GLOBAL SALES DATA ANALYTICS

PROJECT REPORT

Submitted By

Team ID : PNT2022TMID49375

Team Leader : NAVEENKUMAR M

Team Member 1 : SHRISANJAYKUMAR KU

Team Member 2 : SIVABALAN P

Team Member 3 : ABISHEK V

In Partial fulfilment for the award of the degree
of
BACHLOR OF ENGINEERING
in
computer science and engineering

GOVERNMENT COLLEGE OF ENGINEERING
BODINAYAKANNUR-625582

ANNA UNIVERSITY CHENNAI-600025

List Of Contents

6. **PROJECTPLANNING & SCHEDULING**
    a. Sprint Planning & Estimation
    b. Sprint Delivery Schedule
    c. Reports from JIRA
7. **RESULTS**
    a. Performance Metrics
8. **ADVANTAGES & DISADVANTAGES**
9. **CONCLUSION**
10. **FUTURE SCOPE**

GitHub project Demo Link Global Sales Data Analytics

1. **Indroduction:**

    a. **Project Overview:**

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, try to understand a few things like, Customer Analysis and Product Analysis of this Global Super Store.

    b. **Purpose:**

By the end of this Project, you will:

i. Know fundamental concepts and can work on IBM Cognos Analytics.

ii. Gain a broad understanding of plotting different visualizations to providea suitable solution.

iii. Able to create meaningful Visualizations and Dashboard(s).

2.LITRATURE SURVEY

2.1Existing Problem:

Crafting a good sales pitch from sales data analysis can be difficult. Getting the right data, hitting the right client pain points, crystallizing why your services are better than the competitors, all takes hard work. One of the best ways we've found to build a good sales pitch is to use data you already have.

In the digital world, there is no shortage of data, which translates into no shortage of potential competitive insights and advantages. With databases, datawarehouses, corporateintranets, best practicesharing, web analytics, voice of the customer information, and QA or Six Sigma data, you are well-poised for discovering good information.

2.2References:

1. Han Jiawei,Micheline Kamber and Jian Pei, "**Data Mining Concepts and Techniques**" in , MK Publications, 2009.Show in ContextGoogle Scholar

2. M. Tennekes and E. de Jonge, "**Top-down Data Analysiswith Treemaps**" Proceedings of the International Conference on Information Visualization Theory andApplications (IVAPP' 11), pp. 236-241, March 2011. Show in Context Google Scholar

3. P. Hoek, **"Parallel Arc Diagrams: Visualizing Temporal Interactions"**,Journal of Social Structure, vol. 12, 2011. Show in ContextGoogle Scholar

2.3Problem Statement Definition:

Our goal is to design and create a Dashboard using the Superstore Sales data (which is really close to reality)to provide answers to following questions

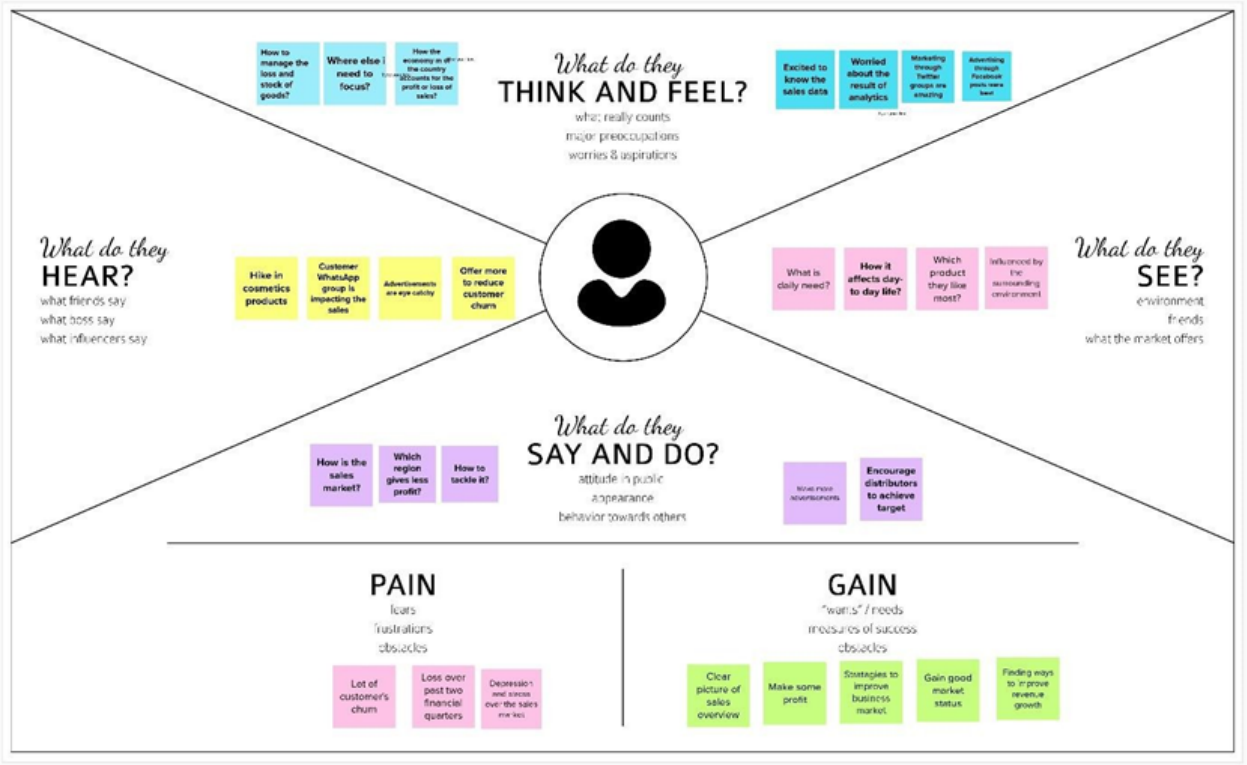
1. What are the performance indicators values for the past month? It's necessary for stock taking and comparing it against the same period last year.

2. What key factors do affect profit growth?

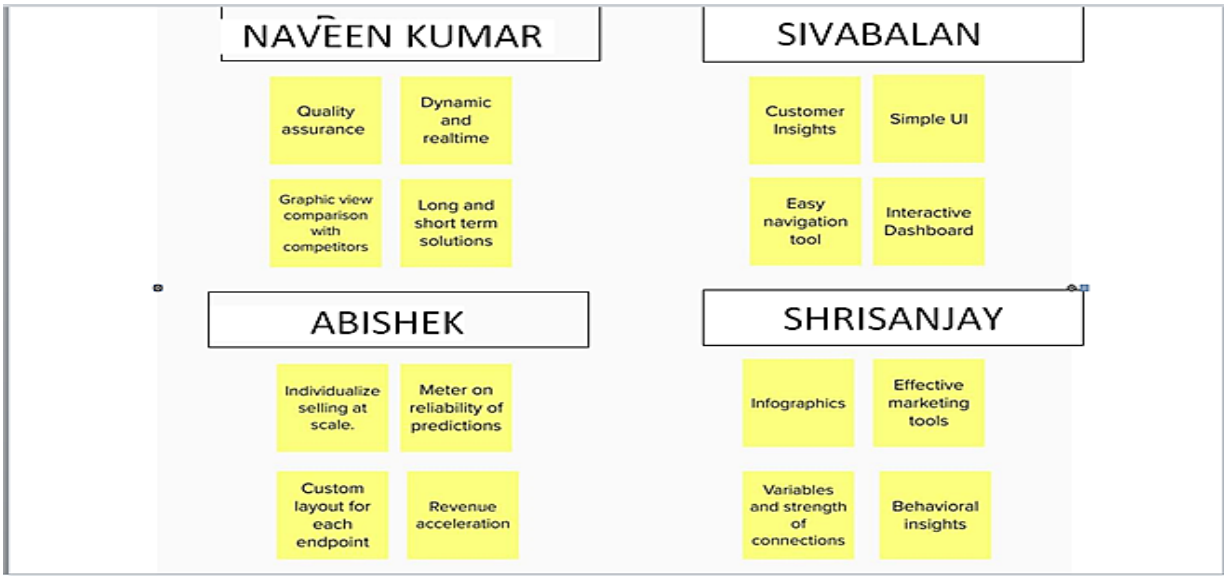
3. What categories, sub categories, products and clients generate more profits, and what ones that bring losses?

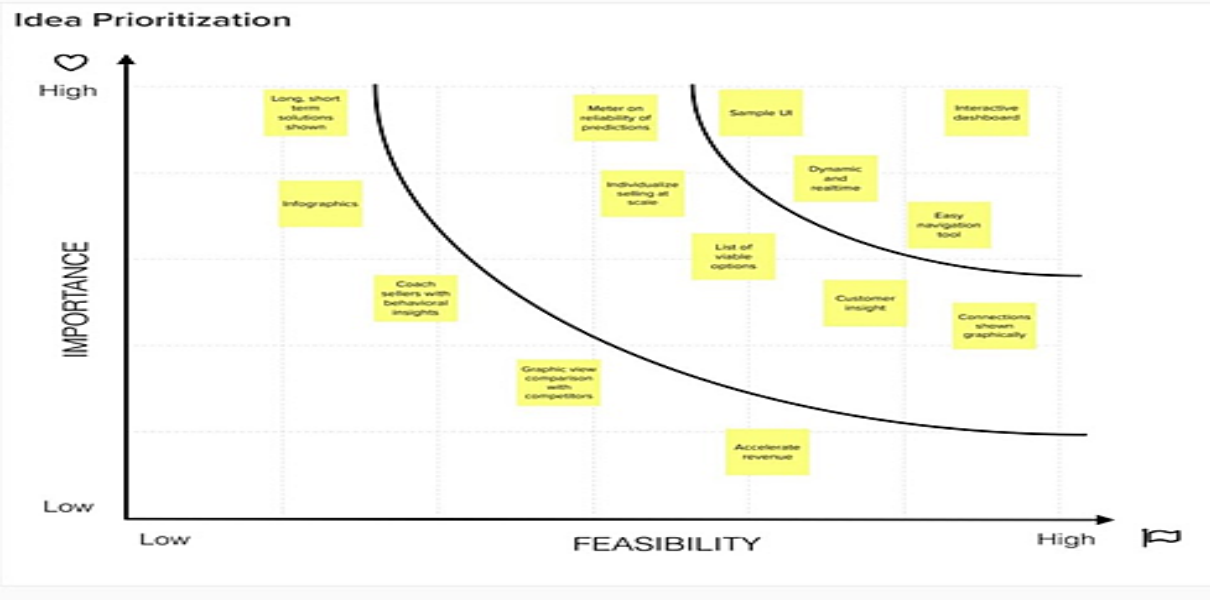
3.IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING

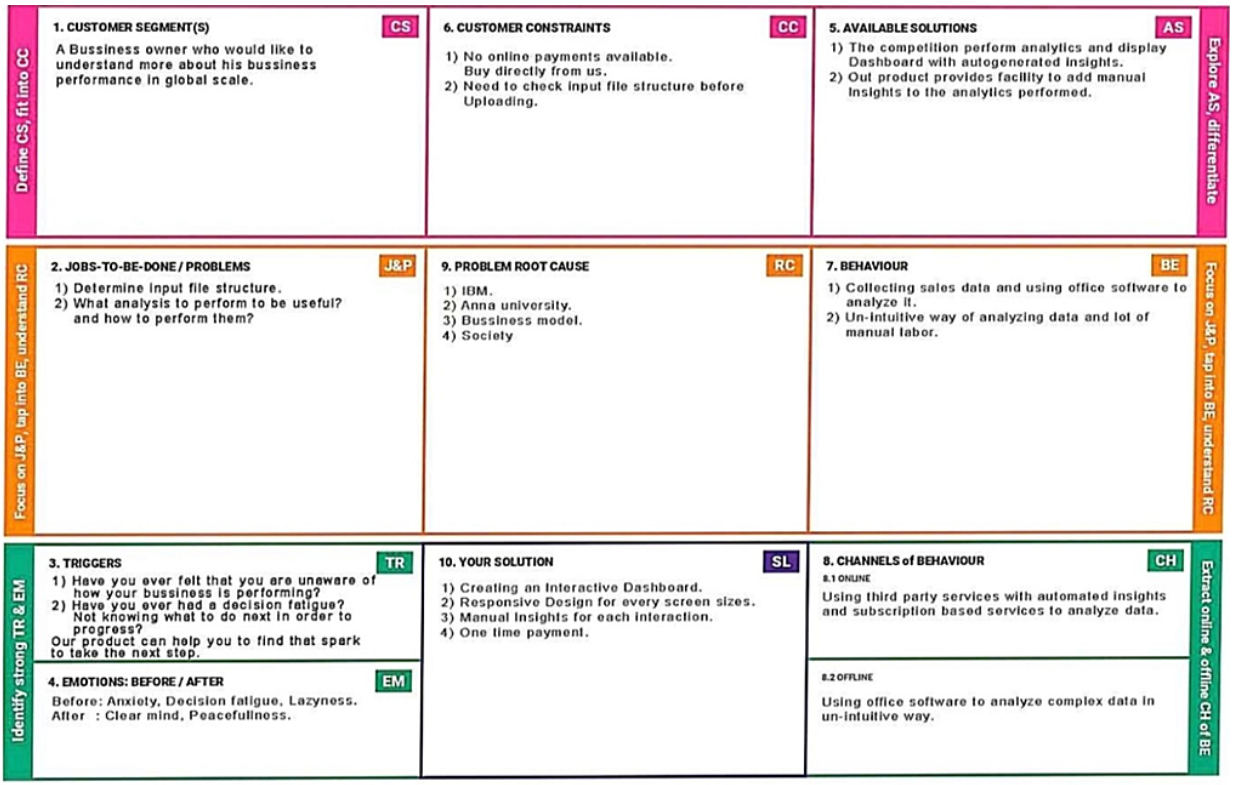




3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	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, try to understand a few things like, Customer Analysis and Product Analysis of this Global Super Store.
2.	Idea / Solution description	The described solution is by using IBM cognos we can display all the records and previous year global sales of product names, category and sub category as a graphical representation.
3.	Novelty / Uniqueness	we are going to provide discounts to the customers to increase the sales by providing free door step delivery of products to customers.
4.	Social Impact / Customer Satisfaction	Customer should know the available products and nearest location of the shops which gives the idea to customer for purchase.
5.	Business Model (Revenue Model)	This method focuses on the actual sales numbers from the customers. This helps to determine which products are top performers and multiplying the shop and increasing the product quantity.

3.4 PROBLEM SOLUTION FIT



4.REQUIREMENT ANALYSIS

PROJECT DESIGN PHASE – II

Solution Requirements (Functional & Non-functional)

Date	11-10-2022
Team ID	PNT2022TMID49375
Project Name	Project- Global Sales Data Analytics
Maximum Mark	4 Marks

Functional Requirements:

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

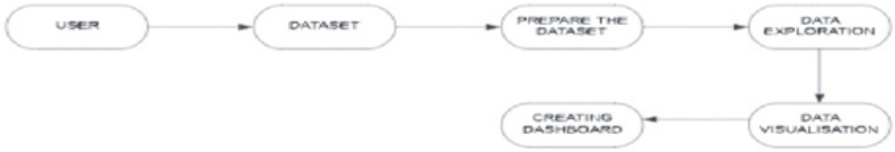
Non-functional Requirements:

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

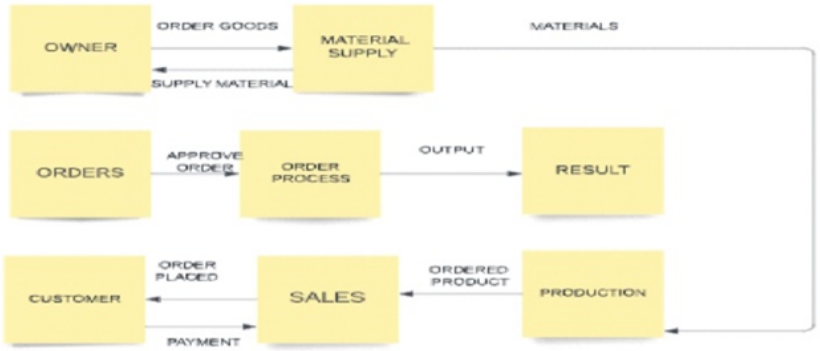
5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

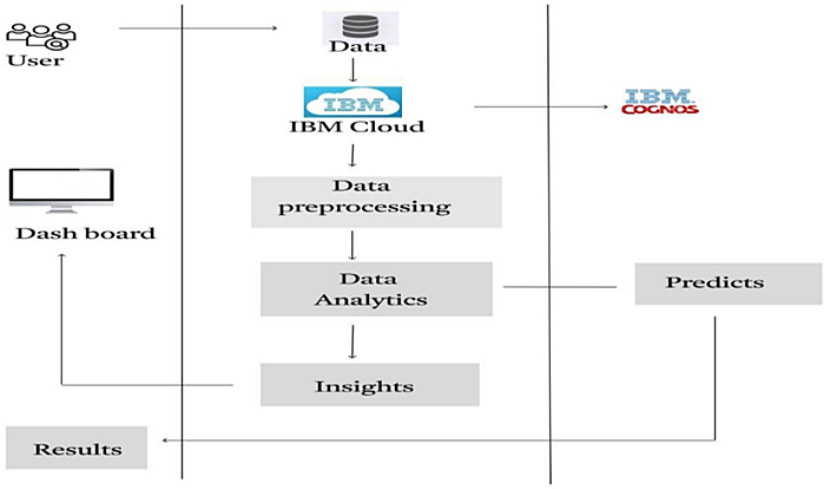
Simple Data flow diagram:



Structural flow diagram:



5.2 SOLUTION & TECHNICAL ARCHITCTURE



5.3 USER STORIES

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 tologin to my account.	Input data fields toenter: 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 filecontaining the sales data.	High	Sprint-1
	Improve performance	USN-3	As a business owner, I want touse the analytics results to make my business performance better	Reflection of the analytics results to myonline store.	High	Sprint-2
Customer (Buyer)	Registration	USN-1	As a buyer, I want to login tomy account	Input data fields toenter: 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 form the online store	Search for the items tobuy in the application	Medium	Sprint-3
Analytics team administrator	Analysis ofsales data	USN-1	As an administrator, Iwant to analysethe sales data for better performance of the store.	Get the sales data formthe business owner.	High	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Naveen Sivabalan Shrisanjay Abishek
		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	Low	
		USN-3	As a user, I will log in to the desired application using login credentials.	1	Medium	
Sprint-2	Pre processing	USN-4	As a user, I can do the data cleaning process.	2	High	Naveen Sivabalan Shrisanjay Abishek
		USN-5	As a user, I can perform Extract, Transform Load (ETL) process.	2	High	
Sprint-3	Dashboard	USN-6	As a user, I can upload the data of global sales for analysis.	1	Medium	

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
	Dashboard	USN-7	As a user, I can analyse the data by performing calculations and executing several visualization charts.	2	High	Naveen Sivabalan Shrisanjay Abishek
		USN-8	As a user, I can gain insights of the data for business analysis	2	High	
		USN-9	As a user, I can get the information for business analysis.	1	Medium	
Sprint-4	Report, Story and customer care	USN-10	As a user, I can generate report for the customer or sales analyst for knowing the insights about the sales.	2	Medium	Naveen Sivabalan Shrisanjay Abishek
		USN-11	As a user, I can clear queries of customers from the analysis of the sales.	1	Medium	
		USN-12	As a user, I can modify report according to the information gathered after analysis.	1	Low	

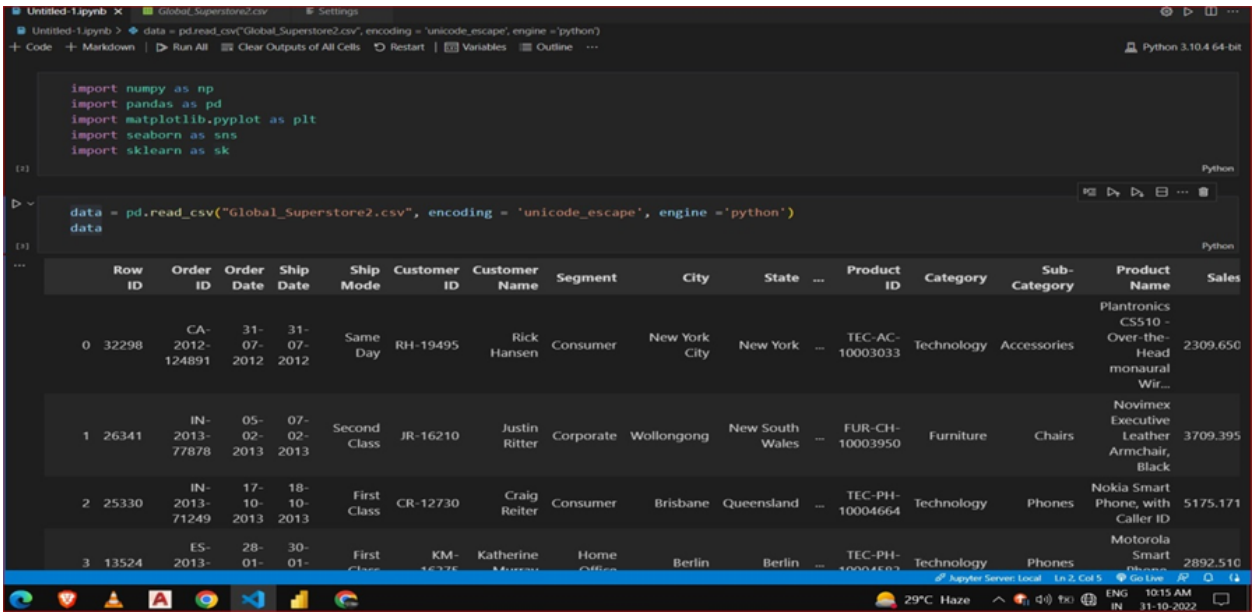
6.2 SPRINT DELIVERY SCHEDULE

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	4	6 Days	04 Nov 2022	10 Nov 2022	4	10 Nov 2022
Sprint-2	4	6 Days	05 Nov 2022	11 Nov 2022	4	11 Nov 2022
Sprint-3	6	6 Days	06 Nov 2022	12 Nov 2022	6	12 Nov 2022

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-4	4	6 Days	07 Nov 2022	13 Nov 2022	4	13 Nov 2022

6.3 REPORTS FROM JIRA

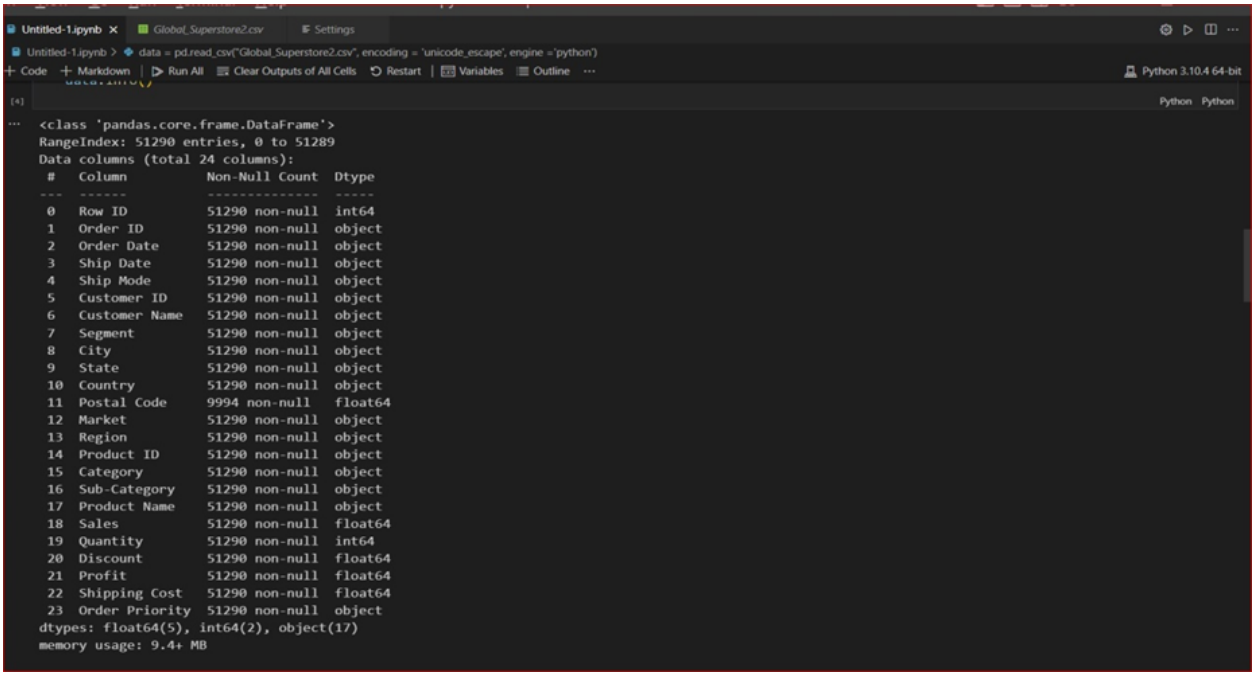
SPRINT 1:



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn as sk
```

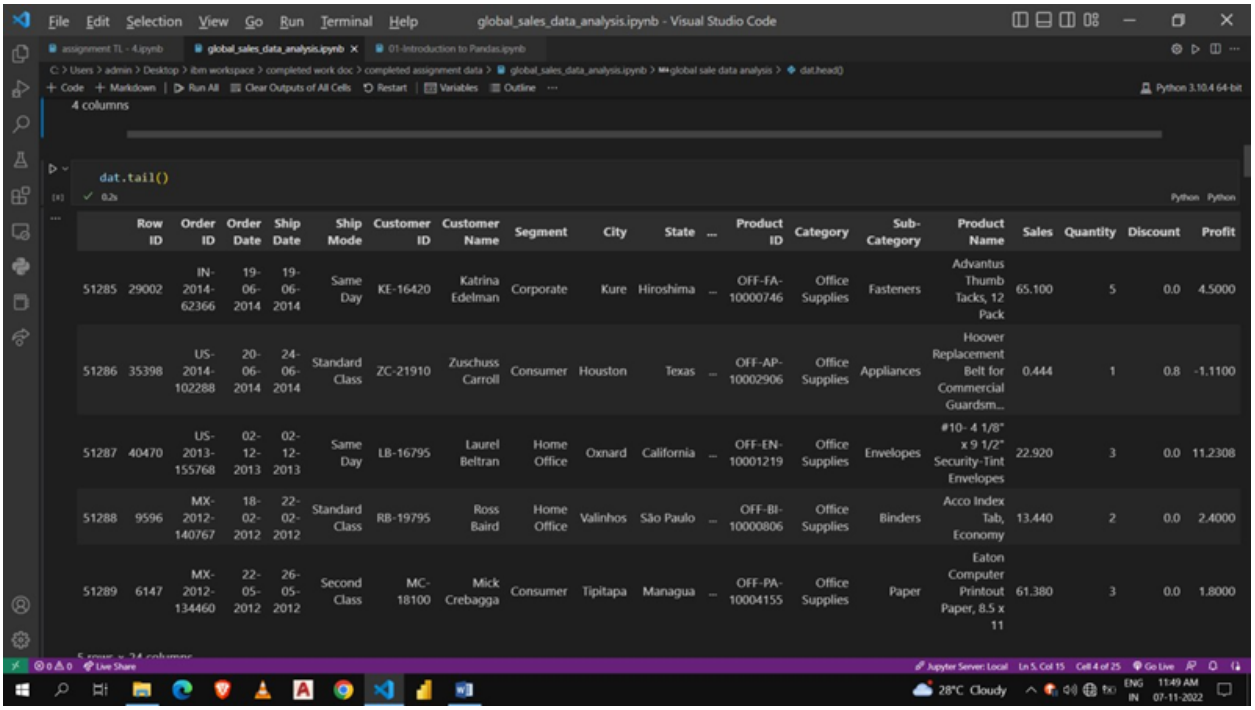
```
data = pd.read_csv("Global_Superstore2.csv", encoding = 'unicode_escape', engine ='python')
data
```

	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-	28-01-	30-01-	First	KM-	Katherine	Home	Berlin	Berlin	...	TEC-PH-	Technology	Phones	Motorola Smart	2892.510



```
<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
```

SPRINT 2:

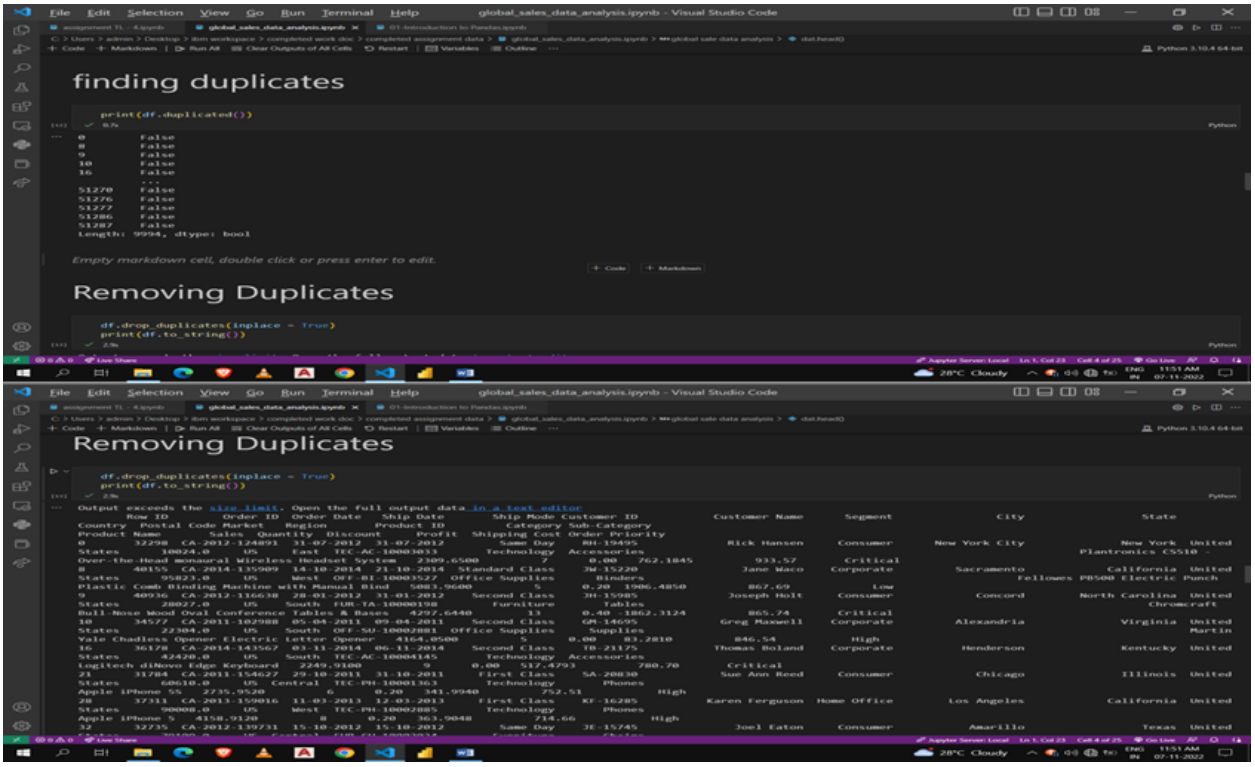


4 columns

```
dat.tail()
```

	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	Quantity	Discount	Profit
51285	29002	IN-2014-62366	19-06-2014	19-06-2014	Same Day	KE-16420	Katrina Edelman	Corporate	Kure	Hiroshima	...	OFF-FA-10000746	Office Supplies	Fasteners	Advantus Thumb Tacks, 12 Pack	65.100	5	0.0	4.5000
51286	35398	US-2014-102288	20-06-2014	24-06-2014	Standard Class	ZC-21910	Zuschuss Carroll	Consumer	Houston	Texas	...	OFF-AP-10002906	Office Supplies	Appliances	Hoover Replacement Belt for Commercial Guardsm...	0.444	1	0.8	-1.1100
51287	40470	US-2013-155768	02-12-2013	02-12-2013	Same Day	LB-16795	Laurel Beltran	Home Office	Oxnard	California	...	OFF-EN-10001219	Office Supplies	Envelopes	#10- 4 1/8" x 9 1/2" Security-Tint Envelopes	22.920	3	0.0	11.2308
51288	9596	MX-2012-140767	18-02-2012	22-02-2012	Standard Class	RB-19795	Ross Baird	Home Office	Valinhos	São Paulo	...	OFF-BI-10000806	Office Supplies	Binders	Acco Index Tab, Economy	13.440	2	0.0	2.4000
51289	6147	MX-2012-134460	22-05-2012	26-05-2012	Second Class	MC-18100	Mick Crebagga	Consumer	Tipitapa	Managua	...	OFF-PA-10004155	Office Supplies	Paper	Eaton Computer Printout Paper, 8.5 x 11	61.380	3	0.0	1.8000

5 rows × 24 columns



finding duplicates

```
print(df.duplicated())
```

```
0        False
8        False
9        False
10       False
16       False
         ...
51270    False
51276    False
51277    False
51286    False
51287    False
Length: 9994, dtype: bool
```

Empty markdown cell, double click or press enter to edit.

Removing Duplicates

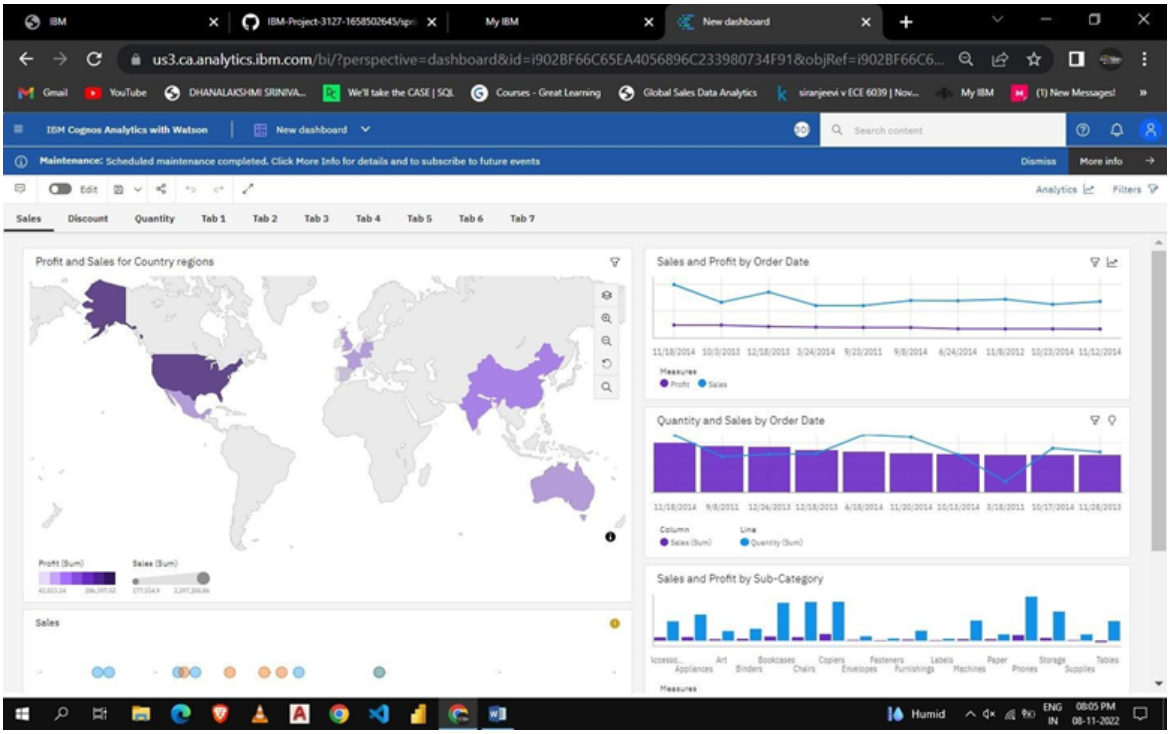
```
df.drop_duplicates(inplace = True)
print(df.to_string())
```

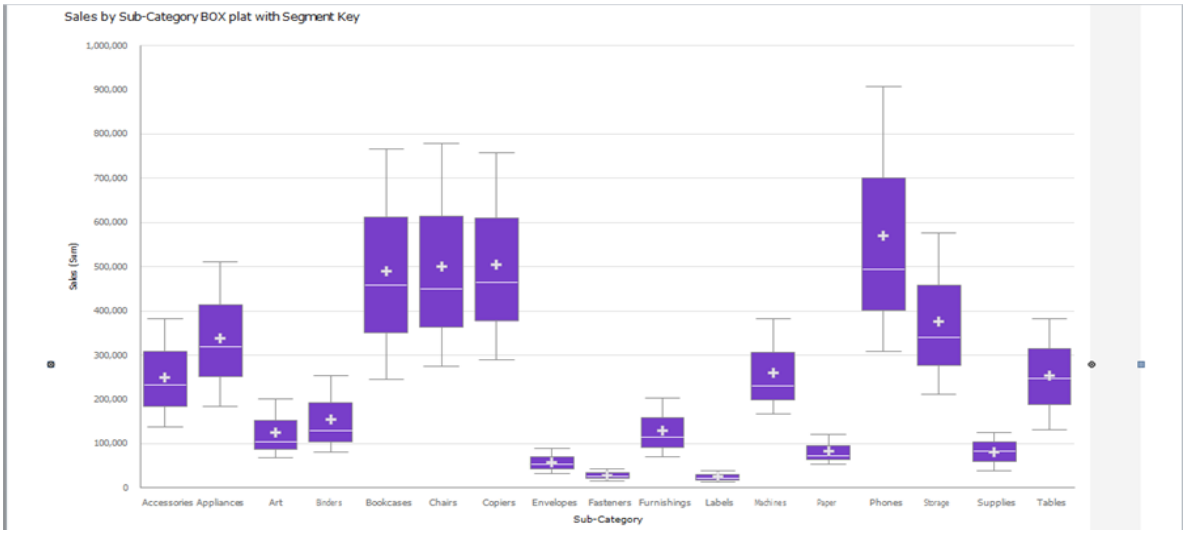
Removing Duplicates

```
df.drop_duplicates(inplace = True)
print(df.to_string())
```

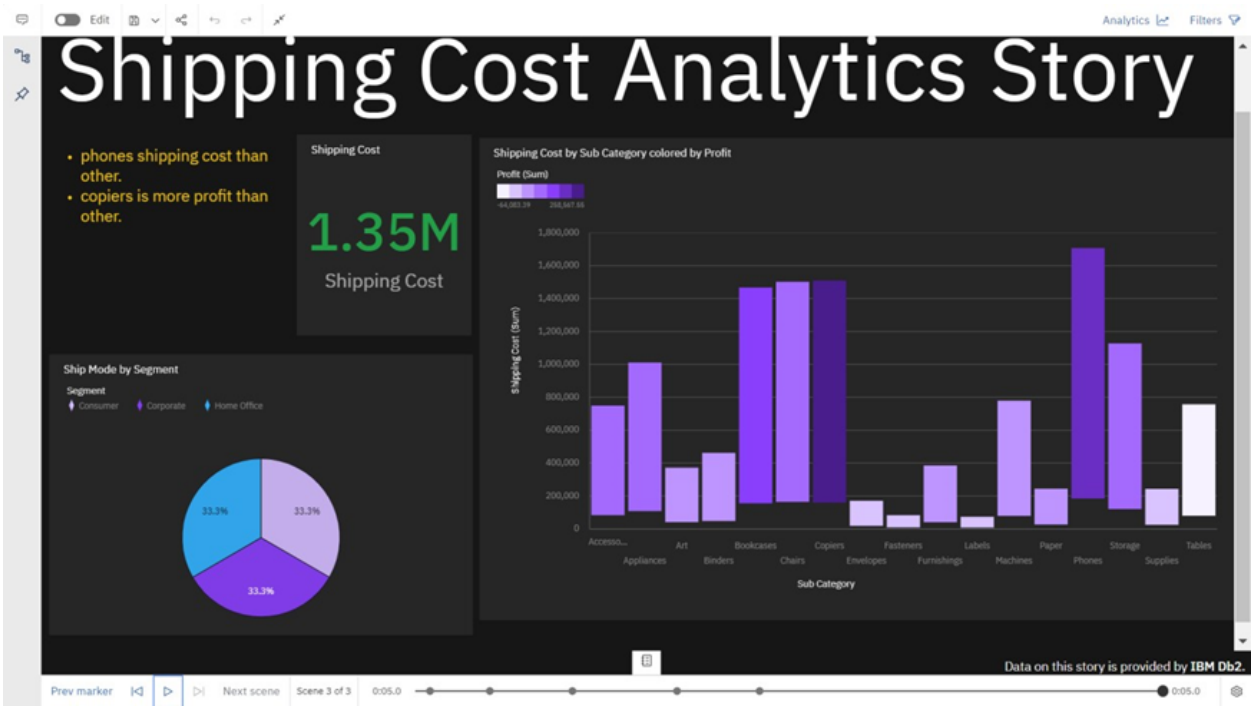
```
Output exceeds the size limit. Open the full output data in a text editor
      Row ID        Order ID  Order Date   Ship Date       Ship Mode Customer ID      Customer Name      Segment            City       State
Country  Postal Code Market    Region  Product ID         Category Sub-Category
Product Name     Sales  Quantity  Discount    Profit  Shipping Cost Order Priority
0          32298  CA-2012-124891  31-07-2012  31-07-2012     Same Day    RH-19495    Rick Hansen    Consumer    New York City   New York  United
States      10024.0        US      East  TEC-AC-10003833         Technology  Accessories
Plantronics CS510 - Over-the-Head monaural Wireless Headset System  2309.6500         7      0.00   762.1845      933.57      Critical
8          48155  CA-2014-135989  14-10-2014  21-10-2014   Standard Class   JM-15220   Jane Waco   Corporate  Sacramento  California  United
States      95823.0        US      West  OFF-BI-10003527  Office Supplies      Binders
Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind  5081.9600         5      0.20   1906.4850      867.69         Low
9          4096  CA-2012-116638  28-01-2012  31-01-2012   Second Class   JH-15985  Joseph Holt    Consumer     Concord  North Carolina  United
States      28027.0        US      South  FUR-TA-10000198         Furniture      Tables
Chromcraft Bull-Nose Wood Oval Conference Tables & Bases  4297.6440        13      0.40   1862.3124      865.74      Critical
10         34577  CA-2011-102988  05-04-2011  09-04-2011   Second Class   GM-14695  Greg Maxwell   Corporate    Alexandria   Virginia  United
States      22304.0        US      South  OFF-SU-10002881  Office Supplies      Supplies
Martin Yale Chadless Opener Electric Letter Opener  4164.0500         5      0.00    83.2830      846.54        High
16         36178  CA-2014-143567  03-11-2014  06-11-2014   Second Class   TB-21175  Thomas Boland   Corporate    Henderson   Kentucky  United
States      42420.0        US      South  TEC-AC-10004145         Technology  Accessories
Logitech diNovo Edge Keyboard  2249.9100         9      0.00   517.4793      780.70    Critical
21         31784  CA-2011-154627  29-10-2011  31-10-2011    First Class   SA-20830  Sue Ann Reed    Consumer     Chicago   Illinois  United
States      60610.0        US    Central  TEC-PH-10003363         Technology      Phones
Apple iPhone 5S  2735.9520         6      0.20    341.9940      752.51        High
28         37311  CA-2013-159016  11-03-2013  12-03-2013    First Class   KF-16285  Karen Ferguson  Home Office  Los Angeles  California  United
States      90008.0        US      West  TEC-PH-10002885         Technology      Phones
Apple iPhone 5  4158.9120         8      0.20    363.9048      714.66        High
32         32735  CA-2012-139731  15-10-2012  15-10-2012     Same Day    JE-15745    Joel Eaton    Consumer     Amarillo      Texas  United
```

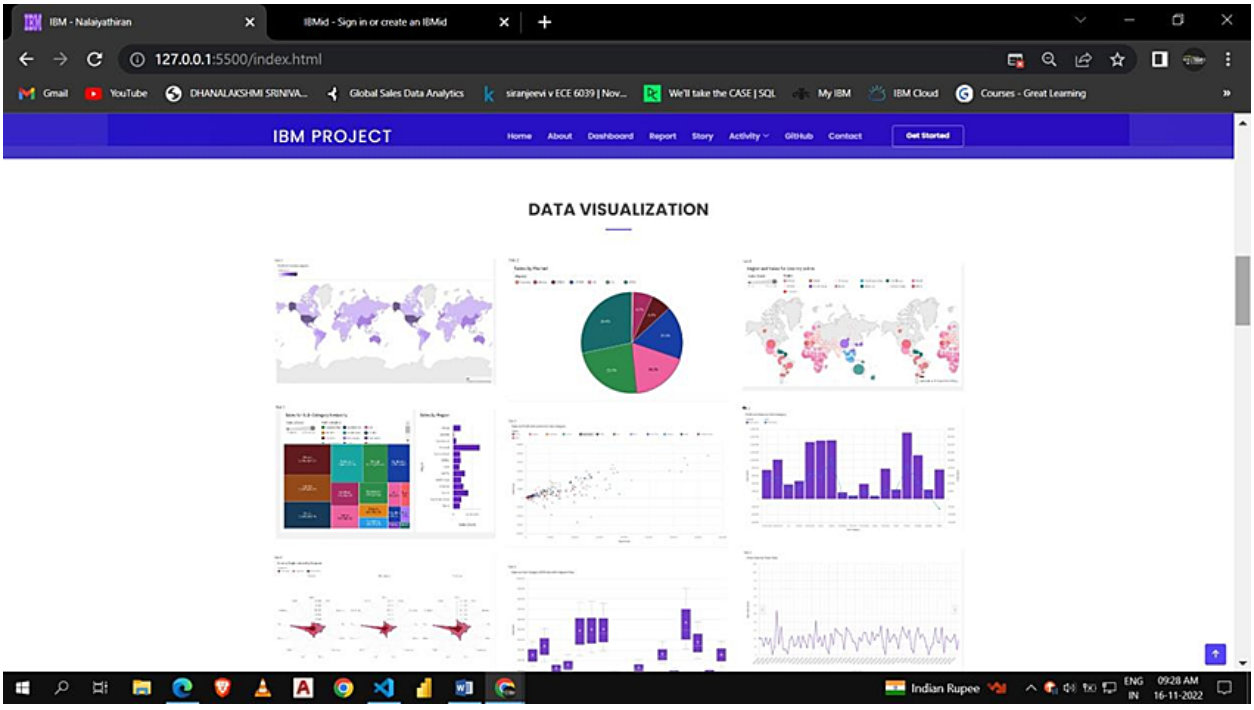
SPRINT 3:


IBM Cognos Analytics with Watson
New dashboard
Profit and Sales for Country regions
Sales and Profit by Order Date
Quantity and Sales by Order Date
Sales and Profit by Sub-Category
Sales


Sales by Sub-Category BOX plat with Segment Key
Sales (Sum)
Sub-Category
Accessories
Appliances
Art
Binders
Bookcases
Chairs
Copiers
Envelopes
Fasteners
Furnishings
Labels
Machines
Paper
Phones
Storage
Supplies
Tables

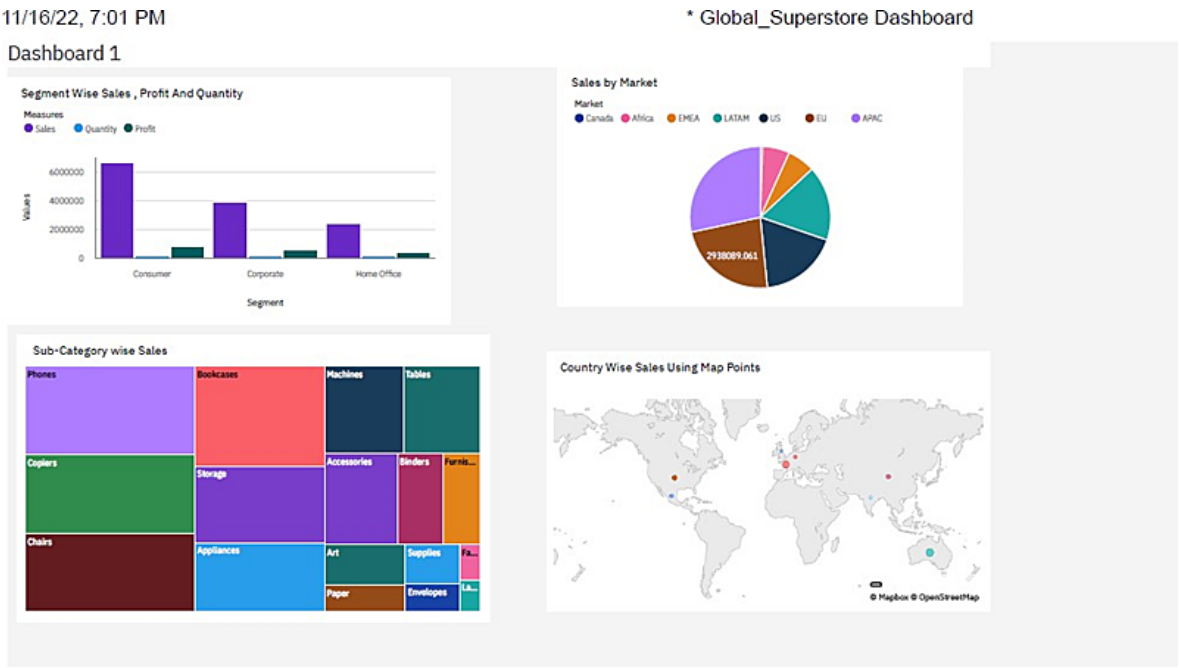
SPRINT 4:

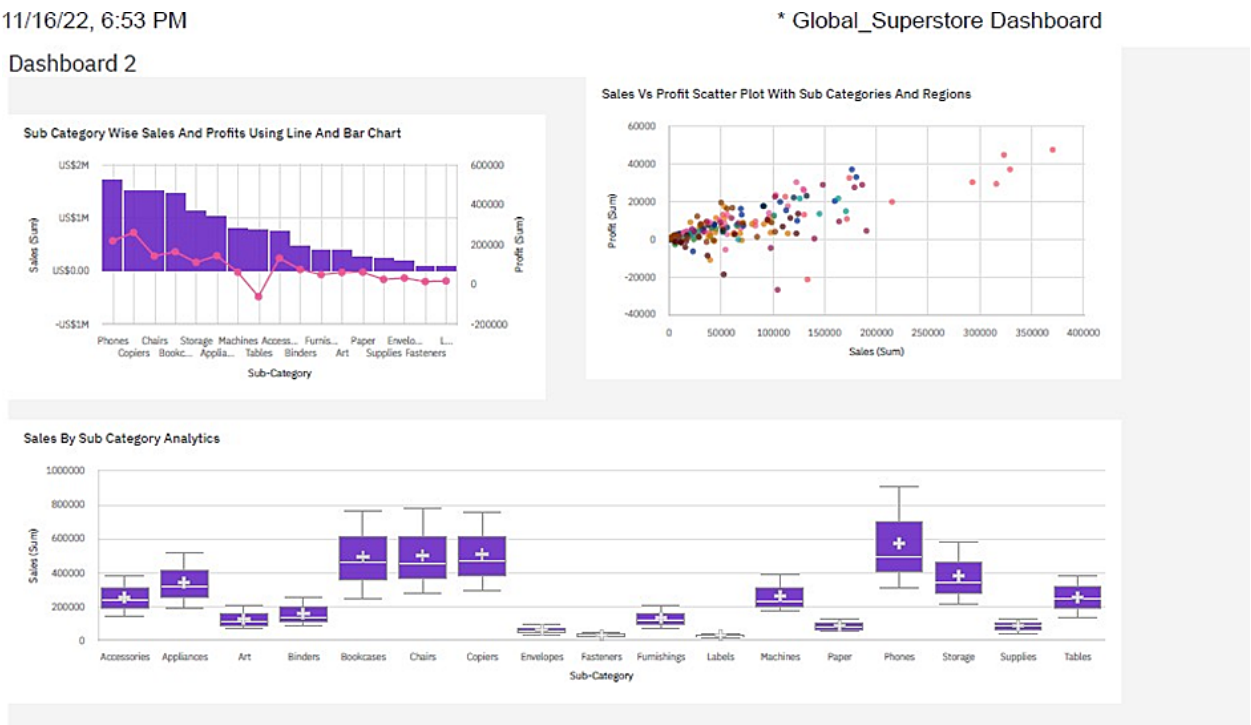

Edit
Analytics
Filters
Shipping Cost Analytics Story
• phones shipping cost than other.
• copiers is more profit than other.
Shipping Cost
1.35M
Shipping Cost
Shipping Cost by Sub Category colored by Profit
Profit (Sum)
Shipping Cost (Sum)
Sub Category
Ship Mode by Segment
Segment
Consumer
Corporate
Home Office
33.3%
33.3%
33.3%
Data on this story is provided by IBM Db2.
Prev marker
Next scene
Scene 3 of 3


IBM - Nalaiyathiran
IBMid - Sign in or create an IBMid
127.0.0.1:5500/index.html
IBM PROJECT
Home
About
Dashboard
Report
Story
Activity
GitHub
Contact
Get Started
DATA VISUALIZATION
Indian Rupee
16-11-2022

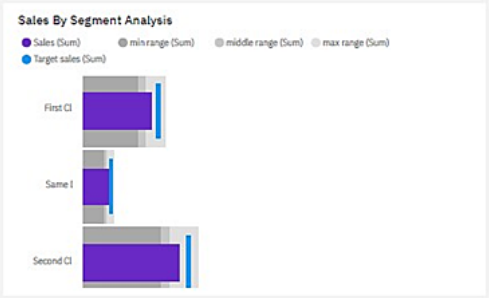
7.RESULTS:

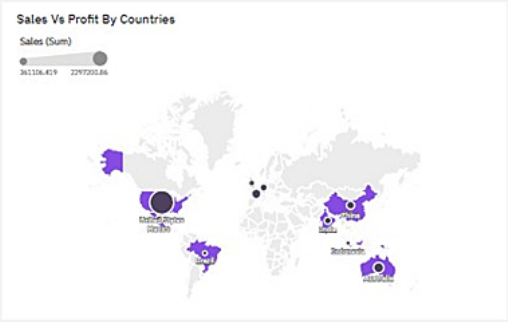
7.1 PERFOMANCE METRICES

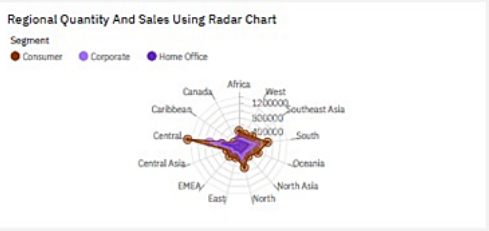




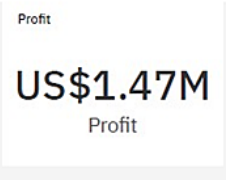
Dashboard3

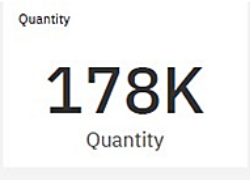


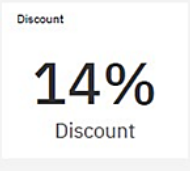


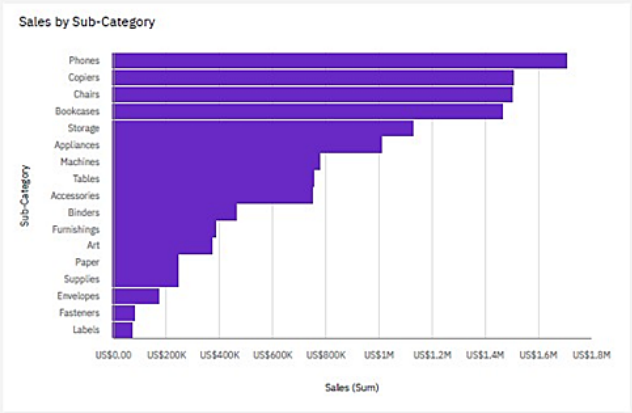


Dashboard 4









8.ADVANTAGES & DISADVANTAGES

ADVANTAGES

a. It was the cost efficiency project.

b. Receive full-scale services Maximize presentation

c. It was the timing saving project for peoples.

DISADVANTAGES

a. The lack of data security is the big disadvantages in this project.

b. Risk of choosing the wrong provider

9.CONCLUSION

By implementing this analytics solution, the company brought their competitive and sales data reporting in-house,cut costs and increased the accuracy of their reporting and analysis. As the company moves forward with this new solution, their sales reporting costs will most likely be reduced by 50 to 70%.

They are now able to analyse raw data themselves, respond more quickly to changes in market trends and perform root cause analysis to determine those shifts in the market. By securing quicker access to their data with the new solution, the company was also able to reduce the risk associated with delayed responses to changes in their markets.

With the new solution, the company can now process sales reports faster than the outsourced solution, reducing turn around time between 50% to 60%. The reporting needs of the company have been streamlined, consolidating over 10 reports into the centralized dashboard solution.

The company's competitive analysis group is also able to more quickly respond to internal data requests given they have the ability to pull the

information themselves. With this quicker response,the company is better able to react to changes in the market and predict opportunities for its sales force.

The business also experienced an increase in the overall understanding of their sales data throughout the organization. The company now has great flexibility in the presentation of their sales and competitive data, while also being able to integrate sales data with other key data points for the organization.

10.FUTURE SCOPE

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. It uses different metrics and KPIs to plan an efficient sales model that generates higher revenue for the business.

GITHUB PROJECT DEMO LINK

Github Link : https://github.com/IBM-EPBL/IBM-Project-5204-1658751561