**“Grocery Demand Prediction”**

**A Major Project Report**

**Submitted in the Partial Fulfillment**

**of**

**the Requirements for the Degree of**

**Bachelors in information technology Engineering**

**at**

**Everest Engineering College**

**Sanepa, Lalitpur**

**Affiliated to Pokhara University**

**By**

**Dipesh Magar (19120054)**

**Jhak Prasad Pun (19120056)**

**Manish Rai (19120060)**

**Milan Bhattarai (19120063)**

****

**July,2023**

# DECLARATION

We hereby declare that the report of the project entitled “**Grocery Demands Predictions**” which is being submitted to the Department of Computer and Information Technology Engineering, Everest Engineering College, Sanepa, in the partial fulfillment of the requirements for a ward of the Degree of Bachelor of Engineering in Information Technology Engineering, is a bonafide report of the work carried out by us. The materials contained in this report have not been submitted to any University or Institution for the award of any degree and we are the only author of this complete work and no sources other than the listed here have been used in this document.

Dipesh Magar(19120054) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jhak Prasad Pun(19120056) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Manish Rai(19120060) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Milan Bhattarai(19120063) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# CERTIFICATE OF APPROVAL

The project report entitled “**Grocery Demands Predictions**”, submitted by Dipesh Magar, Jhak Prasad Pun, Manish Rai and Milan Bhattarai in partial fulfillment of the requirement for the Bachelor's degree in information technology engineering has been accepted as a bonafide record of work independently carried out by the group in the department.

…………………………………… ………………………………………

Name of the External Examiner Santa Bahadur Basnet

Designation Project Supervisor

Affiliation Department of Computer and IT Engineering

……………………………………. ……………………………………….

Nischal Regmi Anuj Ghimire

Project Coordinator Head of Department

Department of Computer and IT Engineering Department of Computer and IT Engineering

# COPYRIGHT

The author has given permission for this report to be made freely available for view by anybody at **Everest Engineering College's library**. Additionally, the author has consented that the lecturers who oversaw the project work documented here, or, in their absence, the head of the department where the project report was completed, may grant permission for widespread copying of this project report for scholarly purposes. It is agreed that credit will be provided to the project report's author and the Department of IT and Computer, EEC, for any usage of the project report's content. This report may not be copied, published, or used in any other way for commercial advantage without the Department's and the author's explicit consent. Requesting authorization to use or copy of the material in this report in whole or in part should be addressed to the Head of Department of IT and Computer.

# ACKNOWLEDGEMENT

This project would not have been possible without the joint efforts of many individuals. It has been

a pleasure for us to acknowledge the assistance and contributions that were very important and

supportive throughout the project. We would like to thank them sincerely.

We are indebted to our project supervisor, **Santa Basnet** for his guidance, support, and follow-ups throughout the project development period which helped our project to grow and foster to a level we didn't think of reaching in such a short period of time. We also express our thanks to our Associate Prof. and project coordinator **Er. Nischal Regmi**, Department of Computer & IT, Everest Engineering College for his encouragement and guidance throughout the project.

Last, but not least we would like to thank our teachers, seniors, colleagues, and family members

who have been knowingly or unknowingly part of our project with their views and instructions

during the entire development time.

Contents

[DECLARATION 1](#_Toc141791258)

[CERTIFICATE OF APPROVAL 2](#_Toc141791259)

[COPYRIGHT 3](#_Toc141791260)

[ACKNOWLEDGEMENT 4](#_Toc141791261)

[Chapter 1: Introduction 8](#_Toc141791262)

[1.1 Introduction 8](#_Toc141791263)

[1.2 Motivation: 10](#_Toc141791264)

[1.3 Objectives: 10](#_Toc141791265)

[1.4 Project Applications and Scope: 10](#_Toc141791266)

[Chapter 2: Literature Review 11](#_Toc141791267)

[Chapter 3: Methodology 12](#_Toc141791268)

[3.1 Data acquisition: 12](#_Toc141791269)

[3.2 Data Preprocessing: 13](#_Toc141791270)

[3.3 Data preparation and EDA: 16](#_Toc141791271)

[Chapter 4: Implementation Details 30](#_Toc141791272)

[4.1 Software Requirement 30](#_Toc141791273)

[4.2 Hardware Requirement 30](#_Toc141791274)

[Chapter 5: Results and Analysis 31](#_Toc141791275)

[Chapter 6: Conclusion 33](#_Toc141791276)

[References 33](#_Toc141791277)

**LIST OF FIGURES**

[Figure 1:Splitting Data into corresponding 13](#_Toc141791227)

[Figure 2: Query to select Date and Quantity 14](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791228)

[Figure 3:Date and Price 15](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791229)

[Figure 4:Quantiy VS month 16](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791230)

[Figure 5:Price Vs month 16](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791231)

[Figure 6:Box Plot(Quantity vs month) 17](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791232)

[Figure 7:Box Plot (Price Vs month) 18](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791233)

[Figure 8:Average price quarterly 19](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791234)

[Figure 9:Average price yearly 19](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791235)

[Figure 10:Average quantity quarterly 20](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791236)

[Figure 11:Average quantity yearly 20](#_Toc141791237)

[Figure 12:average quantity monthly(covt\_pct) 21](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791238)

[Figure 13:average quantity monthly(std) 21](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791239)

[Figure 15:Average monthly(cov\_ppt) 22](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791240)

[Figure 14: Average monthly (Std) 22](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791241)

[Figure 16: train and test data vs month for price 23](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791242)

[Figure 17:train and test data vs month for quantity 23](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791243)

[Figure 18:linear forecast for quantity 24](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791244)

[Figure 19:linear forecast for price 24](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791245)

[Figure 20:Naive forecast for quantity 25](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791246)

[Figure 21:Naive Forecast for price 25](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791247)

[Figure 22:Test MAPE result 26](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791248)

[Figure 23:forecast the future value of price 26](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791249)

[Figure 24:autocorrelation and partial autocorrelation 27](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791250)

[Figure 25:Total quantity for the next month 28](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791251)

[Figure 26:Test MAPE result 31](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791252)

**LIST OF TABLE**

[Table 1:Data set in CSV Format 12](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791221)

[Table 2:Splitted Data in Order 13](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791222)

[Table 3:Date and Quantity 14](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791223)

[Table 4:Date and Price 15](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791224)

[Table 5:Quantity order related to Product\_ID 29](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791225)

[Table 6: Predicted quantities of individual product for next month 32](file:///C:\Users\dipes\OneDrive\Desktop\FinalGROCERYDEAMANDPREDICTION.docx#_Toc141791226)

# Chapter 1: Introduction

## 1.1 Introduction

The grocery industry is constantly evolving and it is becoming increasingly important for retailers to accurately predict demand for products in order to optimize their inventory, reduce waste, and improve customer satisfaction. There has been an increasing demand for Ecommerce sites in the past decade. Online grocers, an offline grocery especially, have increased in popularity. The aim of this report is to create a "GROCERY DEMAND PREDICTION" that needs to be delivered to which customer and come up relevant predictions for the user and display the patterns for some business. This model predicts the grocery items to find which item is most demanded by customers or which item is less in demand. Focusing on grocery retailing has been increasing rapidly especially in developing countries like; Nepal due to the supermarket radical changes. For this project, we will be using different machine learning models, linear Regression Model, Arima Model, Naive Regression and Random forest for supervised learning, classification and regression.

Linear regression is a type of statistical analysis used to predict the relationship between two variables. It assumes a linear relationship between the independent variable and the dependent variable and aims to find the best-fitting line that describes the relationship. The line is determined by minimizing the sum of the squared differences between the predicted values and the actual values. In a simple linear regression, there is one independent variable and one dependent variable. The model estimates the slope and intercept of the line of best fit, which represents the relationship between the variables. The slope represents the change in the dependent variable for each unit change in the independent variable, while the intercept represents the predicted value of the dependent variable when the independent variable is zero. Linear regression is a quiet and the simplest statistical regression method used for predictive analysis in machine learning. Linear regression shows the linear relationship between the independent(predictor) variable i.e. X-axis and the dependent(output) variable i.e. Y-axis, called linear regression. If there is a single input variable X(independent variable), such linear regression is called simple linear regression.

To calculate best-fit line linear regression uses a traditional slope-intercept form which is given below,

(Yi = β0 + β1Xi)

where Yi = Dependent variable, β0 = constant/Intercept, β1 = Slope/Intercept, Xi = Independent variable. This algorithm explains the linear relationship between the dependent(output) variable y and the independent(predictor) variable X using a straight-line Y= B0 + B1 X.

Naive regression refers to a basic form of regression analysis that assumes a simplistic relationship between a dependent variable and one or more independent variables. It is called "naive" because it overlooks important considerations and assumptions that are typically made in more sophisticated regression models. In naive regression, the relationship between the dependent variable and independent variable(s) is assumed to be linear, with no consideration for other factors that may influence the relationship. This means that the model does not account for things like outliers, non-linear relationships, multicollinearity, heteroscedasticity, or autocorrelation, among other potential issues.

Exponential Smoothing is a popular time series forecasting method used to make predictions based on past observations, giving more weight to recent data points while gradually reducing the importance of older observations. The basic idea behind exponential smoothing is to calculate a weighted average of past data, and the weights decrease exponentially as the data gets older.

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods.

## 1.2 Motivation:

This project helps us to analyze Business growth by knowing the future value of quantity and price. It will help in understanding the role of strategy for Business Ideas, Also analyzes to calculate the Business Revenue.

## 1.3 Objectives:

* Forecasting the future average value of quantity and price per month.
* Predicting next month's quantity, which product is needed in what amount.

## 1.4 Project Applications and Scope:

* This model is used to predict what customer demand will be for a product and price.

# Chapter 2: Literature Review

“Predicting sales in a food store department using machine learning” (ROBERT SIWERZ and CHRISTOPHER DAHLEN) [1].In this paper they deal with the SVM (Support Vector Machines) they likewise utilize two more calculation is Multilayer Perceptron and Radial Basis Capacity Networks. There is a measurably huge distinction between the SVM, MLP and RFBN while foreseeing the deals in a food store division. The SVM performed lower mistake measures than the other two techniques. Since this investigation utilized constrained information, hence, one could scarcely make the determination that the SVM is consistently the most exact technique to use for deal expectation in a food store office.

“Predicting Online Grocery Ordering Intention” (Rohit Rathish and Yash Jahagirdar) [2]. In this paper they work on E-Commerce, Grocery, Recommendation System, Neural Networks. They mainly work on the neural network. The recommendation System which has been proposed, is based upon neural networks and is able to provide suitable predictions about the future orders of the customer.

"Applied Machine Learning for Supermarket Sales Prediction"[3]. Sales forecasting is very crucial for every company, especially big ones and this process is very complex because there are lots of factors that should be taken into consideration. In order to implement achievable goals and successfully implement them, supermarket chains always want to forecast sales. In this paper, they used three machine learning algorithms (K-Nearest Neighbor, Random Forest and Gradient Boosting) for sales forecasting, RF performed better, as it had a lower mean absolute error than the other two models.

"A Review on Grocery Management System Using Machine Learning Algorithms"[4].The main goal of this paper is Analyzing the proper requirement of the consumer and suggesting him products accordingly. By comparing Apriori and Max-Miner Algorithm , Max-Miner Algorithm showed more accuracy than Apriori Algorithm. In this paper, we found that Random forest and support vector machine algorithms can be used to classify the products based on purchased product and not purchased product in the grocery list. Comparing the algorithm Support vector machine showed more accuracy than random forest algorithm. This algorithm is used to build a system.

# Chapter 3: Methodology

## 3.1 Data acquisition:

At first, we collected the data from some grocery stores. We obtained nearly 40k data in the csv file. From that data we analyzed that data are in unmanaged form. In that dataset there are nearly 40k rows and nearly 350 columns.

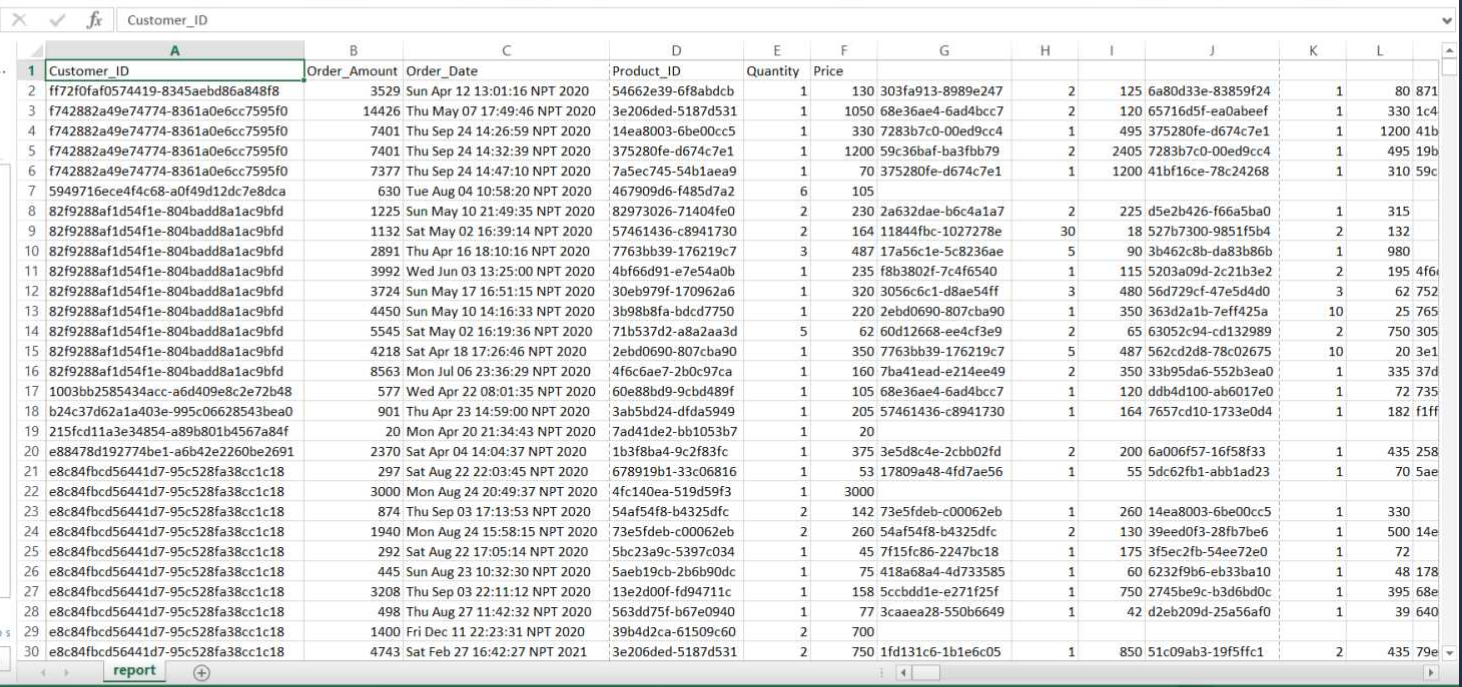


Table 1:Data set in CSV Format

## 3.2 Data Preprocessing:

After we obtained data, we analyzed data and then we sliced it into 5 chunks such as customer id, order amount, order date, product id, quantity and price. After that we obtained nearly 291k data.

A screenshot of a computer program

Description automatically generated

Figure 1:Splitting Data into corresponding

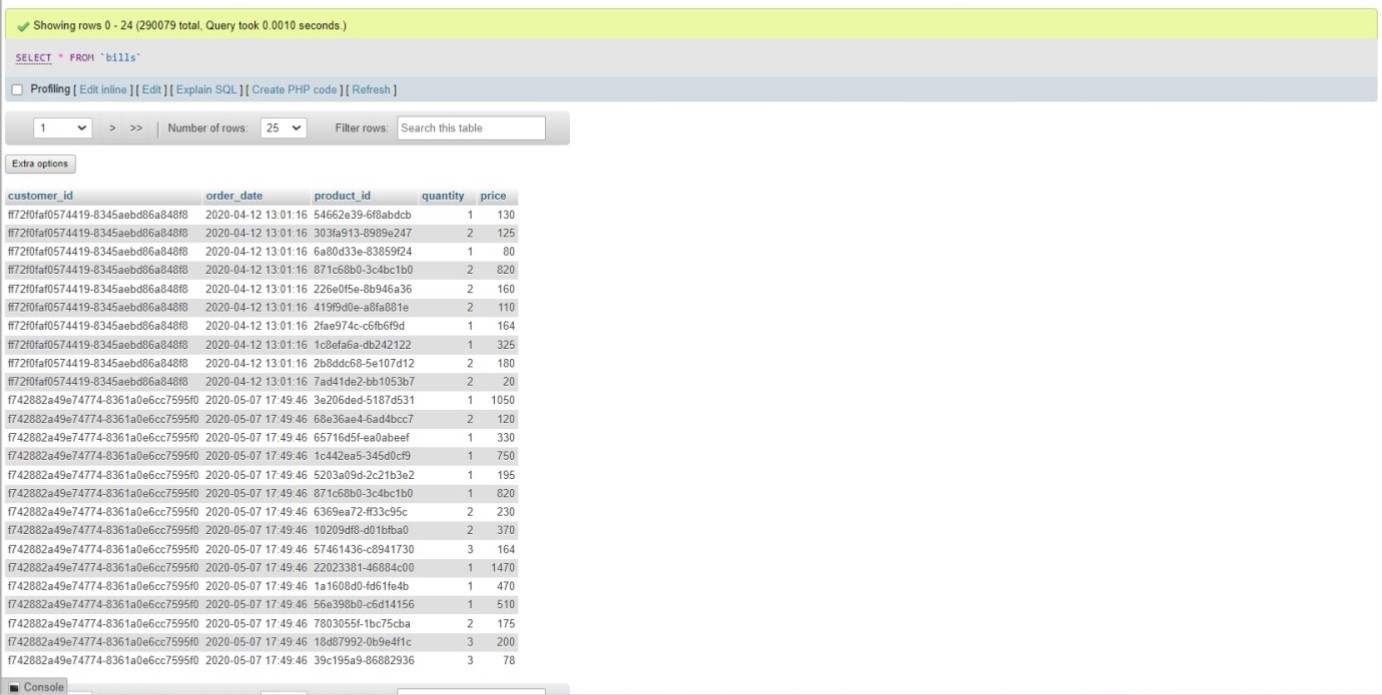


Table 2:Splitted Data in Order

After that we cleaned unmanaged data and then we analyzed data by using aggregate queries according to our project requirements.

Here are some of the examples of aggregate queries which we used during the data analyzed process.

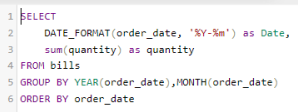
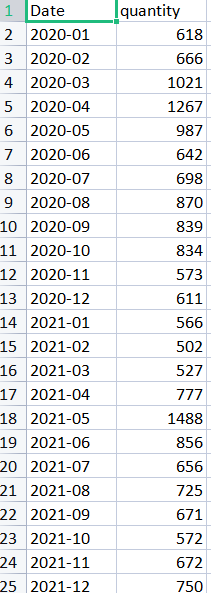


Figure 2: Query to select Date and Quantity

Table 3:Date and Quantity

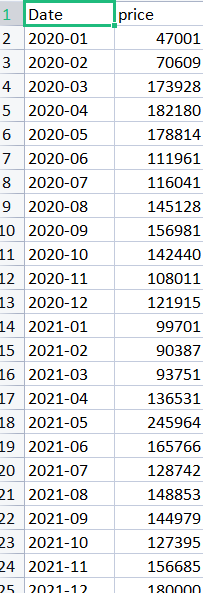
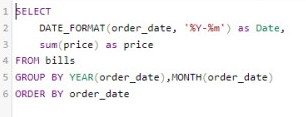


Figure 3:Date and Price

Table 4:Date and Price

**[A] Forecasting the average value of quantity and price per month since 2022**

## 3.3 Data preparation and EDA:

Hereby, the datasets are prepared and preprocessed for Exploratory Data Analysis. The datasets are visualized and analyzed and the parameters or datasets that did not have much effect in our prediction model were neglected and removed. EDA was done to make it easier for us to find anomalies in the datasets or discover patterns in our datasets.

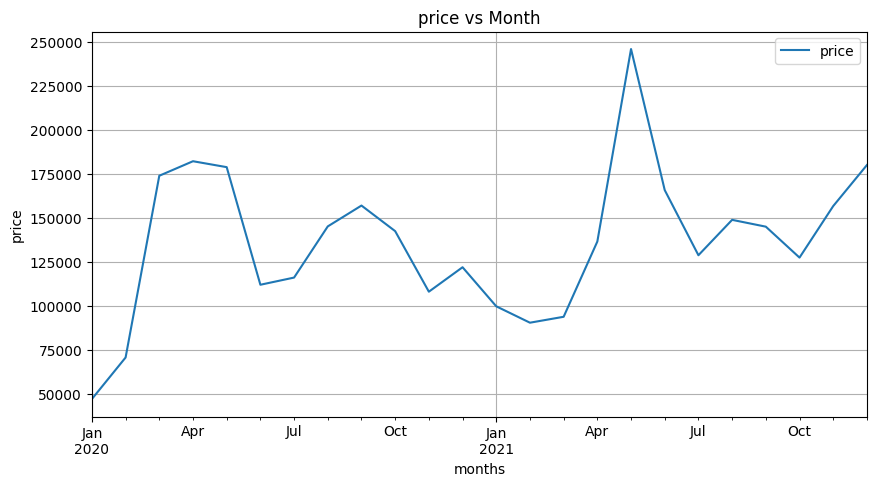
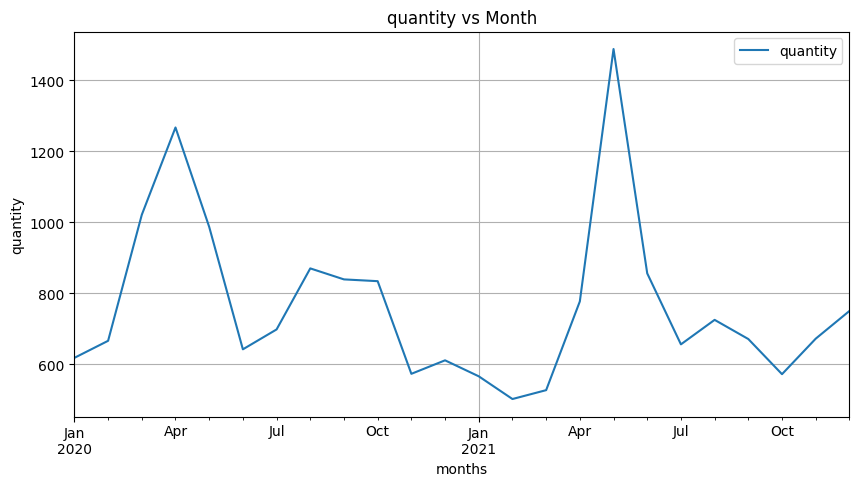
**Step 1**: we did the process of EDA (Exploratory Data Analysis). Here, we plot different graphs

Figure 4:Quantiy VS month

Figure 5:Price Vs month

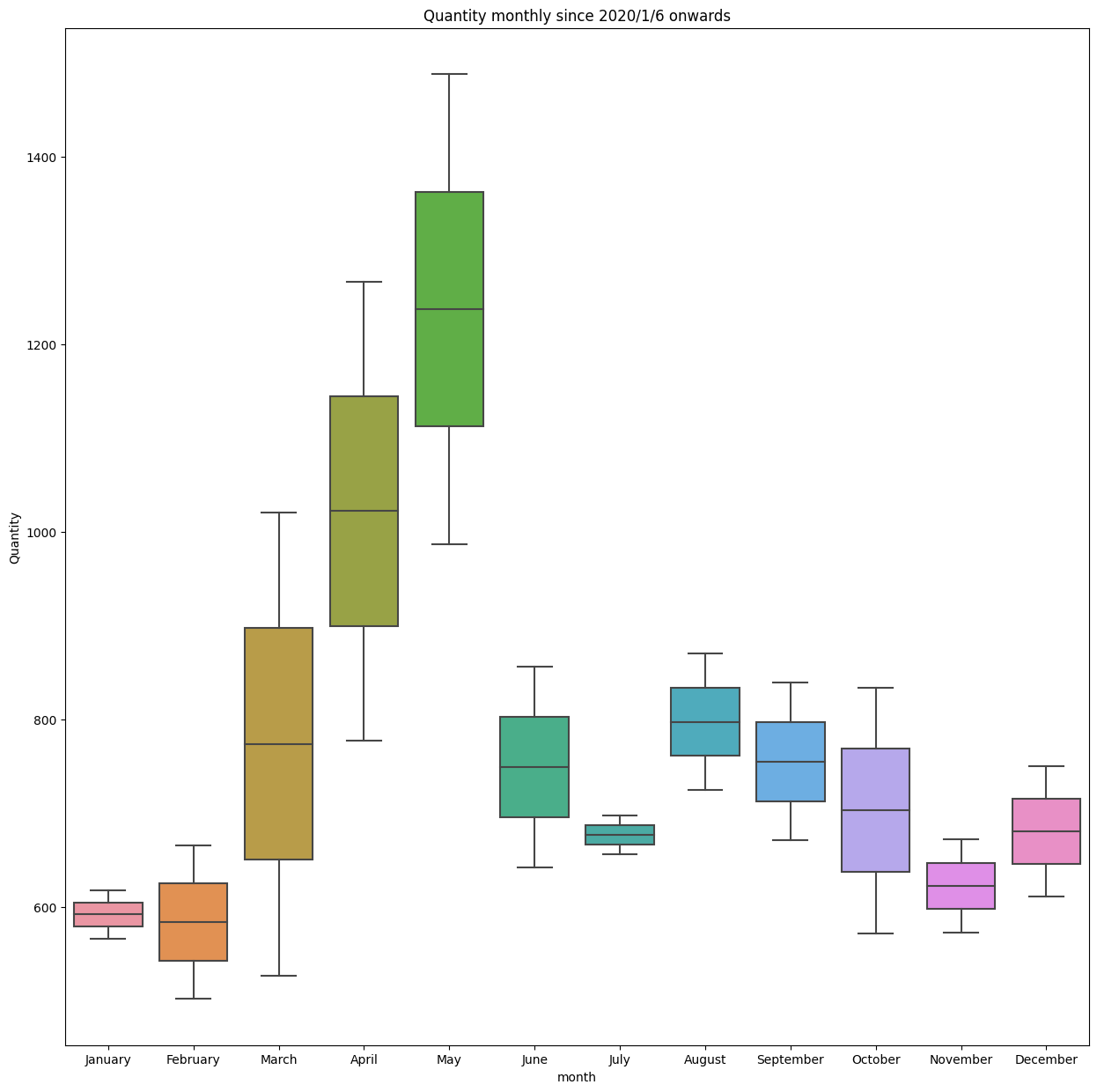


Figure 6:Box Plot(Quantity vs month)

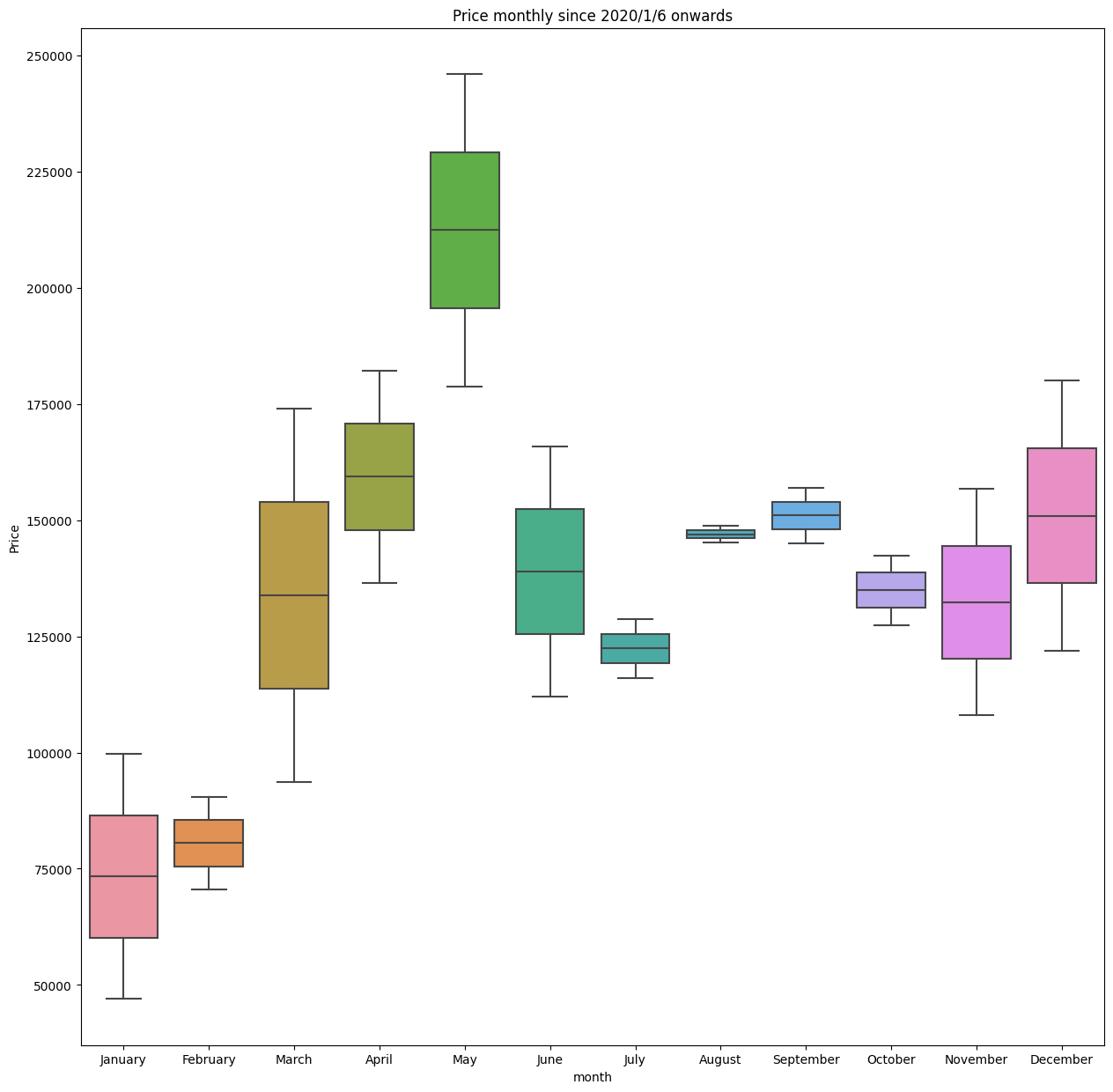


Figure 7:Box Plot (Price Vs month)

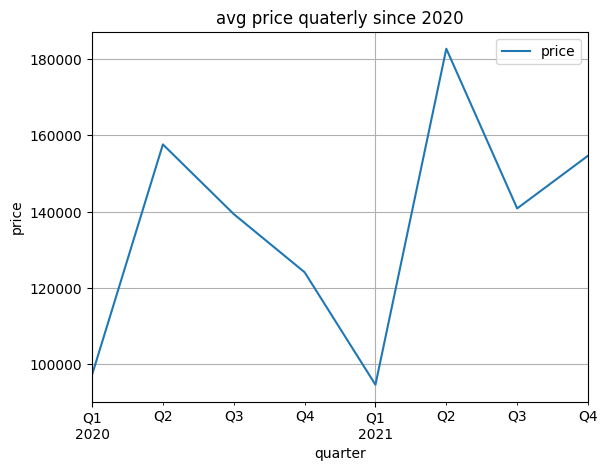
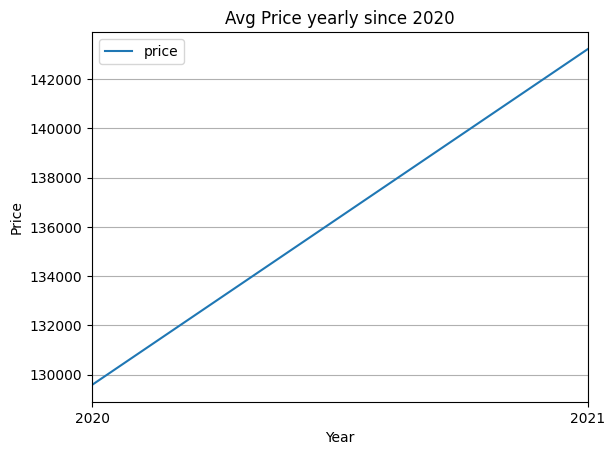


Figure 8:Average price yearly

Figure 9:Average price quarterly

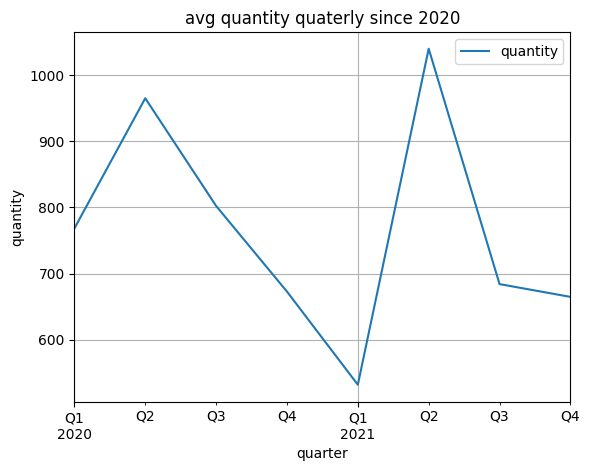


Figure 10:Average quantity quarterly

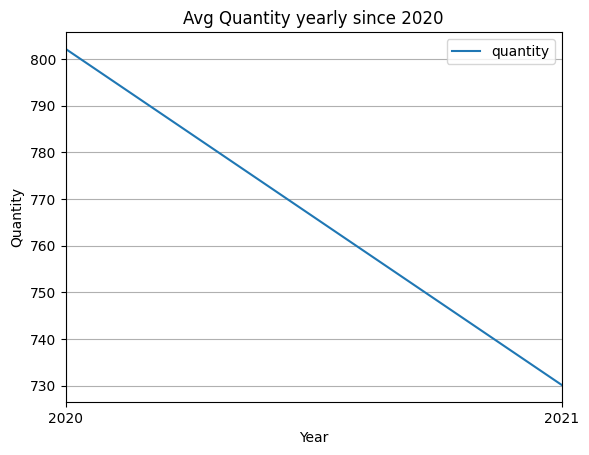


Figure 11:Average quantity yearly

**Step 2**: We Analysis the coefficient of variation.

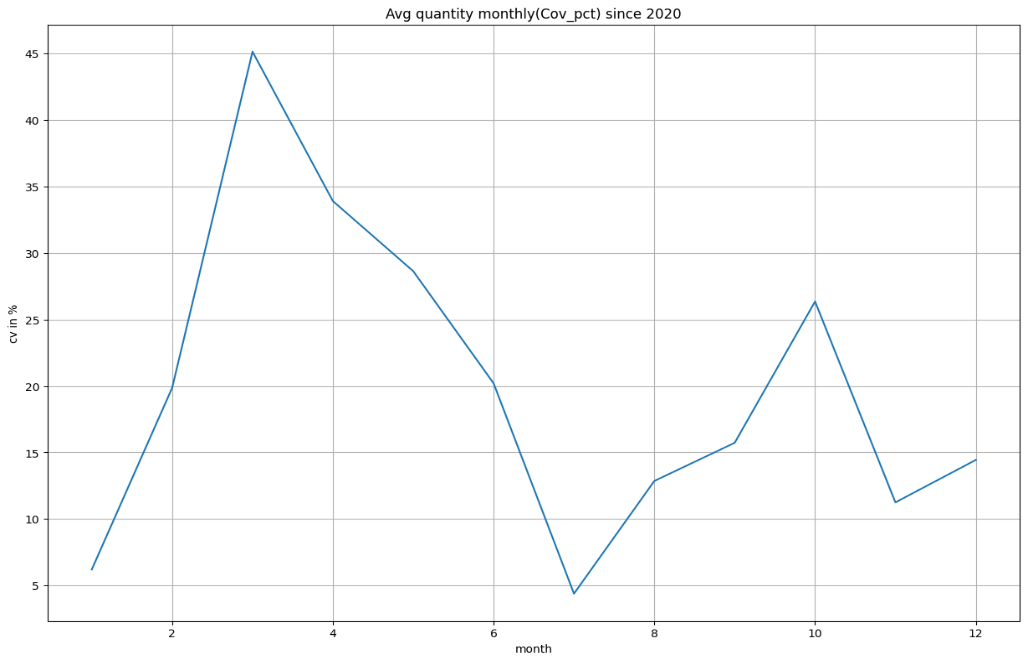


Figure 12:average quantity monthly(covt\_pct)

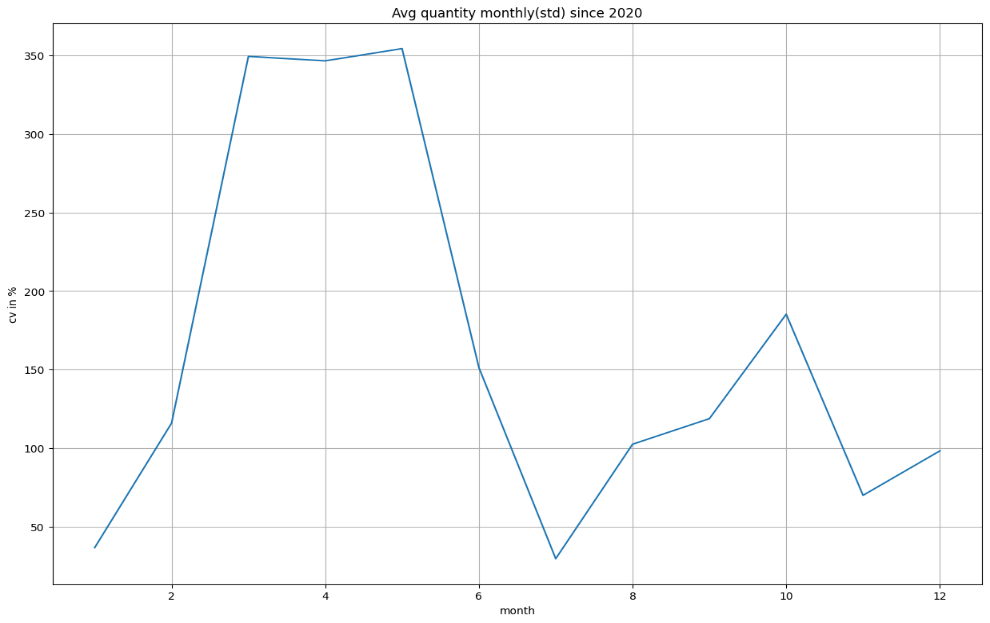


Figure 13:average quantity monthly(std)

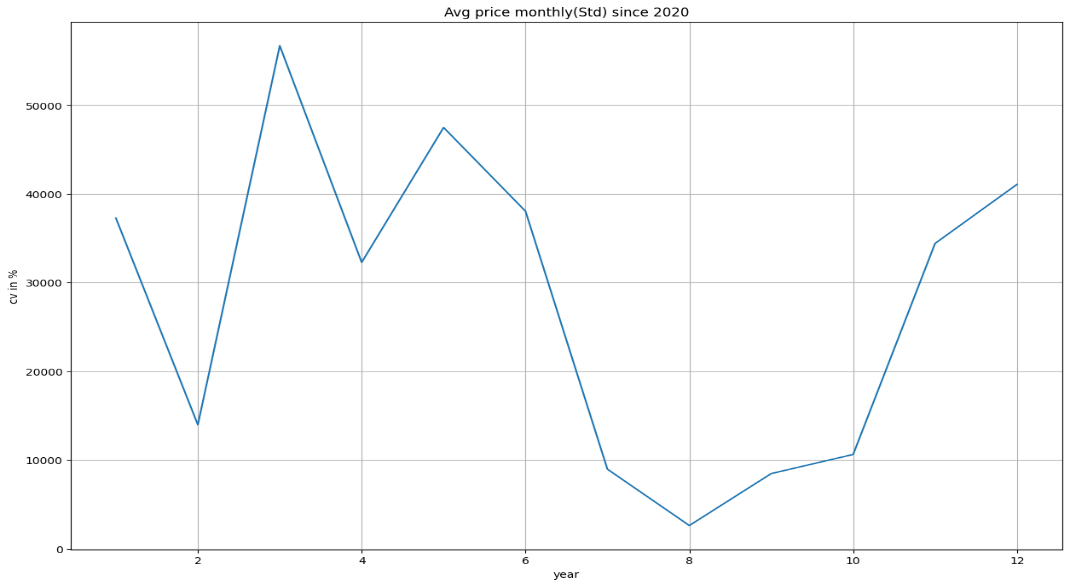
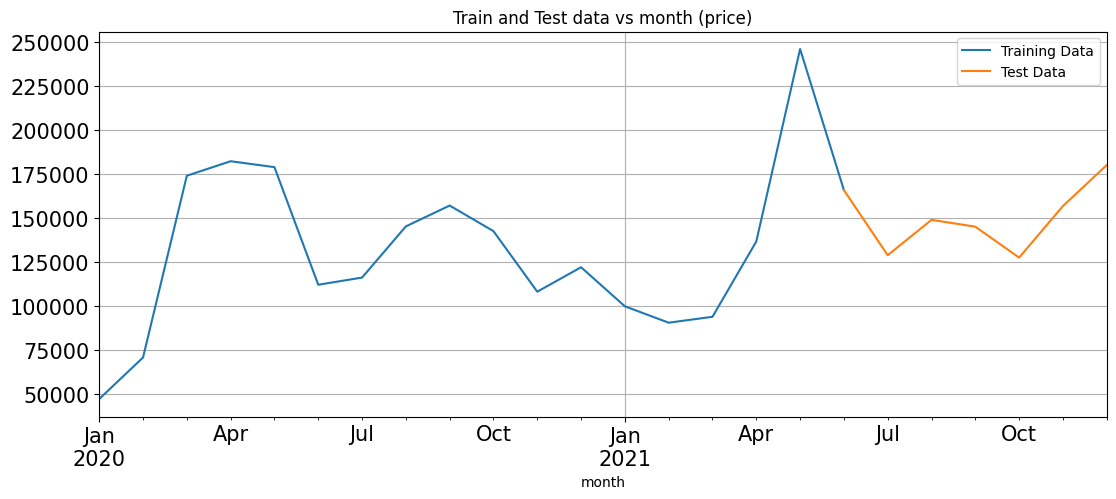




Figure 14:Average monthly(cov\_ppt)

Figure 15: Average monthly (Std)

**Step 3**: We splitted data into train and test.

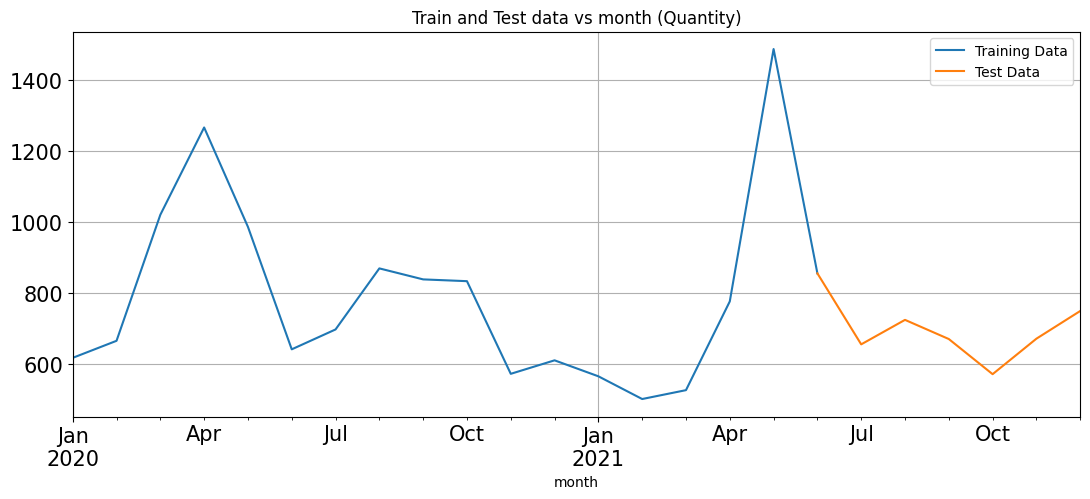
****

Figure 16: train and test data vs month for price

Figure 17:train and test data vs month for quantity

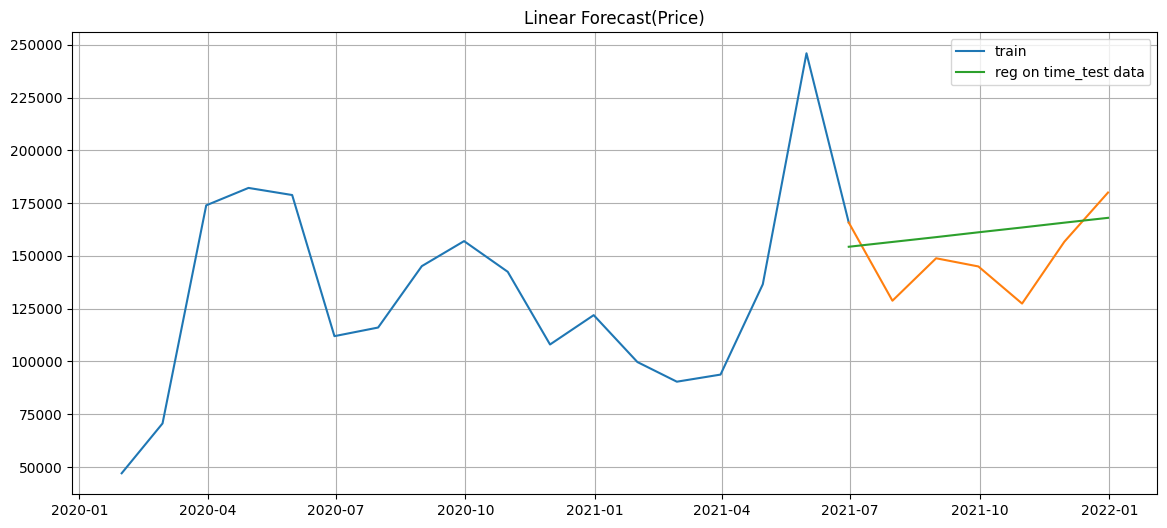
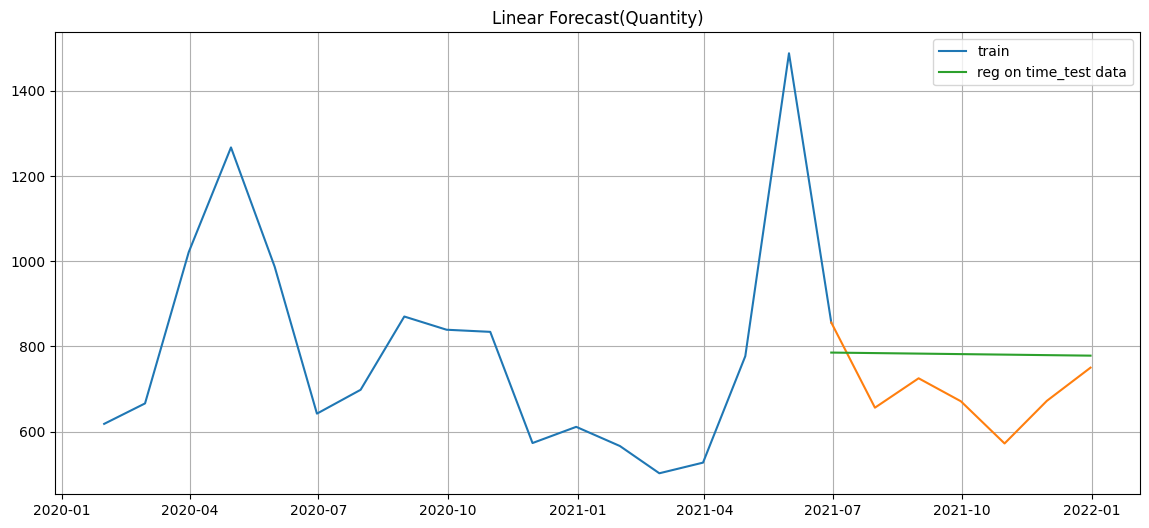
**Step 4**: We did the Model formation. Here, we apply Linear Regression, Naive Regression and ExponentialSmoothing.

Figure 18:linear forecast for quantity

Figure 19:linear forecast for price

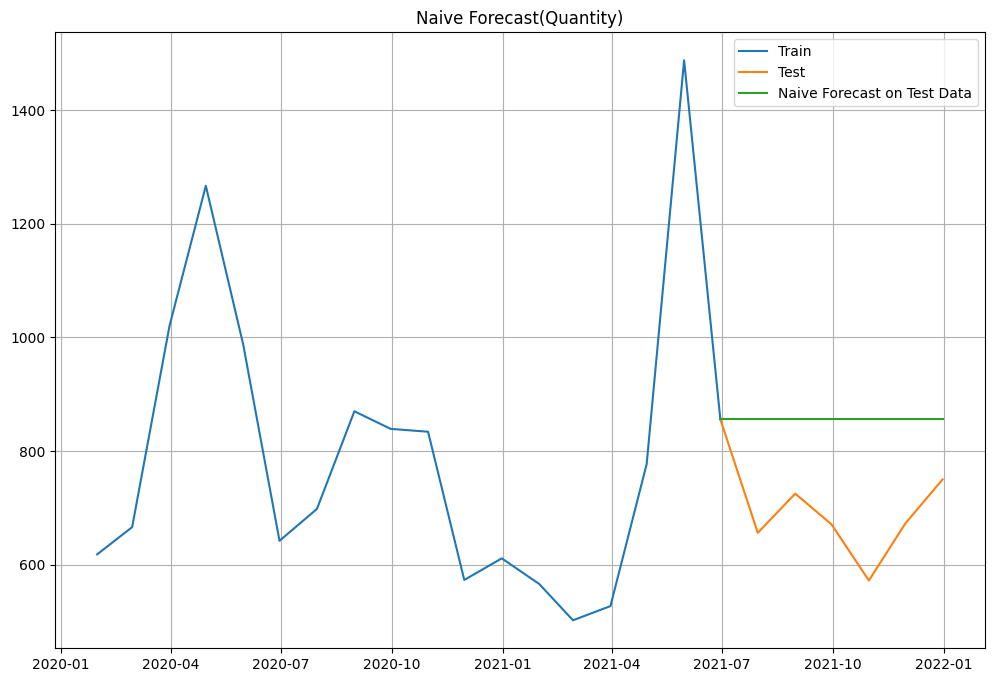


Figure 20:Naive forecast for quantity

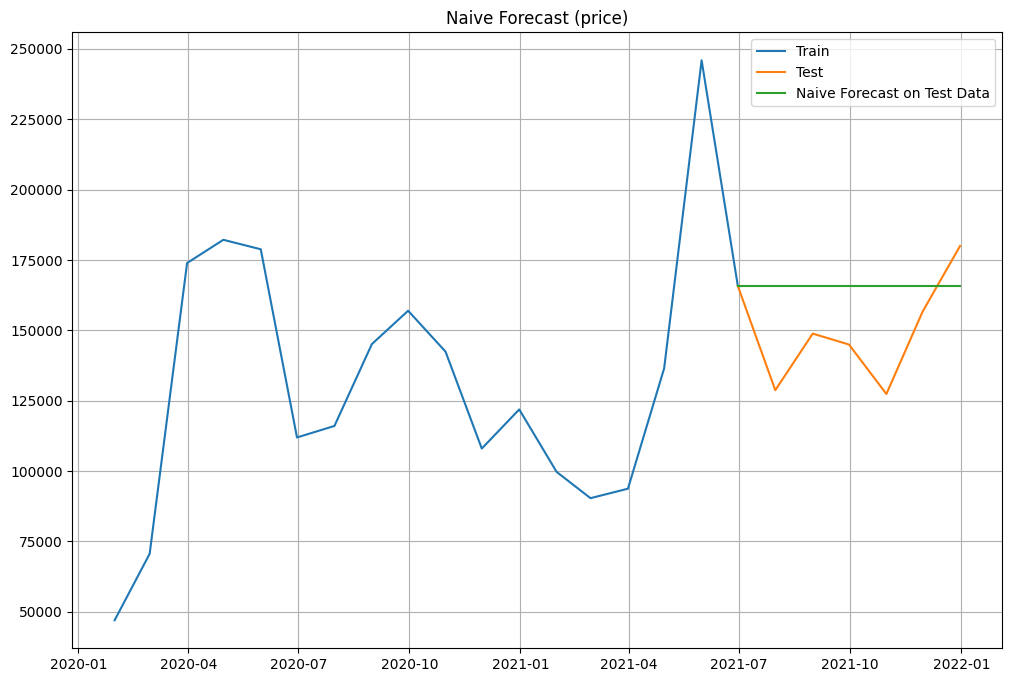


Figure 21:Naive Forecast for price

A screenshot of a test

Description automatically generated

Figure 22:Test MAPE result

After comparing the above three models, we get the least MAPE value on ExponentialSmoothing. So, we choose ExponentialSmoothing for final forecasting.

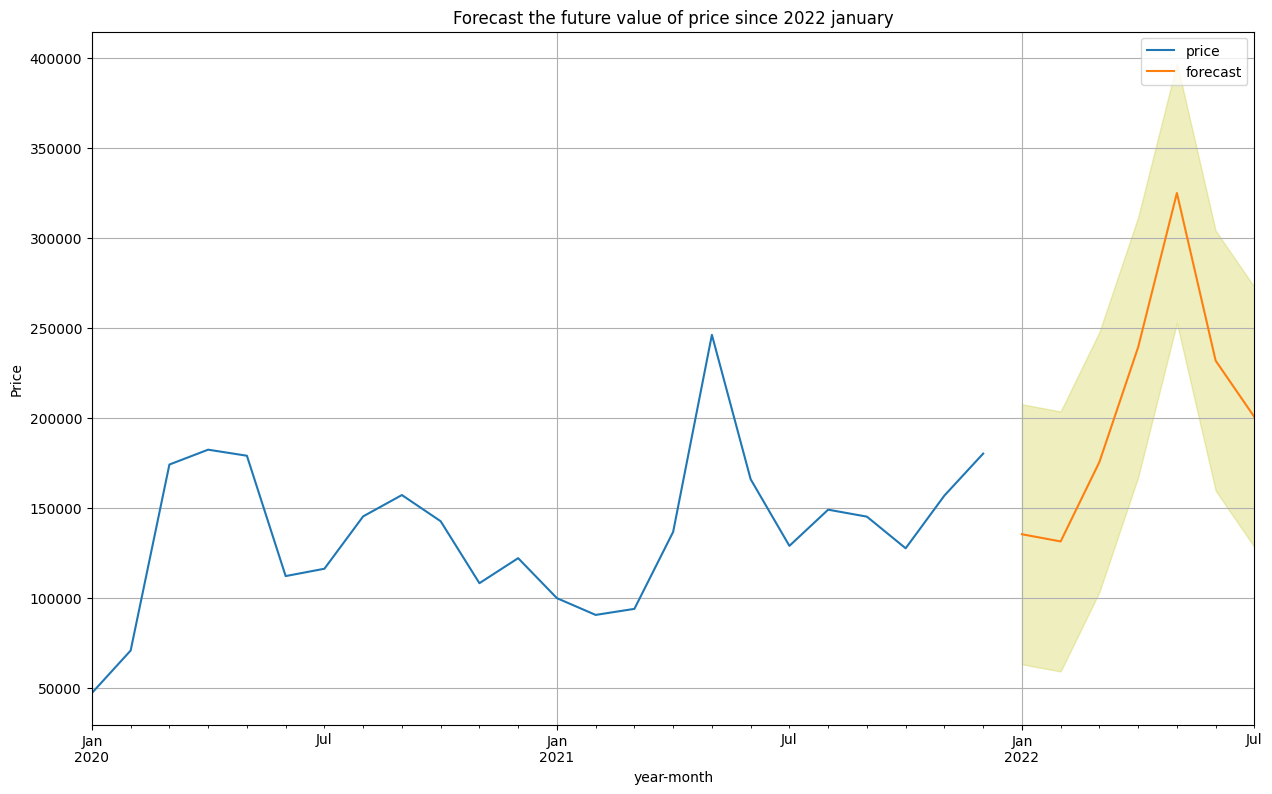
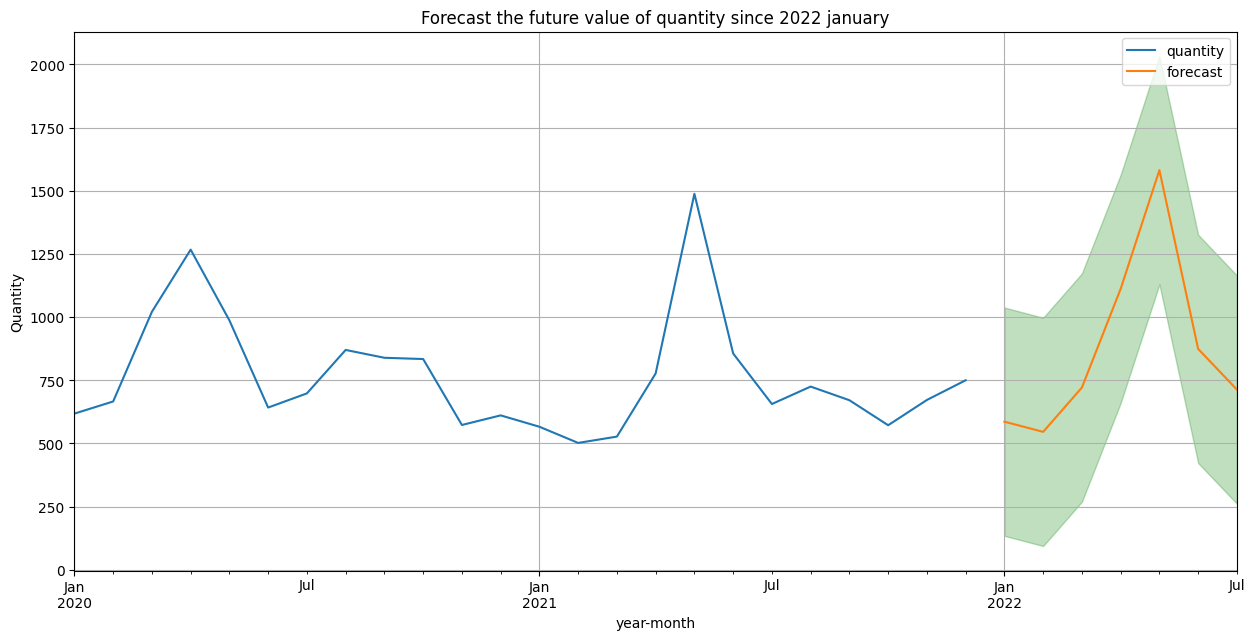
Step 5: we did the final forecasting for price and quantity where we seen the following graph.

Figure 23:forecast the future value of price



**[B] Predicting which quantity is required in what amount using ARIMA model**.

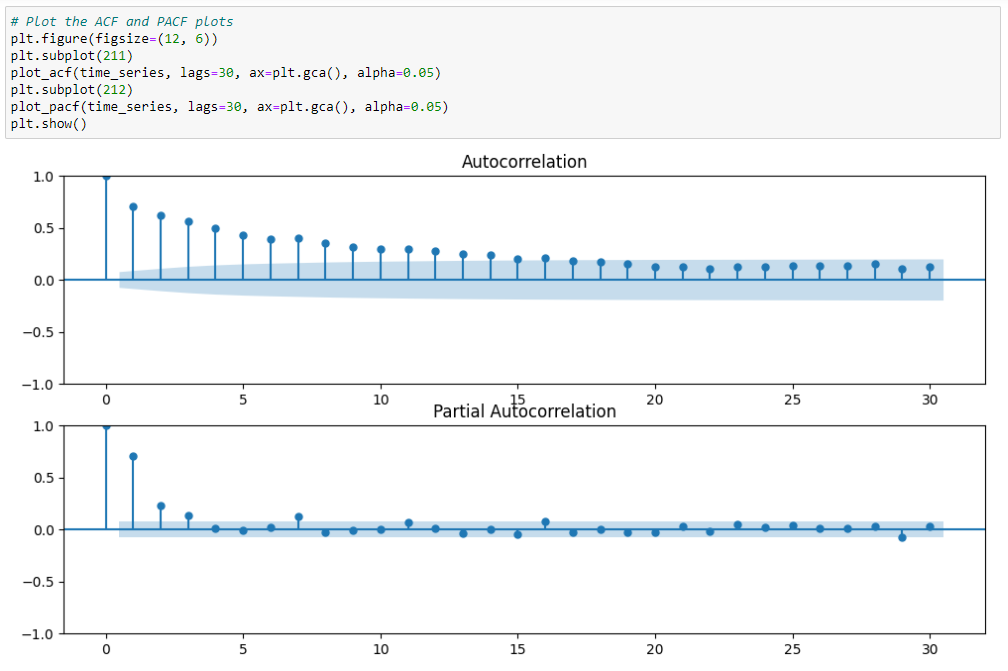
**Step1**: We calculate Autocorrelation and Partial Autocorrelation of data set.

Figure 24:autocorrelation and partial autocorrelation

**Step 2**: We use ARIMA model to forecast the quantity for next month.

A graph of blue lines

Description automatically generated

Figure 25:Total quantity for the next month

**Step 3:** We use table to show which quantity is required in what amount in ascending and descending order.

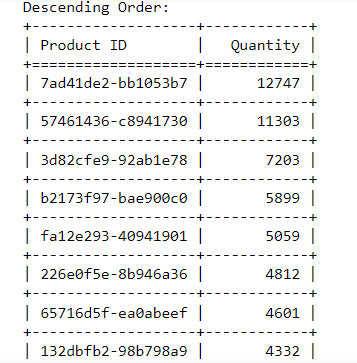


Table 5:Quantity order related to Product\_ID

# Chapter 4: Implementation Details

## 4.1 Software Requirement

The list of software that will be used to run this system is listed below:

* Any operating system (Linux, Windows, MacOS)
* Web Browser
* Code Editor (Jupyter Notebook)

## 4.2 Hardware Requirement

The list of hardware that will be used to run this system is listed below:

* General PC (min RAM 8GB, HDD 500GB, SSD adds value)

# Chapter 5: Results and Analysis

For the first objective of our project that is Forecasting the future average value of quantity and price per month since 2022 we apply Linear regression, Naïve regression and ExponentialSmoothing and comparing the MAPE value of each where we found the least MAPE value of ExponentialSmoothing. So, we choose ExponentialSmoothing for the forecasting of future average value of quantity and price.

A screenshot of a test

Description automatically generated

Figure 26:Test MAPE result

After comparing the above three models, we get the least MAPE value on ExponentialSmoothing. So, we choose ExponentialSmoothing for final forecasting.

For the second objective of our project, we use ARIMA model to predict which product is needed in what amount in coming next month. And show the required quantities in ascending and descending order.

We forecast the future average value of quantity and price per month since 2022 this helps the user to make business strategy accordingly.

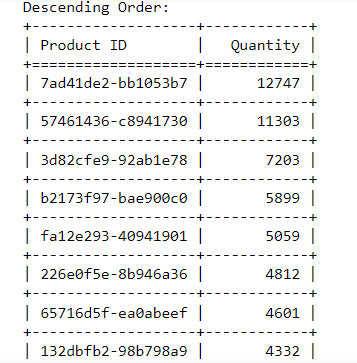


Table 6: Predicted quantities of individual product for next month

# Chapter 6: Conclusion

Grocery Demand Prediction is very crucial for every company, especially big ones and this process is overly complex because there are a lot of factors. In this study, we used three machine learning algorithms (Linear Regression, Navie Regression, Exponential Smoothing) for quantity forecasts, Exponential Smoothing performed better, as it had a lower MAPE value than the other two models.

The Grocery Demand Prediction which has been proposed is based upon ARIMA model and is able to provide suitable predictions about the future quantity of the customer.

# References

1]Dahlen, R. S. (2016 6th). *Predicting sales in a food store department Using machine learning.* International Conference on Information and Communication Technology for Muslim World.

[2] Kumari, N. (n.d.). Predicting Online Grocery Ordering Intention. *Department of CSE, VIT University Chennai Campus, India*.

[3] Odegua, R. (n.d.). *Applied Machine Learning for Supermarket Sales Prediction.* Ekpoma, Edo state, Nigeria: Department of Computer Science Ambrose Alli University.

[4] Pawar, P. M. (n.d.). Retrieved from www.ijera.com