

**DAYANANDA SAGAR COLLEGE OF ENGINEERING**

*(An Autonomous Institute Affiliated to VTU, Belagavi)*

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560078

**DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT**

**A**

**MINI PROJECT**

**Report on**

**TOPIC**: **Sales Forecast & Inventory Management of an E-Commerce Industry Using Data Analytics**

By

|  |  |
| --- | --- |
| NameAJAY BHAT KSUHAS SSURENDRA N | USN1DS19IM0021DS19IM0321DS19IM021 |
|  |  |
|  |  |
|  |  |
|  |  |

**Mentor:** **Dr. S A Vasantha Kumara,**

**HOD Dept OF IEM**

**Department of Industrial Engineering and Management Dayananda Sagar College of Engineering**

**ACKNOWLEDGEMENT**

The success and final outcome of this project required a lot of guidance and assistance from many sources.

I would like to thank Dr. S A Vasantha Kumara, HOD of Department of Industrial Engineering and Management, DSCE , Bangalore for providing us the opportunity to do this project report.

I wish to express gratitude to our project mentor -----( Dr. S A Vasantha Kumara,) for his able guidance and useful suggestions to help us complete our project report on time and I thank department of IEM which helped us in successfully completing the project.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**ABSTRACT**

With the evolution of the internet from a basic tool of communications into an interactive market of products and services, many enterprises across the world attempt to embrace e-commerce. To aim for better performance in e-commerce, measuring the key performance indicators of e-commerce websites are important. By tracking and measuring these indicators, recommendation for operational improvement can be made based on actual data. This paper, focus on the forecasting of product sales based on the items, stores, transaction and other dependent variables. The sales volume forecasting process is a critical one for most businesses, also a difficult area of management. Most of researches and companies use the statistical computer analysis, or sophisticated computer simulation to analyze the sales volume for forecasting. Most of business organization heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business . In this project we have used a sales data set of a UK based e-commerce site and we have done EDA , Forecasting and Inventory Model for the very same and the results obtained are very competitive and can be used to improve their business

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| **CH NO** | **CONTENTS** |
|  | ABSTRACT |
| 1 | INTRODUCTION TO PROJECT  1.1 Introduction of Project  1.2 Module description |
| 2 | TASKS PERFORMED  2.1 Data Cleaning  2.2 Data Exploratory  2.3 Sales Forecasting using ARIMA model  2.4 Inventory Model |
| 3 | RESULTS  3.1 Accuracy of the ARIMA model  3.2 Forecasts made by the ARIMA model  3.3 Suggested Inventory Model |
| 4 | SNAPSHOTS  4.1 EDA Analysis Visualization  4.2 ARIMA Model Statistical Conclusion  4.3 Inventory Model |
| 5 | CONCLUSION |
| 6 | REFERENCES |
|  |  |
|  |  |

**CHAPTER 1**

**INTRODUCTION TO PROJECT**

* 1. **About Project**

In this work we are going to show you how analysis of a e-commerce site is made using MS Excel, SQL, Python (Jupiter),Tableau. With the evolution of the internet from a basic tool of communications into an interactive market of products and services, many enterprises across the world attempt to embrace e-commerce. To aim for better performance in e-commerce, measuring the key performance indicators of e-commerce websites are important. By tracking and measuring these indicators, recommendation for operational improvement can be made based on actual data .So its amazing to get valuable insights with just data and analysis

This project, focus on the forecasting of product sales based on the items, stores, transaction and other dependent variables. The sales volume forecasting process is a critical one for most businesses, also a difficult area of management. Most of researches and companies use the statistical computer analysis, or sophisticated computer simulation to analyze the sales volume for forecasting. Most of business organization heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business . In this project we have used a sales data set of a UK based e-commerce site and we have done EDA , Forecasting and Inventory Model for the very same and the results obtained are very competitive and can be used to improve their business Additionally, the problem becomes more complex as retailers add new locations with unique needs, new products, ever transitioning seasonal tastes, and unpredictable product marketing. In this analysis, a forecasting model is developed using machine learning algorithms to improve the accurately forecasts product sales. A model based on ecommerce's data is developed in order to validate the use of the various machine learning algorithms. In the project , Arima model for sales forecast method are compared to the methods that impact on forecast product availability in store to ensure they have just enough products at right time.

* 1. **Aim**

“To Create a Sales Forecast & Inventory Management for an E-Commerce Industry Using Data Analytics**”**

* 1. **Objective**

To show in-depth analysis of the sales of an e-commerce industry and forecast their sales with an inventory model for business insights

* 1. **Module Description**

In this project work we have been assigned to show how analysis of a e-commerce site is made and to do this project we are using MS Excel, SQL, Python (Jupiter),Tableau to clean, organize ,extract data , forecast data and to create an inventory model using the same . In order to forecast the data we have used a statistical model called ARIMA model , this model is chosen to extract the most accurate forecast model

* 1. **Software Specification**

Minimum software requirements are

Tools: MS Excel , PHP my ADMIN (SQL) , Jupyter Notebooks , Tableau

Operating System: Windows XP 7,8,10,11

Scripting Language : SQL, Python

* 1. **System Design**

**Based on data analysis theory it has four types**

* **Data Cleaning**
* **Data Extraction**
* **Data Visualization**
* **Forecasting**
* **Data Cleaning :** Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.
* **Data Extraction:**  Data extraction is the process of collecting or retrieving disparate types of data from a variety of sources, many of which may be poorly organized or completely unstructured. Data extraction makes it possible to consolidate, process, and refine data so that it can be stored in a centralized location in order to be transformed. These locations may be on-site, cloud-based, or a hybrid of the two.
* **Data Visualization :**  Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data**.**
* **Forecasting:** A forecast is a prediction made by studying historical data and past patterns. Businesses use software tools and systems to analyze large amounts of data collected over a long period. The software then predicts future demand and trends to help companies make more accurate financial, marketing, and operational decisions

**CHAPTER 2**

**TASK PERFORMED**

**2.1 Data cleaning and Transformation**

While the techniques used for data cleaning may vary according to the types of data that company stores, you can follow these basic steps to map out a framework for your organization.

Step 1: Remove irrelevant data.

Step 2: Duplicate your data.

Step 3: Fix structural errors.

Step 4: Deal with missing data.

Step 5: Filter out data outliers.

Step 6: Validate your data.

**2.2 Data Extraction**

In this process we have used SQL in order to organize and extract the required information using sequel coding and the basic steps include

Step 1: Import the cleaned and formatted data into the sequel server

Step 2 : Add the respective data type to the data and then import it

Step 3: Once the data set is imported use sequel coding to extract relevant data and find relationship between the data

Step 4:Export the extracted data in csv format

**2.3 Sales Forecasting using ARIMA model**

ARIMA, short for ‘Auto Regressive Integrated Moving Average’ is actually a class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

An ARIMA model is characterized by 3 terms: p, d, q

where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary

**Formula of an ARIMA MODEL**

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

**2.4 Inventory Model**

(note to ajay, add another page in bw this one and the next, let chapter 3 be a new page)

**CHAPTER 3**

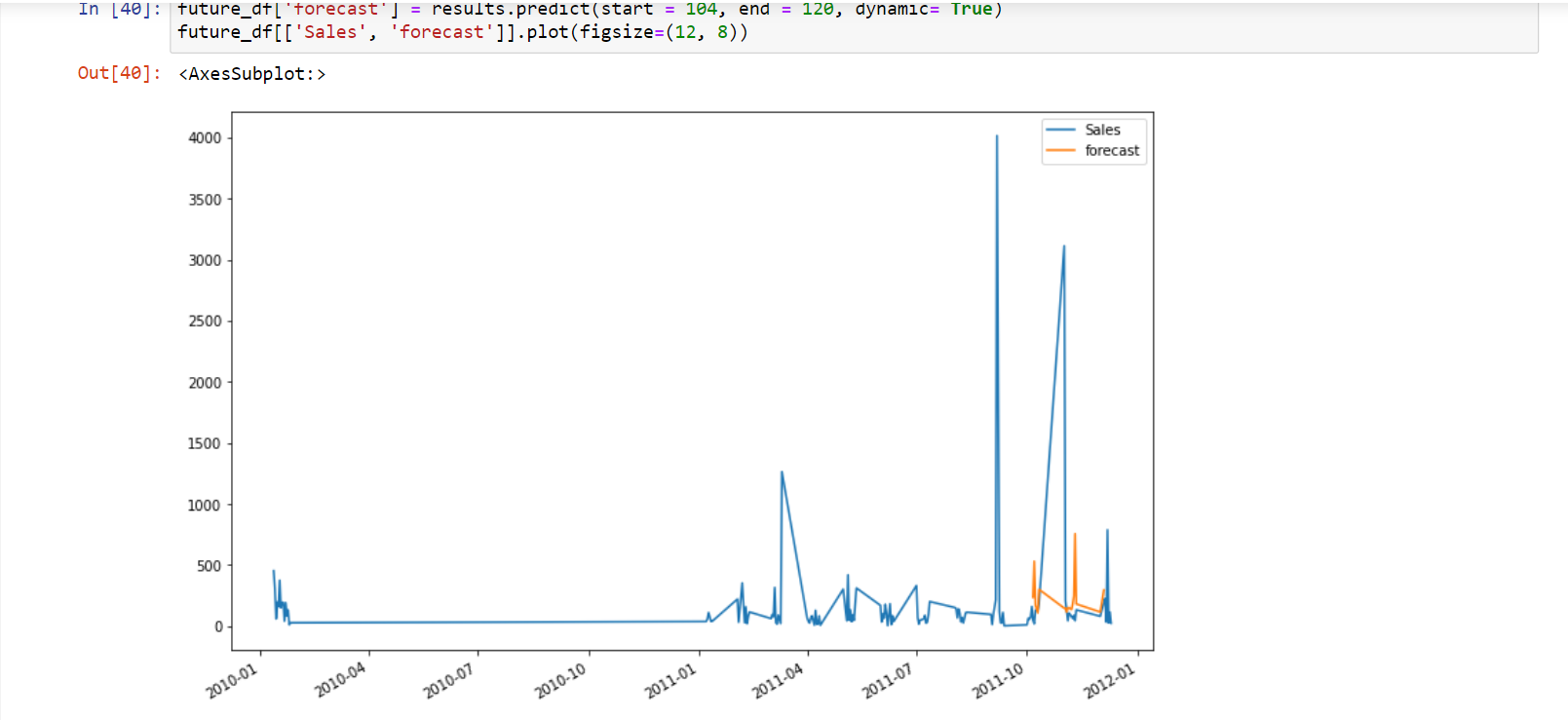
**RESULTS**

**3.1 Accuracy of the ARIMA model**

First, we import the data into the analytics tool (Jupyter notebooks) as follows

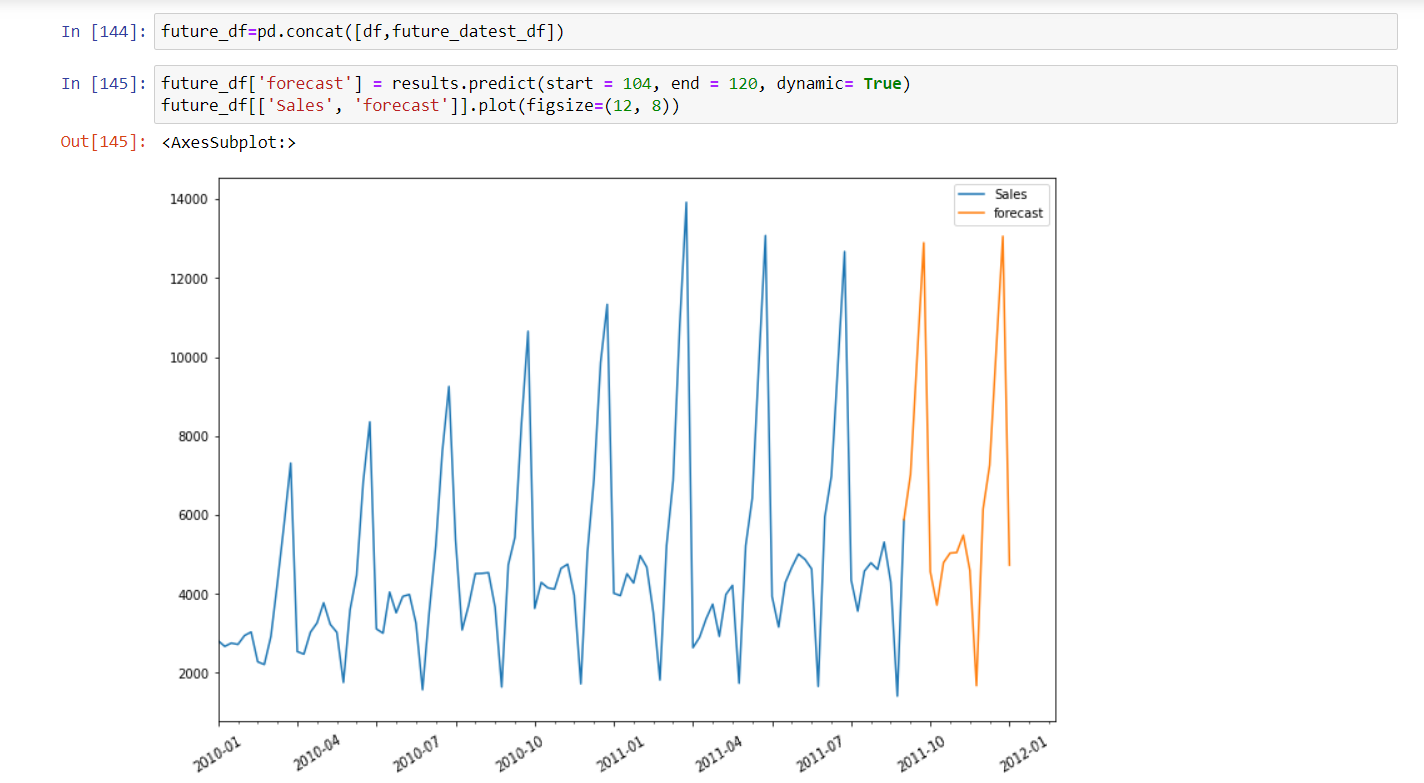


Then for carrying out the analysis of the data, we need to train the model to understand and learn from the data. Hence we test the model to ensure it makes accurate predictions. We do this by splitting the available data into training & testing data. Then we use the training data set to train the model, and then use the testing data set to check the accuracy of predictions. The results are as shown below.

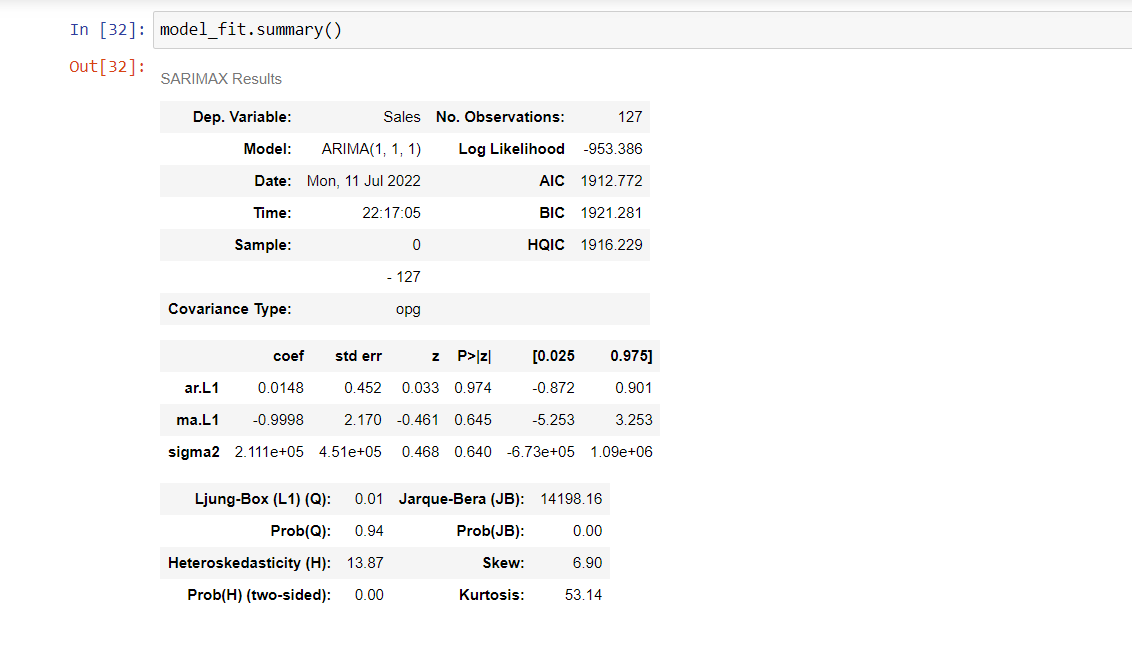


**3.2 Forecasts made by the ARIMA model**

Now that the model is able to make accurate predictions, we then use it to predict the demand for the upcoming years. The prediction made by our model is as follows.



The model we generated & trained is 97.4% accurate. The detailed statistical result is as follows.



**3.3 Suggested Inventory Model**