IBM PROJECT

GLOBAL SALES DATA ANALYTICS

Batch: B8-2A4E

Team ID: PNT2022TMID21528

Team Leader: NAVIN KUMAR V

Team Members:

ASHWIN D

NAVEEN E

BHARAT SRINIVAS R

CONTENTS

l'itle	Page Number
1. **INTRODUCTION**	4
a. Project Overview	4
b. Purpose	4
2. **LITERATURE SURVEY**	4
a. Existing problem	4
b. References	4
c. Problem Statement Definition	4
3. **IDEATION & PROPOSED SOLUTION**	5
a. Empathy Map Canvas	5
b. Ideation & Brainstorming	6
c. Proposed Solution	9
d. Problem Solution fit	10
4. **REQUIREMENT ANALYSIS**	11
a. Functional requirement	11
b. Non-Functional requirements	11
5. **PROJECT DESIGN**	12
a. Data Flow Diagíams	12
b. Solution & Technical Architecture	12
c. User Stories	13
6. **PROJECT PLANNING & SCHEDULING**	14
a. Sprint Planning & Estimation	14
b. Sprint Delivery Schedule	14
c. Burndown chart	15

7. **CODING & SOLUTIONING**	14
a. Dashboard	14
b. Python code	19
8. **TESTING**	23
a. User Acceptance Testing	23
9. **RESULTS**	23
a. Performance Metrics	23
10. **ADVANTAGES & DISADVANTAGES**	24
11. **CONCLUSION**	24
12. **FUTURE SCOPE**	25
13. **APPENDIX**	25
GitHub & Project Demo Link	25

1. INTRODUCTION

1.1 Project Overview:

This project helps understand customer analysis and product analysis of the super store. Segment wise customers are analyzed and the products most bought or liked by them is also analyzed to improve the sales performance of the store.

1.2 Purpose:

To meet the growing need of shopping online by analyzing the customers and also the product to make the process of buying things online easy.

2. LITERATURE SURVEY

2.1 Existing Problem:

The problem of not analyzing the sales performance leads to loss in business and loss of customers.

2.2 References:

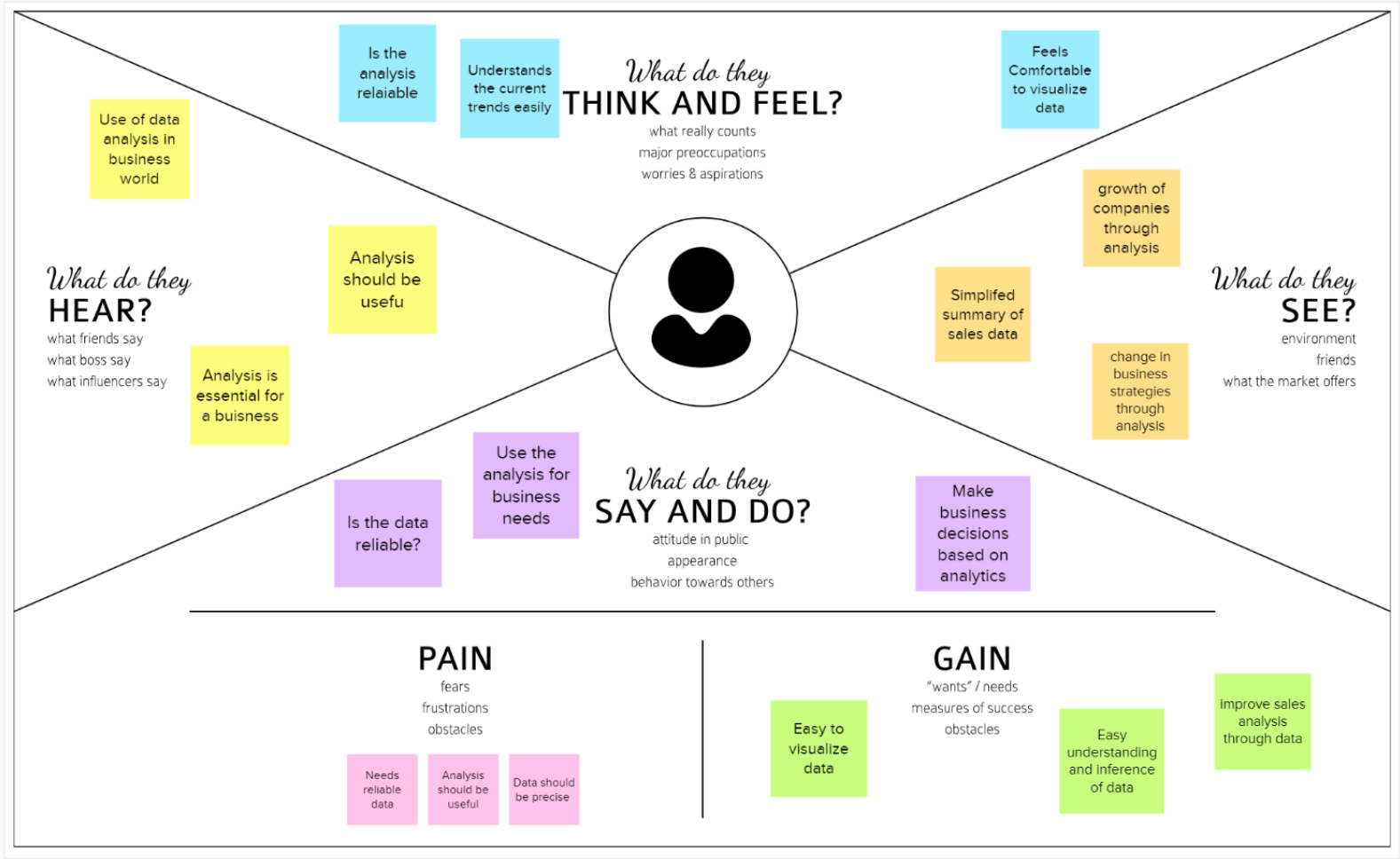
Salespeople Performance Evaluation with Predictive Analytics in B2B - Nelito Calixto and João Ferreira
Predictive Sales Pipeline Analytics - Junchi Yan123, Chao Zhang
Analytics as a Source of Business Innovation - Lorraine Eden

2.3Problem statement definition:

Shopping online is currently the need of the hour. Because of this COVID, it's not easy to walk in a store randomly and buy anything you want. So, understanding a few things like, Customer Analysis and Product Analysis of this Global Super Store is essential

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas:



3.2 Ideation & Brainstorming:

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

Shopping online is currently the need of the hour. Because of this COVID, it's not easy to walk in a store randomly and buy anything you want. Hence understanding things like, Customer Analysis and Product Analysis of this Global Super Store is essential.

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

Navin Kumar V

- Sales analytics refers to the technology and processes used to gather sales data and gauge sales performance.
- Analysis should focus on improvement and developing a strategy for improving your sales performance in both the short- and long-term.
- focus on sales and revenue performance, and a strong (or weak) performance can become a potent rallying force for your entire team.
- display analytics on a sales dashboard which is straight forward, intuitive, and communicate a clear message.
- Sales target evaluates current sales and compares them to your bigger, long-term goals.
- Another positive way to increase transparency and accountability for sales analytics process is to display a sales leaderboard.

Ashwin D

- Perform rapid analysis by comparing your current sales against the previous period, the same period the previous year, and get a sense of historical trends.
- The goal of sales analytics is always to simplify the information available to you.
- Sales analytics can enable your agents to spot key trends, dive deep, predict outcomes, and increase productivity.
- Sales growth shows how much your revenue increases (or decreases) over a specific period of time.
- using free tools like Google Sheets, PowerPoint, or Excel. Or it can be hooked up to a dashboard software solution like PowerMetrics.
- Many global, industry-leading brands are now using their sales data in ingenious ways to make better business decisions

Naveen E

- Sales analytics refers to the use of technology to collect and use sales data to derive actionable insights. It is used to identify, optimize, and forecast sales
- Cash flow may be the top focus of small business owners, but analyzing sales data is equally important
- Without tracking those sales as well as your in-store ones, you won't be able to identify the areas you should focus on.
- gathering, analyzing and leveraging sales data can be a complex task without the right expertise and tools in place
- the ones that aren't moving, your most profitable customers, and potential sales opportunities.
- By analyzing your sales, you can identify your most profitable products

Bharat Srinivas R

- Cash is the lifeblood of any business, and the way to generate cash is through sales.
- There's several benefits to using spreadsheets to gather data. They're free they provide a helpful real-time overview of your current sales operation
- AI gathers information conveyed by customers and the semantics of the conversation.NLP used to analysis users feedback
- The insights you glean from analyzing sales data can change the trajectory of your business, enabling you to take actions that improve your operations.
- analysis of data sets of retailers Store to determine the business drivers and predict which departments are affected by the different scenarios
- Retailers have to create effective promotions and offers to meet its sales and marketing goals

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

Users

Analysis should focus on improvement and developing a strategy for improving your sales performance in both the short- and long-term.

Without tracking those sales as well as your in-store ones, you won't be able to identify the areas you should focus on.

analysis of data sets of retailers Store to determine the business drivers and predict which departments are affected by the different scenarios

The insights you glean from analyzing sales data can change the trajectory of your business, enabling you to take actions that improve your operations.

Others

Retailers have to create effective promotions and offers to meet its sales and marketing goals

the ones that aren't moving, your most profitable customers, and potential sales opportunities.

Many global, industry-leading brands are now using their sales data in ingenious ways to make better business decisions

Perform rapid analysis by comparing your current sales against the previous period, the same period the previous year, and get a sense of historical trends.

Another positive way to increase transparency and accountability for sales analytics process is to display a sales leaderboard.

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

3.3 Proposed Solution:

S.No.	Parameter	Description
1	Problem Statement	Develop an efficient system & an application that can monitor and alert the users(workers)
2	Idea / Solution description	This product helps the industries in monitoring the emission of harmful gasesIn several areas, the gas sensors will be integrated to monitor the gas leakage.If in any area gas leakage is detected the admins will be notified along with the locationIn the web application, admins can view the sensor parameters.
3	Novelty / Uniqueness	Fastest alerts to the workersUser friendly
4	Social Impact / Customer Satisfaction	Cost efficientEasy installation and provide efficient resultsCan work with irrespective of fear
5	Business Model (Revenue Model)	The product is advertised all over the platforms. Since it is economical, it even helps small scale industries from disasters.As the product usage can be understood by everyone, it is easy for them to use it properly for their safest organization.
6	Scalability of the Solution	Since the product is cost-efficient, it can be placed in many places in the industry. Even when the gas leakage is more, the product senses the accurate values and alerts the workers effectively.

3.4 Problem Solution Fit:

4. REQUIREMENT ANALYSIS

4.1 Functional Requirement:

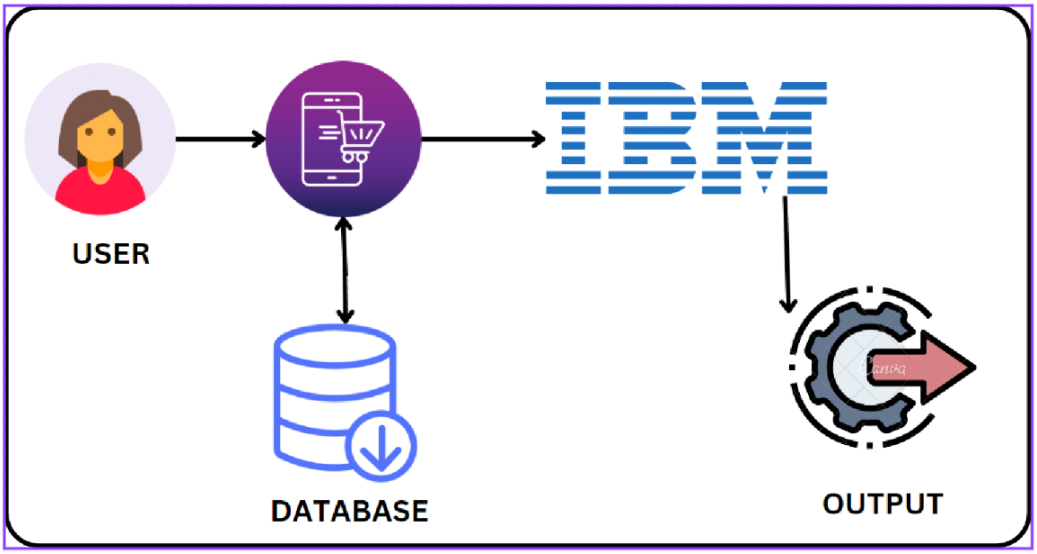
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Google account Registration through user details
FR-2	User Confirmation	Confirmation through one to one google authentication, OTP.
FR-3	User Login	Login through Google account
FR-4	User uploading data(administrative)	To store the data set through the cloud
FR-5	End user benefits	Getting higher state of efficiency and also to know entire data analysis

4.2 Non-Functional Requirement:

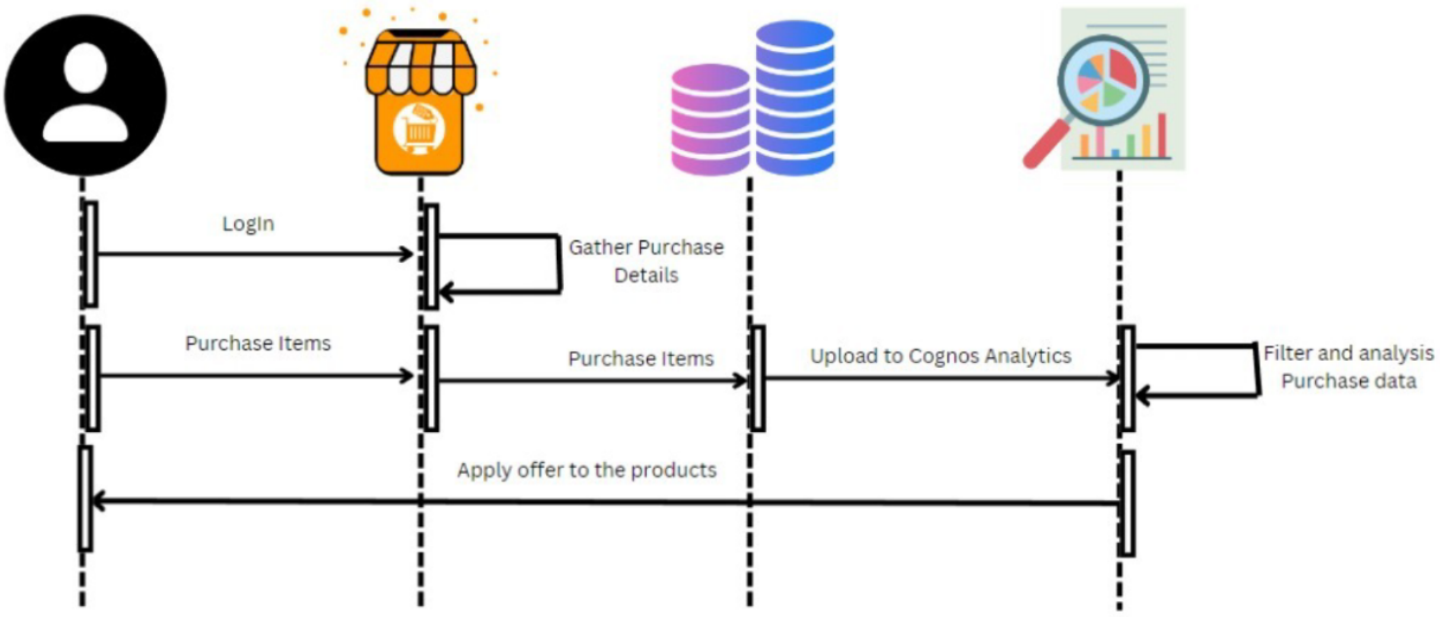
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Easily accessible to all users
NFR-2	Security	Since it has one to one authentication this secured.
NFR-3	Reliability	It is highly reliable
NFR-4	Performance	The performance rate and efficiency rate is high.
NFR-5	Availability	It is available in all platforms 24/7
NFR-6	Scalability	The ability of a hardware and software parallel system to exploit increasing computing resources efficiency in the analysis of the large datasets

5. PROJECT DESIGN

5.1 Data Flow Diagrams:



5.2 Solution & Technical Architecture:



5.3 User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Business owner	Online registration	USN-1	As a business owner, I want to login to my account.	Input data fields to enter: 1.Username/email 2.Password 3.Re-enter password 4.Security question 5.Security answer	High	Sprint-1
	Data upload	USN-2	As a business owner, I want to upload my sales data to perform analytics.	Submission of excel file containing the sales data.	High	Sprint-1
	Improve performance	USN-3	As a business owner, I want to use the analytics results to make my business performance better	Reflection of the analytics results to my online store.	High	Sprint-2
Customer (Buyer)	Registration	USN-1	As a buyer, I want to login to my account.	Input data fields to enter: 1.Username/email 2.Password 3.Re-enter password 4.Security question 5.Security answer	High	Sprint-3
	Buy	USN-2	As a buyer, I want to buy products from the online store.	Search for the items to buy in the application	Medium	Sprint-3
Analytics team administrator	Analysis of sales data	USN-1	As an administrator, I want to analyze the sales data for better performance of the store.	Get the sales data form the business owner	High	Sprint-4

6. PROJECT PLANNING AND SCHEDULING

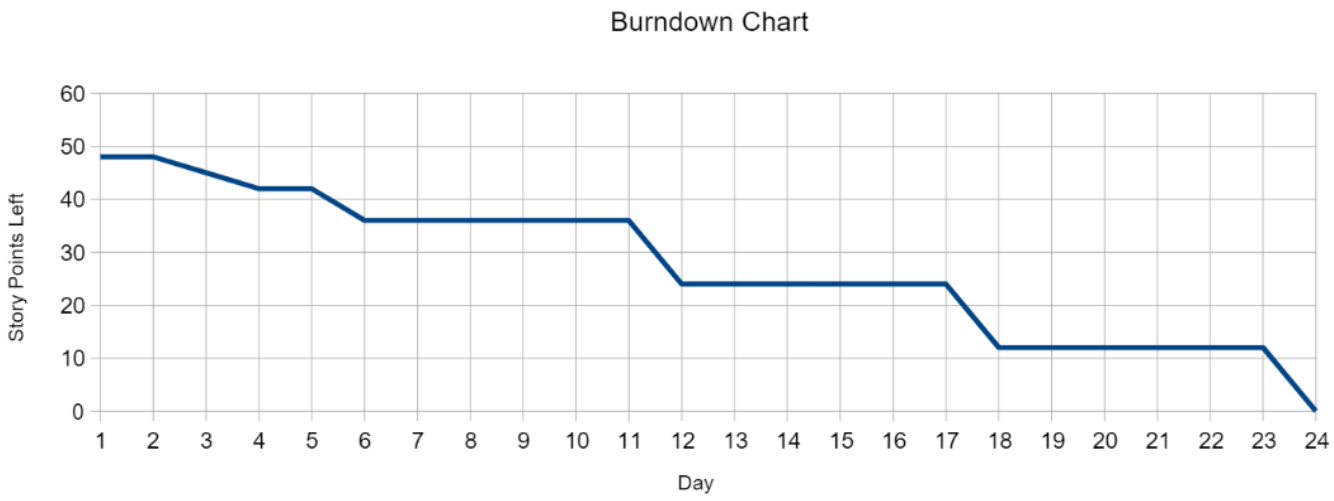
6.1 Sprint Planning & Estimation:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Collection of Data	USN-1	Collect the dataset from kaggle	2	High	Navin Kumar Ashwin D Naveen E Bharat Srinivas
Sprint-2	Data Cleaning	USN-2	Importing the required libraries and Loading Data Cleaning and preparation of dataset	1	Medium	Navin Kumar Ashwin D Naveen E Bharat Srinivas
Sprint-3	Data Analysis and Visualization	USN-3	Analysis of data using different graph and find the trends and relation between the data	2	High	Navin Kumar Ashwin D Naveen E Bharat Srinivas
Sprint-4	Report Building	USN-4	Building report summarizing the data in the dataset with the Dashboard	2	Medium	Navin Kumar Ashwin D Naveen E Bharat Srinivas

6.2 Sprint Delivery Schedule:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	27 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	03 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	11 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	15 Nov 2022

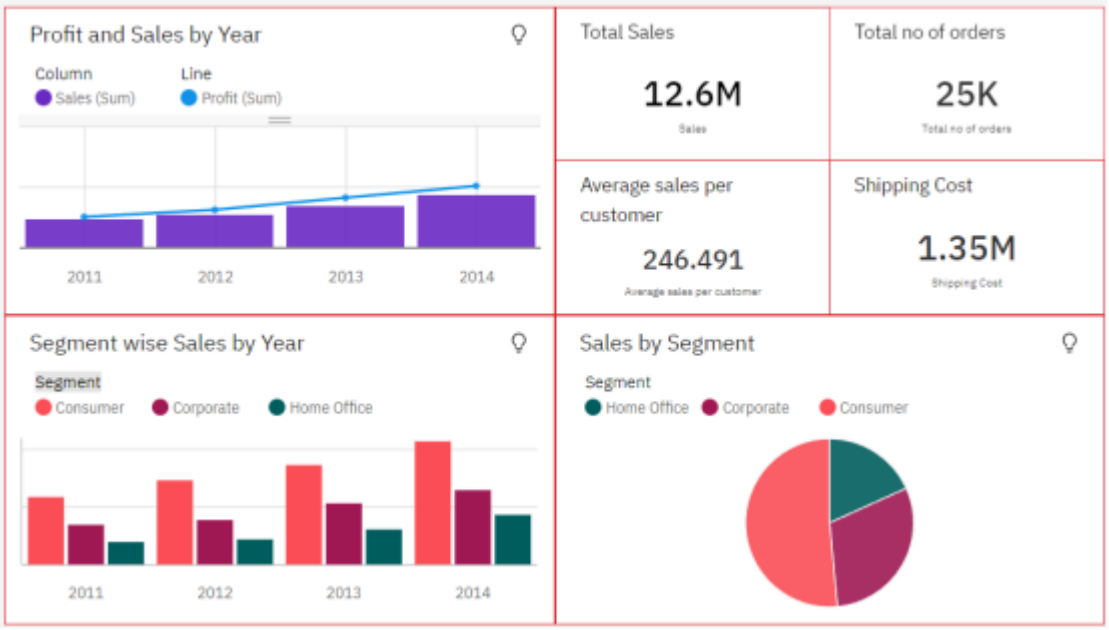
6.3 **Burndown chart:**



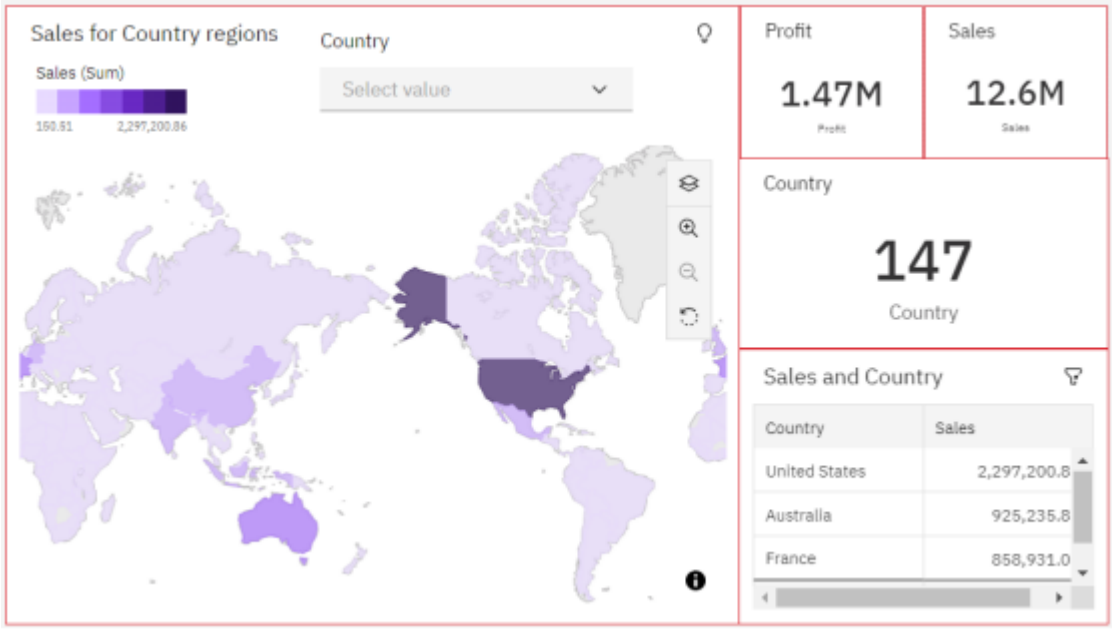
7. CODING AND SOLUTIONING

7.1 Dashboard

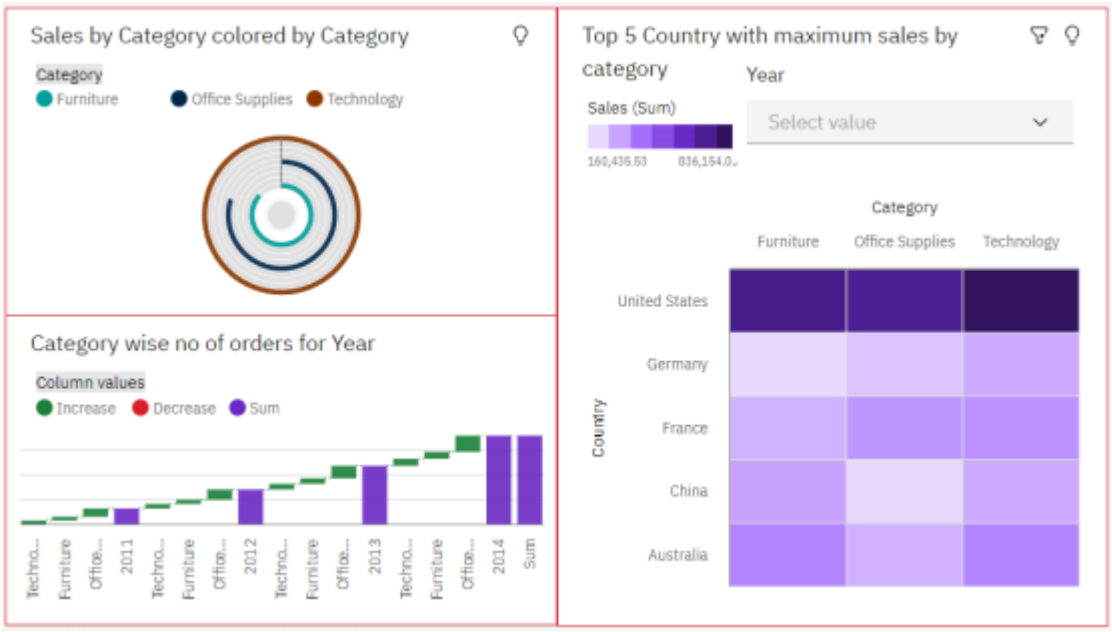
Profit and Sales Analysis



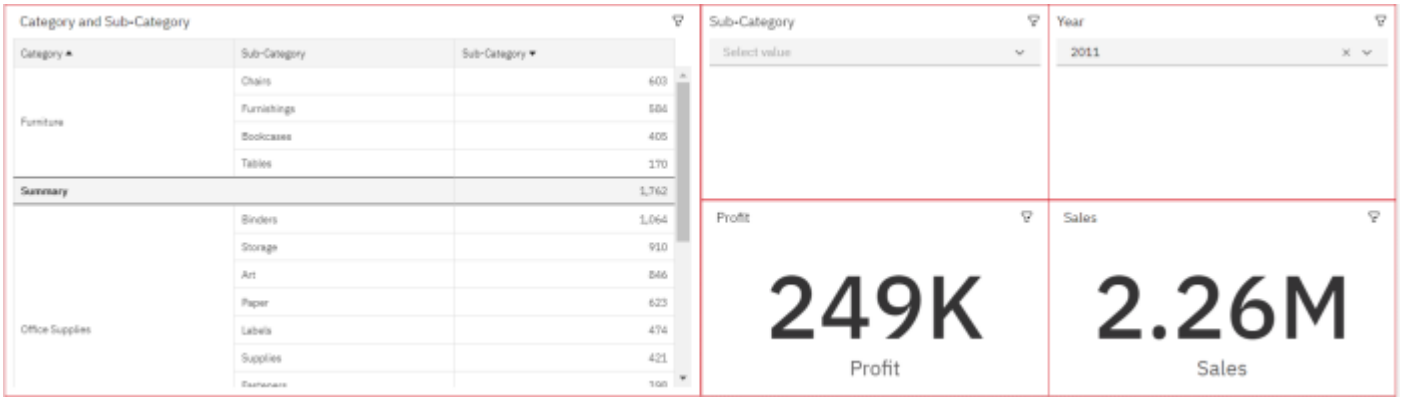
Country Wise Analysis

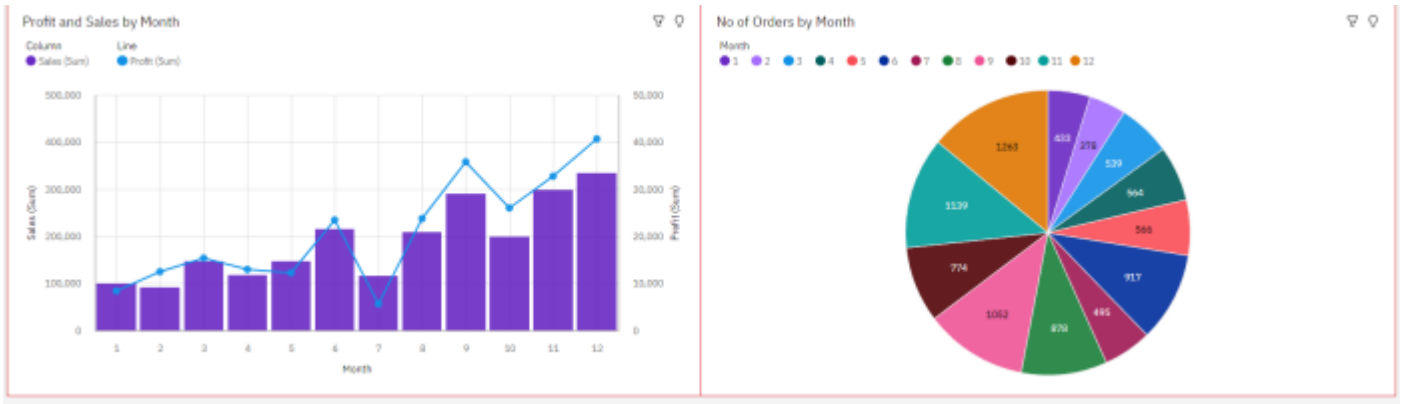


Category Wise analysis

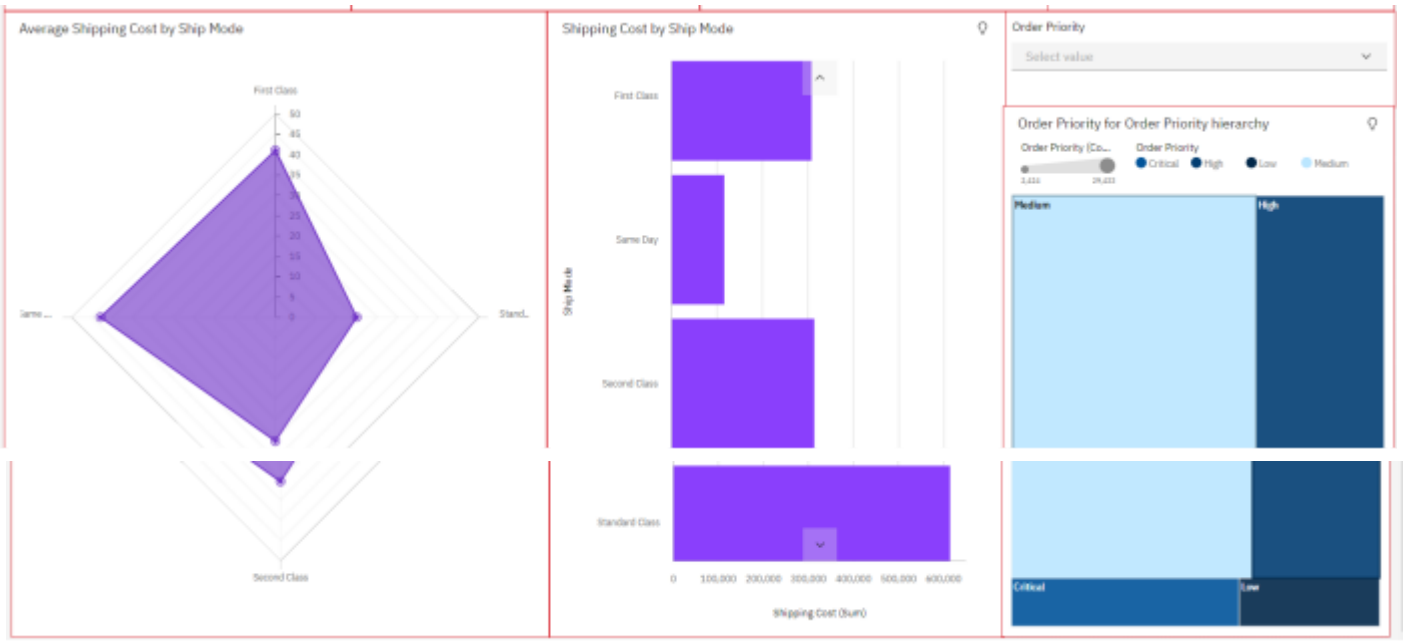


Sub category Wise analysis





Order Priority Wise Analysis



7.2 Python Code

```
In [2]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [3]: df = pd.read_csv('Global_Superstore2.csv',encoding='latin-1')
```

```
In [5]: df.head()
```

Out[5]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	...	Product ID	Category	Sub-Category	Product Name	Sales
0	32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	...	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.650
1	26341	IN-2013-77878	05-02-2013	07-02-2013	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wales	...	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black	3709.395
2	25330	IN-2013-71249	17-10-2013	18-10-2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	...	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID	5175.171
3	13524	ES-2013-1579342	28-01-2013	30-01-2013	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	...	TEC-PH-10004583	Technology	Phones	Motorola Smart Phone, Cordless	2892.510
4	47221	SG-2013-4320	05-11-2013	06-11-2013	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	...	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed	2832.960

5 rows × 24 columns

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51290 entries, 0 to 51289
Data columns (total 24 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Row ID          51290 non-null  int64  
 1   Order ID        51290 non-null  object 
 2   Order Date      51290 non-null  object 
 3   Ship Date       51290 non-null  object 
 4   Ship Mode       51290 non-null  object 
 5   Customer ID     51290 non-null  object 
 6   Customer Name   51290 non-null  object 
 7   Segment         51290 non-null  object 
 8   City            51290 non-null  object 
 9   State           51290 non-null  object 
 10  Country         51290 non-null  object 
 11  Postal Code     9994 non-null   float64
 12  Market          51290 non-null  object 
 13  Region          51290 non-null  object 
 14  Product ID      51290 non-null  object 
 15  Category        51290 non-null  object 
 16  Sub-Category    51290 non-null  object 
 17  Product Name    51290 non-null  object 
 18  Sales           51290 non-null  float64
 19  Quantity        51290 non-null  int64  
 20  Discount        51290 non-null  float64
 21  Profit          51290 non-null  float64
 22  Shipping Cost   51290 non-null  float64
 23  Order Priority  51290 non-null  object 
dtypes: float64(5), int64(2), object(17)
memory usage: 9.4+ MB
```

In [7]:
```
df.shape
```

Out[7]: (51290, 24)

In [8]:
```
df.describe()
```

Out[8]:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit	Shipping Cost
count	51290.00000	9994.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	25645.50000	55190.379428	246.490581	3.476545	0.142908	28.610982	26.375915
std	14806.29199	32063.693350	487.565361	2.278766	0.212280	174.340972	57.296804
min	1.00000	1040.000000	0.444000	1.000000	0.000000	-6599.978000	0.000000
25%	12823.25000	23223.000000	30.758625	2.000000	0.000000	0.000000	2.610000
50%	25645.50000	56430.500000	85.053000	3.000000	0.000000	9.240000	7.790000
75%	38467.75000	90008.000000	251.053200	5.000000	0.200000	36.810000	24.450000
max	51290.00000	99301.000000	22638.480000	14.000000	0.850000	8399.976000	933.570000

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
Row ID            0
Order ID          0
Order Date        0
Ship Date         0
Ship Mode         0
Customer ID       0
Customer Name     0
Segment           0
City              0
State             0
Country           0
Postal Code       41296
Market            0
Region            0
Product ID        0
Category          0
Sub-Category      0
Product Name      0
Sales             0
Quantity          0
Discount          0
Profit            0
Shipping Cost     0
Order Priority    0
dtype: int64
```

In [11]:
```
df.nunique()
```

Out[11]:
```
Row ID            51290
Order ID          25035
Order Date         1430
Ship Date          1464
Ship Mode             4
Customer ID        1590
Customer Name       795
Segment               3
City               3636
State              1094
Country             147
Postal Code         631
Market                7
Region               13
Product ID        10292
Category              3
Sub-Category         17
Product Name       3788
Sales             22995
Quantity             14
Discount             27
Profit            24575
Shipping Cost     10037
Order Priority        4
dtype: int64
```

In [12]:
```
df_customer = df[['Customer ID','Order ID','Order Date', 'Ship Date', 'Ship Mode','Country']]
df_customer.count()
```

Out[12]:
```
Customer ID    51290
Order ID       51290
Order Date     51290
Ship Date      51290
Ship Mode      51290
Country        51290
dtype: int64
```

In [16]:
```
df.drop(['Row ID','Postal Code'],axis=1,inplace=True)
```

In [17]:
```
df.corr()
```

Out[17]:

	Sales	Quantity	Discount	Profit	Shipping Cost
Sales	1.000000	0.313577	-0.086722	0.484918	0.768073
Quantity	0.313577	1.000000	-0.019875	0.104365	0.272649
Discount	-0.086722	-0.019875	1.000000	-0.316490	-0.079056
Profit	0.484918	0.104365	-0.316490	1.000000	0.354441
Shipping Cost	0.768073	0.272649	-0.079056	0.354441	1.000000

In [18]:
```
sns.heatmap(df.corr(),annot= True,fmt='.0%')
```

Out[18]: <AxesSubplot:>

8. TESTING

8.1 User Acceptance Testing:

https://github.com/IBM-EPBL/IBM-Project-27778-1660065423/blob/main/Project%20Development%20Phase/Testing/UAT%20Report.pdf

9. RESULTS

9.1 Performance Testing :

https://github.com/IBM-EPBL/IBM-Project-27778-1660065423/blob/main/Project%20Development%20Phase/Testing/Performance%20Testing.pdf

10. ADVANTAGES AND DISADVANTAGES

Advantages:

- **Analyze segment wise sales**
- **Analyze regional sales and profit forecast**
- **Predict sales performance**
- **Gain customers**
- **Boost sales performance**
- **Boost revenue**

Disadvantages:

- **High technical knowledge is required to perform analytics.**
- **Privacy is compromised by sharing the sales data with the third party.**

11. CONCLUSION

Sales analytics is one of the crucial methods used by the stores to analyze and boost sales performance. WIth the growing technological field, equipping the business with the latest trend by gaining customers and at the same time not losing the existing customers is essential. Analytics helps achieve that goal,

12. FUTURE SCOPE

- **Data Analytics plays a major role in the Financial and Business World. Data in the future provides more information about the Sales which then can be interpreted and can draw many useful conclusions.**
- **After this pandemic people have practiced using Online shopping and the rates has increased exponentially so products and shipment can be delivered fast using these analytics.**
- **The Dashboard can be improved a lot with much data to handle and a lot of useful information and inferences can be made in the future which enhances the sales and contributes to the Financial World.**

13. APPENDIX

GitHub link - https://github.com/IBM-EPBL/IBM-Project-27778-1660065423

Project Demo Link - https://drive.google.com/file/d/1_Dn43qRF6Vqis1RA71hZIRs5czUSGrZT/view?usp=share_link