

# **Store Item Demand Forecasting**

Project report submitted in partial fulfillment of the requirement  
for the degree of Bachelor of Technology

in

**Computer Science and Engineering/Information Technology**

By

**Joshita Sharma (151442)**

Under the supervision of

**Dr. Ruchi Verma**

**Assistant Professor (Senior Grade)**

to



Department of Computer Science & Engineering and  
Information Technology

**Jaypee University of Information Technology Waknaghat,  
Solan-173234, Himachal Pradesh**

## Certificate

I hereby declare that the work presented in this report entitled “**Store Item Demand Forecasting**” in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** submitted in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from July 2019 to May 2020 under the supervision of **Dr. Ruchi Verma**.

The work done embodied in the report has not been appeased for the award of any other degree or diploma.

A handwritten signature in blue ink, appearing to read 'Joshita', with a horizontal line underneath.

Joshita Sharma  
(151442)

This is to certify that the above affirmation made by the candidate is true to the best of my knowledge.

A handwritten signature in blue ink, appearing to read 'Ruchi', with a horizontal line underneath.

Dr. Ruchi Verma  
Assistant Professor (Senior Grade)

Department of Computer Science and Engineering and Information Technology

## **Acknowledgement**

I would like to take the opportunity to thank and express my deep sense of gratitude to my mentor and project guide Dr. Ruchi Verma for her immense support and valuable guidance without which it would not have been possible to reach this stage of my final year project.

I am also obliged to all my faculty members for their valuable support in their respective fields which helped me in reaching at this stage of my project. My thanks and appreciations also go to my colleagues who have helped me out with their abilities in developing the project.

## TABLE OF CONTENT

<b>1. Chapter-1 INTRODUCTION .....</b>	<b>1</b>
1.1. Introduction.....	1
1.2. Problem Statement.....	2
1.3. Objectives.....	2
1.4. Methodology.....	2
1.5. Organization.....	2
<b>2. Chapter-2 LITERATURE SURVEY.....</b>	<b>14</b>
<b>3. Chapter-3 SYSTEM DEVELOPMENT.....</b>	<b>18</b>
3.1. Analytical.....	20
3.2. Mathematical.....	22
<b>4. Chapter-4 PERFORMANCE ANALYSIS.....</b>	<b>24</b>
4.1. Plot of daily, monthly & yearly sales.....	26
4.2. Check for Stationarity .....	28
4.3. ETS Decomposition.....	30
4.4. Grid Search.....	38

<b>4.5. Sales VS Prediction. ....</b>	<b>42</b>
<b>4.6. Final Predictions. ....</b>	<b>44</b>
<b>5. Chapter-5 CONCLUSIONS . ....</b>	<b>45</b>
<b>5.1. Conclusions. ....</b>	<b>45</b>

## **LIST OF ABBREVIATIONS**

- 1) ARIMA – Auto regressive integrated moving average**
- 2) LSTM – Long short term memory**
- 3) SARIMA – Seasonal auto regressive integrated moving average**
- 4) ACF/PCAF- Auto correlation function/ partial auto correlation function**

## **LIST OF FIGURES**

1. Train data and description
2. Sales of 10 stores w.r.t. years
3. Daily, monthly and yearly sales
4. ETS decomposition
5. Stationary plots
6. ACF/PACF plots
7. Prediction graphs

## **LIST OF TABLES**

1. Summary of fitted model (Monthly data)
2. Summary of fitted model (Daily data)



## **ABSTRACT**

Sales is a crucial part of any business. It can make or break a business. Without a powerful forecasting tool and right management, it is impossible to be competent in the industry. One can plan for stock management, inventory and advertisement of products at the right time.

With advancement in technology and data being so readily available in large amounts, it is very expensive or a complex job to get the useful information out of it by previous methods. Machine learning is needed to get the forecasts with more accuracy and with less work. Patterns to be detected are left up to algorithms. This is useful to stand your ground in industry now with less cost. Time consumption is minimised also as training is done by the machine and no manual effort is needed.

Again if one wants to start a new business, sales forecasting might be the most crucial aspect of success ahead. Moreover loans are a must in most business or capital is needed to buy all the machines, inventory they need to get off the ground. One will get nothing with a bright thought and a lot of enthusiasm when you have to go to the bank. Trust is vital to get huge amounts of capital and this is visible if a lot of investors are there. A business plan is all one needs. Thus, our project uses ARIMA model for store item forecasting and trains the model on it for give store item sales.

# **CHAPTER-1**

## **Introduction**

### **1.1 Introduction**

Forecasting helps us making business successful in the long run. It additionally helps one to check, to estimate and to analyse the info to hit the calculable results. It ends up in regular investigation of various aspects of production and management among and out of doors the organisation. Forecasting prepares a ground to figure along and brings higher co-ordination, co-operation and management within the organisation.

Under prognostication, future prospects, stability and also the discrepancies area unit properly weighed and studied. This helps the management to get rid of any hindrances that will are available in the means of management.

Thus, company results area unit compared with the calculable ones, the opposite component that is sort of conspicuous with prognostication. Whenever giant distinction is found, additional investigation is undertaken to search out the explanations for such a discrepancy.

Analysis of time series give us insight in the various aspects of underlying important patterns and breaks down different trends and periodicity in data. The increasing availability of huge amounts of data and firms striving to excel in business has made forecasting a difficult and complex task. Machine learning techniques reduces the cost of labour and automation leading to less time consumption.

Effort is made to predict one year sales of store item from given 5 years of daily sale data. Various trends, seasonal variation are considered for monthly, yearly plots of data items. Machine learning techniques are used to fit Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average models.

## **1.2 Problem Statement**

We are given time series as store item sales and have to forecast the values of sales for three months. Given data is of span of five years.

## **1.3 Objectives**

We will predict the sales for next three months. This will help the shopkeeper maintain an inventory and manage his finances. Also the shopkeeper can derive the marketing strategy.

## **1.4 Methodology**

### **1.4.1 Dataset Information**

- The dataset is taken from the Kaggle in the text form then it is further converted in the .csv form to make the data available in tabular form.
- The system is taking the data of past 5 years.
- Further we divide the data into parts, first training data i.e. 1 year of data. Second is testing data i.e. 4 years of data and the training and testing of the data is done with the cross- validation of five folds.

## **1.4.2 SOLVING THE PROBLEM**

Different machine learning classification algorithms are used to solve the problem and we have chosen the best model for the dataset of course of 5 years.

Parameters used to predict the best model out of the some models are:-

### **1.4.2.1 Correlation Coefficient**

A correlation coefficient is a numerical measure of linear interdependence of given variables. It is also defined for a set of data. The lower limit is -1 whereas upper limit is +1. It shows relative strength of give data sets or variables as well. A negative correlation coefficient means decrease in one variable leads to increase in value of other variable. A positive correlation coefficient means increase in one variable will lead to increase in other variable as well. The exact amount of change in one variable when other variable is increased or decrease is unknown but relative increase decrease can be specified.

### **1.4.2.2 Mean absolute error**

It is difference between the original and predicted value averaged over all the points in datasets. The difference in mean absolute error and root mean squared error is that it take cares of the direction of the error. For every observation in our dataset, we take absolute difference from original observation and get the summation. This is divided by total no of observations to get the result.

### **1.4.2.3 Root Mean squared error**

Root mean squared error is standard deviation of errors which occur when prediction is made. This is similar to mean squared error, only root is taken. Here errors are squared before getting average of them and followed by square root of them. Thus errors are not differenced and it accurately tells large errors in the prediction. This is why it doesn't take absolute error. It can tell model's accuracy if the value is low, model is good to go. Lesser the value, more accurate and close are values predicted to the original values in the future.

### **1.4.2.4 Relative absolute error**

Relative absolute error is another performance matrix used to determine if the predictive model is accurate or not. It is often used as relative error but they are different. Relative error is just to measure the effectiveness of scales or clocks

Relative absolute error is measured by current original observation minus the predicted observations. This is summed up and divided by summation of original values. Thus, result is obtained.

Absolute error is simple difference between prediction and actual observations whereas relative absolute error takes care of this difference by dividing by the original value.

### **1.4.2.5 Root Relative Squared error**

Root relative squared error is comparative with what it would have been if a basic indicator had been utilized. All the more explicitly, this basic indicator is only the normal of the real qualities. In this manner, the squared error is taken. This is further decided by square error of the main predictor. By taking the square foundation of the relative squared error one lessens the mistake to indistinguishable measurements from the amount being anticipated.

### **1.4.3 Language Used**

Python is a programming language which is general purpose, high level language and interpreted as well. Made by Guido van Rossum and first released in 1991, Python's style theory accentuates code meaningfulness with its outstanding utilization of genuine whitespace. Its language builds and article situated methodology intend to help developers compose clear, sensible code for little and enormous scope comes. Python is comparatively straightforward, thus it is easy to be told since it needs a singular syntax that focuses on readability. Developers will scan and translate Python code a lot of easier than different languages. In turn, this reduces the value of program maintenance and development as a result of it permits groups to figure collaboratively while not important language and skill barriers.

- Python could be a problem-oriented language, that means that it abstracts underlying computer-related technical details. for instance, Python doesn't create its users assume an excessive amount of concerning hardware management or correct declaration of variables and uses safe assumptions concerning what the software engineer is attempting to convey. additionally, a high-level,

language may be expressed in a very manner nearer to English prose or mathematical equations. Python is ideal for literate programming due to its light-weight, “low ceremony” nature.

- Python could be a all-purpose language that means that it may be used for all issues that a pc is capable of instead of specializing in a very specific space like applied mathematics analysis. for instance, Python may be used for each computer science and applied mathematics analysis. Python may be used for a spread of heterogeneous tasks inside a given work-flow, because the UCAR person delineate earlier.

- Python is associate degree understood language that means that analysis of code to get results will happen straight off instead of having to travel through a long, compile and run cycle, that thereby hurries up the thinking and experimentation processes. IPython is associate degree interactive kind of the Python language conjointly fictitious by Fernando Pérez. These environments stand out for rapid-prototype of code or fast and straightforward experimentation with new concepts.

Python contains a customary library, and diverse third-party libraries yielding an enormous array of existing codebases and examples for determination issues.

Python has several, several users which implies that programmers will quickly notice solutions and example code to issues with the assistance of Google and Stackoverflow.

These options, perhaps, keep company with a minor price of reduced language performance, however this can be a trade-off the overwhelming majority of users square measure willing to create so as to realize all the benefits Python has got to provide.

In addition, Python contains a made scheme for scientific inquiry within the kind of several proved , and well-liked ASCII text file packages including:

- numpy, scientific computing is done by this package.
- matplotlib, 2D plotting library which provides high quality figures
- cartopy, cartographic tools for Python is provided by this library
- netcdf4-python, a Python interface to the netCDF-C library

## 1.4.4 Tools Used

### 1.4.4.1 Jupyter Notebook

Jupyter Notebook is associate ASCII text file internet application that permits North American nation to make and share codes and documents.

First, in giant companies, finding a selected task may need committal to writing in an exceedingly few languages, say Python, R, Julia, or PHP. rather than putting in completely different interfaces for every language kernel you would like, Jupyter permits you to use an equivalent structure of the notebook style of file. Simply, every notebook you produce can hook up with the language kernel you request. think about conjointly, this file will be simply keep regionally or on a far off server. Therefore, Jupyter facilitates the communication between groups in an exceedingly corporation hugely.

Second, Jupyter isn't a text editor that opens a brand new window on every occasion you execute a special a part of your code, as is that the case with another software system applications. within the same file, you'll have pure text which will communicate a message to the reader, coding system like Python, and output containing wealthy text, like equations,



figures, graphs, pictures, and others. This simplifies the method of the progress vastly, and Jupyter Notebook is progressively most well-liked over alternative software system packages. That's why we'll use it too.

It provides associate atmosphere, wherever you'll document your code, run it, examine the result, visualize knowledge and see the results while not departure the atmosphere. This makes it a handy tool for playing finish to finish knowledge science workflows – knowledge improvement, applied mathematics modeling, building and coaching machine learning models, visualizing knowledge, and many, several alternative uses.

Jupyter has more to offer than just what has been mentioned till now. Let's list some advantages of Jupyter for developers to start using it as a platform for their projects and to enhance their skills with ease.

Some of the Advantages are:

- Everything at same place: The Jupyter Notebook might be an electronic intuitive environment that blends code, rich content, pictures, recordings, liveliness, numerical conditions, plots, maps, intelligent figures and gadgets, and graphical UIs, into one report.
- Easy to Transfer: They are saved as files with JSON format which is easily accessible to other also and can be managed by other party you wish to share with too.
- Easy to convert: Jupyter accompanies an uncommon apparatus, nbconvert, that changes over scratch pad to various organizations like HTML and PDF. Another on-line apparatus, nbviewer, licenses US to render an openly accessible note pad legitimately inside the program.

- Language freelance: There is no language specification or any language that you need to learn to be able to use it. This makes it a more friendlier and powerful tool.
- Easy to make kernel wrappers: There is interface for kernel languages that is written in python. There will be options to implement features for code review and completion.
- Easy to customize: It's easy to work with Jupyter as it has wrapped kernel and various features to work on and console could be run like many other applications.
- Extensions with custom magic commands: These are the features which make Jupyter more attractive. These magic commands can be used like shortcuts to include or exclude codes directly from other notebooks or so.
- Stress-free duplicable experiments: One can leave the task of computing experiments to Jupyter as it can manage the work itself and be free of any stress of not documenting and keeping record of your files.
- interactive add notebooks, place them underneath version management. Also decide the repetitive code to work with it with ease in future.
- Successful instructing cum-learning apparatus: The Jupyter Notebook isn't exclusively a device for research and data examination anyway furthermore a decent device for educating. partner degree model is IPython Blocks – a library that grants you or your understudies to make frameworks of vivid squares.
- Interactive code and data investigation: The ipywidgets bundle gives a few basic program controls to investigating code and data intuitively.

### 1.4.4.2 Colab

Google Colab is free cloud service that can be used to anything ranging from running your machine learning project to coding practice or improve skills and use extensive libraries like Tensor flow and pandas.

One of the most vital feature that distinguishes Colab from different free cloud services is: Colab provides GPU and is completely free.

In Colab, it is similar to using Google Docs, one can manage large size project easily and share with people also. One can work on machine learning libraries and build projects and manage them efficiently. Github files can be transferred too, it is the greatest platform to put open source projects and work on them as a team.

Kaggle, which a well known platform to get big data and work on your machine learning skills is also included where Kaggle files can be shared and transferred and could be worked upon interesting challenges in Colab.

Some benefits of Google Colab includes:-

- GPU support is provided free of cost.
- Just as Google Docs are used, one can share and manage the Jupyter Notebooks as well.
- Supports Bash commands are also supported.
- Libraries like Pandas, Tensorflow and more are already installed on it, making it a perfect platform to start your project in it with these extensive machine learning libraries.
- To match the business standards or if you're period of time and memory intense codes, it's out of the question to put in expensive GPUs on your personal machines. To the simplest currently my data, nowadays, exploitation Colab, every notebook will use free GPU.

- NVIDIA Tesla K80 (~ price £3500.00) in Colab atmosphere, freelance of what personal machine you're exploitation. to boot, if you're ineffective to finish program or task, you'll be able to complete it from operating anyplace simply by sign language into the Colab account.
- Google Colab provides intrinsic version dominant system exploitation lowlife and it's quite straightforward to form notes and documentations, includes figures, and tables exploitation markdown.
- It is loaded with Google Docs features. One can use them with Collab, therefore making it a more powerful tool to process your project with.
- One does not need to check for updates that typically produce problem you're exploitation completely different in operation systems like macintosh, Windows, and Linux. Google servers are the where it runs.
- Last however not least, these days in business, the foremost common necessities within the job packs are data of programming deep learning models exploitation GPUs.
- Link to your GitHub profiles. The GitHub links are helpful for recruiters. One can simply showcase your work from anyplace round the globe and recruiters can explore the work done by you, that facilitate them in distinguishing robust candidates for the roles.

#### **1.4.4.3 Excel(.csv file)**

Microsoft Excel is programme developed by Microsoft for Windows, macOS, humanoid and iOS. It has many features such as fast calculations and already tabular form. One can use it to write weekly chore schedules, plans, loan calculators, attendance reports, project schedule,

loan amortization schedules and for basic financials plans and regression. It's easy to apply formulas to the cells as one wish to. Once you've written something and need it to repeat, there is no need to repeat but just drag the cursor till you've reached your final cell. It has gained popularity as it is very easy to use and learn and serves great purpose.

Pros:

- Easy to use
- worksheets can be multiple
- linking between worksheets
- Great for calculations
- Table operate permits you to filter information handily
- Has tight graphics, graphs, charts, etc. to indicate your information
- Almost anyone will use excel for proper storage in an efficient manner.

### **1.4.5. Output**

We have successfully predicted sales for 3 months by training data of one year and testing on remaining 3 years of given data.

### **1.4.5. Organization**

In the first chapter we have given the introduction of the project and the future expectation from the project that it can under go with.

In the second chapter we have given the information about the literature that we have studied to complete the project and we have also included the future work that we have to undergo to complete to make the project future ready.

In the third chapter we have provided the information about the system development that will help in understanding the development of the project that through which processes it has gone through.

In the fourth chapter we are analysing the performance of the of the system developed and we are comparing the result that we have got from the development of the system.

In the fifth chapter we are including the conclusion of the project that we have got from the work that we have done and what are the scopes of improvement that we can go through.

## **CHAPTER 2**

### **LITERATURE SURVEY**

#### **2.1 Stock Price Prediction Using the ARIMA Mode**

**2.1.1 Author :** Adebisi A. Ariyo, Adewumi O. Adewumi, Charles K. Ayo

**2.1.2 Publication :** March 2014

##### **2.1.3 Summary :**

In finance, Stock Price prediction is a vital aspect. Author explained machine learning models for sales predictive analytics. Author have used uses autoregressive integrated moving average model to successfully forecast stock prices over time. Zenith bank stock index time series is forecasted by ARIMA (1, 0, 1) which fitted best to the data. Machin learning has deduced and used the underlying patterns to predict the future data efficiently. He concluded how ARIMA model is favourable tool for prediction of time series without any seasonal componet and for a short range of time. This appraoch is competitve enough with other existing tools as well.

##### **Advantages :**

- Unlike regeression ARIMA doesn't need a group of variable solely the target variable will be one of it's duty.

- The main advantage of ARIMA prediction is that it needs information on the statistic in question solely. First, this feature is advantageous if one is prediction an oversized range of your time series. Second, this avoids a drag that happens typically with variable models.

**Disadvantages :**

- The ARIMA model yields results that are not distinctive because of their non-linearity. In reality, there is also some dependency between the variables involved.
- Times series in reality is expected to have some irregular rise and fall in observations which is regarded as residual error and not taken as trend or a predicting factor. This leads to forecasting error increase with time.

## **2.2 ARIMA MODEL FOR FORECASTING OIL PALM PRICE**

**2.2.1 Author :** Rangsan Nochai , Titida Nochai

**2.2.2 Publication :** June 2006

**2.2.3 Summary :**

Nowadays, the enhanced oil costs worldwide are having an excellent effect on every economic activity. Several countries, Siam rely heavily on oil for the supply of energy. The govt. has campaigned not just for energy conservation, however conjointly for the employment of different sources of fuel available. Therefore, This paper covers prediction of palm oil prices for 5 year span i.e. 2000 to 2004. Three



varieties of palm oil price is described as farm, wholesale and market price. Here mean absolute relative error is minimized to fit the best model for data.

Main focus is on to search out acceptable ARIMA Model for predicting in 3 sorts of palm value.

## **2.1 Rainfall forecasting by technological machine learning models**

**2.1.1 Author :** Wei-Chiang Hong

**2.1.2 Publication :** June 2008

**2.1.3 Summary :**

Accurate prediction of downfall has been one among the foremost necessary problems in hydrological analysis. because of downfall prognostication consists of a complicated information pattern that are not linear in nature; there are millions of novel prognostication methods for boost the accuracy. Perennial artificial neural networks have vie an important part in prognostication downfall information. Whereas, support vector machines (SVMs) are with success used to unravel nonlinear regression and statistic issues. This paper exposes the feasibleness of hybrid model of RNNs and SVMs combined, particularly RSVR, to predict downfall depth values. Chaotic particle swarm improvement rule (CPSO) is utilized to decide on the parameters of a Foreign Intelligence Service model. Afterwards, example of downfall values throughout hurricane periods of Northern Taiwan is employed let's say the

projected RSVRCPSO model. The final results show that the projected model has well forecasting performance, RSVRCPSO model provides a great way to predict downfall values.

**Advantages :**

They are often combined along so as to allow you the advantages of every approach. statistic will a decent job at mouldering information into trended and seasonal components. This associate analysis will then be used as an input into a machine learning model, which might incorporate the trend and seasonal info into its algorithmic rule.

**Disadvantages :**

- Machine Learning needs large information sets to coach on, and these ought to be inclusive/unbiased, and of fine quality. There also can be times wherever they have to look forward to new information to be generated.
- It conjointly wants large resources to operate. this could mean extra needs of laptop power for you. Training is important in machine learning tools and it take more time and wide range of data to develop a model that accurately forecast for longer period of time

## **CHAPTER 3**

### **SYSTEM DEVELOPMENT**

#### **3.1 Algorithm**

##### **Forecasting with Prophet**

Prophet can be considered as a mechanism for statement statistic information supported.

It is a very powerful and efficient toll when it comes to seasonality whether it's on daily or yearly basis. It accurately depicts the underlying trends and take into account the past information in the data given. It also usually handles outcasts well. It performs better as holidays and major events are taken into consideration.

Facebook's Core information Science team has provided this open source package. It's obtainable for transfer on cubature unit and PyPI.

- **Accurate and fast.**

Prophet is employed in several applications across Facebook for manufacturing reliable forecasts for designing and goal setting. We've found it to perform higher than the other approach within the majority of cases. we have a tendency to match models in Stan in order that you get forecasts in mere many seconds.

- **Fully automatic.**

There's no manual effort required. Missing data, and sudden changes in your time series can be easily manageable.

- **Tunable forecasts.**

The Prophet procedure includes several potentialities for users to tweak and modify forecasts. One can make use of human-interpretable parameters for boosting his forecast by adding the domain information.

- **Available in R or Python.**

Any of the above language can be used as it supports both and is reliable.

### **A brief summary of how Prophet works**

Additional model is called Pophet. It can be represented as follows:

$$Y(T) = G(T) + S(T) + H(T) + \epsilon$$

- $G(T)$  represents increasing or decreasing observation i.e. trend. This rise or fall in values is shown over along interval of time. Prophet can be described by models such as linear model and saturation growth model. The type of model to use is determined by the give data itself and it's nature.
- $s(t)$  represents the periodic nature of the time series and how it gets affected by the season like summer vacation etc.
- $h(t)$  models denotes the underlying relation between the given data and the main events in the zone like new year or holidays.
- $\epsilon$  as the symbol is represents error.

## **Multivariate Time Series Forecasting with LSTMs in Keras**

LSTMs are used for multi-step time forecasting. Following are the advantages :-

- In most forecasting tools, lagged data or past observations are taken as an input but in LSTMS, even sequence is accepted as an input.
- Multiple parallel input sequences are supported by LSTMs for multivariate inputs, in contrast to other models where flat structure has to be presented for multivariate inputs.
- Vector Output is given by LSTMs. In contrast to other methods, it maps sequence data simultaneously to output vector. This consists of various steps to output.

Following steps can be used:

- Transform a raw dataset into usable format for time series foretelling.
- LSTM takes a time series problem and fitted to it to forecast.
- It predict future and previous genuine units are not lost.

### **Autoregressive Integrated Moving Average Model**

An ARIMA model is one of the oldest statistical models for forecasting. ARIMA stands for AutoRegressive Integrated Moving Average. ARIMA is widely used category of applied mathematics models for foretelling and analyzing statistic knowledge.

It expressly caters to a collection of ordinary structures in statistic knowledge, and in and of itself provides an easy however powerful technique for creating practiced statistic forecasts.

Let's break down the different terms of ARIMA as follows:-

**AR stands for Autoregression.** This is a model based on regression where we predict future values with the help of dependent variables. In this case, it is dependency of current values with past values. These are called lagged values

**I is used for Integrated** where we difference the observations from its lagged times series which is the time series stationary. This is an important step as only after that a linear model like ARIMA can be applied. One needs no dependency in the variables involved in the forecasting step.

**MA stands for Moving Average.** This is a model integrated in ARIMA where prediction is based off on residual error as we predict the future values. This model is also applied to lagged values.

Next step is to determine the parameters used in ARIMA which are  $p, q$  and  $d$ . ARIMA is denoted by  $(p, q, d)$  to specify parameters of ARIMA (or SARIMA) model being used.

The meaning of these parameters are related to the parts of ARIMA mentioned above.

**p:** is called lag order. This is no. of past values or total terms we are using in our Autoregression model to predict the future value.

**d:**, Also known as degree of differencing. We take current time series plot and another same times series plot with lag in terms of time. This lag is the value of differencing. This is done to make it stationary which is essential.

**q:**, is known as order of moving average. This is no. of terms in our second model based on residual errors.

## Seasonal Autoregressive Integrated Moving Average Model

SARIMA is seasonal ARIMA. It can be applied to a time series where seasonal factor is prominent. This takes care of the seasonal component which is a limitation of ARIMA.

It has four parameters which are very similar to the ARIMA parameters. They are all written mostly in capitals i.e. P, Q, D and m. These parameters are used for seasonal component of the time series.

The four seasonal parameters that are added to ARIMA which are not included in ARIMA and act as an extension. They are as follows:

- P is order of autoregressive model used for seasonal component.
- D is differencing order of seasonal component.
- Q is no of terms in seasonal moving average model used.
- M is time interval after which periodicity is seen.

These parameters can take various values which can also be added to Seasonal ARIMA models. Thus, it is advisable to brute force various sets of parameter's values so as to see which model fit the given data best.

Therefore SARIMA model is given as:-

$SARIMA(p,d,q)(P,D,Q)m$

Usually  $ARIMA(p,d,q) \times (P,D,Q)$  model is the right representation of a seasonal ARIMA model. To be more precise, parameters denote following :-

P is count of seasonal autoregressive terms involved,

D represents differences required to make it stationary,

Q represents count of moving average as in ARIMA model for raw observations and lagged terms. Here it is for seasonal component.

First step to identify the seasonal model's parameter is differencing. We check if we need some seasonal difference and set the value of periodicity. It is recommended to see by brute force i.e. plot and ACF graphs to try some permutations of 0 or 1 value of both seasonal as well as difference pertaining to non-seasonal component in the model. Commonly, it is not advised to use more than one difference which is seasonal. Also, never exceed it to three or greater differences which means seasonal and non-seasonal together.

If periodicity is clear in some span of time then seasonal difference is a must whereas non-seasonal difference can be included or not. This works better in long time forecasts as they slowly manifest themselves over time and will give good results as expected of ARIMA models.



# CHAPTER-4

## PERFORMANCE ANALYSIS

DATASET:

- **Train Data:** store item sales of one year to test

```
[ ] data.head()
```

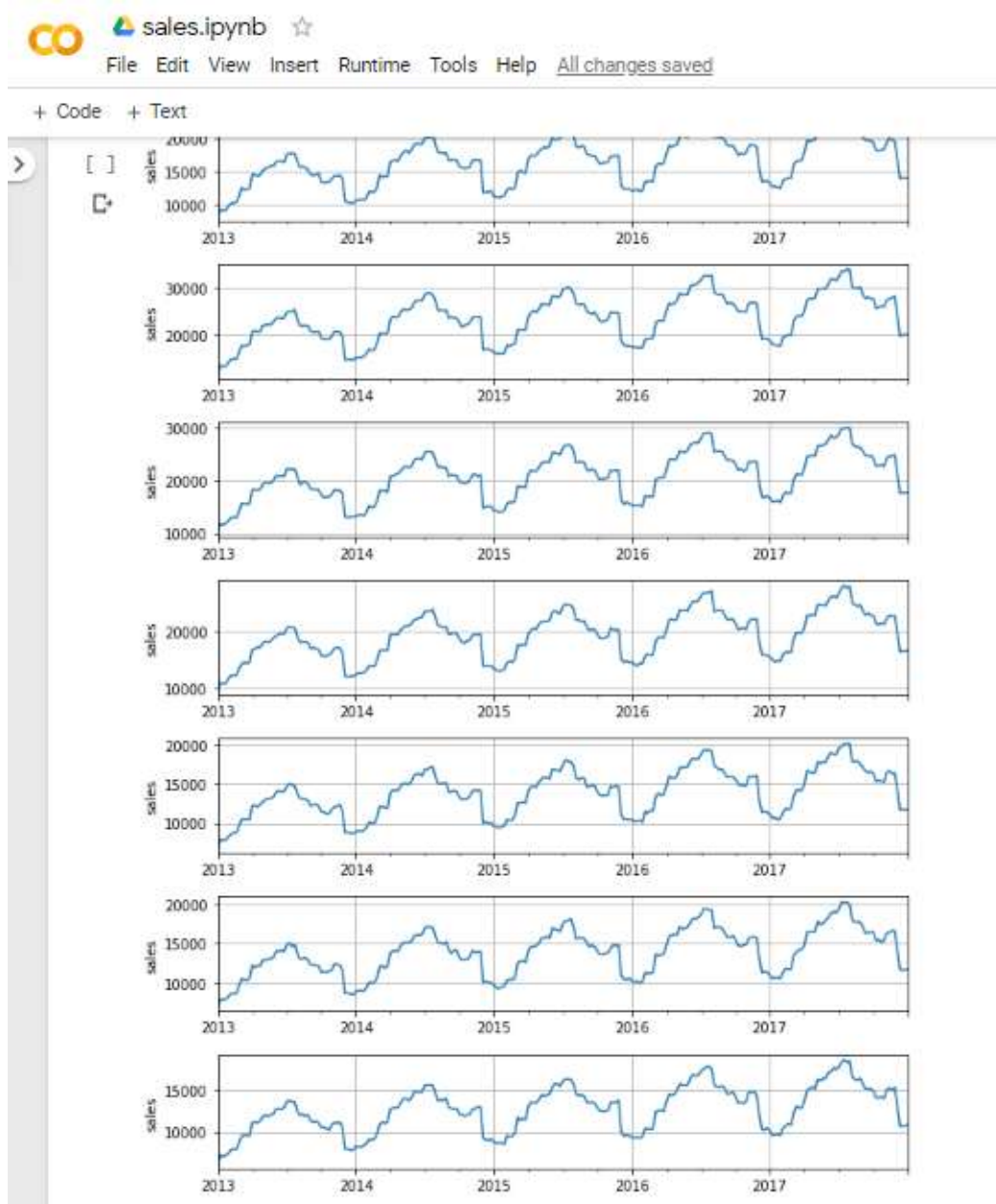
	store	item	sales	year	month	day
date						
2013-01-01	1	1	13	2013	1	1
2013-01-02	1	1	11	2013	1	2
2013-01-03	1	1	14	2013	1	3
2013-01-04	1	1	13	2013	1	4
2013-01-05	1	1	10	2013	1	5

```
[ ] data.describe()
```

	store	item	sales	year	month	day
count	913000.000000	913000.000000	913000.000000	913000.000000	913000.000000	913000.000000
mean	5.500000	25.500000	52.250207	2015.000548	6.523549	15.727820
std	2.872283	14.430878	28.801144	1.414021	3.448535	8.799329
min	1.000000	1.000000	0.000000	2013.000000	1.000000	1.000000
25%	3.000000	13.000000	30.000000	2014.000000	4.000000	8.000000
50%	5.500000	25.500000	47.000000	2015.000000	7.000000	16.000000
75%	8.000000	38.000000	70.000000	2016.000000	10.000000	23.000000
max	10.000000	50.000000	231.000000	2017.000000	12.000000	31.000000

- **Test Data:** sales in terms of specific id

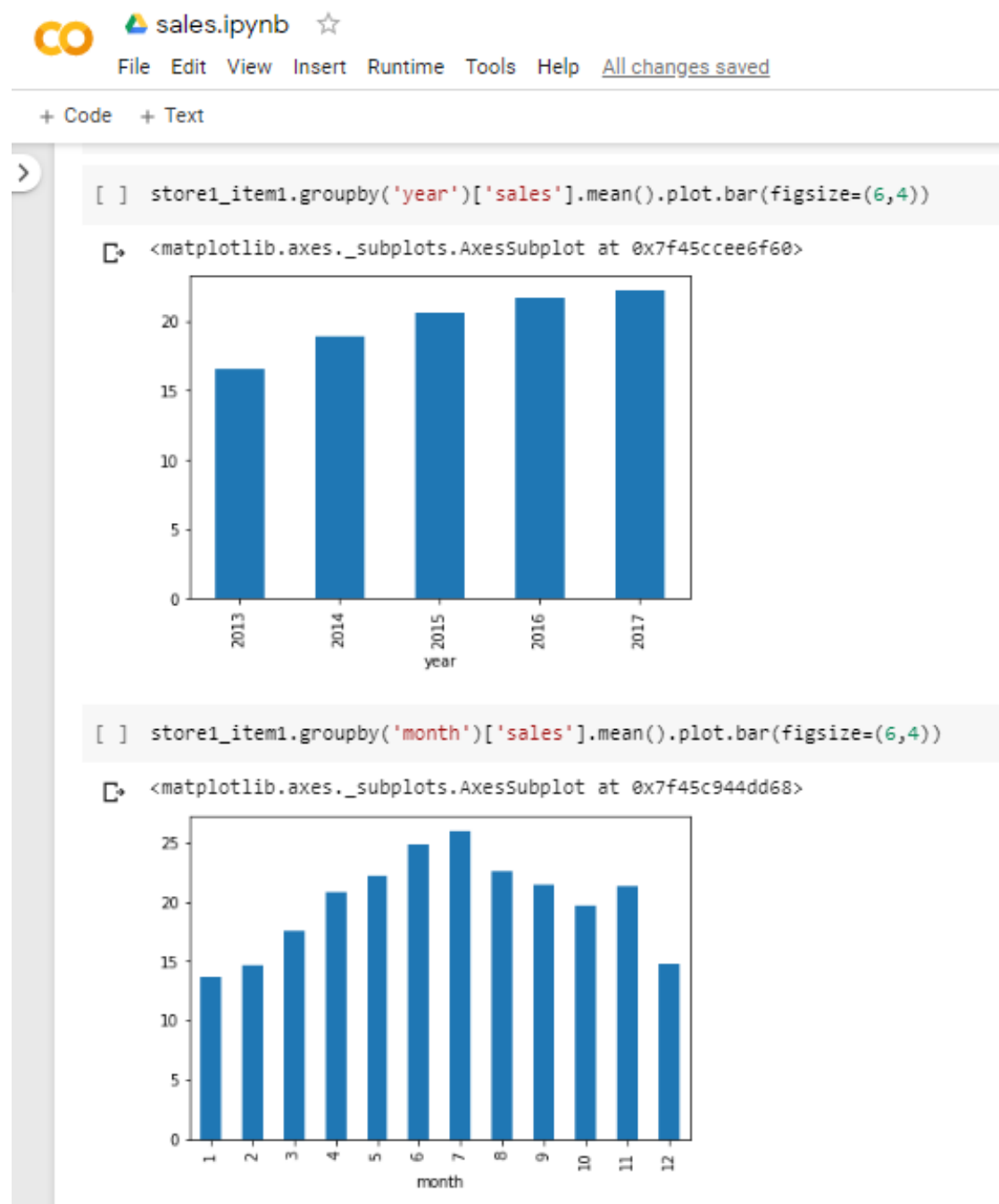
## PLOT SALES FOR 10 STORES WRT YEAR



By looking at the graphs we can say that all stores show same seasonality and trend. So, we make our predictions taking 1 store and 1 item and then apply to all 10 stores.

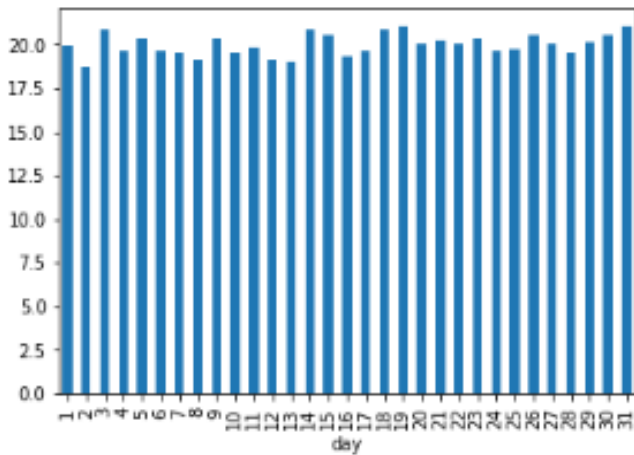
We need to further plot the data and see how sales varies daily, monthly and yearly. This will help us in analysis and we can decide how data can be grouped and predictions can be made.

## 4,1 PLOT OF DAILY, MONTHLY AND YEARLY SALES



```
[ ] store1_item1.groupby('day')['sales'].mean().plot.bar(figsize=(6,4))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f45c932ecf8>
```



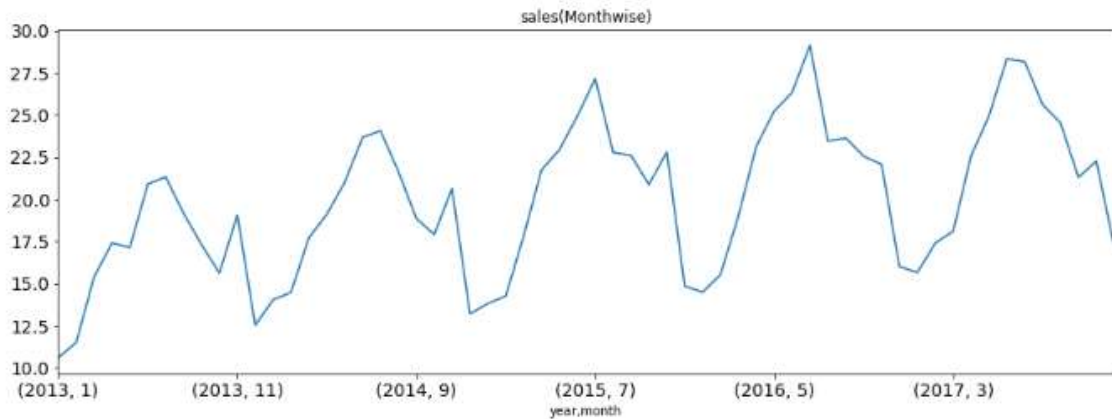
sales.ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

+ Code + Text

```
[ ] temp=store1_item1.groupby(['year', 'month'])['sales'].mean()  
temp.plot(figsize=(15,5), title= 'sales(Monthwise)', fontsize=14)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f45c93acc00>
```



From the above observations we can infer that the sales increases with year, which means there is a trend. Also, the sales is high in the months of June and July, which indicates seasonality.

To get a better understanding we do ETS decomposition.

## **4.2 ETS DECOMPOSITION**

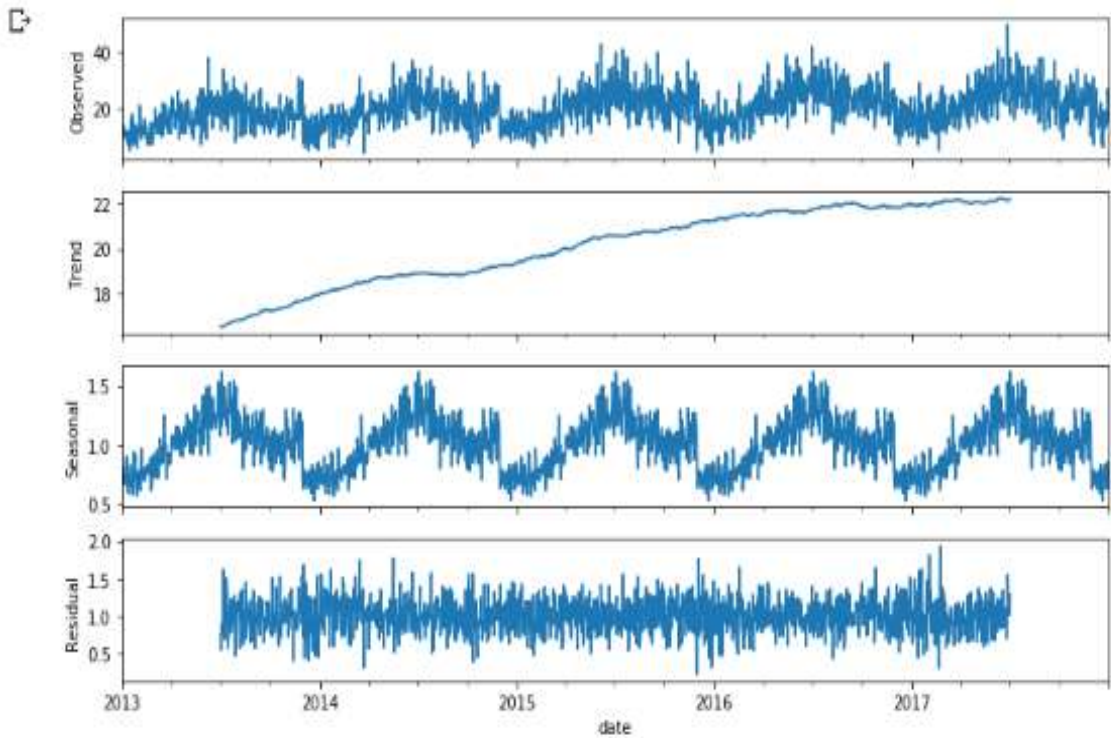
Time series decomposition is a very old mathematical procedure which helps splitting the time series into various different time series. It means extracting trend and all the seasonality components.

The original time series is usually split into:

- **Seasonal:** These are patterns which repeat within fixed periods of time like sales of candles can skyrocket in festival season.
- **Trend:** This shows whether sales are increasing or decreasing for a certain period of time. Sales of a popular products always increases with time meaning upward trend.
- **Random:** This is left behind after above splits. It can be noise or very unnoticeable rise or falls.

## ▼ ETS DECOMPOSITION

```
[ ] plt.rcParams['figure.figsize'] = (10, 6)
fig = seasonal_decompose(store1_item1['sales'], model='multiplicative', freq=365).plot()
```



We can deduce from above plots by ETS decomposition that sales definitely shows seasonality somewhere in middle of year and there is increase in sales i.e. upward trend.

The last step here is to check for stationarity. To fit an ARIMA model, it is a necessity to make the data stationary.

### 4.3 CHECK FOR STATIONARITY

The easiest method to determine time dependent structure or not is directly by eye balling the plots and decide whether stationary or not stationary.

**Check by looking at Plots:** This is the most easiest method but not accurate. It gives insight into whether there is seasonality or not which are easy to notice. Trend can be detected also.

**Summary Statistics:** Reviewing the summary statistics is one of the standard ways and notice differences.

**Statistical Tests:** These are the most accurate method to check stationary in a strict manner like Fuller test.

We use most common **Augmented Dickey Fuller test(ADFT)**

It is used to check whether given time series is stationary or not. For a times series to be stationary, it must have constant mean as well as covariance. It compromises of unit root test which suggest if it has a time dependent structure or not.

Following are the hypothesis mentioned in this test.

**Null Hypothesis** suppose a time dependent structure and deduces that the time series is non-stationary

**Alternate Hypothesis** means the null hypothesis is failed. Thus we have time independent structure and stationary time series.

Results are inferred from p value which is the result of the test. Common upper limit of p value is 5%. A lower value than 5% means it has no time dependent structure rejecting the null hypothesis. Thus times series is stationary. To be precise value should be very less than limit and also not close to it, lower the value better is the parameter for ARIMA.

**p-value is greater than 5%:** Time series is non-stationary and we have to make it stationary. Here we reject the null hypothesis.

**p-value is very less than 5%:** Time series is stationary and no differencing is needed. Thus prediction models like ARIMA can be applied now.



## ▼ CHECK FOR STATIONARITY

```
[ ] def test_stationarity(timeseries):

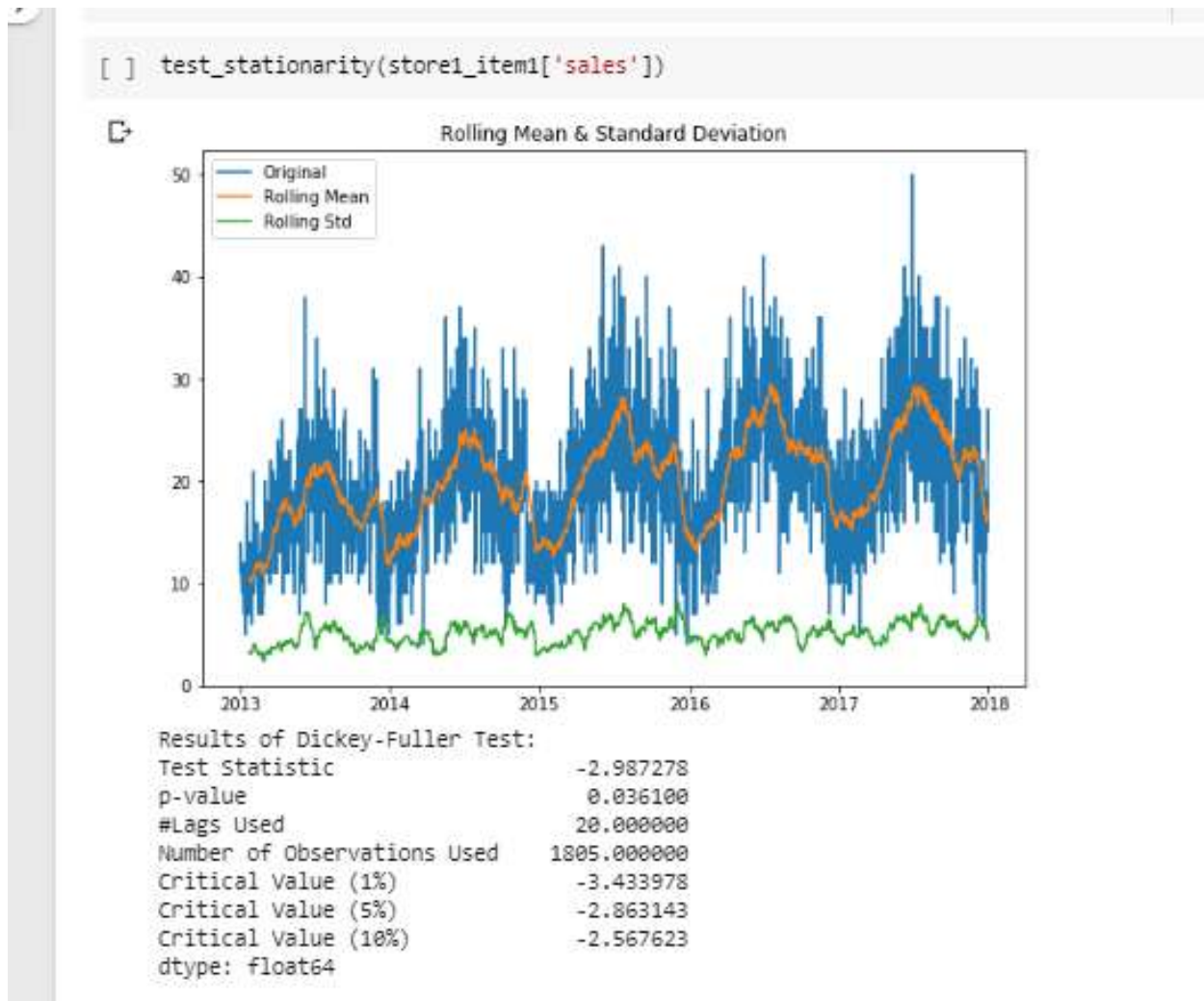
    rolmean = timeseries.rolling(window=24).mean()
    rolstd = timeseries.rolling(window=24).std()

    fig = plt.figure(figsize=(9, 6))
    orig = plt.plot(timeseries, label='Original')
    mean = plt.plot(rolmean, label='Rolling Mean')
    std = plt.plot(rolstd, label = 'Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling Mean & Standard Deviation')
    plt.show()

    print('Results of Dickey-Fuller Test:')
    dftest = adfuller(timeseries, autolag='AIC', maxlag = 20 )
    dfoutput = pd.Series(dftest[0:4], index=['Test Statistic', 'p-value', '#Lags Used', 'Number of Observations Used'])
    for key,value in dftest[4].items():
        dfoutput['Critical Value (%s)'%key] = value

    print(dfoutput)
```

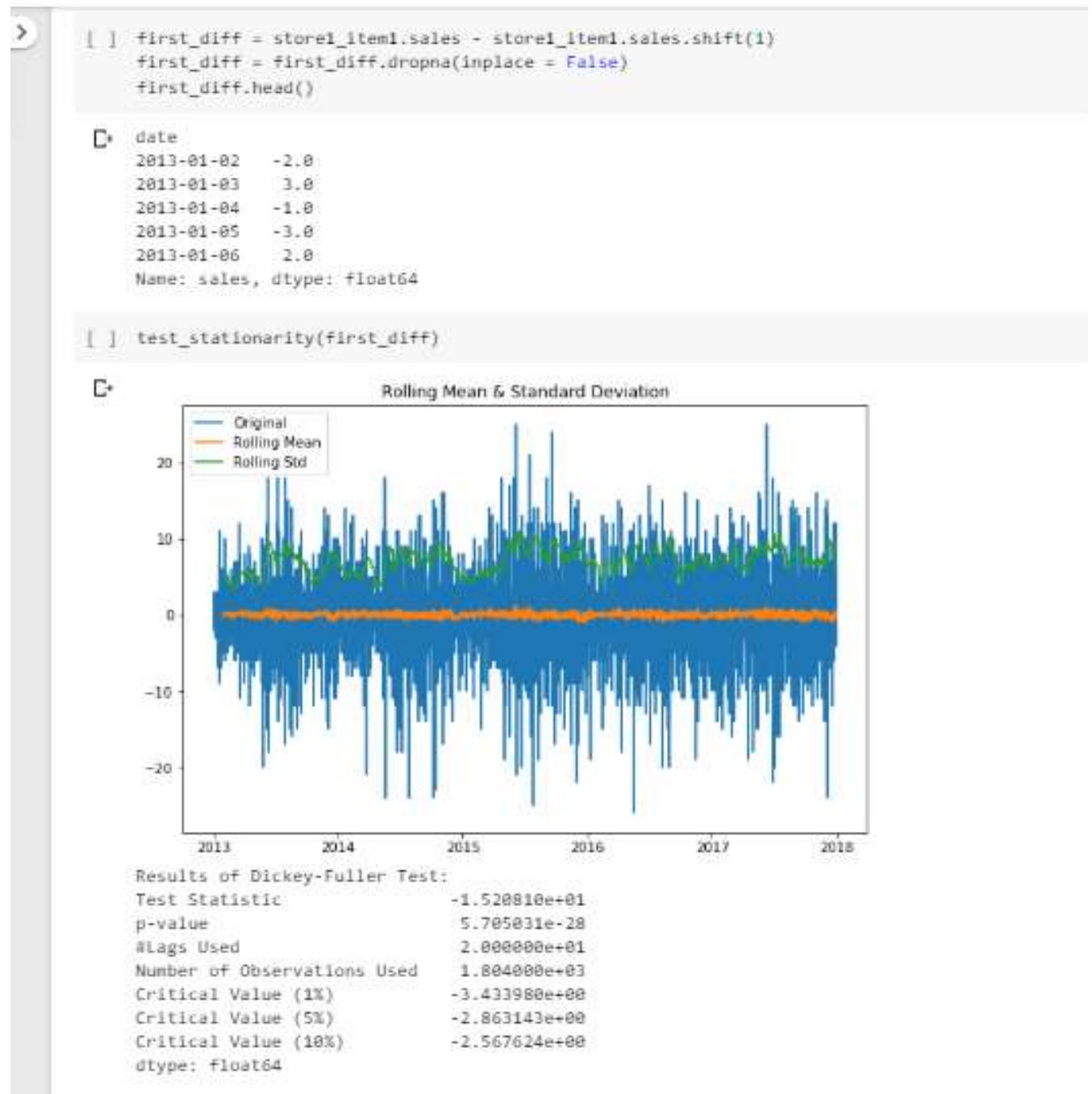
## RESULT



From the p value and critical values we observe that the sales data given to us is not stationary. It should have constant mean as well as variance to fit in the model.

We need to make both mean and variance constant to make future predictions. We can observe it through the p value and critical value.

In our case, we take first difference. Find the values and plot them. This gives us our desired result and gives us constant mean and variance as observed below. Now we can move further with the prediction part.



## APPLYING ALGORITHM (ARIMA AND SARIMA)

ARIMA model is widely used category of applied mathematics models for analysing and foretelling statistic knowledge.

It expressly caters to a collection of ordinary structures in statistic knowledge, and in and of itself provides an easy however powerful technique for creating practiced statistic forecasts. It consists of regression model integrated with moving average model.

Let's break down the different terms of ARIMA as follows:-

**AR stands for *Autoregression*.** This is a model based on regression where we predict future values with the help of dependent variables. In this case, it is dependency of current values with past values. These are called lagged values

**I is used for *Integrated*** where we difference the observations from it's lagged times series which s the time series stationary. This is an important step as only after that a linear model like ARIMA can be applied. One needs no dependency in the variables involved in the forecasting step.

**MA stands for *Moving Average*.** This is a model integrated in ARIMA where prediction is based off on residual error as we predict the future values. This model is also applied to lagged values.

Next step is to determine the parameters used in ARIMA which are p,q and d. ARIMA is denoted by (p,q,d) to specify parameters of ARIMA( or SARIMA) model being used.

The meaning of these parameters are related to the parts of ARIMA mentioned above.

**p:** is called lag order. This is no. of past values or total terms we are using in our Autoregression model to predict the future value.

**d:**, Also known as degree of differencing. We take current time series plot and another same times series plot with lag in terms of time. This lag is the value of differencing. This is done to make it stationary which is essential.

**q:**, is known as order of moving average. This is no. of terms in our second model based on residual errors.

To effectively find the p,d,q parameters we can do ACF/PACF plots or do a genral grid search putting an algorithm.

Here we have done both the methods and chosen the nearest value.

#### ACF/PACF PLOTS

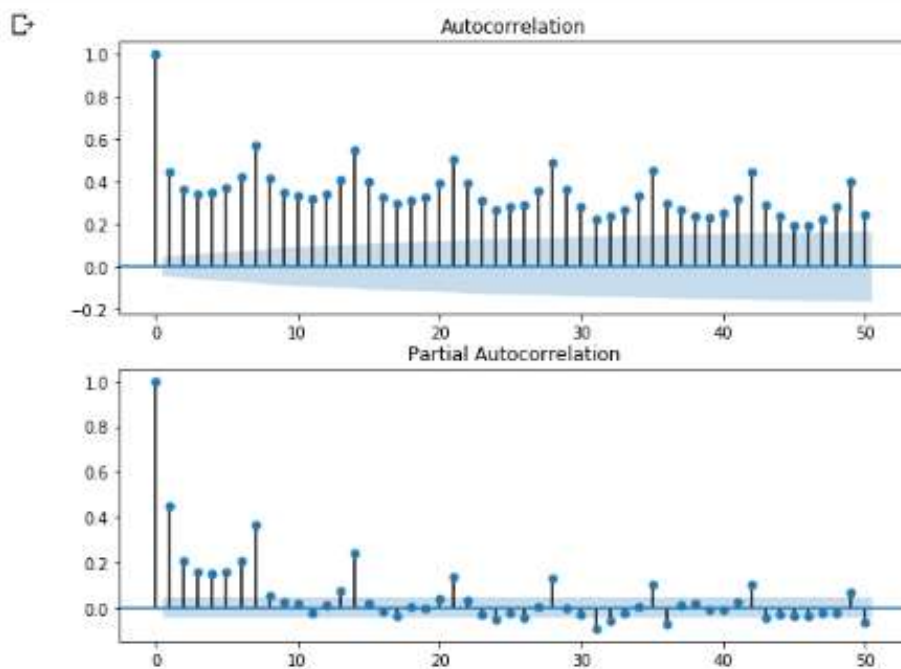
**ACF** stands for auto-correlation function. This is a function of time series with its own lagged version. We first decide on difference which is nth lag values. This are the values we plot with current values of time series. It shows the dependent relationship of current values with lagged values. We can infer components like cyclic, residual by comparison with past values. Thus it shows the relationship with past values which is our basis for future predictions..

**PACF** stands for partial auto-correlation function. As partial describes it, it find relationship with past value at nth lag. Difference between ACF and PACF is that it doesn't consider the values between current and the nth lag but only takes care of two values i.e. current and nth lag value. The correlation is determined between them which clears the their actual

dependency on each other leading to correct future prediction. This will give more insight to the actual relationship and remove any residual errors that might hinder in the ACF plot. These plots can help determine the various parameters that we might need to decide ARIMA model that would fit on our data. Therefore this is the most important step. If one looks into ACF and PACF plots, more accuracy of prediction is made whereas the basic method is to grid search which is brute force and checks on few models by random to decide which fits the best.

### ▼ PLOT ACF AND PACF PLOTS

```
[ ] plt.rcParams['figure.figsize'] = (9, 7)
fig, ax = plt.subplots(2)
ax[0] = sm.graphics.tsa.plot_acf(store1_item1.sales, lags=50, ax=ax[0])
ax[1] = sm.graphics.tsa.plot_pacf(store1_item1.sales, lags=50, ax=ax[1])
```



## 4.4 GRID SEARCH

### ▼ GRID SEARCH (MONTHLY DATA)

```
[ ] from pmdarima import auto_arima

import warnings
warnings.filterwarnings("ignore")

stepwise_fit = auto_arima(s111_month['sales'], start_p=0, start_q=0,
                          max_p=5, max_q=5, m=12, seasonal=True, trace=True)
```

```
↳ Fit ARIMA: order=(0, 1, 0) seasonal_order=(1, 1, 1, 12); AIC=171.706, BIC=179.107, Fit time=0.643 seconds
Fit ARIMA: order=(0, 1, 0) seasonal_order=(0, 1, 0, 12); AIC=167.845, BIC=171.545, Fit time=0.015 seconds
Fit ARIMA: order=(1, 1, 0) seasonal_order=(1, 1, 0, 12); AIC=159.195, BIC=166.596, Fit time=0.137 seconds
Fit ARIMA: order=(0, 1, 1) seasonal_order=(0, 1, 1, 12); AIC=150.423, BIC=157.824, Fit time=0.536 seconds
Fit ARIMA: order=(0, 1, 1) seasonal_order=(1, 1, 1, 12); AIC=152.756, BIC=162.006, Fit time=0.732 seconds
Fit ARIMA: order=(0, 1, 1) seasonal_order=(0, 1, 0, 12); AIC=148.756, BIC=154.306, Fit time=0.196 seconds
Fit ARIMA: order=(1, 1, 1) seasonal_order=(0, 1, 0, 12); AIC=149.579, BIC=156.980, Fit time=0.488 seconds
Fit ARIMA: order=(0, 1, 2) seasonal_order=(0, 1, 0, 12); AIC=149.827, BIC=157.228, Fit time=0.307 seconds
Fit ARIMA: order=(1, 1, 2) seasonal_order=(0, 1, 0, 12); AIC=151.018, BIC=160.269, Fit time=0.504 seconds
Fit ARIMA: order=(0, 1, 1) seasonal_order=(1, 1, 0, 12); AIC=150.482, BIC=157.882, Fit time=0.645 seconds
Total fit time: 4.237 seconds
```

## GRID SEARCH(DAILY DATA)

```
[ ] stepwise_fit = auto_arima(store1_item1['sales'], start_p=0, start_q=0,  
                             max_p=5, max_q=5, m=7, seasonal=True, trace=True)
```

```
Fit ARIMA: order=(0, 1, 0) seasonal_order=(1, 0, 1, 7); AIC=11885.275, BIC=11907.312, Fit time=1.610 seconds  
Fit ARIMA: order=(0, 1, 0) seasonal_order=(0, 0, 0, 7); AIC=12329.035, BIC=12340.054, Fit time=0.032 seconds  
Fit ARIMA: order=(1, 1, 0) seasonal_order=(1, 0, 0, 7); AIC=11791.002, BIC=11813.039, Fit time=0.404 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(0, 0, 1, 7); AIC=11233.966, BIC=11256.004, Fit time=0.627 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(1, 0, 1, 7); AIC=10848.530, BIC=10876.077, Fit time=5.205 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(1, 0, 0, 7); AIC=11178.455, BIC=11200.492, Fit time=0.902 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(1, 0, 2, 7); AIC=10835.378, BIC=10868.434, Fit time=9.327 seconds  
Fit ARIMA: order=(1, 1, 1) seasonal_order=(1, 0, 2, 7); AIC=10843.523, BIC=10882.088, Fit time=10.735 seconds  
Fit ARIMA: order=(0, 1, 0) seasonal_order=(1, 0, 2, 7); AIC=11883.414, BIC=11910.961, Fit time=5.829 seconds  
Fit ARIMA: order=(0, 1, 2) seasonal_order=(1, 0, 2, 7); AIC=10858.247, BIC=10896.812, Fit time=10.898 seconds  
Fit ARIMA: order=(1, 1, 2) seasonal_order=(1, 0, 2, 7); AIC=10854.135, BIC=10898.209, Fit time=11.173 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(0, 0, 2, 7); AIC=11159.229, BIC=11186.776, Fit time=2.368 seconds  
Fit ARIMA: order=(0, 1, 1) seasonal_order=(2, 0, 2, 7); AIC=10838.306, BIC=10876.871, Fit time=14.266 seconds  
Total fit time: 73.410 seconds
```

We give some parameters for grid search and it gives us some results. We print summary and take the parameters with the lowest AIC value.

This helps us get the closest predictions because, we have chosen values accordingly.



## SUMMARY

**Table 1.** Summary of fitted model(monthly data)

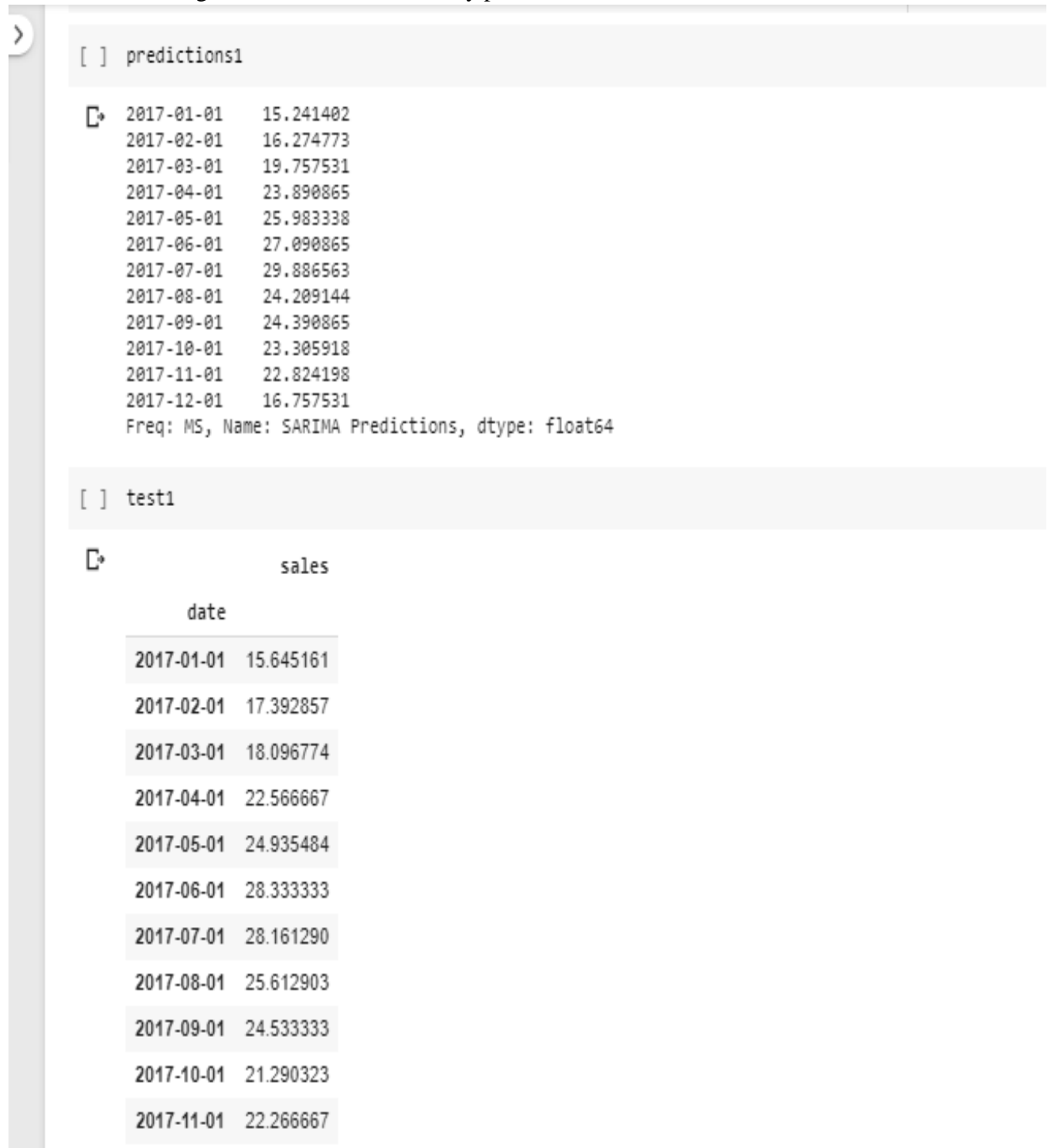
SARIMA Model	Log Likelihood	AIC	BIC	HQIC
(0,1,1)X (0,1,0) <sub>12</sub>	-71.378	148.7 56	154.3 06	150.84 5

**Table 2.** Summary of fitted model(daily data)

SARIMA Model	Log Likelihood	AIC	BIC	HQIC
(0,1,1)X (1,0,2) <sub>7</sub>	-5411.68	10835 .378	10868 .434	10847. 572

## OBTAIN PREDICTED VALUES

Put the values in algorithm and make necessary predictions.



The screenshot displays two data views in a software interface. The first view, titled 'predictions1', shows a list of dates from 2017-01-01 to 2017-12-01 with corresponding predicted values. The second view, titled 'test1', shows a table with 'date' and 'sales' columns, containing data for the first 11 months of 2017.

```
[ ] predictions1
```

2017-01-01	15.241402
2017-02-01	16.274773
2017-03-01	19.757531
2017-04-01	23.890865
2017-05-01	25.983338
2017-06-01	27.090865
2017-07-01	29.886563
2017-08-01	24.209144
2017-09-01	24.390865
2017-10-01	23.305918
2017-11-01	22.824198
2017-12-01	16.757531

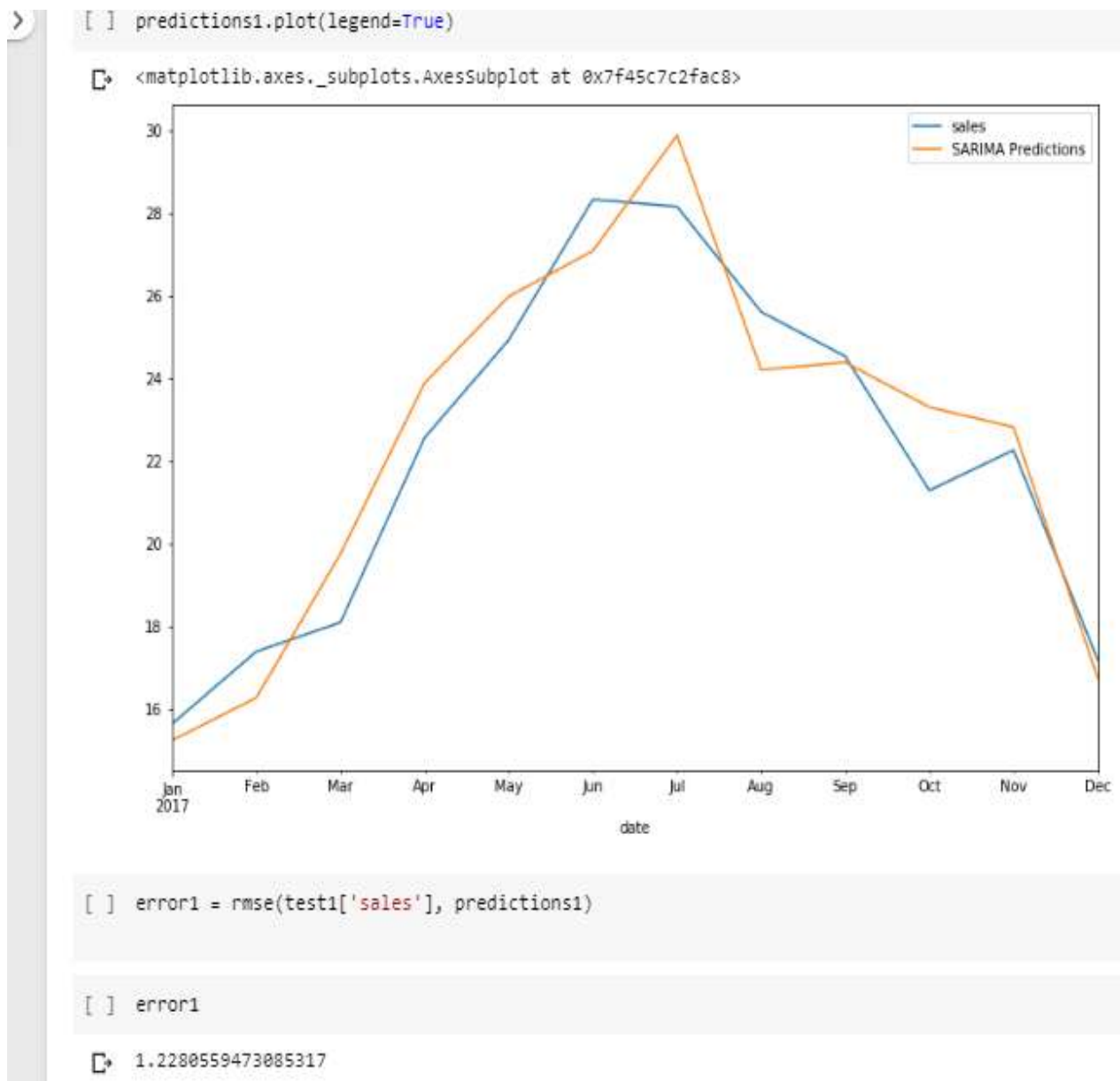
Freq: MS, Name: SARIMA Predictions, dtype: float64

```
[ ] test1
```

	sales
date	
2017-01-01	15.645161
2017-02-01	17.392857
2017-03-01	18.096774
2017-04-01	22.566667
2017-05-01	24.935484
2017-06-01	28.333333
2017-07-01	28.161290
2017-08-01	25.612903
2017-09-01	24.533333
2017-10-01	21.290323
2017-11-01	22.266667

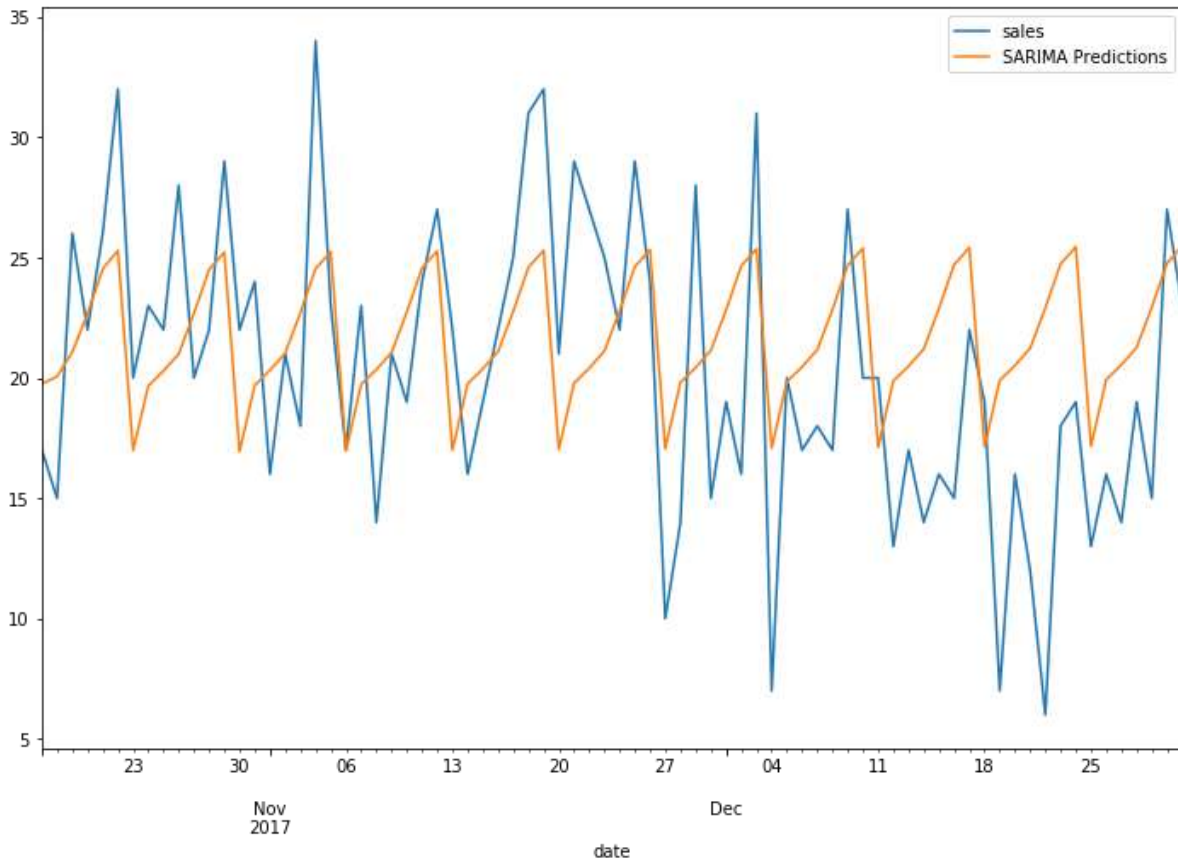
## 4.5 SALES VS PREDICTIONS

Monthly predictions :-



We check error using RMSE (root mean square error) which is 1.228.

**Daily predictions :-**



We check error using RMSE (root mean square error) which is 5.542.

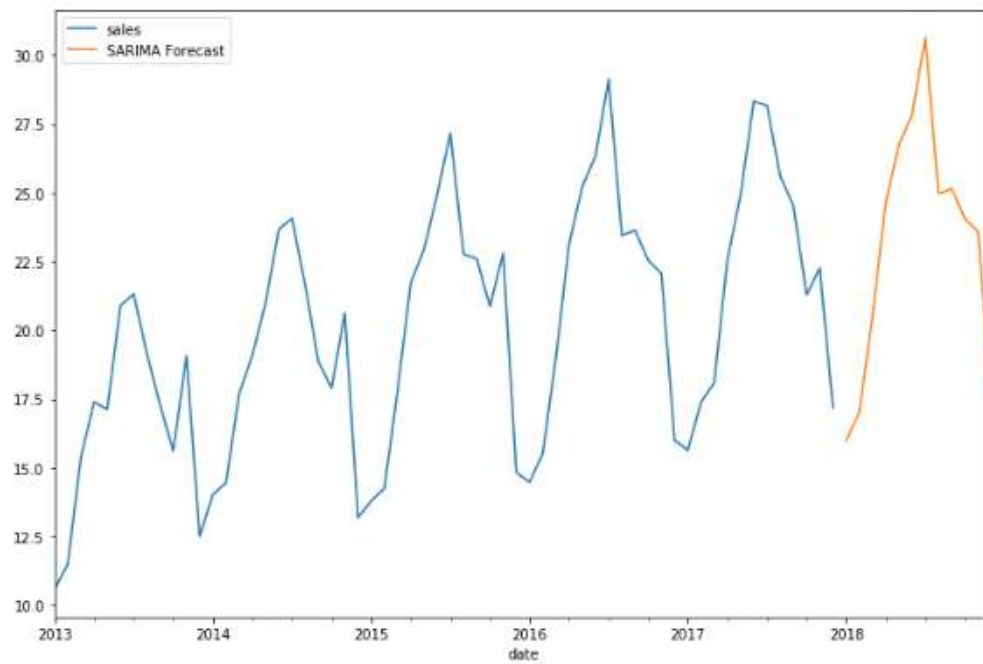
Also from the graph we can see that our predicted values are close to actual sales.

**These are predictions for 1 store and 1 item. Going back to our original problem statement, we further make predictions for 10 stores and 50 items.**

## 4.6 FINAL PREDICTIONS

```
[ ] s111_month['sales'].plot(legend=True, figsize=(12,8))  
    fcast.plot(legend=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fadbc70bb70>
```



```
[ ] rmse(test1['sales'], fcast)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fadbc70bb70>  
1.5920287850364685
```

## **CHAPTER-5**

### **CONCLUSION**

#### **CONCLUSION**

Sales is the crucial part of any business. A business financial plan needs a powerful tool for predicting sales rise or fall. This not only decides the how successful the business will be but it could help to take decisions as to which product to drop or promote. Past data of product sales could be put to use and decide future sales making the business more lively.

Seasonality and trend are two important aspect of sales and with right tools in hand, one can not only up the sales but save money also. Dropping some products is an important decision and with machine learning techniques, the business can bloom to its fullest. Products with upcoming season can be managed and stored beforehand. This will save panic buying or sudden rise in price. Sales forecasting is therefore a must with your business.

In this study, ARIMA(0,1,1) & SARIMA(0,1,0,12) were the best candidate model for predicting next 3 months of store items demand using last 5 years of store data taken from the website kaggle. This shows the power of ARIMA model to accurately predict the sales for small period of time. Although it has its limits but values predicted had very less error.

#### **BENEFITS:**

##### **Sales Planning**

Without any forecasting tools, a business is doomed. One will not be to check whether the products or schemes they are deploying are working in the favour of customers or not.

##### **Demand Forecasting**

With increasing competition in business, efficient sales forecast methods are a must. Moreover with machine learning methods, one is at advantage in relative to others. This is

the best tool to manage and check for loopholes in your business.

### **Higher OTIF Delivery**

With correct sales foretelling, you'll be able to gain an excellent OTIF or delivery. Once demands rise arrangement of inventory must be maintained maximizing the customer satisfaction. Moreover, customers will have lesser complaints and manager could focus on other plans as all prediction is left to machine learning tools.

### **Inventory Controls**

Stable inventory is vital for a successful business. Efficient forecasts of products demands over time can help avoid insufficient stock. Everything can be made according to plan which is desired. In case of rise in demands in season, one can arrange inventory beforehand and maximize the money earned with these stools.

### **Supply Chain Management**

You get a more detailed demand chart for various products, which leads to higher management over your provide chain. This paves path to more opportunities to supply items in time and manage export import as well.

### **Financial Planning**

It gives you more chances of better revenue and profit as you have already anticipated the hurdles as well as the aware of the predictions of various values leading to overall getter profit.

### **Internal Controls**

Internal controls can be managed more efficiently as business have their future predictions. After predicting future sales one will be able to create selections concerning hiring – permanent or temporary – promoting and growth.

### **Continuous Improvement**

With machine learning techniques can be used to predict sales with more accuracy as statistical model produce proper results only for small time span. Business performance will improve as accuracy is increased at a low labour with machine learning.

### **Price Stability**

One can predict demands rise and downfall and arrange inventory according to it. This can help avoid the panic of price instability. Sales can be handled by time series forecasting more accurately.

### **Marketing**

Sales forecasting helps in getting the most profit out of products. This could further provide the opportunity to promote products timely as popularity of some products could be increased. It can also give an insight to which products are not worth investing anymore.

### **FUTURE SCOPE**

The same approach can be applied to various time series problems in our daily life like stock prediction, predicting if a student will be debarred based on attendance, seeing his/her past record etc.



## REFERENCES

- [1] Adebisi, Ayodele & Adewumi, Aderemi & Ayo, Charles. (2014). Stock price prediction using the ARIMA model. Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014. 10.1109/UKSim.2014.67.
- [2] Rangsang Nochai , Titida Nochai. (2006). ARIMA MODEL FOR FORECASTING OIL PALM PRICE. Proceedings – 2<sup>nd</sup> IMT\_GT Regional Conference on Mathematics, Statistics and Applications University Sains Malaysia, Penang, June 13-15,2006.
- [3]Wei-Chiang Hong, “Rainfall forecasting by technological machine learning models”, Applied Mathematics and Computation ( Volume: 200 , Issue: 1 , June 2008 ).
- [4] “Store Item Demand Forecasting Challenge”, kaggle.com, June 2018. [Online]. Available: <https://www.kaggle.com/c/demand-forecasting-kernels-only/overview>.
- [5] Susan Li, “An End-to-End Project on Time Series Analysis and Forecasting with Python”, towardsdatascience.com, July 2018. [Online]. Available: <https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b?gi=b708004bef61>.

ORIGINALITY REPORT

---

5%

SIMILARITY INDEX  
PAPERS

2%

INTERNET SOURCES

2%

PUBLICATIONS

4%

STUDENT

---

PRIMARY SOURCES

---

**1** Wei-Chiang Hong. "Rainfall forecasting by technological machine learning models", Applied Mathematics and Computation, 2008  
Publication 1%

---

**2** Submitted to Coventry University  
Student Paper 1%

---

Submitted to Auckland University of Technology **3**  
Student Paper 1%

---

**4** Submitted to University of Wales Institute,  
Cardiff  
Student Paper 1%

---

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**

**PLAGIARISM VERIFICATION REPORT**

Date: 17-07-20

Type of Document (Tick):  PhD Thesis  M.Tech Dissertation/ Report  B.Tech Project Report  Paper

Name: JOSHITA SHARMA Department: COMPUTER SCIENCE & ENGINEERING (CSE) Enrolment No 151442

Contact No. 9459123565 E-mail. syli.sy22@gmail.com

Name of the Supervisor: DR. RUCHI VERMA

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): STORE ITEM DEMAND FORECASTING

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

**Complete Thesis/Report Pages Detail:**

- Total No. of Pages = 58
- Total No. of Preliminary pages = 9
- Total No. of pages accommodate bibliography/references = 1



(Signature of Student)

**FOR DEPARTMENT USE**

We have checked the thesis/report as per norms and found **Similarity Index** at 5 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.



(Signature of Guide/Supervisor)

Signature of HOD

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
	<ul style="list-style-type: none"><li>• All Preliminary Pages</li><li>• Bibliography/Images/Quotes</li><li>• 14 Words String</li></ul>		Word Counts	
<b>Report Generated on</b>			Character Counts	
		<b>Submission ID</b>	Total Pages Scanned	
			File Size	

Checked by  
Name & Signature

Librarian

.....