Project Report

DATE	14 November 2022
Team ID	PNT2022TMID28002
Team Members	Kishore R (Team Leader) Ganesh Karthik R (Team member 1) Farhan Ahmed R (Team member 2) Prasanna Kumar V (Team member 3)
Project Name	Project - Global Sales Data Analytics

Project Name: Global Sales Data Analytics

Team Id: PNT2022TMID28002

Team Leader: Kishore R

Team Member 1: Ganesh Karthik R

Team Member 2: Farhan Ahmed R

Team Member 3: Prasanna Kumar V

CONTENT

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **Visualization**
8. **ADVANTAGES & DISADVANTAGES**
9. **CONCLUSION**
10. **FUTURE SCOPE**
11. **APPENDIX**
    - Source Code
    - GitHub & Project Demo Link

1. INTRODUCTION

1.1 Project Overview

Even with loads of data available at hand, a lot of companies struggle to identify key information such as how they can boost their sales or where they can minimize their losses. This is when they approach companies that offer analytics services. Through these services, businesses will be able to get useful insights into how well (or bad)

their business is doing and they end up making impactful business decisions.

1.2 Purpose

To analyse and identify trends in given data in order to make impactful business decisions

2. LITERATURE SURVEY

2.1 Existing 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. So try to understand a few things like, Customer Analysis and Product Analysis of this Global Superstore.

2.2 References

S.No	Paper/Title	Year	Journal	Method	Advantage	Disadvantages
1	Effects of 3D Virtual "Try-On" on Online Sales and Customers' Purchasing Experiences	September 2020	IEEE Access (Volume: 8)	1)VTO technology consists of making a virtual body model from the customer's own body size, 3D garment modelling, and interactive try-on and mix-and-match of garments. Many studies explain the method of making a virtual body by scanning or measuring the customer's body 2)We propose a new method, instead of 2D CADpatterns, we use 3D garment photos; to make a 3D garment mesh surface because 2D CAD patterns are always copyrighted and require	1) Virtual try-on (VTO) is a new technology used to help customers try on and mix and match apparel without a fitting room. 2)It appeared when conventional shops were in crisis because of consumers' shift to online retailers. VTO technology has been adopted because it has many advantages in both retail channels	The thing to overcome is the psychological barrier on the side of the customer, and convincing them to use and trust virtual try-on applications. That requires being extra clear about what's happening with user data such as images and videos, where they're processed, how they're used, and so on
2.	Sales Analytics and Big Data Developmen	March 2021	Research Paper	The association of cancer drug costs and total cancer treatment costs per cancer site	This paper provides a commercial/sales analytics,	This paper provides a commercial/sales analytics, big

	ts Needed Now to Address PractitionerI dentified Emerging Biopharmace uti cal Sales Force Strategic and Operational Issues			uses two methods: Kaplan-Meier Sample Average (KMSA) method and an approach similar to the Cox proportional hazard model. Generalised Propensity Score (GPS) based weighting with bootstrap standard errors can be used to estimate the marginal effect of detailing on drug utilisation.	big data management, and organisational blueprint for companies on how to prepare and operate successfully in this evolving sales force pharma landscape. Focuses largely on patient,payer,sa les and marketing analytics	data management, and organisational blueprint for companies on how to prepare and operate successfully in this evolving sales force pharma landscape. Focuses largely on patient,payer,sal es and marketing analytics
3.	Impact of big data analytics on sales performance in pharmaceuti cal organization	April 2021	Research Paper	This study is related to the quantitative research method. This method emphasises the statistical, mathematical,or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre- existing statistical data using computational techniques.	Sales performance effectively and efficiently achieves the targets in the sales process by examining opportunities and improving closing rates	The data were collected from a developing country, and the results of this research may be different in developed countries
4.	Visual Analytics for Decision Support: A Supply Chain Perspective	June 2021	Research Paper	1. Visualisation types suitable for particular analytical goals in each of the SC activities, i.e Visualisation Techniques. 2. Analytical reasoning by analysts involved in the process of analysis, i.e., tactics	1. The lack of identifying specific SC business decisions that can be supported by VA. 2. The lack of exploring various analytical capabilities of SC VA systems. 3. The lack of	First, the application of VA in supporting the source and make processes of the SC may be explored, such as cost modelling of different sources and production scheduling. Second, visualising the impact of

					identifying the state of the art in visualisation techniques and tactics have been resolved.	external variables on different SC operations such as the effect of weather forecast on sales may be investigated more extensively.

Paper links:

1. **Effects of 3D Virtual "Try-On" on Online Sales and Customers' Purchasing Experiences**

https://ieeexplore.ieee.org/document/9189849

2. **Sales Analytics and Big Data Developments Needed Now to Address Practitioner-identified Emerging Biopharmaceutical Sales Force Strategic and Operational Issues**

https://www.pmsa.org/jpmsa-vol05-article01

3. **Impact of big data analytics on sales performance in pharmaceutical organisations**

https://doi.org/10.1371/journal.pone.0250229

4. **Visual Analytics for Decision Support: A Supply Chain Perspective**

https://ieeexplore.ieee.org/document/9445829

2.3 Problem Statement Definition

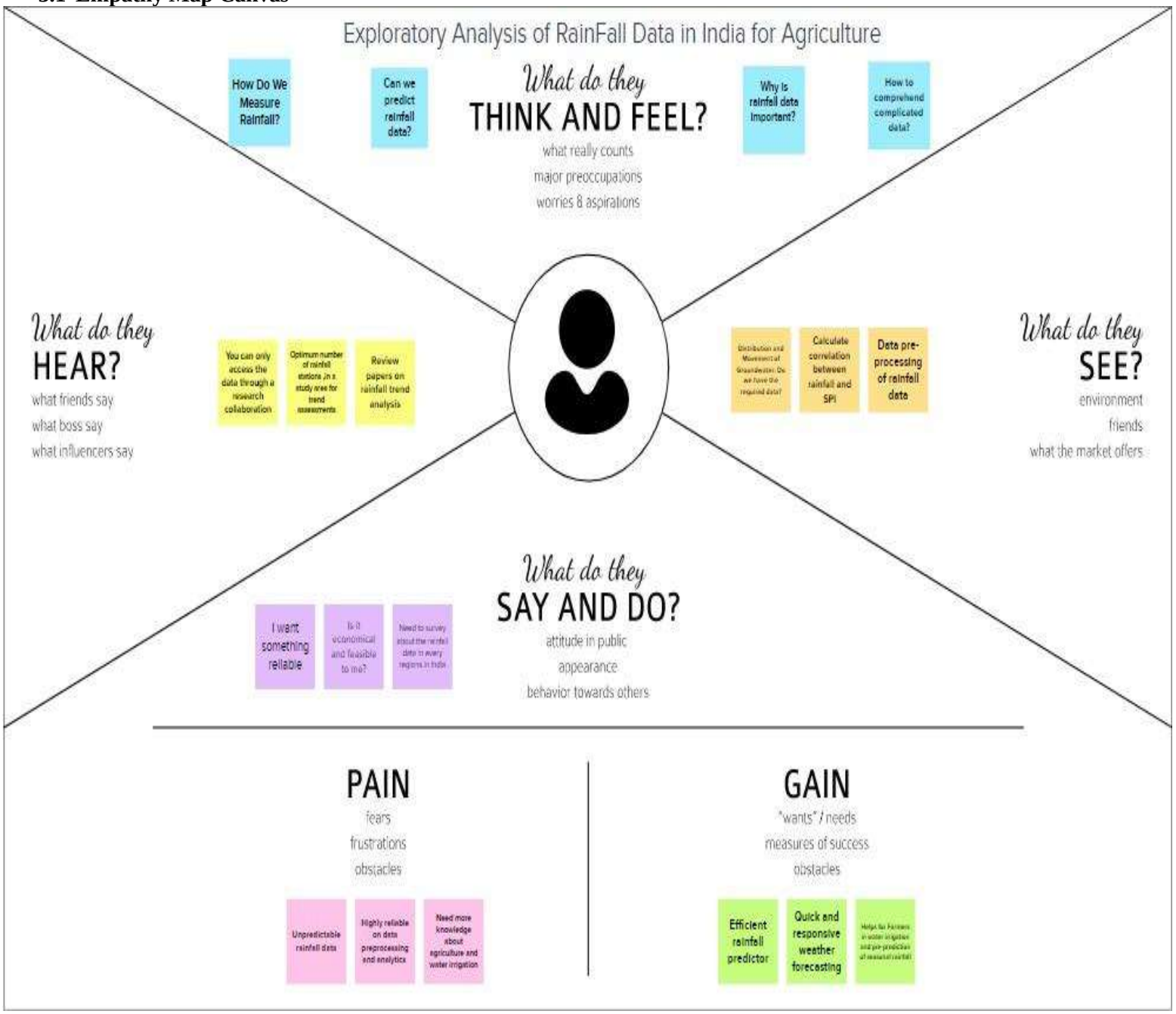
Sales data is a term that includes a large array of metrics but, broadly speaking, if you can measure something in relation to the sales process, its viable sales data. Modern software like IBM cognos can help you collect this data, but it's important to learn how to read this data to understand what it means for your business and where you can improve.

With such an expansive sales data definition, it can be hard to know where to focus your efforts—especially if you also have to spend time selling. With the right sales analysis tools, you'll have a wealth of information and be able to spot trends that will empower you and your team and provide better sales forecasts and goals for the rest of the organization. The trick is in knowing which metrics to use and when.

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	a customer	buy a product	it is not in stock	it is in demand and many people are placing orders	annoyed
PS-2	a customer	replace a product that I bought	customer support is not responding	of too much traffic	disappointed
PS-3	a customer	buy a product	it is overpriced	of its demand	frustated
PS-4	a customer	buy a product	there is something odd about the reviews	the product as such isn't the best in its domain	vexed

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

Top 3 ideas :

1. Identify which markets are thriving in each region/zone
2. Perform segment-wise analysis
3. Identify returning customers (if any)

3.3 Proposed Solution

1. Problem Statement (Problem to be solved)

Customer, Product analysis and detect emerging trends in business using data analytics. (Sales analysis is an important aspect of running a successful business. Through sales analytics, we can decide which products to focus on, where to sell and how best to reach customers.)

2. Idea / Solution description

There are many ready-made analytics software's but that is not the cure to all problems. To tackle this, we are creating a customized analytics model that will help helps businesses of all sizes grow revenue, automate tasks, make smarter decisions and keeps you updated of your customer behavioural changes

3. Novelty / Uniqueness

The customized analytics model will help to get a successful data analytics initiative that will provide a clear picture of where you are, where you have been and where you should go. By using this model. So, this model will stand apart from others.

4. Social Impact / Customer Satisfaction

This model will help various businesses to make informed business decisions, improve efficiency, increase revenue and minimize the financial loss.

5. Business Model (Revenue Model)

This model help companies better understand their customers, evaluate their ad campaigns, personalize content and create content strategies, so it will be a successful model. As it is useful it will definitely attract customers.

6. Scalability of the Solution

Whatever the dataset is, the customized model will produce a clear insight of the data with good and interactive visualization. That data visualization is used to identify patterns, trends, etc.

3.4 Problem Solution fit

CUSTOMER SEGMENT(S)	**CUSTOMER CONSTRAINTS**	**AVAILABLE SOLUTIONS**
1. Companies / Organizations. 2. Retailers. 3. Sales and Marketing Team.	1. Low quality of data 2. High cost 3. Going through same process each time like finding the default and essential thing (Like finding low and high selling product)	Creating a customized analytics software that will automate the repetitive tasks, which will a lot of time for the customers.

JOBS-TO-BE-DONE / PROBLEMS	**PROBLEM ROOT CAUSE**	**BEHAVIOUR**
Each time we have to look for some important things in the dataset like which is the least selling product and all. In the current system we have to do some steps in-order to get the output. It will be time consuming if we do this for each and every datasets	Repetition of some process each time while analyzing a dataset.	Trying to find which is the highest and low selling products by using the analytical tool each time.

TRIGGERS Finding basic details (Like which region has the highest selling, etc.) of a dataset without spending a lot of time. **EMOTIONS** Before: Feeling irate for doing same process again and again After: They will feel calm and better	**YOUR SOLUTION** Automating all the repetitive process that takes place during the analysis of a dataset. By doing the we can able to show the customer some default analysis whenever the dataset is loaded into the software.	**CHALLENGE OF BEHAVIOUR** Try to find a way to avoid time constraint

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	confirmation via Email
FR-3	Login	The user should login to the system by using valid user credentials
FR-4	Dataset	Upload dataset into the analytics tool.
FR-5	Analysis	It involves gathering all the information, processing it, exploring the data, and using it to find patterns and other insights.
FR-6	Create Dashboard	Create Charts, Graphs, Tables, etc.

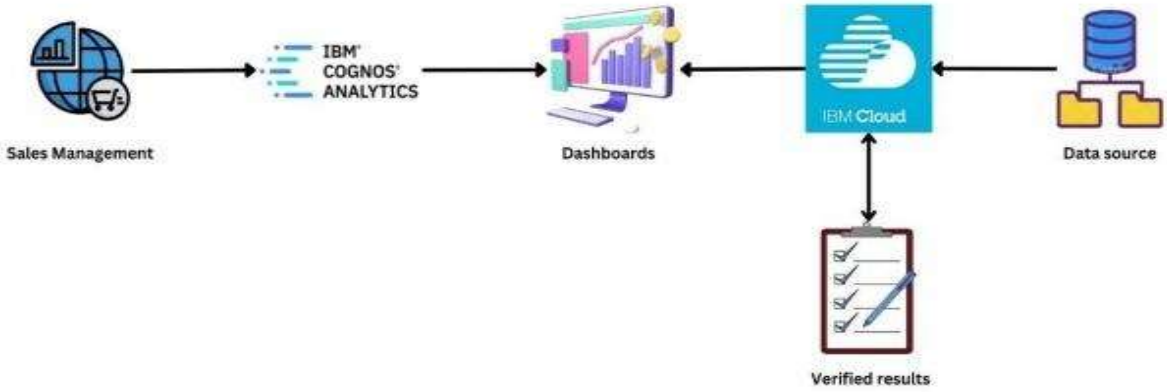
FR-7	Reporting	The reporting function helps users have complete control over their business. The real- time reporting collects current information and displays the data on an intuitive user interface.

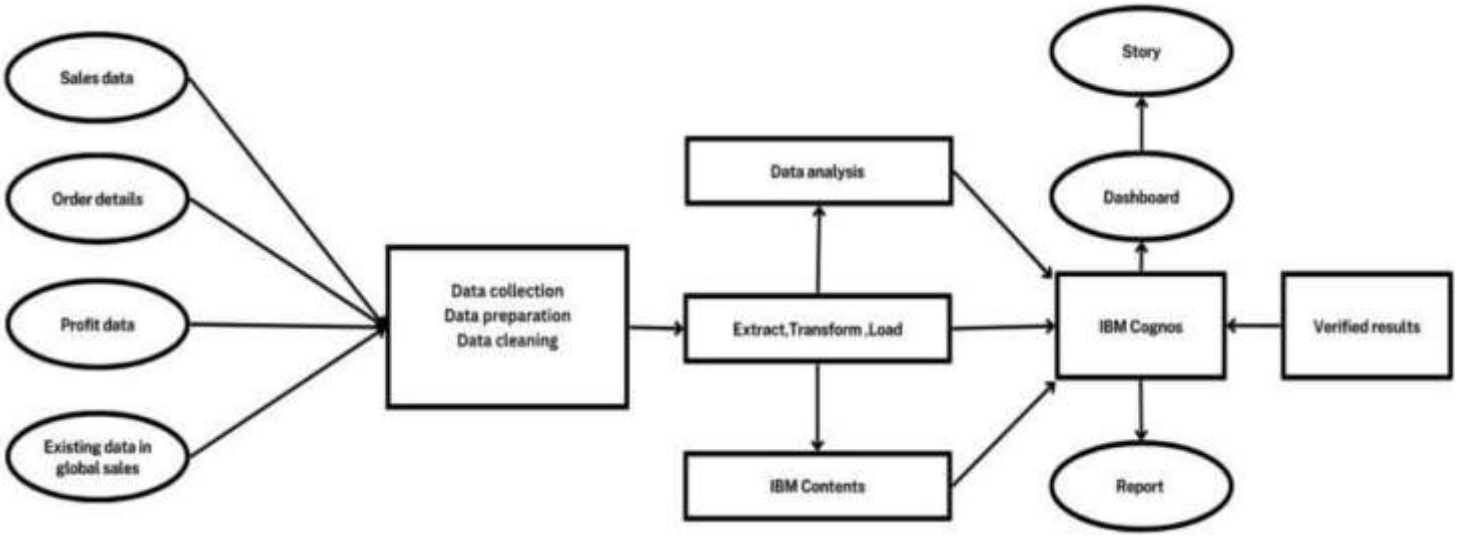
4.2 Non-Functional requirement

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Optimized resources and it can be used by everyone
NFR-2	Security	Anyone with correct Log In credentials can view the Dashboards/Templates.
NFR-3	Reliability	Templates are reliable because we are uploading and accessing it through Cloud
NFR-4	Performance	It has high state of performance and efficiency
NFR-5	Availability	It is free of cost and available to everyone who wants to know about sales data
NFR-6	Scalability	Dashboards/Templates are very much Scalable, the user can modify the metrics whenever they want

5. PROJECT DESIGN

5.1 Data Flow Diagrams


IBM
COGNOS
ANALYTICS
IBM Cloud
Sales Management
Dashboards
Data source
Verified results



5.2 Solution & Technical Architecture

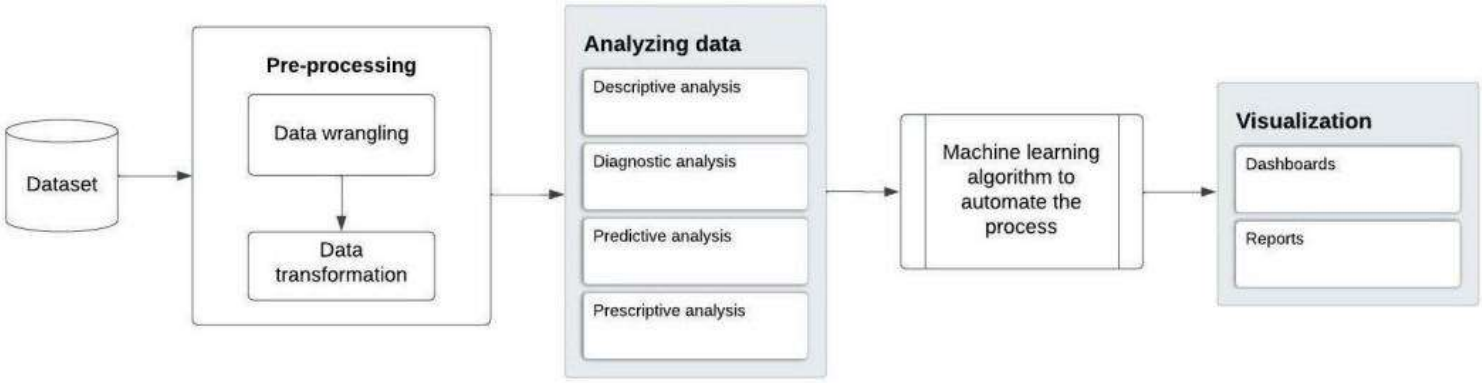


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1	User Interface	How user interacts with application	IBM Cognos Analytics
2	Working with the dataset	Cleaning, extracting process of dataset is done	IBM Cognos Analytics with Watson
3	Data Exploration	Information in the dataset is identified	IBM Cognos Analytics with Watson
4	Data Visualization	Data is represented in form of chart, table and graph in an interactive way	IBM Cognos Analytics with Watson
5	Outcome of analysis process	The user will see the visualization through dashboards, report and story	IBM Cognos Analytics with Watson

6	Cloud Database	Uploaded data are stored in the cloud database (Database Service on Cloud)	IBM DB2, IBM Cloudant
7	File Storage	File storage requirements	IBM Cloud

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	IBM Cognos, IBM Cloud, IBM Watson
2.	Security Implementations	Secure user information and data	Active Directory
3.	Scalable Architecture	Supports various data sizes	Web 3.0 IBM Cloud
4.	Availability	Multi page layout providing various visualizations of data and provide full support irrespective of platform and device specifications	Cognos Business Intelligence Server
5.	Performance	Withstand huge data and process them without crashing	IBM Cognos, Performance Management Hub

5.3 User Stories

Sprint	Functional Requirement (Epic)	UserStory Number	UserStory /Task	StoryPoints	Priority
Sprint-1	Registration and Login	USN-1	Register the application Log into the application after registration	2	High
Sprint-1	Data Collection	USN-1	Gather the data in the form of CSV/XLS files and clean the data to remove the null values	2	High

Sprint-2	Upload dataset	USN -2	Upload data to IBM Cognos	2	High

Sprint-2	Data Preparation and Visualization	USN -2	Filter data for visualization Visualize data in form of charts and graphs	2	High
Sprint-3	Dashboard	USN-3	Create dashboard based on data in IBM Cognos Plan Visualization	2	High
Sprint -4	Prediction and Final Analysis	USN - 4	Predict specific products future sales expectation Analyze the list of categorized products	2	High
Sprint -4	Report	USN -4	Prepare the product and customer description information as a report	2	High

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**
Sprint -1	Registration	USN -1	As a user, I can register for the application by entering my email, password, and confirming my password.	5	High
Sprint -1	Login	USN -2	As a user, I will receive confirmation email onceI have registered for the application, and I can log into the application by entering email & password	5	High

Sprint -1	Data Collection	USN -3	As a user, I need to gather the data in the form of CSV/XLS files and clean the data to remove the null values	10	Low
Sprint -2	Upload Dataset	USN -4	As a user, I will upload the data to IBM Cognosand view the data of the products	5	Medium

Sprint -2	Data preparation	USN -5	As a user, I need to filter the data for visualization in IBM Cognos	5	High
Sprint -2	Data Visualization	USN -6	As a use, I can easily visualize the data in the foím of chaíts and gíaphs thíough IBM Cognos	10	High
Sprint -3	Dashboard	USN -7	As a useí, I will cíeate the dashboaíds based on the given data in IBM Cognos	5	High
Sprint -3	Dashboard	USN -8	As a user, I must plan visualizations in a way that I'm able to gain insights regarding the sales based upon the category of sales and the respective region	5	Medium
Sprint -3	Dashboard	USN -9	As a user, I must be able to gain insights from the charts/graphs through a variety of relationships established in the dashboard.	10	Medium
Sprint -4	Prediction	USN -10	As a user, I will predict the specific product's future sales expectation.	5	Medium
Sprint -4	Final Analysis	USN -11	As a user, I can Analyse the list of categorized products and their details as a report.	5	High

Sprint -4	Report	USN -12	As a user, I can prepare the product and customer description information and more additional information as a Report	10	High

6.2 Sprint Delivery Schedule

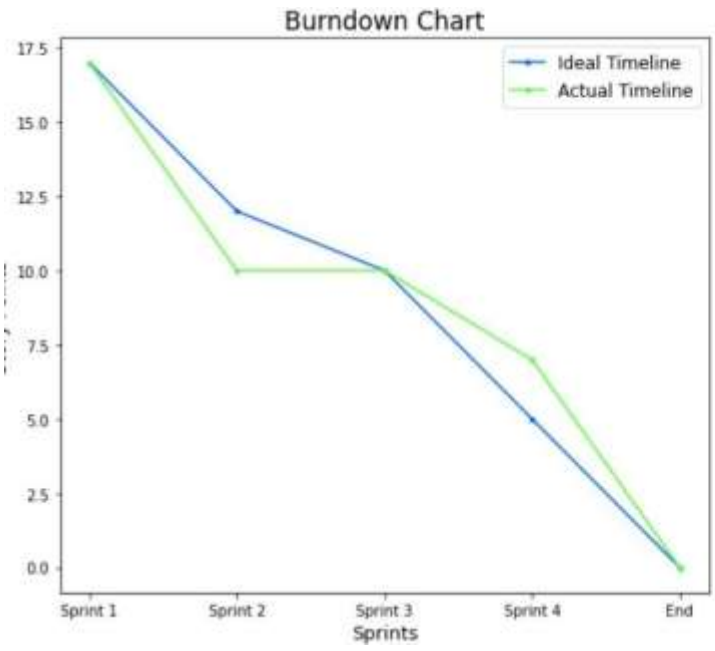
Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date (Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint Release Date (Actual)**
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	30 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	06 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	13 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	20 Nov 2022

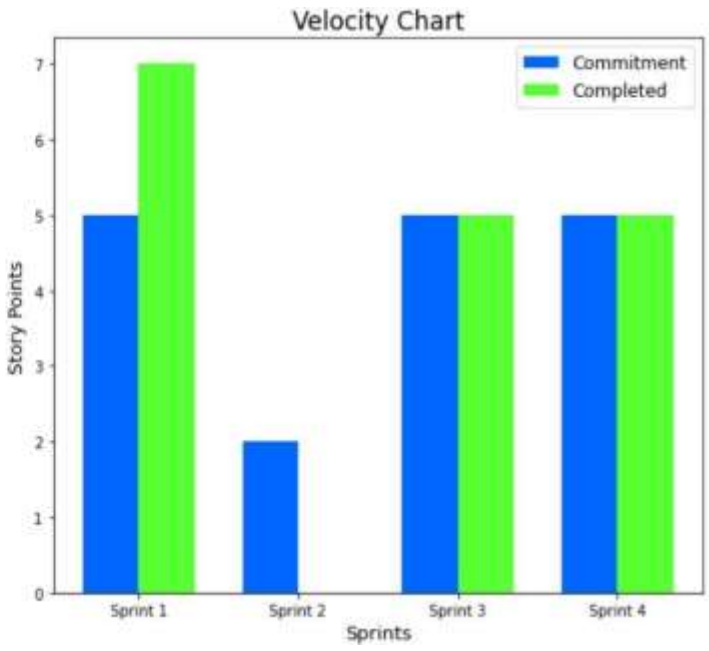
Velocity:

The team's average velocity (AV) per iteration unit (story points per day) :

Sprint 1: AV= Sprint duration/ velocity = 5/6 = 0.87
Sprint 2: AV= Sprint duration/ velocity = 2/6 = 0.34
Sprint 3: AV= Sprint duration/ velocity = 5/6 = 0.87
Sprint 4: AV= Sprint duration/ velocity = 5/6 = 0.87

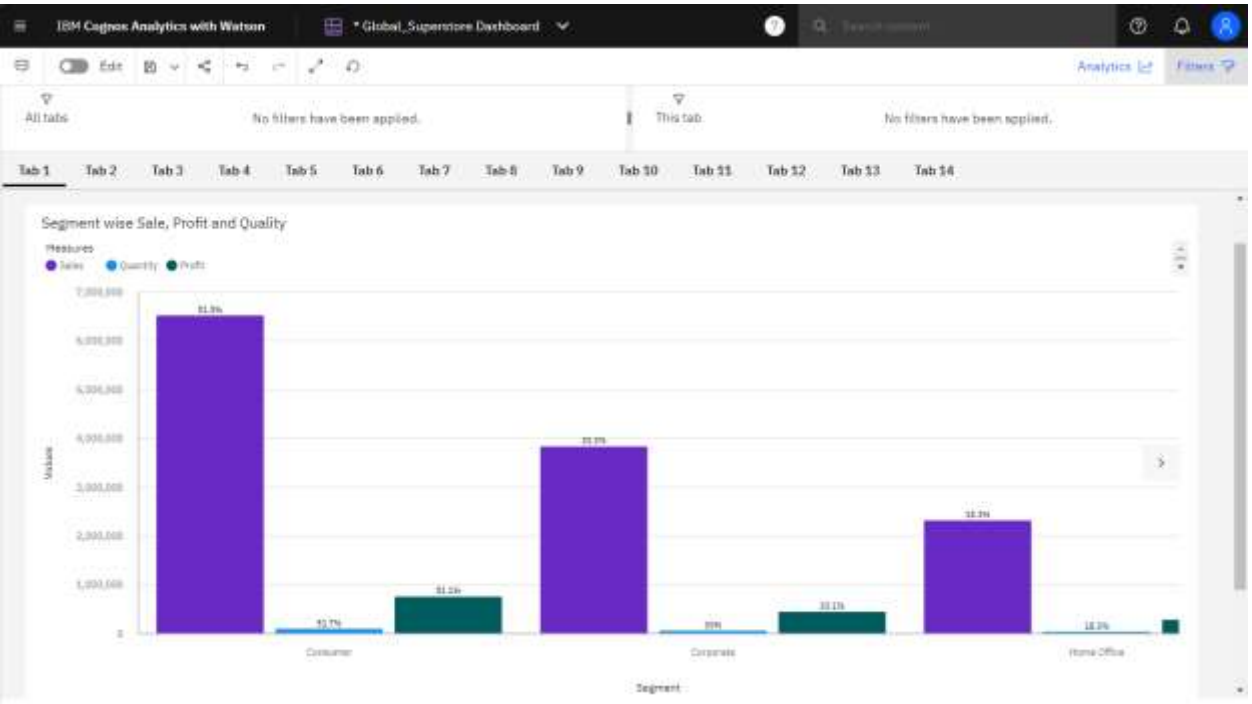
6.3 Reports from JIRA

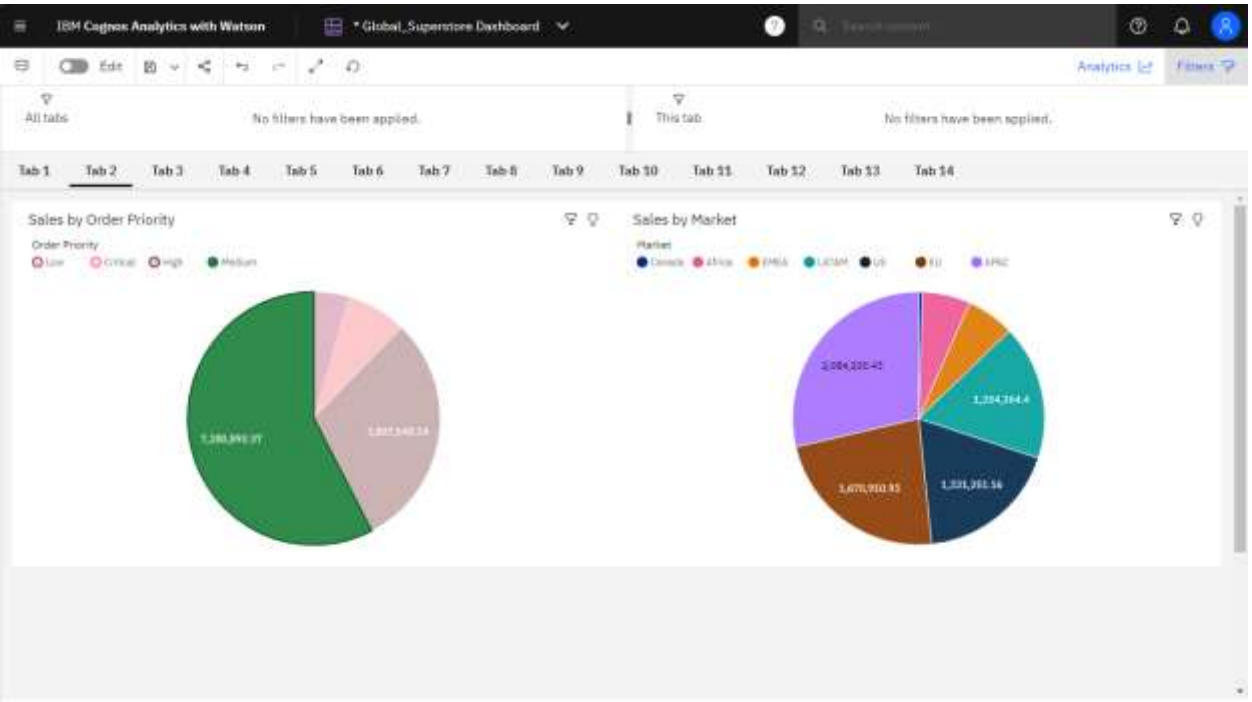


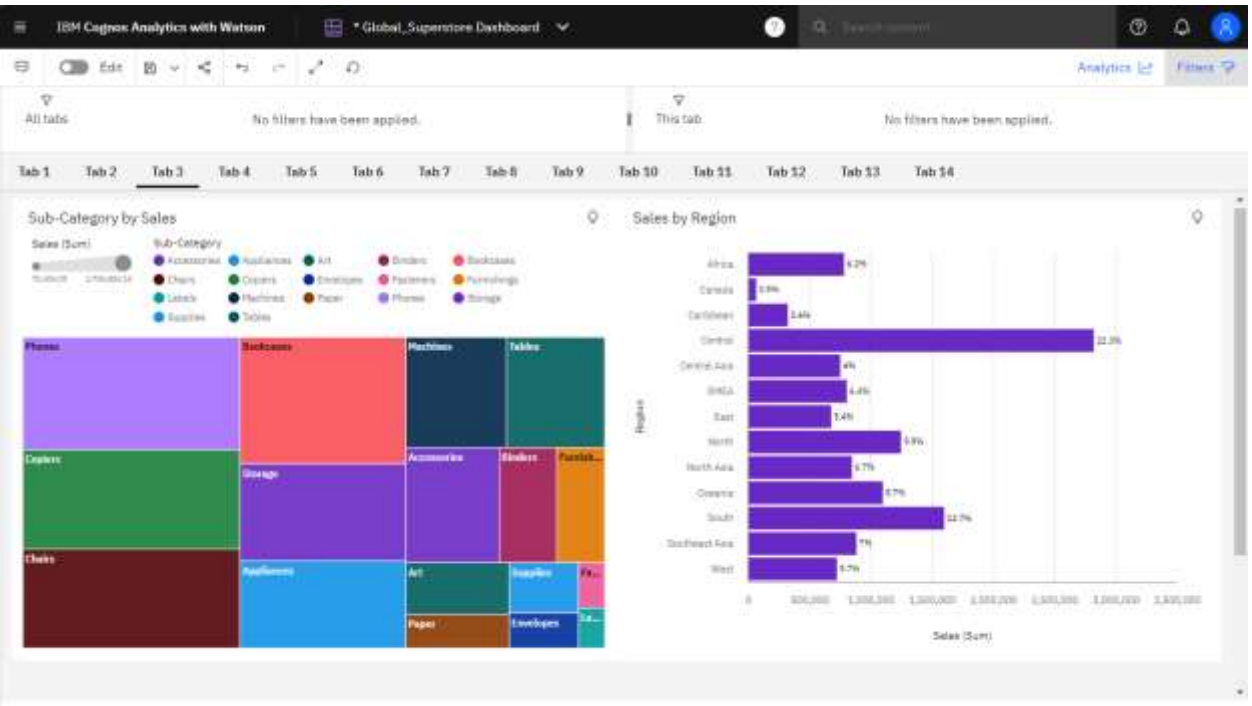


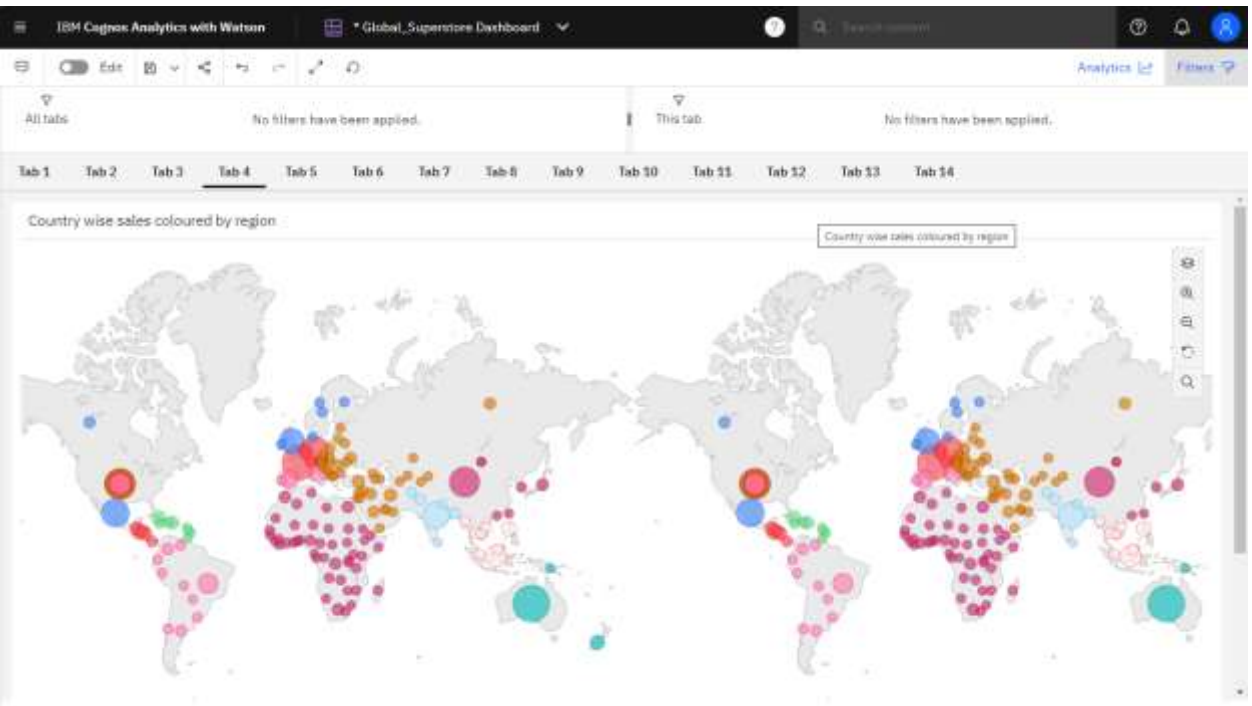
7. VISUALIZATION

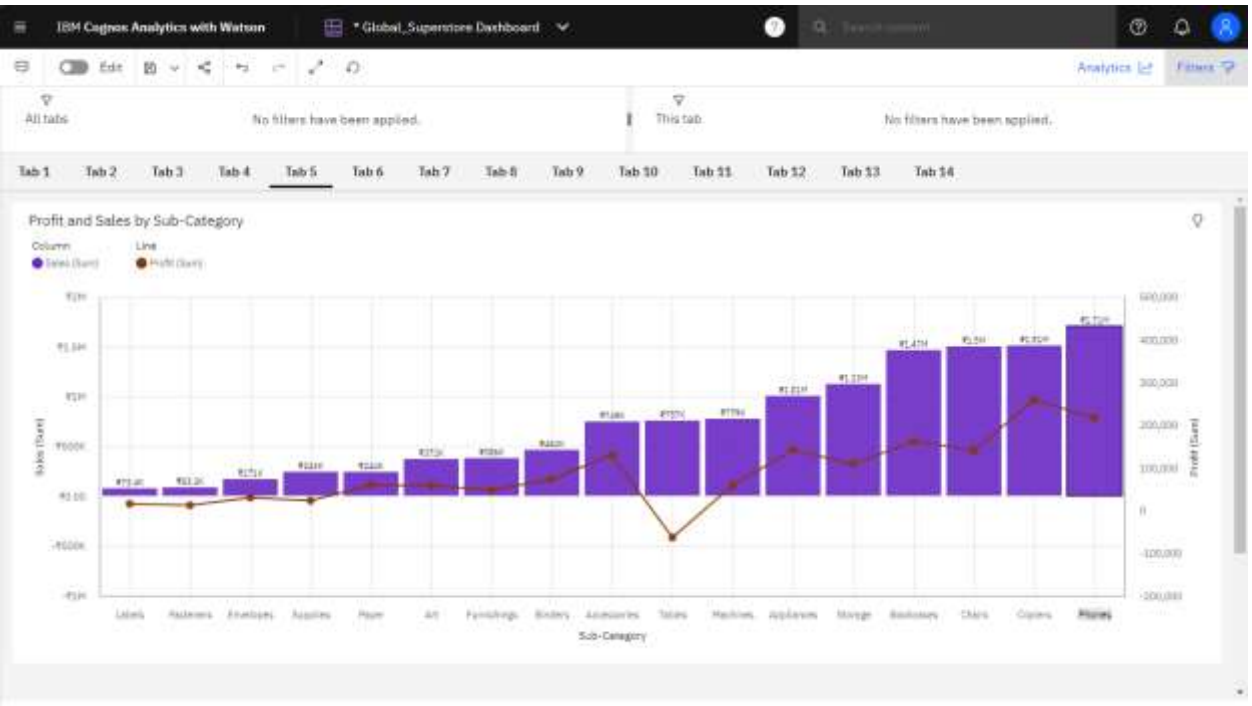
8. TESTING

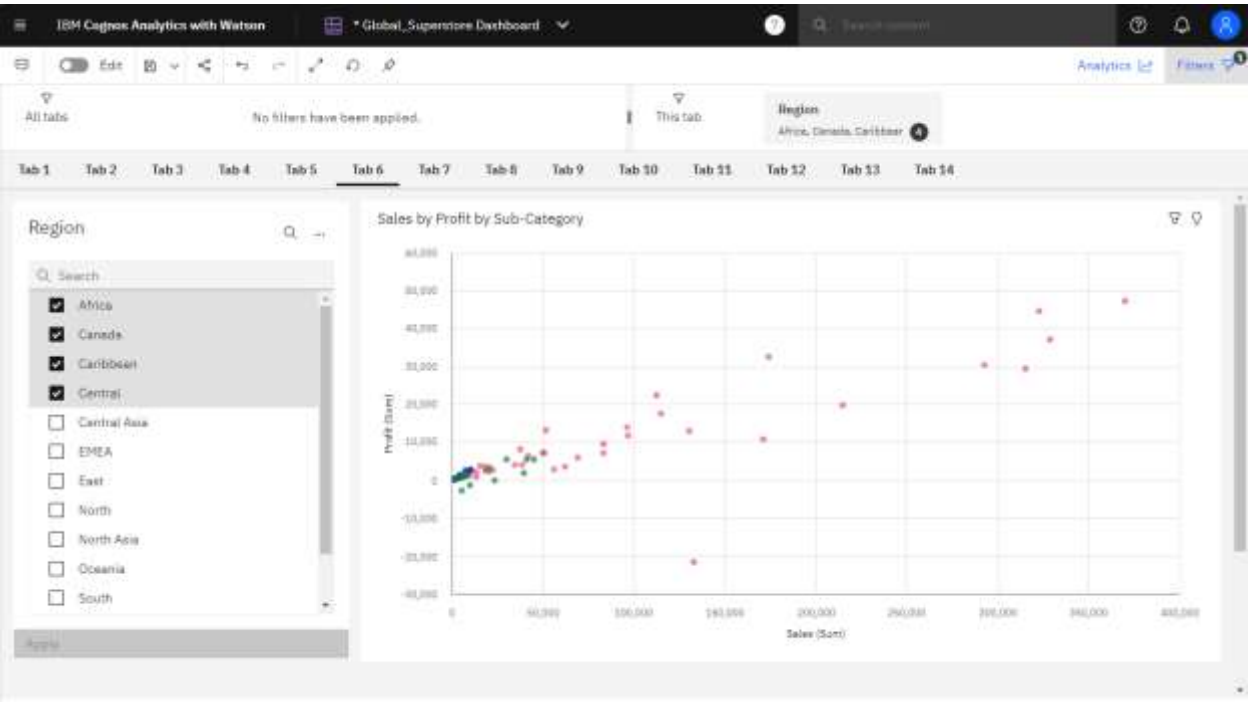


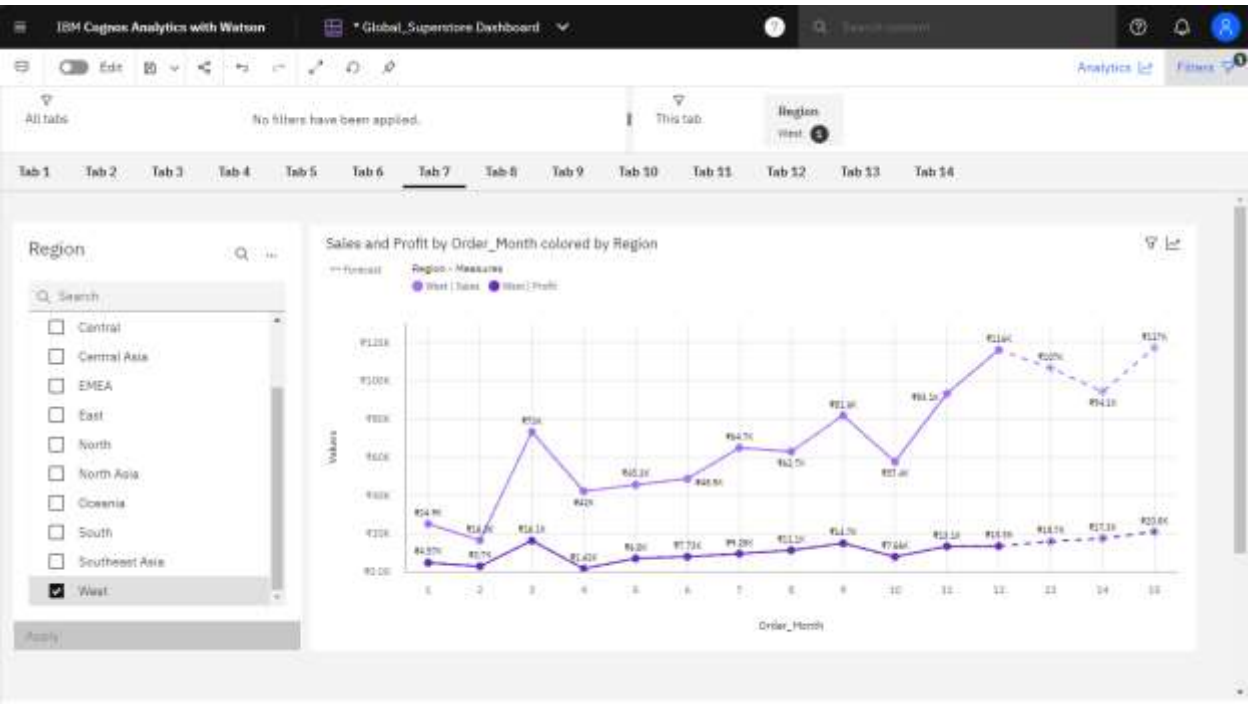


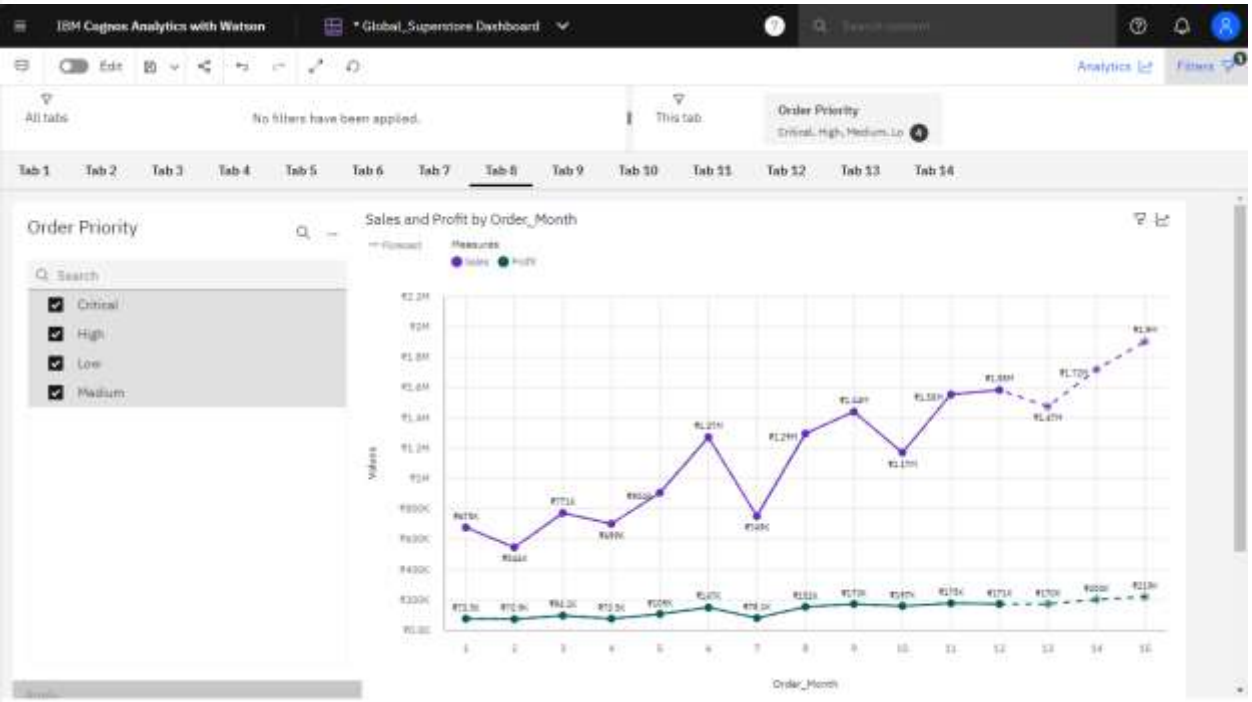

IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Sub-Category by Sales
Sales (Sum)
Sub-Category
Phones
Bookcases
Machines
Tables
Copiers
Storage
Accessories
Binders
Chairs
Appliances
Art
Supplies
Paper
Envelopes
Sales by Region
Region
Africa
Canada
Caribbean
Central
Central Asia
EMEA
East
North
North Asia
Oceania
South
Southeast Asia
West
Sales (Sum)

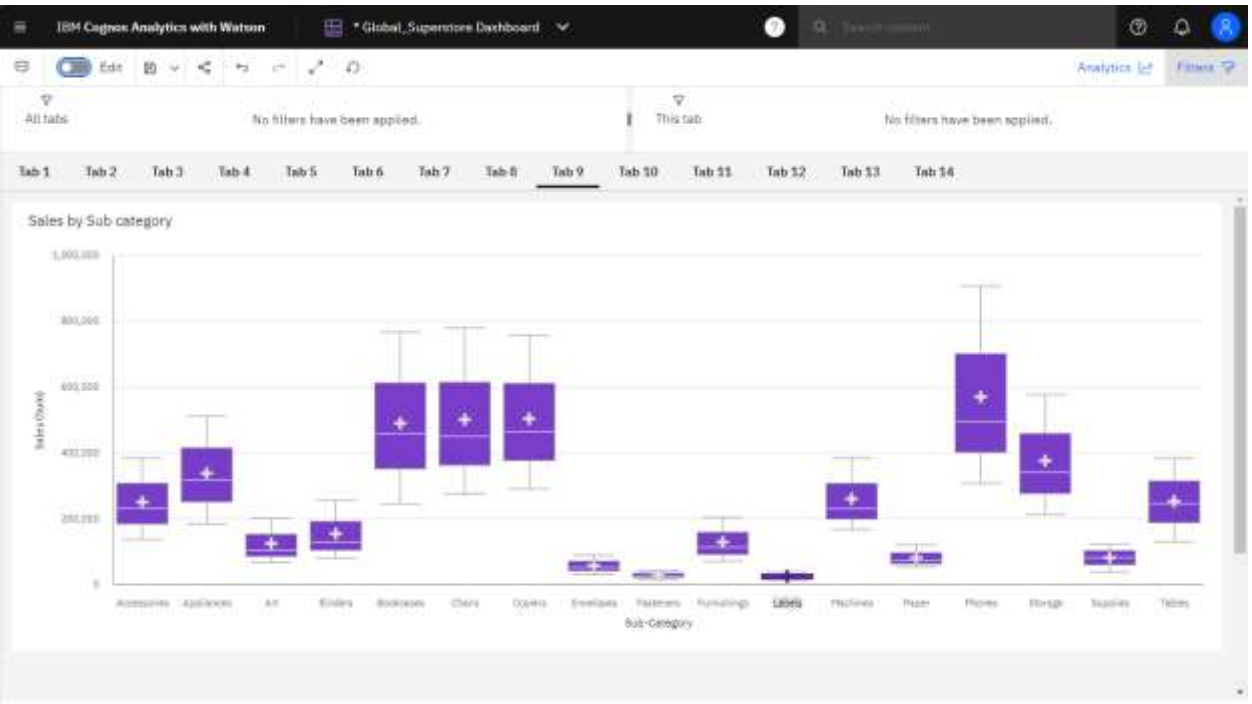

IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Country wise sales coloured by region
Country wise sales coloured by region


IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Profit and Sales by Sub-Category
Column
Line
Sub-Category

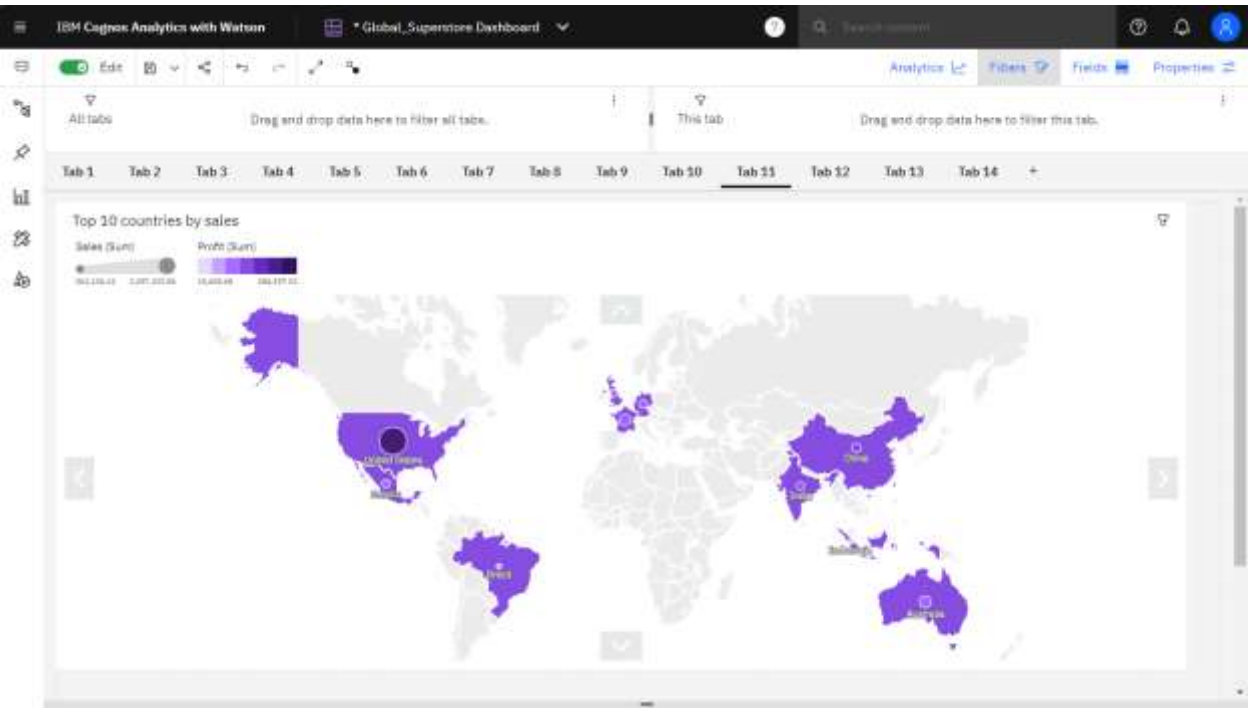

IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
Region
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Region
Search
Africa
Canada
Caribbean
Central
Central Asia
EMEA
East
North
North Asia
Oceania
South
Apply
Sales by Profit by Sub-Category
Profit (Sum)
Sales (Sum)

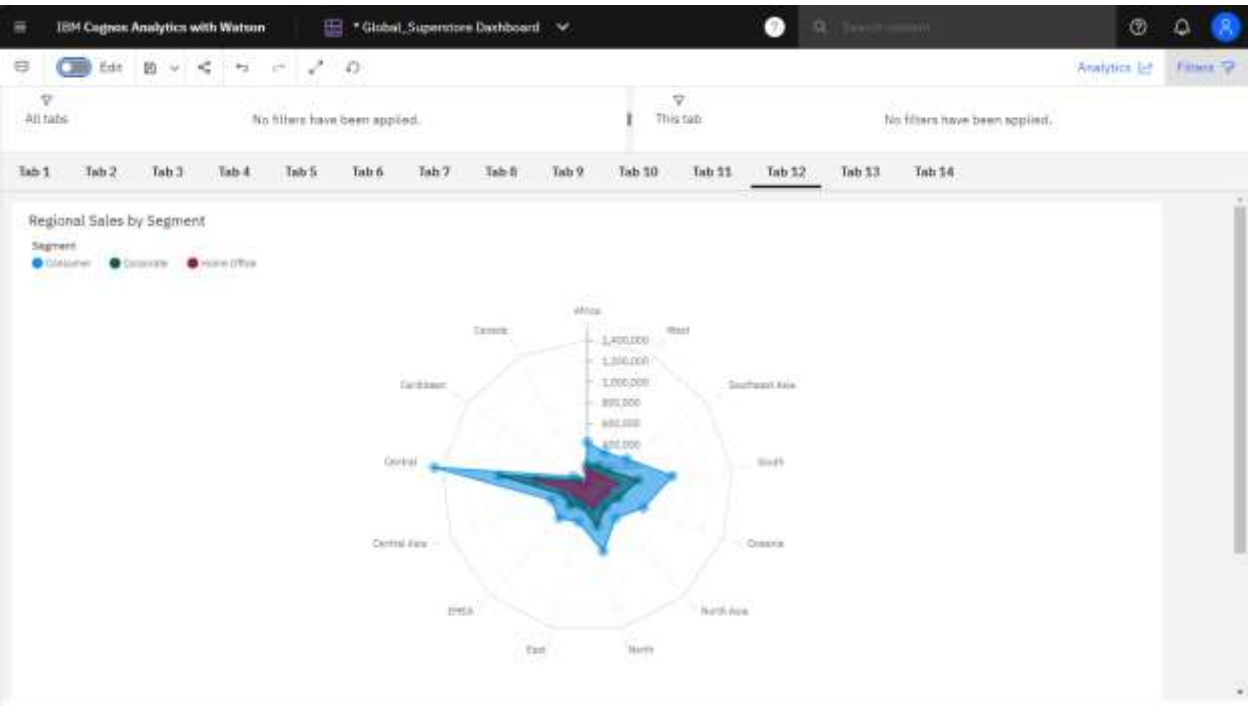

IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
Region
West
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Region
Search
Central
Central Asia
EMEA
East
North
North Asia
Oceania
South
Southeast Asia
West
Apply
Sales and Profit by Order_Month colored by Region
Forecast
Region - Measures
West | Sales
West | Profit
Values
Order_Month


IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
Order Priority
Critical, High, Medium, Lo
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Order Priority
Search
Critical
High
Low
Medium
Sales and Profit by Order_Month
Forecast
Measures
Sales
Profit
Values
Order_Month

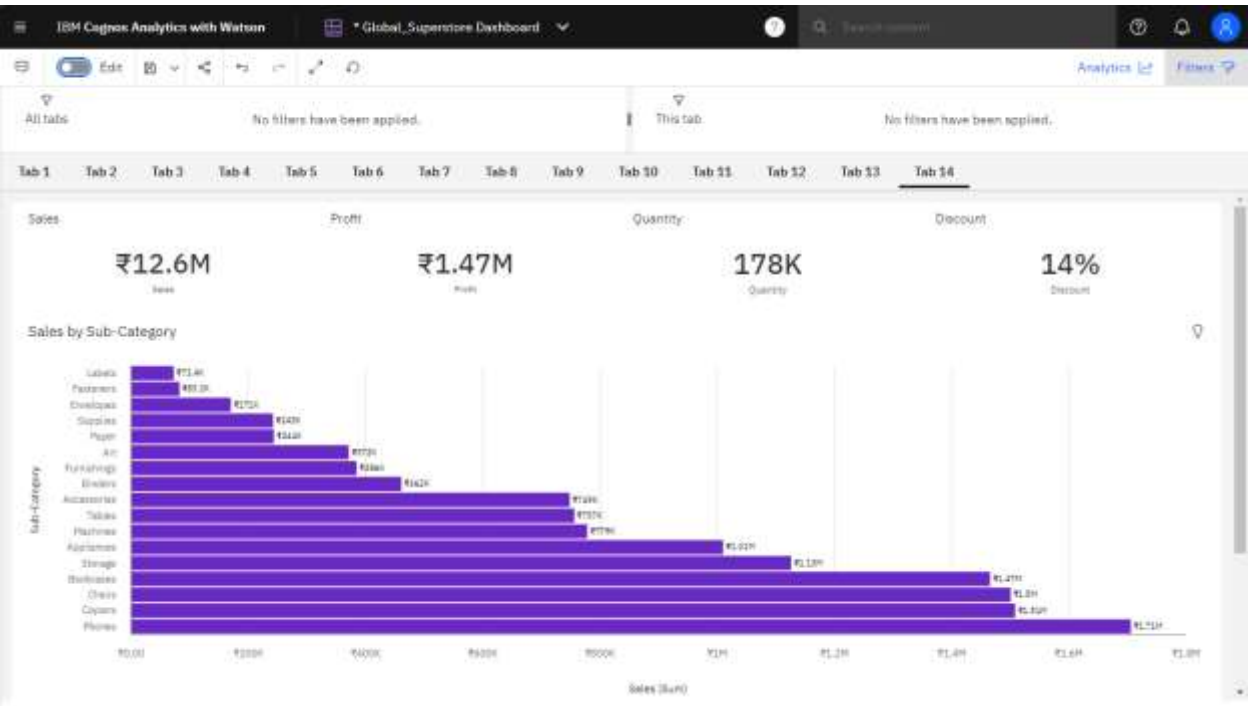

IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Sales by Sub category
Sub-Category

TAB 10 TBD……


IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
Fields
Properties
All tabs
Drag and drop data here to filter all tabs.
This tab
Drag and drop data here to filter this tab.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Top 10 countries by sales
Sales (Sum)
Profit (Sum)


IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Regional Sales by Segment
Segment
Consumer
Corporate
Home Office
Africa
Canada
West
Caribbean
Southeast Asia
Central
South
Central Asia
Oceania
EMEA
North Asia
East
North
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000


IBM Cognos Analytics with Watson
* Global_Superstore Dashboard
Edit
Analytics
Filters
All tabs
No filters have been applied.
This tab
No filters have been applied.
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Tab 7
Tab 8
Tab 9
Tab 10
Tab 11
Tab 12
Tab 13
Tab 14
Country wise Sales
Sales (Sum)
Region
Africa
Canada
Caribbean
Central
Central Asia
EMEA
East
North
North Asia
Oceania
South
Southeast Asia
West
Italy
India
Australia
United Kingdom
United States
United States
Germany
Mexico
China
United States
Indonesia
France
Spain
Brazil
Philippines
Iran
Egypt
Honduras
Bangladesh
Saudi Arabia
Guatemala
Argentina
New Zealand
Vietnam
Pakistan
Ukraine
Nicaragua
El Salvador
Cuba
Thailand
Turkey
Dominican Republic
Colombia
Singapore
Netherlands
Japan
South Africa
Malaysia
Morocco
Russia
Poland
Belgium
Afghanistan
Democratic Republic of the Congo



9.1 Performance Metrics

S.No	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs - 29
2.	Data Responsiveness	No of Visualizations / Graphs - 29
3.	Amount of Data to be Rendered (DB2 Metrics)	No of Visualizations / Graphs - 29No of Visualizations / Graphs - 29
4.	Utilization of Data Filters	Utilized to full effectiveness
5.	Utilization of Data Filters	No of Scenes Added - 7
6.	Descriptive Reports	No of Visualizations / Graphs - 10

10. ADVANTAGES &

DISADVANTAGES Advantages

- *Data Visualizations*
- *Ease of use*
- *Integration capabilities*

Disadvantages

- *No prediction features available as of yet*
- *Need to improve security aspect of the product*

11. CONCLUSION

To conclude, we will say that we are providing revolutionary solutions and insights for businesses and making their job a lot easier.

12. FUTURE SCOPE

We can improve our services that we offer by implementing more prediction based insights. Also, we can further exploit other domains as well and offer analytics services to them

13. APPENDIX

GitHub - https://github.com/IBM-EPBL/IBM-Project-30404-1660145997

Project Demo - https://drive.google.com/file/d/1r5Sz68_-rEA_XR0KG2FYcakgZRj_2aN-/view?usp=sharing

THANK YOU