Global Crude Oil Forecasting by R programming

project report submitted in partial fulfilment of the requirement for the award of degree of

MASTER OF BUSINESS ADMINISTRATION
in
Port and Shipping Management

by

GORAKSH SONAWANE
Registration No: 1903304025

under the guidance of
Dr.A. MOUROUGANE
Associate Professor & Head

School of Maritime Management

INDIAN MARITIME UNIVERSITY

(A Central University, under the Ministry of Ports, Shipping and Waterways)

Chennai Campus, Chennai

July, 2021

DECLARATION

I **Goraksh Sadashiv Sonawane Registration No. 1903304025** student of the School of Maritime Management, Indian Maritime University, Chennai Campus pursuing **a Master of Business Administration Port and Shipping Management,** hereby declare that this report titled **"Global Crude Oil Forecasting by R programming"** has been prepared and submitted by me towards the partial fulfilment of the requirement for the award of degree of "**Master of Business Administration in Port and Shipping Management"** under the guidance of **Dr. A Mourougane,** Associate Professor & Head, School of Maritime Management, Indian Maritime University, Chennai Campus.

I also declare that this project report is my original work and has not been copied from any of the report previously submitted for the award of any Degree, Fellowship, or other in similar titles.

Place: Chennai **Goraksh Sonawane**

Date: 28th June 2021 Reg. No.-1903304025

Forwarded by: **Dr.A. MOUROUGANE**
Project Guide
Associate Professor & Head,
School of Maritime Management
Indian Maritime University
Chennai Campus.

ACKNOWLEDGEMENT

My heartfelt and sincere thanks to **Dr. A Mourougane,** Associate Professor & Head, School of Maritime Management, Indian Maritime University, Chennai Campus who gave me the golden opportunity to do this wonderful project on the topic **"A STUDY ON CRUDE OIL FORECASTING BY R PROGRAMMING"** which also helped me in doing a lot of research and I came to know about so many new things.

I pay him deep sense of gratitude for guiding me in each step of the project, alleviating inspiration, encouraging and kind supervision in the completion of my project in spite of his busy schedule.

And I am also thankful to faculty members, library staffs, my family members, my friends and my well-wishers who were very cooperative during my project in providing appropriate guidance and support without whom this project would not have been completed successfully.

GORAKSH SONAWANE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Sr. No.	Abbreviated Name	Full Name
1	WTI	West Texas Intermediate
2	ARIMA	Auto Regressive Integrated Moving Average
3	GARCH	Generalized Autoregressive Conditional Heteroskedasticity
4	CRAN	Comprehensive R Archive Network
5	ACF & PACF	Autocorrelation Function and Partial Autocorrelation Function
6	Ggplot	Grammar of Graphics plot
7	MAE	Mean Absolute Error
8	MSE	Mean Squared Error
9	RMSE	Root Mean Squared Error
10	MAPE	Mean Absolute percent Error
11	MASE	Mean Absolute Scaled Error
12	AIC	Akaike Information Criterion
13	BIC	Bayesian Information Criteria

EXECUTIVE SUMMARY

Data analysis and Forecasting is required in many situations: deciding whether to create another power generation plant within the next five years requires forecasts of future demand; scheduling staff during a call center next week requires forecasts of call volumes; stocking a listing requires forecasts of stock requirements. Crude oil has been playing an increasingly important role in the world economy since nearly two-thirds of the world's energy demands are met from crude oil.

Forecasting price of oil plays an important role in generating projections of energy use, for modeling investment decisions in the energy sector, in predicting carbon emissions and climate change, and in designing regulatory policies such as automotive fuel standards or gasoline taxes, Crude Oil forecast interest is due mainly to the economic value attached to the product. This is the reason forecasting through the ARIMA model by using R studio for the better result for forecasting.

R is a high-level language (Object-Oriented Language) and an environment for data analysis and graphics. Also, it comprises approx. 12000 packages in its library used for performing tasks on RStudio. RStudio is the software in which the R language will implement various statistical and graphical techniques, including linear and non -linear modeling, classical statistical tests, spatial and time-series analysis… etc.

Auto-Regressive Integrated Moving Average (ARIMA) models are an approach to time series data like GDP, Crude Oil data forecasting. The research shows how the ARIMA model is very appealing and likely to give better results than the other models included as part of this project, i.e., the Naïve Model and the Exponential Smoothening Model.

At last, the project focuses on the application of this R programming in forecasting the Crude Oil prices from the global price of WTI crude and as well as Import of Petroleum Product in India.

CHAPTER- I

1. INTRODUCTION

Forecasting petroleum prices is vital because it affects other key sectors of the economy, including the stock exchange. One of the important areas in economic research is forecasting the trend of price change of international petroleum.

Crude oil has been playing an increasingly important role within the world economy. It's also the world's largest and most actively traded commodity, accounting for over 10% of total world trade. But it also affects individuals too. This is often because the worth increases of petroleum effects on the worth of petrol, which has its attendant effect on goods and services produced within the country and by extension the gross domestic product (GDP).

Oil may be a vital input for the assembly of a good range of products and services because it's used for transportation in businesses of all kinds. Higher oil prices thus increase the value of imports; and final product price increases cause inflation. If the value increases can't be passed on to consumers, economic inputs such as labor and capital stock may be reallocated. Higher oil prices can affect the labor layoffs and the idling of plants, reducing economic output in the short term. In a net importer of oil economy like India, higher oil prices shrink foreign reserves of the economy and affect the economy's purchasing power in terms of international trade. The increased prices of imported oil force the businesses to change their thoughts more into exports, as opposed to satisfying domestic demand for goods and services, therefore cause inflation, even if there is no change in the quantity of foreign oil consumed Oil. There is, therefore, no uniformity or full agreement. Since it's a natural product forming a neighborhood of rocks, geological definition finds more general acceptance. Petroleum is also called "rock oil" or "crude oil". It is a generic term covering a wide range of substances comprising hydrocarbons, which are naturally occurring carbon and hydrogen molecules.

High oil prices have prompted increased investments in the Exploration and Production (E&P) sector, posing new challenges for the sector as the increased cost of operations because of high service costs, exposure to logistically difficult terrain, and lack of technical human resources. The global refining scenario shows very little to negligible addition in capacities in major developed consuming markets like the USA and the European countries. Developing countries just like the Middle East, China, and India are fast emerging as refining hubs. That capacity augmentation in these regions would also result in possible integration of both the refining and petrochemicals business.

India is the world's third-largest crude oil importer and imports 84% of its oil requirements. In February 2021, Iraq remained the highest oil supplier to India, followed by the US as refiners increased purchase of cheaper US crude by 48% (14% of overall imports), relegating Saudi Arabia to being the fourth-largest supplier offsetting OPEC supply cuts strategy.

Due to lack of adequate petroleum reserves, India has to depend mostly on crude oil imports shortly till its renewable energy resources such as solar, wind, hydro, and bio-mass are exploited adequately to realize energy security by replacing the petroleum products consumption which is additionally major contributors to the pollution. In such adverse situations, India has to proactively play a major role in the global crude oil trade as a swing oil produced by using its limited crude oil production base to bring lower the high price of crude oil fixed by OPEC and the Big Oil. International petroleum prices steeply for a little mismatch between global supply and global demand to become a swing oil producer, India should enhance petroleum extraction rate twice of the normally designed rate for continuous extraction from its developed oil fields and extract crude oil intermittently only when crude oil prices exceed a preset upper ceiling value instead of continuously extracting oil.

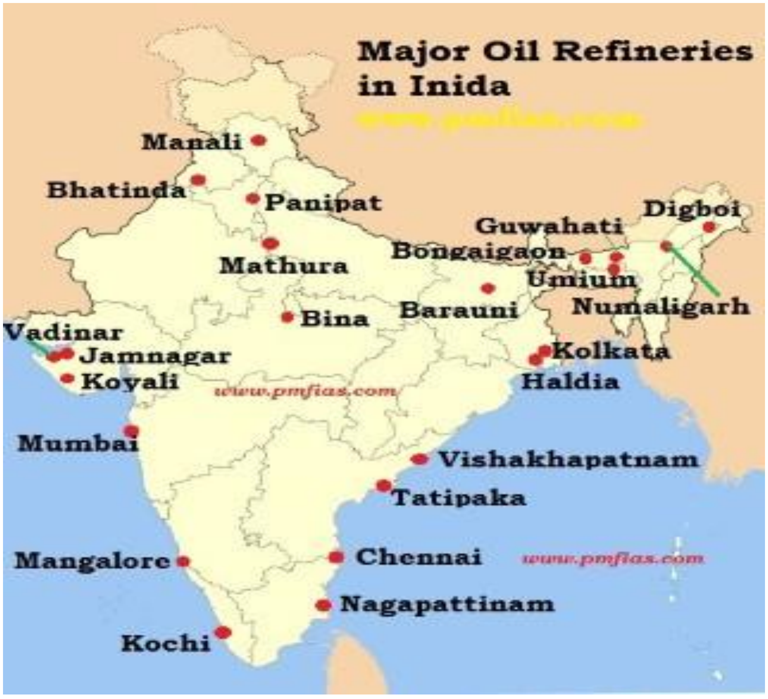


Figure 1.1: Oil Refineries in India

(Source: Oil Refineries in India, Wikipedia)

1.1 Importance of Study

Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge. R programming language is a part of data science where it is a free software environment for statistical computing and graphics. Once we learn the R programming and implement it into R studio software where the tasks are performed. The following procedure applies to achieve forecasting through the R language.

Data Extraction - To perform a task whichever raw data is available that has to be imported in RStudio cannot perform the task on that data that is not properly imported in the software. In this project report, we extracted data from WTI crude oil prices got from the Federal Reserve Bank of St. Louis.

Transformation - Certain data is non-stationary like crude oil prices, once we convert it into stationary data by performing certain iteration on the software and making it time series data, even though the data is to set properly by creating variables naming them properly for further task.

Visualization - Ggplot means Grammar of Graphics in R programming, which transforms the data more visualize. A good visualization will show the things which are not expected or raise new questions during the performing tasks.

Modeling–Model is a tool used after visualizing the data. It is a mathematical or computational tool, so they can scale well in the graph. Selecting the best model or comparing the models which will help in forecasting better results.

Forecasting - once I understand the particular model, the forecasting is done by using tool forecast ().

1.2 Literature Review

I made an attempt in this chapter to review some studies carried out to assess the importance of Crude Oil imports in India. The studies that allowed the researcher to undertake this study are as follows. Crude Oil prices forecasting by using ARIMA model and different models, its importance. Hence, here is a brief review of the studies on Crude Oil prices. Crude Oil prices forecasting has an important role in a nation's the trade and economy because of its significant contribution to the economic prosperity and well-being of any country.

1.2.1 Mohamed Reda Abonazel, Ahmed Ibrahim Abd-Elftah (2019) “Forecasting Egyptian GDP Using ARIMA Models”–

The aim of the study to model and forecast GDP by applying 4 stages of BOX–Jenkins’s approach to finding the appropriate ARIMA model. After getting a better model of ARIMA we understand the minimum values of AIC, BIC, and MSE help in forecasting. Forecasting GDP of EGYPT by using ARIMA MODEL, Measuring GDP purpose is to consider how well the economy is performing, use of statistical techniques in time series analysis to estimate and predict GDP.

1.2.2 S.L. Ho and M. XIE “THE USE OF ARIMA MODELS FOR RELIABILITY FORECASTING AND ANALYSIS,”–

The time-series approach is based on ARIMA models. Duane model and ARIMA model are compared. Duane model cannot detect this change and hence resulted in poor forecasts ARIMA model provides a better fit to the data with relatively low absolute error the ARIMA modeling approach can be a promising alternative for repairable system analysis solute.

1.2.3 Sima Siami-Namini, Neda Tavakoli, Akbar Siami Namin, A Comparison of ARIMA and LSTM in Forecasting Time Series,

The empirical studies conducted and reported in this article show that deep learning-based algorithms such as LSTM outperform traditional-based algorithms such as the ARIMA model, ARIMA is chosen because of the non-stationary property of the data collected and modeled. LSTM method is used for preserving and training features of given data for a longer period. Forecasting time series data is an important topic in economics, business, and finance. Conduct an empirical study and analysis to investigate the performance of traditional forecasting techniques and deep learning-based algorithms. Compare the performance of LSTM and ARIMA concerning minimization achieved in the error rates in prediction.

1.2.4 Rob J Hyndman and George Athanasopoulos Monash University, Australia, Forecasting: Principles and Practice

This textbook should provide a piece of vital information and a comprehensive introduction to forecasting methods and to present enough information about each method for readers to use them. This textbook helped to do forecasting. Various methods and iteration for performing Forecasting, Practical application of R, Detail information of how to perform forecasting.

1.2.5 Hassan Mohammadi and Lixian Su, International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models

Examine the usefulness of several ARIMA-GARCH models is used for modeling and forecasting the conditional mean and volatility of weekly crude oil spot prices in eleven international markets over the 1/2/1997–10/3/2009 period. This study uses four alternative classes of GARCH models to characterize the behavior of weekly oil returns and their conditional variance in eleven global markets comprising oil-exporting and oil-importing countries.

1.2.6 Rana Abdullah Ahmed and Ani Bin Shabri, Daily Crude Oil Price Forecasting Model using ARIMA, Generalized Autoregressive conditional heteroscedasticity and Support vector.

Forecasting crude oil prices is important as it affects other key sectors of the economy, including the stock market. If the price increase of crude oil effects on the price of petrol, which has its attendant effect on goods and services produced in the country and by extension the gross domestic product (GDP). The paper discusses the three techniques that feature prominently in this study. These are ARIMA, GARCH, and SVM. This study adopted two very popular measures for evaluating the forecast accuracy of the series and these are: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

1.2.7 Nazuha Muda & Yusof, Ruzaidah Sulong & A. Rashid Zamzulani Mohamed, Malaysia Crude Oil Production Estimation: An Application of ARIMA Model

The paper explains the ARIMA model and then develops an ARIMA model for crude oil production using monthly data in Malaysia from January 2005 to May 2010. Autocorrelation and partial autocorrelation functions were calculated to examine the stationary of the data. Four primary stages in building a Box Jenkins time series model: a) model, b) Identification, c) model estimation, and d) model validation. Box-Jenkins uses a statistical procedure to identify a model.

1.2.8 Chun-lan Zhao and Bing Wang, Forecasting Crude Oil Price with an Autoregressive Integrated Moving Average (ARIMA) Model

Forecast of the price of oil, play a role in generating projections of energy use, in modeling investment decisions in the energy sector, in predicting carbon emissions and climate change, and in designing regulatory policies such as automotive fuel standards or gasoline taxes, An autoregressive integrated moving average (ARIMA) model generalization of (ARMA) model. We fit these models to time-series data either to better understand the data or to predict future points in the series (forecasting).

1.2.9 Niaz. Bashiri Behmiri, José R. Pires Manso, Crude oil price forecasting techniques: a comprehensive review of literature

The goal of this article is to review the existing literature on crude oil price forecasting. They categorized the existing forecasting techniques into the two major groups of quantitative and qualitative methods. Quantitative methods include econometrics and computational approaches Qualitative methods contain computational and technological Quantitative methods are grouped into Econometrics methods, including (a) time series models (b) financial models, and (c) structural models, and these quantitative methods apply to model the numerical determinants of oil prices Non-standard or computational approaches.

1.2.10 Michael J. Crawley Imperial College London at Silwood Park, UK, The R Book, This edition first published 2013 C 2013 John Wiley & Sons, Ltd

This book is for data science, statistics, Data Analyst… Etc. Complete R information provided. To learn R programming, this book is the latest version to understand each iteration and objectives and how to use it.

1.3 Objective of Study

The primary aim of this article is to investigate which forecasting methods offer the best predictions concerning lower forecast errors and higher accuracy of forecasts.

1. R. programming is used for forecasting the experimental sample with different models to forecast the present data of 10yrs latest of Global Price of West Texas Intermediate (WTI) crude oil spot price (in US dollars per barrel) series
2. Plot the forecast data for visualizing future patterns

1.4 Research Methodology

The study involves using the R programming language to identify the best model for forecasting the present data of Global Prices of WTI Crude Oil and another data of Import of Petroleum products in India.

I collected the data from

I. Global Prices WTI crude retrieved from FRED, Federal Reserve Bank of St. Louis

II. Petroleum Product Import in India by Department of Commerce, Ministry of Commerce and Industry, Government of India

III. R programming from the R book and website of R studio

IV. Various Research Papers about Crude oil Prices forecasting and use of the ARIMA model

V. Online such as YouTube for further learning

1.5 Scope of Study

Crude Oil is the most sensitive commodity and Forecast of the price of oil, play a role in generating projections of energy use, in modeling investment decisions in the energy sector, in predicting carbon emissions and global climate change, and in designing regulatory policies like automotive fuel standards or gasoline taxes, Now the question arises why we want to forecast petroleum prices by machine learning?

The reason behind this is often its price is that the most notorious for having the ability to travel in the exact opposite direction after one market event.

The present data rarely supported fundamental oil prices or we are saying real-time data, instead, it's driven by externalities making our attempt for forecasting very challenging.

Why should we care about oil prices? The reason is simple to create a model because a slight deviation in prices can cause a tremendous impact on the economy or the health of the economy is linked.

The aforementioned reasons predict oil prices as an imperative task to decrease the impact of price fluctuations and assist policymakers and individuals to form informed decisions that might help in dealing with price fluctuations arising from the energy markets

As a result, achieving reliable and highly accurate forecasting and answering the complexities of crude oil prices would be important to policymakers. For this purpose, I have tried various techniques to forecast the movements, fluctuations, or volatility of crude oil prices.

1.6 Limitations of the study

Crude oil price is one of the foremost complexes and difficult to model because the fluctuation of the petroleum price is sort of irregular, nonlinear, nonstationary, and with high volatility. Thus, accurate forecasting of the petroleum price statistic is one of the simplest challenges and among the foremost important issues facing energy economists towards better decisions at several managerial levels. For this reason, many researchers have devoted considerable effort to the event of varied kinds of models for petroleum price forecasting.

According to this report, the application of the Auto-Regressive Integrated Moving Average (ARIMA) model is that the classical statistic model, also as

Naïve and Exponential Smoothening Model (ETS) for petroleum price forecasting, has received much attention within the last decade. However, the above models can provide good prediction results, when the worth series under study is essentially linear or near-linear, and have a limited ability to capture nonlinearity and nonstationary oil prices data.

Artificial neural network (ANN) techniques have shown exceptional ability in modeling and forecasting nonlinear and sophisticated statistics. ANN offers an efficient approach for handling large amounts of dynamic, nonlinear, and noise data. Many papers have already presented a successful application of ANN for modeling and forecasting the crude oil price series.

So we won't be able to get the perfect forecasting which we are looking forward to but can have a better forecasting result as compared to other models used in among these three models.

CHAPTER – II

2. Introduction to R and Forecasting

R, the open-source data analysis environment and programming language allow users to conduct several tasks that are essential for the effective processing and analysis of big data. One advantage of R is that it helps in facilitating data management processes such as transformations, sub-setting, and cleaning, and helps users carrying out exploratory data analysis and prepare the data for statistical testing. R also contains many ready-to-use machine learning and statistical modeling algorithms which allow users to create reproducible research and develop data products.

2.1 So why R?

As there is the availability of alternative tools such as Python being one example, R has the advantage of being the only software open-source programming language that has been built specifically for statistical analysis. It contains its inbuilt statistical algorithms–the sheer number of mathematical models and machine learning algorithms available to users in base R and third-party packages is staggering and continues to grow—reaching almost 12,000 add-on libraries on the Comprehensive R Archive Network (CRAN).

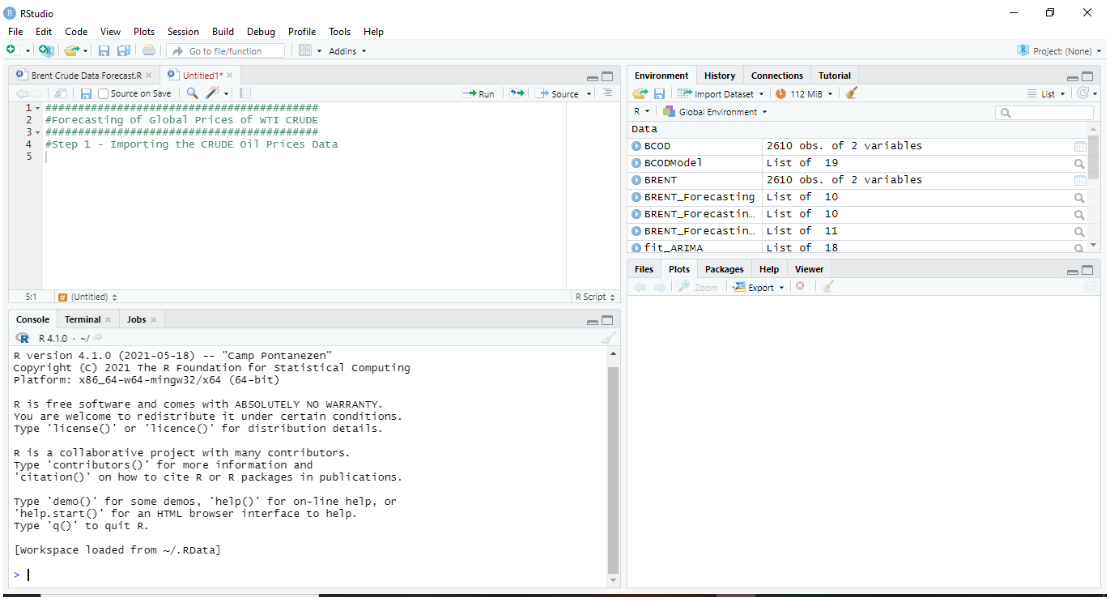


Fig.2.1. Overview of Rstudio Software.

2.2 Forecasting

Forecasting is a common statistical task in business. It helps to provide information regarding decisions to be taken for scheduling of production unit, transportation, and personnel and provides a guide to long-term strategic planning. However, business forecasting is usually done poorly and is usually confused with planning and goals.

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organizations require short-term, medium-term, and long-term forecasts, counting on the precise application.

2.2.1 Short-term forecasts

We need it to schedule personnel, production, and transportation. As a part of the scheduling process, forecasts of demand are often also required.

2.2.2 Medium-term forecasts

It helps to work out future resource requirements, purchase raw materials, hire personnel, or buy machinery and equipment.

2.2.3 Long-term forecasts

Are used in strategic planning. Such decisions must appreciate market opportunities, environmental factors, and internal resources.

An organization must develop a forecasting system that involves several approaches to predicting uncertain events. Such forecasting systems require the event of experience in identifying forecasting problems, applying a variety of forecasting methods, selecting appropriate methods for every problem, and evaluating and refining forecasting methods. It is also important to possess strong organizational support to use formal forecasting methods if they're to be used successfully.

2.3 Method For Forecasting

Forecast methods could also be broadly classified into qualitative and quantitative techniques. Qualitative methods are not purely guesswork—there are well-developed structured approaches to getting good forecasts without using historical data. Quantitative methods use mathematical or statistical models to get an inexpensive prediction from the knowledge of the past. Compared to qualitative methods, quantitative methods have the advantage of being supported by mathematical and statistical theory and maybe fully reproduced by any forecaster.

Each method has its own properties, accuracies, which must be considered while choosing a specific method for particular data.

Most quantitative prediction problems use either time-series data (collected at regular intervals over time) or cross-sectional data (collected in time). In this book, we are concerned with forecasting future data, and that we consider the time-series domain.

Forecasting, especially statistic forecasting, a primary sort of data in business and economics, the quantitative methods are widely applied. A time series is a set of observations $\{y_t: t = 1,2,3,4,\cdots,T\}$. Usually, time series is considered as discrete series in which observations are recorded at predetermined, equal-interval time point such as hourly, daily, monthly, quarterly, or yearly.

2.4 The Basic Steps for Forecasting in Rstudio

A forecasting task usually involves five basic steps.

Step 1: Problem definition.

Often this is the foremost tough part of forecasting. Defining the matter carefully requires an understanding of the way the forecasts are going to be used, who requires the forecasts, and the way the forecasting function fits within the organization requiring the forecasts. A forecaster must collect data, maintaining databases, and using the forecasts for future planning.

Step 2: Gathering information.

There is always a minimum of two sorts of information required:

(a) Statistical data, and

(b) The accumulated expertise of the people that collect the info and use the forecasts.

Often, it'll 'e difficult to get enough historical data to be ready to fit an honest statistical model. In that case, judgmental forecasting methods will be used. Occasionally, old data are going to be less useful thanks to structural changes within the system being forecast; then we may prefer to use only the foremost recent data. However, we remember that a good statistical model will handle evolutionary changes in the system.

Step 3: Preliminary analysis.

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is there seasonality in data? Is there evidence of the presence of economic cycles? Are there any outliers within the data that require to be explained by those with expert knowledge? How strong variables relationship is available for analysis? They develop various tools to assist with this analysis.

Step 4: Choosing and fitting models.

The best model to use depends on the availability of past data, the strength of relationships between the forecast variable and any explanatory variables, and how the forecasts are to be used. It is common to match two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters that must be estimated using the known historical data. We will discuss regression models, exponential smoothing methods, Box-Jenkins ARIMA models, Dynamic regression models, Hierarchical forecasting, and several advanced methods, including neural networks and vector auto-regression.

Step 5: Using and evaluating forecasting model.

Once a model has been selected and its parameters estimated, they employed the model to form forecasts. The performance of the model can only be properly evaluated after the info for the forecast period became available. Several methods have been developed to help in assessing the accuracy of forecasts.

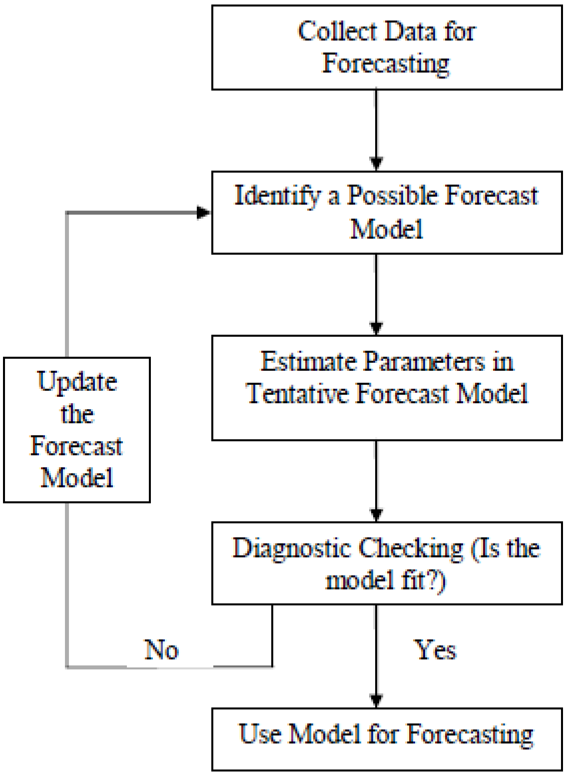


Fig 2.2. Forecast Approach.

2.5 Conclusion

Forecasting with R programming requires certain steps and models, various iterations. As R is the free open software and it has various libraries and tools for performing forecasting.

We have got the knowledge about the forecasting required for the statistical analysis of the databases to plan according to for the future aspect in businesses and it is essential for scheduling production, transportation and for short term and long-term strategic goals to be estimated.

Forecasts types are short term, medium-term, and long-term as per the requirement for modern businesses by selecting and applying appropriate methods.

Forecasting a requires step-by-step approach to implement in R studio software, by problem definition, gathering data, preliminary analysis, diagnosing fitted models, using and evaluating forecast models.

CHAPTER – III

3 Forecasting Techniques

Quantitative approach includes, ARIMA model, Exponential smoothing and Naive models are the three most widely used approaches to time series forecasting, and provide complementary approaches to the problem. While it based exponential smoothing models on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

3.1 ARIMA Model

An autoregressive integrated moving average (ARIMA) model generalization of an autoregressive moving average (ARMA) model. It fit these models to time-series data either to better understand the data or two forecast points in the series. They are applied sometimes where data shows possible being on stationary, where an initial differencing step i.e., Integrated can remove the non-stationary.

We refer the model to as an ARIMA (p, d, q) model where p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modeling. Given a time series of data $\boldsymbol{x_t}$ where t is an integer index and the $\boldsymbol{x_t}$ are real

numbers, then an ARMA (p, q) model is given by:

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) x_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

where L is the lag operator, the $\boldsymbol{\alpha_i}$ are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the $\boldsymbol{\varepsilon_t}$ are error terms. The error terms $\boldsymbol{\varepsilon_t}$ are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

An ARIMA (p, d, q) process expresses this polynomial factorization property, and is given by:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

And thus, can be thought as a particular case of an ARMA (p + d, q) process having the auto-regressive polynomial with some roots in the unity. For this reason, every ARIMA model with d > 0 is not wide sense stationary.

3.1.1 ARIMA modelling in R

ARIMA modelling in R is denoted by function called **auto. arima (),** the function is in R uses a variation of the Hyndman-Khandakar algorithm (Hyndman & Khandakar, 2008), which combines unit root tests, minimisation of the AICc and MLE to get an ARIMA model. The arguments to **auto. Arima ()** provides for many variations on the algorithm.

3.1.2 The four stages modeling in the Box-Jenkins iterative approach:

a) **Model identification**:

Identification involves looking at the graph of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) to determine whether the series is stationary and then decide on which functional form best fits an appropriate model for the data. The ACF and PACF are random variables and cannot give an equivalent picture because of the theoretical functions. This makes the model identification harder and may involve much trial and error. The most common method to check stationery is by examining the time plot of the data. Time series can meet this condition, but as long as the data can be transformed into a stationary series, a Box-Jenkins model can be developed. If the time series data was non stationary, we need to differentiate the data by simply subtracting each value of the data with its previous data to remove the variation to achieve stationary in the mean.

b) **Model Estimation** - Once a suitable ARIMA (p, d, q) structure is identified, the second step is the parameter estimation or fitting stage. Most of the identified ARIMA models are nonlinear, requiring a nonlinear estimation procedure. There are many other estimation techniques and the main approaches to fitting Box-Jenkins models are nonlinear fewest squares and the very popular maximum likelihood estimation. Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC) are two goodness-of-fit statistics that are most commonly used for model selection.

The least squares approach is the most common technique, whereas the maximum likelihood estimation is the preferred technique. The ARIMA (1, 0, 0) model can be written as the equation below:

$$y_t = \delta + a_1 \cdot y_{t-1} + u_t$$

where δ is a constant and u_t as the white noise

c) **Model Validation:** Once the model is estimated, the next step is to assess how well the model fits the data. A good way to check the adequacy of an overall Box-Jenkins model is to analyse the residuals gets from the model. Therefore, a plot of the auto correlogram should immediately die out from one lag on.

d) **Forecasting**: when the selected ARIMA model conforms to the specifications of a stationary inferred process, then we can use this model for forecasting.

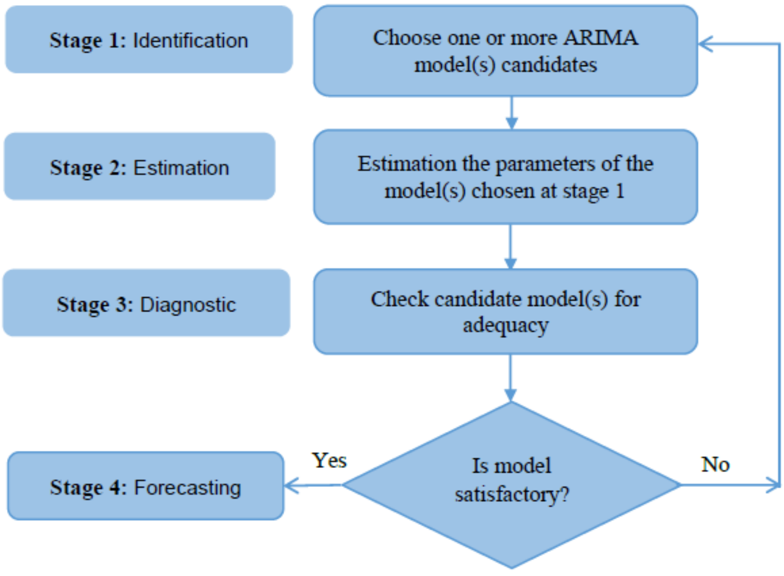


Fig 3.1. ARIMA Model Approach or Box–Jenkins Iterative approach.

3.2 Exponential smoothing

A simple exponential smoothing is one of the few ways to forecast a statistic. This model assumes that the future will be more or less the same as the (recent) past. The only pattern that this model is going to be ready to learn from demand history is its level.

The level shown is that the average value around which the demand varies. We can observe in the figure below that the level is a smoothed version of the demand.

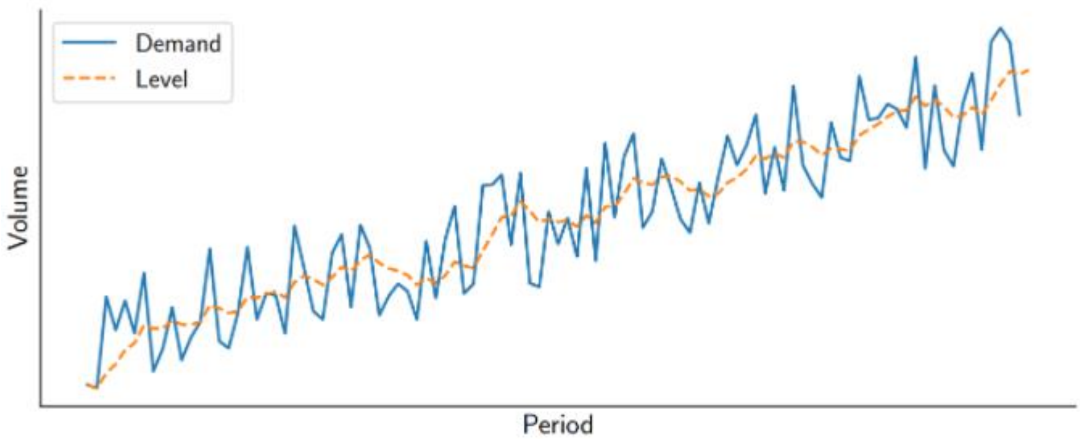


Fig 3.2. Demand Level.

The exponential smoothing model will then forecast the future demand level is estimated. It is essential to know that there's no definitive mathematical definition of the extent. Instead, it's up to our model to estimate it.

The exponential smoothing model has some advantages compared to a simpler forecast model (such as a naïve or a moving average):

• The weight that is put on each observation decreases exponentially. This is often better than moving average models where an equivalent weight is given to all or any of the relevant historical months.

• Outliers and noise have less impact than with the naïve method.

3.2.1 ETS Model,

The underlying idea of an exponential smoothing model is that, at each period, the model will learn from the recent demand observation and remember a bit of the last forecast it did. The magic about this is that the last forecast the model did was including a part of the previous demand observation and a part of the previous forecast. And so forth. That means that this previous forecast includes everything the model learned, supported demand history. The smoothing parameter (or learning rate) alpha will determine what proportion of importance is given to the foremost recent demand observation. Let's represent this mathematically.

$$f_t = \alpha d_{t-1} + (1-\alpha) f_{t-1}$$

$$0 < \alpha \leq 1$$

3.2.2 Intuition behind this formula

alpha is a ratio (or a percentage) of how much importance the model will allocate to the most recent observation compared to the importance of demand history.

• αd_{t-1} represents the previous demand observation times the learning rate. You could say that the model attaches a particular weight (alpha) to the last demand occurrence.

• $(1-\alpha) f_{t-1}$ represents how much the model remembers from its previous forecast. Note that this is often where the recursive magic happens asf_{t-1} was itself defined as partially d_{t-2} and f_{t-2}.

3.2.3 Limitations

This simple exponential smoothing model is slightly smarter than the moving average model because of its smarter weighting of the historical demand observation.

But it's many limitations:

- It doesn't project trends.
- It doesn't recognize any seasonal pattern.

In conclusion, this first exponential smoothing model is going to be presumably too simple to realize excellent results, but it's a honest foundation block to make more complex models later.

3.3 Naïve Method

The simplest forecasting method is to use the most recent observation as the forecast for the next observation. This is called a naïve forecast and may be implemented using the 'naïve ()' functio'. This method might not be the simplest forecasting technique, but it often provides a useful benchmark for other, more advanced forecasting methods.

Forecasts produced employing a naïve approach are adequate for the ultimate observed value. The naïve forecasting method easy to use as compared to other models, and it is familiar with the finance and sales department because it helps to ensure that these departments should work well for the growth of the company.

For naïve forecasts, we simply set all forecasts to be the worth of the last observation.

$$\hat{y}_{T+h|T} = y_T$$

This method works remarkably well for many economic and financial time series.

Function of the naïve model denoted as - naïve (y, h)

y = contains time series

h = forecast horizon

3.3.1 Seasonal naïve method

A similar method is beneficial for highly seasonal data. Here, we set each forecast to be adequate to the last observed value from an equivalent season of the year (e.g., an equivalent month of the previous year). so, the forecast for time T+h is written as

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$$

Where m= the seasonal period and k is the integer part of (h−1)/m (i.e., the number of complete years in the forecast period before time T+h). This looks more complicated than it is. For example, with monthly data, the forecast for all future September values is adequate to the last observed September value. With quarterly data, the forecast of all future Q3 values is equal to the last observed Q3 value (where Q3 means the third quarter).

We denote the function of the seasonal naïve method as–Snaive (y, h)

3.4 Conclusion

Forecasting in Rstudio requires a function models to run the forecast on the particular data which has to be forecast and the model which included for forecasting plays a vital role in providing a better forecasting comparison that which model suits better for forecasting, the model included in the report are ARIMA, Exponential Smoothening Model and Naïve model, each has its particular role in performing task, we have to find out which gives the better result.

CHAPTER – IV

4. CRUDE OIL PRICE FORECASTING

In the previous chapters we have learned briefly about the various models which we are going to apply for forecasting. The aim of this project is to compare between the forecasting techniques provided via the forecast package in R.

These models include, ARIMA, ETS and Naïve model, given that crude oil price fluctuations remain a widely researched topic, this research topic can help researchers determine which of the models provided via the forecast package are most suitable for this purpose, and rule out the models which are unlikely to be useful, and save time.

In order to achieve the aim of this paper, the yearly WTI spot price data, which spans from 5th May 2011 to 5th May 2021, got from FRED, Federal Reserve Bank of St. Louis. To attain the forecasting by R studio step-by-step approach is necessary, after which we can apply the techniques.

4.1 Data Import

Before starting the algorithm, the most important step required is to import the data from the sources where we can find the data which is valid to attempt the forecasting. In this paper we have included data on the yearly Global Price of West Texas Intermediate (WTI) crude oil price (in US dollars per barrel).

Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)
2011-05-01	101.25	2014-01-01	95.00	2016-09-01	45.17	2019-05-01	60.73
2011-06-01	96.25	2014-02-01	100.70	2016-10-01	49.89	2019-06-01	54.68
2011-07-01	97.19	2014-03-01	100.57	2016-11-01	45.62	2019-07-01	57.51
2011-08-01	86.32	2014-04-01	102.18	2016-12-01	52.01	2019-08-01	54.84
2011-09-01	85.62	2014-05-01	102.00	2017-01-01	52.56	2019-09-01	56.86
2011-10-01	86.41	2014-06-01	105.24	2017-02-01	53.40	2019-10-01	53.98
2011-11-01	97.07	2014-07-01	102.99	2017-03-01	49.58	2019-11-01	57.11
2011-12-01	98.61	2014-08-01	96.38	2017-04-01	51.17	2019-12-01	59.86

Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)	Observation Date	Actual Price (US Dollars per Barrel)
2012-01-01	100.15	2014-09-01	93.35	2017-05-01	48.56	2020-01-01	57.71
2012-02-01	102.26	2014-10-01	84.40	2017-06-01	45.17	2020-02-01	50.60
2012-03-01	106.15	2014-11-01	75.70	2017-07-01	46.67	2020-03-01	29.88
2012-04-01	103.28	2014-12-01	59.10	2017-08-01	48.03	2020-04-01	16.81
2012-05-01	94.51	2015-01-01	47.60	2017-09-01	49.71	2020-05-01	28.79
2012-06-01	82.36	2015-02-01	50.72	2017-10-01	51.56	2020-06-01	38.30
2012-07-01	87.89	2015-03-01	47.78	2017-11-01	56.71	2020-07-01	40.75
2012-08-01	94.11	2015-04-01	54.20	2017-12-01	57.96	2020-08-01	42.36
2012-09-01	94.61	2015-05-01	59.26	2018-01-01	63.56	2020-09-01	39.61
2012-10-01	89.52	2015-06-01	59.80	2018-02-01	62.15	2020-10-01	39.53
2012-11-01	86.69	2015-07-01	51.16	2018-03-01	62.86	2020-11-01	41.52
2012-12-01	88.19	2015-08-01	42.86	2018-04-01	66.32	2020-12-01	47.09
2013-01-01	94.65	2015-09-01	45.48	2018-05-01	69.89	2021-01-01	51.94
2013-02-01	95.30	2015-10-01	46.20	2018-06-01	67.52	2021-02-01	59.08
2013-03-01	93.12	2015-11-01	42.65	2018-07-01	70.99	2021-03-01	62.35
2013-04-01	92.02	2015-12-01	37.24	2018-08-01	67.99	2021-04-01	61.70
2013-05-01	94.72	2016-01-01	31.70	2018-09-01	70.19	2021-05-01	65.23
2013-06-01	95.79	2016-02-01	30.35	2018-10-01	70.75		
2013-07-01	104.55	2016-03-01	37.77	2018-11-01	56.57		
2013-08-01	106.55	2016-04-01	40.96	2018-12-01	48.64		
2013-09-01	106.31	2016-05-01	46.85	2019-01-01	51.36		
2013-10-01	100.50	2016-06-01	48.75	2019-02-01	54.99		
2013-11-01	93.81	2016-07-01	44.89	2019-03-01	58.15		
2013-12-01	97.90	2016-08-01	44.75	2019-04-01	63.88		

Table.1 Global Crude WTI Price Data from May 2011 to May 2021

(Source - https://fred.stlouisfed.org/series/POILWTIUSDM)

4.2 How to Import Data in Rstudio

Once the required sample of data is saved in excel sheet, it is then imported by executing a certain code i.e.

Simple coding is implemented in which the Rstudio runs a library (read xl) so that it can read the data from a excel sheet, as u can see in the below figure Excel sheet name is Global Crude Oil Price Data which is being read by software and from that the data is extracted from the sheet sample data, in which the column is read as date and numeric to understand by the software properly.

View (Crude) is the function so that we can view that data in the software in the top left side section and the right-side section shows the Number of observations which is 121 in our data and variables that are 2 variables.

Input

```
CRUDE <- read_excel("Global Crude Oil Price Data.xls", sheet = "Sample Data",
col_types = c("date", "numeric"))
View (CRUDE)
```

Fig. 4.1

Output

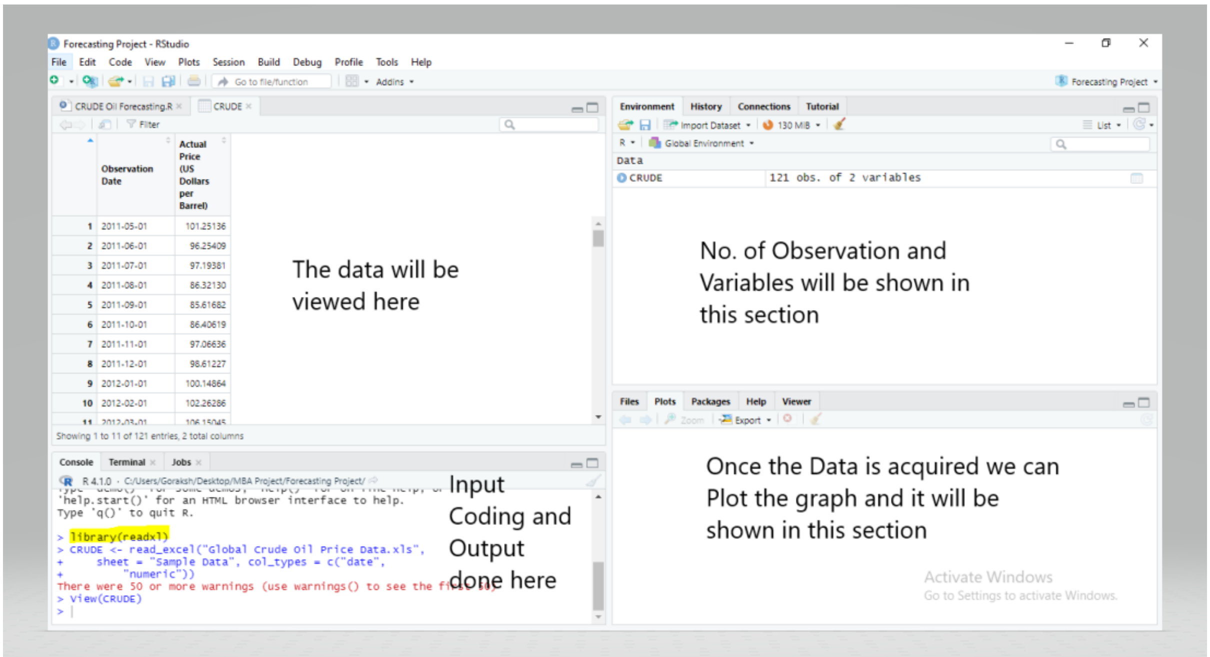


Fig 4.2. Data Extraction in Rstudio software

Rstudio software is easy to use once we learn the certain attributes, from the above figure.8 we can see there are four sections, each section has its importance.

Section classifieds we show the input and output of whichever code we implement or perform any task, the result in a particular section.

1. **Top Left side section** - the section which has two functions, it can be used as a note for the coding which we are going to write and we can give it as input once the coding is correct. (Note–the same coding can be written in the bottom left side but it once written cannot be erased) in this section we can erase and change the coding before implementing and it can be saved for the further studies, second, we can view the data n this section once it is imported and read by console.

2. **Top Right section–**This section comprises four options out of which only two are important for project point of view, first, we can import the data from this section by clicking on the import tab so that whichever data it is as Excel or CSV i.e., comma-separated values data. Second, it helps in viewing the data values form means we can see how many observations are there in the data and how many values, values we can consider in the excel sheet as the columns.

3. **Bottom Left Section–**This section is the most important because it is called as R console and this place where each code is running. This is where we can perform tasks and which gives us the output, to keep the track on the coding and perform with no mistake the codes are written on the top left side so that there should not receive any error while performing coding.

4. **Bottom Right Section–**The importance of this section is that we get to see the forecasting graph, any graph whether the actual data or whether the future data and various other graph which we plot in the R console is seen in this section.

4.3 Visualizing Data

Data visualisation is done once the data is dispersed or the graph is not properly seen. For example, the class of the data is Tibble (tbl) when we plot the data without visualisation, it is shown as fig.9.

Below.

Input

```
Step 2 - Check the class of data
class(CRUDE)
plot(CRUDE)
```

Fig. 4.3. Checking Class of Data

Output

```
class(CRUDE)
[1] "tbl_df"    "tbl"       "data.frame"
> plot(CRUDE)
```

Fig. 4.4 Output class of data

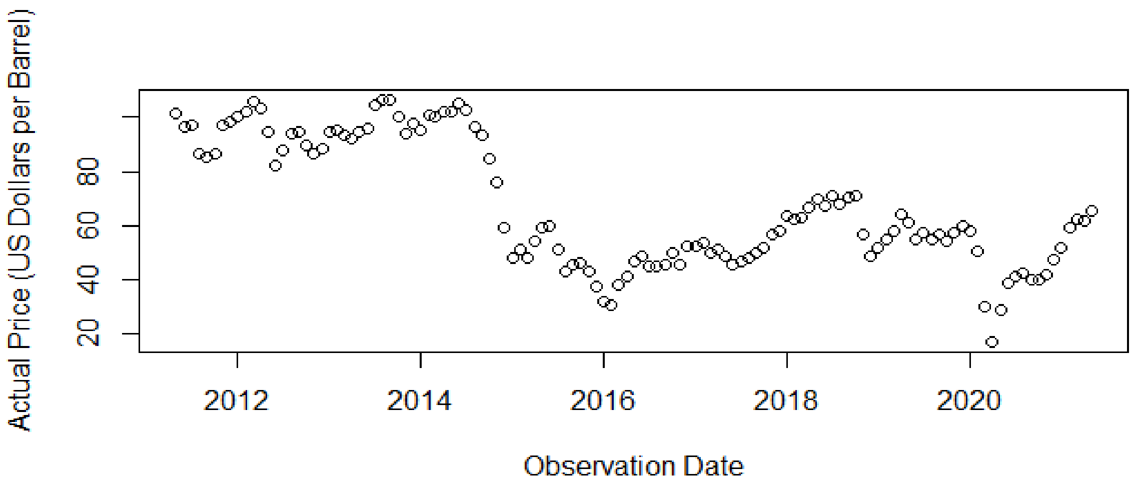


Fig 4.5. Disperse Data before visualization

Visualizing the Data by applying Ggplot function, which plays a vital role to pot the data in visible format which can be understood properly.

4.3.1 Ggplot for Visualizing

Grammar of Graphics (ggplot). We denote it as "ggplot ()" initializes a ggplot object. It can declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden.

The ggplot is a dedicated package in the R library for data visualization. It is very helpful to improve the quality and aesthetics of graphics and make it more efficient to view any graph which will be plotted in the console.

ggplot (data = NULL, mapping = aes (),…, environment = parent. Frame ())

Before applying ggplot we have to convert our columns or two variables into their respective forms that is dated in Date format and Price into Integer by using as. Date () and as.Integer ().

Ggplot comprises data, and aesthetic mapping of both variables i.e., aes (Date, Price), geometric line on which the graph is plotted, scale-x-date which denotes in a graph as x axis value and Ylab for y axis denoted.

Input

```
#Aesthetic Mapping
CRUDE$DATE= as.Date(CRUDE$`Observation Date`)
CRUDE$PRICE = as.numeric(CRUDE$`Actual Price (US Dollars per Barrel)`)
#Applying GGPLOT
ggplot(CRUDE,aes(DATE,PRICE)) + geom_line() + scale_x_date("Time Period")   +
ylab("Price US Dollar(per barrell)") + xlab("")
summary(CRUDE)
```

Fig. 4.6 Input of aesthetic mapping

Output

```
> summary(CRUDE)
 Observation Date      Actual Price (US Dollars per Barrel)      DATE                VALUE
 Min.   :2011-05-01   Min.   : 16.00                          Min.   :2011-05-01   Min.   : 16.00
 1st Qu.:2013-11-01   1st Qu.: 48.00                          1st Qu.:2013-11-01   1st Qu.: 48.00
 Median :2016-05-01   Median : 59.00                          Median :2016-05-01   Median : 59.00
 Mean   :2016-05-01   Mean   : 66.59                          Mean   :2016-05-01   Mean   : 66.59
 3rd Qu.:2018-11-01   3rd Qu.: 93.00                          3rd Qu.:2018-11-01   3rd Qu.: 93.00
 Max.   :2021-05-01   Max.   :106.00                          Max.   :2021-05-01   Max.   :106.00
     PRICE
 Min.   : 16.00
 1st Qu.: 48.00
 Median : 59.00
 Mean   : 66.59
 3rd Qu.: 93.00
 Max.   :106.00
```

Fig 4.7 Summary of CRUDE Data

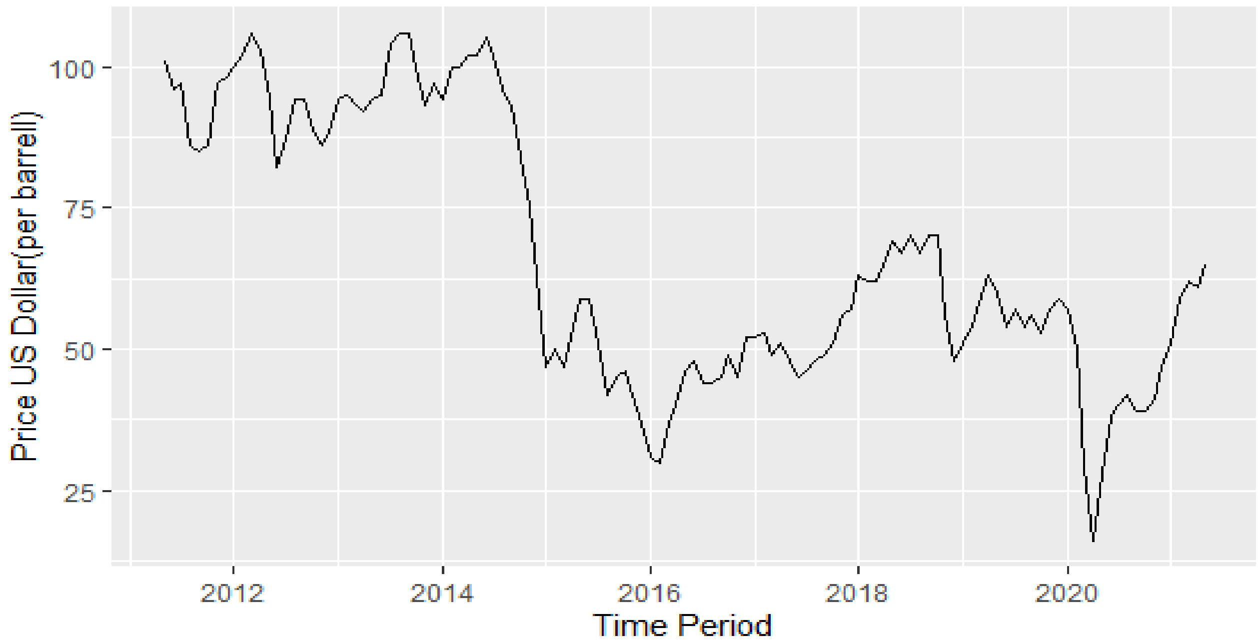


Fig 4.8 - Crude data graph after ggplot

Once the ggplot function is applied, we can see the graph in the fig (4.3.1).3. the graph we can see the visuals of the data which will help further and summary of the data shows the minimum and maximum value and mean medium of the data.

From the fig we can see the minimum price of the crude oil i.e., $16 per barrel and maximum price of $106 per barrel. That helps us to show that our data is valid and we can now convert it to the time series and further steps for forecasting.

4.4 To convert Data into Time Series

First thing for data analyst is to plot the data, graphs enable features to visualize, patterns and unusual observations, changes over the period and the variables relationships.so the data which is incorporated maximum as possible for methods to be used, but to do so we need our data to be based on time series.

Past and present tend decides future tends of any economic cycle, or the data which we want to forecast.

To run the forecasting models in ‘R’, we need to convert the data into a time series object which is done in the first line of code below. The ‘start’ and ‘end’ argument specify the time of the first and the last observation, respectively. The argument ‘frequency’ specifies the number of observations per unit of time.

Input

```
CRUDE_ts = ts(CRUDE[,2], start = c(2011,1), end = c(2021,1),frequency = 12)
plot(CRUDE_ts)
```

Fig. 4.9 Converting into Time series Data

Output

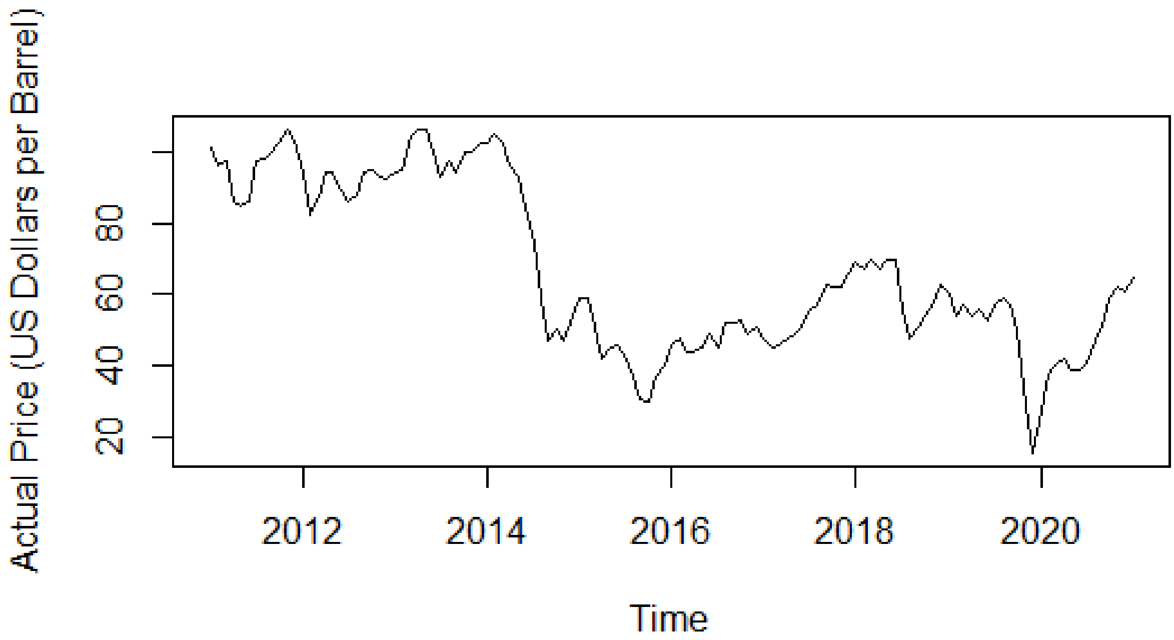


Fig. 4.10 Time series CRUDE-ts data

4.5 Differentiating to remove Trend

Plotting time series data is a crucial initiative for analyzing their various components. Beyond that, however, we'd like a more formal means for identifying and removing characteristics like a trend or seasonal variation. So, the method is using differentiating to remove the trend in the time series data.

Input

```
CRUDE_Trend = diff(CRUDE_ts)
plot(CRUDE_Trend)
```

Fig.4.11 Differentiating to remove Trend

Output

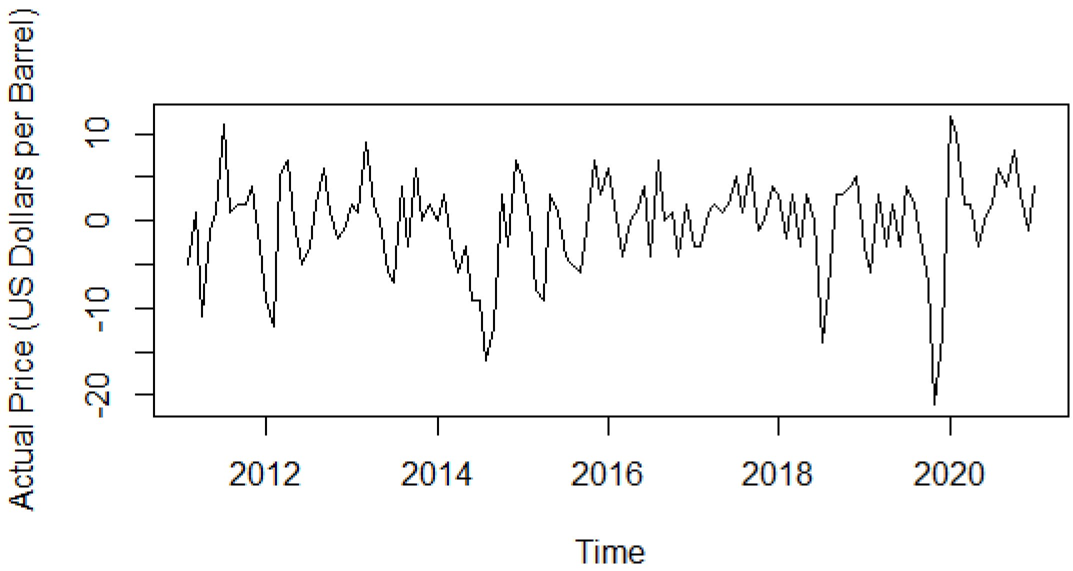


Fig. 4.12 CRUDE_Trend Data after differencing

4.6 To check Seasonality of Data

Along with time plots, there are other useful ways of plotting data to stress seasonal patterns and show changes in these patterns.

A seasonal plot is analogous to a time plot except that we plot the info against the individual "seasons" during which the info was observed. You can create one using the ggseasonplot () function an equivalent way you are doing with autoplot (). An interesting variant of a season plot uses polar coordinates, where the time axis is circular instead of horizontal; to form one, simply add a polar argument and set it to TRUE. A subseries plot comprises mini time plots for every season. Here, the mean for every season is shown as a blue horizontal line.

Input

```
ggseasonplot(CRUDE_Trend) + ggtitle("CRUDE Oil Price seasonality ") +
ylab("Yearly Demand")
```

Fig. 4.13 Data Seasonality

Output

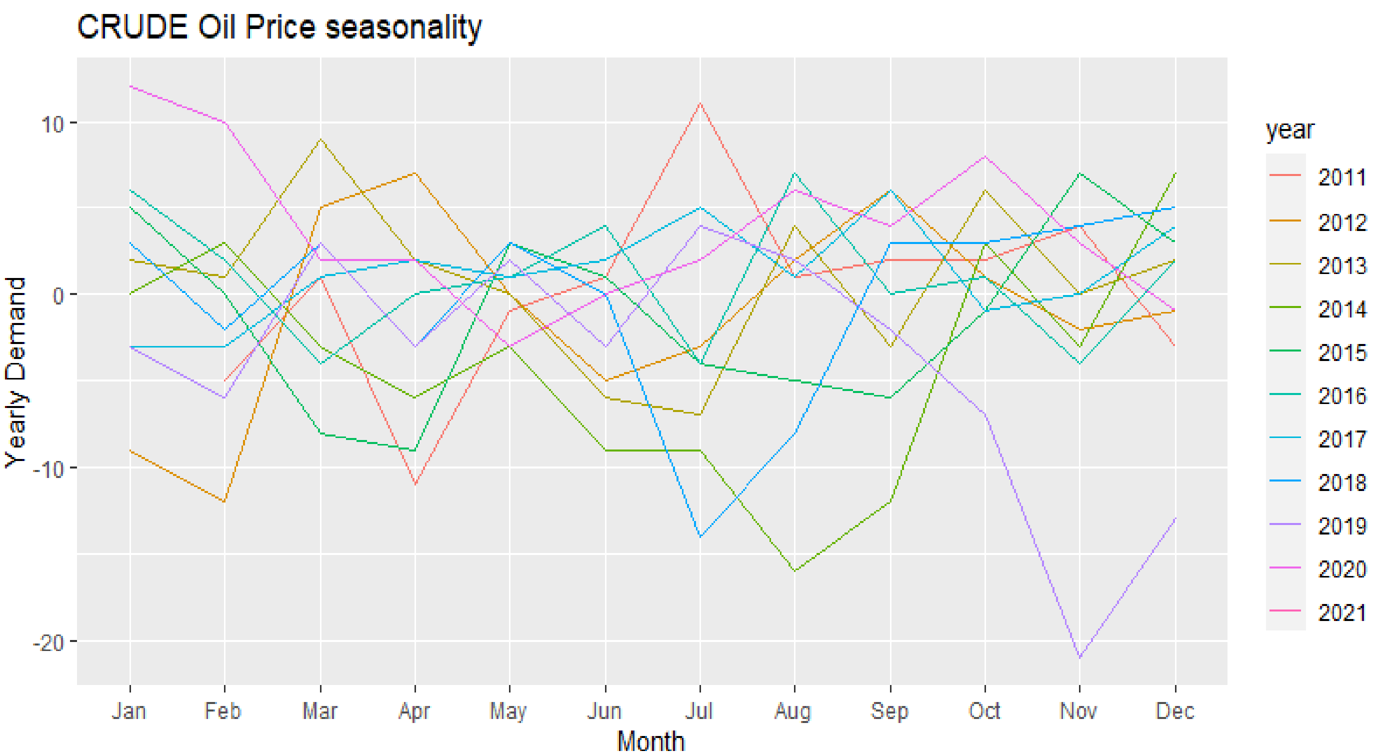


Fig. 4.14 Crude Oil seasonality Graph.

4.7 GG Subseries Plot

An alternative plot that emphasises the seasonal patterns is where the data for each season are collected together in separate mini time plots.

The horizontal blue lines show the means for each month. This form of plot enables the underlying seasonal pattern to be seen clearly and also shows the changes in seasonality. It is especially useful in identifying changes within particular seasons.

Input

```
#ggsubseriesplot for seasonality patterns
ggsubseriesplot(CRUDE_Trend)
```

Fig.4.15 GGSubseries Data

Output

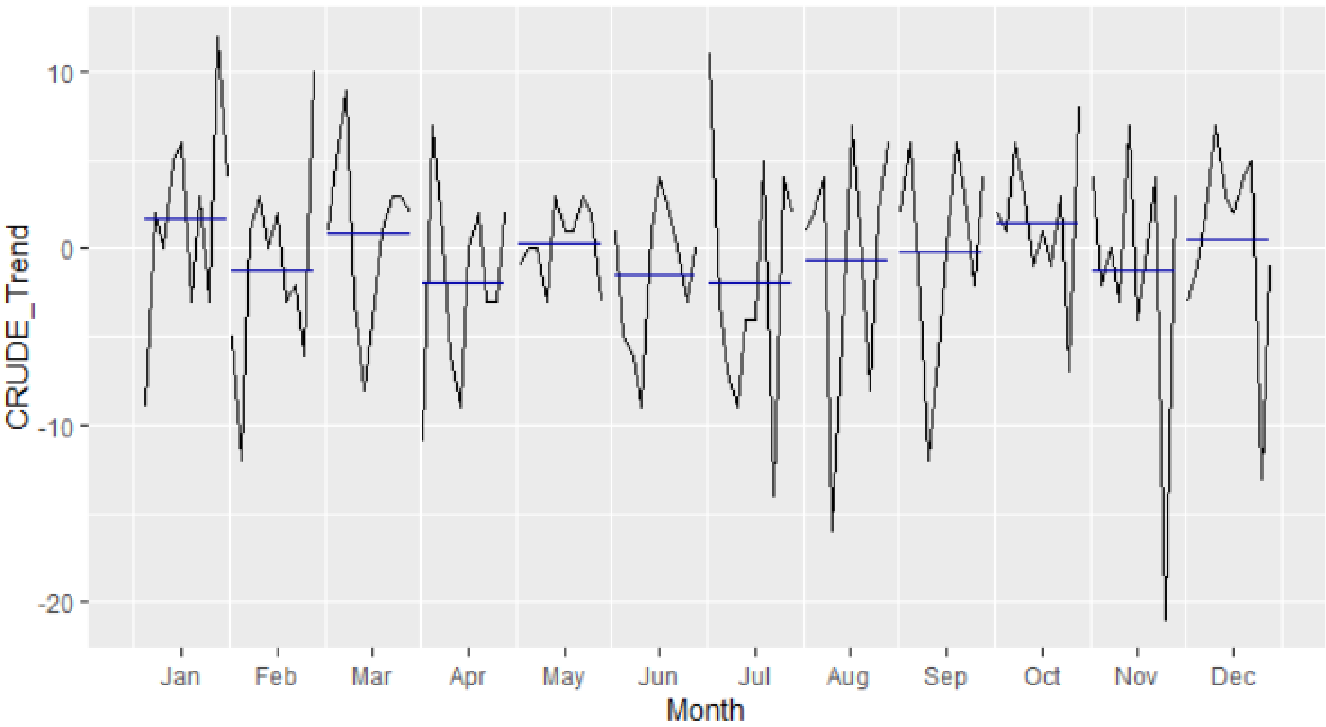


Fig. 4.16–GgSubseries Graphical Representation

4.8 Preliminary Test of Models

Now the most important part of the project is to introduce to our data with the three models which are chosen for forecast, Crude Oil Price.

ARIMA model, Exponential Smoothening Model (ETS) and Naïve benchmarking method for forecasting. As we know, our data which is now been titled with CRUDE_ts used for forecasting with these models.

4.8.1 ARIMA Model

Auto Regression Integrated Moving Average (ARIMA) as we learned a detailed theoretical knowledge about arima and auto arima function in chapter 3. Its time to implement. As we know the auto. Arima () function requires a data.

Returns best ARIMA model according to either AIC, AICc or BIC value. The function conducts a search over viable model within the order constraints provided.

auto. arima (y, d = NA, D = NA, stationary = FALSE, seasonal = TRUE, trace = FALSE, approximation = (length (x) > 150 | frequency (x) > 12)

Arguments,

y- a inferred time series

D- Order of first-differencing. If missing, will choose a value based on the test.

D - Order of seasonal-differencing. If missing, will choose a value based on season test.

Stationary - If TRUE, restricts search to stationary models.

Seasonal - If FALSE restricts search to non-seasonal models.

Trace - If TRUE, the list of ARIMA models considered will be reported.

Approximation - If TRUE, estimation is via conditional sums of squares and the information criteria used for model selection are approximated. The ultimate model is still computed using maximum likelihood estimation. Approximation should be used for a long time series or a high seasonal period to avoid excessive computation times.

Input

```
crude2_ARIMA = auto.arima (CRUDE_ts, d = 1, D = 1, stepwise = FALSE,
approximation = FALSE, trace = TRUE)
print(summary(crude2_ARIMA))
checkresiduals(crude2_ARIMA)
```

Fig. 4.17 - Applying Auto ARIMA model

Output

Estimation Result and the list of ARIMA models from that which model will be suitable to plot ARIMA.

```
> ###########################################
> crude2_ARIMA = auto.arima(CRUDE_ts,d = 1, D = 1, stepwise = FALSE, approximation = FALSE, trace = TRUE)

 ARIMA(0,1,0)(0,1,0)[12]                    : 757.3749
 ARIMA(0,1,0)(0,1,1)[12]                    : Inf
 ARIMA(0,1,0)(0,1,2)[12]                    : Inf
 ARIMA(0,1,0)(1,1,0)[12]                    : 739.3438
 ARIMA(0,1,0)(1,1,1)[12]                    : Inf
 ARIMA(0,1,0)(1,1,2)[12]                    : Inf
 ARIMA(0,1,0)(2,1,0)[12]                    : 732.4806
 ARIMA(0,1,0)(2,1,1)[12]                    : Inf
 ARIMA(0,1,0)(2,1,2)[12]                    : Inf
 ARIMA(0,1,1)(0,1,0)[12]                    : 744.1171
 ARIMA(0,1,1)(0,1,1)[12]                    : Inf
 ARIMA(0,1,1)(0,1,2)[12]                    : Inf
 ARIMA(0,1,1)(1,1,0)[12]                    : 723.8595
 ARIMA(0,1,1)(1,1,1)[12]                    : Inf
 ARIMA(0,1,1)(1,1,2)[12]                    : Inf
 ARIMA(0,1,1)(2,1,0)[12]                    : 716.6436
 ARIMA(0,1,1)(2,1,1)[12]                    : Inf
 ARIMA(0,1,1)(2,1,2)[12]                    : Inf
 ARIMA(0,1,2)(0,1,0)[12]                    : 744.796
 ARIMA(0,1,2)(0,1,1)[12]                    : Inf
 ARIMA(0,1,2)(0,1,2)[12]                    : Inf
 ARIMA(0,1,2)(1,1,0)[12]                    : 723.8671
 ARIMA(0,1,2)(1,1,1)[12]                    : Inf
 ARIMA(0,1,2)(1,1,2)[12]                    : Inf
 ARIMA(0,1,2)(2,1,0)[12]                    : 716.8493
 ARIMA(0,1,2)(2,1,1)[12]                    : Inf
 ARIMA(0,1,3)(0,1,0)[12]                    : 745.2775
 ARIMA(0,1,3)(0,1,1)[12]                    : Inf
 ARIMA(0,1,3)(0,1,2)[12]                    : Inf
 ARIMA(0,1,3)(1,1,0)[12]                    : 724.8714
 ARIMA(0,1,3)(1,1,1)[12]                    : Inf
 ARIMA(0,1,3)(2,1,0)[12]                    : 717.3556
 ARIMA(0,1,4)(0,1,0)[12]                    : 744.944
 ARIMA(0,1,4)(0,1,1)[12]                    : Inf
 ARIMA(0,1,4)(1,1,0)[12]                    : 725.6804
```

```
ARIMA(2,1,0)(0,1,1)[12]                    : Inf
ARIMA(2,1,0)(0,1,2)[12]                    : Inf
ARIMA(2,1,0)(1,1,0)[12]                    : 723.4279
ARIMA(2,1,0)(1,1,1)[12]                    : Inf
ARIMA(2,1,0)(1,1,2)[12]                    : Inf
ARIMA(2,1,0)(2,1,0)[12]                    : 716.038
ARIMA(2,1,0)(2,1,1)[12]                    : Inf
ARIMA(2,1,1)(0,1,0)[12]                    : 743.7822
ARIMA(2,1,1)(0,1,1)[12]                    : Inf
ARIMA(2,1,1)(0,1,2)[12]                    : Inf
ARIMA(2,1,1)(1,1,0)[12]                    : 724.1421
ARIMA(2,1,1)(1,1,1)[12]                    : Inf
ARIMA(2,1,1)(2,1,0)[12]                    : 716.4147
ARIMA(2,1,2)(0,1,0)[12]                    : Inf
ARIMA(2,1,2)(0,1,1)[12]                    : Inf
ARIMA(2,1,2)(1,1,0)[12]                    : 723.2061
ARIMA(2,1,3)(0,1,0)[12]                    : Inf
ARIMA(3,1,0)(0,1,0)[12]                    : 741.9783
ARIMA(3,1,0)(0,1,1)[12]                    : Inf
ARIMA(3,1,0)(0,1,2)[12]                    : Inf
ARIMA(3,1,0)(1,1,0)[12]                    : 722.7241
ARIMA(3,1,0)(1,1,1)[12]                    : Inf
ARIMA(3,1,0)(2,1,0)[12]                    : 715.4192
ARIMA(3,1,1)(0,1,0)[12]                    : 744.1759
ARIMA(3,1,1)(0,1,1)[12]                    : Inf
ARIMA(3,1,1)(1,1,0)[12]                    : 724.8674
ARIMA(3,1,2)(0,1,0)[12]                    : Inf
ARIMA(4,1,0)(0,1,0)[12]                    : 744.1745
ARIMA(4,1,0)(0,1,1)[12]                    : Inf
ARIMA(4,1,0)(1,1,0)[12]                    : 724.8072
ARIMA(4,1,1)(0,1,0)[12]                    : 746.2449
ARIMA(5,1,0)(0,1,0)[12]                    : 746.0071


Best model: ARIMA(3,1,0)(2,1,0)[12]
```

Fig 4.18 Output Parameters Values

From the above fig list of arima models we can see in which it gave the values of AICc, we know the lowest AIC value will be the best suited model which will be selected by the Auto ARIMA function.

1) Evaluation of ARIMA model

Sr No	Models	AIC
1	ARIMA (0,1,0) (0,1,0)	778.0239
2	ARIMA (0,1,0) (1,1,0)	756.0993
3	ARIMA (0,1,0) (2,1,0)	749.2022
4	ARIMA (1,1,0) (0,1,0)	775.9559
5	ARIMA (1,1,0) (1,1,0)	754.3458
6	ARIMA (1,1,0) (2,1,0)	747.9945
7	ARIMA (2,1,0) (0,1,0)	773.9969
8	ARIMA (2,1,0) (1,1,0)	752.1172
9	ARIMA (2,1,0) (2,1,0)	745.534
10	ARIMA (3,1,0) (0,1,0)	764.2953
11	ARIMA (3,1,0) (1,1,0)	742.3888
12	ARIMA (3,1,0) (2,1,0)	736.8119
13	ARIMA (4,1,0) (0,1,0)	757.3925
14	ARIMA (4,1,0) (1,1,0)	737.295
15	ARIMA (5,1,0) (0,1,0)	755.5521

Table No. 2 Evaluation list of ARIMA model

So the best model is ARIMA (3,1,0) (2,1,0) with AICc value = 736.8119, which is the lowest value as compare to the other models.

2) Summary of ARIMA model

```
 Best model: ARIMA(3,1,0)(2,1,0)[12]

> print(summary(crude2_ARIMA))
Series: CRUDE_ts
ARIMA(3,1,0)(2,1,0)[12]

Coefficients:
         ar1      ar2      ar3     sar1    sar2
      0.4196  -0.1126  -0.1645  -0.5903  -0.344
s.e.  0.0948   0.1021   0.0965   0.0978   0.105

sigma^2 estimated as 38.87:  log likelihood=-351.29
AIC=714.59   AICc=715.42   BIC=730.68

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 0.1314101 5.752467 4.297951 -0.111003 8.175442 0.2712662 -0.002981059
```

Fig 4.19 Summary of ARIMA model

3) Parameters from the Above figure

Parameters	Value
Sigma square root	**38.87**
AIC	**714.42**
AICc	**715.42**
BIC	**730.68**
Log Likely hood	**-351.29**
ME	**0.1314101%**
RMSE	**5.752467%**
MAE	**4.297951%**

Table .3 ARIMA Model Parameters

Standard deviation Value = $\sqrt{38.87}$ = **6.23**

The standard deviation value decides whether which model shows the lowest value and that value is the sign of the forecast model, that means we can choose that model for our final forecasting.

4) Checking Residuals of ARIMA

```
> checkresiduals(crude2_ARIMA)

        Ljung-Box test

data:  Residuals from ARIMA(3,1,0)(2,1,0)[12]
Q* = 14.81, df = 19, p-value = 0.7346

Model df: 5.   Total lags used: 24
```

Fig 4.20 ARIMA Residuals values

The above output while checking the residual of the ARIMA model, in which we get the parameter values by Ljung–Box test.

Ljung-Box test - Compute the Box–Pierce or Ljung–Box test statistic for examining the null hypothesis of independence in a time series. These are sometimes known as 'portmanteau' tests.

Box.test(x, lag = 1, type = c (" Box-Pierce", "Ljung-Box"), fitdf = 0)

The above code can check Ljung test but in ARIMA we can use check Residual which directly gives the test value which comprise Model of, Total Lags and most important P-value.

Significance of P value talks about the hypothesis test result.

- High P-values: Your sample results are consistent with a null hypothesis that is true.
- Low P-values: Your sample results contradict a null hypothesis that is true

5) **Overview of ARIMA in Rstudio**

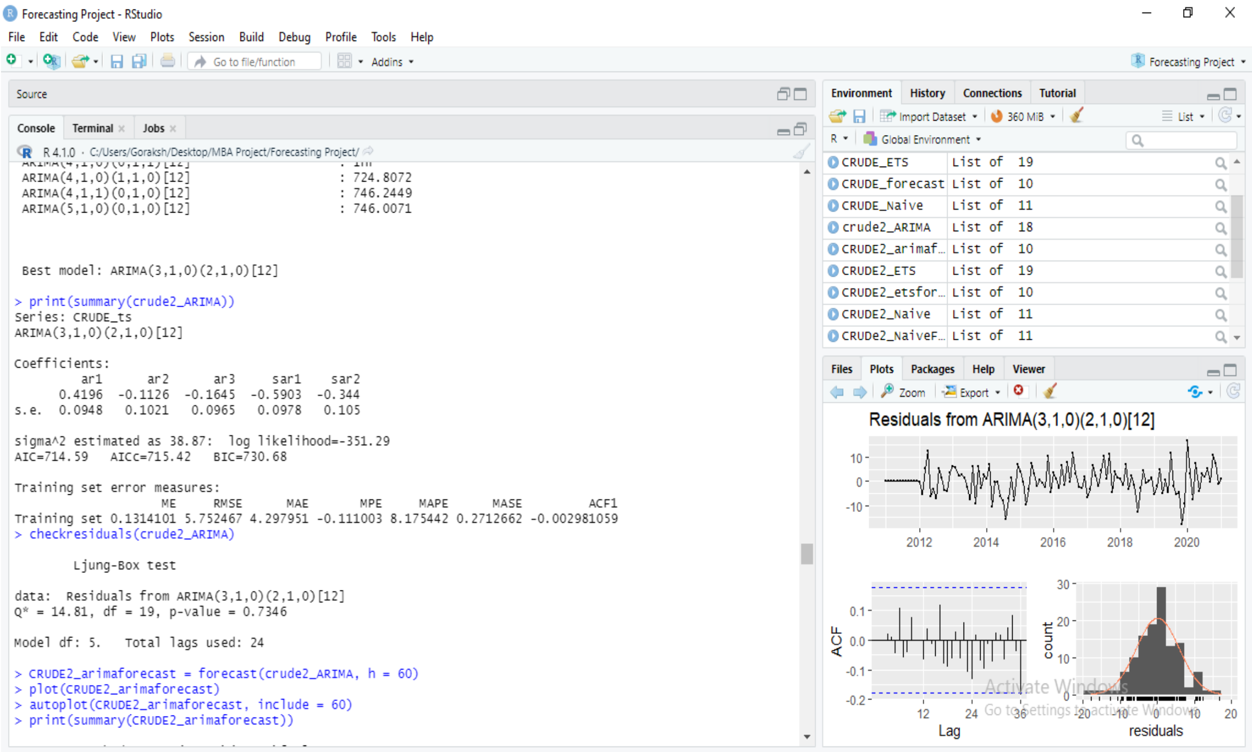


Fig. 4.21 Dashboard view of Rstudio while performing Residuals

6) Graphical Representation of ARIMA model

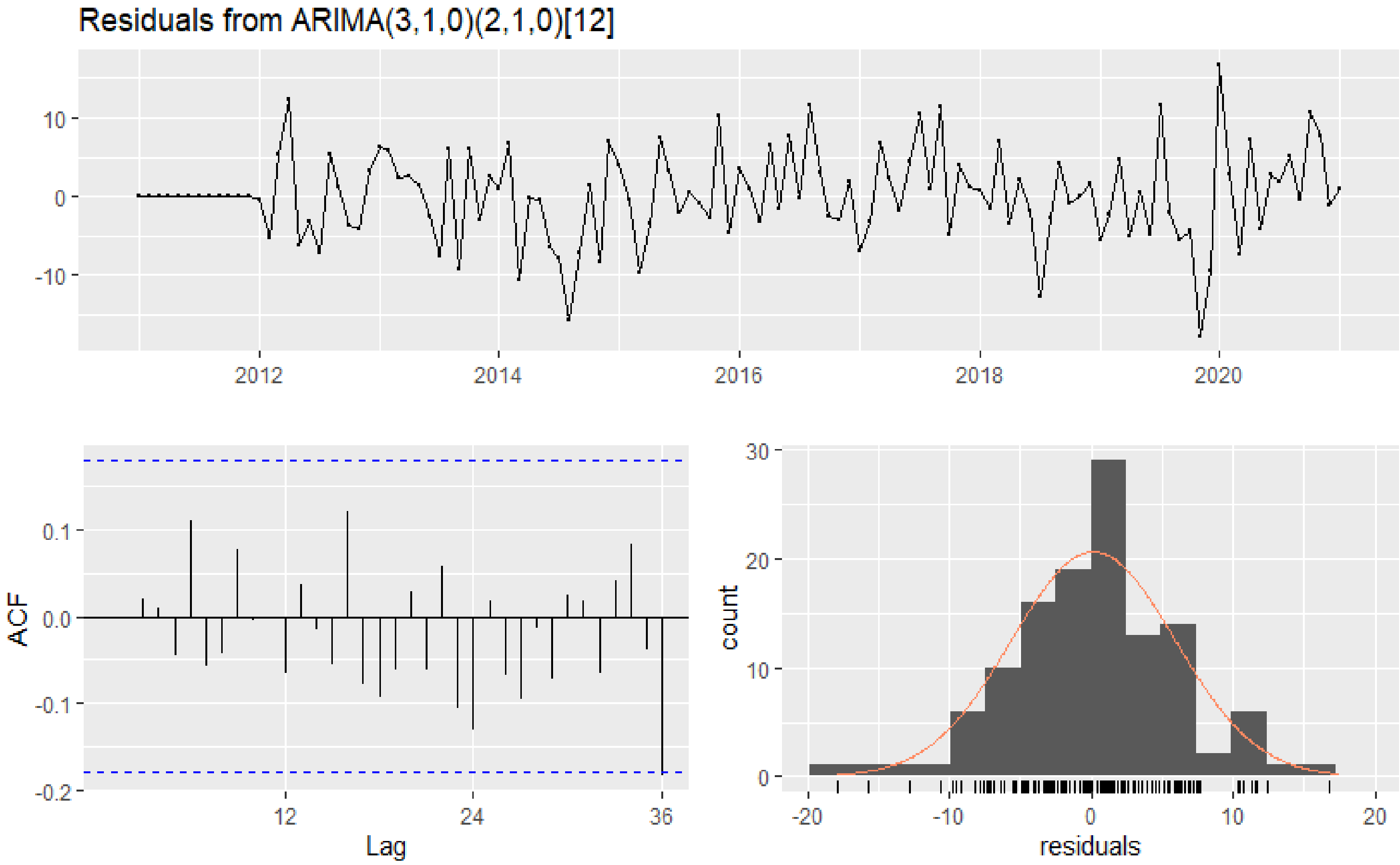


Fig 4.22 Graphical Representation of ARIMA model

4.8.2 Exponential Smoothening Model

The detailed theory of this model we learned already in chapter 3. the ets () function does not produce forecasts. It estimates the model parameters and returns information about the fitted model. By default, it uses the AICc to pick an appropriate model, although other information criteria are often selected.

A significant advantage of the ETS framework for exponential smoothers is that we can use the model selection criterion that we have used previously, namely AIC, AICc, and BIC.

The below table is the model of ets function. Each model comprises a measurement equation that describes the observed data, and some transition equations that describe how the unobserved components or states (level, trend, seasonal) change by time. Hence, these are referred to as "state-space models".

Name	Model
Simple Exponential Smooth with Additive Errors	(And, N, N)
SES with multiplicative errors	(M, N, N)
Holt's linear method with additive errors	(And, A, N)
Holt's linear method with multiplicative errors	(M, A, N)

Table no.4 ETS model information

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N, N)	(N, A)	(N, M)
A (Additive)	(A, N)	(A, A)	(A, M)
A_d (Additive Damped)	(A_d, N)	(A_d, A)	(A_d, M)

Fig 4.23. Taxonomy of Exponential Smoothening model.

1. Exponential Smoothening Model in R

Input

```
CRUDE2_ETS = ets(CRUDE_ts)
print(summary(CRUDE2_ETS))
checkresiduals(CRUDE2_ETS)
```

Fig 4.24 Applying ETS model

Output

```
> CRUDE2_ETS = ets(CRUDE_ts)
> print(summary(CRUDE2_ETS))
ETS(A,N,N)

Call:
 ets(y = CRUDE_ts)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 100.9913

  sigma:  5.5928

     AIC     AICc      BIC
1000.872 1001.077 1009.260

Training set error measures:
                     ME    RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set -0.2974818 5.54639 4.165507 -1.162133 7.929244 0.2629069 0.3414568
```

Fig 4.25 Output of ETS model

2. Parameters from the Output

Parameters	**Values**
Sigma	**5.5928**
AIC	**1000.872**
AICc	**1001.077**
BIC	**1009.260**
ME	**-0.2974818%**

RMSE	**5.54639%**
MAE	**4.1655%**
ACF1	**0.3414568%**

Table No. 5 ETS Output Parameters Values

From the output result we can see ets function not only provided the best fitted model i.e., ETS (A, N, N) model as well as it has given the sigma value which is more important for comparison sigma–5.6084 standard deviation value.

3. Checking Residuals of ETS model

```
> checkresiduals(CRUDE2_ETS)

        Ljung-Box test

data:  Residuals from ETS(A,N,N)
Q* = 33.689, df = 22, p-value = 0.05282

Model df: 2.   Total lags used: 24
```

Fig. 4.26 Residual checking for ETS model

From the residuals check we can see the Ljung-Box test has been implemented to check the value of P, as we know the significance of P in ARIMA model. Now we can see the main graphical representation of exponential smoothening method for forecasting.

4. Overview of ETS in R studio

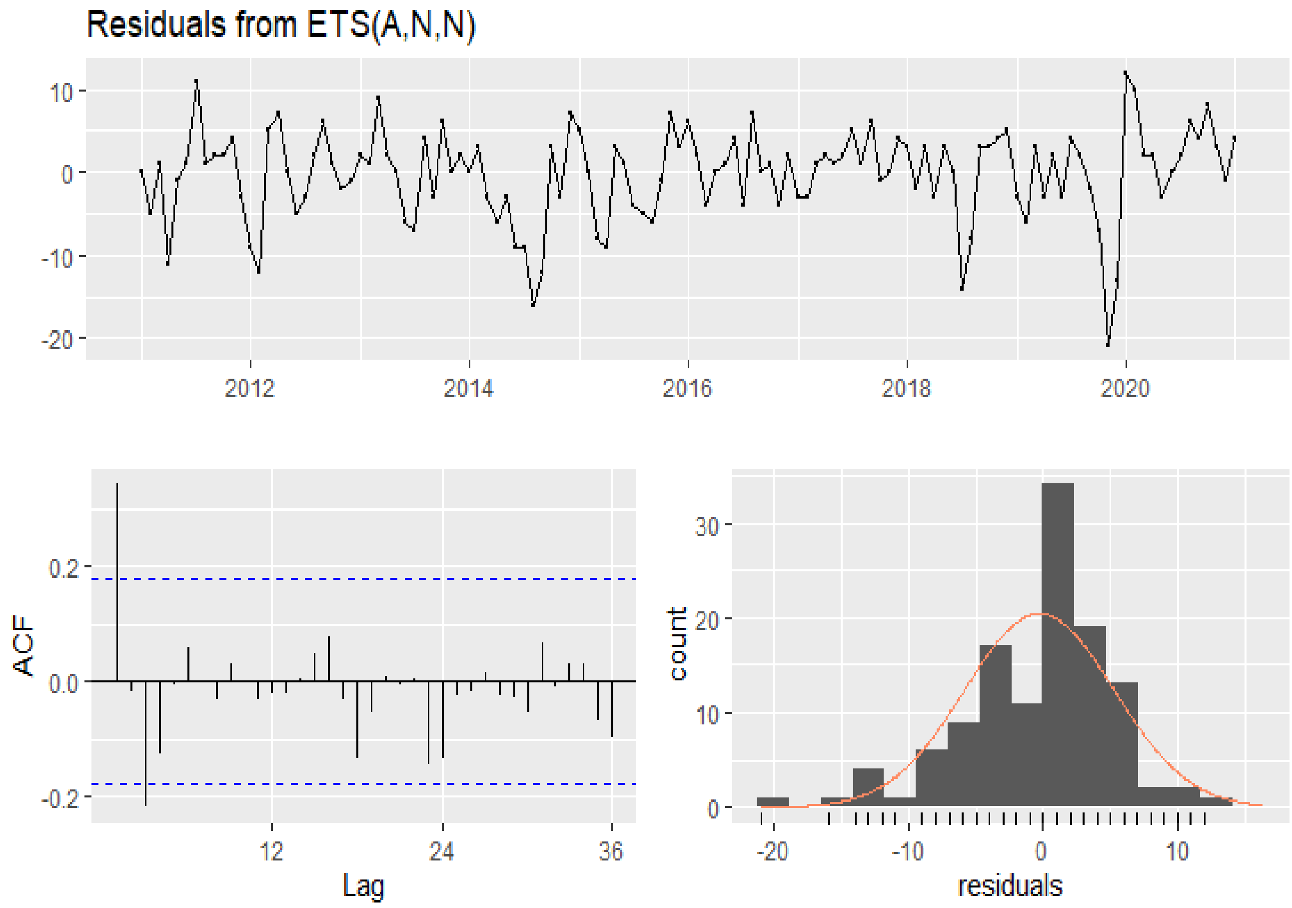


Fig. 4.27 Graphical Representation of ETS (A, N, N) model

4.8.3 NAÏVE Benchmarking Method

Naïve benchmarking model is the easiest way for forecasting but it has certain limitations which we have learned in the chapter 3 theory. Now we will apply a Naïve model for forecasting our CRUDE oil data.

Input

```
CRUDE2_Naive = snaive(CRUDE_ts)
print(summary(CRUDE2_Naive)) # Standard Deviation - 20.8518
checkresiduals(CRUDE2_Naive)
```

Fig 4.28 Applying Naïve Method

Output

```
Forecast method: Seasonal naive method

Model Information:
Call: snaive(y = CRUDE_ts)

Residual sd: 20.8518

Error measures:
                    ME     RMSE      MAE       MPE     MAPE MASE      ACF1
Training set -5.256881 20.85181 15.84404 -15.86799 31.70833    1 0.9022023

Forecasts:
         Point Forecast    Lo 80     Hi 80        Lo 95     Hi 95
Feb 2021             38 11.27732  64.72268   -2.8688053  78.86881
Mar 2021             40 13.27732  66.72268   -0.8688053  80.86881
Apr 2021             42 15.27732  68.72268    1.1311947  82.86881
May 2021             39 12.27732  65.72268   -1.8688053  79.86881
Jun 2021             39 12.27732  65.72268   -1.8688053  79.86881
Jul 2021             41 14.27732  67.72268    0.1311947  81.86881
Aug 2021             47 20.27732  73.72268    6.1311947  87.86881
Sep 2021             51 24.27732  77.72268   10.1311947  91.86881
Oct 2021             59 32.27732  85.72268   18.1311947  99.86881
Nov 2021             62 35.27732  88.72268   21.1311947 102.86881
Dec 2021             61 34.27732  87.72268   20.1311947 101.86881
Jan 2022             65 38.27732  91.72268   24.1311947 105.86881
Feb 2022             38  0.20843  75.79157  -19.7972187  95.79722
Mar 2022             40  2.20843  77.79157  -17.7972187  97.79722
Apr 2022             42  4.20843  79.79157  -15.7972187  99.79722
May 2022             39  1.20843  76.79157  -18.7972187  96.79722
Jun 2022             39  1.20843  76.79157  -18.7972187  96.79722
Jul 2022             41  3.20843  78.79157  -16.7972187  98.79722
Aug 2022             47  9.20843  84.79157  -10.7972187 104.79722
Sep 2022             51 13.20843  88.79157   -6.7972187 108.79722
Oct 2022             59 21.20843  96.79157    1.2027813 116.79722
Nov 2022             62 24.20843  99.79157    4.2027813 119.79722
Dec 2022             61 23.20843  98.79157    3.2027813 118.79722
Jan 2023             65 27.20843 102.79157    7.2027813 122.79722
```

Fig 4.29 Naïve Method Output Values

a. Parameters from the Output

Parameters	Values
Residual Std. Deviation	**20.8518**
ME	**-5.256881%**
RMSE	**20.85181%**
MAE	**15.8440%**
ACF1	**0.9022023%**

Table 6 Naïve Parameters Values

From the above output we can see the naïve () function is used as our data is seasonal so seasonal naïve method is used and the value in which return, we found that the standard deviation value sigma–7.9884 which will be saved for the comparison for the best fitted model for forecasting.

b. Checking the Residuals of Naïve Method

```
> checkresiduals(CRUDE2_Naive)

        Ljung-Box test

data:  Residuals from Seasonal naive method
Q* = 411, df = 24, p-value < 2.2e-16

Model df: 0.   Total lags used: 24
```

Fig. 4.30 Naïve Model Residual Checking

After all, the residual work is to give the value by doing Ljung-Box test and it gives the P value which is very low as compared to other models and not at all near to P = 0.05 the significant value of.

c. Overview of Naïve Model in Rstudio

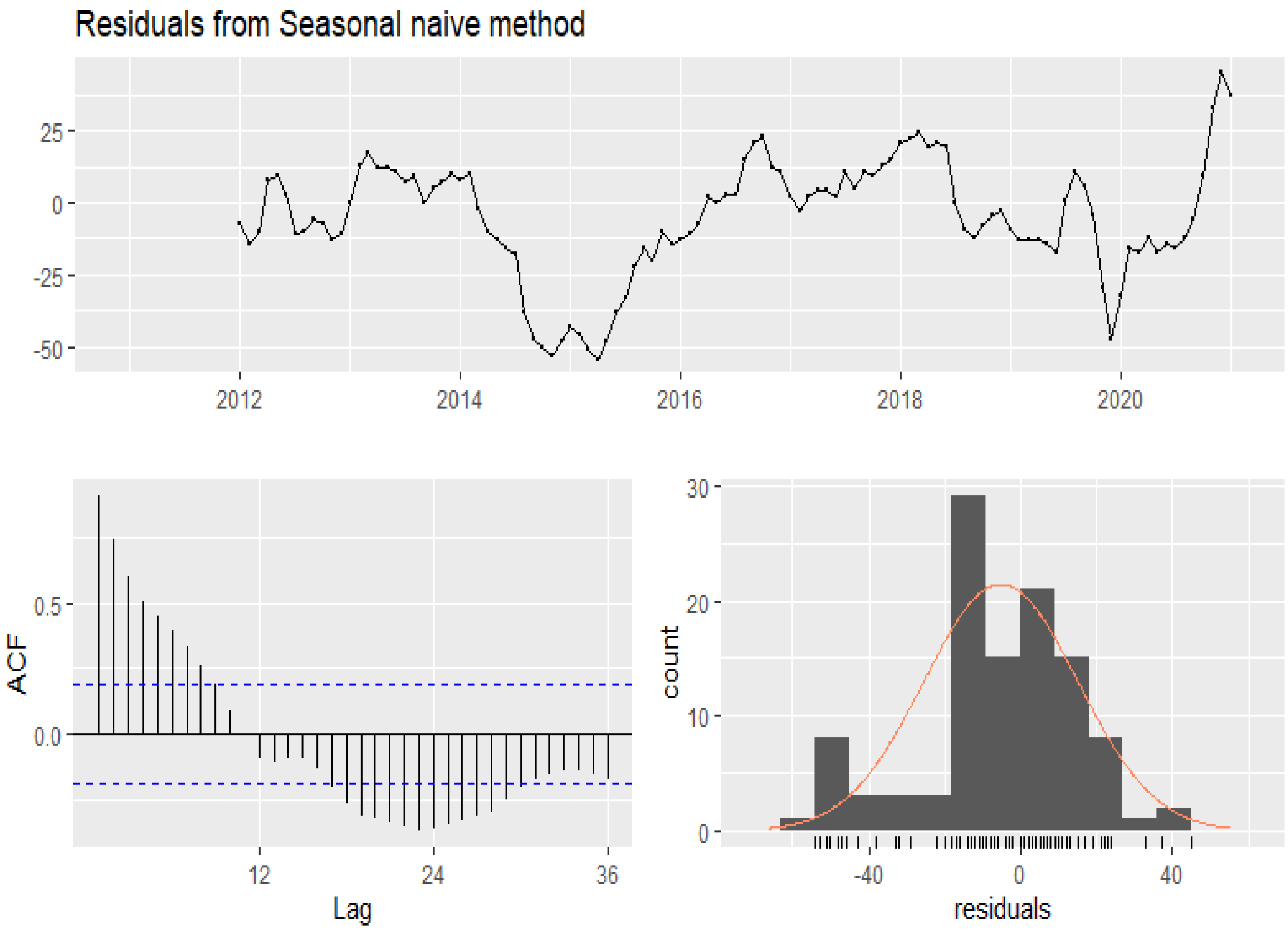


Fig. 4.31 Graphical Representation of Naïve Method

4.9 Selecting the Best Model for Forecasting

As we have checked all three models ARIMA, Naïve and ETS for the forecasting and we will now compare their results to choose the last model for Forecasting the CRUDE oil. Before selecting model, we need to check their parameters which will give the best value for forecasting.

a) Comparing P value

Model	P value
ARIMA	0.7346
NAÏVE	2.2e-16
ETS	0.05282

Table. .7 Comparing P value

Why we have chosen P value for comparison? Because P value is significant value P values tell you whether your hypothesis test results are statistically significant. Statistics use them all over the place. P values are the probability of observing a sample statistic.

- P value < 0.05 statistically significant and
- P value < 0.001 is statistically highly significant

b) Comparing Standard Deviation

Model	δ Sigma Value
ARIMA	6.23
NAÏVE	20.8518
ETS	5.5928

Table No. 8 Comparing Standard Deviation

We chose standard deviation value, which is lowest for the final forecasting of our data.

4.10 Forecasting Models

As we have reached finally to the last step for forecasting the best model by using the forecast function which is denoted by forecast ().

The forecast () function works with many inputs. It takes a time series or time series model as its principal argument, and produces forecasts appropriately. It always returns objects of class forecast.

Objects of class forecast contain information about the forecasting method, the data used, the point forecasts obtained, prediction intervals, residuals and fitted values. There are several functions designed to work with these objects including autoplot (), summary () and print ().

A. ARIMA Model for Forecasting

ARIMA model is the most significant as compare to other models the result which we received while performing task through ARIMA is close enough to give the best result for forecasting.

From the start by applying ARIMA model to our data i.e., to time series data CRUDE_ts the values we received in the form of the most fitted model ARIMA (3, 1, 0) (2, 1, 0).

This model has the better result of lowest AIC value of 714.59. this is how the forecast function gives the result,

h = is the period for which we have to forecast, if the data is of 12 months than 12 multiply by 5years forecasting we required h = 60.

Print () function gives the summary result for the parameters value for diagnose the errors appeared and certain parameters

Input

```
CRUDE2_arimaforecast = forecast(crude2_ARIMA, h = 60)
plot(CRUDE2_arimaforecast)
autoplot(CRUDE2_arimaforecast, include = 60)
print(summary(CRUDE2_arimaforecast))
```

Fig 4.32. Applying Forecast function in ARIMA model

Output

```
Forecast method: ARIMA(3,1,0)(2,1,0)[12]

Model Information:
Series: CRUDE_ts
ARIMA(3,1,0)(2,1,0)[12]

Coefficients:
         ar1      ar2      ar3     sar1    sar2
      0.4196  -0.1126  -0.1645  -0.5903  -0.344
s.e.  0.0948   0.1021   0.0965   0.0978   0.105

sigma^2 estimated as 38.87:  log likelihood=-351.29
AIC=714.59   AICc=715.42   BIC=730.68

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.1314101 5.752467 4.297951 -0.111003 8.175442 0.2712662 -0.00298

Forecasts:
         Point Forecast       Lo 80     Hi 80      Lo 95     Hi 95
Feb 2021       64.48334 56.49301567  72.47366  52.263194  76.70348
Mar 2021       65.90000 52.02518841  79.77481  44.680308  87.11969
Apr 2021       64.93070 46.68424042  83.17715  37.025150  92.83624
May 2021       65.75365 44.76592730  86.74136  33.655701  97.85159
Jun 2021       65.43119 42.53909734  88.32329  30.420755 100.44163
Jul 2021       62.53758 38.04494795  87.03021  25.079331  99.99583
Aug 2021       62.65208 36.58948870  88.71466  22.792788 102.51136
Sep 2021       64.71331 37.05993896  92.36669  22.421127 107.00550
Oct 2021       67.23992 38.02189232  96.45794  22.554804 111.92503
Nov 2021       64.67550 33.96978129  95.38123  17.715154 111.63585
Dec 2021       62.81118 30.70788924  94.91446  13.713436 111.90892
Jan 2022       66.39395 32.96894772  99.81896  15.274819 117.51309
Feb 2022       66.58621 30.87909378 102.29333  11.976888 121.19553
Mar 2022       68.68672 30.35317516 107.02027  10.060621 127.31282
```

Fig 4.33 Forecast Values

The figure shows the list of the forecast value which we received while performing forecasting function as we can see the level of output i.e., Low and High of 80% and 90% these are the confidence level for the forecasting value.

Mar 2022	68.68672	30.35317516	107.02027	10.060621	127.31282
Apr 2022	67.74417	26.87557705	108.61277	5.241050	130.24729
May 2022	68.02771	24.96423315	111.09118	2.167807	133.88761
Jun 2022	66.86374	21.86075589	111.86672	-1.962384	135.68986
Jul 2022	67.54827	20.73614686	114.36040	-4.044695	139.14124
Aug 2022	69.76192	21.18699725	118.33684	-4.527012	144.05084
Sep 2022	70.90370	20.59241191	121.21498	-6.040773	147.84816
Oct 2022	71.50063	19.49120633	123.51006	-8.040922	151.04219
Nov 2022	63.96400	10.30905750	117.61894	-18.094153	146.02215
Dec 2022	58.48144	3.23697586	113.72591	-26.007679	142.97057
Jan 2023	65.06273	8.27914814	121.84631	-21.780263	151.90572
Feb 2023	68.45458	9.34129176	127.56787	-21.951393	158.86055
Mar 2023	70.35215	8.60996121	132.09433	-24.074377	164.77867
Apr 2023	70.41533	6.09120732	134.73945	-27.959926	168.79059
May 2023	69.70208	3.05757603	136.34658	-32.221892	171.62605
Jun 2023	69.14578	0.38605898	137.90549	-36.013136	174.30469
Jul 2023	69.40164	-1.36716605	140.17045	-38.829910	177.63319
Aug 2023	72.40092	-0.34034244	145.14218	-38.847239	183.64907
Sep 2023	74.75244	0.05952147	149.44536	-39.480522	188.98541
Oct 2023	78.37146	1.75914782	154.98377	-38.796959	195.53987
Nov 2023	75.68424	-2.80139565	154.16987	-44.349180	195.71765
Dec 2023	72.63488	-7.67404143	152.94380	-50.187016	195.45678
Jan 2024	77.58967	-4.49667909	159.67602	-47.950569	203.12991
Feb 2024	78.84891	-6.10659939	163.80442	-51.079329	208.77715
Mar 2024	80.63101	-7.63781882	168.89984	-54.364516	215.62654
Apr 2024	80.09131	-11.45486845	171.63748	-59.916486	220.09910
May 2024	80.15203	-14.33013304	174.63420	-64.345970	224.65004
Jun 2024	79.52653	-17.62446876	176.67752	-69.053099	228.10615
Jul 2024	78.80446	-20.88368410	178.49261	-73.655401	231.26433
Aug 2024	80.61782	-21.57120328	182.80684	-75.666803	236.90244
Sep 2024	82.57154	-22.10291927	187.24601	-77.514232	242.65732
Oct 2024	85.07047	-22.05649514	192.19744	-78.766087	248.90704
Nov 2024	81.23120	-28.29455774	190.75696	-86.273993	248.73640
Dec 2024	77.99029	-33.87377633	189.85435	-93.091034	249.07161
Jan 2025	82.87364	-31.27317256	197.02045	-91.698844	257.44612

Fig 4.33 Forecast Values

Mar 2023	70.35215	8.60996121	132.09433	-24.074377	164.77867
Apr 2023	70.41533	6.09120732	134.73945	-27.959926	168.79059
May 2023	69.70208	3.05757603	136.34658	-32.221892	171.62605
Jun 2023	69.14578	0.38605898	137.90549	-36.013136	174.30469
Jul 2023	69.40164	-1.36716605	140.17045	-38.829910	177.63319
Aug 2023	72.40092	-0.34034244	145.14218	-38.847239	183.64907
Sep 2023	74.75244	0.05952147	149.44536	-39.480522	188.98541
Oct 2023	78.37146	1.75914782	154.98377	-38.796959	195.53987
Nov 2023	75.68424	-2.80139565	154.16987	-44.349180	195.71765
Dec 2023	72.63488	-7.67404143	152.94380	-50.187016	195.45678
Jan 2024	77.58967	-4.49667909	159.67602	-47.950569	203.12991
Feb 2024	78.84891	-6.10659939	163.80442	-51.079329	208.77715
Mar 2024	80.63101	-7.63781882	168.89984	-54.364516	215.62654
Apr 2024	80.09131	-11.45486845	171.63748	-59.916486	220.09910
May 2024	80.15203	-14.33013304	174.63420	-64.345970	224.65004
Jun 2024	79.52653	-17.62446876	176.67752	-69.053099	228.10615
Jul 2024	78.80446	-20.88368410	178.49261	-73.655401	231.26433
Aug 2024	80.61782	-21.57120328	182.80684	-75.666803	236.90244
Sep 2024	82.57154	-22.10291927	187.24601	-77.514232	242.65732
Oct 2024	85.07047	-22.05649514	192.19744	-78.766087	248.90704
Nov 2024	81.23120	-28.29455774	190.75696	-86.273993	248.73640
Dec 2024	77.99029	-33.87377633	189.85435	-93.091034	249.07161
Jan 2025	82.87364	-31.27317256	197.02045	-91.698844	257.44612
Feb 2025	84.29102	-32.92665610	201.50870	-94.977946	263.55999
Mar 2025	86.21110	-34.37921463	206.80141	-98.215871	270.63807
Apr 2025	85.68128	-38.24986977	209.61244	-103.855059	275.21762
May 2025	85.62805	-41.39498355	212.65108	-108.636915	279.89301
Jun 2025	84.83434	-45.07817759	214.74686	-113.849710	283.51839
Jul 2025	84.83702	-47.85785465	217.53190	-118.102280	287.77633
Aug 2025	87.08015	-48.35935978	222.51967	-120.056707	294.21701
Sep 2025	88.85252	-49.30964124	227.01467	-122.448270	300.15330
Oct 2025	90.97297	-49.87745137	231.82338	-124.439158	306.38509
Nov 2025	86.14543	-57.34381246	229.63467	-133.302429	305.59329
Dec 2025	82.18051	-63.89320387	228.25422	-141.219957	305.58098
Jan 2026	87.66559	-60.94195625	236.27313	-139.610038	314.94121

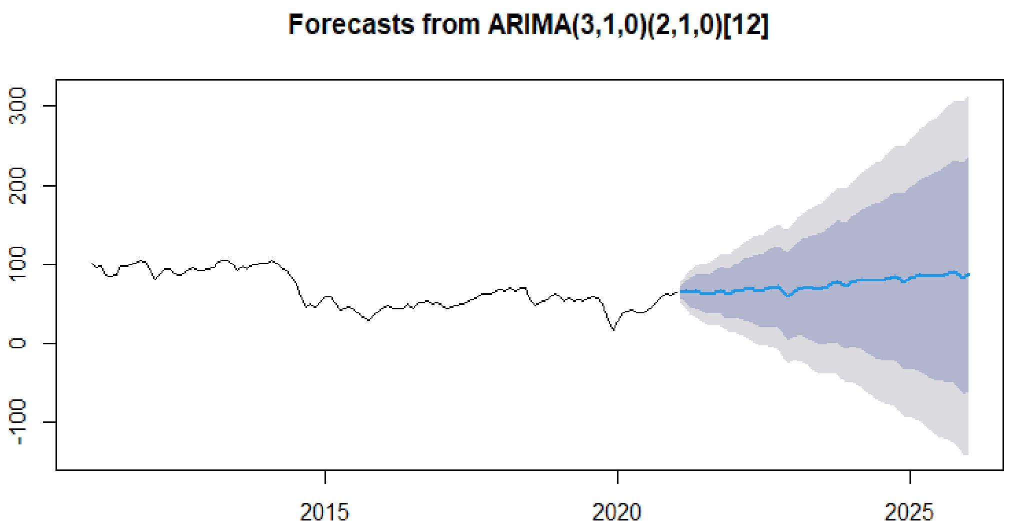
Fig.4.33 Forecast Values

4.11 Forecast Graph Plotting,

Once the values are received after forecasting, we can see the visuals showing trend, future trends, this is what the most important part we can see the CRUDE oil prices fluctuations in the future too.

The below figures are the graph plotted in the R studio by plot () function to which the data we want to forecast, the blue line is the forecast trend which we can see the fluctuations.

Second graph shows the better visual in geometric lines and we can see that for future 5 years till 2026 the Crude oil price forecasting will be fluctuate not more than the past 5years period.



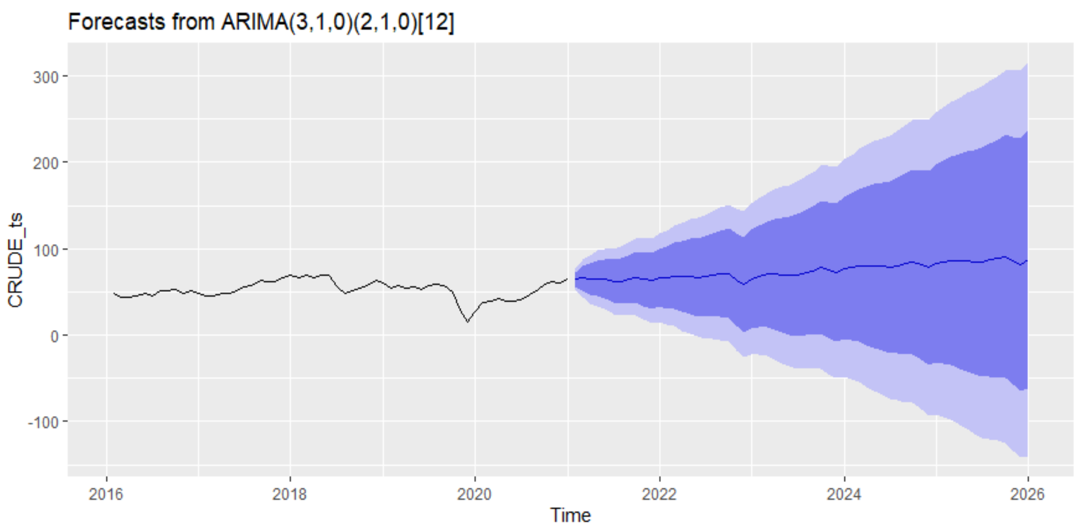


Fig 4.34 Graphical Representation of Forecasting

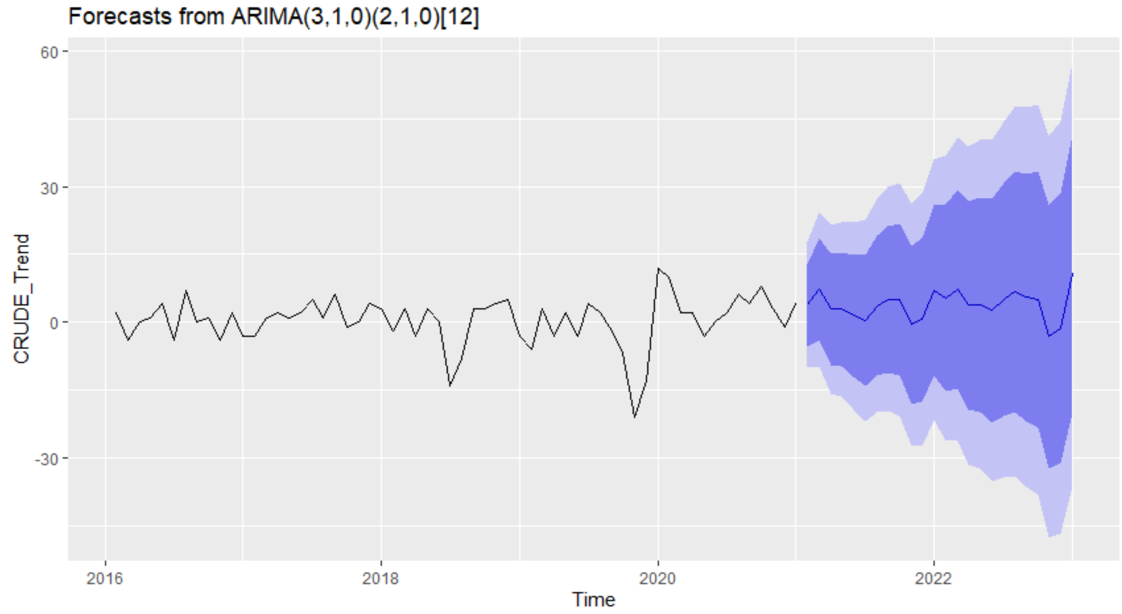


Fig 4.35 Forecasting Pattern with Trend

```
Error measures:
                     ME     RMSE      MAE MPE MAPE      MASE        ACF1
Training set 0.07693069 6.54784 4.810253 NaN  Inf 0.8041909 -0.06762502

Forecasts:
         Point Forecast      Lo 80     Hi 80      Lo 95     Hi 95
Feb 2021      3.7812823  -5.320462 12.88303 -10.138635 17.70120
Mar 2021      7.3134091  -3.905761 18.53258  -9.844831 24.47165
Apr 2021      2.8588625  -9.341223 15.05895 -15.799558 21.51728
May 2021      2.8304675  -9.759661 15.42060 -16.424473 22.08541
Jun 2021      1.4488517 -12.047334 14.94504 -19.191784 22.08949
Jul 2021      0.4186408 -14.112989 14.95027 -21.805570 22.64285
Aug 2021      3.6536312 -11.812342 19.11960 -19.999534 27.30680
Sep 2021      5.0477260 -11.094959 21.19041 -19.640380 29.73583
Oct 2021      4.9615334 -11.843510 21.76658 -20.739562 30.66263
Nov 2021     -0.6396740 -18.136120 16.85677 -27.398179 26.11883
Dec 2021      0.5073969 -17.692968 18.70776 -27.327659 28.34245
Jan 2022      7.0518199 -11.797697 25.90134 -21.776029 35.87967
Feb 2022      5.3793510 -15.264603 26.02330 -26.192851 36.95155
Mar 2022      7.3541850 -14.583208 29.29158 -26.196162 40.90453
Apr 2022      3.6840164 -19.276551 26.64458 -31.431142 38.79917
May 2022      3.8792223 -19.883306 27.64175 -32.462429 40.22087
Jun 2022      2.5819730 -22.132720 27.29667 -35.215888 40.37983
Jul 2022      4.9854582 -20.698042 30.66896 -34.294065 44.26498
Aug 2022      6.7509404 -19.858213 33.36009 -33.944248 47.44613
Sep 2022      5.4486692 -21.991300 32.88864 -36.517142 47.41448
Oct 2022      4.9489506 -23.301869 33.19977 -38.256950 48.15485
Nov 2022     -3.1820210 -32.238274 25.87423 -47.619725 41.25568
Dec 2022     -1.2984794 -31.150999 28.55404 -46.953968 44.35701
Jan 2023     10.7395387 -19.879319 41.35840 -36.087964 57.56704
```

Fig 4.36 Point Forecast without Trend

4.12 Conclusion

The below table is the value of forecasted price for the future ears i.e., from Feb 2021 to January 2026, almost 5-year future forecasting and the Low and High 80 to 95% is the confidence level to attain the forecast.

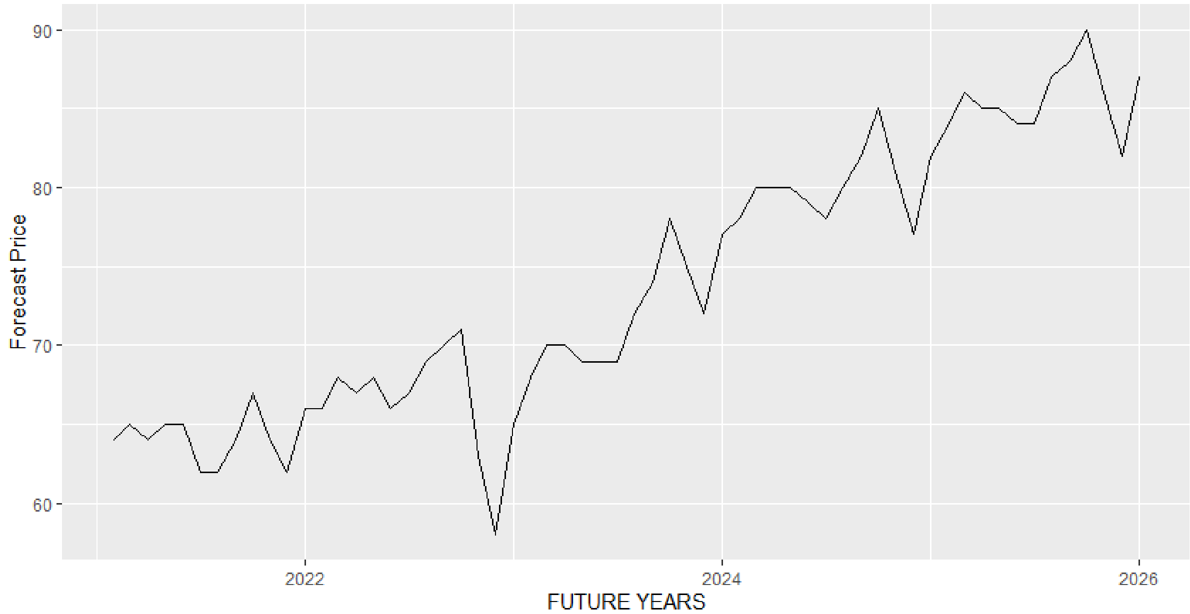
So that in that respective month it may increase from low to high. In the chapter 5 we will plot the graph of the forecasted price of Crude.

CRUDE Oil Forecasts by ARIMA model

Month	Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95,
Feb	2021	64.48334	56.49301567	72.47366	52.263194	76.70348
Mar	2021	65.9	52.02518841	79.77481	44.680308	87.11969
Apr	2021	64.9307	46.68424042	83.17715	37.02515	92.83624
May	2021	65.75365	44.7659273	86.74136	33.655701	97.85159
Jun	2021	65.43119	42.53909734	88.32329	30.420755	100.44163
Jul	2021	62.53758	38.04494795	87.03021	25.079331	99.99583
Aug	2021	62.65208	36.5894887	88.71466	22.792788	102.51136
Sep	2021	64.71331	37.05993896	92.36669	22.421127	107.0055
Oct	2021	67.23992	38.02189232	96.45794	22.554804	111.92503
Nov	2021	64.6755	33.96978129	95.38123	17.715154	111.63585
Dec	2021	62.81118	30.70788924	94.91446	13.713436	111.90892
Jan	2022	66.39395	32.96894772	99.81896	15.274819	117.51309
Feb	2022	66.58621	30.87909378	102.29333	11.976888	121.19553
Mar	2022	68.68672	30.35317516	107.02027	10.060621	127.31282
Apr	2022	67.74417	26.87557705	108.61277	5.24105	130.24729
May	2022	68.02771	24.96423315	111.09118	2.167807	133.88761
Jun	2022	66.86374	21.86075589	111.86672	-1.962384	135.68986
Jul	2022	67.54827	20.73614686	114.3604	-4.044695	139.14124
Aug	2022	69.76192	21.18699725	118.33684	-4.527012	144.05084
Sep	2022	70.9037	20.59241191	121.21498	-6.040773	147.84816
Oct	2022	71.50063	19.49120633	123.51006	-8.040922	151.04219
Nov	2022	63.964	10.3090575	117.61894	-18.094153	146.02215
Dec	2022	58.48144	3.23697586	113.72591	-26.007679	142.97057
Jan	2023	65.06273	8.27914814	121.84631	-21.780263	151.90572

Feb	2023	68.45458	9.34129176	127.56787	-21.951393	158.86055
Mar	2023	70.35215	8.60996121	132.09433	-24.074377	164.77867
Apr	2023	70.41533	6.09120732	134.73945	-27.959926	168.79059
May	2023	69.70208	3.05757603	136.34658	-32.221892	171.62605
Jun	2023	69.14578	0.38605898	137.90549	-36.013136	174.30469
Jul	2023	69.40164	-1.36716605	140.17045	-38.82991	177.63319
Aug	2023	72.40092	-0.34034244	145.14218	-38.847239	183.64907
Sep	2023	74.75244	0.05952147	149.44536	-39.480522	188.98541
Oct	2023	78.37146	1.75914782	154.98377	-38.796959	195.53987
Nov	2023	75.68424	-2.80139565	154.16987	-44.34918	195.71765
Dec	2023	72.63488	-7.67404143	152.9438	-50.187016	195.45678
Jan	2024	77.58967	-4.49667909	159.67602	-47.950569	203.12991
Feb	2024	78.84891	-6.10659939	163.80442	-51.079329	208.77715
Mar	2024	80.63101	-7.63781882	168.89984	-54.364516	215.62654
Apr	2024	80.09131	-11.45486845	171.63748	-59.916486	220.0991
May	2024	80.15203	-14.33013304	174.6342	-64.34597	224.65004
Jun	2024	79.52653	-17.62446876	176.67752	-69.053099	228.10615
Jul	2024	78.80446	-20.8836841	178.49261	-73.655401	231.26433
Aug	2024	80.61782	-21.57120328	182.80684	-75.666803	236.90244
Sep	2024	82.57154	-22.10291927	187.24601	-77.514232	242.65732
Oct	2024	85.07047	-22.05649514	192.19744	-78.766087	248.90704
Nov	2024	81.2312	-28.29455774	190.75696	-86.273993	248.7364
Dec	2024	77.99029	-33.87377633	189.85435	-93.091034	249.07161
Jan	2025	82.87364	-31.27317256	197.02045	-91.698844	257.44612
Feb	2025	84.29102	-32.9266561	201.5087	-94.977946	263.55999
Mar	2025	86.2111	-34.37921463	206.80141	-98.215871	270.63807
Apr	2025	85.68128	-38.24986977	209.61244	-103.855059	275.21762
May	2025	85.62805	-41.39498355	212.65108	-108.636915	279.89301
Jun	2025	84.83434	-45.07817759	214.74686	-113.84971	283.51839
Jul	2025	84.83702	-47.85785465	217.5319	-118.10228	287.77633
Aug	2025	87.08015	-48.35935978	222.51967	-120.056707	294.21701
Sep	2025	88.85252	-49.30964124	227.01467	-122.44827	300.1533
Oct	2025	90.97297	-49.87745137	231.82338	-124.439158	306.38509
Nov	2025	86.14543	-57.34381246	229.63467	-133.302429	305.59329
Dec	2025	82.18051	-63.89320387	228.25422	-141.219957	305.58098
Jan	2026	87.66559	-60.94195625	236.27313	-139.610038	314.94121

<u>Table. 9 Global CRUDE WTI Prices Forecasted Values.</u>



<u>Fig 4.37 Graphical representation of Forecasted Data</u>

CHAPTER- V

5.1 Summary and Findings

CRUDE Oil they say is highly volatile commodity and no doubt it is because forecasting such commodity requires a presence of mind and deep knowledge about the commodity and related points to it. Researchers around the world with the deep machine learning to research about the CRUDE oil price fluctuations.

The present data rarely support fundamental oil prices or we are saying real-time data, instead, it's driven by externalities making our attempt for forecasting very challenging

CRUDE oil prices fluctuations can affect the economy of any nation whose requirement or demand is high.

There are several software's which are improving timely to give the exact result in forecast, such as Python, R, Tableau and several others. Each has its improvement and chosen for the better result to be got.

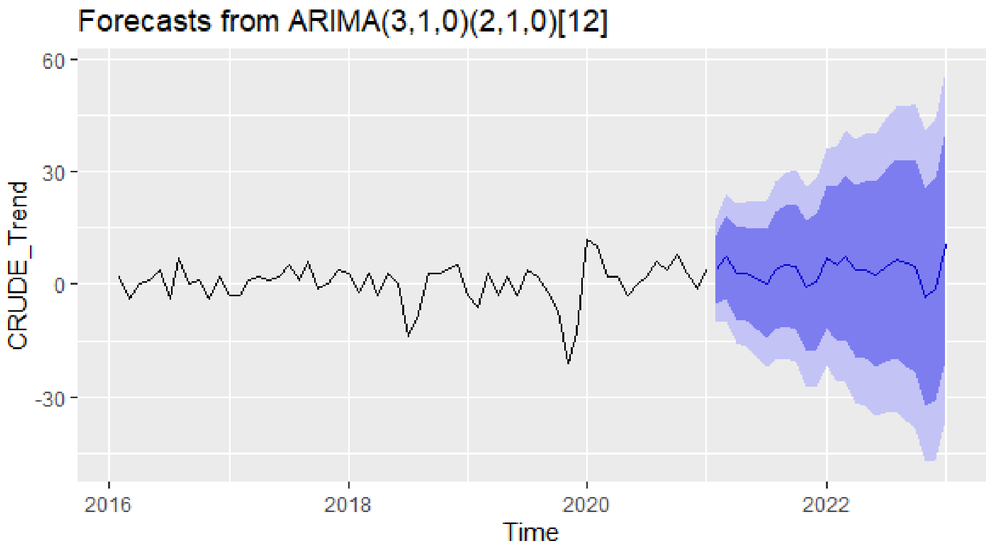
We learnt about the R programming how the data is imported and converted it to the time series so that the data should be stationary enough to forecast. Once the stationary data is plotted and trend is removed, all the three models are diagnosed to check the parameters which has having the best model to forecast. ARIMA model has shown the best result compare to other models with ARIMA (3, 1, 0) (2, 1, 0) [12].

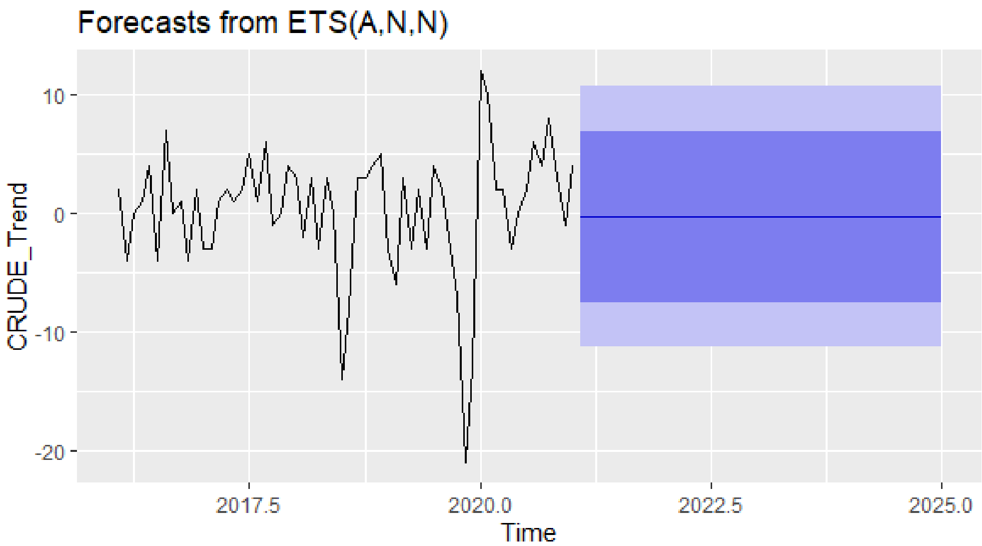
Naïve benchmarking and ETS model have not worked properly with the data because of standard deviation and the Akaike model value, which seem to be not valid as per the data.

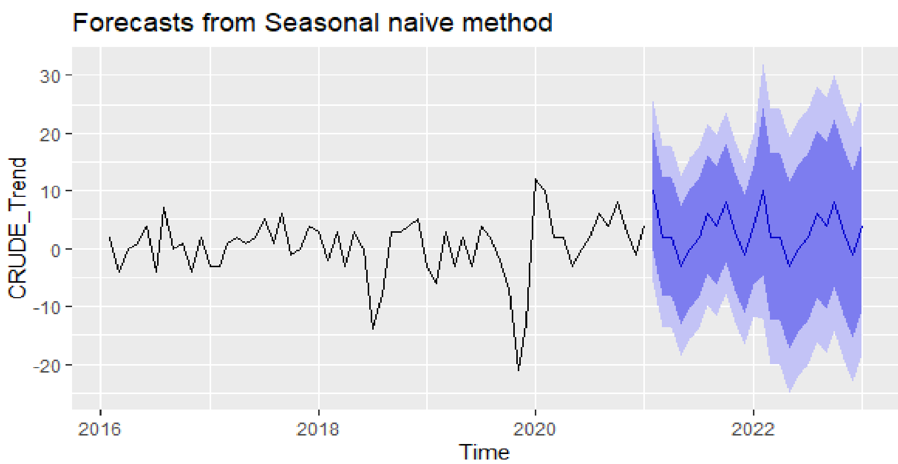
I can show the forecast model of the Naïve and ETS model as the findings to which is why ARIMA model is best for forecasting Crude oil data.

Suggestions–The R studio does gave the good result for forecasting but if we do forecasting from Python which is nothing but advanced R studio kind of software which can work with the same library packages of R but has a number of packages as compare to R and better visualization.

Comparison of the Forecasting Patterns







<u>Fig 4.38 Comparison of Three Models Graphs</u>

5.2 Conclusion

The movement of oil price seems to be uncertain and arbitrary, since its influential Factors are complex. However, in this report, we use data from 2011 to 2021 to. develop a reasonable model to describe and predict the average annual price of world crude oil with the aid of R software by extracting the statistic features. We also give a practical example. By comparison, the value of prediction is good too. Worth with actual data. The model we established has the preferable approach ability. and predication performance, particularly for the short-term forecast.

BIBLIOGRAPHY

1. https://fred.stlouisfed.org/series/POILWTIUSDM
2. https://otexts.com/fpp2/simple-methods.html
3. https://www.pluralsight.com/guides/time-series-forecasting-using-r
4. https://www.wiley.com/en-us/The+R+Book%2C+2nd+Edition-p-9780470973929
5. https://www.hindawi.com/journals/mpe/2014/201402/
6. https://www.monash.edu/business/marketing/marketing-dictionary/n/naive-forecasting
7. https://www.rpubs.com/shubh2565/pvalue-rsquared#:~:text=P%20values%20tell%20you%20whether,the%20null%20hypothesis%20is%20true.
8. https://sci-hub.se/https://doi.org/10.1016/j.ijforecast.2018.04.007
9. https://sci-hub.se/http://www.iaee.org/en/publications/ejarticle.aspx?id=2158
10. https://sci-hub.se/https://www.iaee.org/energyjournal/article/2158
11. https://sci-hub.se/10.1109/CSSR.2010.5773729
12. https://scihub.se/https://www.sciencedirect.com/science/article/pii/S0140988310000654