Project Report on

PHARMACEUTICAL SALES ANALYTICS

Submitted for the award of the Bachelor of Technology degree

by

ANSHUMAN SINHA (Reg. No: 15030141CSE008)

Under the guidance of

Prof. Srinatha D K
Computer Science & Engineering Department,
Alliance University

Department of Computer Science and Engineering
Alliance College of Engineering and Design
Alliance University
Bangalore

May 2019

ALLIANCE UNIVERSITY

ALLIANCE COLLEGE OF ENGINEERING & DESIGN
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Chikkahagade Cross
Chandapura - Anekal Main Road, Anekal
Bangalore – 562106

CERTIFICATE

This is to certify that the report entitled, **"Pharmaceutical Sales Analytics"** is a bonafide record of work done during the academic year 2018-2019 by **Anshuman Sinha (Reg. No: 15030141CSE008)** in partial fulfillment of the requirements for the award of Bachelor degree in Computer Science and Engineering, Faculty of Engineering, Alliance University.

Srinatha 27/5/19

Prof. Srinatha D K
Internal Guide

27/5/19

Dr. Shekhar R.
Head of the Département
(Incharge)

External Examiners :

1. Name : Dr. R. Selvarani Signature : Selva
2. Name : Prof. Srinatha. D. K Signature : Srinatha

ALLIANCE UNIVERSITY
Alliance College of Engineering and Design

DECLARATION

This is to declare that the report titled **'Pharmaceutical Sales Analytics'** has been made for the partial fulfilment of the Course Bachelor of Techonology in Computer Science and Engineering, under the guidance of **Prof. Srinatha D K.**

I confirm that this report truly represents my work undertaken as a part of my project work. This work is not a replication of work done previously by any other person. I also confirm that the contents of the report and the views contained therein have been discussed and deliberated with the faculty guide.

NAME	REG. NO.	SIGNATURE
Anshuman Sinha	15030141CSE008	Anshumansinha

ACKNOWLEDGEMENT

First and foremost, I praise and thank ALMIGTHY GOD whose blessings have bestowed in me the will power and confidence to carry out my Project work. I owe my heartful gratitude to Alliance University, Alliance College of Engineering & Design for giving me this opportunity to undergo this course.

We avail this opportunity to express my deep sense of gratitude and hearty thanks to **Management**, Alliance University, for providing world class infrastructure, congenial atmosphere and encouragement.

I wish to extend my sincere thanks to **Dr. Reeba Korah**, **Associate Dean**, Alliance College of Engineering and Design, Alliance University, for giving me the opportunity to do this project in this prestigious institution.

I owe a deep sense of gratitude **to Dr. Shekhar. R , Head of the Department**, Department of Computer Science & Engineering, ACED, for his keen interest on me at every stage of my project. His prompt inspirations, timely suggestions have enabled me to complete my project.

I feel it a pleasure to be indebted to my **Project Guide Prof. Srinatha D K** for his invaluable support, advice, enthusiasm and dynamism have encouraged me to complete magnificently my project.

We express my deep sense of gratitude and thanks to the teaching and non-teaching staff at our department who stood with me during the project and helped me to make it a successful venture.

We place highest regards to my parents, my friends and well-wishers who helped a lot in making the report of this project.

Anshuman Sinha (15030141CSE008)

ABSTRACT

Machine learning (ML) is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. It can be used to predict the Weather prediction, Sales Prediction, Hormone Imbalance Predictions. Here we are considering **Pharmaceutical Sales Data** through which we can predict the models. Prediction helps to analyse the trained data and gives prediction on untrained data.

We will be analysing few problems and perform operation on related data sets, such as:

1. Business Improvement
2. Sales Prediction
3. Sales Forecasting

To solve such problems, **regression model** are used. They predict missing values based on other independent variables about which information is available.

- **Predictive model:** To get a prediction of the missing values, the data set has to be divided into two parts. The first part with the existing data and the second part with missing data. The first part will become the train set, while the missing values in the second part become the forecast target. A binary classifier in this case answers the question whether an event happened or not (for example, whether the products were present on the shelf). A categorical classifier assigns a product to a particular segment.
- **Linear Regression:** In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used.

TABLE OF CONTENTS

Acknowledgement…… i

Synopsis……. ii

List of Figures and Screenshots……... vii

1. Introduction……... 1

1.1. About the Project……...……... 2

1.2. Objective of the project……...…….. 3

1.3. Modules……. 4

1.3.1. Data Analytics Model……. 4

1.3.2. Profit and Loss function…….. 5

1.3.3. Graph Generation…… 5

1.3.4. Model Generation…… 5

1.4. Proposed System……...……. 6

1.4.1. Expected Advantages of Proposed System……. 6

1.5. Feasibility Study……. 7

1.5.1. Technical Feasibility…… 7

1.5.2. Economic Feasibility…… 7

1.5.3. Behavioral Feasibility…….. 8

2. Analysis……. 9

2.1. Methodology……... 11

2.1.1. Data Analytics…….. 11

2.1.2. Machine Learning (ML)……... 12

2.1.3. Supervised Machine Learning…….. 13

2.1.4. Linear Regression……. 13

2.1.5. Python……... 14

2.1.6. Anaconda Navigator……. 15

2.1.7. Jupyter Notebook…….. 16

3. Design and Implementation……. 17

3.1. Logical Design…… 18

3.2. Context Flow Diagram…… 19

3.2.1. Flow diagram for dataset of a region…… 19

3.2.2. Flow Diagram for dataset of a month……... 20

3.2.3. Flow Diagram for finding Profit or Loss……..……. 21
3.2.4. Flow Diagram for Prediction Model…… 22
4. Outputs and Results……. 23
4.1. First Month Data Analytics Output……. 24
4.1.1. Barchart for Region 1……... 24
4.1.2. Piechart for Region 1…… 24
4.1.3. Barchart for Region 2…….. 25
4.1.4. Piechart for Region 2……...… 25
4.1.5. Barchart for Region 3…….. 26
4.1.6. Piechart for Region 3…….. 26
4.1.7. Barchart for Region 4……..…… 27
4.1.8. Piechart for Region 4…….. 27
4.1.9. Barchart for First Month…….… 28
4.1.10. Piechart for First Month…….. 28
4.1.11. Barchart for Profit/Loss in Region 1……... 29
4.1.12. Barchart for Profit/Loss in Region 2……... 29
4.1.13. Barchart for Profit/Loss in Region 3……... 30
4.1.14. Barchart for Profit/Loss in Region 4……... 30
4.1.15. Barchart for Profit/Loss in First Month ……. 31
4.2.Second Month Data Analytics Output…… 31
4.2.1. Barchart for Region 1…….. 31
4.2.2. Piechart for Region 1……... 32
4.2.3. Barchart for Region 2…….. 32
4.2.4. Piechart for Region 2……... 33
4.2.5. Barchart for Region 3……... 33
4.2.6. Piechart for Region 3……..... 34
4.2.7. Barchart for Region 4……..……. 34
4.2.8. Piechart for Region 4……... 35
4.2.9. Barchart for Second Month……. 35
4.2.10. Piechart for Second Month…….. 36
4.2.11. Barchart for Profit/Loss in Region 1…… 36
4.2.12. Barchart for Profit/Loss in Region 2…… 37
4.2.13. Barchart for Profit/Loss in Region 3…… 37
4.2.14. Barchart for Profit/Loss in Region 4…… 38

4.2.15. Barchart for Profit/Loss in Second Month ……. 38
4.3.Third Month Data Analytics Output…… 39
4.3.1. Barchart for Region 1……..... 39
4.3.2. Piechart for Region 1……... 39
4.3.3. Barchart for Region 2…….. 40
4.3.4. Piechart for Region 2…….. 40
4.3.5. Barchart for Region 3……. 41
4.3.6. Piechart for Region 3…….. 41
4.3.7. Barchart for Region 4……. 42
4.3.8. Piechart for Region 4…….. 42
4.3.9. Barchart for Third Month…… 43
4.3.10. Piechart for Third Month…… 43
4.3.11. Barchart for Profit/Loss in Region 1…….. 44
4.3.12. Barchart for Profit/Loss in Region 2……... 44
4.3.13. Barchart for Profit/Loss in Region 3……... 45
4.3.14. Barchart for Profit/Loss in Region 4……... 45
4.3.15. Barchart for Profit/Loss in Third Month ……..... 46
4.4.Fourth Month Data Analytics Output…….. 46
4.4.1. Barchart for Region 1…….. 46
4.4.2. Piechart for Region 1……... 47
4.4.3. Barchart for Region 2…….…. 47
4.4.4. Piechart for Region 2……... 48
4.4.5. Barchart for Region 3…….……. 48
4.4.6. Piechart for Region 3……... 49
4.4.7. Barchart for Region 4…….. 49
4.4.8. Piechart for Region 4……... 50
4.4.9. Barchart for Fourth Month……... 50
4.4.10. Piechart for Fourth Month……... 51
4.4.11. Barchart for Profit/Loss in Region 1……... 51
4.4.12. Barchart for Profit/Loss in Region 2……... 52
4.4.13. Barchart for Profit/Loss in Region 3……..... 52
4.4.14. Barchart for Profit/Loss in Region 4……..... 53
4.4.15. Barchart for Profit/Loss in Fourth Month ……... 53
4.5.Fifth Month Data Analytics Output…….. 54

4.5.1. Barchart for Region 1…….. 54
4.5.2. Piechart for Region 1……... 54
4.5.3. Barchart for Region 2…….. 55
4.5.4. Piechart for Region 2……... 55
4.5.5. Barchart for Region 3……... 56
4.5.6. Piechart for Region 3……... 56
4.5.7. Barchart for Region 4…….. 57
4.5.8. Piechart for Region 4……... 57
4.5.9. Barchart for Fifth Month……. 58
4.5.10. Piechart for Fifth Month…….. 58
4.5.11. Barchart for Profit/Loss in Region 1……... 59
4.5.12. Barchart for Profit/Loss in Region 2……... 59
4.5.13. Barchart for Profit/Loss in Region 3……... 60
4.5.14. Barchart for Profit/Loss in Region 4……... 60
4.5.15. Barchart for Profit/Loss in Fifth Month ……. 61
4.6.Sixth Month Data Analytics Output…… 61
4.6.1. Barchart for Region 1……... 61
4.6.2. Piechart for Region 1……... 62
4.6.3. Barchart for Region 2……... 62
4.6.4. Piechart for Region 2……..... 63
4.6.5. Barchart for Region 3…….. 63
4.6.6. Piechart for Region 3……... 64
4.6.7. Barchart for Region 4…….. 64
4.6.8. Piechart for Region 4……... 65
4.6.9. Barchart for Sixth Month……. 65
4.6.10. Piechart for Sixth Month……. 66
4.6.11. Barchart for Profit/Loss in Region 1…… 66
4.6.12. Barchart for Profit/Loss in Region 2…… 67
4.6.13. Barchart for Profit/Loss in Region 3…… 67
4.6.14. Barchart for Profit/Loss in Region 4…… 68
4.6.15. Barchart for Profit/Loss in Sixth Month ……. 68
4.7. Predicted Output using Linear Regression…….. 69
5. Conclusion and Future Scope…….. 70
6. References…….. 71

LIST OF FIGURES AND SCREENSHOTS

Fig. 3.2.1. Flow Diagram for one Region…… 19

Fig. 3.2.2. Flow Diagram for one Month……..….. 20

Fig, 3.2.3. Flow Diagram for Profit/Loss……..... 21

Fig. 3.2.4. Flow Diagram for Prediction Model……..... 22

Fig. 4.1.1. Barchart for Region 1, First Month…… 24

Fig. 4.1.2. Piechart for Region 1, First Month……...…….. 24

Fig. 4.1.3. Barchart for Region 2, First Month……...……. 25

Fig. 4.1.4. Piechart for Region 2, First Month……...…….. 25

Fig. 4.1.5. Barchart for Region 3, First Month……...……. 26

Fig. 4.1.6. Piechart for Region 3, First Month……...….. 26

Fig. 4.1.7. Barchart for Region 4, First Month……...……. 27

Fig. 4.1.8. Piechart for Region 4, First Month…….. 27

Fig. 4.1.9. Barchart for all Region, First Month……... 28

Fig. 4.1.10. Piechart for all Region, First Month…….. 28

Fig, 4.1.11. profit/loss graph for Region 1, First Month…….. 29

Fig, 4.1.12. profit/loss graph for Region 2, First Month…….. 29

Fig, 4.1.13. profit/loss graph for Region 3, First Month…….. 30

Fig, 4.1.14. profit/loss graph for Region 4, First Month…….. 30

Fig, 4.1.15. profit/loss graph for all Region, First Month…… 31

Fig. 4.2.1. Barchart for Region 1, Second Month……. 31

Fig. 4.2.2. Piechart for Region 1, Second Month……... 32

Fig. 4.2.3. Barchart for Region 2, Second Month…….. 32

Fig. 4.2.4. Piechart for Region 2, Second Month……... 33

Fig. 4.2.5. Barchart for Region 3, Second Month…… 33

Fig. 4.2.6. Piechart for Region 3, Second Month……. 34

Fig. 4.2.7. Barchart for Region 4, Second Month…… 34

Fig. 4.2.8. Piechart for Region 4, Second Month……. 35

Fig. 4.2.9. Barchart for all Region, Second Month…….. 35

Fig. 4.2.10. Piechart for all Region, Second Month……. 36

Fig, 4.2.11. profit/loss graph for Region 1, Second Month……. 36

Fig, 4.2.12. profit/loss graph for Region 2, Second Month……. 37

Fig, 4.2.13. profit/loss graph for Region 3, Second Month……. 37

Fig, 4.2.14. profit/loss graph for Region 4, Second Month……. 38

Fig, 4.2.15. profit/loss graph for all Region, Second Month…… 38

Fig. 4.3.1. Barchart for Region 1, Third Month…….. 39

Fig. 4.3.2. Piechart for Region 1, Third Month…… 39

Fig. 4.3.3. Barchart for Region 2, Third Month……..… 40

Fig. 4.3.4. Piechart for Region 2, Third Month……... 40

Fig. 4.3.5. Barchart for Region 3, Third Month……..… 41

Fig. 4.3.6. Piechart for Region 3, Third Month…….. 41

Fig. 4.3.7. Barchart for Region 4, Third Month……..… 42

Fig. 4.3.8. Piechart for Region 4, Third Month……..…. 42

Fig. 4.3.9. Barchart for all Region, Third Month……... 43

Fig. 4.3.10. Piechart for all Region, Third Month……..…. 43

Fig, 4.3.11. profit/loss graph for Region 1, Third Month……..…. 44

Fig, 4.3.12. profit/loss graph for Region 2, Third Month……..…. 44

Fig, 4.3.13. profit/loss graph for Region 3, Third Month……..…. 45

Fig, 4.3.14. profit/loss graph for Region 4, Third Month…….. 45

Fig, 4.3.15. profit/loss graph for all Region, Third Month……. 46

Fig. 4.4.1. Barchart for Region 1, Fourth Month…… 46

Fig. 4.4.2. Piechart for Region 1, Fourth Month……. 47

Fig. 4.4.3. Barchart for Region 2, Fourth Month…… 47

Fig. 4.4.4. Piechart for Region 2, Fourth Month……. 48

Fig. 4.4.5. Barchart for Region 3, Fourth Month…… 48

Fig. 4.4.6. Piechart for Region 3, Fourth Month……. 49

Fig. 4.4.7. Barchart for Region 4, Fourth Month…… 49

Fig. 4.4.8. Piechart for Region 4, Fourth Month……. 50

Fig. 4.4.9. Barchart for all Region, Fourth Month…….. 50

Fig. 4.4.10. Piechart for all Region, Fourth Month……. 51

Fig, 4.4.11. profit/loss graph for Region 1, Fourth Month……. 51

Fig, 4.4.12. profit/loss graph for Region 2, Fourth Month……. 52

Fig, 4.4.13. profit/loss graph for Region 3, Fourth Month……. 52

Fig, 4.4.14. profit/loss graph for Region 4, Fourth Month……. 53

Fig, 4.4.15. profit/loss graph for all Region, Fourth Month……..… 53

Fig. 4.5.1. Barchart for Region 1, Fifth Month……..… 54

Fig. 4.5.2. Piechart for Region 1, Fifth Month……. 54

Fig. 4.5.3. Barchart for Region 2, Fifth Month…… 55

Fig. 4.5.4. Piechart for Region 2, Fifth Month……. 55

Fig. 4.5.5. Barchart for Region 3, Fifth Month…… 56

Fig. 4.5.6. Piechart for Region 3, Fifth Month……. 56

Fig. 4.5.7. Barchart for Region 4, Fifth Month…… 57

Fig. 4.5.8. Piechart for Region 4, Fifth Month……. 57

Fig. 4.5.9. Barchart for all Region, Fifth Month……. 58

Fig. 4.5.10. Piechart for all Region, Fifth Month……. 58

Fig, 4.5.11. profit/loss graph for Region 1, Fifth Month……. 59

Fig, 4.5.12. profit/loss graph for Region 2, Fifth Month……. 59

Fig, 4.5.13. profit/loss graph for Region 3, Fifth Month……. 60

Fig, 4.5.14. profit/loss graph for Region 4, Fifth Month……. 60

Fig, 4.5.15. profit/loss graph for all Region, Fifth Month……..…… 61

Fig. 4.6.1. Barchart for Region 1, Sixth Month…… 61

Fig. 4.6.2. Piechart for Region 1, Sixth Month……. 62

Fig. 4.6.3. Barchart for Region 2, Sixth Month…… 62

Fig. 4.6.4. Piechart for Region 2, Sixth Month……. 63

Fig. 4.6.5. Barchart for Region 3, Sixth Month…… 63

Fig. 4.6.6. Piechart for Region 3, Sixth Month……. 64

Fig. 4.6.7. Barchart for Region 4, Sixth Month…… 64

Fig. 4.6.8. Piechart for Region 4, Sixth Month……. 65

Fig. 4.6.9. Barchart for all Region, Sixth Month…….. 65

Fig. 4.6.10. Piechart for all Region, Sixth Month……. 66

Fig, 4.6.11. profit/loss graph for Region 1, Sixth Month……. 66

Fig, 4.6.12. profit/loss graph for Region 2, Sixth Month……..….. 67

Fig, 4.6.13. profit/loss graph for Region 3, Sixth Month……... 67

Fig, 4.6.14. profit/loss graph for Region 4, Sixth Month……..….. 68

Fig, 4.6.15. profit/loss graph for all Region, Sixth Month……..…… 68

Fig. 4.7. Predicted Output Using Linear Regression Model……..… 69

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

The project titled “Pharmaceutical Sales Analytics” is a data analytics and Machine Learning (ML) based project. Various data analytics techniques have been applied to process and anlyse the data. Different types of graphs have also been generated for each and every datasets. Finally a Machine Learning algorithm, “Linear Regression” has been applied to predict the next values of the dataset.

In this project, datasets from different regions of the city has been collected for six months. The different data analytics and preprocessing methods have been applied on the data for each region. The whole data for a month is combined and profit and loss function is applied to find the profit or loss for the month. Finally, after all the data preprocessing techniques, the Machine Learning algorithm, “Linear Regression” is applied and the prediction is made. The detailed description of the project and the modules are discussed further.

1.1. ABOUT THE PROJECT:

The project is titled “Pharmaceutical Sales Analytics”. The area of study of the project is Data Analytics and Machine Learning (ML) and the language used is “Python”. We have used “Jupyter Notebook” from the Anaconda Navigator as the python Interpreter. In this project, we have divided the city into four different regions. Now, three different types of diseases are taken under the study and the medicine for each disease has been chosen. The three different diseases and their respective medicines taken under ther study are discussed as follows: -

1. Malaria – larigo DS , Chloroquine phosphate
2. Typhoid – Ciprofoxacin , Chloramphenicol
3. Cold – Sinarest , Flucold

The tablets sold for malaria are termed as T1, tablets sold for Typhoid are termed as T2 and tablets sold for Cold are termed as T3.

Now, let us take the four different regions. In each region, total of ten medical stores are taken under the survey for a month and data for all the number of tablets for Malaria (T1), Typhoid (T2) and Cold (T3) ordered by the medical stores and the total number of tablets sold in each medical store is noted and a CSV (comma separated values) dataset is created, thus creating a dataset for one month. Similarly, CSV datasets for the other five months are also created and in total we get data for the number of tablets sold for all the three diseases for the total of six months.

Now, different data analytics techniques are applied such as reading and printing the dataset, finding maximum and minimum number of tablets sold in each region, finding the total number of tablets (T1, T2, T3) sold in each region, generating different graphs for the total number of tablets sold and the total percentage of tablets sold in each region.

After doing analytics for all the four regions in a month, the dataset of different regions are combined together making it one whole dataset for the month and the total number of tablets sold and the total percentage of the tablets sold for the whole month in all the four region as a whole is analysed and the graphs are generated.

After the data preprocessing, the dataset of number of tablets ordered are taken and are compared with the datasets that contains the number of tablets sold. A Profit and Loss function is written and the profit or loss is found out by comparing the two types of datasets i.e. ordered and sold tablets data and the graphs for the profit or loss is generated for each region. After all this, the graph for profit or loss for the whole month in different regions is generated and data analytics is completed. These methods are repeated on the datasets for all the six months and graphs for all the months are generated.

Now, after doing all the data analytics processes discussed above, the data for all the six month are collected together and the total of the tablets – T1, T2, T3 are found out for each month. A Linear Regression algorithms is generated for the data of all six months to predict the sales of tablet for the further months. This Linear Regression model predicts about the number of tablets sold in the seventh months and further as requested by the user and the graph is generated along with the prediction.

1.2. OBJECTIVE OF THE PROJECT:

The main objective of this project is to collect data of the number of tablets ordered and sold from different regions in the city and analyse the data using different data analytics method to see the total profit and loss generated for a period of six month. The data is analysed properly and graphs are generated according to the outcome of the analytics.

After performing the data analytics and generating the profit and loss function, all the data for six months are collected and combined together and are trained using the machine learning algorithm – “Linear Regression model” to predict the number of tablets sold for the next months and further, thus giving us a brief idea of the number of tablets which will be sold in future months. This will allow the tablet manufacturing companies to make and distribute the tablets in a limited quantity and less wastage will be there. Due to this, the loss will be reduced, both for the company as well as the pharmaceutical shops and more profits will be generated which will allow the business to run more smoothly and efficiently.

We will be analysing few problems and perform operation on related data sets, such as:

4. Business Improvement
5. Sales Prediction
6. Sales Forecasting

Due to this model prediction, we can also find where the demand of what type of tablets are more and where the demands are less with which we can conclude that in which area what type of diseases are spreading more and general public can be made aware of the type of diseases and the prevention from them.

Thus, due to sales prediction and forecasting of the number of tablets sold, the business for both the producer and the retailer will improve greatly and work will be carried out more efficiently with reduced loss and increased profit. Also, we can predict that what type of disease is spreading in which area and awareness in the society can be spread about the diseases and their ways of prevention.

1.3. MODULES:

There are 4 modules in this project, which is discussed as follows: -

1. Data Analytics
2. Profit and Loss Function
3. Graph Generation
4. Model Generation.

1.3.1. Data Analytics:

In this module, the data collected from the different regions for the number of tablets sold is stored into a CSV format. This data is read and printed in the python interpreter and different analytics have been performed such as finding the maximum and minimum number of tablets sold in each region, finding the sum of each tablets sold in particular region, concatenating the dataset of all the four regions and finding the sum of each tablets sold for the particular month, finding the sum of tablets sold in each store as an individual etc.

1.3.2. Profit and Loss Function:

In this module, a Profit and Loss function is generated. First of all, two sets of datasets are taken in this module for one region – one dataset of tablets sold and other dataset of tablets ordered. Then, sum of tablets sold and tablets ordered are found out for each medical store as an individual. After performing data analytics on both the datasets, a new data set is generated by concatenating both the datasets i.e. dataset for tablets sold and dataset for tablets ordered rowwise for the comparison. Now,

the profit and loss function is written and the profit and loss is found for each of the tablets (T1, T2, T3) and then final output is taken out wheather it was a total profit or total loss for that region and by how many tablets. This function is applied in all the regions separately to find the profit or loss for each region so that proper analytics is being done.

1.3.3. Graph Generation:

In this module, graph is generated for the analytics on datasets. The different types of graphs are generated for different data analytics like, barchart is generated to show the total sum of the tablets shown and the piechart is generated to show what percentage of tablets are sold in a particular region and the particular month. Barchart is also generated to show the profit and loss trends in the data. Finally, Linechart is generated to show the prediction of the tablets to be sold in future after the model generation is done.

1.3.4. Model Generation:

The model generation module is the main and very important module of the project. Here a Linear Regression model is generated based on Machine Learning algorithm. This linear regression model will combine the data for all the six months and the model will train the data so that it can predict the sales of number of tablets for the next months. A Linechart will also be made to show the trends in the prediction done by the predictor. This module will give us the final output of the project.

1.4. PROPOSED SYSTEM:

Through this project, we are proposing a system in which the data is trained for some months and then the system predicts what will further happen. For example, here the data from all the different regions is for all the six months have been trained by a Machine Learning (ML) model and prediction is made that how many tablets will be sold in the upcoming months.

Due to this system, we can predict that what amount of tablets of what kind of disease will be sold in how many numbers which will help both the producer in producing the specific amount of tablets and retailer in ordering and selling the specific amount of tablets for the particular disease, thus reducing the total loss which can happen if the prediction is not done. This model will save time and money for both the producer and retailer and will allow them to work efficiently and smoothly.

Due to the prediction of this system, we can see at which region, what types of tablets are sold more through which we can infer that what type of disease is spreading more in which of the regions.

1.4.1. Expected advantages of Proposed System:

The system is very simple in design and to implement. The system requires very low system resources and the system will work in almost all configurations. It has got following features:

- Ensure data accuracy.
- Minimize manual data entry.
- Minimum time needed for the various processing.
- Greater efficiency.
- Better prediction.
- Increased profit and reduced loss.
- Minimum time required.
- Simple to implement.

1.5. FEASIBILITY STUDY:

Feasibility study is made to see if the project on completion will serve the purpose of the organization for the amount of work, effort and the time that spend on it. Feasibility study lets the developer foresee the future of the project and the usefulness. A feasibility study of a system proposal is according to its workability, which is the impact on the organization, ability to meet their user needs and effective use of resources. Thus when a new application is proposed it normally goes through a feasibility study before it is approved for development.

The document provide the feasibility of the project that is being designed and lists various areas that were considered very carefully during the feasibility study of this project such as Technical, Economic and Operational feasibilities. The following are its features:

1.5.1. Technical Feasibility:

The system must be evaluated from the technical point of view first. The assessment of this feasibility must be based on an outline design of the system requirement in the terms of input, output, programs and procedures. Having identified an outline system, the investigation must go on to suggest the type of equipment, required method developing the system, of running the system once it has been designed.

The project should be developed such that the necessary functions and performance are achieved within the constraints. The project is developed within latest technology. Through the technology may become

obsolete after some period of time, due to the fact that never version of same software supports older versions, the system may still be used. So there are minimal constraints involved with this project.

1.5.2. Economic Feasibility:

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication of the system is economically possible for development.

1.5.3. Behavioral Feasibility:

The project would be beneficial because it satisfies the objectives when developed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible.

CHAPTER 2

ANALYSIS

CHAPTER 2:

ANALYSIS

The observations related to the project were made and anaysed based on the datasets collected and the analytics done. The study was done in the area of supervised Machine Learning (ML) algorithms and different data analytics techniques. The programming language Python was also studied in which the algorithm was developed and the project was made. The proper data analytics was done before on each and every data individually as well as combined and all the graphs were made according to the analytics obtained. Different types of functions were used to perform data analytics such as plot(), sum(), max(), min() etc. A Profit and Loss function was also made to find the total Profit and Loss obtained by the medical store for the particular month in a particular region. Finally, the supervised Machine Learning (ML) algorithm – the Linear Regression Model was made and the desired prediction was made to complete the project.

The observations were made based on the study done are as follows:

- During the duration of the project, a detailed study on data science, data analytics and Supervised Machine Learning Algorithm was done and observations were made based on the above study on how to draw the basic outline of the project.

- While studying the topics related to the project, observations were made about the feasibility and design of the product to be made during the given time.

- While studying about Machine Learning algorithms, difference between Supervised and Unsupervised Machine Learning algorithm were clearly understood and supervised Machine Learning algorithm was taken as per the dataset obtained to further move on with the project.

- The programming language “Python” was used in the Jupyter Notebook Interpreter. The platform used was Anaconda Navigator.

- At the time of working on this project, the data flow was observed and different data analytics techniques was also seen and tested. Graphs were also made on the basis of the analytics performed on the dataset.

- After doing the data analytics and making graphs for data of all the six months, all of the data was combined together and Linear Regression model was made and data was trained on the prediction model to predict the data for future months and get the output.

- Finally, the predictions were made and readings were stored to further analyse about how much tablets will be sold in further months.

2.1. METHODOLOGY

The methodology section discusses and describes elaborately about the basic tools, techniques, technologies, algorithms and models used during the making and completion of the project. In this section, we discuss about the techniques and tools used for design and implementation in the project. The methodology is discussed in the upcoming topics.

2.1.1. Data Analytics:

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.

Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

The process involved in data analysis involves several different steps:

1. The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demographic, income, or gender. Data values may be numerical or be divided by category.
2. The second step in data analytics is the process of collecting it. This can be done through a variety of sources such as computers, online sources, cameras, environmental sources, or through personnel.
3. Once the data is collected, it must be organized so it can be analyzed. Organization may take place on a spreadsheet or other form of software that can take statistical data.
4. The data is then cleaned up before analysis. This means it is scrubbed and checked to ensure there is no duplication or error, and that it is not incomplete. This step helps correct any errors before it goes on to a data analyst to be analyzed.

Predictive analytics moves to what is likely going to happen in the near term. What happened to sales the last time we had a hot summer? How many weather models predict a hot summer this year?

2.1.2. Machine Learning (ML):

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: ***The ability to learn***. Machine learning is actively being used today, perhaps in many more places than one would expect.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Some Machine Learning algorithms:

- Supervised Machine Learning alogorithms
- Unsupervised Machine Learning algorithms
- Semi – Supervised Machine Learning algorithms
- Reinforcement Machine Learning algorithms.

2.1.3. Supervised Machine Learning:

Supervised learning as the name indicates the presence of a supervisor as a teacher. Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

Supervised learning classified into two categories of algorithms:

- **Classification**: A classification problem is when the output variable is a category, such as "Red" or "blue" or "disease" and "no disease".
- **Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

2.1.4. Linear Regression:

In statistics, **linear regression** is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called **multiple linear regression**. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.
- If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response.

2.1.5. Python:

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

It is used for:

- web development (server-side),
- software development,
- mathematics,
- system scripting.

What can python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a functional way.

2.1.6. Anaconda Navigator:

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments

and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, macOS, and Linux.

Why use Navigator:

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages, and use multiple environments to separate these different versions.

The command line program conda is both a package manager and an environment manager, to help data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages and update them, all inside Navigator.

2.1.7. Jupyter Notebook:

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Language of choice

The Notebook has support for over 40 programming languages, including Python, R, Julia, and Scala.

Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer.

Big data integration

Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.

CHAPTER 3

DESIGN AND IMPLEMENTATION

CHAPTER 3

DESIGN AND IMPLEMENTATION

Design is the first step into the development phase for any engineered product or system. A good design is the key to effective system. The term “design” is defined as “the process of applying various techniques and principles for the purpose of defining a process or a system in sufficient detail to permit its physical realization”. It may be defined as a process of applying various techniques and principles for the purpose of defining a device, a process or a system in sufficient detail to permit its physical realization. Software design sits at the technical kernel of the software engineering process and is applied regardless of the development paradigm that is used. The system design develops the architectural detail required to build a system or product. As in the case of any systematic approach, this software too has undergone the best possible design phase fine-tuning all efficiency, performance and accuracy levels. The design phase is a transition from a user-oriented document to a document to the programmers or database personnel.

In this chapter, we will discuss about various designs using the context flow diagram to show the flow of different parts and modules of the project. This will tell us how the project has been made and implemented and the flow of data in the project.

a. LOGICAL DESIGN:

The logical flow of a system and define the boundaries of a system. It includes the following steps:

- Reviews the current physical system – its data flows, file content, volumes, Frequencies etc.
- Prepares output specifications – that is, determines the format, content and Frequency of reports.
- Prepares input specifications – format, content and most of the input functions.
- Prepares edit, security and control specifications.
- Specifies the implementation plan.
- Prepares a logical design walk through of the information flow, output, input, Controls and implementation plan.

b. CONTEXT FLOW DIAGRAM:

The context flow diagram or the flowchart was prepared to observe and analyze the flow of the data, so that the application should be developed properly and without any problem and hustle.

i. Flow Diagram for Dataset of a Region:

Here we show the flow of the data for one region. If we get the analytics for one region, then we can get the analytics of other three regions by performing the same steps.

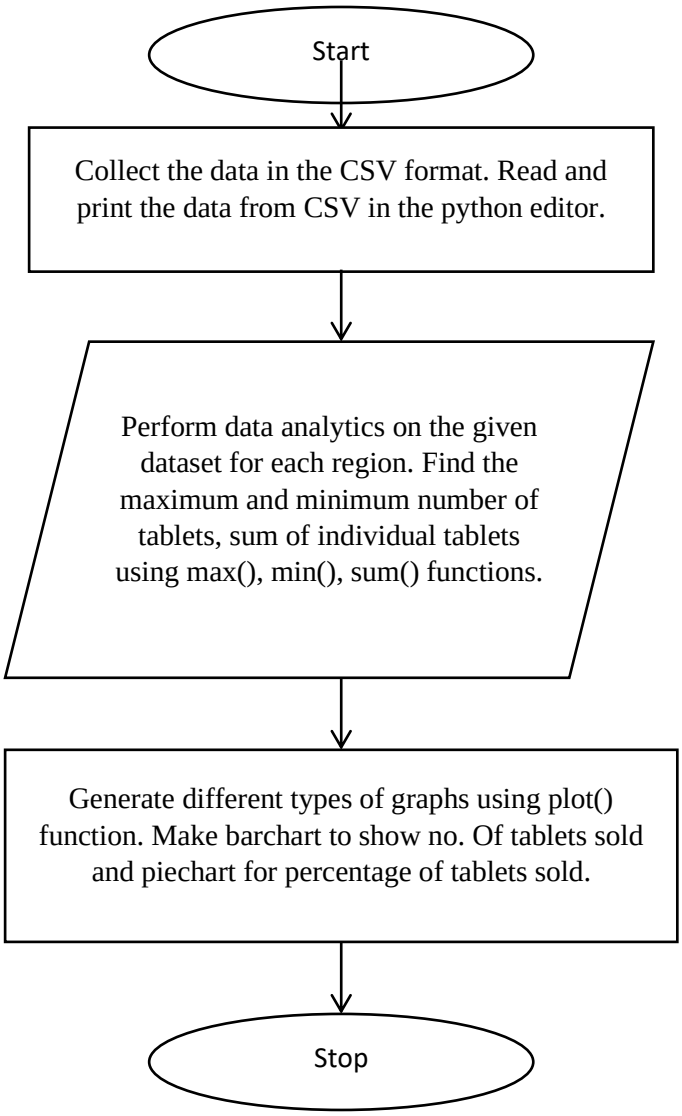


Fig. 3.2.1 Flow diagram for one region.

ii. Flow Diagram of the Dataset for a Month:

Previously, we have done the data flow for one region and reiterated it to rest of the three regions. From this we get the data analytics of the four regions along with the graphs.

Now, we will combine the dataset of all the four region by concat() function and get the data for full one month and then perform the data analytics and graph genration on the dataset obtained.

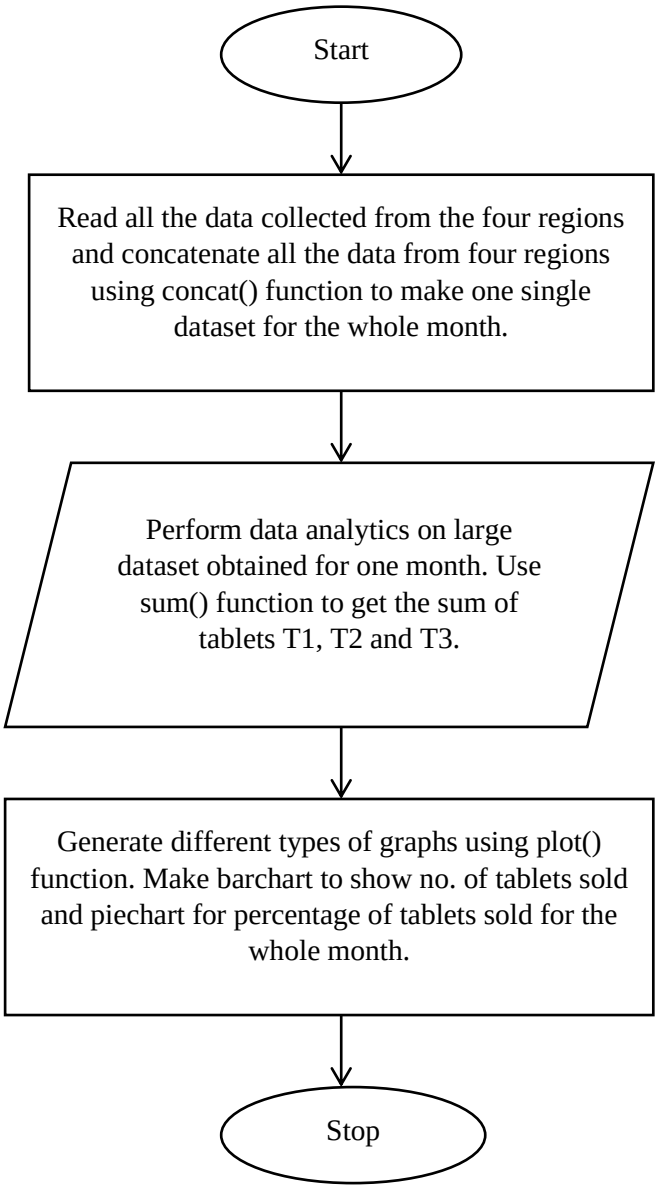


Fig. 3.2.2. Flow Diagram for One Month.

iii. Flow Diagram for finding Profit and Loss:

Now in this part, we read two types of data sets: Dataset for number of tablets ordered and Dataset for number of tablets sold.

Here, we will first read the sold tablets dataset for one region. Then we will add the number of tablets sold in each of the particular stores by adding the data tablets T1, T2 and T3 and store the sum of the individual stores in “total tablets” column. Now we will again use sum() function to find out the total sum of the total tablets sold in each store. We will repeat the same process with the dataset of number of tablets Ordered.

After doing the above analytics, we will concatenate the two datasets obtained by the concat() function for the comparison and apply the profit and loss function to get the profit or loss in that particular region. Generate graphs for the profit and loss.

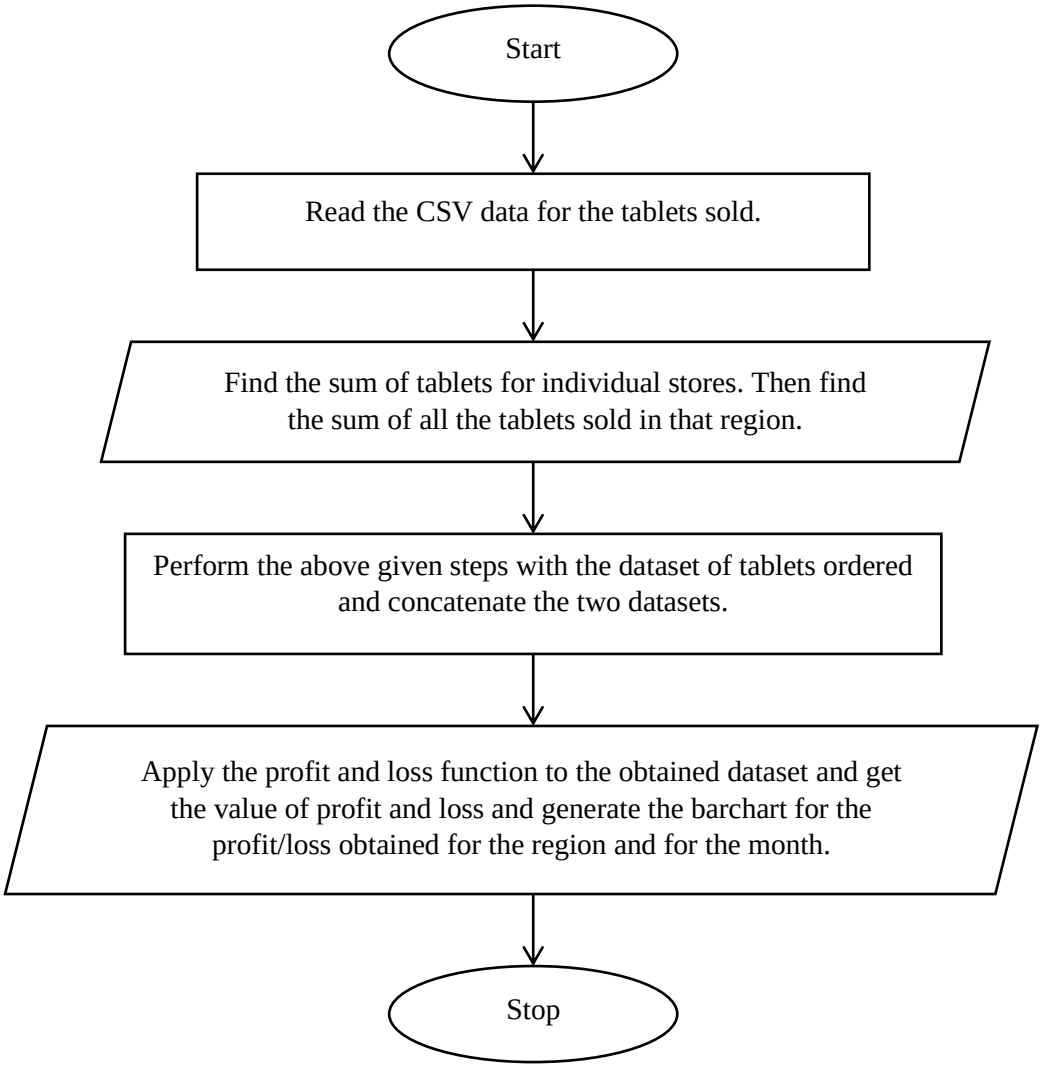


Fig. 3.2.3. Flow diagram for Profit and Loss.

iv. Flow diagram for the Prediction Model:

In this part, we will read the dataset for all the six months. Then we will find out the sum of all the individual tablets per month. By doing so, we will get the sum of all the tablets i.e. T1, T2, T3 for all the six months.

Now, we will apply the Linear Regression model of the Supervised Machine Learning algorithm to get the prediction for the sales of the tablets for the future months. Draw the Linechart graph to show the value of actual prediction graphically.

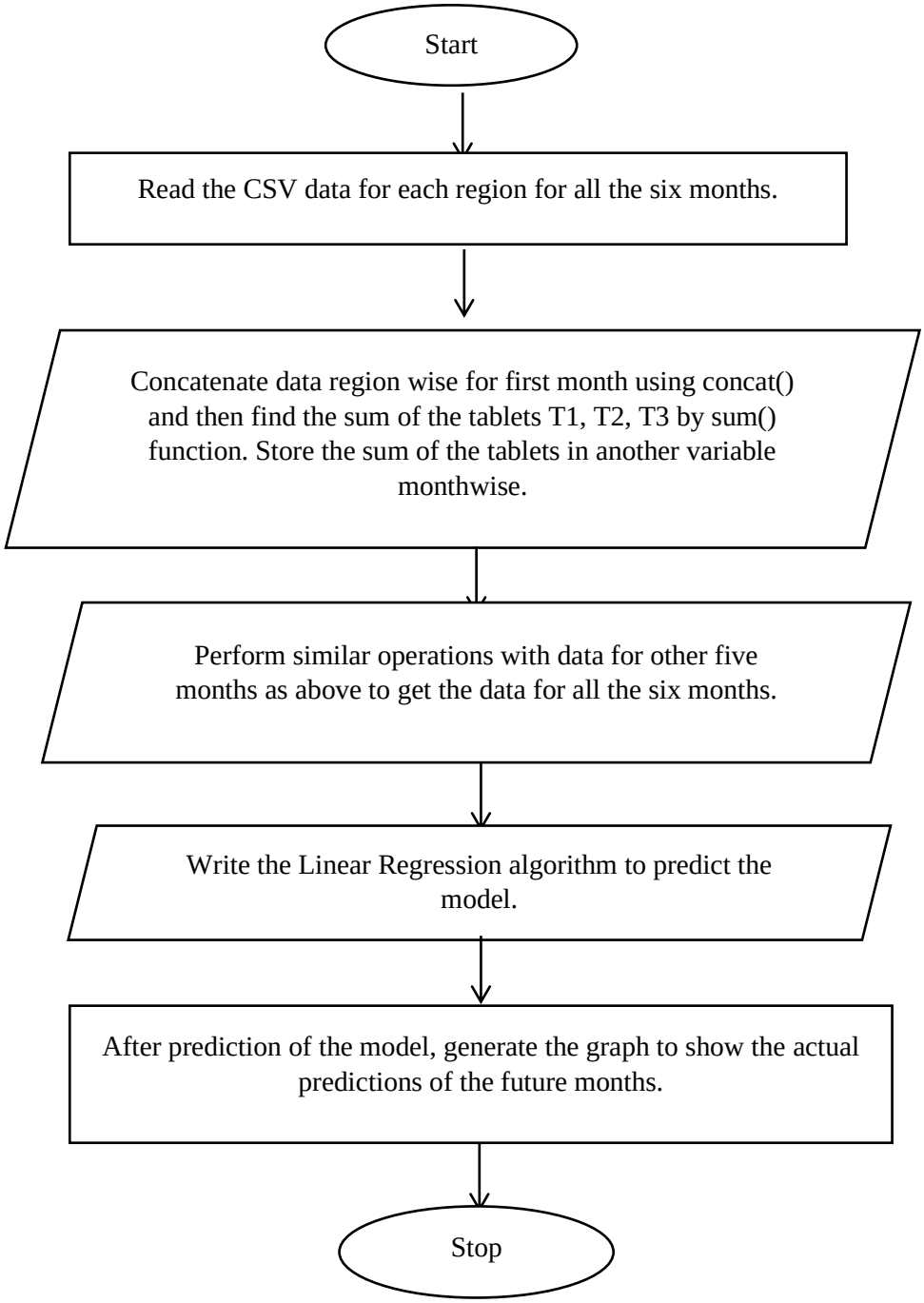


Fig.3.2.4. Flowchart for the Prediction Model.

CHAPTER 4

OUTPUTS AND RESULTS

4.1. FIRST MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of first month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in first month and then the combined graph for the first month and the profit and loss graph for the month.

4.1.1. Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the first month.

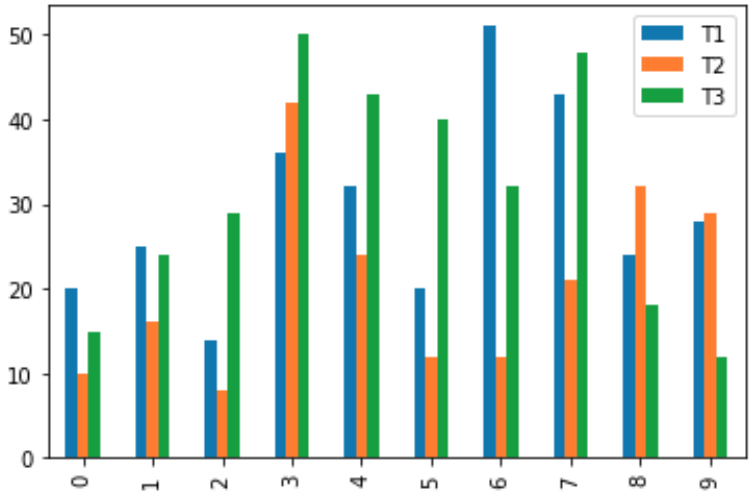


Fig 4.1.1. Barchart for region 1 first month.

4.1.2. Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the first month.

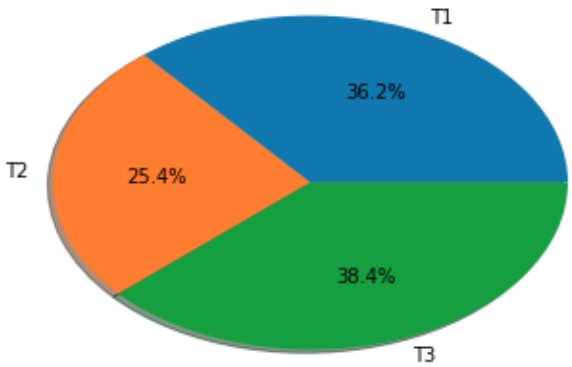


Fig.4.1.2. Piechart for region 1 first month.

4.1.3. Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the first month.

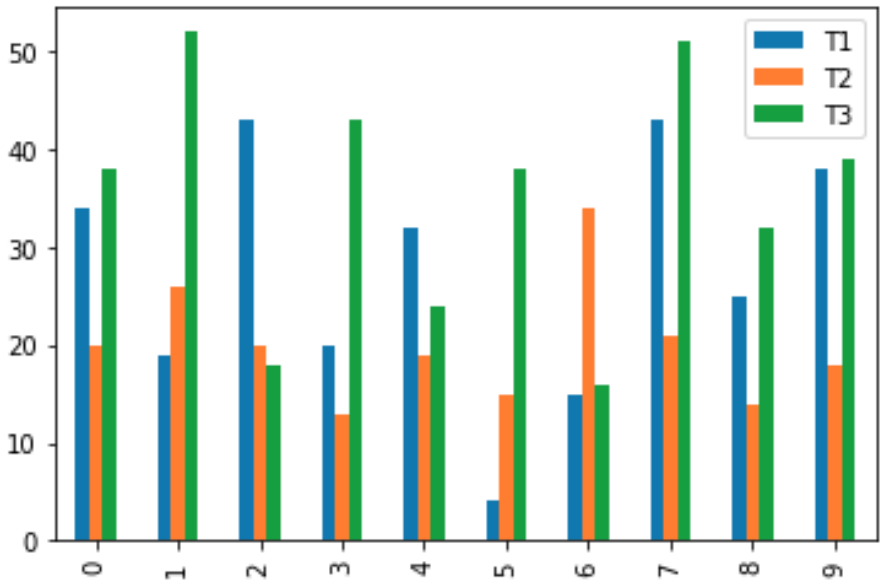


Fig.4.1.3. Barchart for Region 2 first month

4.1.4.Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the first month.

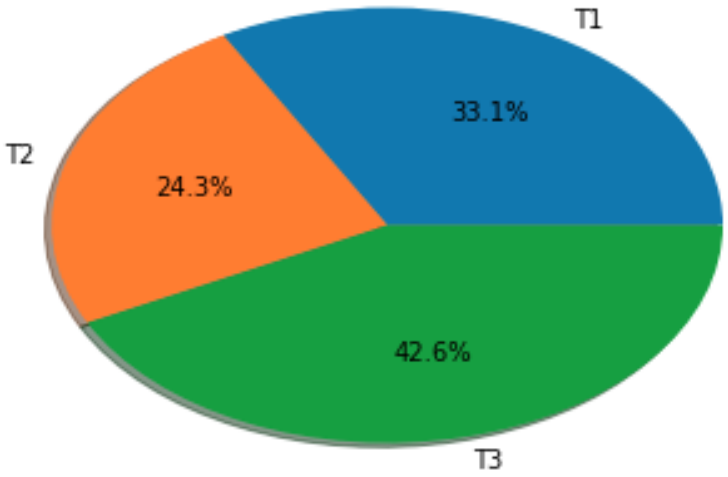


Fig.4.1.4. Piechart for Region 2 first month.

4.1.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the first month.

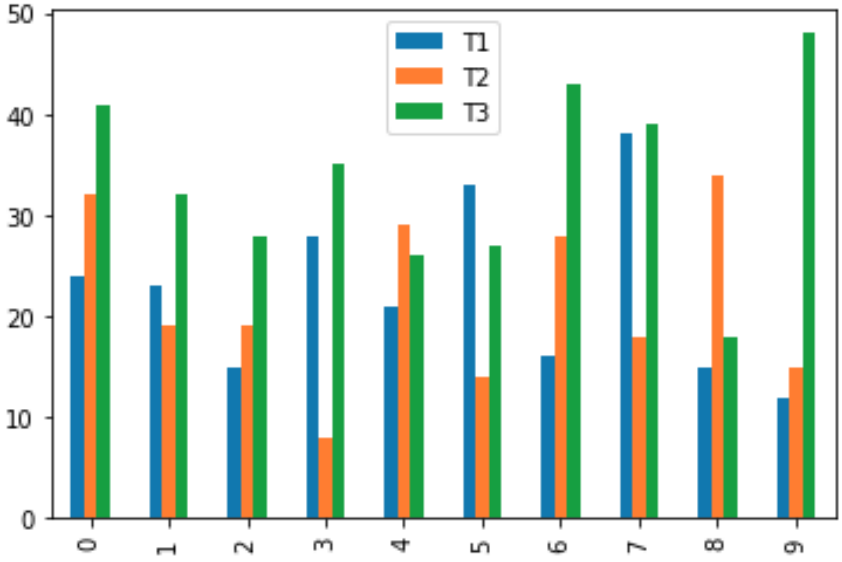


Fig.4.1.5. Barchart for region 3 first month.

4.1.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the first month.

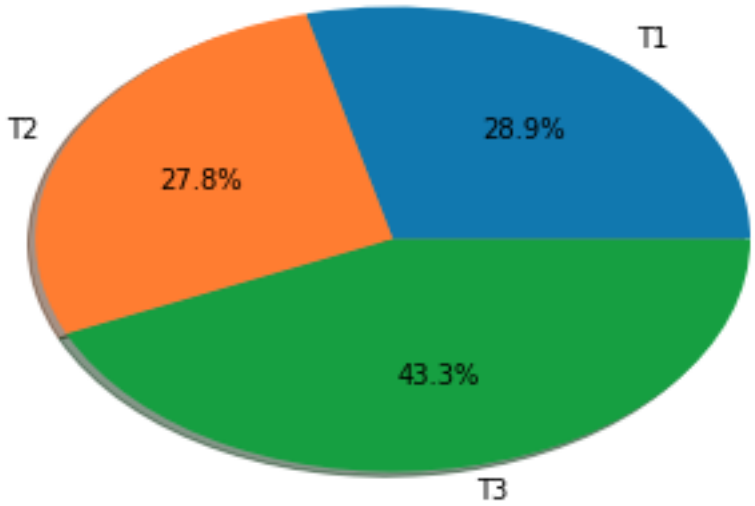


Fig.4.1.6. Piechart for region 3 first month.

4.1.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the first month.

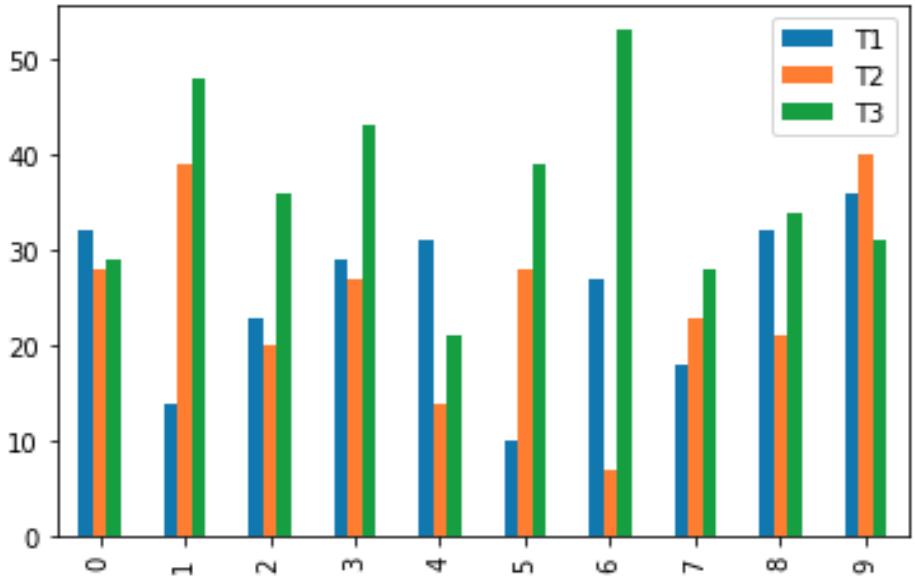


Fig.4.1.7. Barchart for region 4 first month

4.1.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the first month.

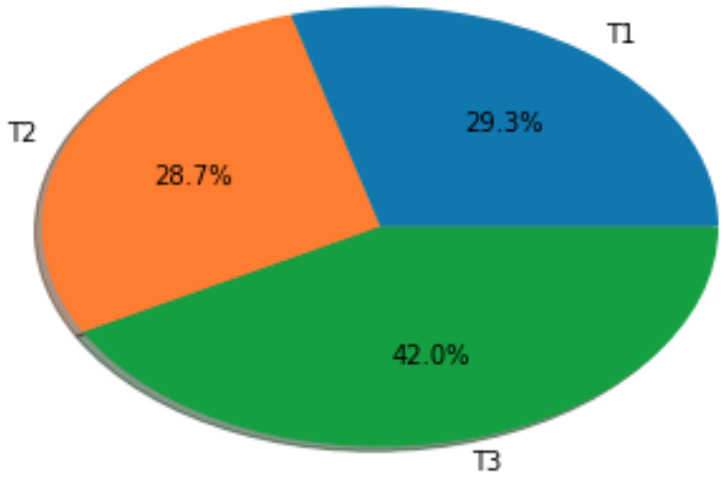


Fig.4.1.8. Piechart for region 4 first month.

4.1.9. Barchart for First Month:

This shows the Barchart for the total number of tablets sold in all region in the first month.

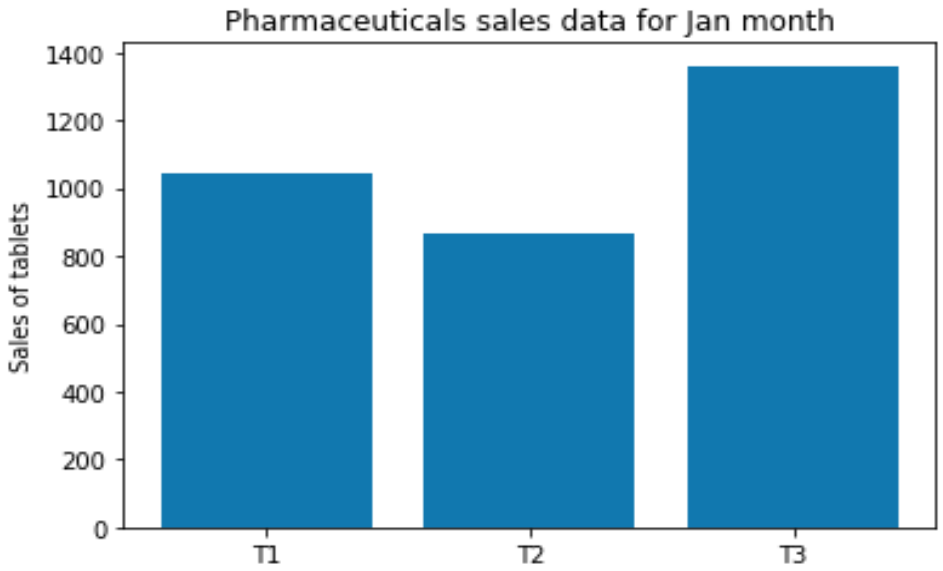


Fig.4.1.9 Barchart for all regions First Month

4.1.10. Piechart for First Month

This shows the Piechart for the total percentage of tablets sold in all region in the first month.

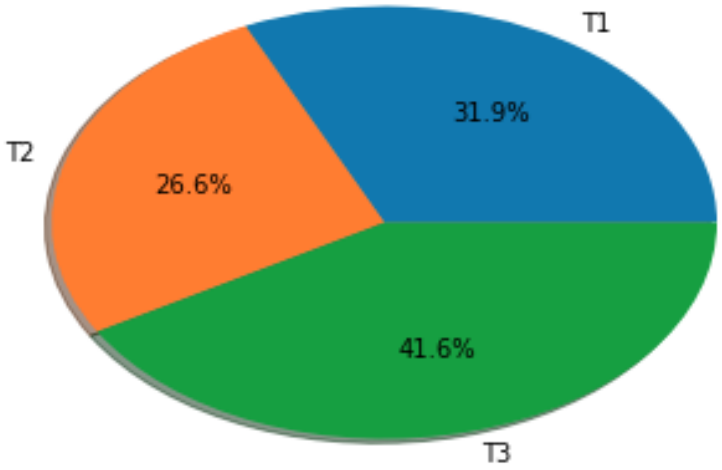
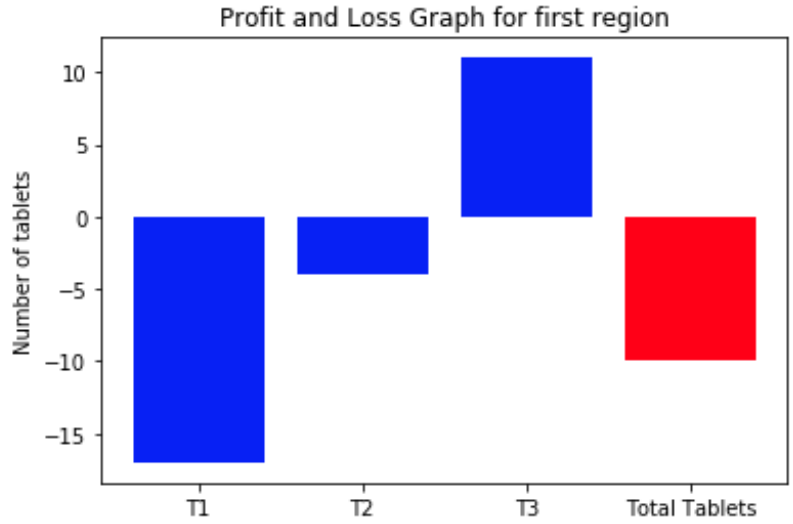


Fig.4.1.10. Piechart for all regions First Month.

4.1.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the first month.

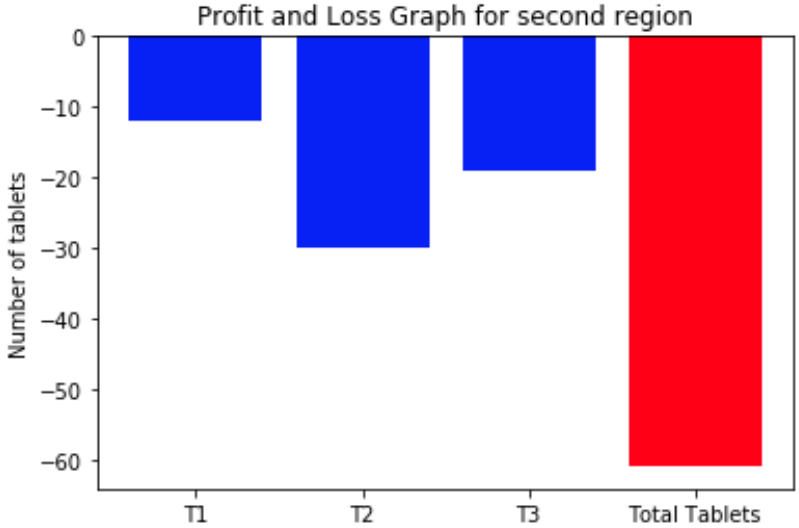



```
loss of 10 tablets
```

Fig.4.1.11. Profit/loss graph for region 1 first month.

4.1.12. Barchart for Profit/Loss in region 2:

This shows the profit or loss happened in region 2 in the first month.




```
loss of 61 tablets
```

Fig,4.1.12. Profit/loss graph for region 2 first Month.

4.1.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the first month.

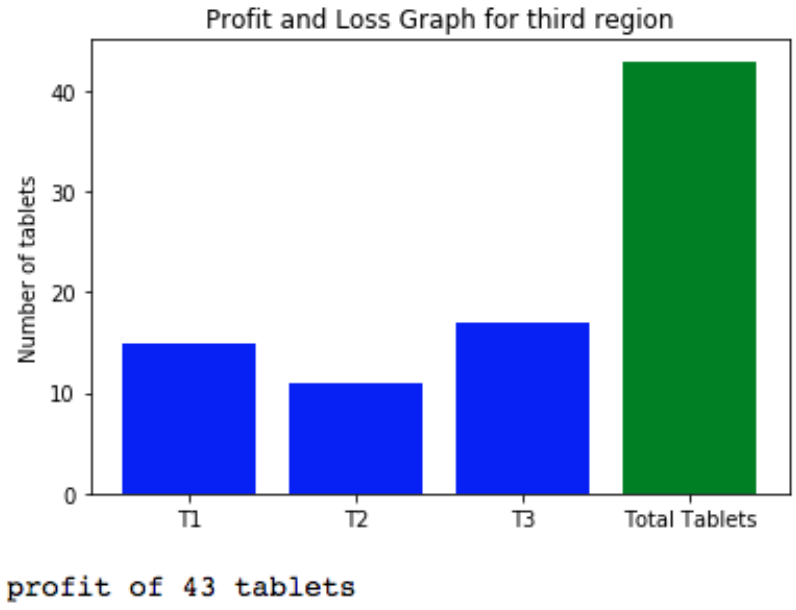


Fig.4.1.13. Profit/loss graph for region 3 first month.

4.1.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the first month.

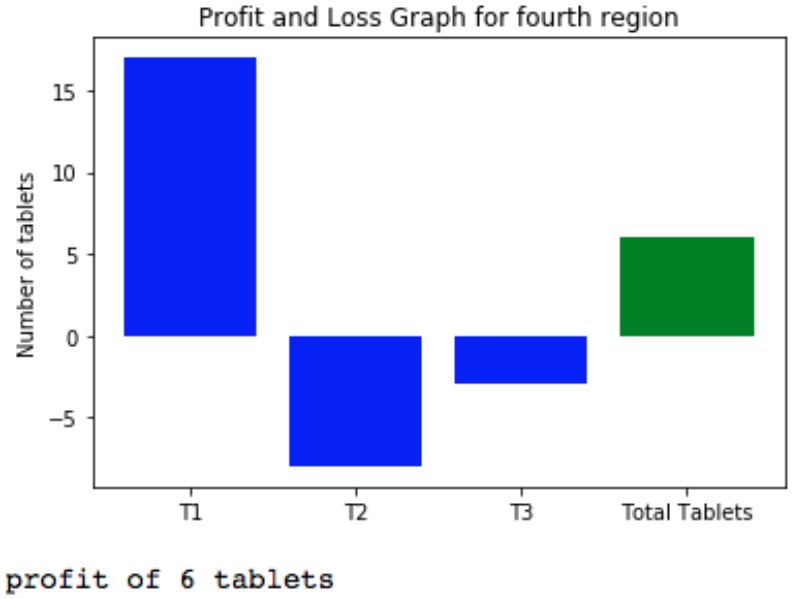


Fig. 4.1.14. Profit/loss graph for region 4 first month.

4.1.15. Barchart for profit/loss in first month:

This shows the profit or loss happened in all regions in the first month.

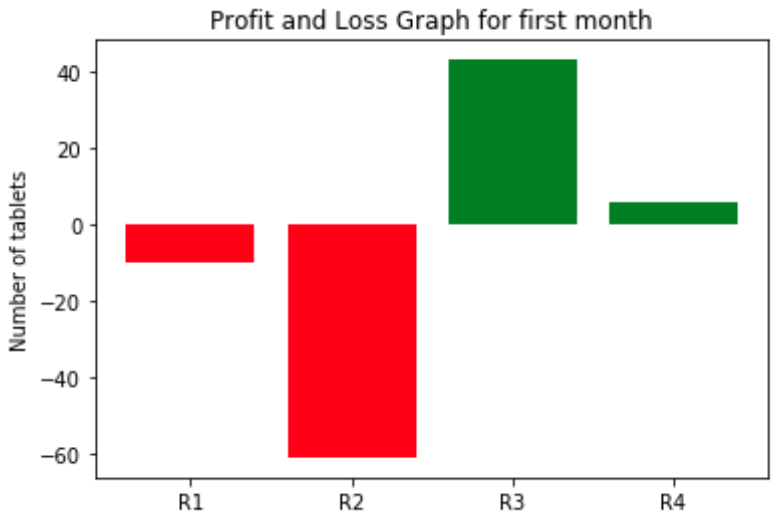


Fig.4.1.15. Profit/loss graph for all regions first month.

4.2 SECOND MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of second month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in second month and then the combined graph for the second month and the profit and loss graph for the month.

4.2.1 Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the second month.

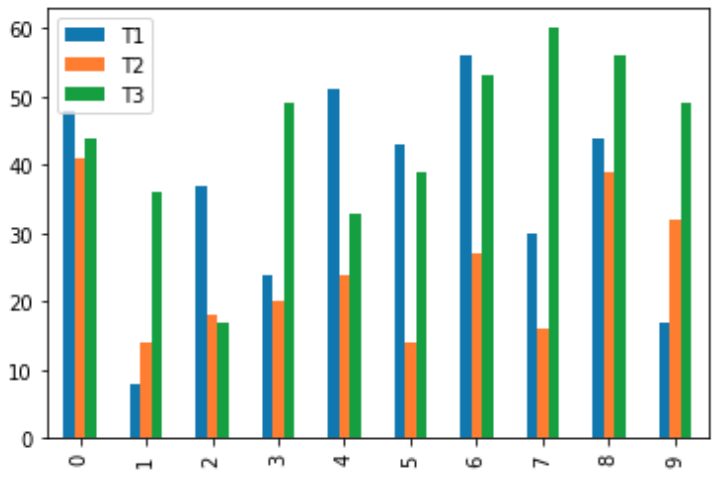


Fig.4.2.1. Barchart for region 1, second month

4.2.2 Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the second month.

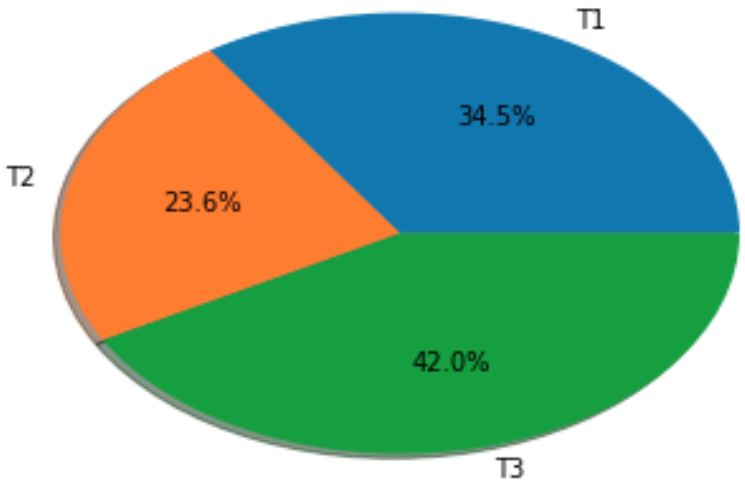


Fig.4.2.2. Piechart for region 1, second month.

4.2.3 Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the second month.

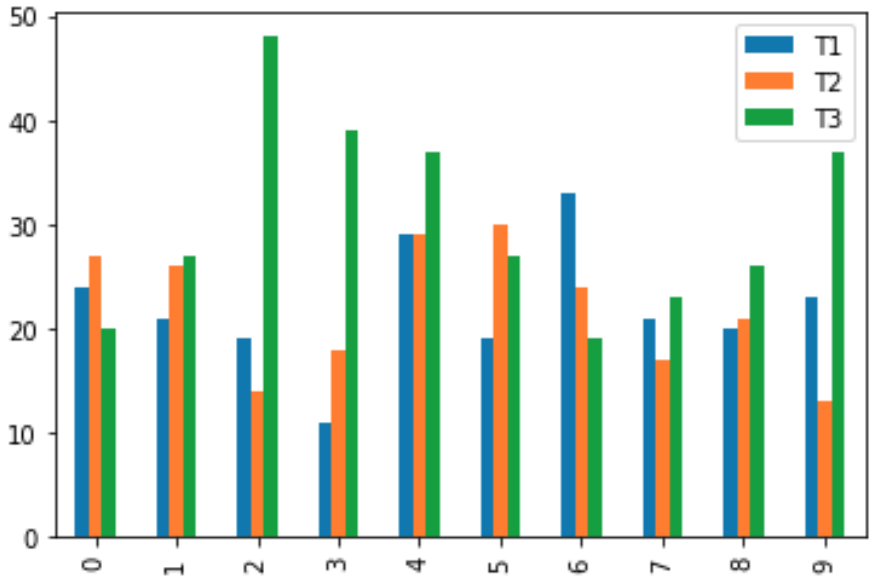


Fig.4.2.3. Barchart for Region 2, second month

4.2.4 Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the second month.

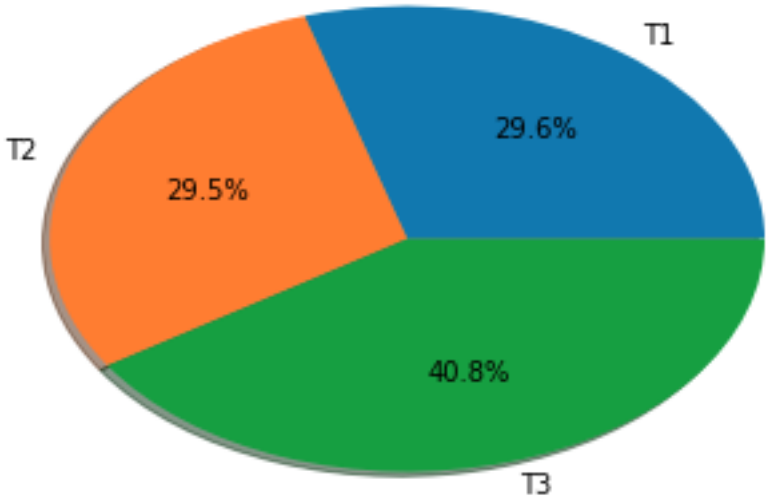


Fig.4.2.4. Piechart for Region 2, second month.

4.2.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the second month.

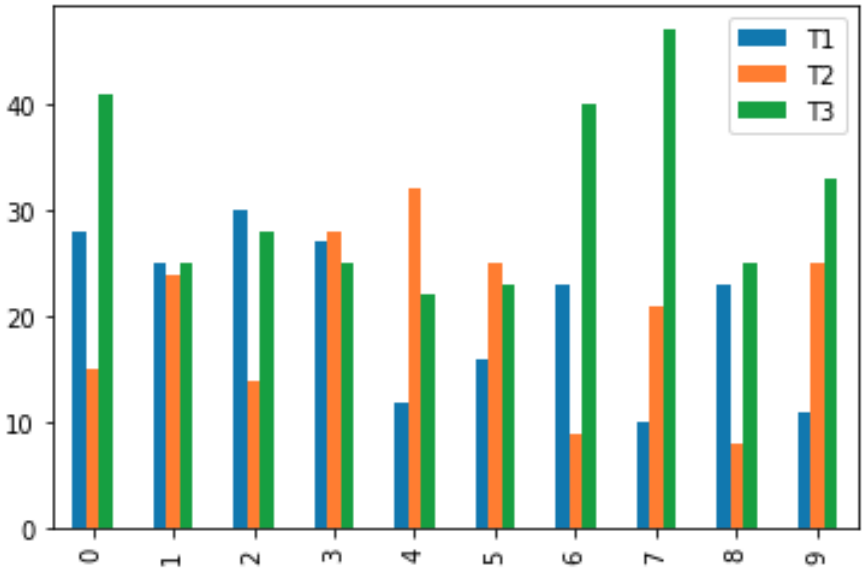


Fig.4.2.5. Barchart for region 3, second month.

4.2.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the second month.

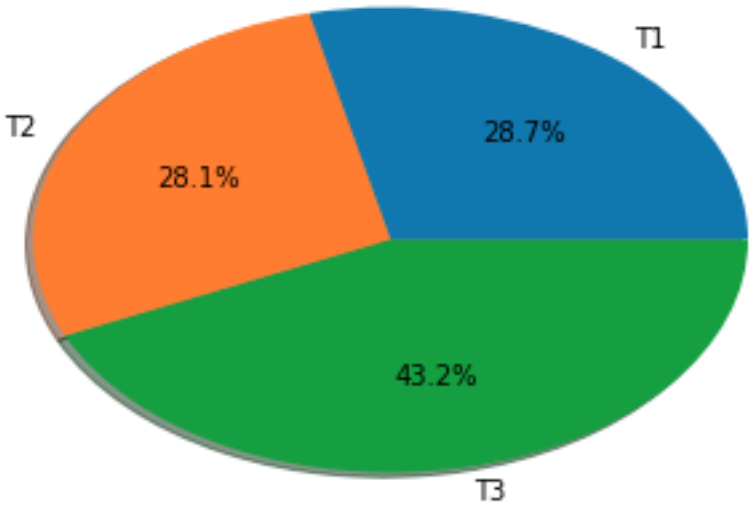


Fig.4.2.6. Piechart for region 3, second month.

4.2.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the second month.

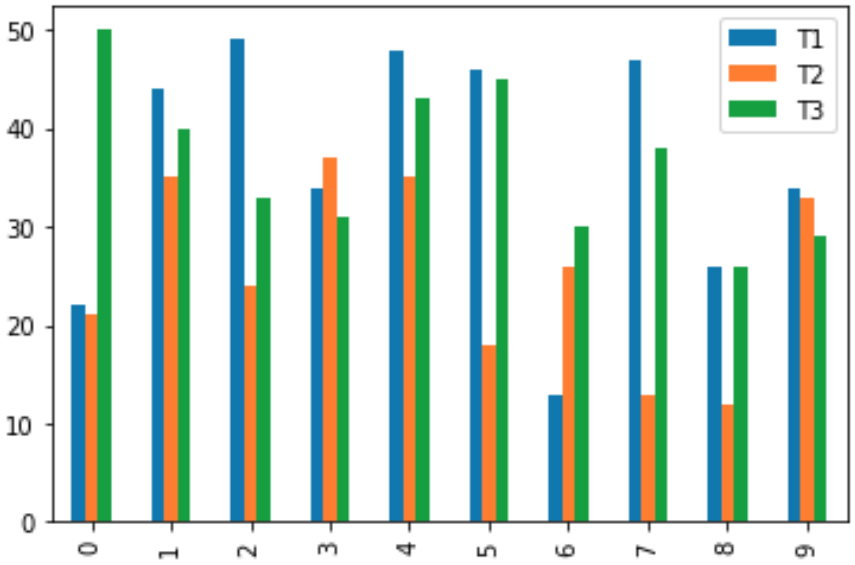


Fig.4.2.7. Barchart for region 4, second month

4.2.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the second month.

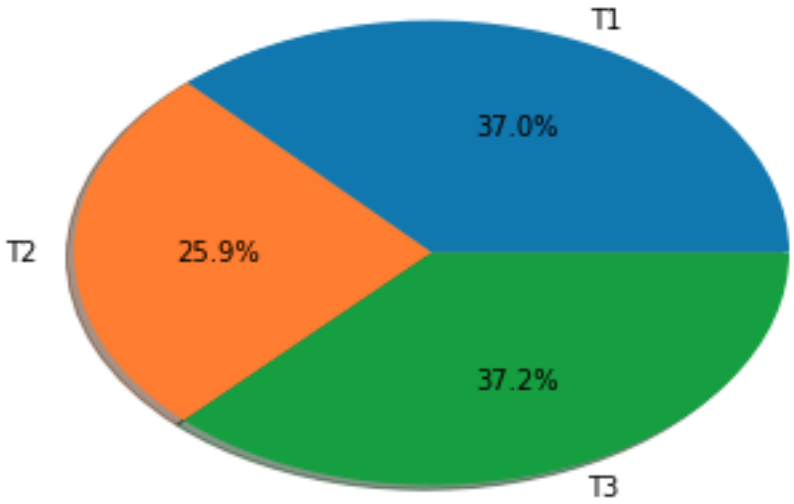


Fig.4.2.8. Piechart for region 4, second month.

4.2.9. Barchart for Second Month:

This shows the Barchart for the total number of tablets sold in all regions in the second month.

Fig.4.2.9 Barchart for all regions second Month

4.2.10. Piechart for Second Month

This shows the Piechart for the total percentage of tablets sold in all regions in the second month.

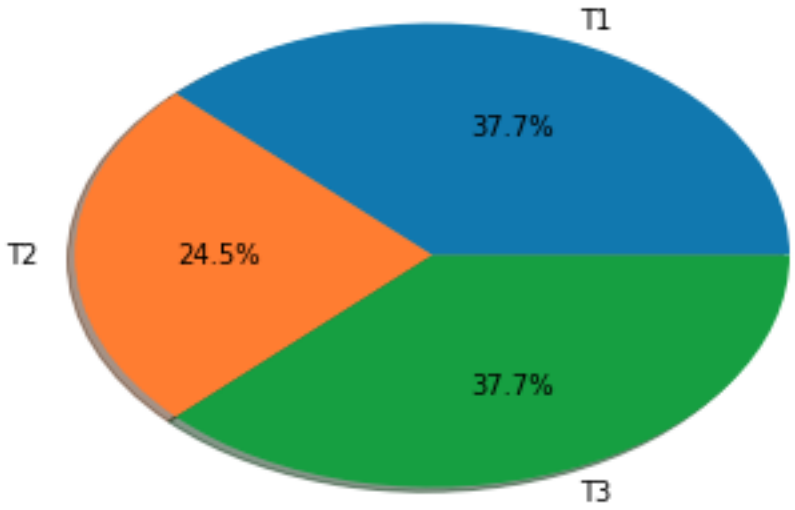
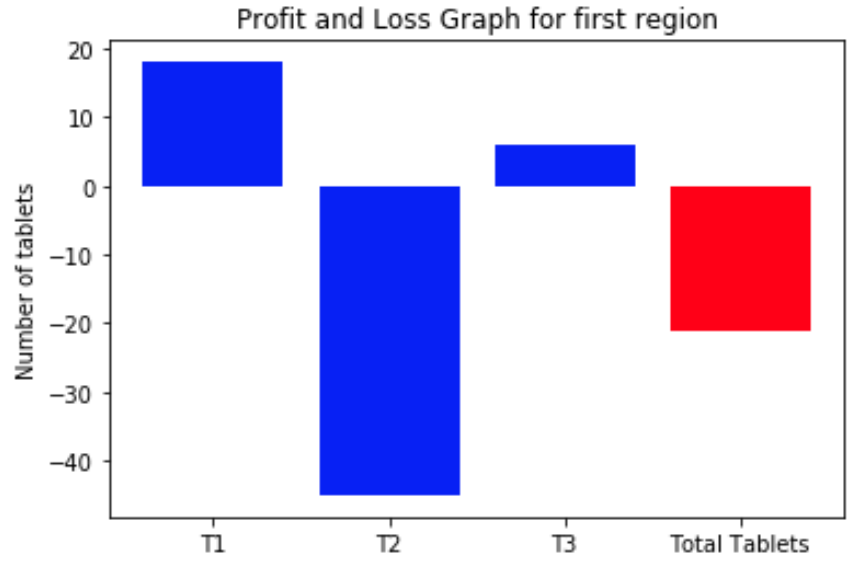


Fig.4.2.10. Piechart for all regions Second Month.

4.2.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the second month.

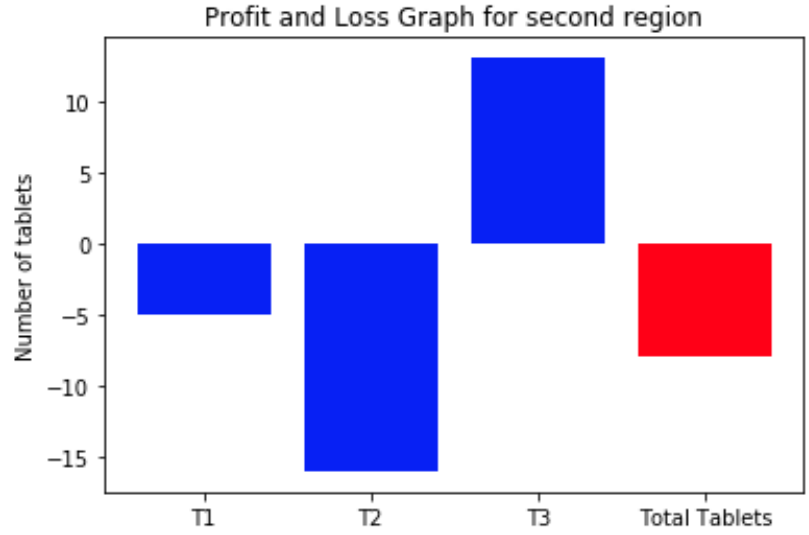



```
loss of 21 tablets
```

Fig.4.2.11. Profit/loss graph for region 1, second month.

4.2.12. Barchart for Profit/Loss in region 2:

This shows the profit or loss happened in region 2 in the second month.

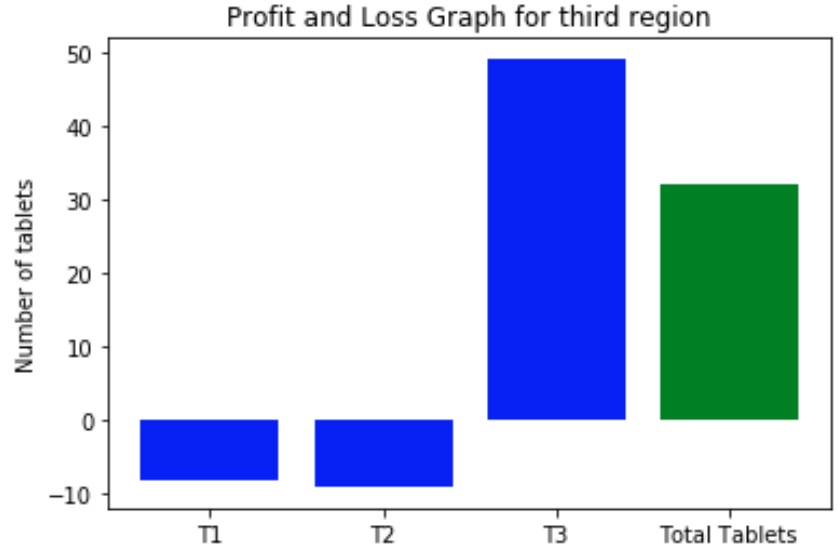



```
loss of 8 tablets
```

Fig,4.2.12. Profit/loss graph for region 2 second Month.

4.2.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the second month.




```
profit of 32 tablets
```

Fig.4.2.13. Profit/loss graph for region 3, second month.

4.2.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the second month.

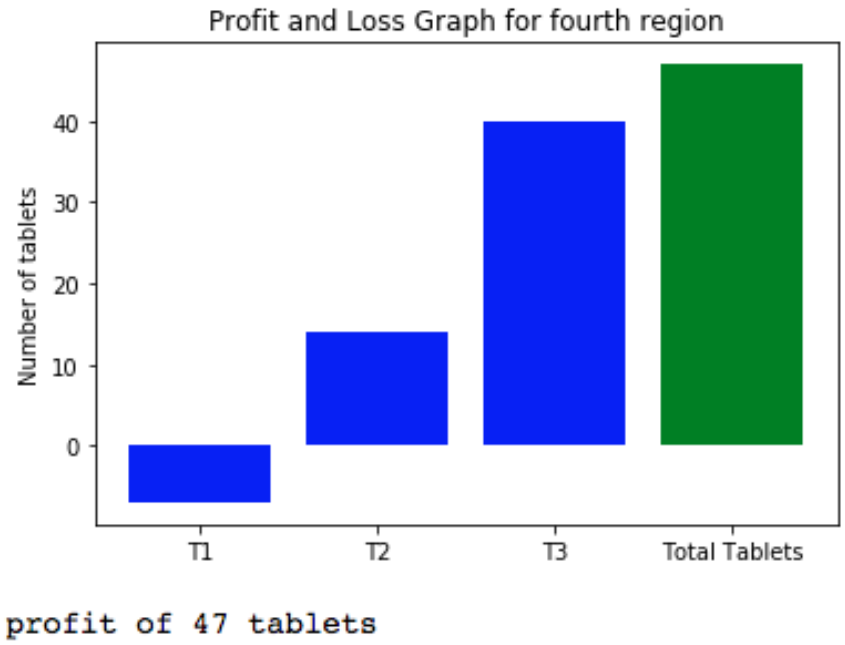


Fig. 4.2.14. Profit/loss graph for region 4, second month.

4.2.15. Barchart for profit/loss in second month:

This shows the profit or loss happened in all regions in the second month.

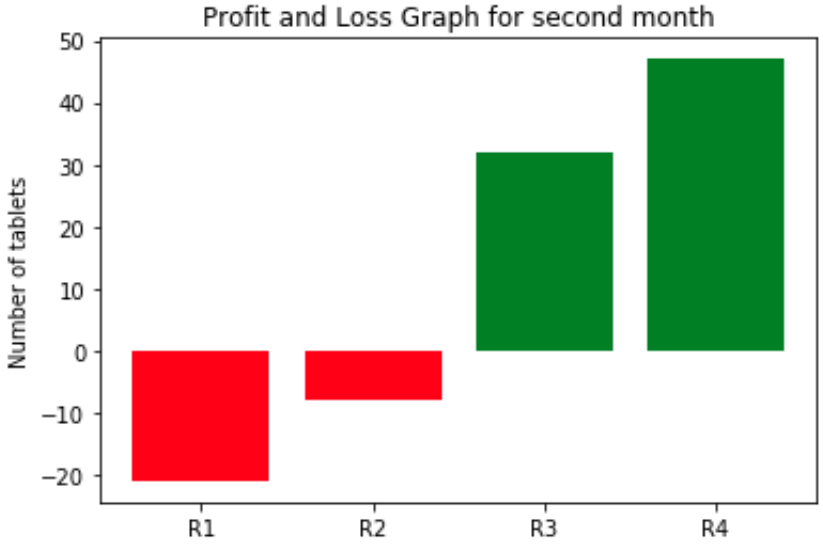


Fig.4.2.15. Profit/loss graph for all regions, second month.

4.3 THIRD MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of third month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in third month and then the combined graph for the third month and the profit and loss graph for the month.

4.3.1 Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the third month.

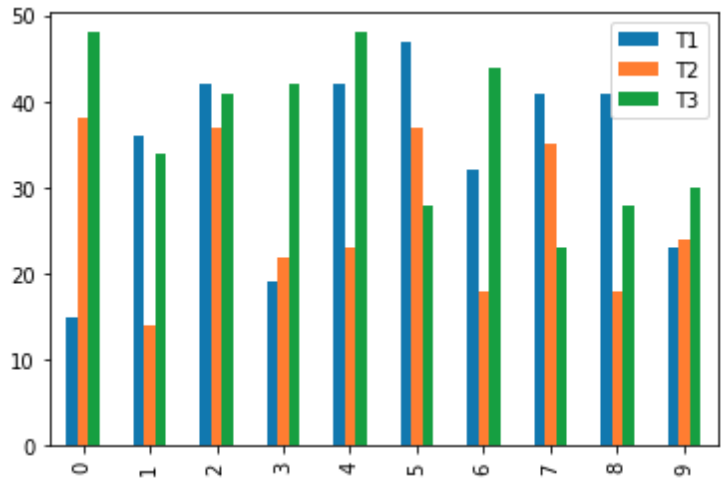


Fig 4.3.1. Barchart for region 1, third month.

4.3.2 Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the third month.

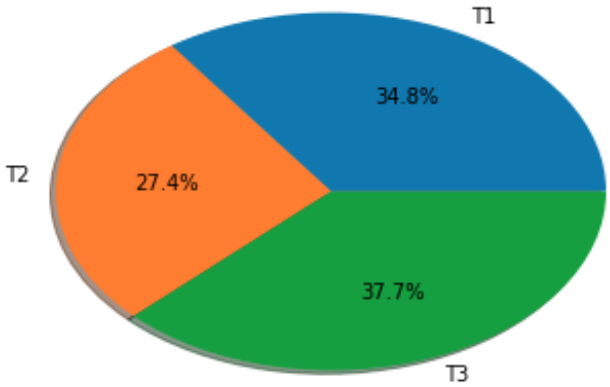


Fig.4.3.2. Piechart for region 1, third month.

4.3.3 Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the third month.

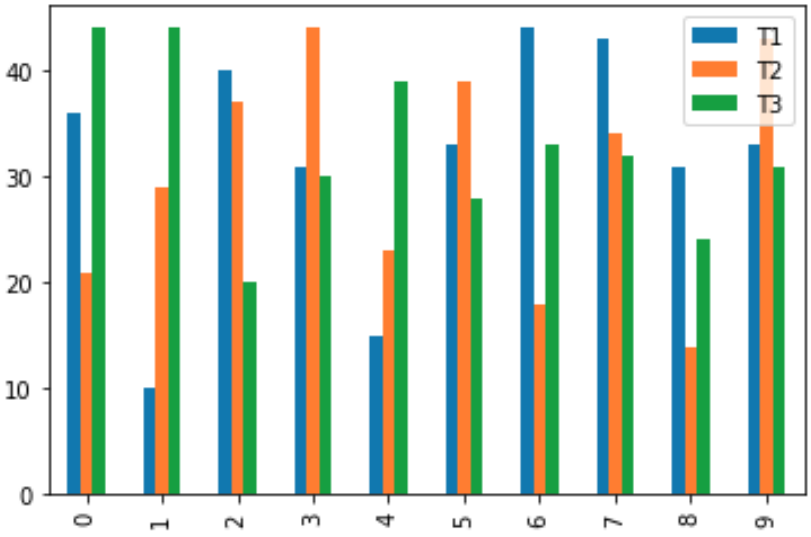


Fig.4.3.3. Barchart for Region 2, third month

4.3.4.Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the third month.

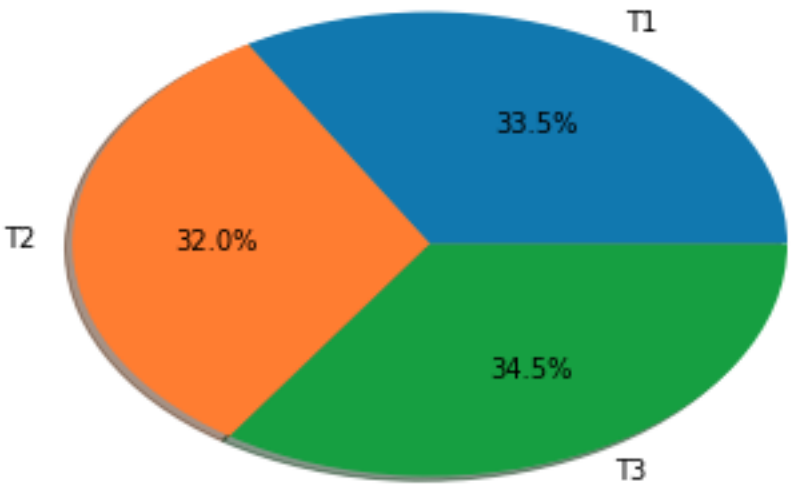


Fig.4.3.4. Piechart for Region 2, third month.

4.3.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the third month.

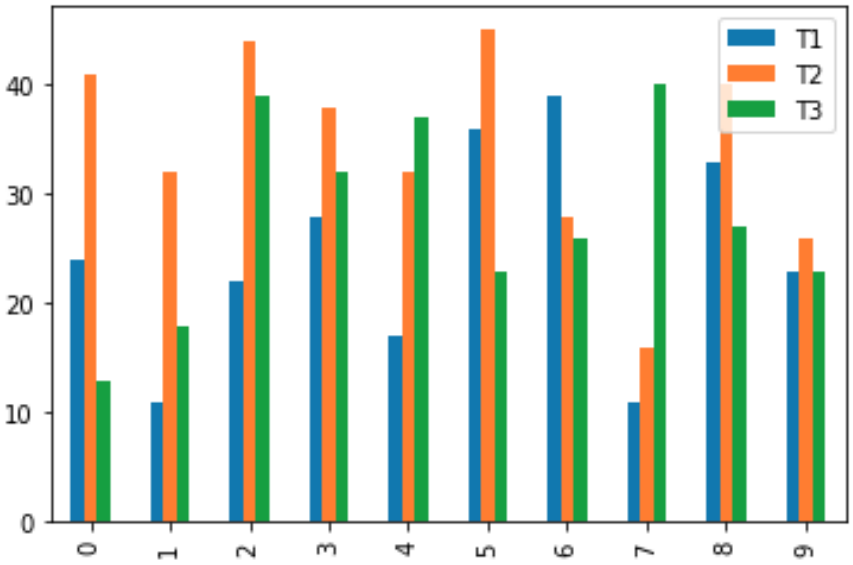


Fig.4.3.5. Barchart for region 3, third month.

4.3.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the third month.

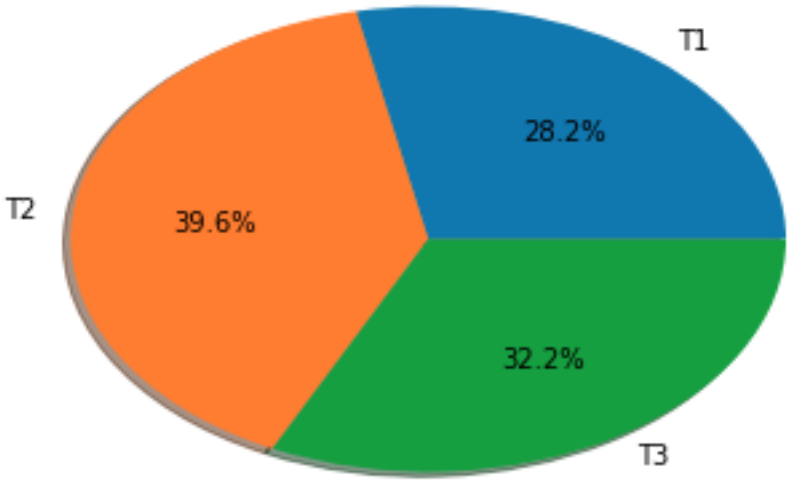


Fig.4.3.6. Piechart for region 3, third month.

4.3.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the third month.

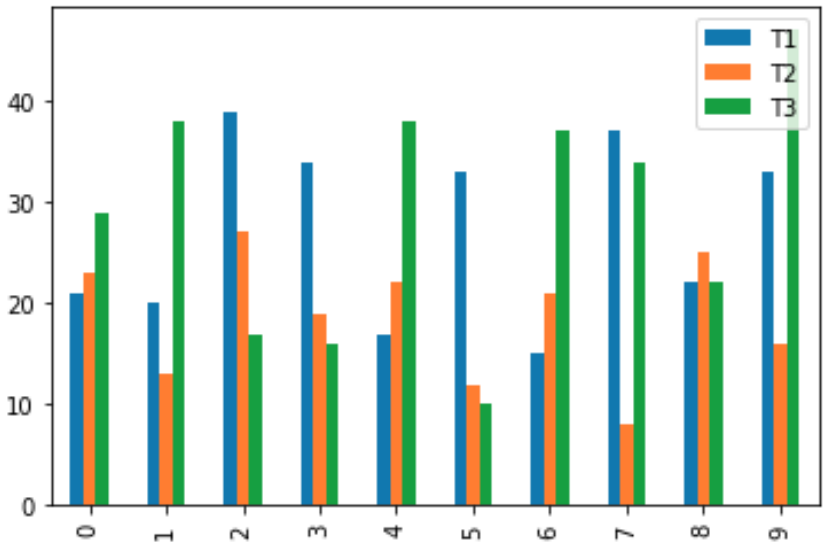


Fig.4.3.7. Barchart for region 4, third month

4.3.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the third month.

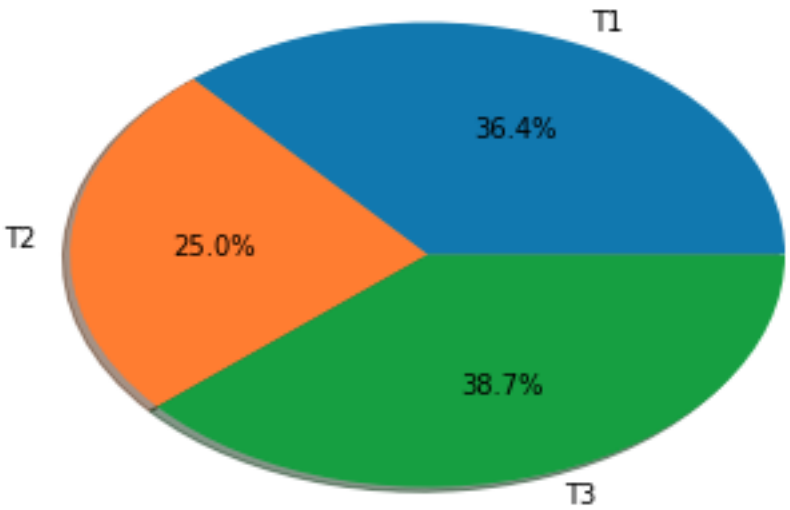


Fig.4.3.8. Piechart for region 4, third month.

4.3.9. Barchart for Third Month:

This shows the Barchart for the total number of tablets sold in all regions in the third month.

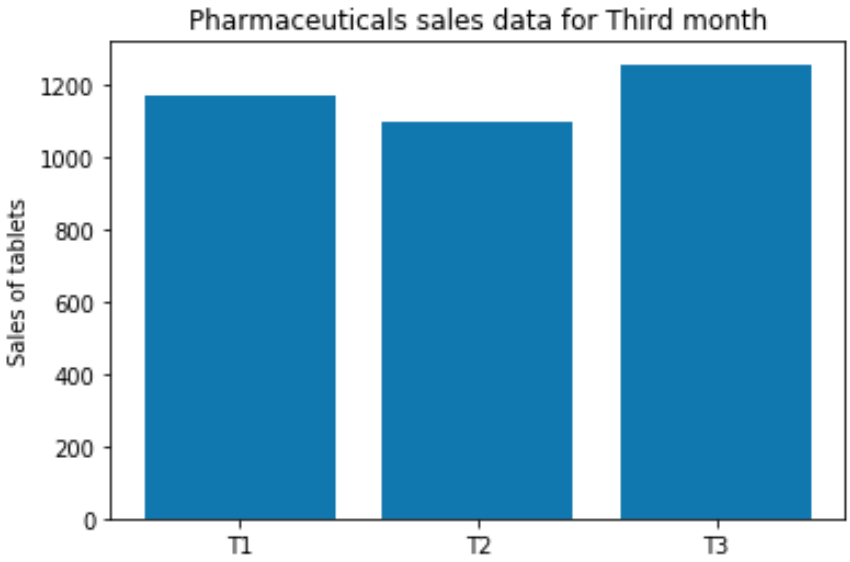


Fig.4.3.9 Barchart for all regions Third Month

4.3.10. Piechart for third Month

This shows the Piechart for the total percentage of tablets sold in all regions in the third month.

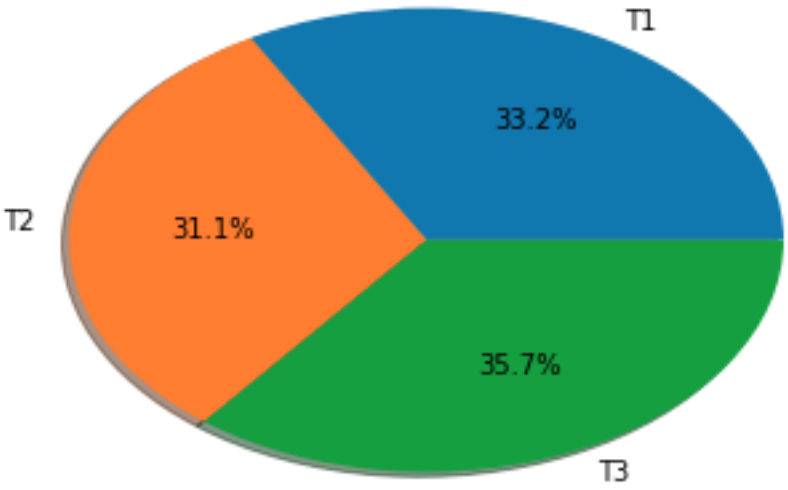


Fig.4.3.10. Piechart for all regions Third Month.

4.3.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the third month.

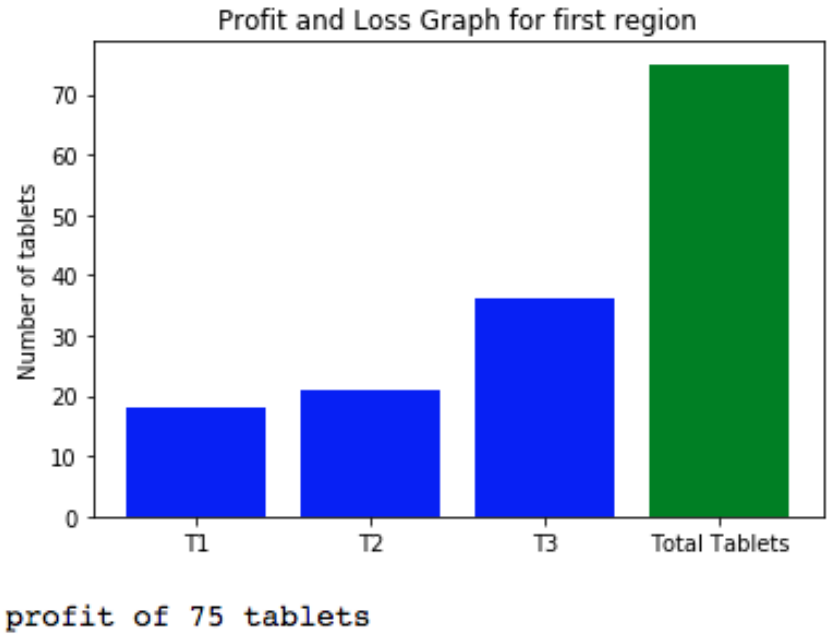
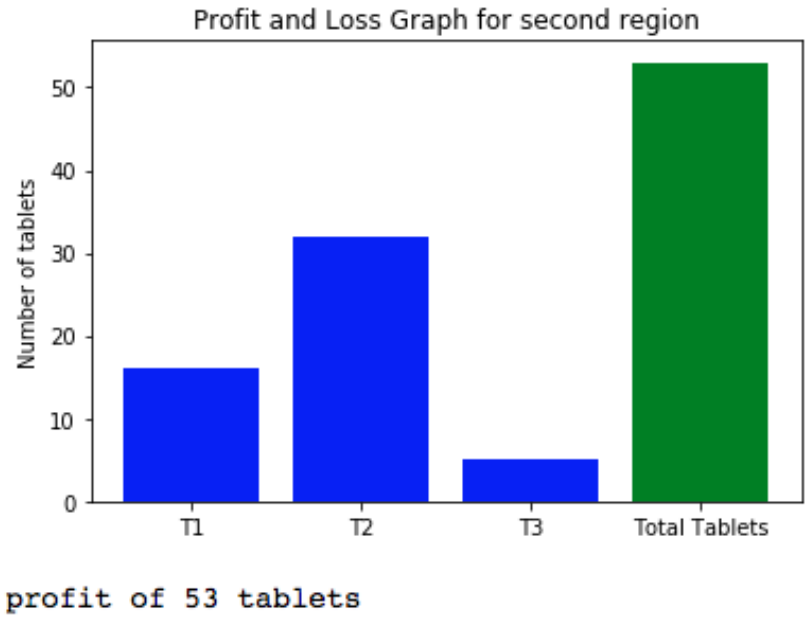


Fig.4.3.11. Profit/loss graph for region 1 third month.

4.3.12. Barchart for Profit/Loss in region 2:

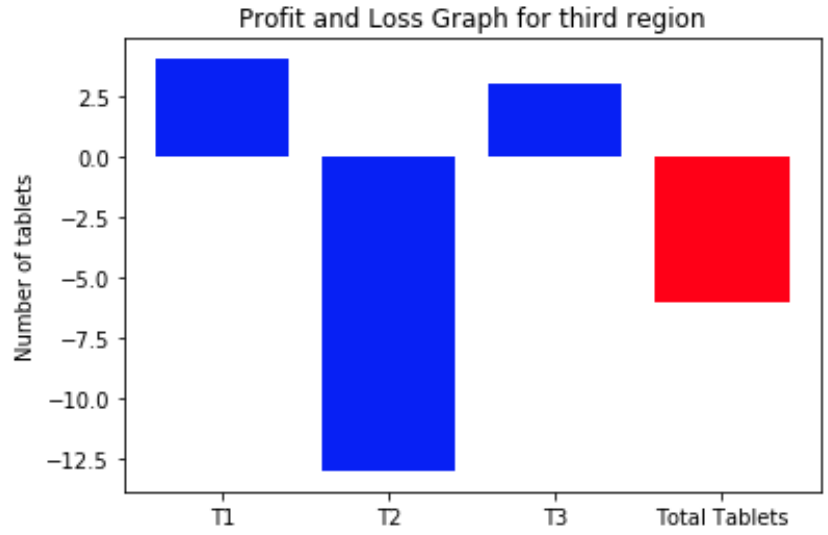
This shows the profit or loss happened in region 2 in the third month.



Fig,4.3.12. Profit/loss graph for region 2 third Month.

4.3.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the third month.

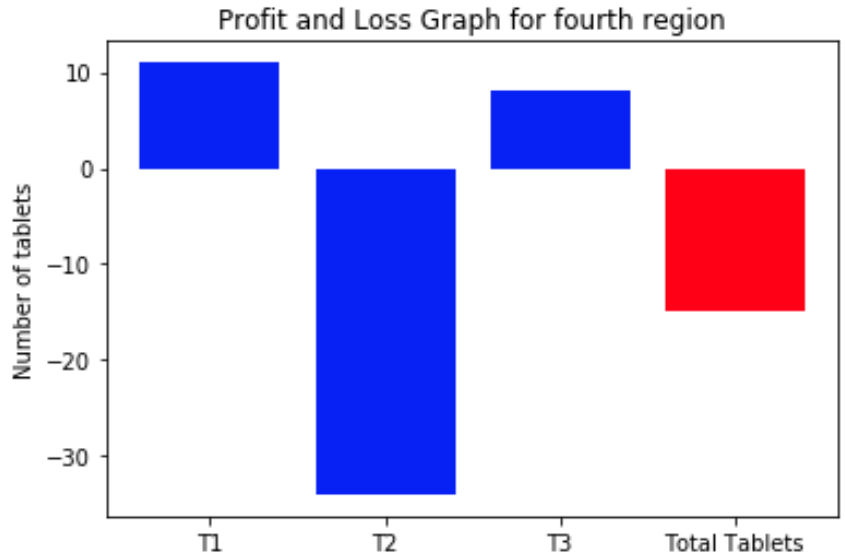



```
loss of 6 tablets
```

Fig.4.3.13. Profit/loss graph for region 3, third month.

4.3.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the third month.




```
loss of 15 tablets
```

Fig. 4.3.14. Profit/loss graph for region 4, third month.

4.3.15. Barchart for profit/loss in third month:

This shows the profit or loss happened in all regions in the third month.

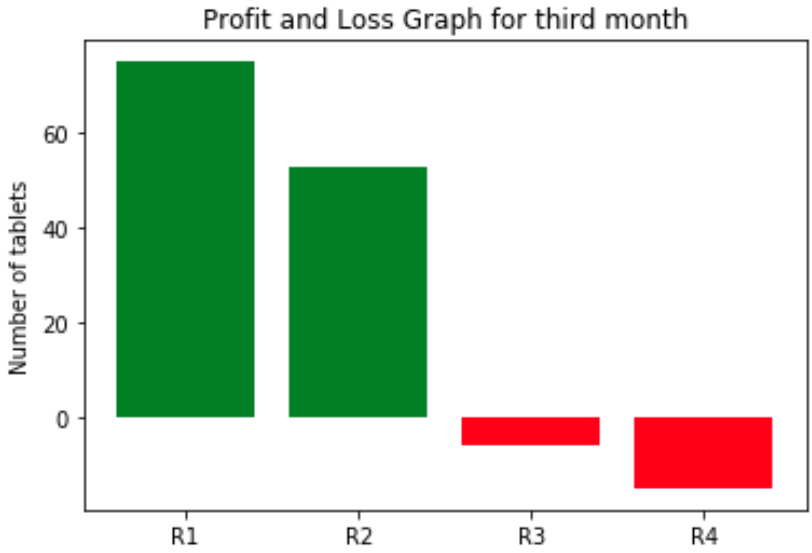


Fig.4.3.15. Profit/loss graph for all regions third month.

4.4 FOURTH MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of fourth month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in fourth month and then the combined graph for the fourth month and the profit and loss graph for the month.

4.4.1 Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the fourth month.

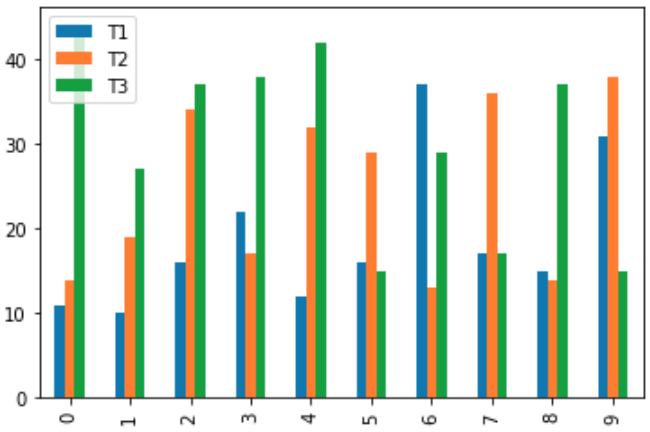


Fig.4.4.1. Barchart for region 1, fourth month

4.4.2 Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the fourth month.

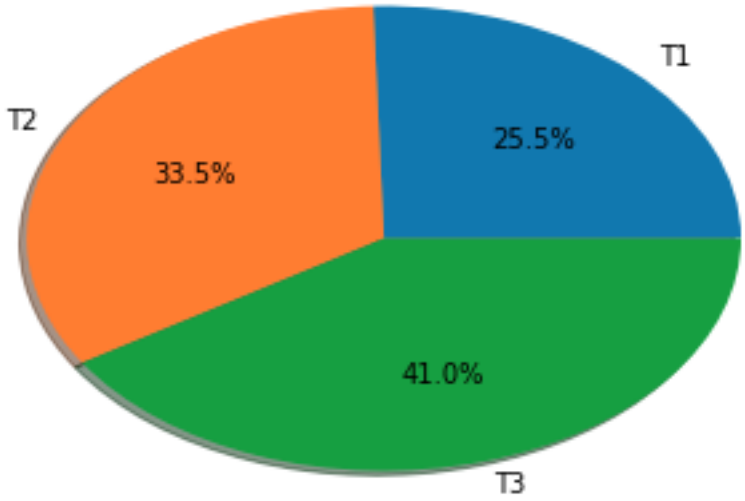


Fig.4.4.2. Piechart for region 1, fourth month.

4.4.3 Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the fourth month.

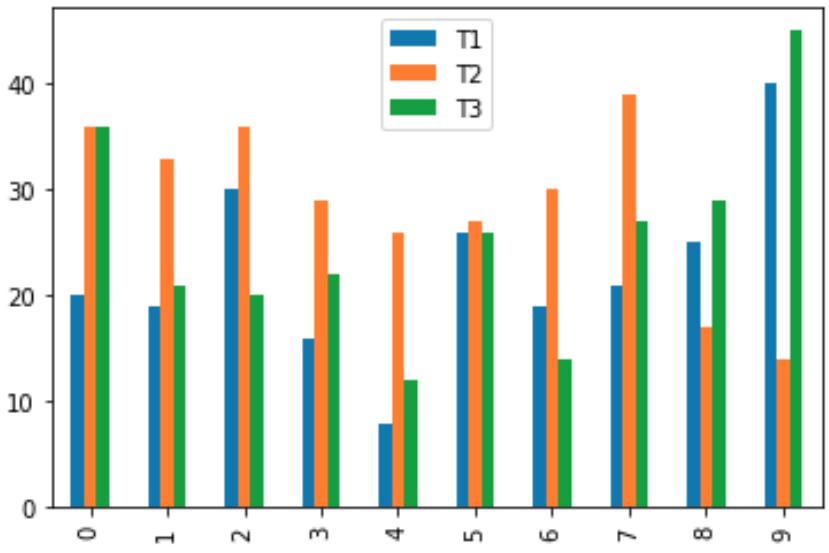


Fig.4.4.3. Barchart for Region 2, fourth month

4.4.4 Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the fourth month.

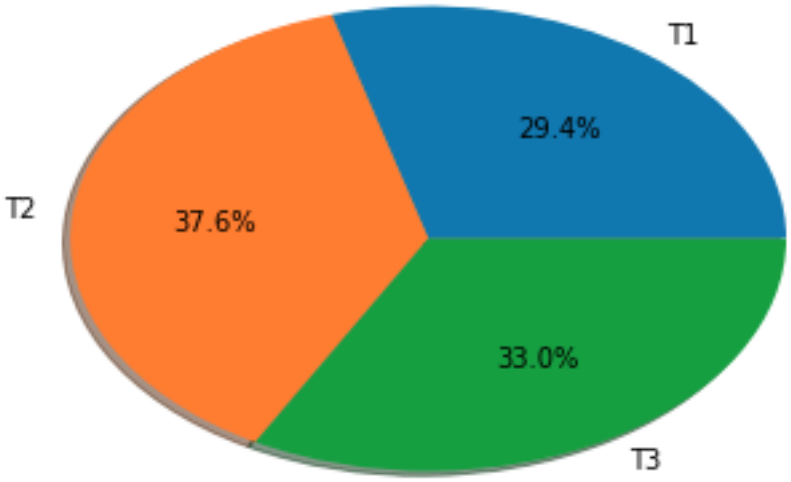


Fig.4.4.4. Piechart for Region 2, fourth month.

4.4.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the fourth month.

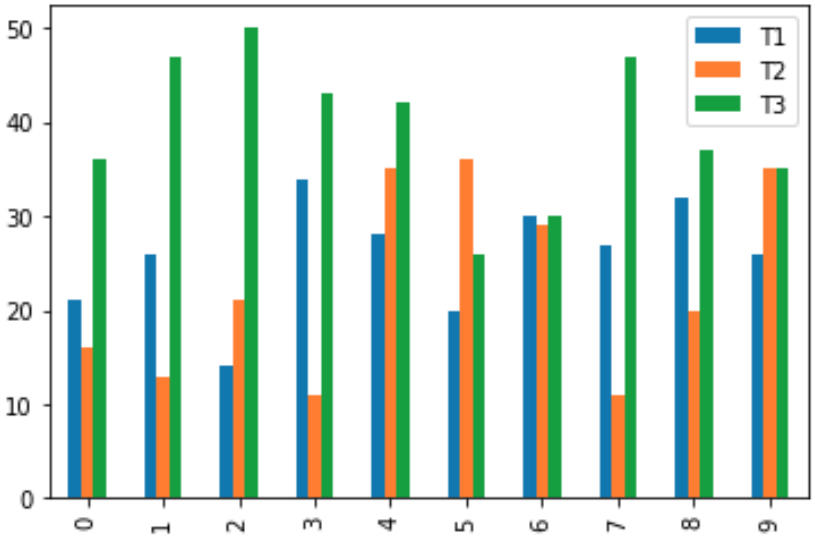


Fig.4.4.5. Barchart for region 3, fourth month.

4.4.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the fourth month.

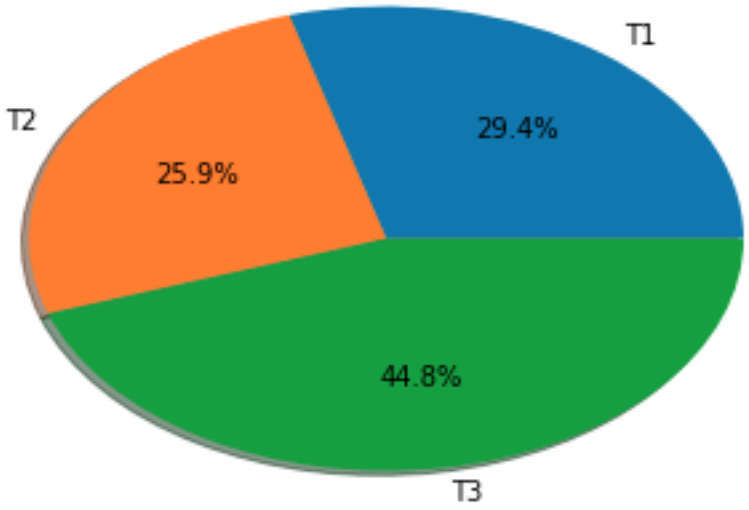


Fig.4.4.6. Piechart for region 3, fourth month.

4.4.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the fourth month.

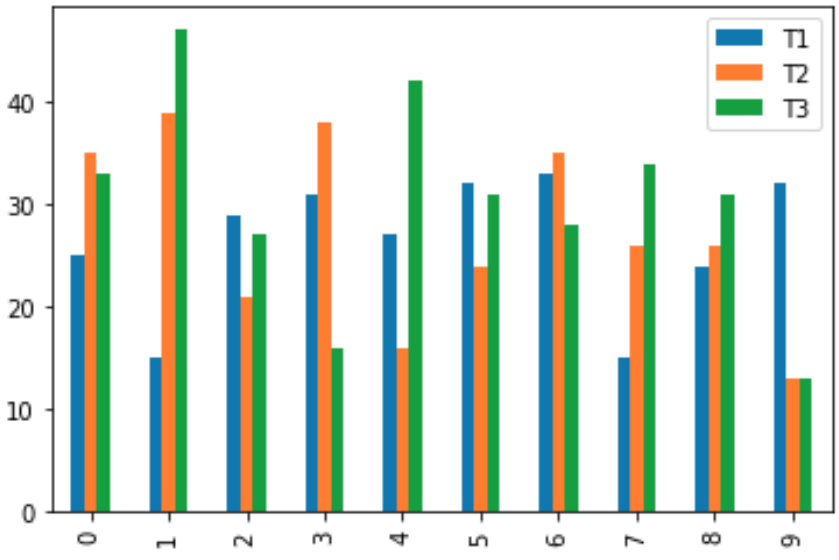


Fig.4.4.7. Barchart for region 4, fourth month

4.4.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the fourth month.

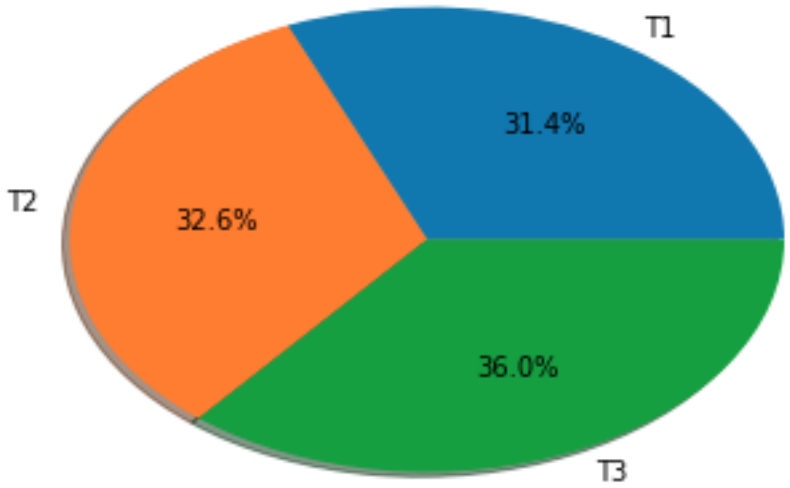


Fig.4.4.8. Piechart for region 4, fourth month.

4.4.9. Barchart for Fourth Month:

This shows the Barchart for the total number of tablets sold in all regions in the fourth month.

Fig.4.4.9 Barchart for all regions fourth Month

4.4.10. Piechart for Fourth Month

This shows the Piechart for the total percentage of tablets sold in all regions in the fourth month.

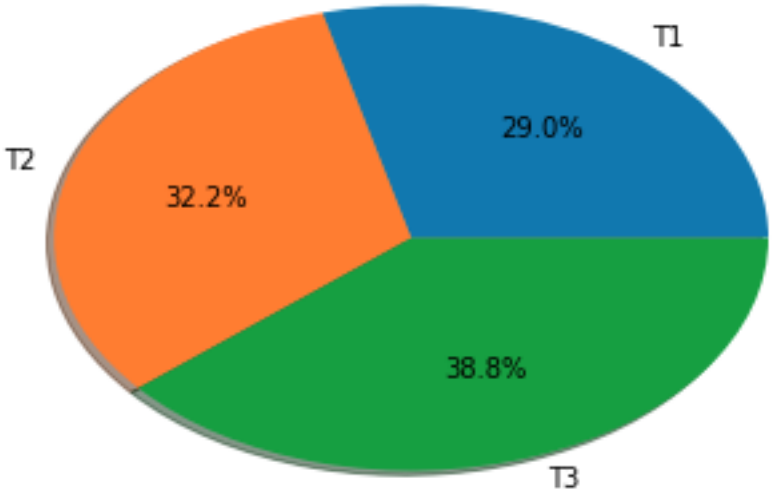
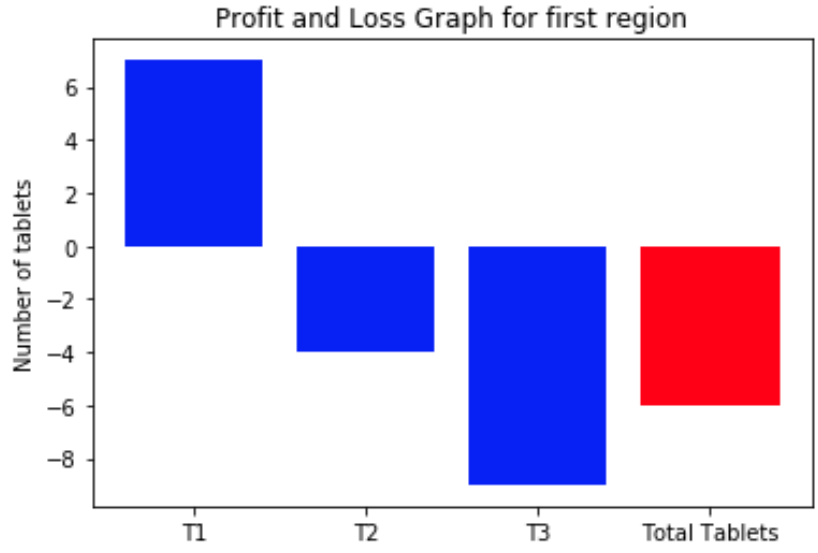


Fig.4.4.10. Piechart for all regions Fourth Month.

4.4.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the fourth month.

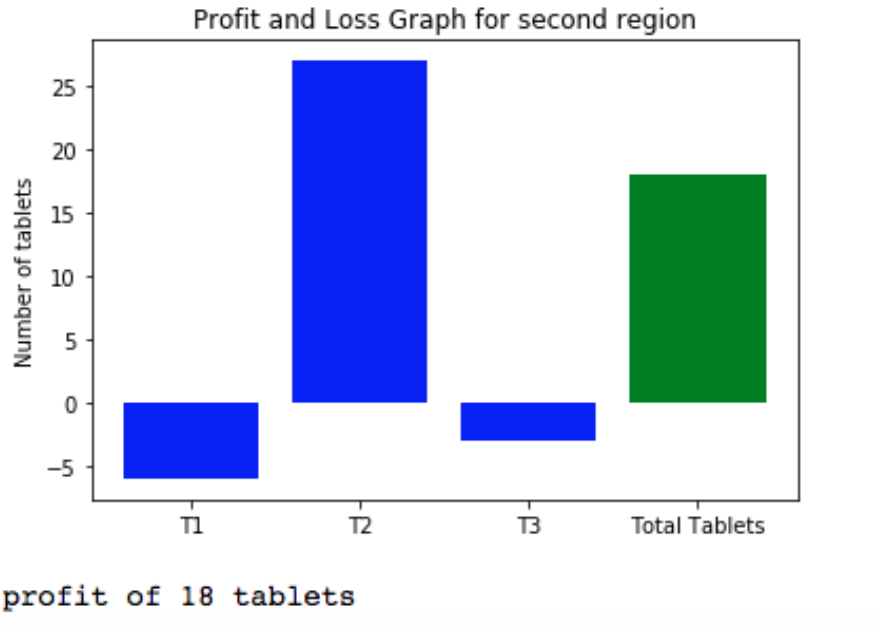



```
loss of 6 tablets
```

Fig.4.4.11. Profit/loss graph for region 1, fourth month.

4.4.12. Barchart for Profit/Loss in region 2:

This shows the profit or loss happened in region 2 in the fourth month.



Fig,4.4.12. Profit/loss graph for region 2, fourth Month.

4.4.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the fourth month.

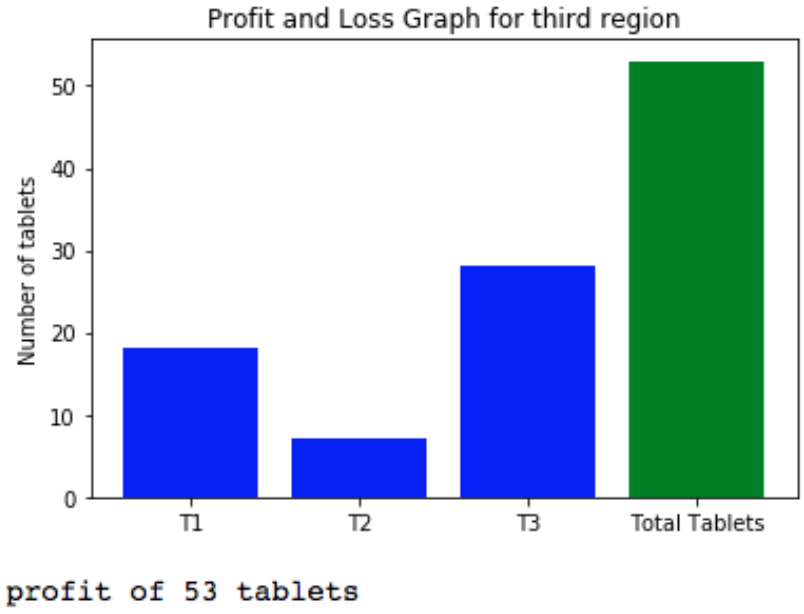
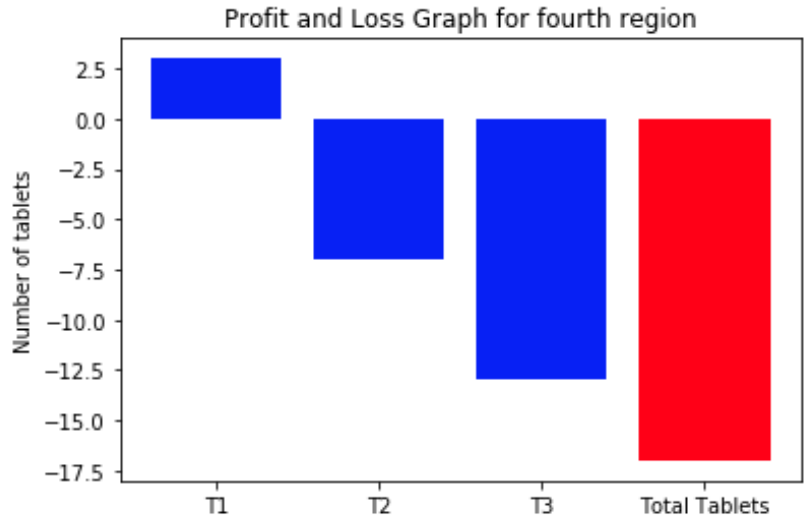


Fig.4.4.13. Profit/loss graph for region 3, fourth month.

4.4.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the fourth month.




```
loss of 17 tablets
```

Fig. 4.4.14. Profit/loss graph for region 4, fourth month.

4.4.15. Barchart for profit/loss in fourth month:

This shows the profit or loss happened in all regions in the fourth month.

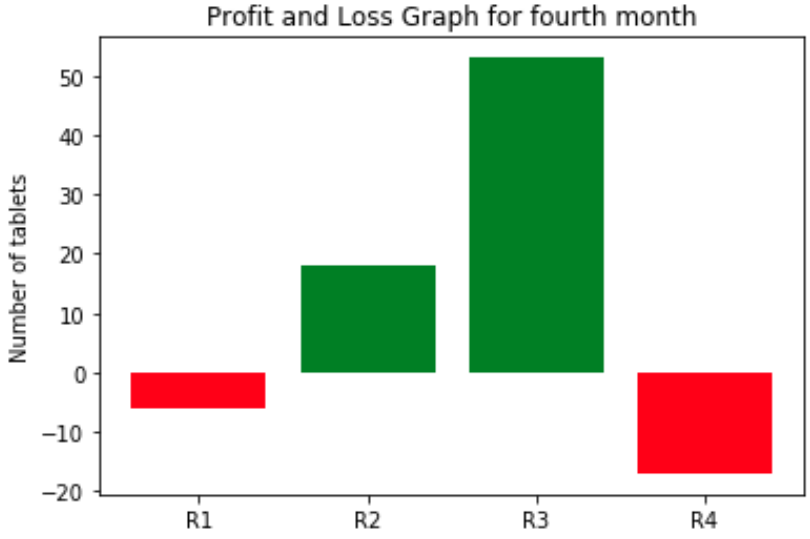


Fig.4.4.15. Profit/loss graph for all regions, fourth month.

4.5 FIFTH MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of fifth month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in fifth month and then the combined graph for the fifth month and the profit and loss graph for the month.

4.5.1 Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the fifth month.

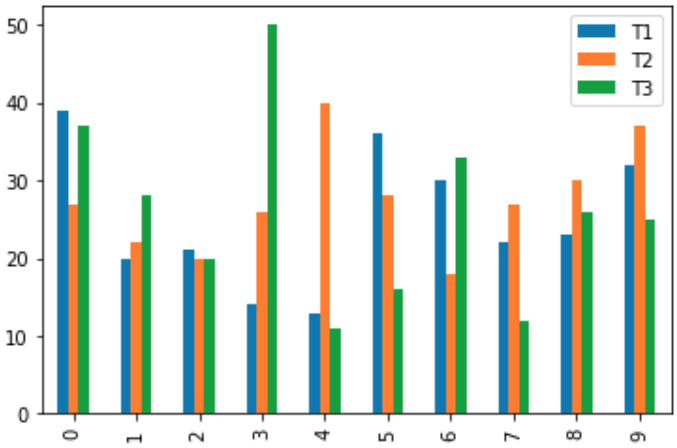


Fig 4.5.1. Barchart for region 1, fifth month.

4.5.2 Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the fifth month.

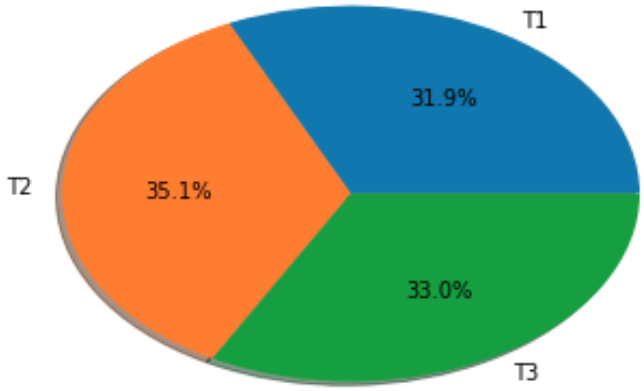


Fig.4.5.2. Piechart for region 1, fifth month.

4.5.3 Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the fifth month.

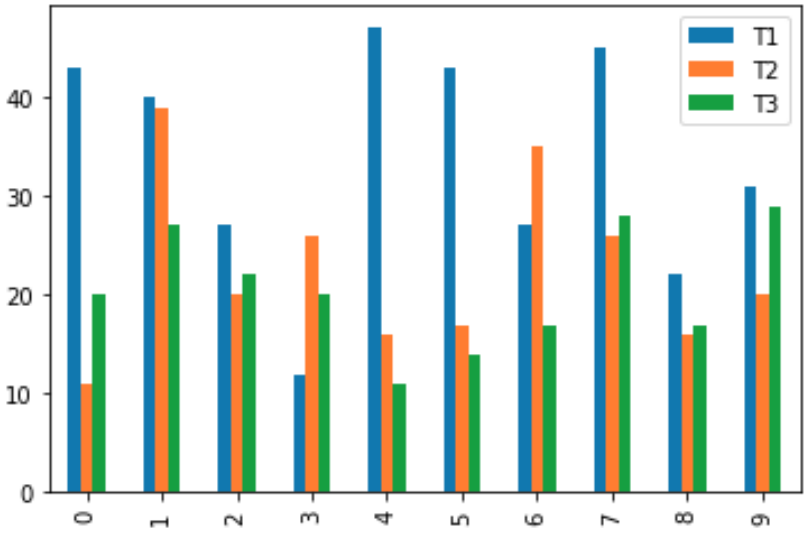


Fig.4.5.3. Barchart for Region 2, fifth month

4.5.4.Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the fifth month.

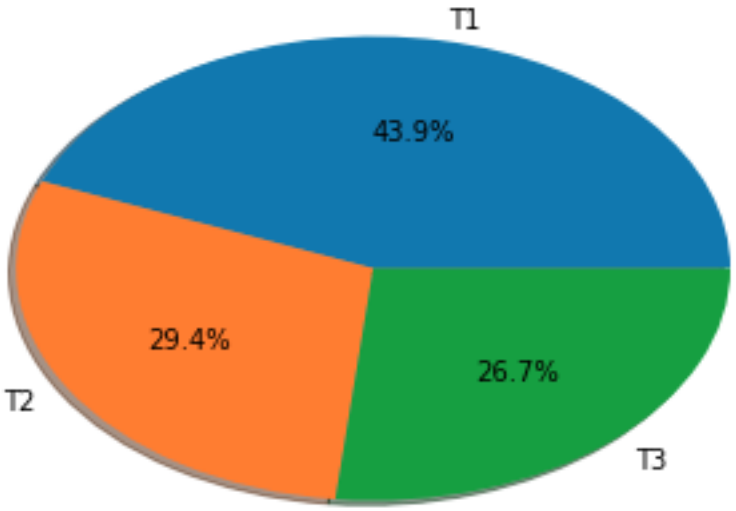


Fig.4.5.4. Piechart for Region 2, fifth month.

4.5.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the fifth month.

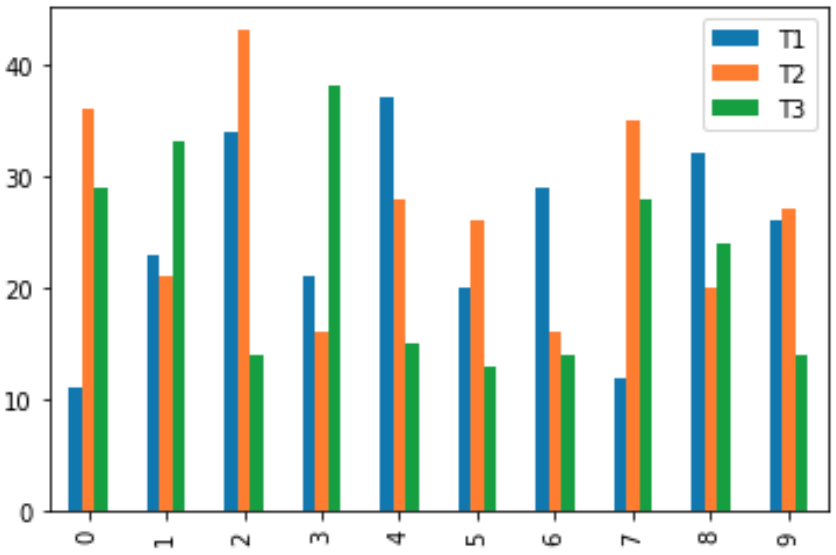


Fig.4.5.5. Barchart for region 3, fifth month.

4.5.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the fifth month.

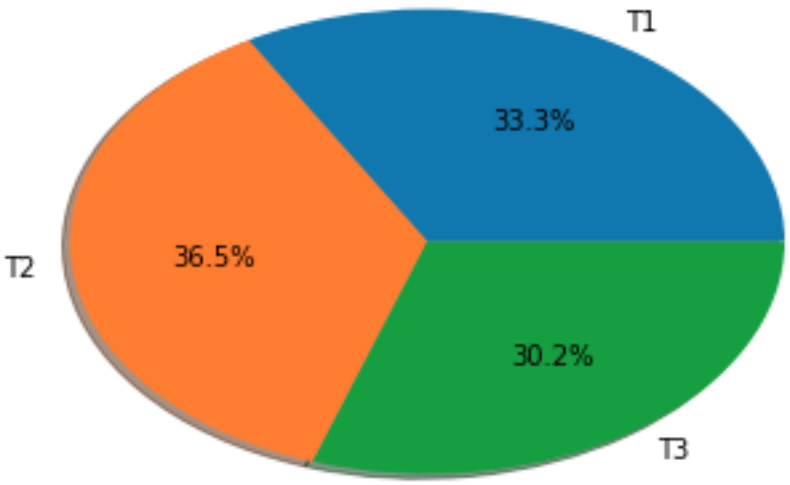


Fig.4.5.6. Piechart for region 3, fifth month.

4.5.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the fifth month.

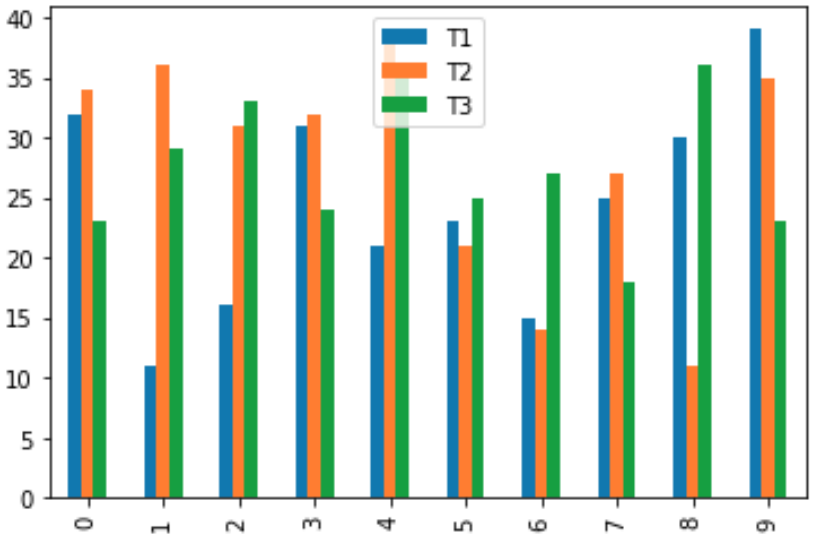


Fig.4.5.7. Barchart for region 4, fifth month

4.5.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the fifth month.

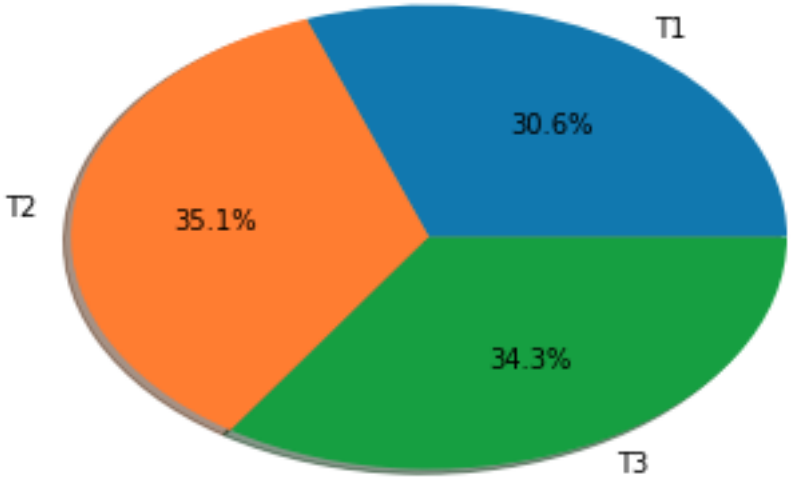


Fig.4.5.8. Piechart for region 4, fifth month.

4.5.9. Barchart for Fifth Month:

This shows the Barchart for the total number of tablets sold in all regions in the Fifth month.

Fig.4.5.9 Barchart for all regions Fifth Month

4.5.10. Piechart for Fifth Month

This shows the Piechart for the total percentage of tablets sold in all regions in the Fifth month.

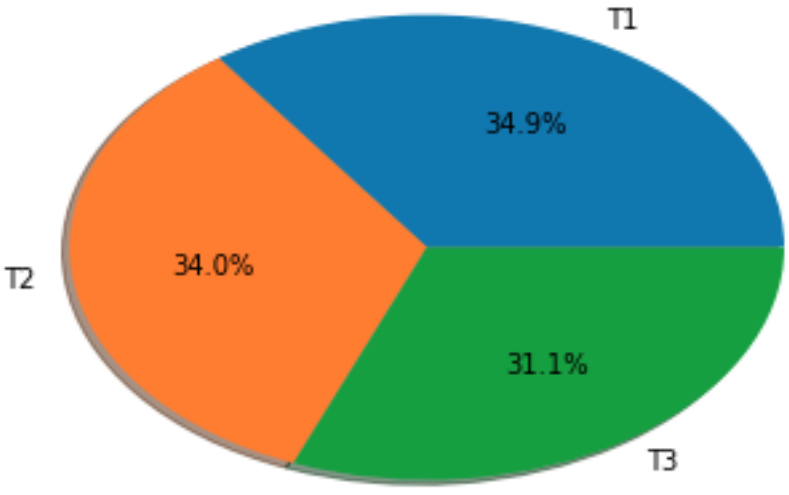
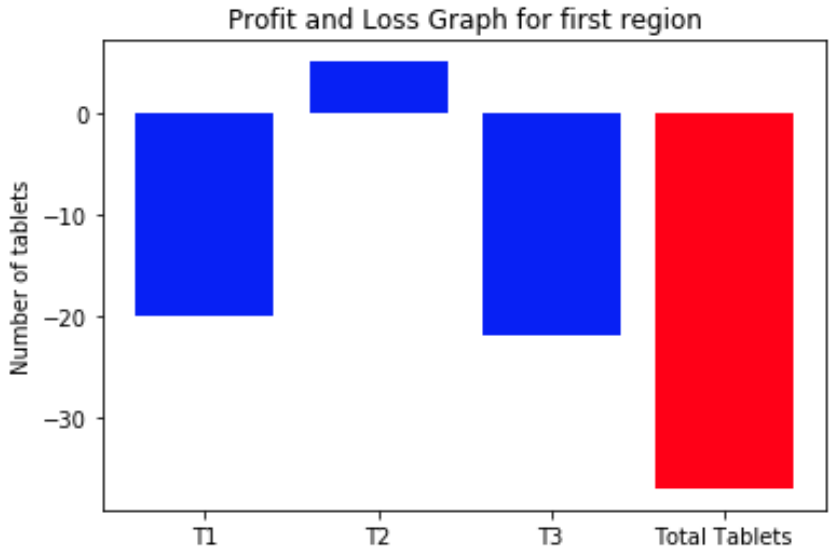


Fig.4.5.10. Piechart for all regions Fifth Month.

4.5.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the fifth month.

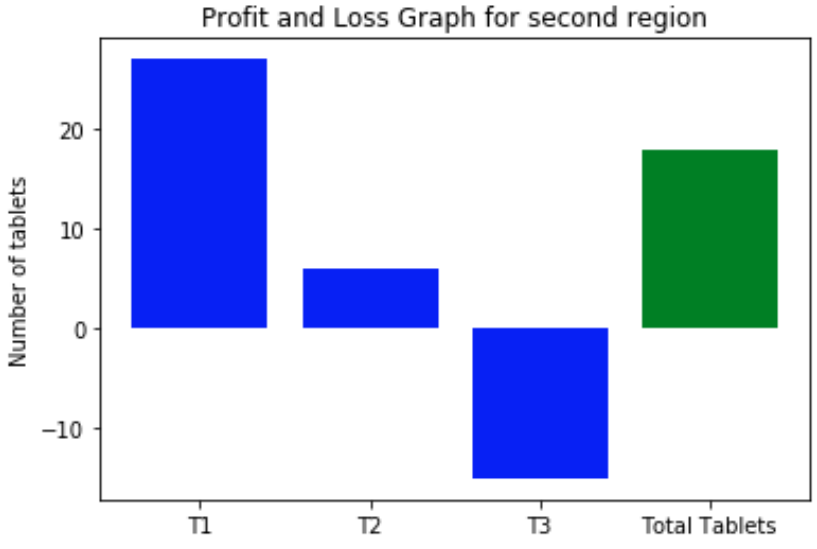



```
loss of 37 tablets
```

Fig.4.5.11. Profit/loss graph for region 1 fifth month.

4.5.12. Barchart for Profit/Loss in region 2:

This shows the profit or loss happened in region 2 in the fifth month.




```
profit of 18 tablets
```

Fig,4.5.12. Profit/loss graph for region 2 fifth Month.

4.5.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the fifth month.

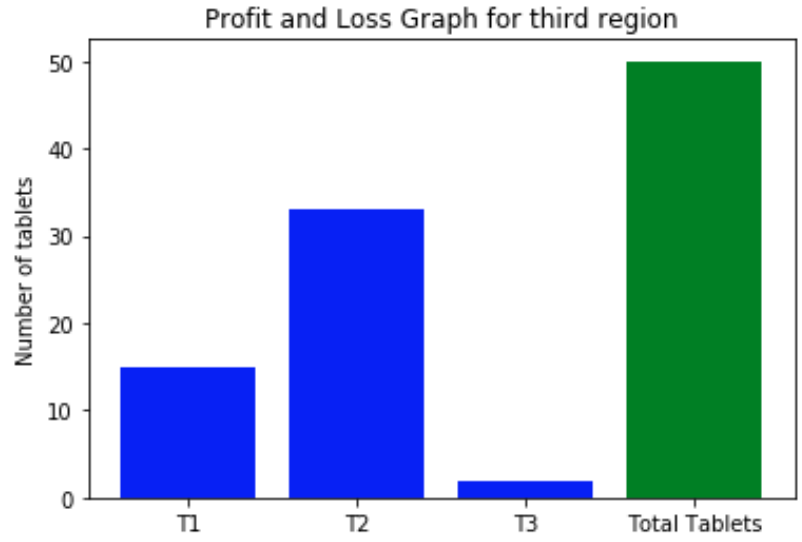


Fig.4.5.13. Profit/loss graph for region 3, fifth month.

4.5.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the fifth month.

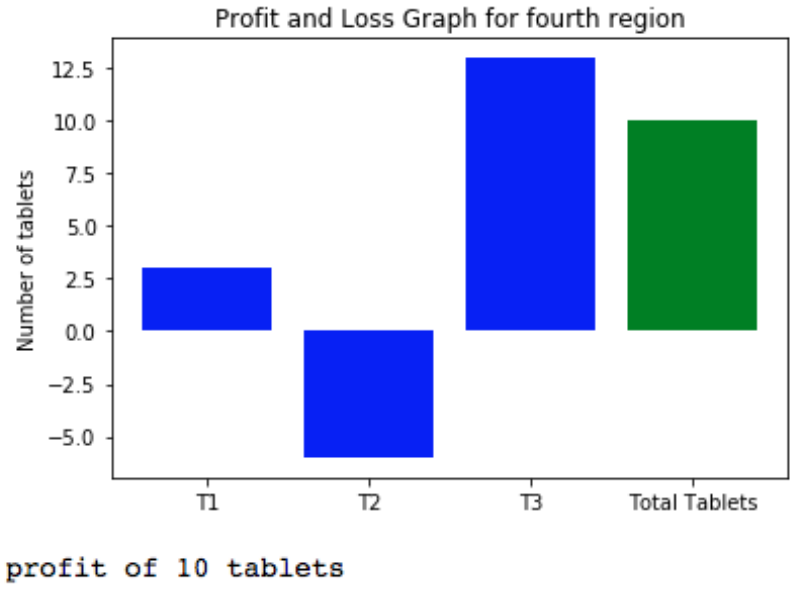


Fig. 4.5.14. Profit/loss graph for region 4, fifth month.

4.5.15. Barchart for profit/loss in fifth month:

This shows the profit or loss happened in all regions in the fifth month.

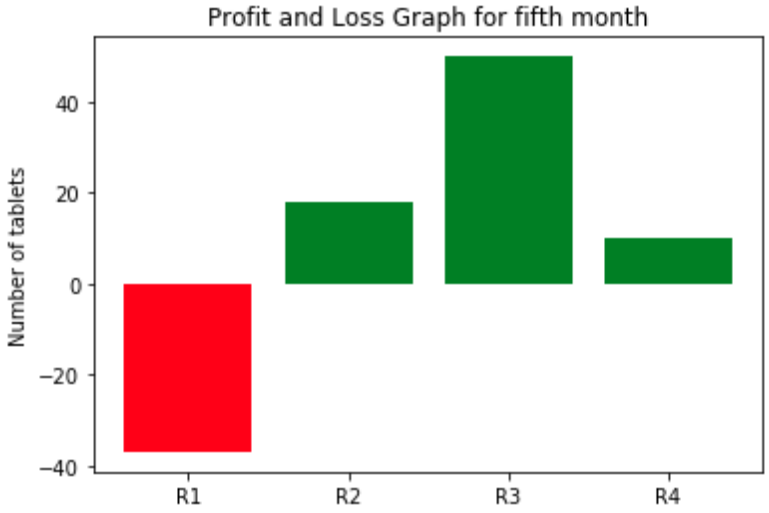


Fig.4.5.15. Profit/loss graph for all regions fifth month.

4.6 SIXTH MONTH DATA ANALYTICS OUTPUT:

This shows the graphs generated after doing the analytics on the dataset of sixth month. Here, we show the graphs of total number and percentage of tablets sold for all four regions in sixth month and then the combined graph for the sixth month and the profit and loss graph for the month.

4.6.1 Barchart for region 1:

This shows the Barchart for the total number of tablets sold in region 1 in the sixth month.

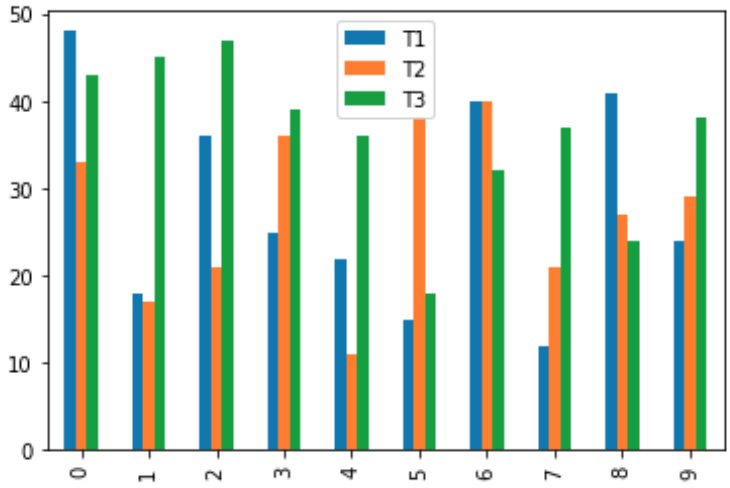


Fig.4.6.1. Barchart for region 1, sixth month

4.6.2 Piechart for region 1:

This shows the Piechart for the total percentage of tablets sold in region 1 in the sixth month.

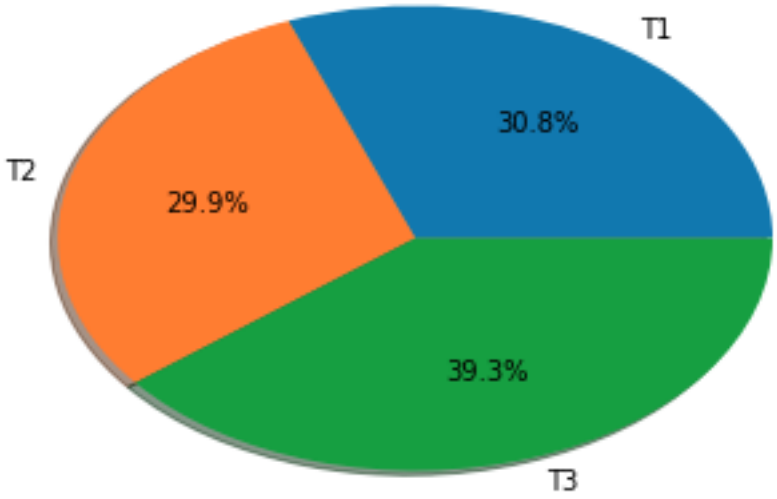


Fig.4.6.2. Piechart for region 1, sixth month.

4.6.3 Barchart for region 2:

This shows the Barchart for the total number of tablets sold in region 2 in the sixth month.

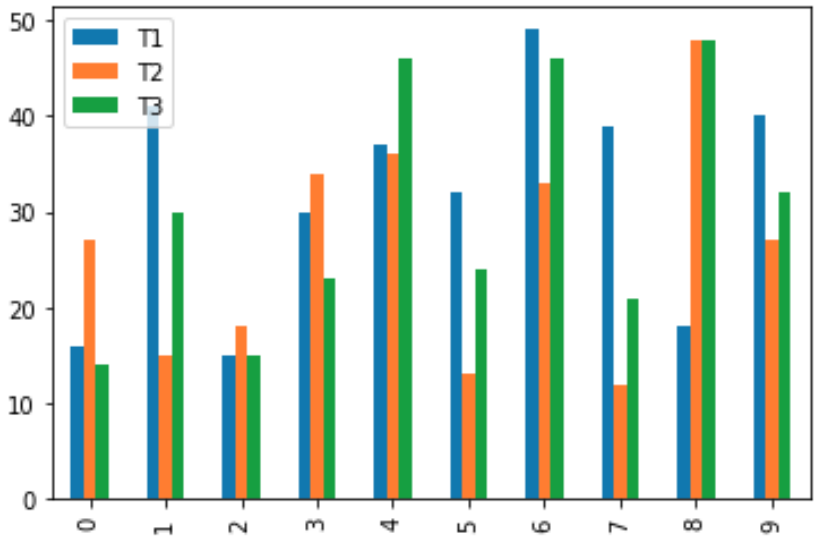


Fig.4.6.3. Barchart for Region 2, sixth month

4.6.4 Piechart for region 2:

This shows the Piechart for the total percentage of tablets sold in region 2 in the sixth month.

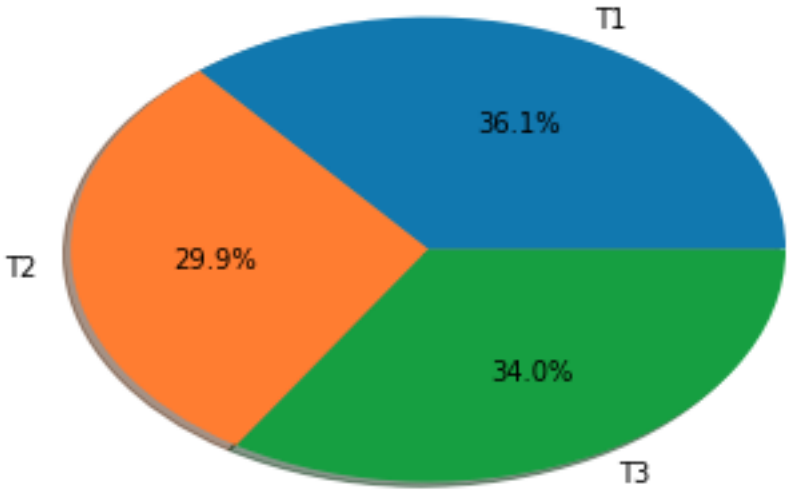


Fig.4.6.4. Piechart for Region 2, sixth month.

4.6.5. Barchart for region 3:

This shows the Barchart for the total number of tablets sold in region 3 in the sixth month.

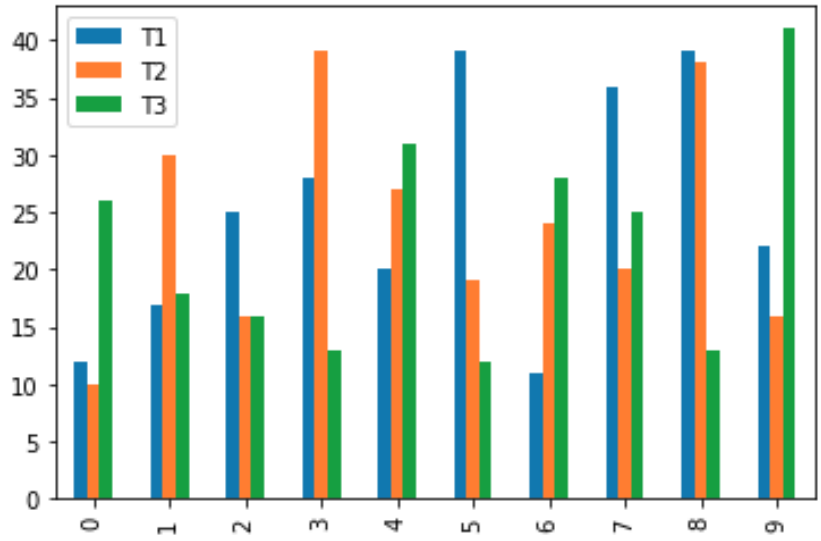


Fig.4.6.5. Barchart for region 3, sixth month.

4.6.6. Piechart for region 3:

This shows the Piechart for the total percentage of tablets sold in region 3 in the sixth month.

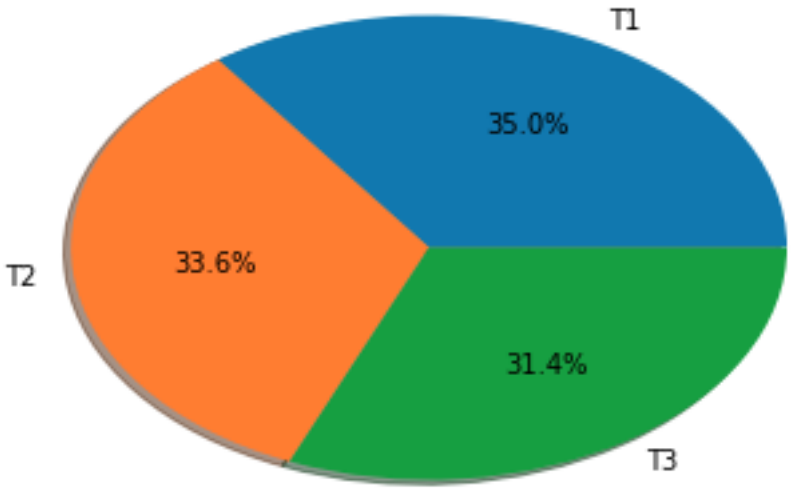


Fig.4.6.6. Piechart for region 3, sixth month.

4.6.7. Barchart for Region 4:

This shows the Barchart for the total number of tablets sold in region 4 in the sixth month.

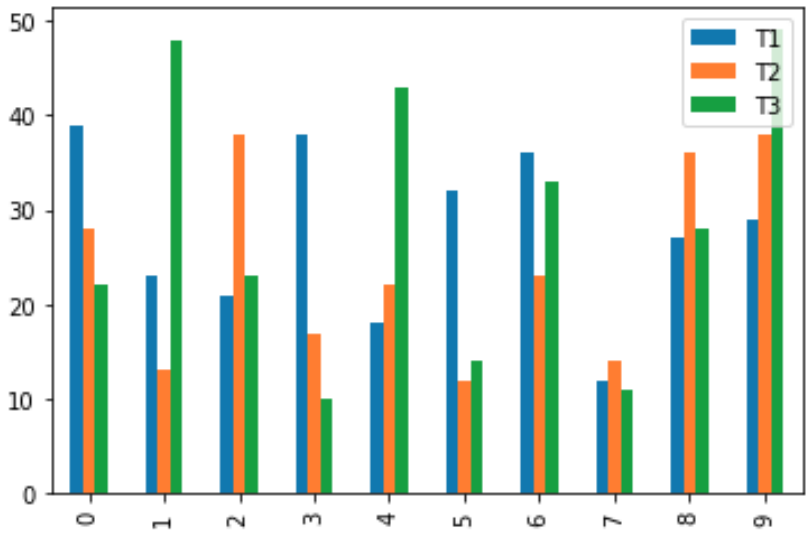


Fig.4.6.7. Barchart for region 4, sixth month

4.6.8. Piechart for Region 4:

This shows the Piechart for the total percentage of tablets sold in region 4 in the sixth month.

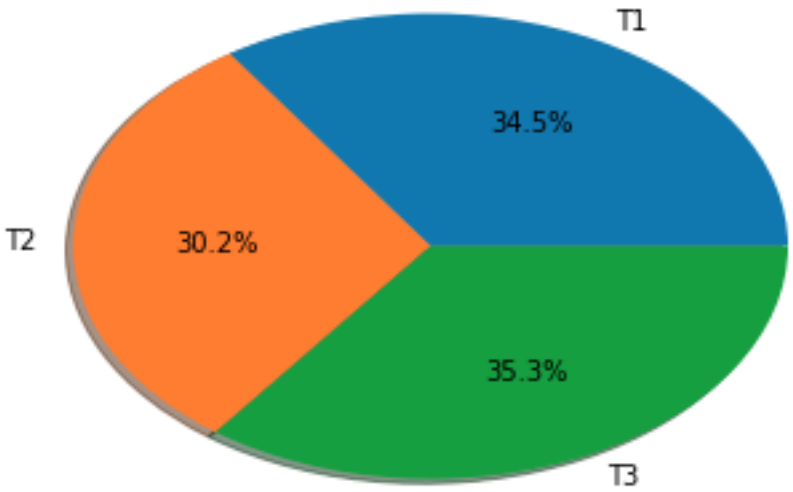


Fig.4.6.8. Piechart for region 4, sixth month.

4.6.9. Barchart for Sixth Month:

This shows the Barchart for the total number of tablets sold in all regions in the sixth month.

Fig.4.6.9 Barchart for all regions sixth Month

4.6.10. Piechart for Sixth Month

This shows the Piechart for the total percentage of tablets sold in all regions in the sixth month.

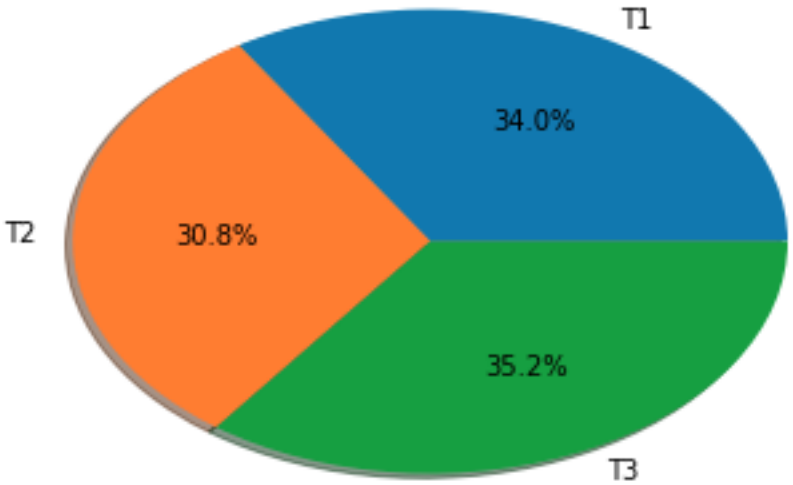
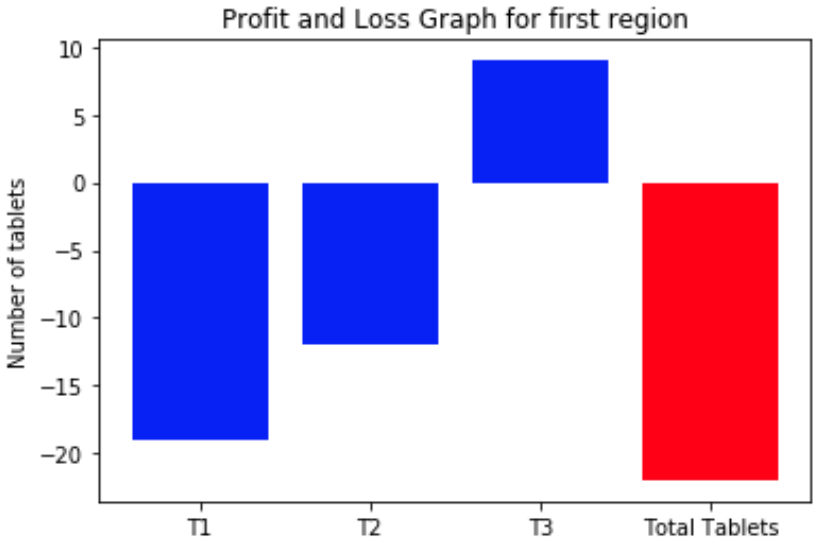


Fig.4.6.10. Piechart for all regions Sixth Month.

4.6.11. Barchart for Profit/Loss in region 1:

This shows the profit or loss happened in region 1 in the sixth month.

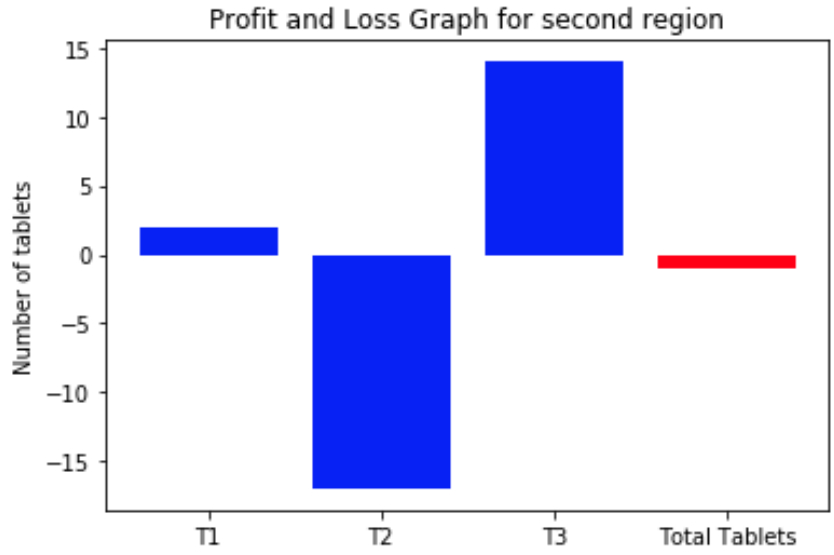



```
loss of 22 tablets
```

Fig.4.6.11. Profit/loss graph for region 1, sixth month.

4.6.12. Barchart for Profit/Loss in region 2:

This shows the profit or loss happened in region 2 in the sixth month.

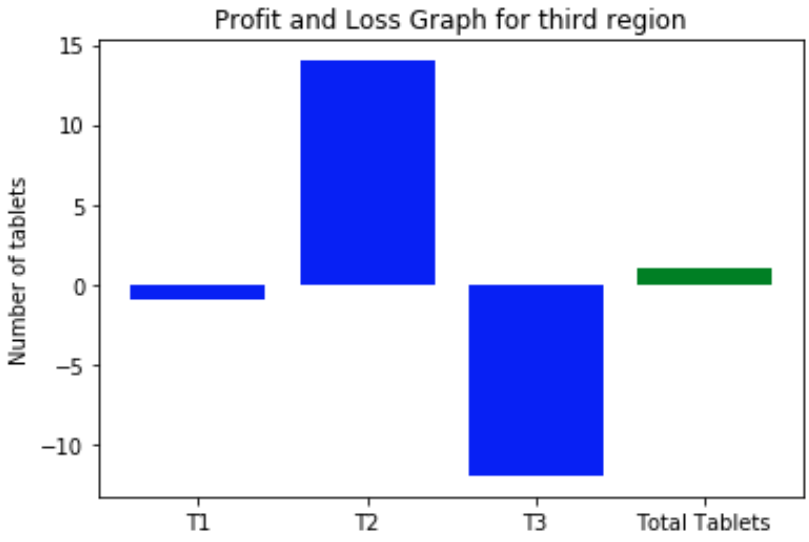



```
loss of 1 tablets
```

Fig,4.6.12. Profit/loss graph for region 2, sixth Month.

4.6.13. Barchart for Profit/Loss in region 3:

This shows the profit or loss happened in region 3 in the sixth month.




```
profit of 1 tablets
```

Fig.4.6.13. Profit/loss graph for region 3, sixth month.

4.6.14. Barchart for Profit/Loss in region 4:

This shows the profit or loss happened in region 4 in the sixth month.

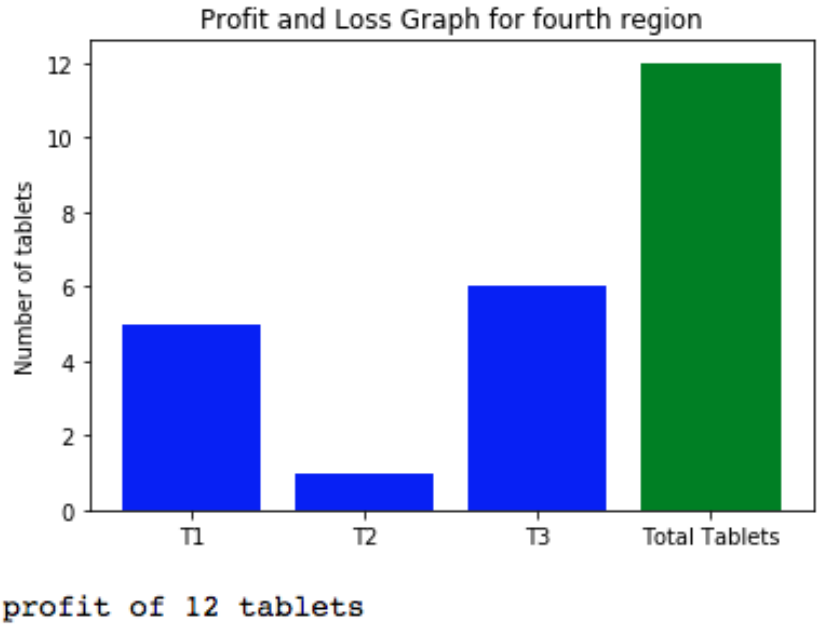


Fig. 4.6.14. Profit/loss graph for region 4, sixth month.

4.6.15. Barchart for profit/loss in sixth month:

This shows the profit or loss happened in all regions in the sixth month.

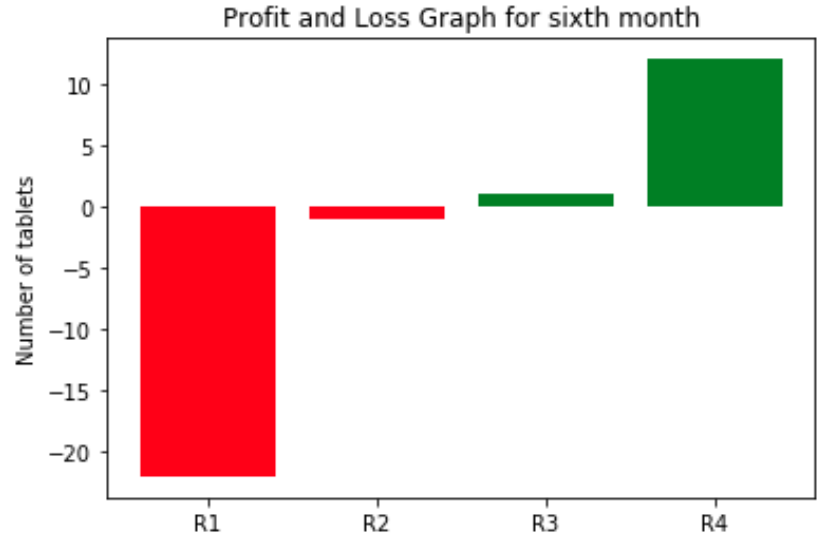


Fig.4.6.15. Profit/loss graph for all regions, sixth month

4.7 PREDICTED OUTPUT USING LINEAR REGRESSION:

This show the prediction of the total number of tablets generated by training the data for all six months by using Linear Regression predictive model.

Here the prediction is done for the seventh month using the linear regression model.

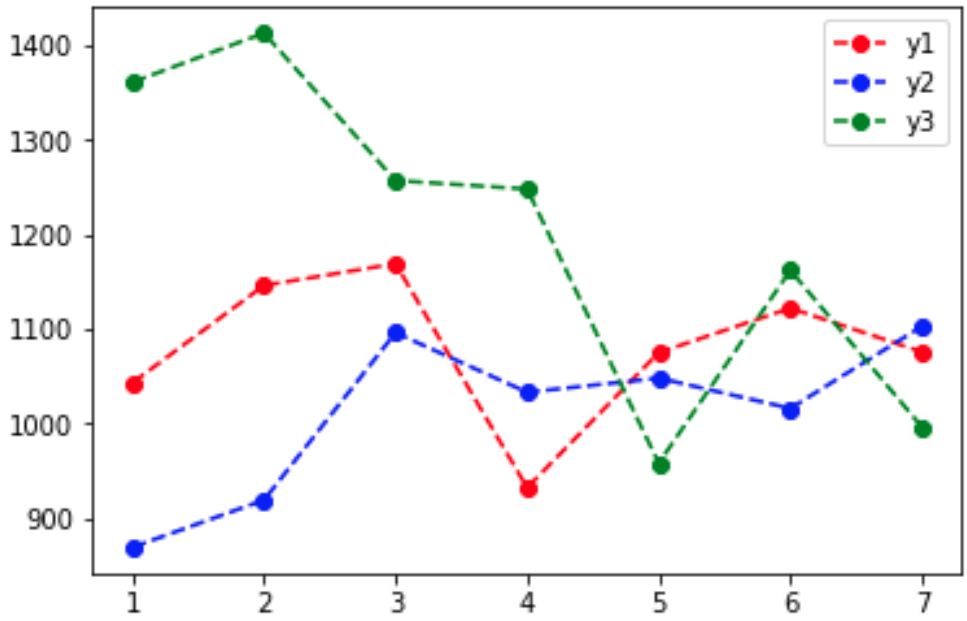


Fig 4.7. Predicted output using linear regression model.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

The project titled - “Pharmaceutical Sales Analytics” was overtaken as the Final Year Project during the course of the Bachelor’s of Technology (B-Tech) degree in Alliance College of Engineering and Design, Alliance University. The area of Study in the project was “Supervised Machine Learning Algorithm”.

This project is developed after thoroughly studying about Data Analytics and its methods, Supervised Machine Learning algorithms and Python programming language.

The main purpose of the project is to predict the future output from the given dataset of medical stores of the tablets sold using the supervised machine learning algorithms. Here, we have used the Linear Regression model to predict the outcome. In this project, first the data is collected and read and then data analytics are performed on the dataset for each region and for month as a whole. After this, the profit and loss function is written for each region and the profit or loss in every region is found. Graphs are also generated for the data analytics and profit and loss function. Finally, the data of all the six months are read together and sum of every tablets monthwise is found. Then, the linear regression model is developed and the model predicts the outcome for the seventh month about what the sales will be there and so on for further months. The graph is also generated for the linear regression model.

The project is made on the Anaconda Navigator Platform. The programming language used here is Python and the Interpreter used is Jupyter Notebook. The Python is an easy and most sought after language for data science.

We can further enhance it in future by,

1. Training the data based on more available datasets.
2. Finding predictions for more tablets to understand the sales trends and to get the knowledge of the disease spreading in the particular region.
3. To make the regression model such that it show trends in the data for continuous 6 months to 1 year.

REFERENCES

[1] https://hal.archives-ouvertes.fr/hal-01881362/document

[2] https://python-graph-gallery.com/

[3] https://towardsdatascience.com/data-cleaning-with-python-and-pandas-detecting-missing-values-3e9c6ebcf78b

[4] https://realpython.com/python-data-cleaning-numpy-pandas/

[5] https://python-graph-gallery.com/3-control-color-of-barplots/

[6] https://www.databriefing.com/coding/predict-sales-1/

[7] https://medium.com/we-are-orb/multivariate-linear-regression-in-python-without-scikit-learn-7091b1d45905

[8] https://stackabuse.com/linear-regression-in-python-with-scikit-learn/

[9] https://towardsdatascience.com/a-beginners-guide-to-linear-regression-in-python-with-scikit-learn-83a8f7ae2b4f

[10] https://www.johnwittenauer.net/machine-learning-exercises-in-python-part-2/

[11] https://realpython.com/linear-regression-in-python/

[12] https://www.thegeekstuff.com/2013/06/python-list/embed/

[13] https://www.johnwittenauer.net/machine-learning-exercises-in-python-part-1/

[14] https://www.youtube.com/

[15] https://www.wikipedia.org/

[16] https://www.edureka.co/blog/what-is-data-science/

[17] https://www.techopedia.com/definition/26418/data-analytics

[18] https://www.geeksforgeeks.org/machine-learning/

[19] https://www.geeksforgeeks.org/supervised-unsupervised-learning/

[20] https://res.mdpi.com/data/data-04-00015/article_deploy/data-04-00015-v2.pdf?filename=&attachment=1

PROFORMA FOR APPROVAL OF PROJECT REPORT

Name of the students	: Anshuman Sinha
Name of the Department	: Computer Science and Engineering
Branch	: Computer Science and Engineering (CSE)
Year	: 2018-2019
Date of Submission	: 29th May 2019
Batch Number	: 16

Items	Yes	No
Whether all pages are as per instruction?	✓	
Whether the contents of report are in correct order as given in the manual?	✓	
Whether the margin specification and spacing for the text correctly followed as per instruction in the manual?	✓	
Whether typed in black?	✓	
Whether numbers given for figures,tables & other pages serially and in correct position?	✓	
Whether tables numbered properly with captions given as top of tables?	✓	
Whether figures & diagram numbered properly with caption given below?	✓	
Whether binding is OK?	✓	
Whether references are marked in text within square brackets?	✓	
Whether CD containing executable s/w, required libraries & source code?	✓	
Comment: Good 10/10		**Rating:**

Srinatha
Internal Guide

Head of the Department