ANALYTICAL TOOLS FOR E-COMMERCE BUSINESS

A UG PROJECT PHASE-1 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

Inpartial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

REPALA PRAVALIKA	**19UK1A0531**
MAMIDALA KRISHINA SRI	**19UK1A0529**
SYED ABDUL MUJEEB	**19UK1A0569**
CHIDIRALA SAI KIRAN	**19UK1A0566**

Under the esteemed guidance of

Ms.A.Swathi

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTUH,Hyderabad)

Bollikunta,Warangal–506005

2019-2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

BOLLIKUNTA,WARANGAL–506005

2019 –2023

CERTIFICATE OF COMPLETION

UG PROJECT PHASE

This is to certify that the UG Project Phase-1 entitled **"ANALYTICAL TOOLS FOR E-COMMERCE BUSINESS"** is being submitted by **REPALA PRAVALIKA (H.NO:19UK1A0531) , MAMIDALA KRISHINA SRI (H.NO :19UK1A0529), SYED ABDUL MUJEEB (H.NO:19UK1A0569) ,CHIDIRALA SAI KIRAN (H.NO:19UK1A0566)** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University Hyderabad** during the academic year**2022-23**,is a record of work carried out by them under the guidance and supervision.

ProjectGuide
Ms.A.Swathi.
(AssistantProfessor)

Headof theDepartment
Dr.R.NaveenKumar
(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR,** Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **Ms.A.Swathi** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

REPALA PRAVALIKA (19UK1A0531)
MAMIDALA KRISHNA SRI (19UK1A0529)
SYED ABDUL MUJEEB (19UK1A0569)
CHIDIRALA SAI KIRAN (19UK1A0566)

ABSTRACT

The drivers for electronic commerce are both technological (under the tremendous pressure of innovation) and business oriented. This paper will highlight some guidelines for companies who are entering into E- commerce to create an E-commerce strategy or who already have an E-commerce presence to revise their existing strategy. E-Commerce is now seen as a reality for many businesses and a normal part of a business plan. The immediate benefits, in terms of cost savings, efficiencies and enhanced profitability are clear at every stage in the supply chain. Adopting e-business is no longer a competitive advantage, but a normal business process, without which an enterprise is unlikely to survive in the New economy . Year 2000 saw many Dot-com companies built upand many companies going into E-commerce however now it is a different story, more and more companies are failing, and investors are becoming cautious to invest money into Internet ventures. There is more cash needed then was expected.

TABLE OF CONTENTS:-

1.INTRODUCTION

“Electronic commerce, commonly written as E-Commerce, is the trading in products or services using computer networks, such as the Internet. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web for at least one part of the transaction’s life cycle, although it may also use other technologies such as E-Mail.

E-Commerce businesses may employ some or all of the following:

- Online shopping websites for retail sales direct to consumers,
- Providing or participating in online marketplaces, which process third-party businessto-consumer or consumer-to-consumer sales,
- Business-to-business buying and selling,
- Gathering and using demographic data through Web contacts and social media,
- Business-to-business electronic data interchange.

The role of analytics in e-commerce is If you have worked in financial industry, you will probably be aware of analytics playing a crucial role into risk and marketing strategy.However, E-Commerce industry goes beyond these two pillars. The primary job of E- Commerce industry isto make user experience on their website is delightful. Other than that they are simply a platform between sellers and buyers. With such focus on user experience, analytics itself becomes a product instead of just being business enabler. For instance, Recommender Enginesyou see on Amazon sidebar is a classic product. Now, you can appreciate the much broader role of analytics in E-Commerce industry. In the following section, we will talk more about broad functions where analytics is being actively used.

1.1 INTRODUCTION TO E-BUSINESS

"Electronic business, or E-Business, is the application of information and communication technologies (ICT) in support of all the activities of business. Commerce constitutes the exchange of products and services between businesses, groups and individuals and can be seen as one of the essential activities of any business. Electronic commerce focuses on the use of ICT to enable the external activities and relationships of the business with individuals, groups and other businesses or E-Business refers to business with help of Internet i.e. doing business with the help of Internet network. The term was coined by IBM's marketing and Internet teamin 1996."

E-Business is a more general term than E-Commerce. However, in this book we will only use the term "E-Commerce", because every business transaction finally is involved in selling or buying of products or services. And the term "E-Commerce" obviously is more widespread than the term "E-Business".

10 Key Benefits Of Ecommerce For Your Business

- Low costs. A significant advantage of ecommerce is that launching an online store ismuch less costly than opening a physical store. ...
- Speed & Flexibility. ...
- Faster Buying Process ...
- Product Catalogue. ...
- Wider Customer Base. ...
- Customer Data Insights. ...
- Scalability. ...
- Reviews & Ratings

1.2 OVERVIEW

Ecommerce analytics is the process of gathering data from all areas that have an impact on online stores and using this informaction to understand the trends and the shift in consumer behavior to make datadrive decisions that will drive more online sales.

1.3 PURPOSE

Ecommerce analytics is the process of discovering ,interpreting,and communicating data pattern related to online business.ecommerce analytics helps measure user behavior, performance trends. Ecommerce analytics helps centralize and mange data.

The focus of analytics is on issues that matter most to the business,and the performance metrics are helpful in identifying and solving problems in real time.

Because technology evolves rapidly and shopping trends shift on a daily basis, ecommerce is in a constant state of adaptation that can leave brands fumbling in the dark. To keep up, ecommerce businesses must anticipate changes in the market using reliable data insights. In short, they need effective ecommerce analytics.Ecommerce analytics simply refers to any tool or strategy designed to analyse large amounts of data in order to produce actionable insights. Because it exists in an almost entirely virtual space, ecommerce generates complex, comprehensive datasets — particularly those related to client behaviour. More data was created in 2017 than was created during the previous 5,000 years combined. That is a lot of data to measure, parse, and analyse. Finding the right ecommerce analytics tools for the job, however, can offer your brand an immeasurable advantage over the competition.

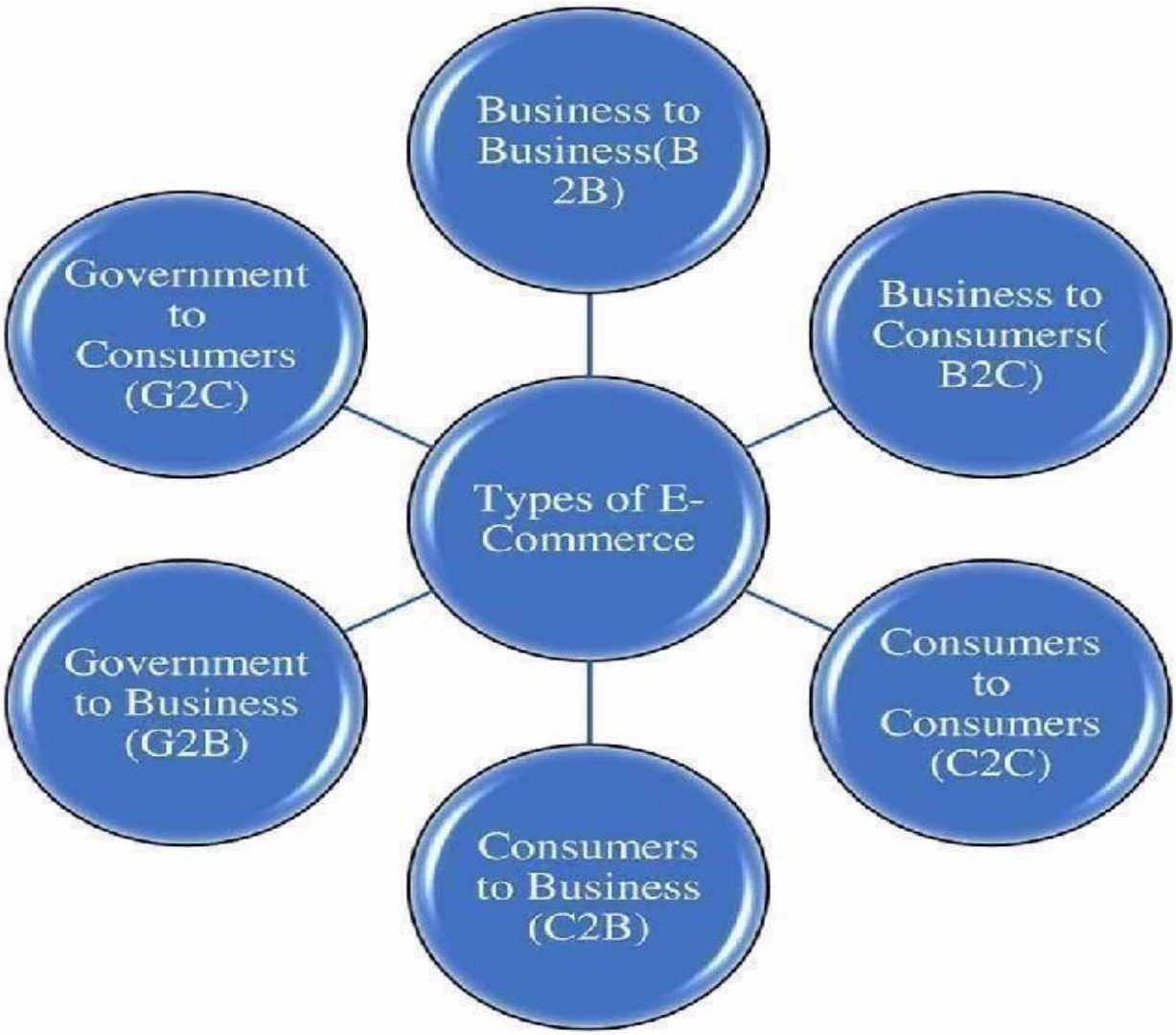
2.LITERATURE SURVEY

Electronic commerce (i.e., e-commerce) has been referred to as a popular business mode that firms can sell products, information and/or services through online channels and consumers can visit these Internet shops at any place at any time . E-commerce enterprises can build their online channels and review their marketing processes to provide more competitive products/services, thereby attracting more consumers . With the rapid development of the Internet and information technology, global e-commerce transactions have grown rapidly and witnessed a fast expanding trend in recent decades.

Systematically examines the use and value of business analytics in e-commerce through quantitative analysis. The research papers relevant to consumer-to-consumer e-commerce research collaboration and pinpoin t the research trend by using a content analysis approach.

Data analysis plays a crucial role as an intermediate step in an e-commerce platform. For example, business data analysis of e-commerce data can not only improve decision-making on external sales, customer profiles, and satisfaction, but also enhance internal product development, technical, and organizational workflows (Li, 2021b). Nowadays, the rapid development of block chain technology, artificial intelligence and machine learning not only helps enterprises to create and capture value, but also significantly affects the online shopping environment of consumers

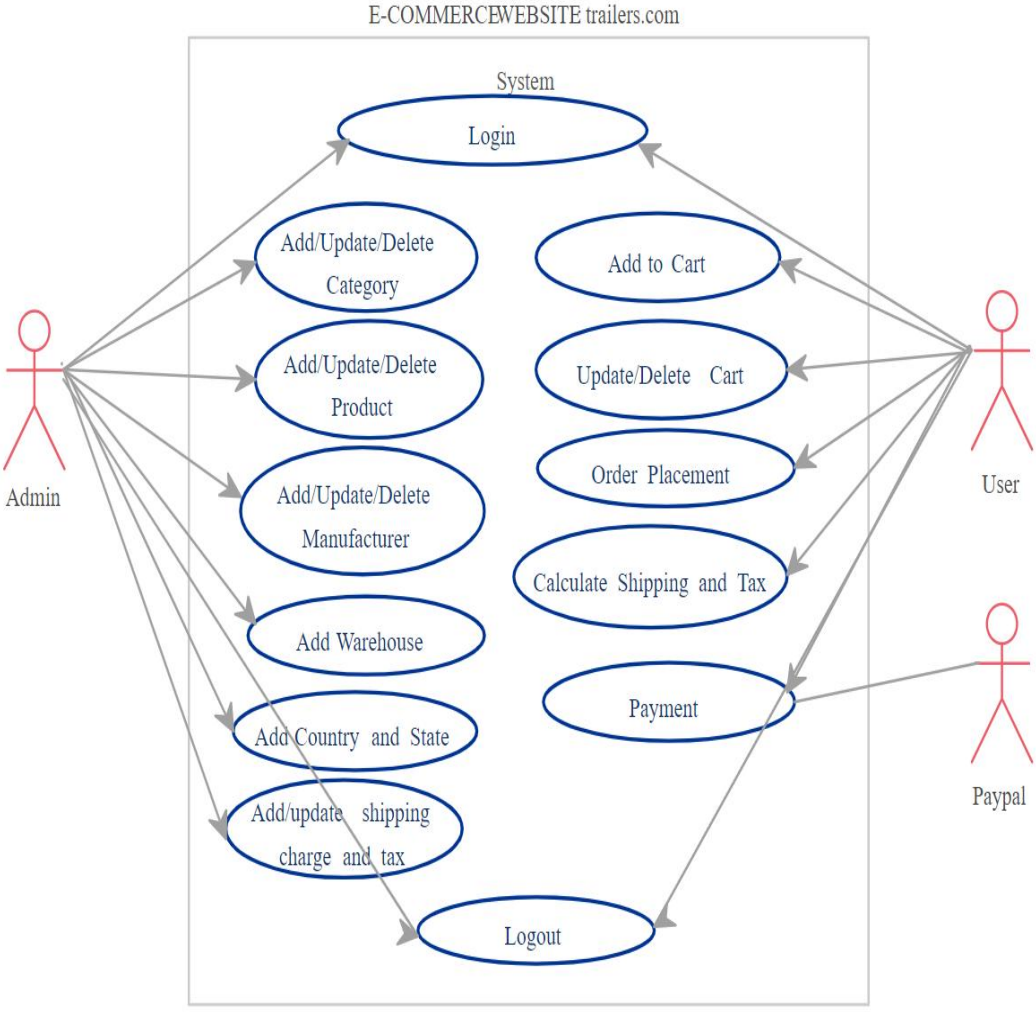
3.TYPES OF E-COMMERCE



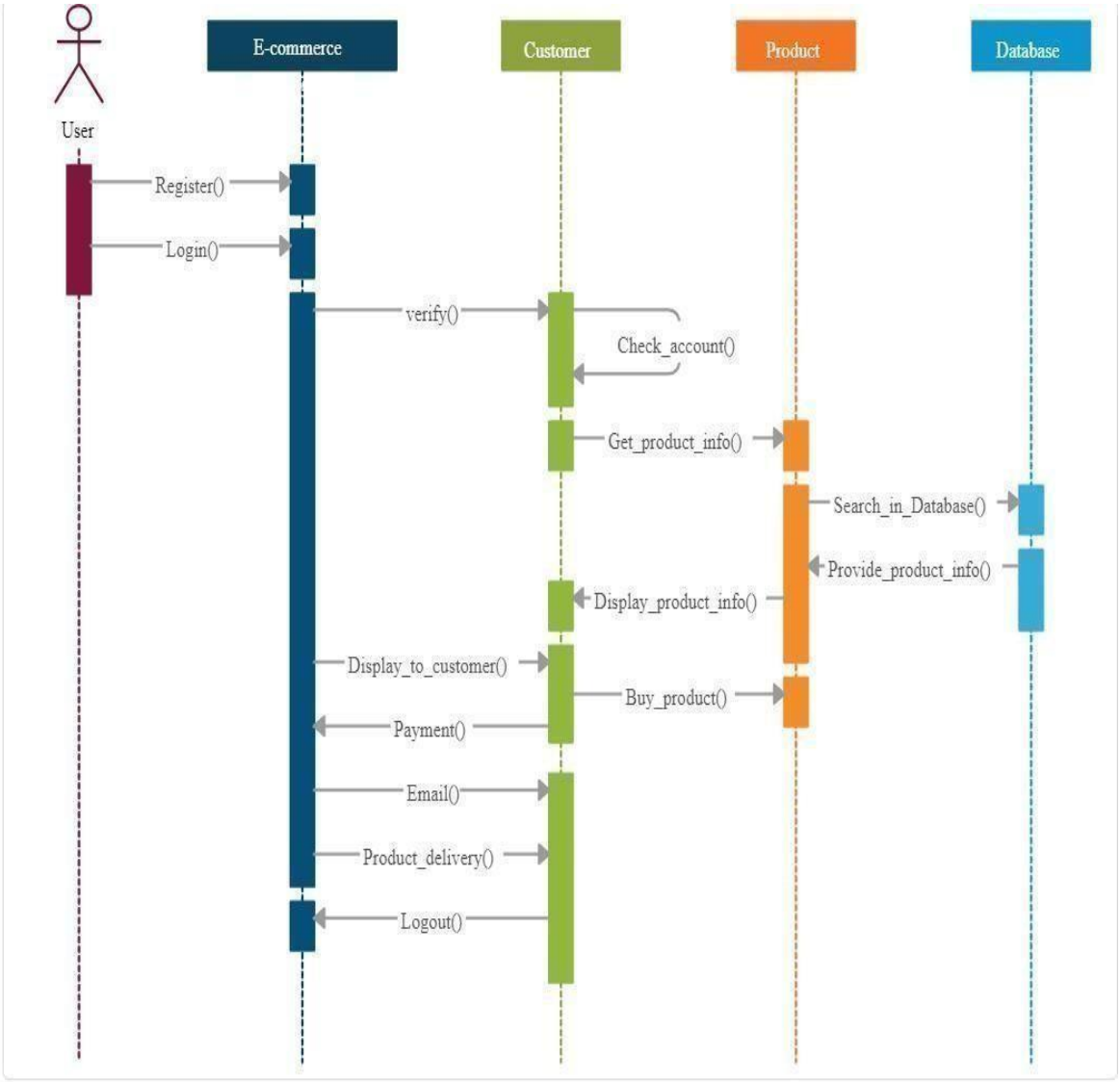
4.THEORITICAL ANALYSIS

4.1 UML DIAGRAM

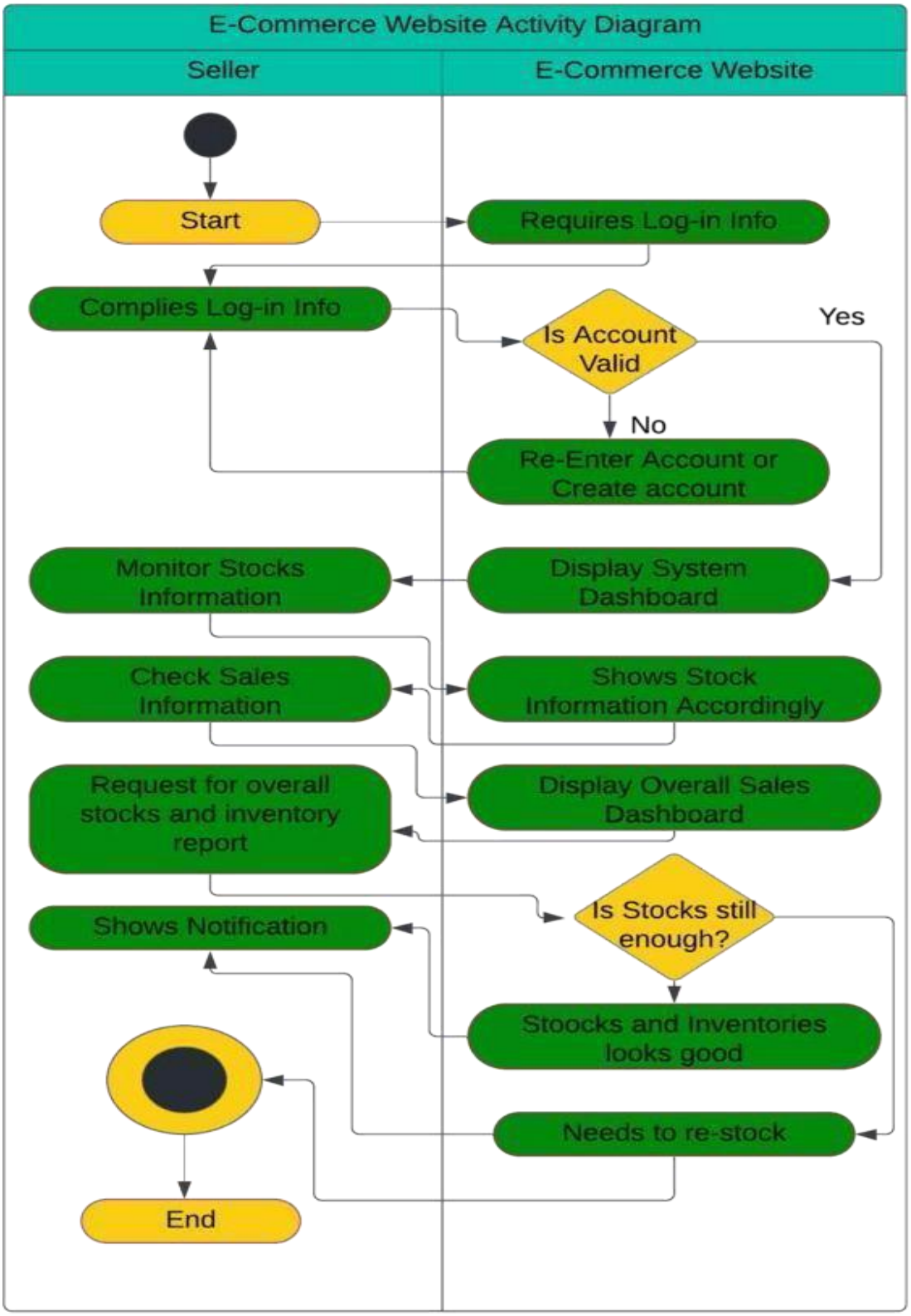
Usecase Diagram



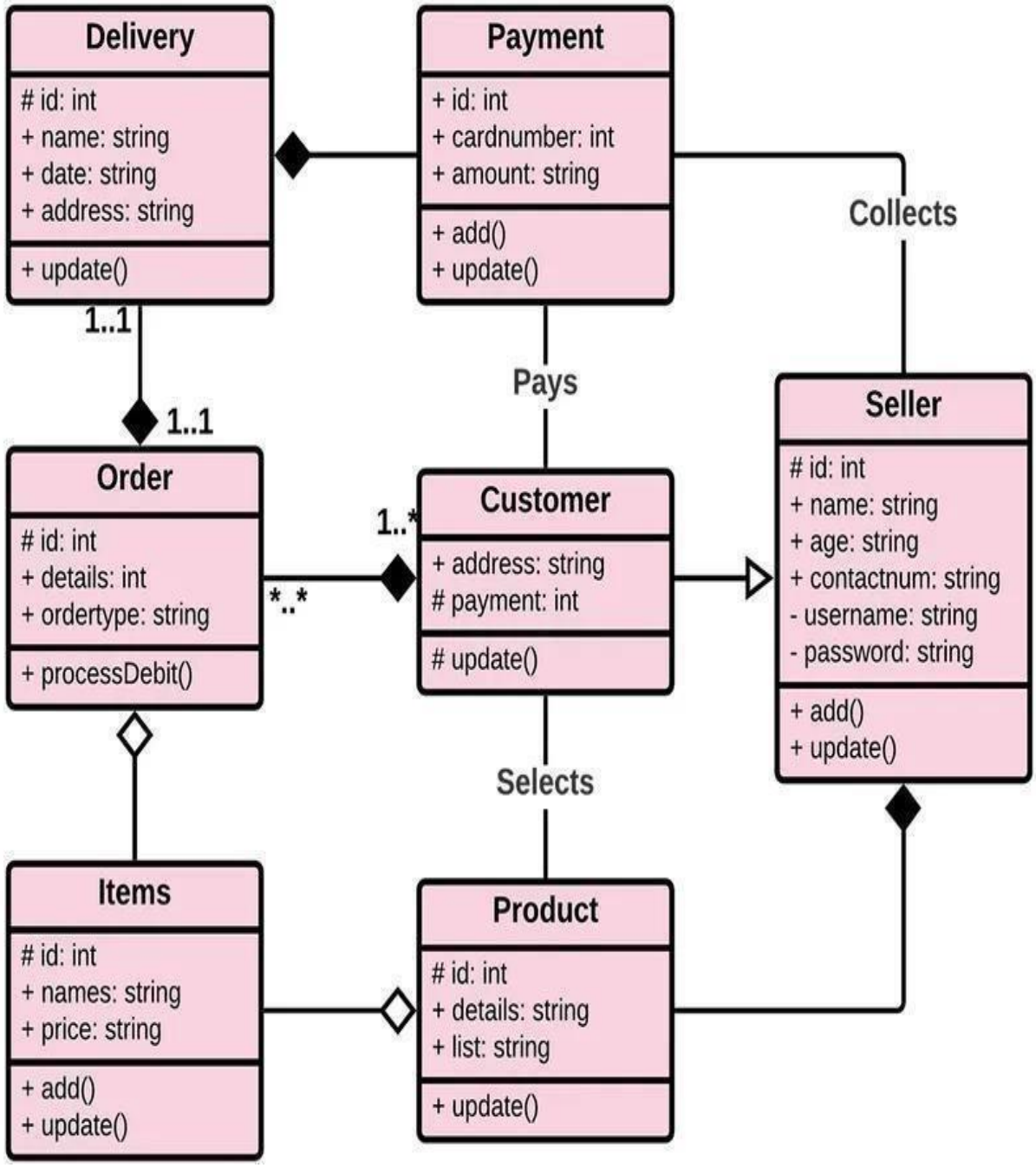
Sequence diagram



Activity Diagram



Class Diagram



4.1 SOFTWARE AND HARDWARE REQUIREMENTS

Hardware Requirement for E-commerce - Pentium II/III based Intel server running Linux can serve hundreds of unique customers each day. Low traffic sites can be easily served from a single machine depending on the needs of the business. High traffic sites require a backup of servers which automatically takes over operations in case of failure of primary ones.

Software Requirements for E-commerce – Several software are available free on theinternet that can be used to build e-commerce exchanges. Ex:- Linux OS, mySQL database ,Apache web server etc.,

E-commerce software Catalogue display

A catalogue is an organized list of goods & services being sold. An e-catalogue is a simplelist of goods and services in HTML form that appears on a webpage on the website of an ecommerce company.

Two types

(i)static catalogue -providing simple list of goods and services on offer.

(ii)dynamic catalogue -providing a detailed feature about items on sale in a databaseDetailed descriptions,shipment time etc.,

5.FLOWCHART


1
Seller details like name and mobile number
2
Create account of seller
3
Update store details
4
Add product details
5
Update product details
6
Submit the information
7
Store and product list on site

6.ADVANTAGES AND DISADVANTAGES

Advantages	
...for the customer	**...for the provider**
• Flexible shopping hours (7·24h) • No waiting queues (if net is available and software appropriately designed) • Shopping at home (we don't have to leave our apartment, refuel our car or buy a subway ticket, look for a parking place, etc.) • Individual needs can be covered (if customization is offered) • Global offers, more competition, pressure on prices	• Better customer service can be offered • Fast communication with customer • New customer potential through global visibility • No (traditional) intermediaries, who take away margins

Disadvantages	
...for the customer	**... for the provider**
• Security risks: o Data theft (e.g. stealing account or credit card numbers) o Identity theft (acting under our name or user identity) o Abuse (e.g. third person orders goods with our identity, gets them delivered and we have to pay for it) • Crime: o Bogus firm (firm does not really exist) o Fraud (e.g. order is confirmed, invoice has to be paid, but goods are never delivered) • Uncertain legal status (if something goes wrong, can we accuse the provider?)	• Higher logistics cost (goods have to be sent to the customer's location) • Anonymity of customers (how to make targeted advertisements?)

7.APPLICATIONS

- Data analytical is used in tracking customer's behavior towards product or service. You can use it to identify why sales are low,what product people buy,why they are buying them,how much they are spending on these products,how you can sell your product better,and many other queries.
- **HELPS TO BUILD A REBUST SUPPLY CHAIN**

Customers shop online for two reasons – convenience and better prices. This is why your supply chain needs to be robust. Are the products being advertised on your site available in adequate quantities in your inventory? If not, it will result in slower delivery and the level of customer satisfaction is bound to drop and the customer lifetime value will be negatively affected. On the flip side, having too many units in stock will take up space in your inventory and increase your costs. Striking this balance, between supply and demand is tricky, and this is why using analytics to forecast future sales is vital.

- Analyze information to detect fraud.
- Predict what's in store for you.
- Measure your marketing
- Personalize the customer's shopping experience

8.CONCLUSION

In UG Project Phase-1, we have worked on problem statement, literature survey and also done the experimental analyses which are required for the project to move forward. In experimentalanalysis we have discussed about the machine learning concepts and models and explained the algorithms to be used in the project. We also discussed about the flowcharts, use case diagrams, decision tree and sequence diagrams which are used in the project. Based on the experimental analysis we have designed the model for the project. Entire designing part is involved in UGProject Phase-1.

9.FUTURE SCOPE

UG Project Phase-2 is the extension of UG Project Phase-1. UG Project Phase-2 involves allthe coding and implementation of the design which we have retrieved from UG Project Phase-1. All the implementation is done and conclusions will be retrieved in the phase. We will also work on the applications, advantages, and disadvantages of the project in this phase. Future scope of the project will be also discussed in the UG Project Phase-2.

ANALYTICAL TOOLS FOR E-COMMERCE BUSINESS

A UG PROJECT PHASE-2 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

Inpartial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

REPALA PRAVALIKA	**19UK1A0531**
MAMIDALA KRISHINA SRI	**19UK1A0529**
SYED ABDUL MUJEEB	**19UK1A0569**
CHIDIRALA SAI KIRAN	**19UK1A0566**

Under the esteemed guidance of

Ms.A.Swathi

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated toJNTUH,Hyderabad)

Bollikunta,Warangal–506005

2019-2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

BOLLIKUNTA,WARANGAL–506005

2019 –2023

CERTIFICATE OFCOMPLETION

UG PROJECT PHASE-2

This is to certify that the UG Project Phase-2 entitled **"ANALYTICAL TOOLS FOR E-COMMERCE BUSINESS"** is being submitted by **REPALA PRAVALIKA(H.NO:19UK1A0531) , MAMIDALA KRISHINA SRI(H.NO:19UK1A0529),SYED ABDUL MUJEEB (H.NO:19UK1A0569),CHIDIRALA SAI KIRAN(H.NO:19UK1A0566)** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University Hyderabad** during the academic year**2022-23**,is a record of work carried out by them under the guidance and supervision.

Project Guide
Ms.A.Swathi
(AssistantProfessor)

Head of the Department
Dr.R.NaveenKumar
(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-2 in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR,** Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-2.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-2 and for their support in completing the UG Project Phase-2.

We express heartfelt thanks to the guide, **Ms.A.Swathi** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-2.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

REPALA PRAVALIKA (19UK1A0531)
MAMIDALA KRISHNA SRI (19UK1A0529)
SYED ABDUL MUJEEB (19UK1A0569)
CHIDIRALA SAI KIRAN (19UK1A0566)

ABSTRACT

The drivers for electronic commerce are both technological (under the tremendous pressure of innovation) and business oriented. This paper will highlight some guidelines for companies who are entering into E- commerce to create an E-commerce strategy or who already have an E-commerce presence to revise their existing strategy. E-Commerce is now seen as a reality for many businesses and a normal part of a business plan. The immediate benefits, in terms of cost savings, efficiencies and enhanced profitability are clear at every stage in the supply chain. Adopting e-business is no longer a competitive advantage, but a normal business process, without which an enterprise is unlikely to survive in the New economy . Year 2000 saw many Dot-com companies built up and many companies going into E-commerce however now it is a different story, more and more companies are failing, and investors are becoming cautious to invest money into Internet ventures. There is more cash needed then was expected.

TABLE OF CONTENTS:-

LIST OF FIGURES PAGE NO

LIST OF FIGURES **PAGE NO**

1.INTRODUCTION

“Electronic commerce, commonly written as E-Commerce, is the trading in products or services using computer networks, such as the Internet. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web for at least one part of the transaction’s life cycle, although it may also use other technologies such as E- Mail.

E-Commerce businesses may employ some or all of the following:

> Online shopping websites for retail sales direct to consumers,

> Providing or participating in online marketplaces, which process third-party businessto-consumer or consumer-to-consumer sales,

> Business-to-business buying and selling,

> Gathering and using demographic data through Web contacts and social media,

> Business-to-business electronic data interchange.

The role of analytics in e-commerce is If you have worked in financial industry, you will probably be aware of analytics playing a crucial role into risk and marketing strategy.However, E- Commerce industry goes beyond these two pillars. The primary job of E- Commerce industry is to make user experience on their website is delightful. Other than that they are simply a platform between sellers and buyers. With such focus on user experience, analytics itself becomes a product instead of just being business enabler. For instance, Recommender Engines you see on Amazon sidebar is a classic product. Now, you can appreciate the much broader role of analytics in E-Commerce industry. In the following section, we will talk more about broad functions where analytics is being actively used.

1.1 INTRODUCTION TO E-BUSINESS

"Electronic business, or E-Business, is the application of information and communication technologies (ICT) in support of all the activities of business. Commerce constitutes the exchange of products and services between businesses, groups and individuals and can be seen as one of the essential activities of any business. Electronic commerce focuses on the use of ICT to enable the externalq activities and relationships of the business with individuals, groupsand other businesses or E- Business refers to business with help of Internet i.e. doing business with the help of Internet network. The term was coined by IBM's marketing and Internet teamin 1996."

E-Business is a more general term than E-Commerce. However, in this book we will only use the term "E-Commerce", because every business transaction finally is involved in selling or buying of products or services. And the term "E-Commerce" obviously is more widespread than the term "E-Business".

10 Key Benefits Of Ecommerce For Your Business

- Low costs. A significant advantage of ecommerce is that launching an online store is much less costly than opening a physical store. ...
- Speed & Flexibility. ...
- Faster Buying Process ...
- Product Catalogue. ...
- Wider Customer Base. ...
- Customer Data Insights. ...
- Scalability. ...
- Reviews & Ratings

1.2OVERVIEW

Ecommerce analytics is the process of gathering data from all areas that have an impact on online stores and using this informaction to understand the trends and the shift in consumer behavior to make datadrive decisions that will drive more online sales.

1.3PURPOSE

Ecommerce analytics is the process of discovering ,interpreting,and communicating data pattern related to online business.ecommerce analytics helps measure user behavior, performance trends. Ecommerce analytics helps centralize and mange data.

The focus of analytics is on issues that matter most to the business,and the performance metrics are helpful in identifying and solving problems in real time. Because technology evolves rapidly and shopping trends shift on a daily basis ecommerce is in a constant state of adaptation that can leave brands fumbling in the dark. To keep up, ecommerce businesses must anticipate changes in the market using reliable data insights. In short, they need effective ecommerce analytics .Ecommerce analytics simply refers to any tool or strategy designed to analyse large amounts of data in order to produce actionable insights. Because it exists in an almost entirely virtual space, ecommerce generates complex, comprehensive datasets — particularly those related to client behaviour. More data was created in 2017 than was created during the previous 5,000 years combined. That is a lot of data to measure, parse, and analyse. Finding the right ecommerce analytics tools for the job, however, can offer your brand an immeasurable advantage over the competition.

2.CODE SNIPPETS

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load in

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# Any results you write to the current directory are saved as output.
```

Lets look at the data of a superstore in US and get some insights from it.

Figure 1: ..ipynb code describing importing libraries and reading the data.

IMPORTING LIBRARIES

```
[ ] import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as sns
    %matplotlib inline
    import warnings
    warnings.filterwarnings('ignore')
    from sklearn.preprocessing import LabelEncoder
```

READING THE DATA

Lets us read the data and take a look at first few rows.

```
[ ] import pandas as pd
    url='https://docs.google.com/spreadsheets/d/1Flg3KjnKIQvToqbGm2HAp39vjRsxujnI/edit?usp=sharing&ouid=113831357595098568214&rtpof=true&sd=true'
    # input = dataset - us superstore.xls
    url='https://drive.google.com/uc?id=' + url.split('/')[-2]
    df = pd.read_excel(url)
```

EXPLORATORY DATA ANALYSIS

```
[ ]  #row,column count of data
     df.shape
```

```
     (9994, 21)
```

```
[ ]  #column names of table
     df.columns
```

```
     Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
            'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State',
            'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category',
            'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit'],
           dtype='object')
```

```
City                      object
State                     object
Postal Code                int64
Region                    object
Product ID                object
Category                  object
Sub-Category              object
Product Name              object
Sales                    float64
Quantity                   int64
Discount                 float64
Profit                   float64
dtype: object
```

Lets check if there are any missing values in the data

Figure 2: .ipynb code describing exploratory data analysis .

```
[ ]  df.isnull().sum()
```

```
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
Product ID      0
Category        0
Sub-Category    0
Product Name    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
```

There are no missing values. Hence we can go exploratory analysis part directly

WAIT.... Check for unnecessary columns and drop them if not required

'Row ID' column is nothing but the serial number so we can drop this column.

Figure 3: .ipynb code describing isnull( ) and sum( ) methods .

```
#Dropping  Row ID column and assigning to df
df=df.drop('Row ID',axis=1)
df.head()
```

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
CA-2016-52156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136
CA-2016-52156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400	3	0.00	219.5820
CA-2016-38688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	0.00	6.8714
US-2015-08966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	0.45	-383.0310
US-2015-08966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	0.20	2.5164

Figure 4: .ipynb code describing row id column and assigning to df from the dataset

```
df['Country'].value_counts()
```

```
United States    9994
Name: Country, dtype: int64
```

Clearly the data is for US country only, so we can drop the 'Country' column as we dont need any analysis to be done based on it.

```
[13] #dropping Country column
     df=df.drop('Country',axis=1)
     df.head()
```

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
0	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136
1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400	3	0.00	219.5820
2	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	Los Angeles	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	0.00	6.8714
3	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	0.45	-383.0310
4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	0.20	2.5164

We can analyse the data further in 3 different ways

- PRODUCT LEVEL ANALYSIS
- CUSTOMER LEVEL ANALYSIS
- ORDER LEVEL ANALYSIS

images.jfif

Lets look at the product categories available to shop for customers

Figure 5: .ipynb code describing country column from the dataset

```
[14] df['Category'].unique()
```

```
array(['Furniture', 'Office Supplies', 'Technology'], dtype=object)
```

```
[15] #number of products in each category
     df['Category'].value_counts()
```

```
Office Supplies    6026
Furniture          2121
Technology         1847
Name: Category, dtype: int64
```

```
[16] #number of Sub-categories products are divided.
     df['Sub-Category'].nunique()
```

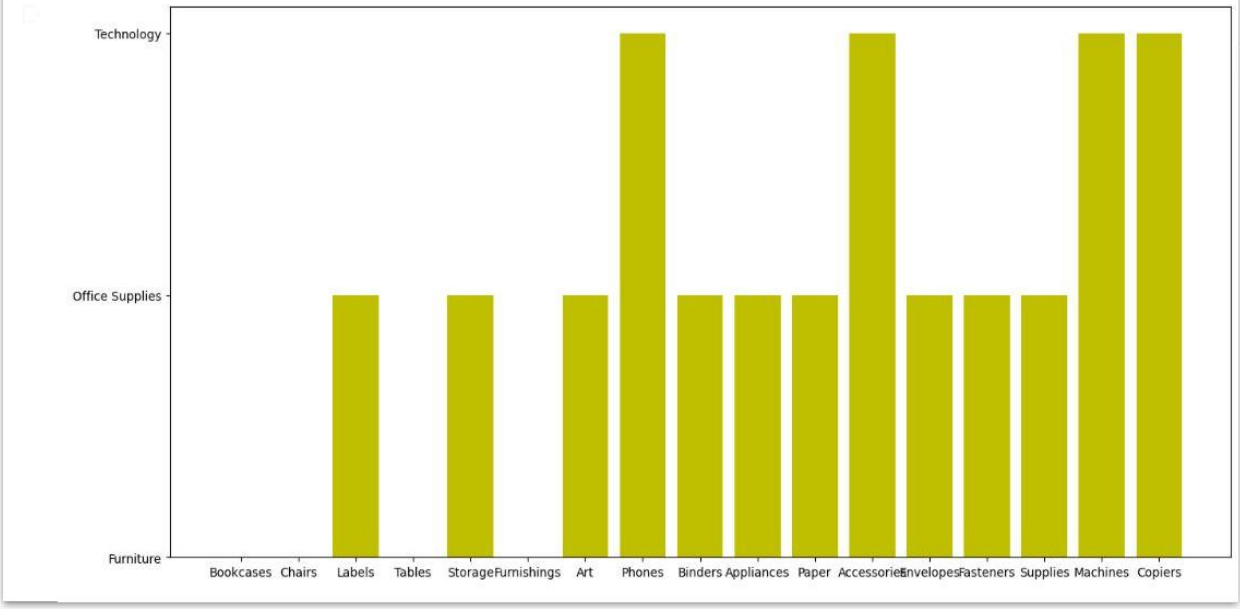
```
17
```

```
[17] #number of products in each sub-category
     df['Sub-Category'].value_counts()
```

```
Binders        1523
Paper          1370
Furnishings     957
Phones          889
Storage         846
Art             796
Accessories     775
Chairs          617
Appliances      466
Labels          364
Tables          319
Envelopes       254
Bookcases       228
Fasteners       217
Supplies        190
Machines        115
Copiers          68
Name: Sub-Category, dtype: int64
```

Figure 6: .ipynb code describing number of products in each category and number of products ineach sub-category

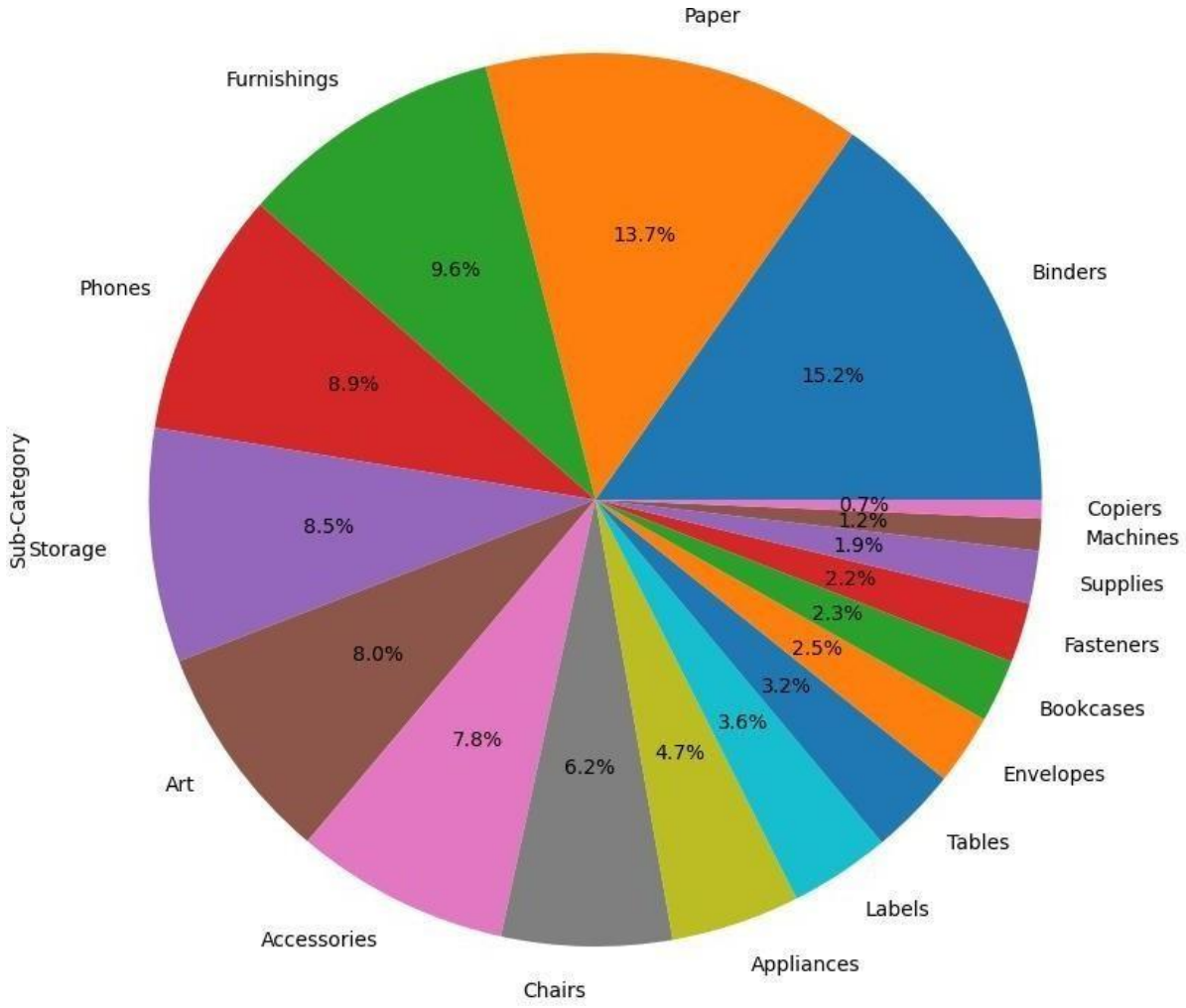
```
#Lets see how sub-categories are distributed wrt to category
plt.figure(figsize=(16,8))
plt.bar('Sub-Category','Category',data=df,color='y')
plt.show()
```



From this graph, one can easily makeout which Category & Sub-Category to choose when they are looking to purchase a product

Figure 7: .ipynb code describing bar plot

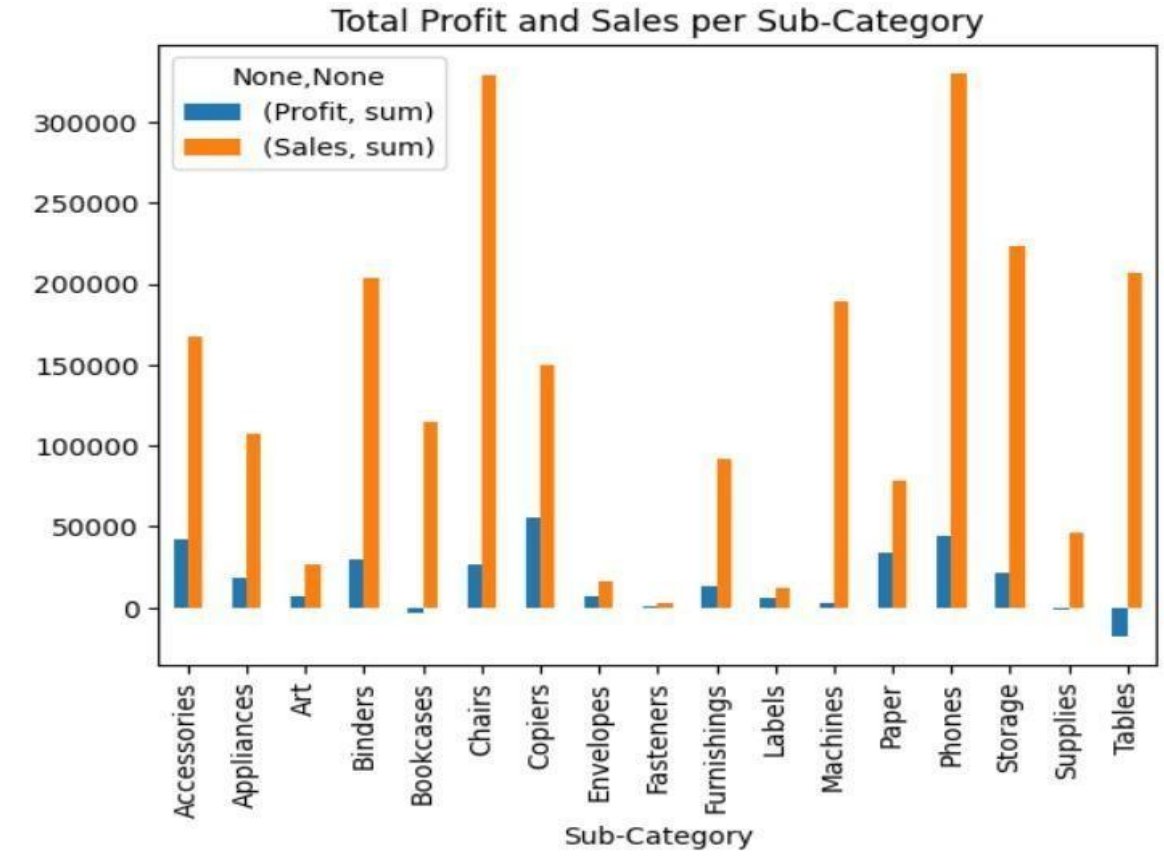
```
plt.figure(figsize=(12,10))
df['Sub-Category'].value_counts().plot.pie(autopct="%1.1f%%")
plt.show()
```



The store has wide variety of Office Supplies especially in Binders and Paper department.

Figure 8: .ipynb code discribing the pie plot

```
[20] df.groupby('Sub-Category')['Profit','Sales'].agg(['sum']).plot.bar()
     plt.title('Total Profit and Sales per Sub-Category')
     # plt.legend('Profit')
     # plt.legend('Sales')
     plt.show()
```



Highest profit is earned in Copiers while Selling price for Chairs and Phones is extremely high compared to other products.

Another interesting fact- people dont prefer to buy Tables and Bookcases from Superstore. Hence these departments are in loss.

Figure 9: .ipynb code describing the bar plot

```
[21] #number of products available in store
     df['Product Name'].nunique()

     1850
```

```
[22] df['Product Name'].value_counts()

     Staple envelope                                                                    48
     Staples                                                                            46
     Easy-staple paper                                                                  46
     Avery Non-Stick Binders                                                            20
     Staples in misc. colors                                                            19
                                                                                        ..
     Boston 1900 Electric Pencil Sharpener                                               1
     RCA ViSYS 25423RE1 Corded phone                                                     1
     Canon Color ImageCLASS MF8580Cdw Wireless Laser All-In-One Printer, Copier, Scanner 1
     Newell 342                                                                          1
     Eldon Jumbo ProFile Portable File Boxes Graphite/Black                              1
     Name: Product Name, Length: 1850, dtype: int64
```

```
[23] #Distribution of Top 10 Products
     plt.figure(figsize=(12,10))
     df['Product Name'].value_counts().head(10).plot.pie(autopct="%1.1f%%")

     <Axes: ylabel='Product Name'>
```

Figure 10: .ipynb code describing number of products available in store and product name from the data set

```
[ ]  #Distribution of Top 10 Products
     plt.figure(figsize=(12,10))
     df['Product Name'].value_counts().head(10).plot.pie(autopct="%1.1f%%")
```

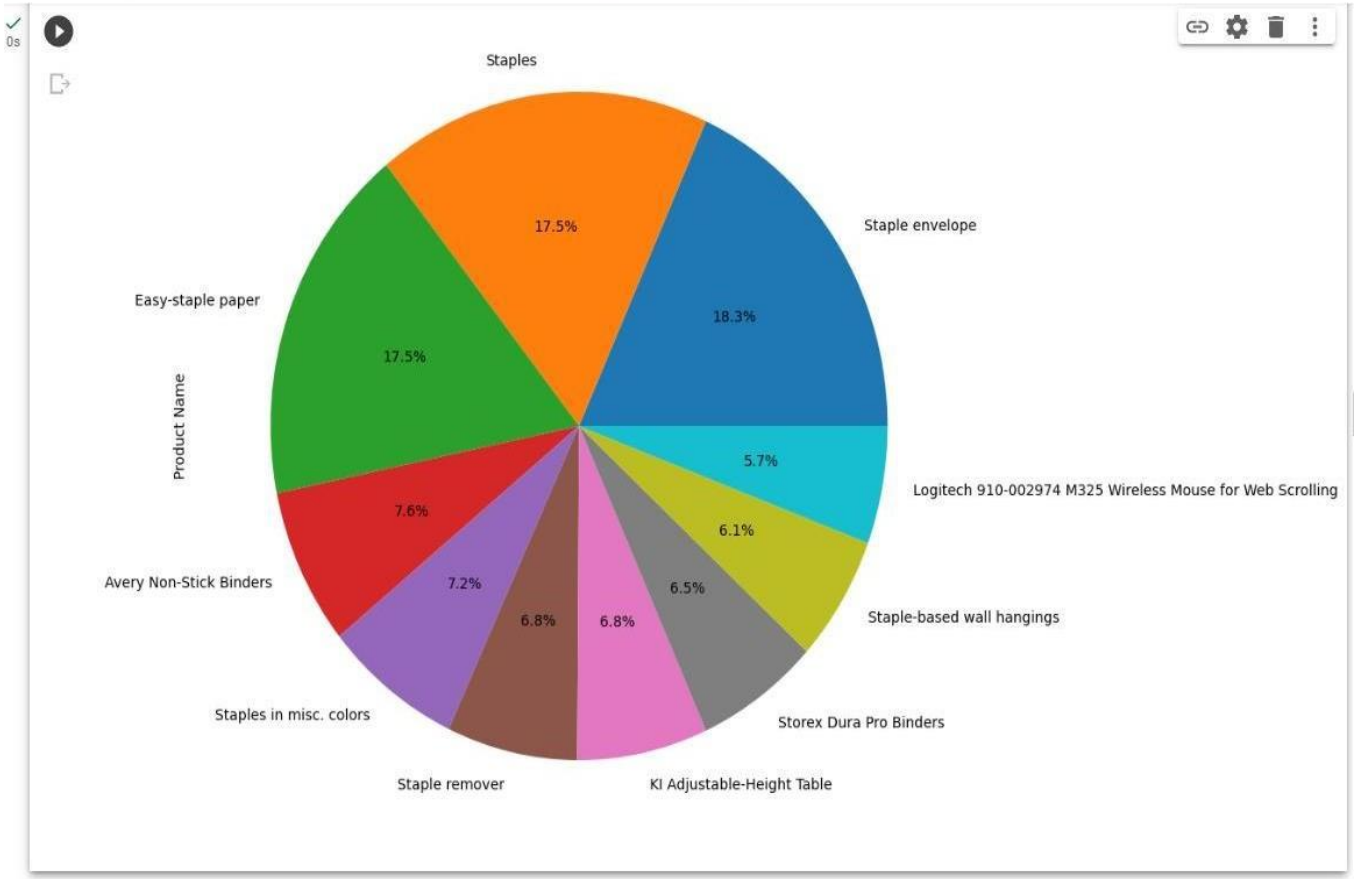
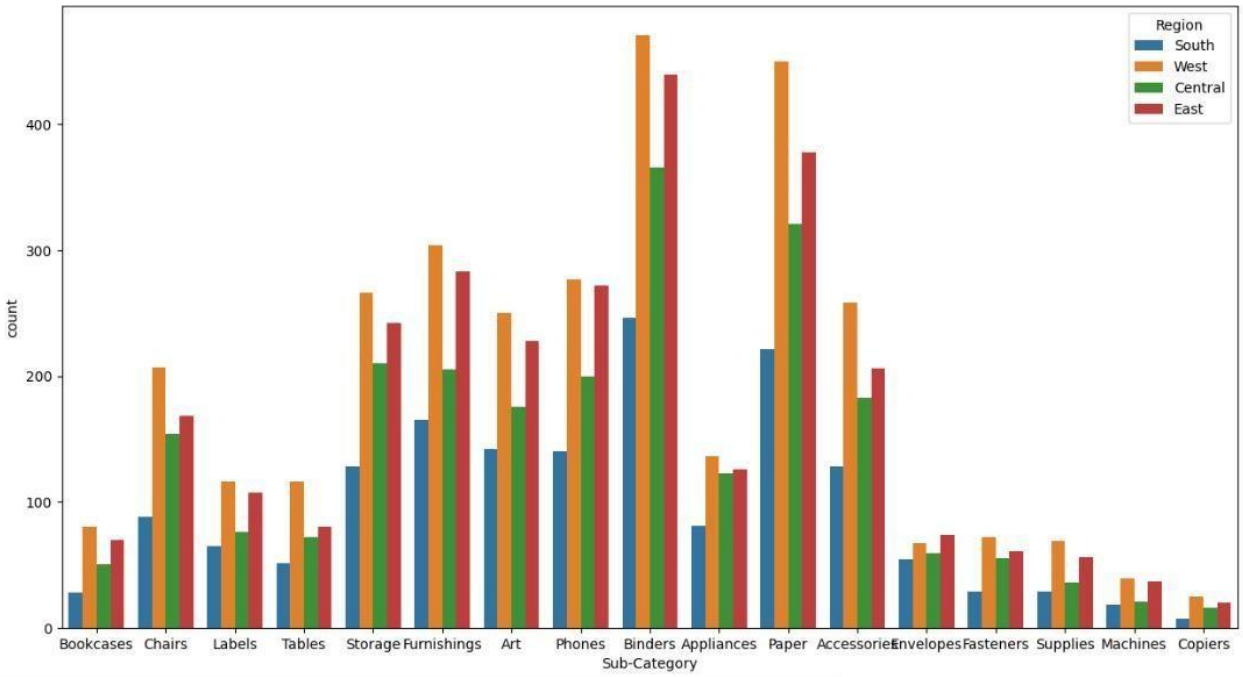


Figure 11: .ipynb code describing pie plot

```
[24] #Count of Sub-Category region wise
     plt.figure(figsize=(15,8))
     sns.countplot(x="Sub-Category", hue="Region", data=df)
     plt.show()
```



People residing in Western part of US tend to order more from superstore.

To understand the data better. Lets create some new columns like Cost,Profit%

Figure 12: .ipynb code describing bar plot

```
[25] df['Cost']=df['Sales']-df['Profit']
     df['Cost'].head()
```

```
0     220.0464
1     512.3580
2       7.7486
3    1340.6085
4      19.8516
Name: Cost, dtype: float64
```

```
[26] df['Profit %']=(df['Profit']/df['Cost'])*100
```

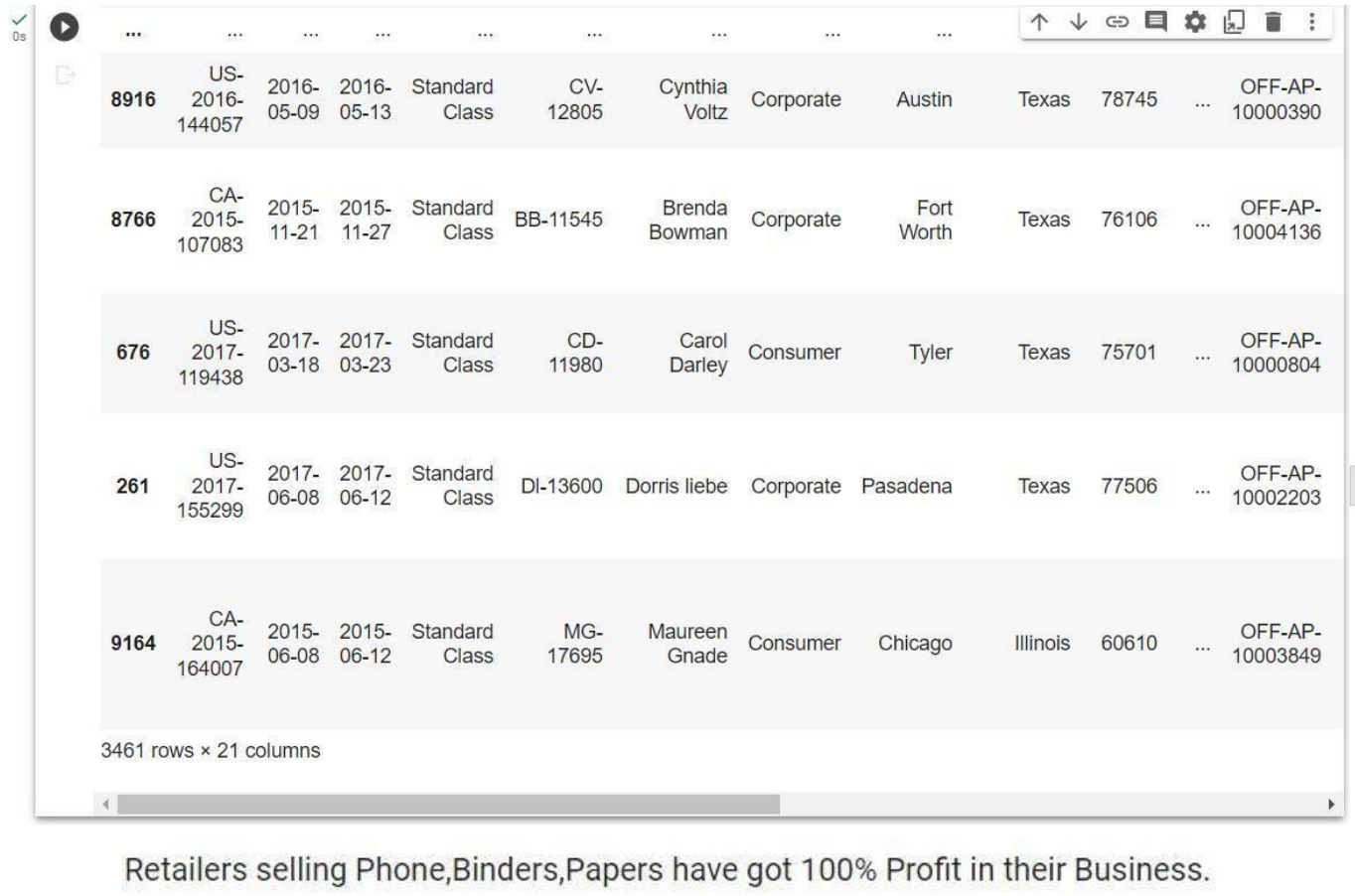
```
[27] #Profit Percentage of first 5 product names
     df.iloc[[0,1,2,3,4],[14,20]]
```

	Product Name	Profit %
0	Bush Somerset Collection Bookcase	19.047619
1	Hon Deluxe Fabric Upholstered Stacking Chairs,...	42.857143
2	Self-Adhesive Address Labels for Typewriters b...	88.679245
3	Bretford CR4500 Series Slim Rectangular Table	-28.571429
4	Eldon Fold 'N Roll Cart System	12.676056

Figure 13: .ipynb code describing profit percentage of first 5 product names

```
#Products with high Profit Percentage
df.sort_values(['Profit %','Product Name'],ascending=False).groupby('Profit %').head(5)
```

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Postal Code	...	Product ID
1696	CA-2015-156335	2015-09-25	2015-09-28	Second Class	PO-19195	Phillina Ober	Home Office	Bayonne	New Jersey	7002	...	TEC-PH-10002726
4611	CA-2016-109407	2016-01-24	2016-01-26	Second Class	HG-14965	Henry Goldwyn	Corporate	Gulfport	Mississippi	39503	...	OFF-BI-10000285
3984	CA-2016-135265	2016-07-07	2016-07-09	Second Class	CC-12370	Christopher Conant	Consumer	Los Angeles	California	90045	...	OFF-PA-10002923
4651	CA-2017-150987	2017-04-08	2017-04-12	Standard Class	AH-10120	Adrian Hane	Home Office	San Francisco	California	94110	...	OFF-PA-10002923
6600	CA-2014-154095	2014-12-02	2014-12-07	Standard Class	ON-18715	Odella Nelson	Corporate	Clinton	Maryland	20735	...	OFF-PA-10002923
...	...	...	...	...	...	...	...	...	...	...	...	...
	US-											



...	...	...	...	...	...	...	...	...	...	...	...	...
8916	US-2016-144057	2016-05-09	2016-05-13	Standard Class	CV-12805	Cynthia Voltz	Corporate	Austin	Texas	78745	...	OFF-AP-10000390
8766	CA-2015-107083	2015-11-21	2015-11-27	Standard Class	BB-11545	Brenda Bowman	Corporate	Fort Worth	Texas	76106	...	OFF-AP-10004136
676	US-2017-119438	2017-03-18	2017-03-23	Standard Class	CD-11980	Carol Darley	Consumer	Tyler	Texas	75701	...	OFF-AP-10000804
261	US-2017-155299	2017-06-08	2017-06-12	Standard Class	DI-13600	Dorris liebe	Corporate	Pasadena	Texas	77506	...	OFF-AP-10002203
9164	CA-2015-164007	2015-06-08	2015-06-12	Standard Class	MG-17695	Maureen Gnade	Consumer	Chicago	Illinois	60610	...	OFF-AP-10003849

3461 rows × 21 columns

Retailers selling Phone,Binders,Papers have got 100% Profit in their Business.

Figure 14: .ipynb code describing products with high profit percentage

LETS LOOK AT THE DATA WRT TO CUSTOMER LEVEL

```
[29] df['Customer ID'].nunique()
```

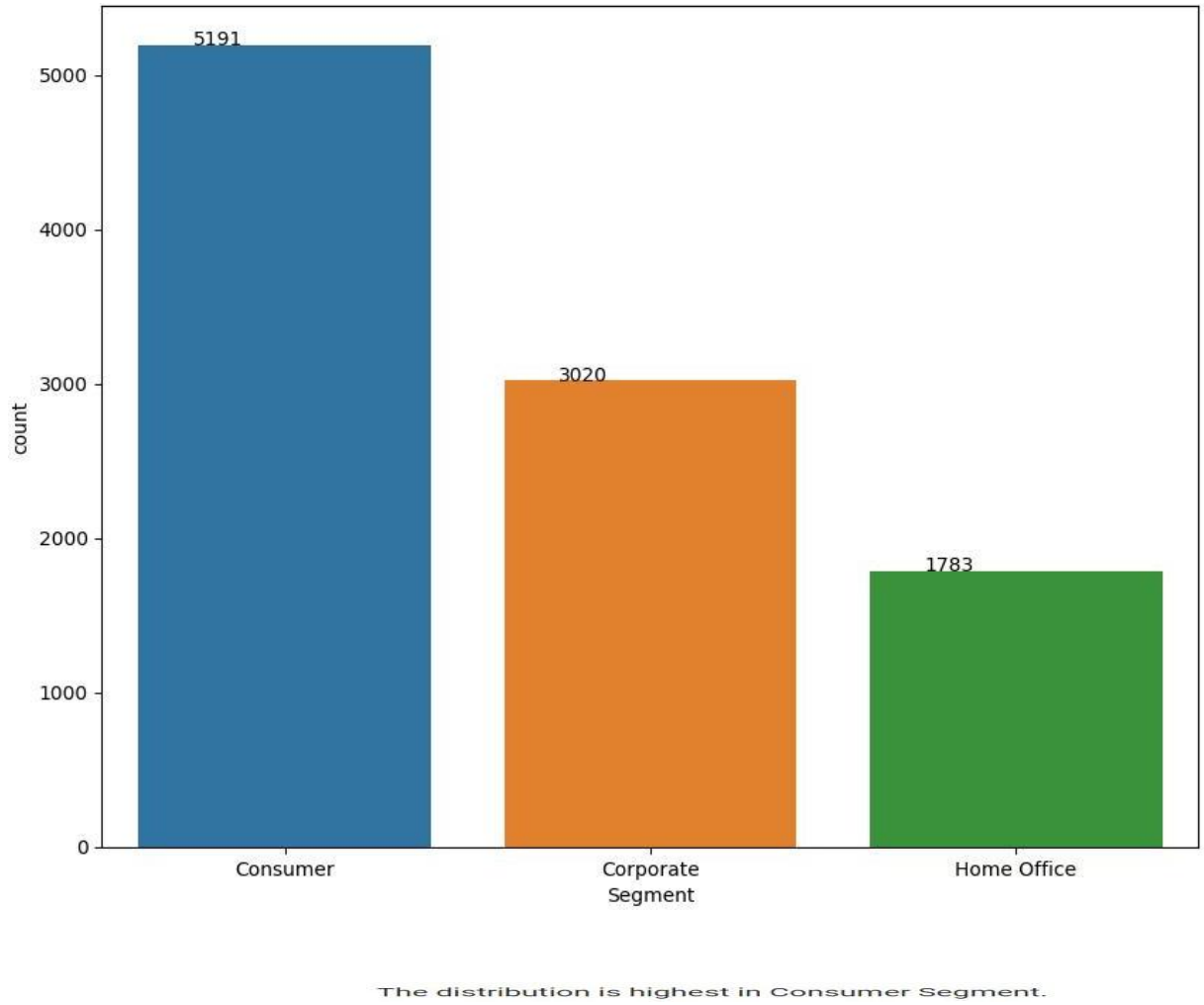
```
793
```

```
[30] #Top 10 customers who order frequently
     df_top10=df['Customer Name'].value_counts().head(10)
     df_top10
```

```
William Brown          37
John Lee               34
Matt Abelman           34
Paul Prost             34
Chloris Kastensmidt    32
Seth Vernon            32
Jonathan Doherty       32
Edward Hooks           32
Zuschuss Carroll       31
Emily Phan             31
Name: Customer Name, dtype: int64
```

Figure 15: .ipynb code describing top 10 customers who order frequently

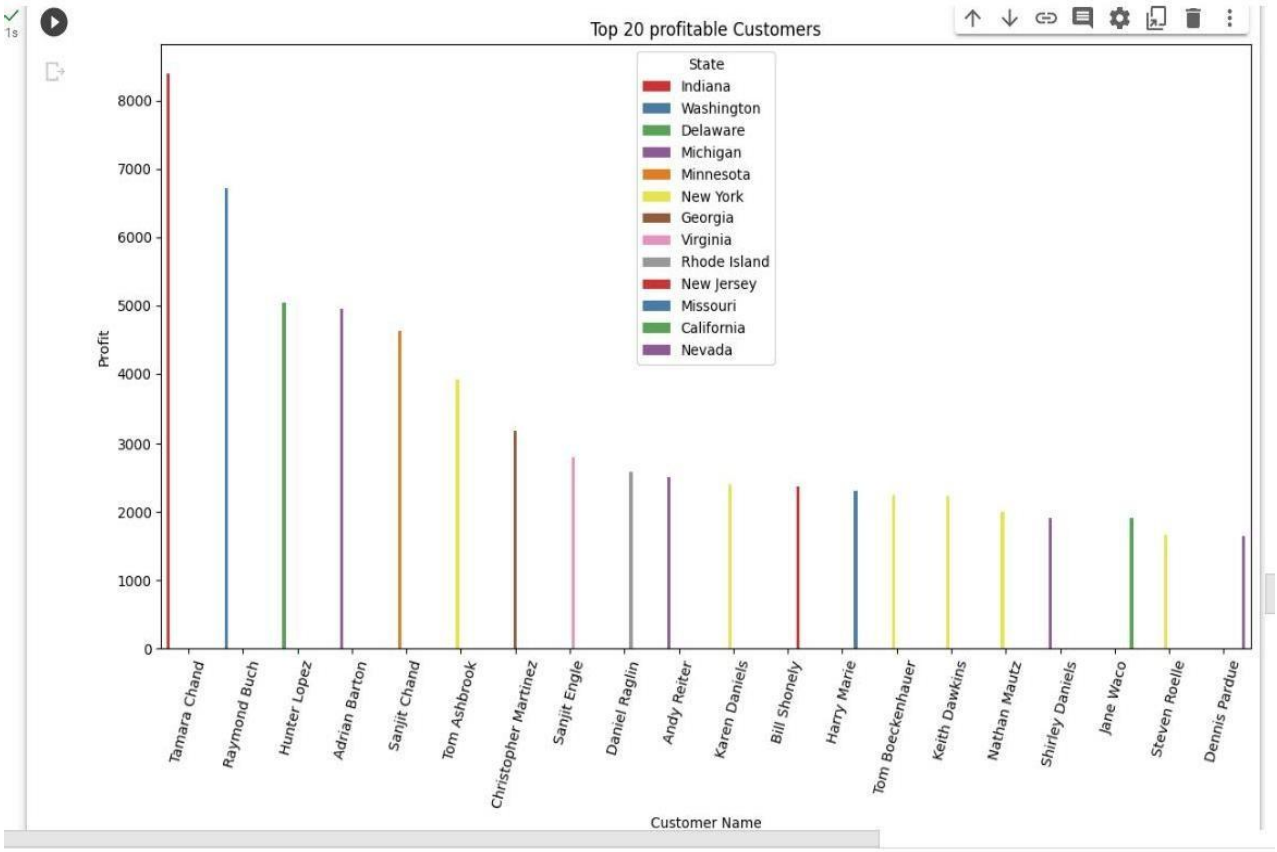
```
fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111)
sns.countplot(x='Segment', data=df, ax=ax)

for patch in ax.patches:
    ax.annotate('{:.0f}'.format(patch.get_height()), (patch.get_x()+0.15, patch.get_height()+1))

plt.show()
```



The distribution is highest in Consumer Segment.

Figure 16: .ipynb code describing the bar plot

```
[32] #Top 20 Customers who benefitted the store
     sortedTop20 = df.sort_values(['Profit'], ascending=False).head(20)
     fig = plt.figure(figsize=(12,8))
     ax = fig.add_subplot(111)
     p = sns.barplot(x='Customer Name', y='Profit',hue='State',palette='Set1', data=sortedTop20, ax=ax)
     ax.set_title("Top 20 profitable Customers")
     ax.set_xticklabels(p.get_xticklabels(), rotation=75)
     plt.tight_layout()
     plt.show()
```



We see that majority of the Profitable Customers are from New York and Michigan State.

Lets do some do some Analysis with Order details of the data

Figure 17: .ipynb code describing the bar plot

```
[33] #number of unique orders
     df['Order ID'].nunique()
```

```
5009
```

```
[34] #Calculating the time taken for an order to ship and converting the no. of days in int format
     df['Shipment Duration']=(pd.to_datetime(df['Ship Date'])-pd.to_datetime(df['Order Date'])).dt.days
     df['Shipment Duration']
```

```
0       3
1       3
2       4
3       7
4       7
       ..
9989    2
9990    5
9991    5
9992    5
9993    5
Name: Shipment Duration, Length: 9994, dtype: int64
```

```
[35] df.iloc[:,[0,3,21]]
```

	Order ID	Ship Mode	Shipment Duration
0	CA-2016-152156	Second Class	3
1	CA-2016-152156	Second Class	3
2	CA-2016-138688	Second Class	4
3	US-2015-108966	Standard Class	7
4	US-2015-108966	Standard Class	7
...	...	...	...
9989	CA-2014-110422	Second Class	2
9990	CA-2017-121258	Standard Class	5
9991	CA-2017-121258	Standard Class	5
9992	CA-2017-121258	Standard Class	5
9993	CA-2017-119914	Second Class	5

9994 rows × 3 columns

Lets find out some more details about each Customer like total products purchased,Products they purchase,First Purchase Date,Last Purchase Date,Location from where the Customer placed an order.

Figure 18: .ipynb code describing the no. Of unique orders and calculating the time taken for an order

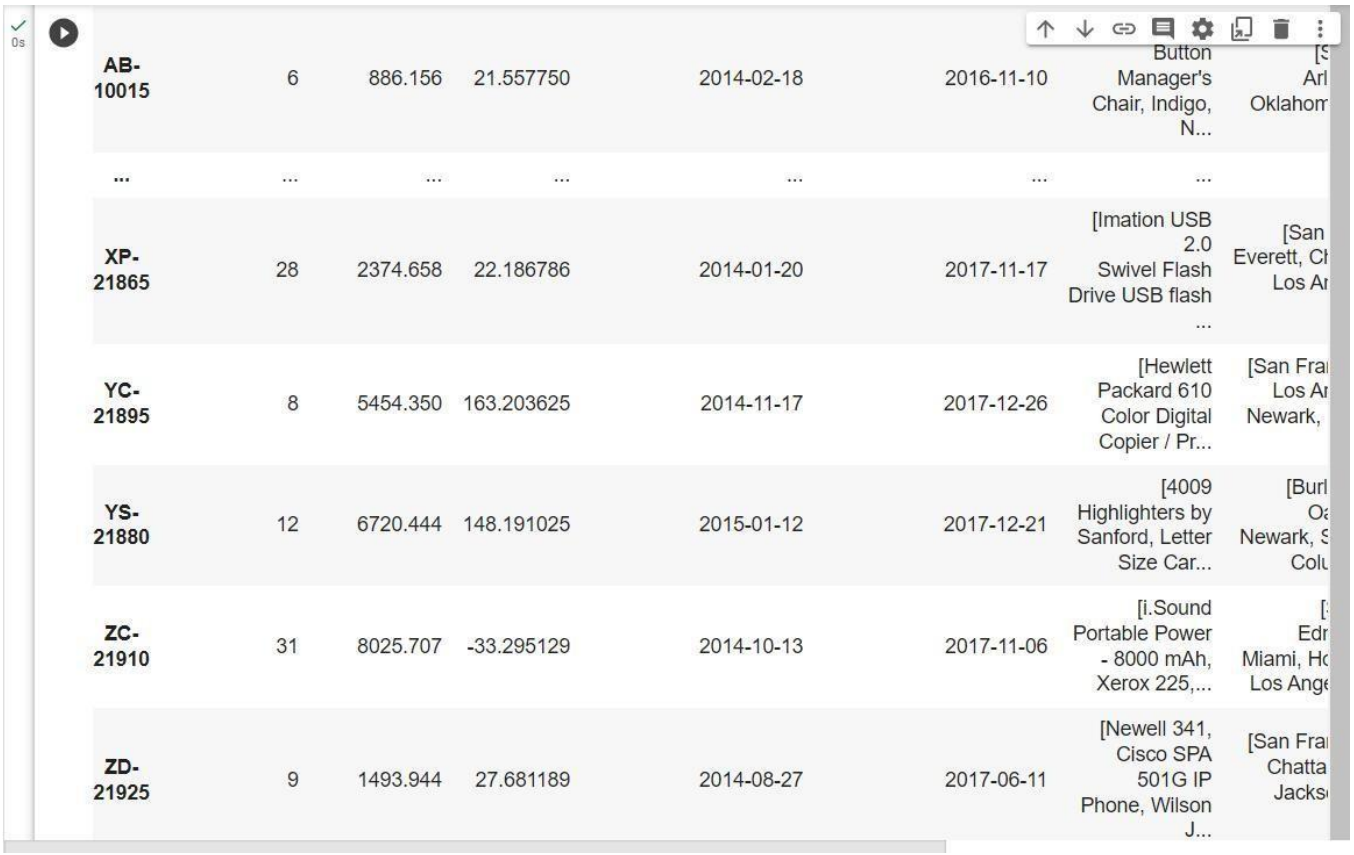
```
[36] print(df.columns)
```

```
Index(['Order ID', 'Order Date', 'Ship Date', 'Ship Mode', 'Customer ID',
       'Customer Name', 'Segment', 'City', 'State', 'Postal Code', 'Region',
       'Product ID', 'Category', 'Sub-Category', 'Product Name', 'Sales',
       'Quantity', 'Discount', 'Profit', 'Cost', 'Profit %',
       'Shipment Duration'],
      dtype='object')
```

```
[37] def agg_customer(x):
         d = []
         d.append(x['Order ID'].count())
         d.append(x['Sales'].sum())
         d.append(x['Profit'].mean())
         d.append(pd.to_datetime(x['Order Date']).min())
         d.append(pd.to_datetime(x['Order Date']).max())
         d.append(x['Product Name'].unique())
         d.append(x['City'].unique())
         return pd.Series(d, index=['#Purchases', 'Total_Sales', 'Average Profit', 'First_Purchase_Date', 'Latest_Pu
```

```
#grouping based on Customer ID and applying the function we created above
df_agg = df.groupby('Customer ID').apply(agg_customer)
df_agg
```

Customer ID	#Purchases	Total_Sales	Average Profit	First_Purchase_Date	Latest_Purchase_Date	Products Purchased	Locati
AA-10315	11	5563.560	-32.989318	2014-03-31	2017-06-29	[Belkin 325VA UPS Surge Protector, 6', Avery B...	[Mi San Rc
AA-10375	15	1056.390	18.492160	2014-04-21	2017-12-11	[Sterilite Officeware Hinged File Box, Colored...	[Mes Lo Sale
AA-10480	12	1790.512	36.318950	2014-05-04	2017-04-15	[Xerox 1967, DAX Wood Document Frame, Strathmo...	M S
AA-10645	18	5086.935	47.655739	2014-06-22	2017-11-05	[Padded Folding Chairs, Black, 4/Carton, Panas...	[Ch Ge Sa



AB-10015	6	886.156	21.557750	2014-02-18	2016-11-10	Button Manager's Chair, Indigo, N...	[S Arl Oklahom
...	...	...	...	...	...	...	
XP-21865	28	2374.658	22.186786	2014-01-20	2017-11-17	[Imation USB 2.0 Swivel Flash Drive USB flash ...	[San Everett, Cl Los Ar
YC-21895	8	5454.350	163.203625	2014-11-17	2017-12-26	[Hewlett Packard 610 Color Digital Copier / Pr...	[San Fra Los Ar Newark,
YS-21880	12	6720.444	148.191025	2015-01-12	2017-12-21	[4009 Highlighters by Sanford, Letter Size Car...	[Burl Oa Newark, S Colu
ZC-21910	31	8025.707	-33.295129	2014-10-13	2017-11-06	[i.Sound Portable Power - 8000 mAh, Xerox 225,...	[Edr Miami, Hc Los Ange
ZD-21925	9	1493.944	27.681189	2014-08-27	2017-06-11	[Newell 341, Cisco SPA 501G IP Phone, Wilson J...	[San Fra Chatta Jacks

Figure 19: .ipynb code describing grouping based on the customer ID and applying the function

```
[39] #extracting the year of order
     df['order year']=df['Order Date'].dt.year
     df['order year'].head()
```

```
0    2016
1    2016
2    2016
3    2015
4    2015
Name: order year, dtype: int64
```

```
[40] df.columns
```

```
Index(['Order ID', 'Order Date', 'Ship Date', 'Ship Mode', 'Customer ID',
       'Customer Name', 'Segment', 'City', 'State', 'Postal Code', 'Region',
       'Product ID', 'Category', 'Sub-Category', 'Product Name', 'Sales',
       'Quantity', 'Discount', 'Profit', 'Cost', 'Profit %',
       'Shipment Duration', 'order year'],
      dtype='object')
```

```
[41] fig = plt.figure(figsize=(16, 8))
     ax = fig.add_subplot(111)
     sns.barplot(x='order year', y='Profit', hue='Sub-Category', palette='Paired', data=df, ax=ax)
     for o in ax.patches:
         ax.annotate('{:.0f}'.format(o.get_height()), (o.get_x()+0.15, o.get_height()+1))
     plt.show()
```

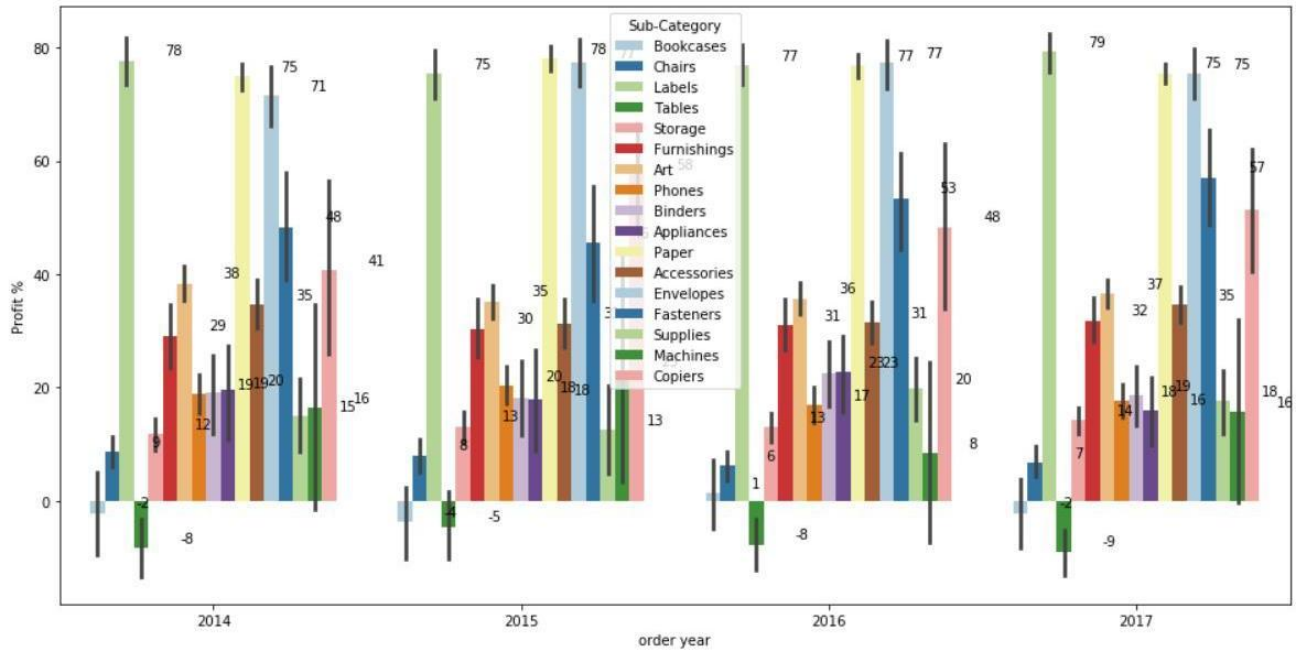
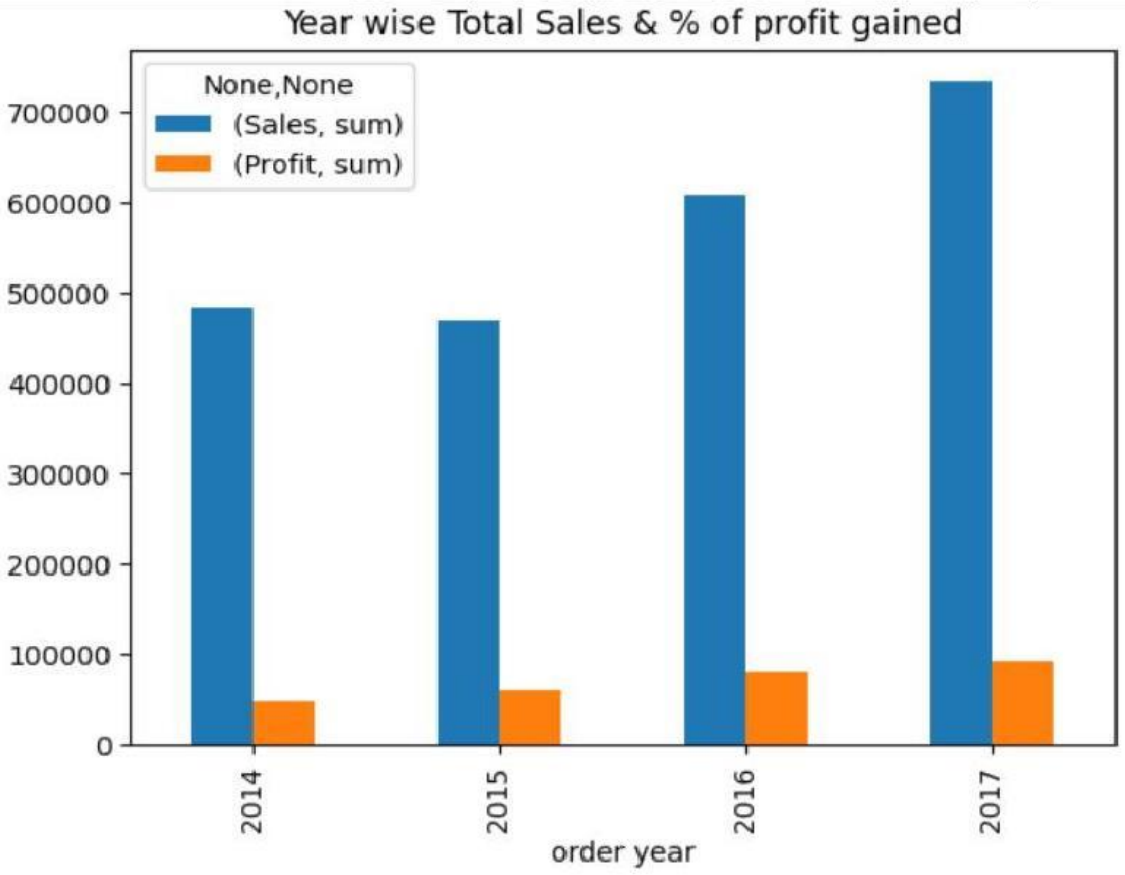


Figure 20: .ipynb code describing the bar plot

```
[42] #Sales per year
     df.groupby('order year')['Sales','Profit'].agg(['sum']).plot.bar()
     plt.title('Year wise Total Sales & % of profit gained')
```

```
Text(0.5, 1.0, 'Year wise Total Sales & % of profit gained')
```



Sales of the store has increased every year resulting in high profit margin by the end of 2017.

Figure 21: .ipynb code describing bar plot

3.CONCLUSION

Todays business must always strive to create the next best thing that consumers will want because consumers continue to desire their products,services etc. To continuously be better ,faster, and cheaper.In this world of new technology,businesses need to accommodate to the new types new types ofconsumers behaviour and trends because it will prove to be vital to their business success and survival. E-commers is continuously progressing and is becoming more and more important to businesses as technology continues to advance and is something that should be taken advantage of and implemented.

From the inception of the internet and e-commers,the possibilities have become endless for both businesses and consumers. Creating more opportunities for profit and advancements for businesses,while creating more option for consumers.however,just like anything else,e-commerce have its disadvantages including consumer uncertainties,but nothing that can not be resolved or avoided by good decision- making and business practices.

There are several factors and variable that need to be considered and decided upon when starting an e- commerce business. Some of these include : types of e- commerce, marketing strategies ,and countless more. If the correct methods and practices are followed, a business will prosper in an e-commerce settingwith much success and profitability.

4.APPLICATION

Data analytical is used in tracking customer's behavior towards product or service. You can use it to identify why sales are low,what product people buy,why they are buying them,how much they are spending on these products,how you can sell your product better,and many other queries.

HELPS TO BUILD A REBUST SUPPLY CHAIN

Customers shop online for two reasons – convenience and better prices. This is why your supply chain needs to be robust. Are the products being advertised on your site available in adequate quantities in your inventory? If not, it will result in slower delivery and the level of customer satisfaction is bound to drop and the customer lifetime value will be negatively affected. On the flip side, having too many units in stock will take up space in your inventory and increase your costs. Striking this balance, between supply and demand is tricky, and this is why using analytics to forecast future sales is vital.

- Analyze information to detect fraud.
- Predict what's in store for you.
- Measure your marketing
- Personalize the customer's shopping experience

5.ADVANTAGES AND DISADVANTAGES

Advantages	
...for the customer	**...for the provider**
• Flexible shopping hours (7·24h) • No waiting queues (if net is available and software appropriately designed) • Shopping at home (we don't have to leave our apartment, refuel our car or buy a subway ticket, look for a parking place, etc.) • Individual needs can be covered (if customization is offered) • Global offers, more competition, pressure on prices	• Better customer service can be offered • Fast communication with customer • New customer potential through global visibility • No (traditional) intermediaries, who take away margins

Disadvantages	
...for the customer	**... for the provider**
• Security risks: o Data theft (e.g. stealing account or credit card numbers) o Identity theft (acting under our name or user identity) o Abuse (e.g. third person orders goods with our identity, gets them delivered and we have to pay for it) • Crime: o Bogus firm (firm does not really exist) o Fraud (e.g. order is confirmed, invoice has to be paid, but goods are never delivered) • Uncertain legal status (if something goes wrong, can we accuse the provider?)	• Higher logistics cost (goods have to be sent to the customer's location) • Anonymity of customers (how to make targeted advertisements?)

6. FUTURE SCOPE

In 2020, the e-commerce sector was at an estimated US$ 50 billion. The e-commerce business is responsible for driving 1.2 million transactions every day, according to NASSCOM. By 2023, this sector is expected to pass the US and become the 2nd largest retail market. These sectors own a 5.9% share in this industry and have used data analytics to predict trends, provide better customer service, and streamline their warehouse operations.

Data analytics and data analytics bootcamp is a huge industry and is predicted to keep growing. It is expected to touch US$11.87 billion by 2026 as it keeps growing at a steady pace. This industry will disrupt the market, causing a great shift in it and bringing several job opportunities with it.

7.BIBILOGRAPHY

1) Dr. C. s. Ryudu, “E-commerce & e-commerce.”Himalaya Publishing House.
2) D.S. Yadhav, “Foundation of Information Technology.” New Age International Publishers,New Delhi.
3) Harini liat and Dalit Tzafrur.”electronic commerce.”
4) C.B.Memoria,”Personnel management .” himalaya publishing,1988.
5) Chatterjee N.N “Management of principle in India enterprises.”Agency culcutta,1980.
6) C.B. Memoria:”Industrial Organization,” Jain Brothers ,Jodhapur,1977.

7)Parag Diwan and Sunil sharma: ”Electronic commerce A managers guide to e-business”vanity books international,new Delhi.

8)Timmers .P.” Electronic commerce: Strategies and model for Business to Business Trading “NewYark:John Wiley and sons inc.2000.

9)Devis and keith A : “Human Behaviour at work : organization behavior”M.C.Grew Hill Publshing1989.

10)Claude .S ,George : “The History of Management thought” Prentice hall of India pvt Itd.New Delhi,1974.

11)Kamlesh k Bajaj & Debjani Nag : “E-commerce: the Cutting Edge of business.” Tata mcgrewHill,New Delhi,2000.

12)Bhaskar B.: “electronic commerce:Framework Technologies and applications”New Delhi ,Tata McGraw Hill,2023.

13)korper. S. and Eliis, J.: “E-commerce Book,” san Diego : Academic press.2001.

14)Bhaskar B.: “ electronic consumer.Framework Technologies and Applications” New Delhi : mcGraw Hill 2003.

8. HELP FILE

The project is OTT Platform Analysis Tools, so to build the project we need to follow the following steps:

STEP-1:

• Collect the datasets for the project from the online website (Kaggle etc).

STEP 2:

• Create an IBM Academic Initiative Account and IBM Cognos account.

• (https://www.youtube.com/watch?v=x6143M7BAqE-Referral video for creating IBM Academic Initiative Account)

STEP 3:

• Open IBM Cognos Analytics in Google and log in with your IBM Academic Initiative Account.

• Then click on Dashboard Creation and upload your datasets.

STEP 4:

• Select a Dashboard design for dashboard creation.

• Now by using your datasets and visualizations design as shown select a visualization to display your dataset in a pictorial form.

STEP 5:

• Same as previous step you can create as many visualization charts as you wish or as you want for your project.

STEP 6:

• And then you can add any background colors or any changes you need.

• Now the project is finished.