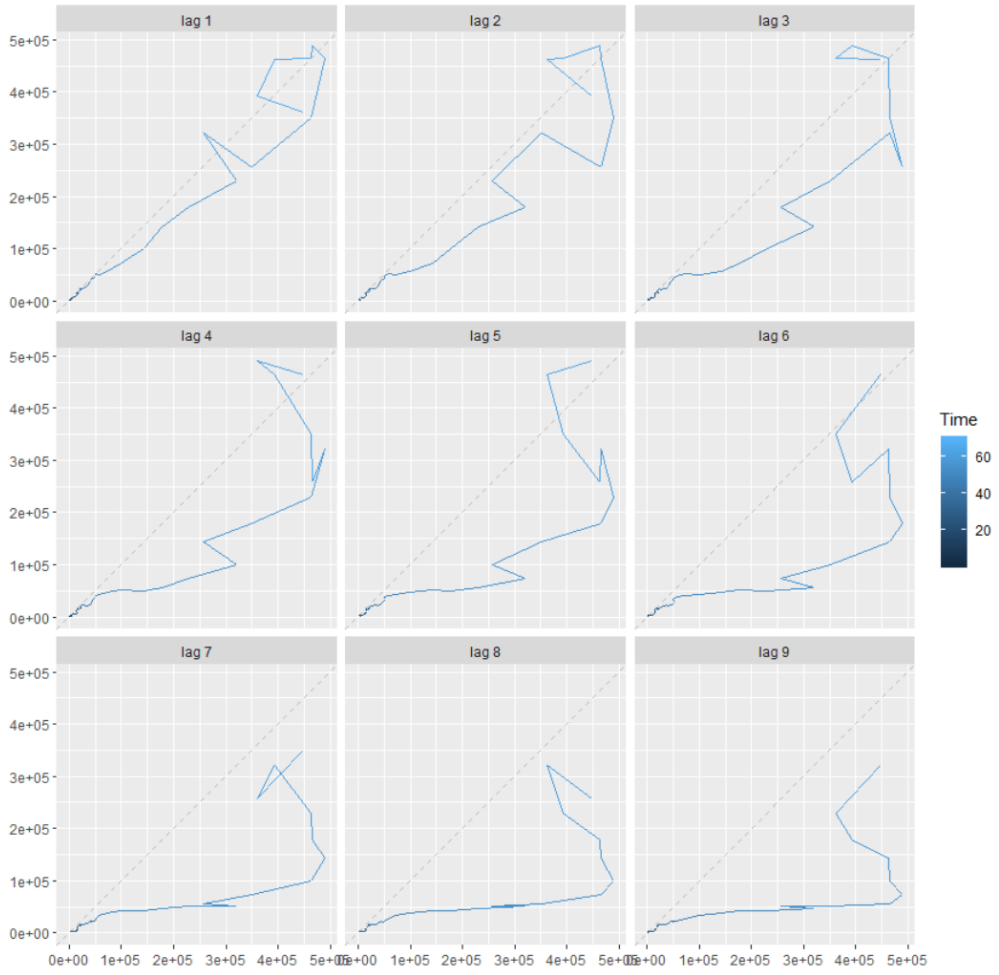

Figure 19.a: Lag plot of Import data

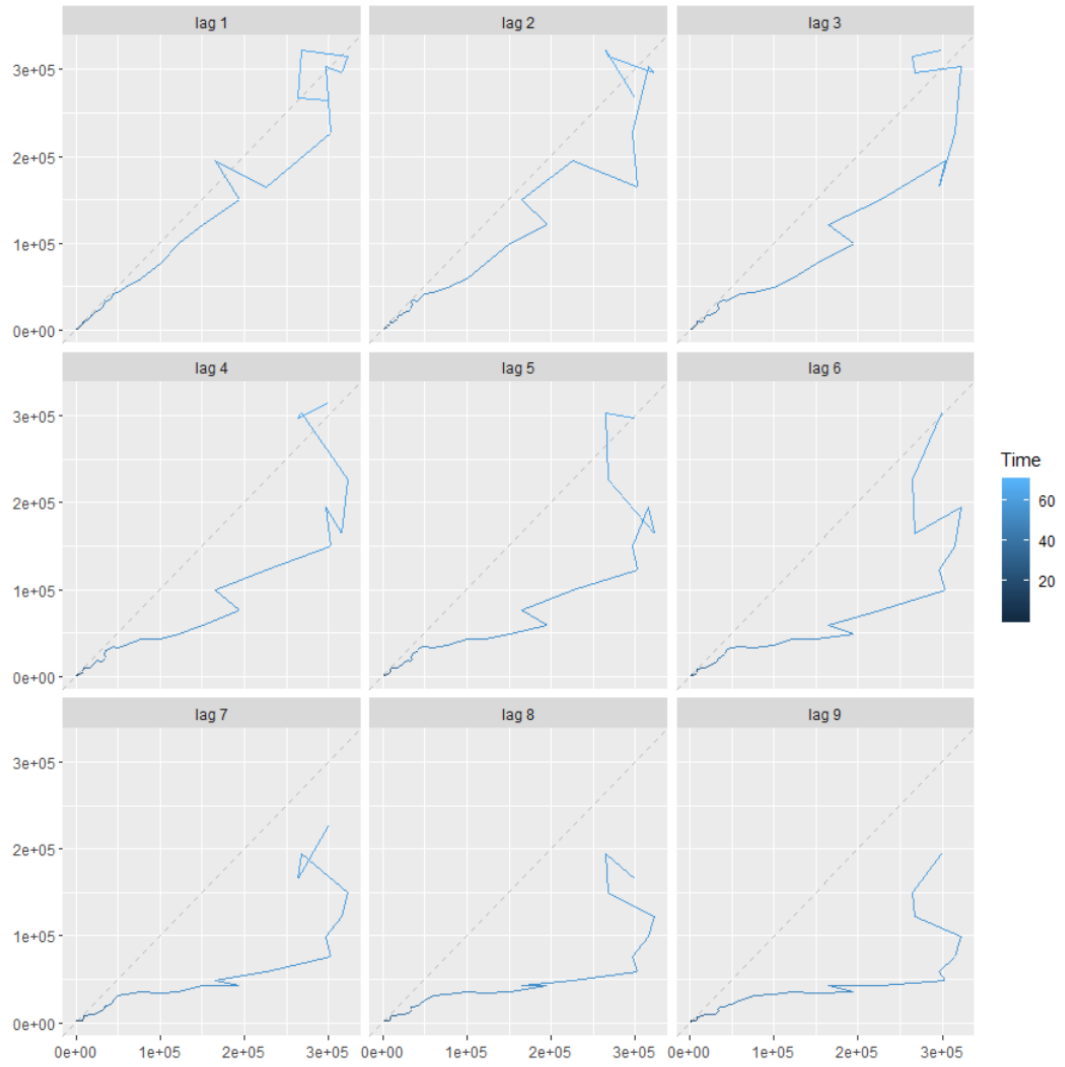

Figure 19.b: Lag plot of Export data

Now lags are measured in the data with the help of lag plot which explores the time series data to understand its nature. The lag plot is a scatter plot drawn between actual value of current observation and lagged value of the same observation i.e. current value vs lag value of current observation. The lag plot shown in the figure 19.a & 19.b is obtained using gglagplot(ts_data). It can be seen in figure that each lag plot looks different which conveys that current data only depends upon the data of the previous month only. That's why each lag plot is different from each other and doesn't exhibits any trend in the graphs. It also denotes that the data lacks seasonality and cyclicity as it doesn't repeat for any lag.

Since the data is yearly data and use case is International trade other components of forecast data such seasonality, cyclicity did not exist. This may be also due to higher degree of causation of external variables and other environmental variables to the trade.

Box plots are drawn for both the export data and import data to check for outliers. The figure 20.a & 20.b show the box plot for import and export respectively. From the figures it is observed that the data has many outliers, but these outliers cannot be treated since the data is of time series nature which may lose it accuracy in forecasting if treated.

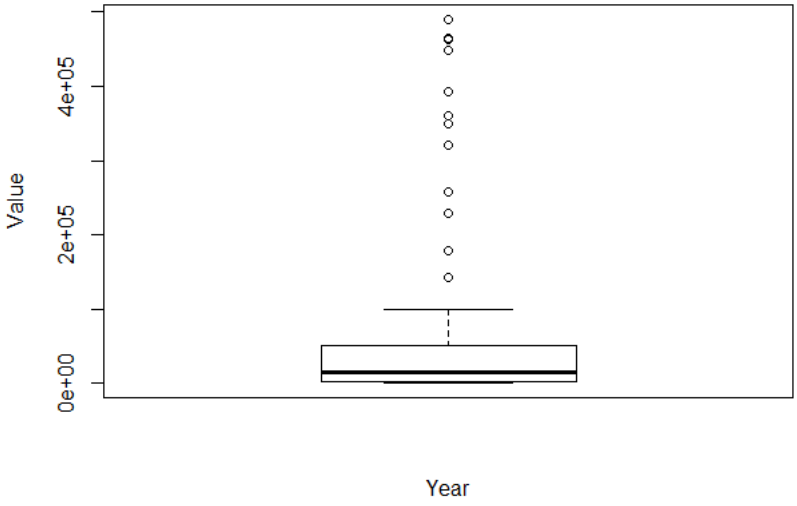

Figure 20.a: Box plot of Total Imports

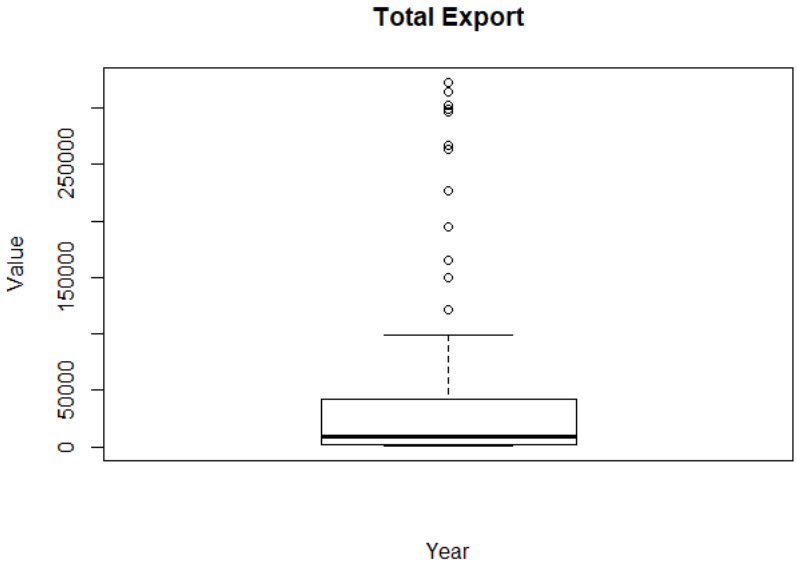

Figure 20.b: Box plot of Total Exports

DATA PREPROCESSING

The data is pre-processed before creating model. First step in pre-processing the data is data decomposition. Since there is no seasonality and cyclicity in the data it is enough to detrend the data. Here the data is detrended by technique called differencing. In detrending by differencing technique the series is detrended by constructing a new time series by calculating the difference between the original observation and observation at the previous step. Here First order differencing is applied where difference of current observation and immediately previous observation is calculated to form the new time series. The Figure 21.a and 21.b shows the detrended data

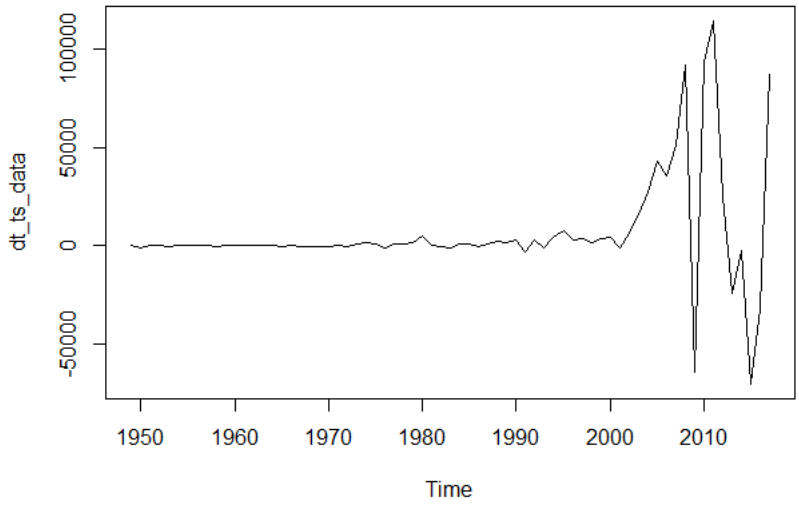

Figure 21.a: Detrended plot of Import data

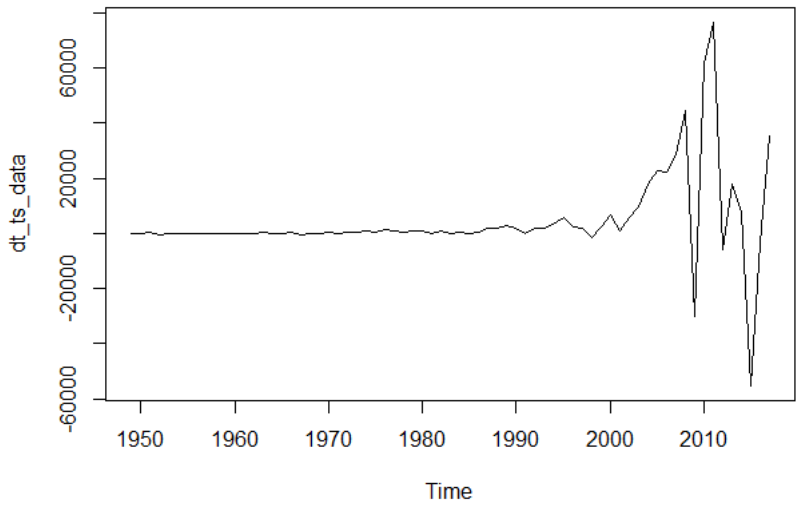

Figure 21.b: Detrended plot of Export data

From the figure above it is evident that for many time series models the data has to be stationary which means the mean and variance of the data over the time period has to be constant. To check for the stationarity of the data Augmented Dickey fuller test is used. Augmented dickey fuller test is a statistical test that checks the stationarity of the series. It has two hypothesis

H0: Series is stationary

H1: Series is not stationary

The Augmented dickey fuller test is run on the detrended data with the help of command adf.test(ts_data). The results of the test is

```
        Augmented Dickey-Fuller Test

data:  dt_ts_data
Dickey-Fuller = -4.0808, Lag order = 4, p-value = 0.01132
alternative hypothesis: stationary
```

For import data and

```
        Augmented Dickey-Fuller Test

data:  dt_ts_data
Dickey-Fuller = -3.5542, Lag order = 4, p-value = 0.04384
alternative hypothesis: stationary
```

For Export data. From both the results it is seen that the p-value < 0.05 thus alternate hypothesis is rejected and the series is observed to be stationary. Now this series is fed to all the models for processing.

MODEL BUILDING

SIMPLE EXPONENTIAL SMOOTHING

Simple exponential smoothing (SES), is a timeseries forecasting method for univariate data and it can support data with trend and seasonality. In this method the prediction of future value is based on weighted sum of past observations and the weights of past observation decays exponentially with decrease in time periods i.e. for the older observations.

Here single exponential smoothing is applied, since the data is of univariate nature, it requires single smoothing parameter "α". It can take values from 0 to 1, higher the value means the model pays attention to most recent values only.

The equation for simple exponential smoothing is

$$\hat{y}_{T+1|T} = \alpha\, y_T + \alpha(1-\alpha)\, y_{T-1} + \alpha(1-\alpha)^2\, y_{T-2} + \cdots$$

Where

$\hat{y}_{T+1|T}$ = Forecast value for time T+1

y_T = Actual value for the time period T

α = Smoothening parameter.

y_{T-1} = Actual value of previous observation T-1.

This model is applied to both the import data and export data.

```
Simple exponential smoothing

Call:
 ses(y = ts_data, h = 1)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 1588.2727

  sigma:  28673.4

     AIC     AICc      BIC
1738.287 1738.651 1745.033
```

```
Simple exponential smoothing

Call:
 ses(y = ts_data, h = 1)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 1300.5592

  sigma:  17000.93

     AIC     AICc      BIC
1665.109 1665.472 1671.854
```

Figure 22: Model for import data (on left) and export data (on right)

From the above built models it is observed that the smoothing parameter α for both the models are nearly 1 (0.9999). So, the value of forecast depends only on the value of previous observation. Thus there are more chances of error. The model is applied to forecast for out of sample value it predicts the value as shown in the figure 23.a for import data and 23.b for export data.

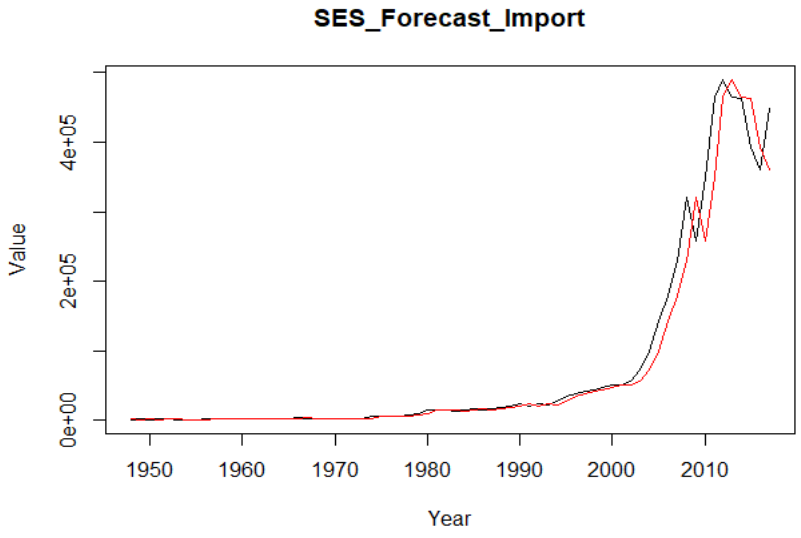

Figure 23.a: SES Forecast vs actual of import data

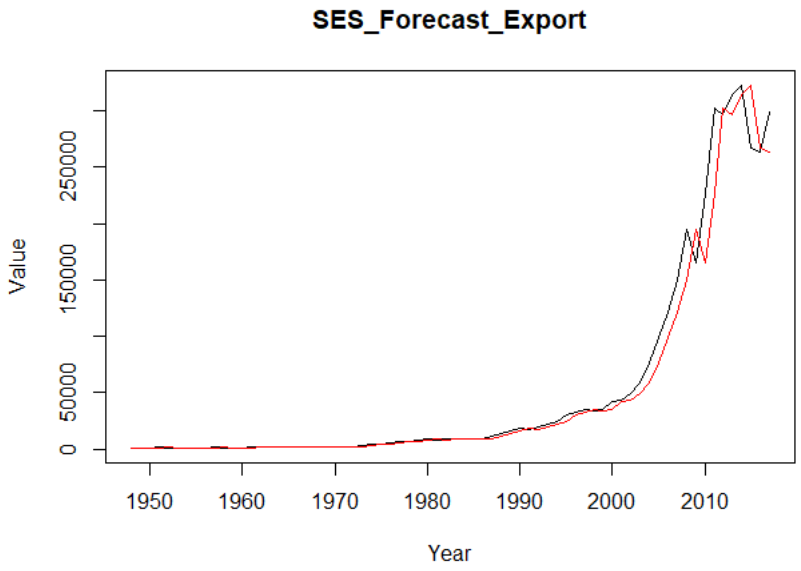

Figure 23.b: SES Forecast vs actual of export data

Observing the above figures it can be understood that the models have higher error. On evaluating the models across the metrics it is found that

ME	RMSE	MAE	MPE	MAPE	MASE
6383.87	28260.81	12191.49	6.18	15.03	0.99

Table 02.a: Error values of Import data.

ME	RMSE	MAE	MPE	MAPE	MASE
4257.16	16756.30	7067.16	6.76	11.01	0.99

Table 02.b: Error values of Export data.

Error across all the chosen metrics is very high with mean error of 6383.87/4245.16, Mean percentage error of 6.18%/6.76% and Mean absolute Percentage Error of 15.03/11.01 for import/export respectively.

AUTO REGRESSIVE MODEL

In Auto regression (AR) model the value is forecasted based on the linear combination of past values and the auto regression means the regression of the variable to be forecasted is regressed with itself. The auto regression model is characterised by the term p which means order of auto regression.

The equation of AR model is

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

Here

Y_t = Forecast value

α = Intercept term

β_1 = Coefficient of lag1

β_2 = Coefficient of lag 2 etc.

ϵ_1 = Variance of the error term

The order of regression is determined from the ACF and PACF plot. ACF – Auto correlation function will give values for auto correlation in a lagged series. Auto correlation of different orders gives the insider information regarding the series. ACF (0) is always one and it ranges from -1 to +1. The ACF value derived from ACF plot determine the order q of the series used as one of the parameter among p, d, q in building an ARIMA model. Following is the code executed for getting the ACF plot. PACF – Partial correlation function will eliminate the value of intermediate lags and it's basically adjusting for the intervene series. PACF (1) = ACF (1). PACF (2) tells us that the value of correlation between original observation and Lag2 series. The PACF value derived from PACF plot determine the order q of the series used as one of the parameter among p, d, q in building an ARIMA model. The number of lines outside the blue line in the PACF plot will determine the values for p.

The ACF and PACF plot of import data and export data are

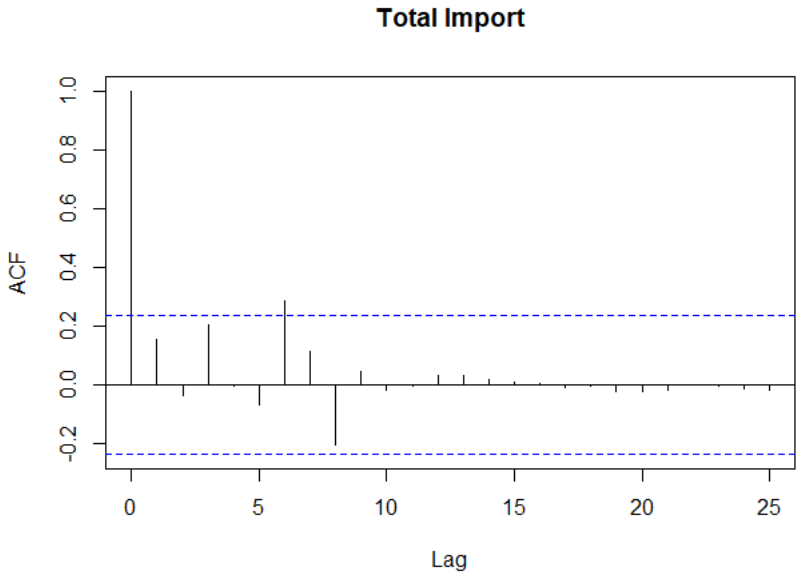

Figure 24.a: ACF plot of Import data

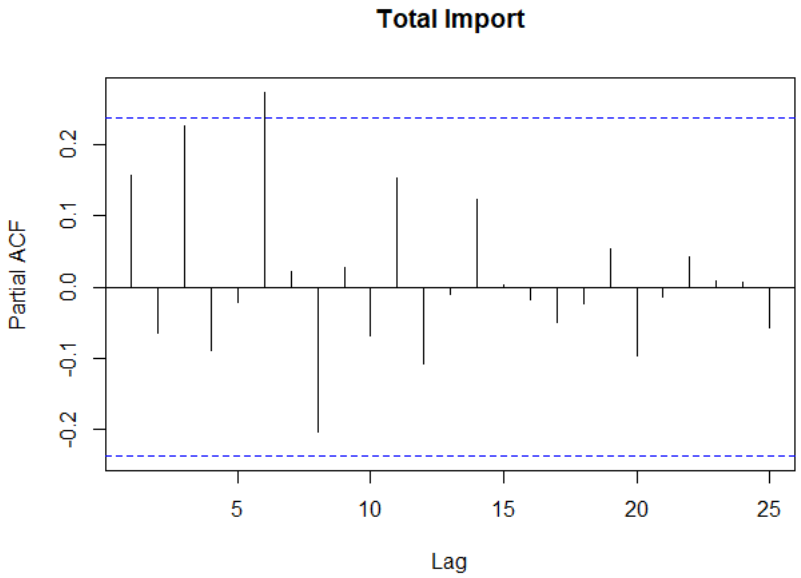

Figure 24.b: PACF plot of Import data

By analysing the above figures (24.a & 24.b) it can be seen that order regression for import data is 1, since only one lag line is outside the significance levels.

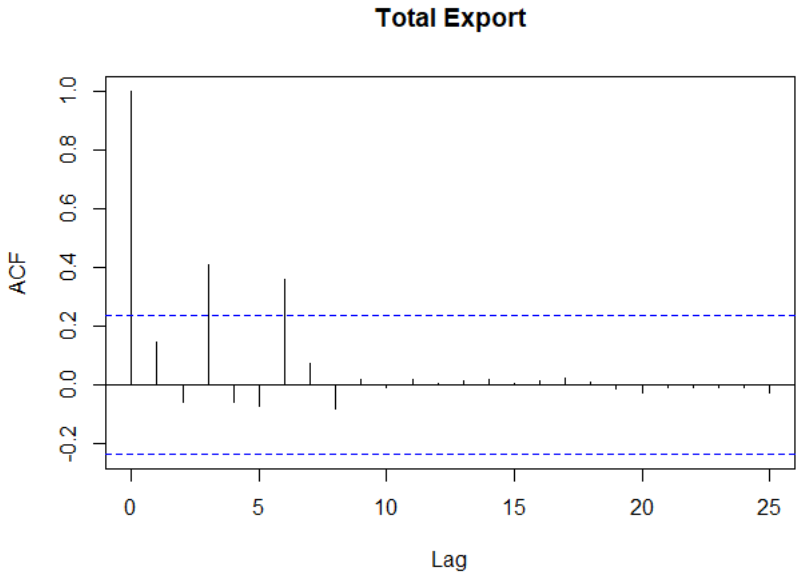

Figure 25.a: ACF plot of Export data

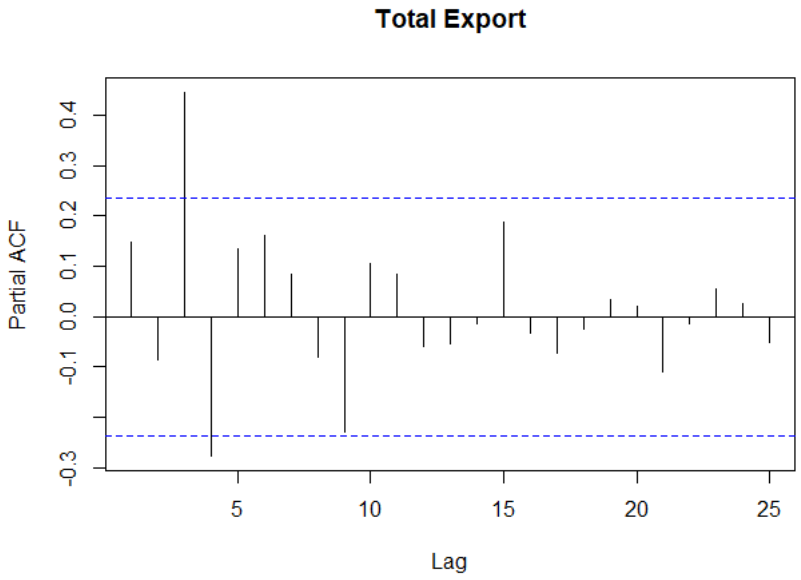

Figure 25.b: PACF plot of Export data

By analysing above figures (25.a & 25.b) it can be seen that the order of regression for export data is 3 since two lines are outside significance levels.

Using the p characteristic value i.e. order of regression based on the analysis of ACF and PACF plots the model is built on the detrended data and is shown in the figure below

```
Call:
arima(x = ts_data, order = c(1, 1, 0))

Coefficients:
         ar1
      0.2160
s.e.  0.1256

sigma^2 estimated as 776454633:  log likelihood = -804.15,  aic = 1612.31
```

AR model for Import data

```
Call:
arima(x = ts_data, order = c(3, 2, 0))

Coefficients:
          ar1      ar2     ar3
      -0.5691  -0.5974  0.1877
s.e.   0.1229   0.1278  0.1301

sigma^2 estimated as 210194623:  log likelihood = -748.95,  aic = 1505.91
```

AR model for Export data

The built model is fitted upon the actual values to the see the fit of the model in the figures 26.a & 26.b below. It can be observed that the model does not fit the actual value is good enough and there are discrepancies in the fit. Though the fit is not good on comparing it with simple exponential smoothing models we can say it fits better.

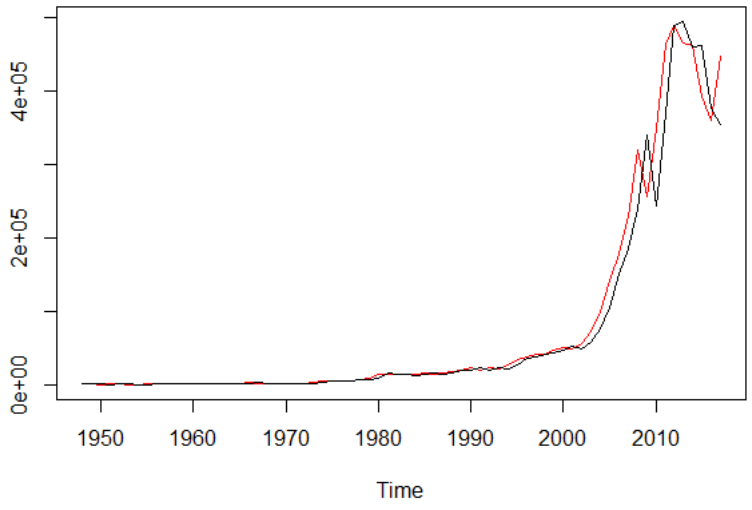

Figure 26.a: AR Fitted values for Import data

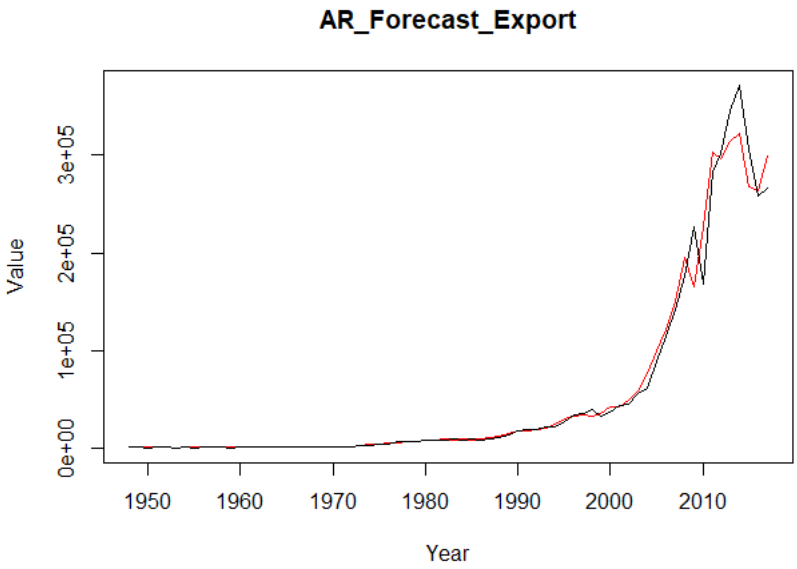

Figure 26.b: AR fitted values for export data

With model been built and tested on out of sample forecast the model performs better than that of simple exponential smoothing. The forecast of the model is shown in the figure 27.a and 27.b for import data and figure for export data

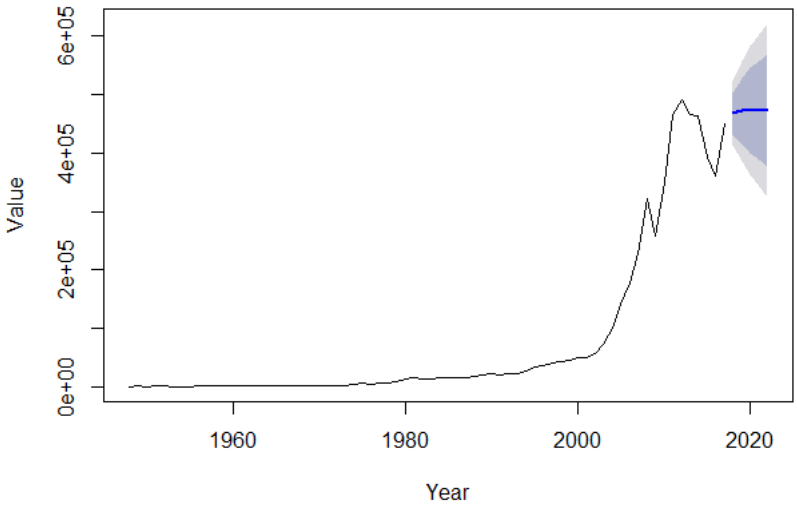

Figure 27.a: AR Forecast of Total Imports

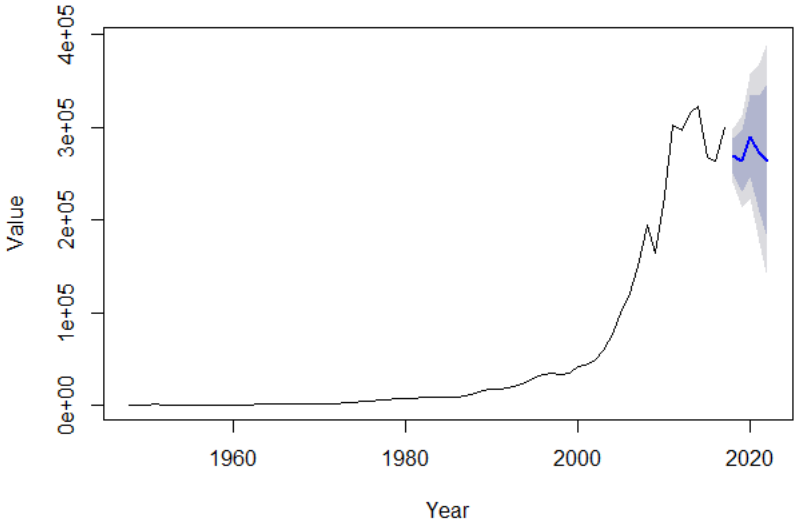

Figure 27.a: AR Forecast of Total Exports

On validating the model for out of sample forecast the accuracy metrics are found as shown in table below

ME	RMSE	MAE	MPE	MAPE
5275.31	27665.18	11384.08	5.07	14.83

Table 3.a: Forecast accuracy of AR model for import data

ME	RMSE	MAE	MPE	MAPE	MASE
-16.43	14289.47	5793.35	1.28	10.65	0.81

Table 3.b: Forecast accuracy of AR model for Export data

ARIMA

Auto Regressive Integrated Moving Average (ARIMA) model predicts the value of future based on its own past timeseries values, its own lag and lagged forecast errors. The model is characterised by three terms namely p, q, d

Where

p =order of the AR term

q = Order of MA term

d = number of differencing required to make the series stationary

The equation of ARIMA model is

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

Here

Y_t = Forecast

α = Intercept term

$\beta_1, \beta_2, \ldots, \beta_p$ = Coefficient of lag1, lag2, … ,lag p

$\phi_1, \phi_2, \ldots, \phi_q$ = Coefficient of error terms of 1,2, …,q

ϵ_t = Error term of time period t

The with these characteristics we can model manual ARIMA and auto ARIMA by following the steps listed below

Step 1: ACF and PACF plots - on differenced TS

As discussed in the Auto regression (sec. 4.4.2) we are considering the same autocorrelation and partial auto correlation functions and the plots for both the data are drawn both import and export data as shown in the figure 28.a, 28.b.

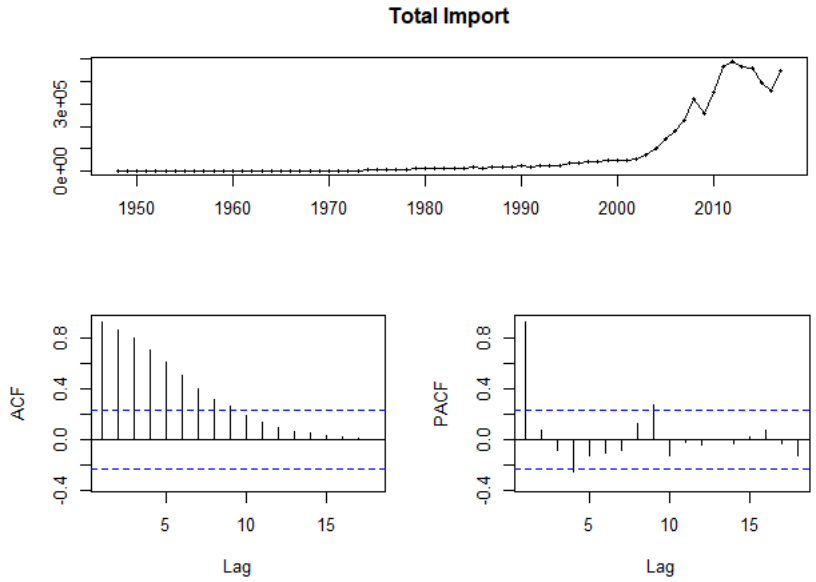

Figure 28.a: Combined ACF and PACF plots for import data

By analysing both the figures we can come to the conclusion that the p, d, q values for the import data is 2,1,2 and for total export it is 3, 2, 2.

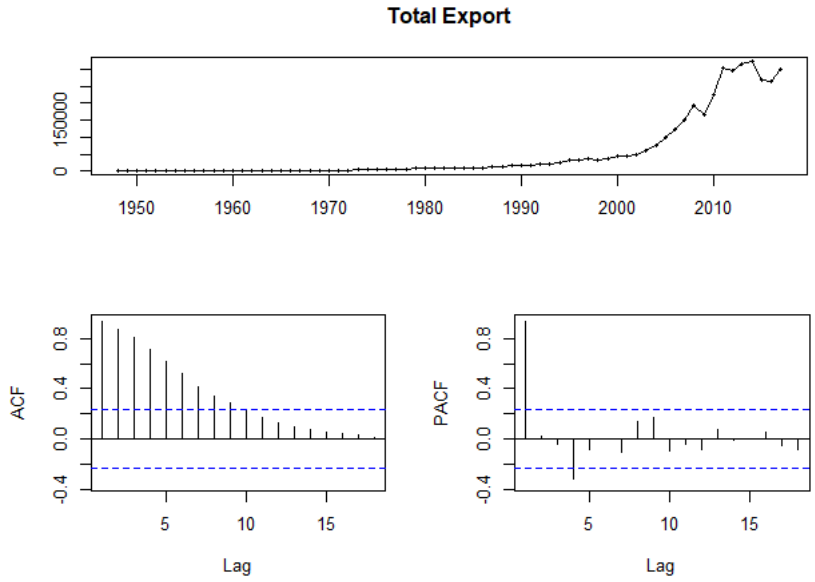

Figure 28.b: Combined ACF and PACF plots for Total Export

Step 2: Build and fit the manual ARIMA model with p, d, q

From the analysis of the ACF and PACF plots the values of p, d, q are chosen in the step 1 with those values the model is built for both the total imports and exports as shown below and also the coefficients can be found.

```
Call:
arima(x = ts_data, order = c(1, 3, 3))

Coefficients:
          ar1      ma1      ma2     ma3
      -0.4409  -1.3010  -0.3295  0.6317
s.e.   0.3468   0.3806   0.5213  0.3568

sigma^2 estimated as 727670520:  log likelihood = -784.33,  aic = 1578.66
```

For total Imports

```
Call:
arima(x = ts_data, order = c(3, 2, 2))

Coefficients:
          ar1      ar2     ar3      ma1      ma2
      -0.0984  -0.1155  0.5362  -0.5893  -0.3751
s.e.   0.1782   0.1234  0.1534   0.1902   0.1480

sigma^2 estimated as 187128041:  log likelihood = -745.59,  aic = 1503.17
```

For total exports

Step 3: Ljung - Box Test - Residual Analysis

Ljung- Box Test is a test applied on the residuals of any time series data once the model is built. It helps us in examining the autocorrelations existing in the residuals. Ljung box test is used to test the following hypothesis

Ho: Residuals are independent

Ha: Residuals are not independent

When the residuals are independent and if there's no pattern persists in residuals plot it confirms the goodness of the model.

```
        Box-Ljung test

data:  arma_mod$residuals
X-squared = 13.865, df = 25, p-value = 0.964
```

```
        Box-Ljung test

data:  arma_mod$residuals
X-squared = 9.179, df = 25, p-value = 0.9983
```

It can be noticed from the above output that Box-pierce test has been applied on the residuals of two models built. The p-value of manual ARIMA model in both the cases the p value is greater than 0.05 thus accepting the alternate hypothesis that the residuals are independent which depends on any other observations reveal the goodness of model. So, the model can predict the values with good amount of accuracy.

Step 4: Building and fitting Auto ARIMA model

With learnings from the manual ARIMA model and the learnings are applied to auto ARIMA the model built is

```
Series: ts_data
ARIMA(0,2,1)

Coefficients:
          ma1
      -0.9151
s.e.   0.0575

sigma^2 estimated as 787569707:  log likelihood=-793.36
AIC=1590.73   AICc=1590.91   BIC=1595.17
```

For total imports

```
Series: ts_data
ARIMA(5,2,0)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.5712  -0.6366  -0.1360  -0.4186  -0.3607
s.e.   0.1173   0.1320   0.1551   0.1413   0.1259

sigma^2 estimated as 192715525:  log likelihood=-743.87
AIC=1499.74   AICc=1501.11   BIC=1513.05
```

For total exports.

Then the values are fitted to the auto ARIMA very well and the same is represented graphically in the figures 29.a & 29.b.

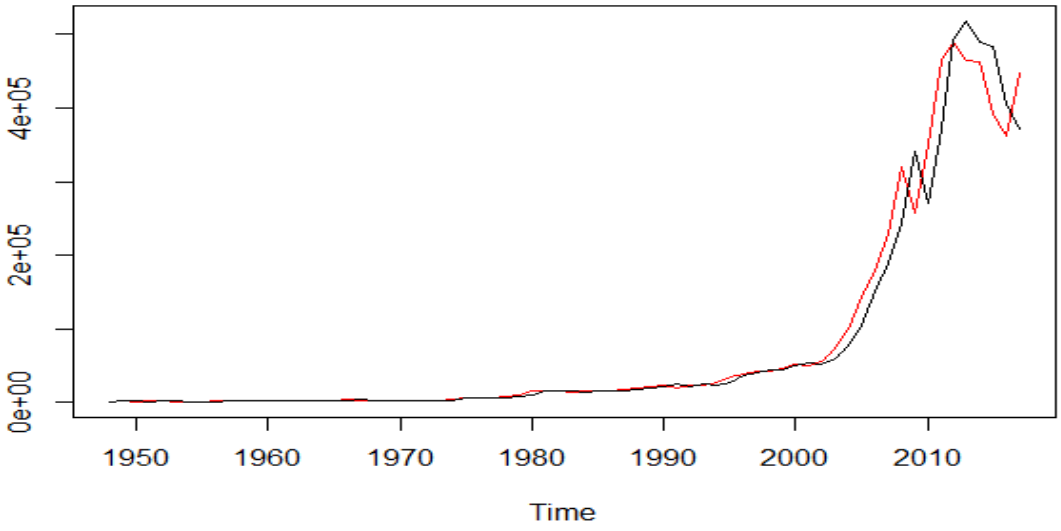

Figure 29.a: Auto ARIMA fit of Total imports

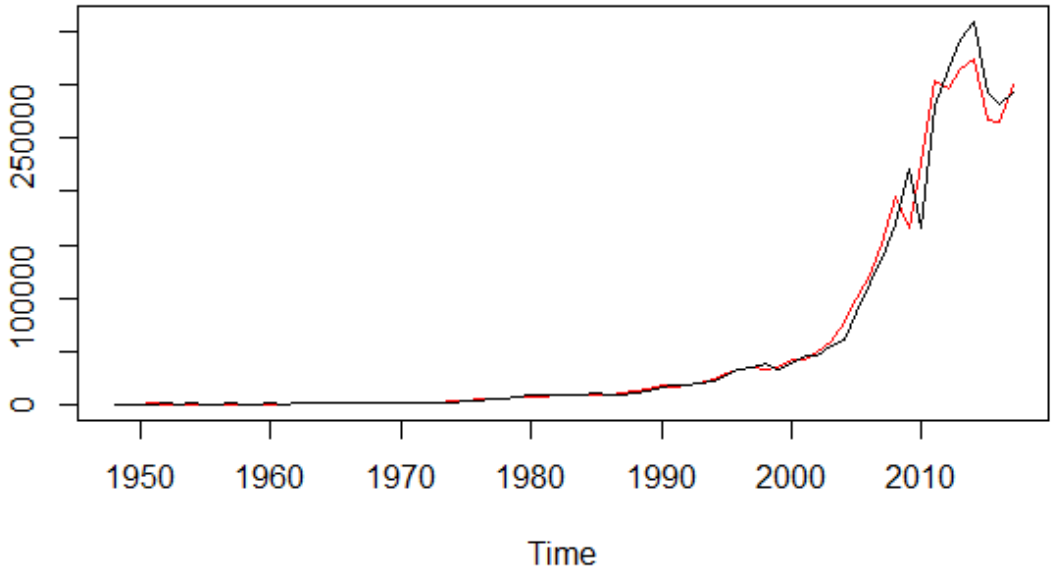

Figure 29.b: Auto ARIMA fit of Total exports

Step 5: Forecasting with the Manual ARIMA model

The built in ARIMA model has been used to forecast the values for next 5 years that is for year 2019, 2020, 2021, 2022, 2023. The plot of the forecast values are

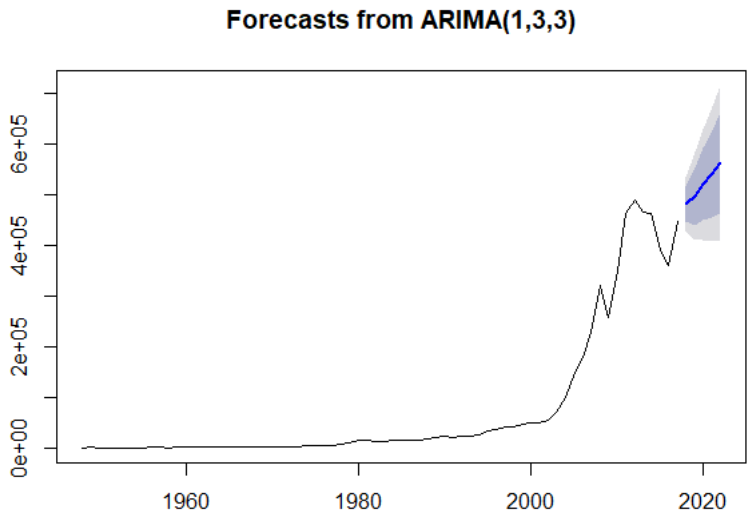

Figure 30.a: Manual ARIMA Forecast for Total Imports

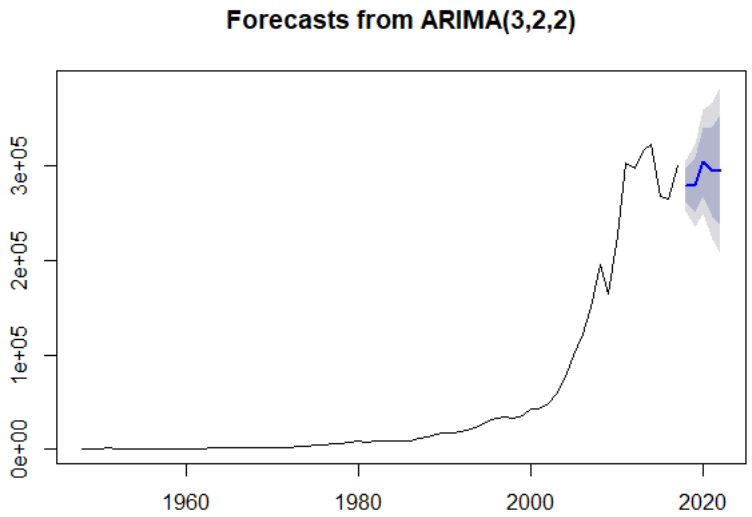

Figure 30.b: Manual ARIMA Forecast for Total Exports

Step 6: Forecasting with Auto ARIMA Model

The built auto ARIMA model is used to forecast for next three year out of sample i.e. year 2019, 2020, 2021 and the plot of the same is shown in the figure 31.a & 31.b

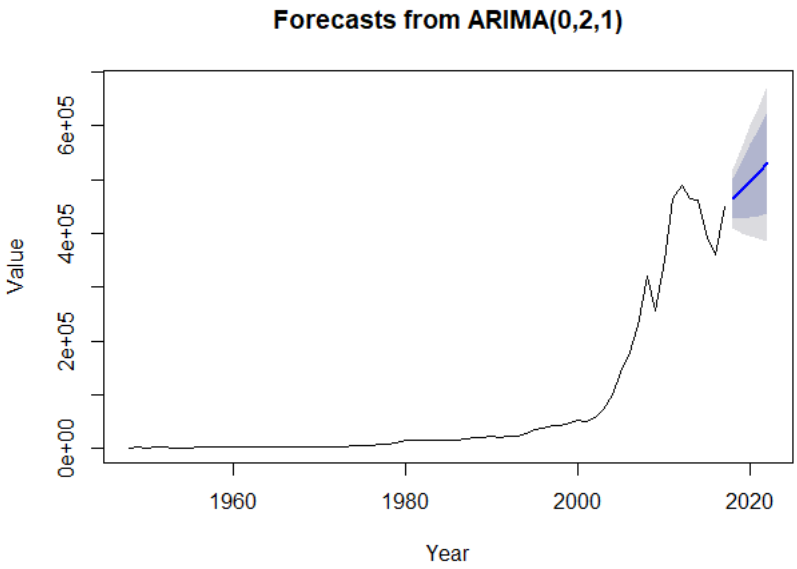

Figure 31.a: Auto ARIMA Forecast for Total Imports

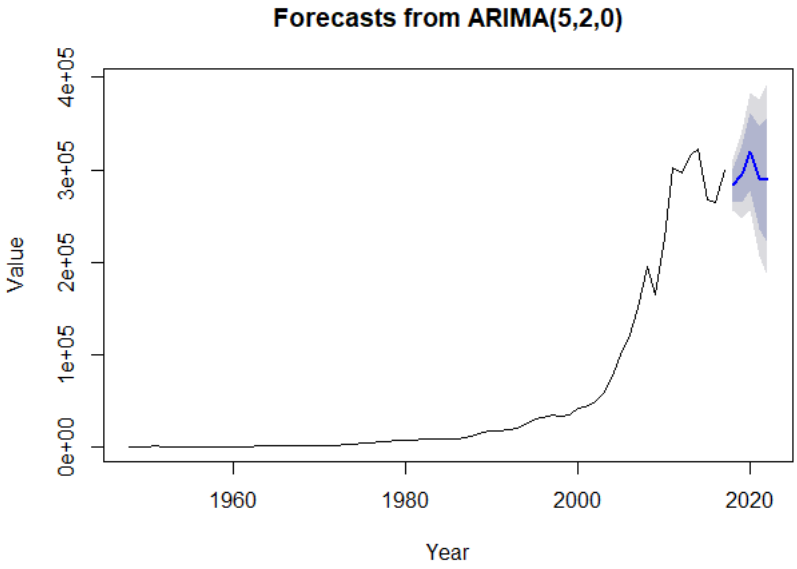

Figure 31.b: Auto ARIMA forecast for Total Exports.

Step 7: Compute accuracy of the forecast.

The models have been built and fitted with the actual values and optimised with the residual values based on Ljung-Box test. These models are used for forecasting and the forecasted values are represented in the graphs as shown above. Now to validate the model accuracy of the models are checked and the results are shown in the table below

Model	ME	RMSE	MAE	MPE	MAPE
Total Imports					
ARIMA	2262.14	26391.00	11015.17	3.48	13.03
Auto Arima	2751.49	27455.72	11965.47	2.69	14.23
Total Exports					
ARMA	-16.43	14289.47	5793.35	1.28	10.65
Auto Arima	64.05	13169.81	5574.27	1.64	10.23

Table 04: Accuracy measures of ARIMA model

ARTIFICIAL NEURAL NETWORK MODEL

Artificial neural networks are simple forecasting methods that are based on mathematical models of the brain. They establishes a complex nonlinear relationships between the predictor variable and response variables. A neural network is a network of "neurons" which are organised in layers. The inputs (or predictors) form the bottom layer, and the outputs (or forecasts) form the top layer. They may also contain intermediate layers which has "hidden neurons".

Here the Neural Network Auto Regression is used which are the lagged values of the time series that are used as inputs to a neural networks, just as lagged values used in a linear auto regression model in section 4.4.2. Feed forward networks with one hidden layer is alone used to build the model here and a NNAR(p,0) model is equivalent to an ARIMA(p,0,0) model or simple AR model, but it don't have the restrictions on the parameters to ensure stationarity. The built model is shown below in the figure

```
Series: ts_data
Model:  NNAR(1,1)
Call:   nnetar(y = ts_data, P = 1, repeats = 30, lambda = "auto")

Average of 30 networks, each of which is
a 1-1-1 network with 4 weights
options were - linear output units

sigma^2 estimated as 0.04547
```

For total Imports

```
Series: ts_data
Model:  NNAR(1,1)
Call:   nnetar(y = ts_data, P = 1, repeats = 30, lambda = "auto")

Average of 30 networks, each of which is
a 1-1-1 network with 4 weights
options were - linear output units
```

For Total Exports

The nnetar() function used here fits NNAR(p,0) model.

When applying the learning for forecasting, the network is applied iteratively. For forecasting one step ahead, we simply use the available historical inputs and for forecasting two steps ahead, we use one-step forecast as an input, along with historical data. This process proceeds until all the required forecasts are computed. Here the forecast done for 3 periods ahead and the forecast is shown in the figure32.a & 32.b below for both the import and the export.

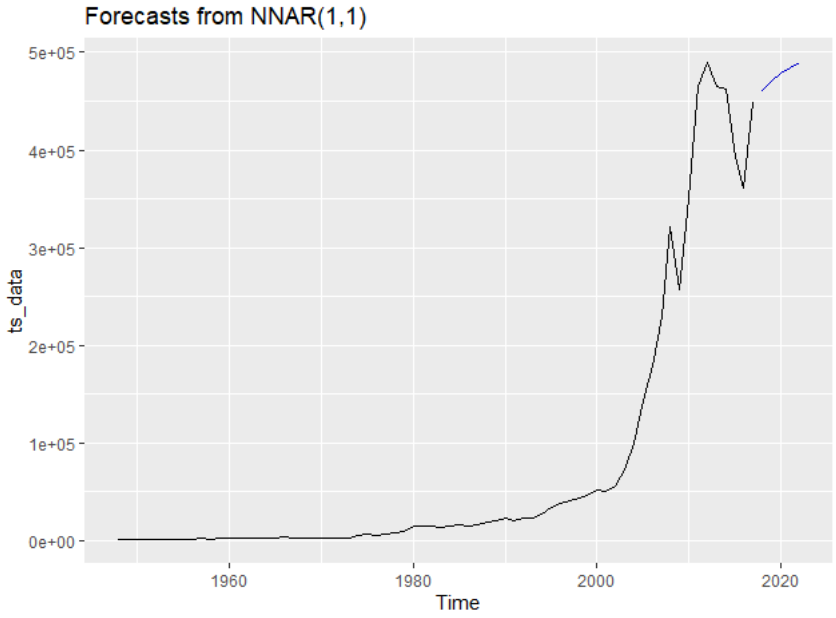

Figure 32.a: ANN forecast for Total Imports

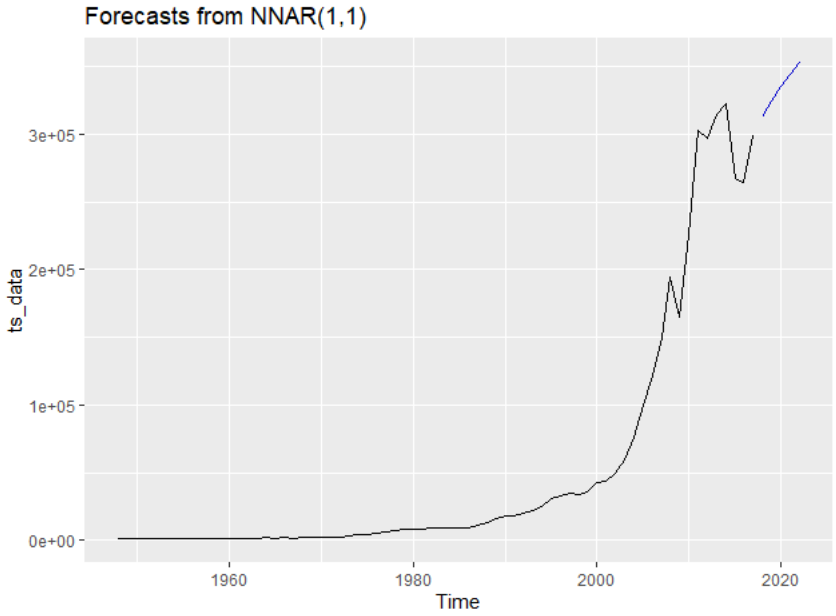

Figure 32.b: ANN forecast for total Exports

The above forecasted values are checked for their accuracy to ensure comparison with other models and the values of accuracy are given in the table below

Model	ME	RMSE	MAE	MPE	MAPE
Total Imports					
ANN	609.03	23604.67	9966.16	-1.58	13.65
Total Exports					
ANN	275.13	14478.17	5729.64	-0.68	9.19

Table 05: ANN Forecast accuracy measures

It is observed that the model predicts well for Exports but it fails to predict the import data since the MAPE values are more than 10% while a MAPE value of less than 10% is considered to be a good model.

MULTI LAYER PERCEPTRON NEWURAL NETWORK MODEL

Multilayer perceptron is a feed forward neural network, it consists of three components to it namely input layer, hidden layer and an output layer. It utilises supervised learning technique known as back propagation method. It can distinguish the data that are not linearly separable. Perceptron is an algorithm which classifies vector of numbers and decides whether the input belongs to some specific class.

In MLP a linear function maps the weighted inputs to the output of each neuron and then reduces the number of layer in the model. MLP consists at least of three layers of non-linearly activating nodes and each node in the layer connects with a certain weight to every other node in the layers. The learning occurs in the perceptron by changing weights after each piece of data is processed and optimises it based on the error in the output compared to the expected result. Thus building a superior model.

Here MLP model is built with the code

```
mlp(ts_data,m = 1, hd= NULL, reps = 30, comb = c("median", "mean", "mode")
, sel.lag = F, retrain = T, outplot =T)
```

The output for the code is shown in the figure

The output indicates that the resulting network has 5 hidden nodes and 4 input nodes, it was trained 30 times and the different forecasts were combined using the median operator. The mlp() function automatically generates ensembles of networks and the training of which starts with different random weights initially. Furthermore the inputs were included in the network. The grey nodes are auto regressions. The mlp() function accepts several arguments to fine-tune

the resulting network. The hd argument defines a fixed number of hidden nodes here it is given as null so that model trains itself and comes up with the best number of hidden nodes to accurately predict. The argument `reps` define the number of training repetitions are used, here there are two models with 20 and 30 repetitions used to make the model better predict the output. Best one is selected based on least MSE values because the algorithm uses least squares to decrease the error.

The argument retrain is used to retrain the neural network by using the model specifications mentioned in the model. Using regressors in the model decreases the increases the MSE of the model hence the use of regressors is avoided in the model.

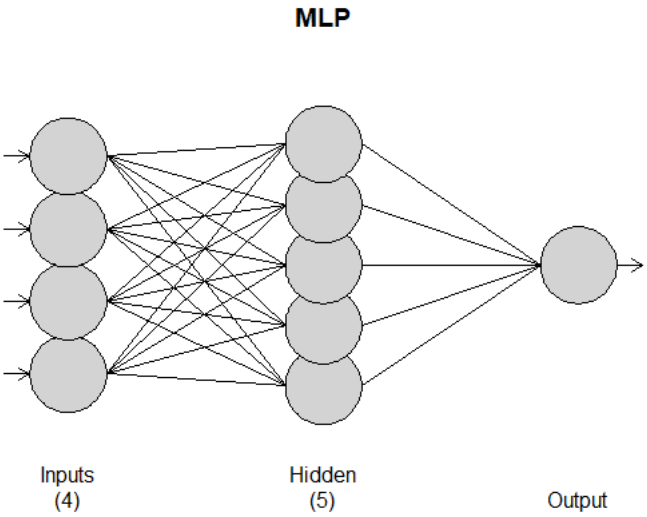

Figure 33: MLPNN model built for both Total Import and Total Export

Thus the build model is used to forecast for 5 time periods ahead and a plot of it is obtained The plot of the forecasts provides the forecasts of all the ensemble members in grey. The output of the function forecast() is of class forecast and those familiar with the forecast package will find familiar elements.

The plot of forecast using MLP is shown in the figure 34.a & 34.b for both total Import and Total export data.

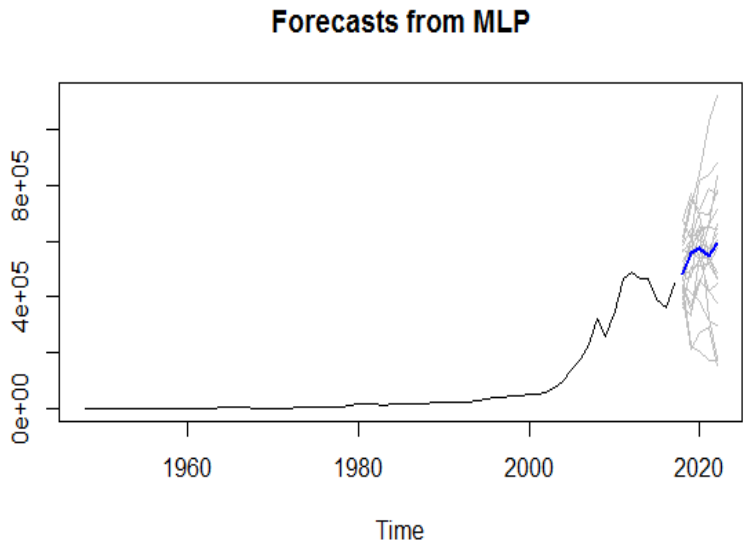

Figure 34.a: MLP forecast of Total Imports

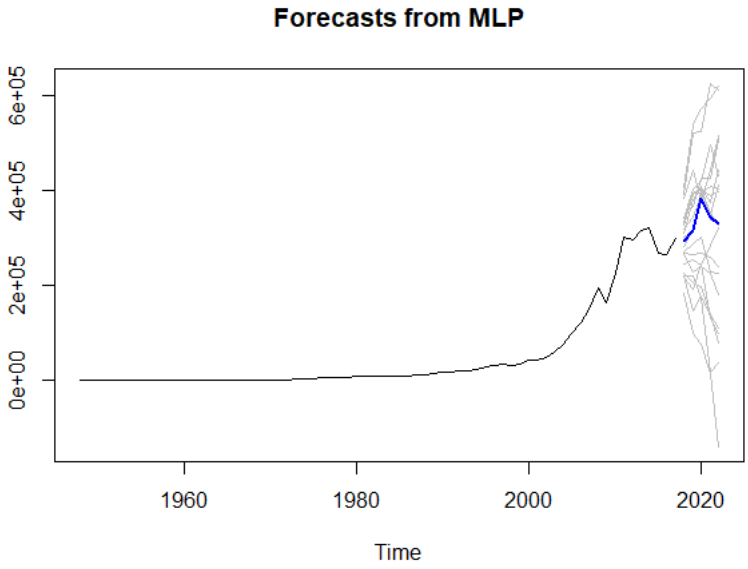

Figure 34.b: MLP forecast of Total Exports

Now the model forecast and fitted values are compared to check the accuracy of the model. The model accuracy is found with the help of accuracy function in the forecast package and the evaluation metrics for the model found to be.

Model	ME	RMSE	MAE	MPE	MAPE
Total Imports					
MLPNN	-64.61	2568.85	1673.20	-5.60	9.89
Total Exports					
MLPNN	-32.94	1719.16	1042.52	-6.33	8.98

Table 06: Forecast accuracy of MLPNN model

TBATS MODEL

TBATS stands for

T: Trigonometric seasonality

B: Box-Cox transformation

A: ARIMA errors

T: Trend

S: Seasonal components

The model was introduced by De Livera et al (JASA, 2011) [21]. It uses exponential smoothing, allows automatic Box-Cox transformation and ARMA errors. TBATS model has capability to deal with complex data with seasonality with no seasonality constraints, making it possible to\ forecast for long term.

The modelling of the algorithm is completely automated in R. In modelling the TBATS model it is fed with box cox transformation by using argument use.box.cox, with the inclusion of use.damped.trend argument which decreases the effect of trend in the prediction. To increase the accuracy of the model the ARMA errors are also used and these data on errors are fed to the learning of the model.

	Length	Class	Mode
lambda	1	-none-	numeric
alpha	1	-none-	numeric
beta	1	-none-	numeric
damping.parameter	1	-none-	numeric
gamma.values	0	-none-	NULL
ar.coefficients	0	-none-	NULL
ma.coefficients	0	-none-	NULL
likelihood	1	-none-	numeric
optim.return.code	1	-none-	numeric
variance	1	-none-	numeric
AIC	1	-none-	numeric
parameters	2	-none-	list
seed.states	2	-none-	numeric
fitted.values	70	ts	numeric
errors	70	ts	numeric
x	140	-none-	numeric
seasonal.periods	0	-none-	NULL
y	70	ts	numeric
call	5	-none-	call
series	1	-none-	character
method	1	-none-	character

Here due to lack of complex seasonality in the data the TBATS model built on the non-seasonal data. The data is passed on to the model and results are observed as shown above. The built BATS model is applied to forecast the data for out of sample horizon of 5 time periods i.e. for five years after 2018. The plot of the forecast is shown below in the figure

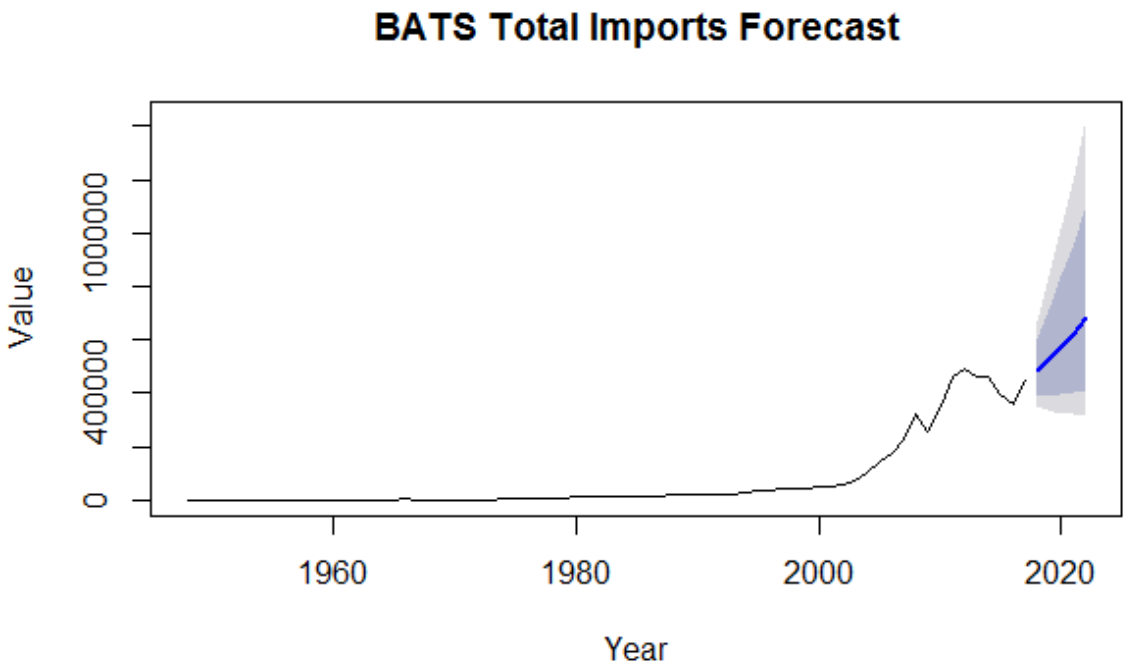

Figure 35.a: BATS Forecast for Total Imports

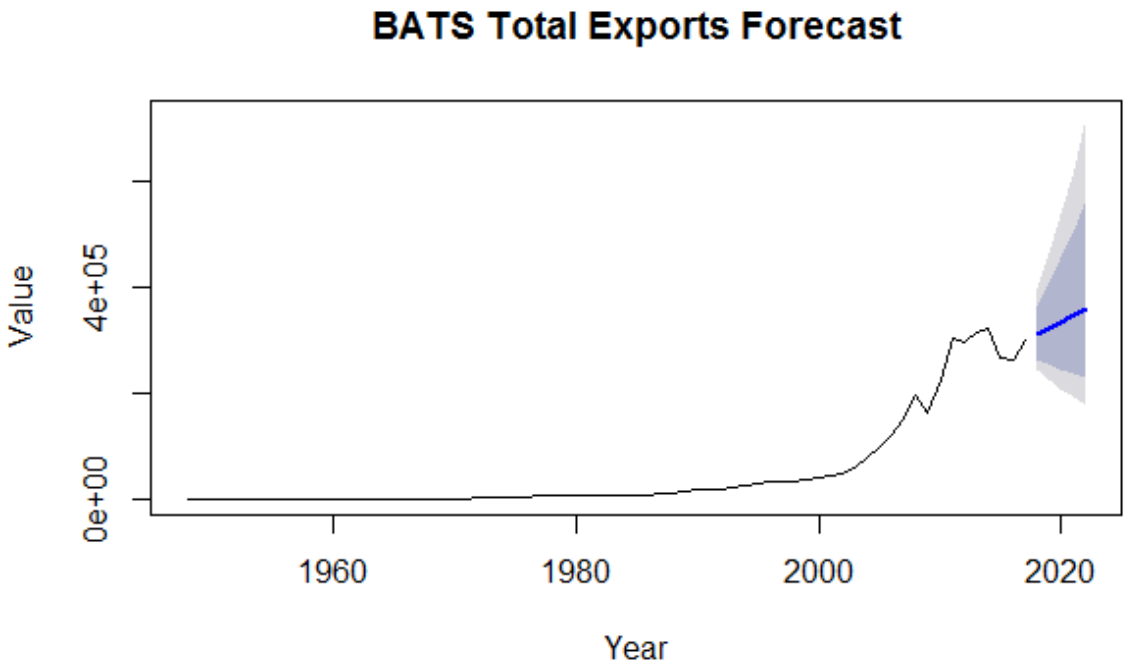

Figure 35.b: BATS forecast for total exports

To evaluate the model the accuracy function of forecast package is used and the results of evaluations are found to be satisfying. The model accuracy measures are given in the table below

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**
	Total import				
TBATS	-208.81	28821.14	12643.82	0.42	14.39
	Total export				
TBATS	-1168.92	17692.00	7207.49	1.08	9.68

Table 07: Accuracy measures of TBATS model.

It can be analysed that the model performs well in predicting the exports data but fails to forecast import data which is evident from the high value of mean absolute percentage error.

ACCURACY OF MODELS – COMPARISION

To compare the performances of various models built on international trade of exports and imports from India time series dataset, the accuracy measure- Mean Error (ME), Root mean square error (RMSE), Mean absolute error (MAE), Mean Percentage error (MPE) and Mean Absolute Percentage Error (MAPE) on in sample data is been tabulated below to understand the effectiveness of various time series modelling techniques.

TOTAL IMPORT	ME	RMSE	MAE	MPE	MAPE
SES	6383.87	28260.8	12191.5	6.18	15.03
AR	5275.31	27665.2	11384.1	5.07	14.83
Manual ARIMA	2262.14	26391	11015.2	3.48	13.03
Auto ARIMA	2751.49	27455.7	11965.5	2.69	14.23
ANN	609.03	23604.7	9966.16	-1.58	13.65
MLPNN	-64.61	2568.85	1673.2	-5.6	9.89
TBATS	-208.81	28821.1	12643.8	0.42	14.39

Table 08.a: Model Comparison for Total Imports data

By analysing all the metrics it can be seen from the table that Multi-layer perceptron performs well across all metrics for import data.

TOTAL EXPORT	ME	RMSE	MAE	MPE	MAPE
SES	4257.16	16756.3	7067.16	6.76	11.01
AR	-16.43	14289.5	5793.35	1.28	10.65
Manual ARIMA	-16.43	14289.5	5793.35	1.28	10.65
Auto ARIMA	64.05	13169.8	5574.28	1.64	10.23
ANN	275.13	14478.2	5729.64	-0.68	9.19
MLPNN	-32.94	1719.16	1042.52	-6.33	8.98
TBATS	-1168.9	17692	7207.49	1.08	9.68

Table 08.b: Model comparison for Total Exports data

By analysing all the metrics it can be seen from the table that Multi-layer perceptron performs well across all metrics for import data.

So for both imports and exports data the multi-layer perceptron neural network predicts well across all metrics.

CHAPTER 05

5 FINDINGS AND RESULTS

5.1 RESULTS OF VISUALISATION OF WEB ANALYTICS:

The key findings and results of the visualisation dashboard are listed below

1. There are totally 48,669 users of the website until 18-sept-2019 with an increase of 2.6% traffic compared to previous month.
2. It is observed that average time on page by an uses is about 51 seconds
3. Webpage is listed in the search engine by organic searches for about 58.8%, referrals about 15.1%, Direct visitors about 11.9%.
4. It can be seen that there is an increase in organic searches but at the same time there was a decrease in the number of page views.
5. The decrease in the number of page views can be attributed to the increase in the bounce rate i.e. number of customers leaving the webpage immediately after visiting the site.
6. To view the webpage desktop is the most used device by the visitors of the page, followed by mobile and tablet at the bottom.
7. It is seen that there is a increasing trend in the visitors.
8. Seasonality is observed in the data since the data repeats itself for every seven days.
9. On comparing the data of the current year vs previous year we can observe that there is a decrease in users of the webpage.
10. Number of unique visitors to the webpage is high for the home page alone that is most visitors see only the home page and leave the site.
11. Bounce rate is very high for the webpage number 5 and 7, thus both the pages has to be relooked and optimised for superior experience to decrease bounce rate.

The visualisation and the dashboard to monitor the website visitors is shown in the next page

5.2 RESULTS OF AIR PASSENGERS FORECASTING:

The results of the air passenger forecast across the metrics are shown below

```
> accuracy(fcast_Air)
                     ME       RMSE        MAE         MPE      MAPE      MASE      ACF1
Training set 0.0005730623 0.03504883 0.02626034 0.01098898 0.4752815 0.2169522 0.0144393
```

The model forecasts well since the error in the forecast are very negligible. The mean absolute percentage error is observed to be of very less about 0.475%. thus the ARIMA model can forecast the data very well.

The key results are

1. The air passengers for the upcoming years will see a tremendous increase due to steep trend in the forecast.
2. The ARIMA model performs better when optimising the model with different p. d, q characteristics of the model based on the analysis from the residuals of the model building.
3. During the month of June to August each year airline industry sees a spike in demand so it will be best period to increase their revenue by offering more attractive services to the air travellers.

5.3 FINDINGS AND RESULTS OF INTERNATIONAL TRADE FORECASTING:

Findings:

The model on the performance is tested on the data of three broad categories of merchandised goods traded over the world and as classified by the World Trade Organisation are

1. Agricultural products
2. Fuels and Mining products
3. Manufactures

So, the import and export values are forecasted for these categories as designated by World Trade Organization, the findings are

For agricultural products the performance of all the models are found to be as shown in the table

	ME	RMSE	MAE	MPE	MAPE
	AGRICULTURAL PRODUCTS IMPORTS				
SES	744.66	1477.37	989.73	5.68	15.98
AR	55.8152	1130.03	727.486	0.04573	14.8051
Manual ARIMA	50.9815	1100.13	700.646	0.03564	13.1061
Auto ARIMA	136.794	1138.7	754.858	0.27835	15.4779
ANN	27.8058	776.512	575.968	-2.7142	13.9595
MLPNN	-4.5381	300.831	197.562	-1.3017	9.18191
TBATS	127.025	1200.05	775.711	1.27497	14.6039
	AGRICULTURAL PRODUCTS EXPORTS				
SES	829.89	3245.96	1792.86	5.46	11.03
AR	-93.003	3125.35	1601.23	0.74225	10.4516
Manual ARIMA	-93.003	3125.35	1601.23	0.74225	10.4516
Auto Arima	572.917	2833.82	1528	3.92702	9.96005
ANN	147.59	2803.05	1552.75	-0.5573	10.3288
MLPNN	27.769	1356.03	842.655	0.41039	8.84506
TBATS	-96.683	3027.81	1593.72	0.88422	10.6935

Table 09: Model performance on Agricultural products.

For Fuels and mining products the performance of all the models are found to be as shown in the table below

Model	ME	RMSE	MAE	MPE	MAPE
	FUELS AND MINING PRODUCTS IMPORTS				
SES	2754.15	20470.78	11057.75	4.67	20.53
AR	-1753.97	21326.67	10814.61	-0.62	21.13
Manual ARIMA	-1605.97	19989.67	9808.61	-0.59	19.68
Auto ARIMA	2732.94	20469.84	11036.40	4.37	20.23
ANN	1275.49	18106.01	10141.76	-2.43	18.50
MLPNN	-55.76	6022.90	3341.40	-0.42	12.94
TBATS	-1362.69	21141.19	10947.35	0.96	21.01
	FUELS AND MINING PRODUCTS EXPORTS				
SES	964.35	7953.52	3924.52	3.96	24.45
AR	720.50	7650.43	3601.81	0.60	26.11
Manual ARIMA	-710.50	9990.43	4508.10	0.60	23.42
Auto Arima	966.54	7953.17	3921.86	4.32	24.09
ANN	434.68	7132.87	3795.52	-6.34	22.96
MLPNN	11.43	1440.19	784.52	-7.56	17.42
TBATS	115.45	7617.78	3615.65	-3.62	30.17

Table 10: Model performance of Fuels and mining products

For Fuels and mining products the performance of all the models are found to be as shown in the table below

	ME	RMSE	MAE	MPE	MAPE
	MANUFACTURES IMPORTS				
SES	4800.62	11672.24	6598.58	6.73	13.88
AR	38.56	9512.33	5477.40	1.34	10.74
Manual ARIMA	37.67	9815.36	5747.40	1.85	11.00
Auto ARIMA	38.56	9512.33	5477.40	1.34	10.74
ANN	148.81	7521.08	4316.20	-0.57	10.06
MLPNN	0.67	1838.05	1303.82	-1.01	9.28
TBATS	-546.75	11539.43	7397.91	1.03	13.30
	MANUFACTURES EXPORTS				
SES	4967.74	12546.22	7345.49	8.81	11.08
AR	31.18	9176.58	5281.21	2.48	7.99
Manual ARIMA	34.78	9712.64	5364.53	2.86	7.12
Auto Arima	31.18	9176.58	5281.21	2.48	7.99
ANN	307.60	10032.77	5314.72	-0.47	8.13
MLPNN	-11.13	1721.36	1216.66	-1.61	7.06
TBATS	-774.70	12470.70	6540.10	0.05	9.62

Table 11: Model performance of Manufactures

From above three tables it can be seen that the model Multi-layer perceptron model works well since the model performs well across all the metrics compared to other models. The reason for ht model performing better when compared to other models are the model uses back propagation technique and ordinary least squares methods in which the model is self-adjusted for best fitted value based on the median of the data. Due to this optimization possible in the perceptron the model is able to do fairly well than other models under consideration.

Now, it is clear that the Multi-layer perceptron method forecasts the data very well. So this training of the model is applied to out of sample to check the real time working of the model. For convenience and availability of the data the year 2018 which was in the sample is made out of sample and the model is retrained on the data. This training of the models are used to forecast for the out of sample year 2018. To evaluate the error in the forecast is used. Since this is deployment we have to concentrate on error to the predicted value. Error is calculated using the formula

$$\text{Error} = \text{Actual value} - \text{Forecast value}$$

The results of the out of sample prediction are tabulated in the table in the next page

MODEL	AGRICULTURE			FUEL AND MINING			MANUFACTURES			TOTAL MERCHANDISE		
	IMPORTS											
	Forecast value	Actual value	Error	Forecast value	Actual value	Error	Forecast value	Actual value	Error	Forecast value	Actual value	Error
SES	29031.6	33253.8	4222.2	108094.9	150892.3	42797.4	186022.8	234895.0	48872.2	448414.3	510664.7	62250.4
AR	29353.0	33253.8	3900.8	74678.4	150892.3	76213.9	188759.3	234895.0	46135.7	467263.4	510664.7	43401.3
Manual ARIMA	30260.0	33253.8	2993.7	84916.2	150892.3	65976.1	189580.0	234895.0	45315.0	481677.8	510664.7	28986.9
Auto ARIMA	30550.7	33253.8	2703.1	108092.7	150892.3	42799.6	188759.3	234895.0	46135.7	464845.9	510664.7	45818.8
ANN	28462.0	33253.8	4791.8	127674.1	150892.3	23218.2	189613.9	234895.0	45281.1	461001.9	510664.7	49662.8
MLPNN	31330.4	33253.8	1923.4	128787.5	150892.3	22104.8	190067.2	234895.0	44827.8	499779.1	510664.7	10885.6
TBATS	31287.4	33253.8	1966.4	103117.2	150892.3	47775.1	201171.9	234895.0	33723.1	487959.1	510664.7	22705.6
	EXPORTS											
	Forecast value	Actual value	Error	Forecast value	Actual value	Error	Forecast value	Actual value	Error	Forecast value	Actual value	Error
SES	33544.7	39495.5	5950.9	36442.8	48457.9	12015.1	188809.6	208329.6	19520.0	299271.9	325562.2	26290.3
AR	33546.7	39495.5	5948.8	32363.4	48457.9	16094.5	199392.1	208329.6	8937.5	269657.8	325562.2	55904.4
ARMA	35516.2	39495.5	3979.3	43297.3	48457.9	5160.7	197504.8	208329.6	10824.8	278974.3	325562.2	46587.9
Auto Arima	34814.5	39495.5	4681.1	36441.9	48457.9	12016.1	199392.1	208329.6	8937.5	283531.8	325562.2	42030.4
ANN	35458.5	39495.5	4037.0	44895.9	48457.9	3562.0	195646.4	208329.6	12683.2	312486.2	325562.2	13076.0
MLPNN	35920.4	39495.5	3575.1	47269.7	48457.9	1188.3	206562.9	208329.6	1766.7	315858.6	325562.2	9703.6
TBATS	35640.7	39495.5	3854.8	36785.3	48457.9	11672.7	205593.3	208329.6	2736.3	309905.4	325562.2	15656.8

Highlighted values are best values. *Values are in millions

Table 12: Out of sample prediction of all models across imports and exports

From the table it can be observed that error values are very less for the multi-layer perceptron model. Thus the trained model when deployed for out of sample prediction it performs well and decision makers can use the model with high confidence.

Results:

From the analysis the key findings and results are

1. Multi-layer perceptron neural networks forecasts well for all the trade data and errors are less for both in-sample data and out-sample data.
2. Auto regression models are useful in modelling ARIMA models since we run through the AR models and find the p characteristic of the model. So, it becomes easy in the ARIMA model to incorporate the p characteristic of that model.
3. Differencing not only detrends the data but also it makes the data stationery so there is a multiple use for that technique.
4. Out of sample testing results are also good only for multi-layer perceptron model.
5. Median optimisation of Multi-layer perceptron model yields best result when compared to mean and mode of the data.
6. So, International trade value of merchandise goods can be forecasted with good accuracy with the help of machine learning since the built model forecasts the value with least error.
7. From the experience of building all these models it can be understood that the timeseries modelling of international trade the value of forecast will always depend on the immediate previous year value to a greater extent.
8. Though ANN have less accuracy measures it can predict the values out of the sample with good accuracy.
9. When information about trend and the errors from the ARMA model is fed into neural networks the model optimises the result to near perfect.
10. The TBATS model also predicts well in certain cases so its better to have the TBATS model as an alternate model and can be used to compare with MLPNN model for gaining better insights into the forecasts.
11. All the built models hold good for only short term forecasts and in long term qualitative forecasts provide a good result on the trade value.

CHAPTER 06

6 CONCLUSION

The tools like google analytics captures many datapoints about the websites. Due to the presence of lot of variables in the data captured it often becomes difficult with conventional software to process the data and generate insights for making decisions. So, Google data studio a business reporting software has been used to analyse the data. With the software, the analysis about the website data from google analytics tool is used to create different types of visualisation. Form the visualisation a interactive dashboard is created. the key insights from the data are generated from the dashboard. This dashboard can be used to generate reports about the website visitors for any time frame with a few clicks. These reports helps the company to make decisions about the website and redesign the websites according to the preferences of the visitors.

Forecasting air travel demand is important for airline operators and airports for better serving the people. So, better facilities airports like infrastructure facilities, logistics handling, runway etc. can be offered. Thus, data on international Air passengers flow is taken for the purpose of forecasting the number of international flyers. To Forecast the passengers demand ARIMA model is used which absorbs the seasonality and cyclicity in the data. This ARIMA model is evaluated across the evaluation metrics and the results are found to be very good with MAPE value of 0.475%. It can be inferred that the model forecasts the value with high accuracy. The model also performs well on the out of sample forecast too.

International trade one of the key component in the GDP of a nation has to be monitored and decisions are to be made to on the international trade to increase the economy of the nation. To make better decisions data about the future performance of the trade is important for the decision makers and these decision makers forecasts trade of a nation across all of their trading partners and trade categories. In this project the timeseries data on International trade of merchandise goods are taken for India across 25 traded categories with all the trading partners in the world to create a timeseries model for forecasting. For this purpose SES, AR, ARIMA, ANN, MLPNN, TBATS models are used. On building models and analysing them for best forecasting MLPNN- Multilayer perceptron Neural network model is found to be better forecasting than other models. So, this model when deployed for out of sample forecasting it forecasts the data with very less error compared to other models. Hence, this timeseries model can be used for trade forecasting.

7. REFERENCE

1. A.J. Samimi and K.D. Darabi, Forecasting government size in Iran using artificial neural network, Journal of Economics and Behavioral Studies, 3(5), (2011), 274–278.
2. A.P.N. Refenes, A. N. Burgess, and Y. Bentz, "Neural networks in financial engineering: A study in methodology," IEEE Transactions on Neural Networks, vol. 8, no. 6, pp. 1222 - 1267, 1997.
3. ACZEL, A. D., Complete Business Statistics, Irwin, 1989, ISBN 0-256-05710-8
4. Alam, H. M. (2011). An Econometrics Analysis of Export-Led Growth Hypothesis: Reflection from Pakistan. IJCRB , 2 (12), 732-744.
5. Ali, E., & Talukder, D. K. (2009). Preferential Trade among the SAARC Countries: Prospects and Challenges of Regional Integration in South Asia. JOAAG , 4 (1).
6. Armstrong, S., red. (2001),Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston: Kluwer Academic Publishing.
7. Blum, M., Riedmiller, M. (2013),Electricity Demand Forecasting Using Gaussian Processes. Palo Alto, California, AAAI Press, pp. 10-13.
8. Box, G. E. P., Jenkins, G. M. (1976),Time Series Analysis: Forecasting and Control. San Francisco: Holden Day.
9. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. i Ljung, G. M. (2016),Time Series nalysis: Forecasting and Control.Fifth red. New Jersey: John Wiley & Sons.
10. Brockwell, P. J., Davis, R. A. (1996),Introduction to Time Series and Forecasting. New York: Springer.
11. Brożyna, J., Mentel, G., Szetela, B. (2016),A Mid-Term Forecast of Maximum Demand for Electricity in Poland. Montenegrin Journal of Economics, 12(2), pp. 73-88.
12. Chen, J. F., Wang, W. M., Huang, C. M. (1995),Analysis of an Adaptive Time-Series Autoregressive Moving-Average (ARMA) Model for Short-Term Load Forecasting. Electric Power Systems Research, vol. 34, pp. 187-196.
13. Cottet, R., Smith, M. (2003),Bayesian Modelling and Forecasting of Intraday Electricity Load. Journal of the American Statistical Association, vol. 98, p. 839–849.
14. de Andrade, L. C. M., da Silva, I. N. (2009),Very Short-Term Load Forecasting Based on ARIMA Model and Intelligent Systems. Curitiba, IEEE, pp. 1-6.
15. De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American statistical association, 106(496), 1513-1527.

16. Dongxiao, N., Yongli, W., Desheng, D. W. (2010),Power Load Forecasting Using Support Vector Machine and Ant Colony Optimization. Expert Systems with Applications, Issue 37, pp. 2531-2539.
17. Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction. Neurocomputing, 172, 446-452. 4r
18. Hoptroff, R. G., Bramson, M. J., & Hall, T. J. (1991, July). Forecasting economic turning points with neural nets. In IJCNN-91-Seattle International Joint Conference on Neural Networks (Vol. 1, pp. 347-352). IEEE.
19. Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D.(2008), Forecasting with Exponential Smoothing: The State Space Approach. New York: Springer.
20. Keck, A., Raubold, A., & Truppia, A. (2010). Forecasting international trade. OECD Journal: Journal of Business Cycle Measurement and Analysis, 2009(2), 157-176.
21. Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with applications, 37(1), 479-489. 2r
22. Kuan-Yu Chen and Chia-Hui Ho, "An Improved Support Vector Regression Modeling for Taiwan Stock Exchange Market Weighted Index Forecasting", ICNN&B '05: International Conference on Neural Networks and Brain, Volume 3, , 2005
23. Küçükdeniz, T. (2010),Long Term Electricity Demand Forecasting: An Alternative Approach With Support Vector Machines. İstanbul Üniversitesi Mühendislik Bilimleri Dergisi, Issue 1, pp. 45-54.
24. L. Cao and F. Tay, "Support Vector Machine with adaptive parameters in financial time series forecasting," IEEE Transactions on Neural Networks, vol. 14, no. 6, pp. 1506–1518, 2003.
25. Lee, C. M. , Ko, C. N. (2011),Short–term Load Forecasting Using Lifting Scheme and ARIMA Models. Expert Systems with Applications, Tom 38, pp. 5902-5911.
26. M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decis.SupportSyst.37(4) (2004)567–581.
27. Manwa, F., & Wijeweera, A. (September 2016). Trade liberalisation and economic growth link: The case of Southern African Custom Union countries. Economic Analysis and Policy, 51, 12e21.
28. Minghui, H., Saratchandran, P., & Sundararajan, N. (2003, October). A sequential learning neural network for foreign exchange rate forecasting. In SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference

Theme-System Security and Assurance (Cat. No. 03CH37483) (Vol. 4, pp. 3963-3968). IEEE.

29. MONTGOMERY, D. C., JOHNSON, L. A., GARDINER, J. S., Forecasting and Time Series Analysis, McGraw-Hill, Inc., 1990, ISBN 0-07-042858-1
30. Pappas, S. S. et. al. (2008),Electricity Demand Loads Modeling Using Auto Regressive Moving Average (ARMA) Models. Energy, September, 33(9), pp.1353–1360.
31. Rahman, M. M., & Mamun, S. A. K. (October 2016). Energy use, international trade and economic growth nexus in Australia: New evidence from an extended growth model. Renewable and Sustainable Energy Reviews, 64, 806e816.
32. Rob JH(2014).Testing for trend in arima models
33. Robert HS, David SS(2011). Time Series Analysis and its Applications. ARIMA Models Springer Texts in Statistics 2011. Pp. 83-171.
34. Robert N(2014)(c). All Rights Reserved. Statistical forecasting: notes on regression and time series analysis. website. http://people.duke.edu/~rnau/411home.htm.last updated 10/30/2014.
35. r-project https://www.r-project.org/about.html
36. Ruffin, R. J. (2002). David Ricardo's Discovery of Comparative Advantage. History of Political Economy , 34 (4), 727-748.
37. Stock, J. H. and Watson, M. W. (2003) Introduction to Econometrics, Boston, MA: Pearson Education, Inc.
38. Taylor, J. W. (2006),Comparison of Univariate Methods for Forecasting Electricity Demand Up to a Day Ahead. International Journal of Forecasting, Issue 22, pp. 1-16.
39. Taylor, J. W., De Menezes, L. M., McSharry, P. E. (2006), A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day ahead. International Journal of Forecasting, 22(1), pp. 1-16.
40. Wong, H. T. (2010). Terms of trade and economic growth in Japan and Korea: An empirical analysis. Empirical Economics, 38, 139. http://dx.doi.org/10.1007/s00181-009-0259-9.
41. World bank trade statistics. https://wits.worldbank.org/CountryProfile/en/IND
42. World Trade Organization (WTO) (2008) World Trade Report 2008: Trade in a globalizing world, Geneva: WTO.
43. Zeliaś, A., Pawełek, B., Wanat, S. (2016), Prognozowanie ekonomiczne: teoria, przykłady, zadania.Warszawa: Wydawnictwo Naukowe PWN.

44. Zhou, P., Ang, B. W., Poh, K. L. (2006),A Trigonometric Grey Prediction Approach to Forecasting Electricity Demand. Energy, Issue 31, pp. 2839-2847.

8 ANNEXURE

The machine learning code used for the project is attached below:

```
## Project

Setwd("……..")

##Essential libraries

library(xlsx)

library(tseries)

library(forecast)

library(ggplot2)

library(stats)

library(Metrics)

##reading data

prodata<- read.xlsx("tsdata.xlsx", sheetName = "Total_export")

str(prodata)

ts_data<-ts(prodata[,2], start = 1948, end = 2017, frequency = 1)

str(ts_data)

## When data starts and Ends

start(ts_data)

end(ts_data)

frequency(ts_data)

##Summary statistics

summary(ts_data)

##visual of data

plot.ts(ts_data, xlab = "Year", ylab = "Value", main = "Total Export")
```

```
## visualising trend in the dataset

abline(reg=lm(ts_data~time(ts_data)), xlab = "Year", ylab = "Value")

cycle(ts_data)

##aggregate plot

plot(aggregate(ts_data, FUN = mean), main = "Total Export"

## box plot

boxplot(ts_data~ cycle(ts_data), xlab = "Year", ylab = "Value", main = "Total Export")

##More plots to analyse the data

gglagplot(ts_data)

tsdisplay(ts_data, main = "Total Export")

## XTS Plot

require(xts)

plot(as.xts(ts_data), major.format = "YYYY", cex=0.6, main = "Total Export")

##data decomposition

##testing for stationery

adf.test(ts_data) #p-value less than 0.05 means data is stationary

kpss.test(ts_data)
```

##SES

```
ts_ses<- ses(ts_data, h=1)

ts_ses

ts_ses$model

ts.plot(ts_data, ts_ses$fitted, col= c("black", "red"), xlab = "Year", ylab = "Value", main = "SES_Forecast_Export")

metrics_ses<-forecast::accuracy(ts_ses)
```

```
round(metrics_ses,2)
```

##AR Model

```
##data decomposition

## detrending the data

dt_ts_data<-diff(ts_data, differences = 1)

plot(dt_ts_data)

## stationary test

adf.test(dt_ts_data)

acf(dt_ts_data, lag= 25, main ="Total Export")

pacf(dt_ts_data, lag = 25, main ="Total Export")

##fitting AR model

ar_mod <- arima(ts_data, order = c(3,2,0)) ## order = c(p= its AR based on pacf and acf plot, d = differencing term to make series stationary, q = Moving average part)

ar_mod

ar_fit <- cbind(ts_data,fitted(ar_mod))

ar_fit

ts.plot(ts_data, ar_fit, col= c("black", "red"),xlab = "Year", ylab = "Value", main = "AR_Forecast_Export")

## Check model adequacy by acf test

acf(ar_mod$residuals, lag =25)

Box.test(ar_mod$residuals, lag = 25, type = "Ljung-Box")## p>0.05 acepct null hypothesis

ar_forecast <- forecast(ar_mod, h= 5)

ar_forecast
```

```
plot(ar_forecast, xlab ="Year", ylab ="Value")

metric_ar <- forecast::accuracy(ar_forecast)

metric_ar
```

##Manual ARIMA model

```
arma_mod <- arima(ts_data, order = c(3,2,2)) ## order = c(p= its AR based on pacf and acf plot, d = differencing term to make series stationary, q = Moving average part)

arma_mod

ar_fit <- cbind(ts_data,fitted(arma_mod))

ar_fit

ts.plot(ts_data, ar_fit, col= c("black", "red"), xlab = "Year", ylab = "Value", main = "AR_Forecast_Export")

## Check model adequacy by acf test

acf(arma_mod$residuals, lag =25)

Box.test(arma_mod$residuals, lag = 25, type = "Ljung-Box")## p>0.05 acepct null hypothesis

arma_forecast <- forecast(arma_mod, h= 5)

arma_forecast

plot(arma_forecast)

metric_arma <- forecast::accuracy(ar_forecast)

metric_arma
```

##AUTOARIMA

```
autoarma_mod <- auto.arima(ts_data) ## order = c(p= its AR based on pacf and acf plot, d = differencing term to make series stationary, q = Moving average part)

autoarma_mod
```

```
autoar_fit <- cbind(ts_data,fitted(autoarma_mod))

autoar_fit

ts.plot(ts_data, autoar_fit, col= c("black", "red"))

## Check model adequacy by acf test

acf(autoarma_mod$residuals, lag =25)

Box.test(autoarma_mod$residuals, lag = 25, type = "Ljung-Box")## p>0.05 acepct null hypothesis

autoar_forecast <- forecast(autoarma_mod, h= 5)

autoar_forecast

plot(autoar_forecast, xlab = "Year", ylab= "Value")

metric_autoar <- forecast::accuracy(autoar_forecast)

metric_autoar
```

##ANN

```
nn_mod <- nnetar(ts_data,P=1, lambda = "auto", repeats = 30)

nn_mod

nn_fit <- forecast(nn_mod, h=5)

nn_fit

autoplot(nn_fit)

forecast::accuracy(nn_fit)
```

MLP Neural Network

```
library(nnfor)

set.seed(1234)
```

```
fit_mlp <- mlp(ts_data,hd= NULL, reps = 20, comb = c("median", "mean", "mode"), sel.lag = F, retrain = T)

print(fit_mlp)

fit_mlp1 <- mlp(ts_data,m = 1, hd= NULL, reps = 30, comb = c("median", "mean", "mode"), sel.lag = F, retrain = T, outplot =T)

print(fit_mlp1)

plot(fit_mlp)

frc_mlp <- forecast(fit_mlp, h=5)

frc_mlp

forecast::accuracy(frc_mlp)

plot(fit_mlp)

frc_mlp1 <- forecast(fit_mlp1, h=5)

frc_mlp

plot(frc_mlp)

predict(fit_mlp, ts_data_Test)
```

#TBATS

```
tbats_mod<- tbats(ts_data, use.box.cox = T, use.damped.trend = T, use.arma.errors = T)

summary(tbats_mod)

for_tbats <- forecast::forecast(tbats_mod, h = 5)

for_tbats

df_tbats = as.data.frame(for_tbats)

plot(for_tbats)

forecast::accuracy(for_tbats)
```