TEAM ID : PNT2022TMID03233

REPORT

GLOBAL SALES DATA ANALYTICS

1. INTRODUCTION

1.1 Project Overview

Data is being generated very rapidly due to an increase in information in everyday life. Huge amount of data gets accumulated from various organizations that is difficult to analyze and exploit. Processing, analyzing and communicating this data are a challenge. Online shopping websites get flooded with voluminous amounts of sales data every day. Analyzing and visualizing this data for information retrieval is a difficult task. Therefore, a system is required which will effectively analyze and visualize data. This paper focuses on a system which will visualize sales data which will help users in applying intelligence in business, revenue generation, and decision making, managing business operation and tracking progress of tasks. By using IBM Cognos Analytics and the global sales data we are going to identify patterns, relationships, connections using dataset, exploring relationships in the data, and visualizing the data.

1.2 Purpose

The purpose of this paper is to provide a conjecture and guesswork of events and it helps to find answers that can be sufficiently disguised for a particular problem to come up with an optimal conclusion and a convincing solution. Based on analytics, we may determine which orders, items, and products were most likely purchased by customers in the past, and we can project future sales based on the profit information that was examined by the provided data.

2. LITERATURE SURVEY

2.1 Existing problem

Despite the fact that there are numerous existing ways to forecast insights, they are not very effective at predicting the right solution to the issue. The expected insights might not always be accurate and lack the necessary precision.Even if our devices provide us with a lot of data, we still need to compare, evaluate, and interpret it in order to use it..

2.2 References

1. Data mining with its role in marketing, sales support and customer identification data analysis [Mohammed Bin Ali Al Atif, Ahmed H. Shakir, et al, 2022]

In today's technologically advanced age, every company wants to equip its sales force with a sustainable sales force automation system to improve sales performance and customer relationship capabilities. This study examines the impact of big data analytics on the sales performance of the organizations. Current advances in information technology and the development trend of social networks have had changed the way salespeople perform their daily activities. Data related to customer buying behavior is being generated at an unprecedented rate due to the technological revolution and the advent of sources such as social networks. Sales performance helps to efficiently and effectively achieve sales process goals by looking at opportunities and improving close rates. The data analytics established in this study as a technology or system provides useful insights into customer behavior by uncovering hidden patterns in BD to aid in the development of effective strategies for sales. In the era of the big data revolution, the method of strategy formulation in sales has changed, and organizations need to use data analytics systems to meet their needs. Individual characteristics is said to be the individual perception of big data analytics. In order to improve the objectivity of the comparison results, companies can add other models to participate in the comparison, so as to obtain accurate data analysis results. Data analytics is of great significance in this era of data overflow, and can provide unforeseen insights and benefits to decision makers in various areas. If properly exploited and applied, big data analytics has the potential to provide a basis for advancement. By applying such analytics to the data, valuable information can be extracted and exploited to enhance decision making and support informed decisions.

2. Impact of big data analytics on sales performance in pharmaceutical organizations: The role of customer relationship management capabilities [Muhammad Shahbaz, Lili Zhai, et al, 2021]

The technique of studying raw data to conclude a specific piece of information is known as data analytics. It is employed to assist people and organizations in making sense of data. They are applied to the analysis of raw data to discover trends and insights. We can infer conclusions about the information they contain by looking at select datasets and identifying trends. Data analytics is carried out using specialized hardware and software. These tools and methods are frequently employed in a variety of commercial sectors to empower businesses to take wise business decisions. Additionally, the analytics give companies the ability to react quickly to changing market trends and acquiring an advantage over rival companies. Various efforts can benefit from some of the components of this analytics process. A good data analytics initiative will give you a clear picture of where you are, where you have been, and where you should go by merging these elements. To improve corporate performance, however, is data analytics' ultimate objective. Depending on the specific application, the data that is evaluated may be made up of new data that has been processed for real-time analytics or historical records. For the most effective data manipulation, data analytics uses a variety of software tools, including spreadsheets, data visualization, and reporting tools, data mining software, or open-source programming languages. Inside the data analytics process, the data analytics applications involve more than just analyzing data, especially on advanced analytics projects. After the data are analyzed, it will produce charts and other infographics that can be designed to make findings easier to understand. Data visualizations often are incorporated into BI dashboard applications that display data on a single screen and can be updated in real-time as new information becomes available.

3. Data Analysis and Visualization of Sales Dataset using Power BI [Ms. Sarika Singh, Ms. Lavina Jadhav, 2022]

Data analytics enables organizations to analyze all of their data to identify patterns and generate insights to inform and, in some cases, automate decisions by relating Smart and actionable. Today's best solutions support end-to-end analytics, from accessing, preparing and analyzing data to operating analysis and monitoring results. When analyzing data, the main task is defined objects to analyze and separate data time period analyzed, to ensure the eccentricity of data analysis results. Data is useless if it cannot be analyzed, understood and applied in context. A picture is worth a thousand words, and business analytics can help create a picture by visualizing data to provide retailers with business insights. With this information, businesses can make meaningful changes to their future plans to maximize profits and success. Most raw data, especially large-scale databases, are worthless in their unprocessed state. We can extract valuable insights from this bit store using Power BI tools. The main goal here is to read and

analyze the available data sets to generate business insights and overviews. The success of any organization, company or business depends on its business division as it is. The only part of the organization that earns revenue and money and delivers profits. The importance of selling is as follows: Sales data is a broad word that includes many types of metrics, but in general if You can measure something based on the sales process is the actionable sales data. Through visualization, data analysis helps students understand concepts. Much technology is available to perform business data analysis, but Power BI visualization technique is the most popular techniques to learn the basics of data analysis. With the help of visualization techniques, data interpretation and data representation can be done quickly and easily. This strategy is useful for a more solid conceptual design.

4. Survey on Growth of Business using Data Analytics for Business Intelligence in Real-Time world [Madamanchi Brahmaiah, Talluri Sreekrishna, 2021]

Data analytics strategies can screen developments and metrics data might in any other case be misplaced withinside the mass of facts. These facts can then be used to optimize procedures to growth the overall performance of a commercial enterprise or system. Data analytics is the technology of studying uncooked records to make conclusions approximately that facts. Many of the strategies and procedures of records analytics has been automatic into mechanical procedures and algorithms that paint over uncooked records for human consumption. Data analytics is the technology of reading uncooked statistics to make conclusions approximately data information. The strategies and approaches of statistics analytics were computerized into mechanical approaches and algorithms data paintings over uncooked statistics for human consumption. Data analytics assist a business optimize its performance. Companies everywhere in the international try and get the advantages from get entry to the statistics to improve their overall performance and boom their revenue, however processing heterogeneous varieties of information to extract the precious information is a massive hassle that many businesses try and solve. One of the most essential developments is “Big Data Analytics”, a generation for Storing, Processing, and analyzing the information, groups are Managing information to apply it in new ranges and direct decision-makers. Companies can use the insights they advantage from statistics analytics to tell their decisions, main to higher outcomes. Data analytics removes lots of the guesswork from planning marketing campaigns, selecting what content material to create, growing products, and more. Data analytics additionally offers you beneficial insights into how your campaigns are appearing so that you could fine-track them for top of the line outcomes. Data analytics offer you with extra insights into your customers, permitting you to tailor customer support to their needs, offer extra personalization, and construct more potent relationships with them.

2.3 Problem Statement Definition

Customer Problem Statement Template:

Create a problem statement to understand your customer's point of view. The Customer Problem Statement template helps you focus on what matters to create experiences people will love.

A well-articulated customer problem statement allows you and your team to find the ideal solution for the challenges your customers face. Throughout the process, you'll also be able to empathize with your customers, which helps you better understand how they perceive your product or service.

Global Sales Data Analytics problem statement:

Problem Statement (PS)	Iam	I'm trying to	But	Because	Which makes Me feel
PS-1	Company	Discover a reliable analytical tool to enhance sales by making smarter business judgements	It can be challenging to find out the best tool.	There are several incorrect interpreted data available on the internet.	Torment
PS-2	Customer	Find the best product based on the previous reviews.	Suggestions and reviews are displayed irrespective of their personal choice and interests.	The new software tool used to analyze the customer data might not provide the right suggestion of the product.	Confused
PS-3	Customer	Analyze sales across multiple sources	Found many inaccuracies in sales	Sources are from different systems	Dissatisfied

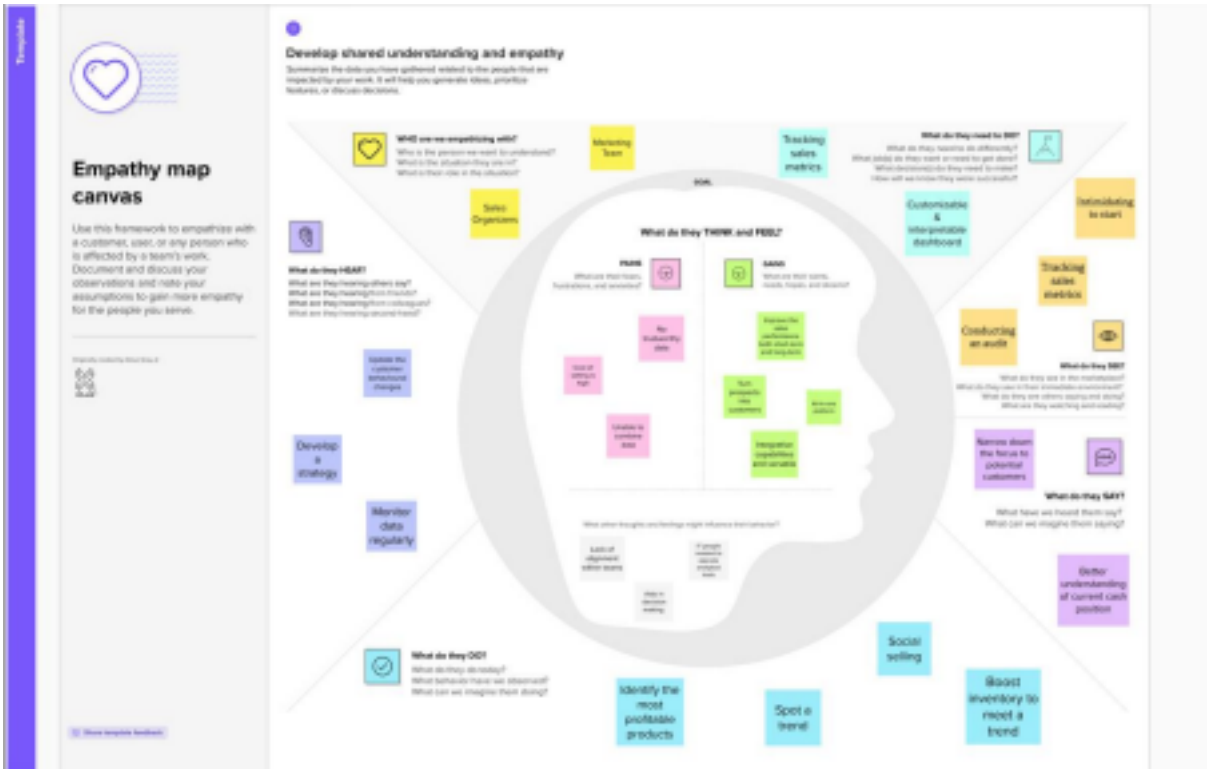
3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes.

It is a useful tool to help teams better understand their users.

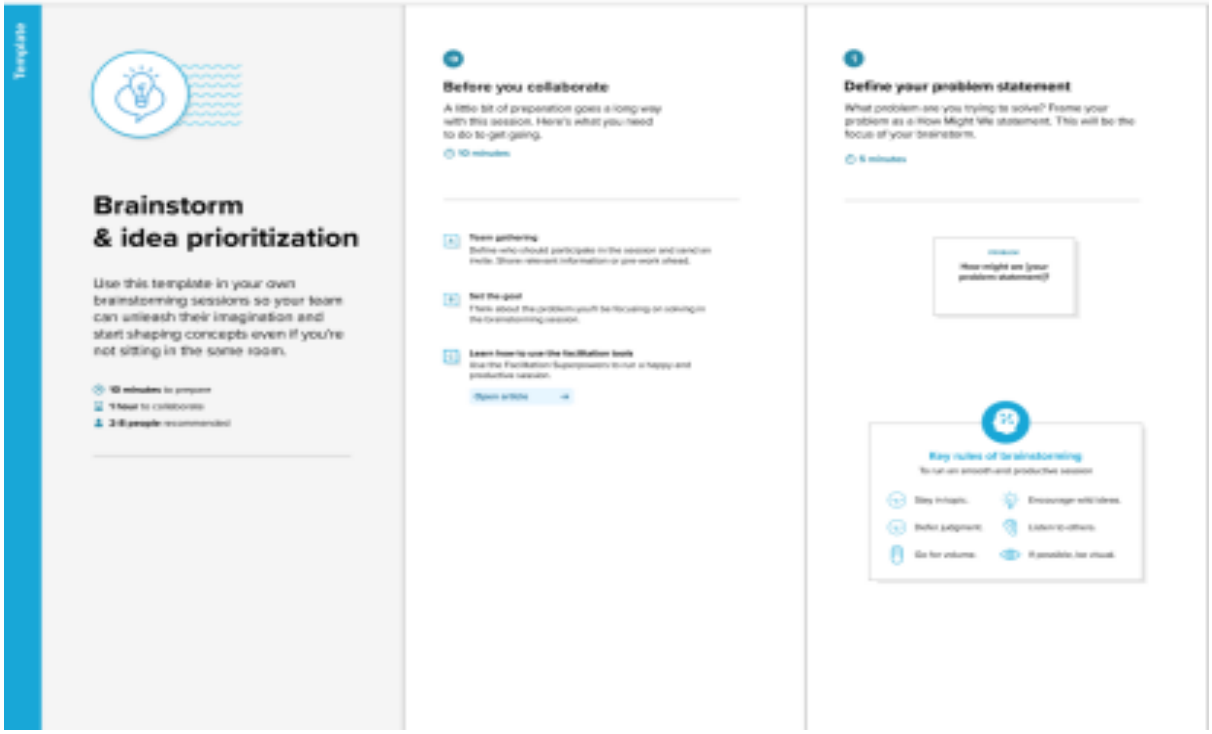
Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.



3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.



2

Brainstorm

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

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

Abinayaa

- Customer insights
- Easy Navigation tool
- Simple UI
- Interactive Dashboard
- AI based prediction

Shree raghav

- Easy accessible helpline and support
- Infographics
- Effective Tool
- Behavioral insights
- All variables and strength of connection

Tanmayee

- Individualize selling at scale
- Provide a list of variable option
- Meter reliability on predictions
- Customer layout for each end point
- Revenue acceleration

Sarath Kumar

- Quality Assurance
- Dynamic and real time
- Graphic view comparisons with competitors
- Long and short term solutions
- Headline followed by detail analysis

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

5

After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick additions

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template →

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

Share template feedback

3.3 Proposed Solution

Project team shall fill the following information in the proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Due to the pandemic's impact, online purchasing has become indispensable. Given the growing popularity of online shopping, it is crucial to examine how the industry is developing.
2.	Idea / Solution description	The idea is to build a dashboard using IBM Cognos for the visualizations, in which the user may understand the sales and product values. Machine learning algorithm is used to identify, optimize and forecast sales.
3.	Novelty / Uniqueness	Provide dominant insights, so that the user could make decisions on the recommendation Provided. Dynamic and real time analysis.
4.	Social Impact / Customer Satisfaction	The customer will understand the sales and purchases clearly. Can see the visible profit driven by informed decisions.
5.	Business Model (Revenue Model)	Stunning graphics and Impressive visualizations to create a successful marketing strategy. This enables the user to determine the actions that need to be made to improve the Business.
6.	Scalability of the Solution	The solution can work effectively on large and small datasets. User-friendly.

3.4. Problem Solution fit

Problem-Solution fit

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email
FR-3	Dataset	Upload and prepare the dataset provided.
FR-4	Visualizing data	Using IBM cognos analytics, create several visualization charts to analyze the trends in sales in the dashboard.
FR-5	Analysis	Generation report can be viewed by the users
FR-6	Log out	Logging out after analyzing the visualizations or even after downloading the dashboard.

4.2 Non-Functional requirements

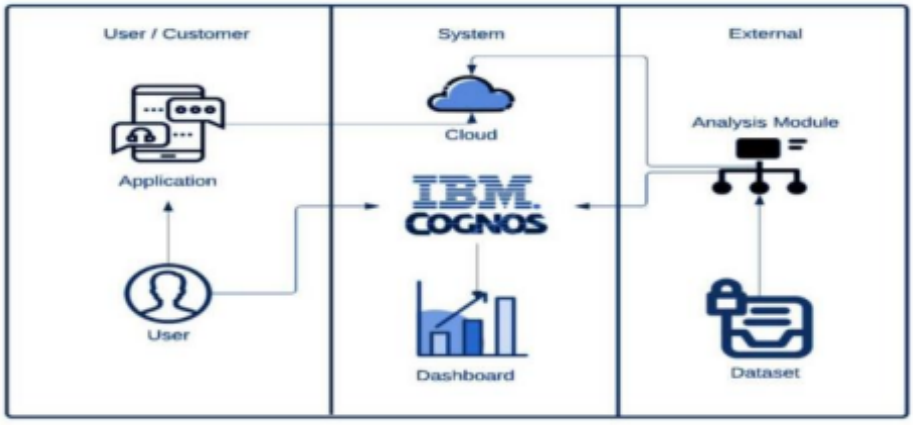
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	User Friendly website with interactive UI and UX.
NFR-2	Security	The user authentication and data is secured by end to end encryption.
NFR-3	Reliability	Will work without any glitches.
NFR-4	Performance	Due to the efficient use of data, performance and efficiency levels are high.
NFR-5	Availability	Available for the users 24 x 7 without any interruptions.
NFR-6	Scalability	Works with a high number of users and large datasets. Dashboard metrics can be change according to user demands.

5. PROJECT DESIGN

5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 Solution & Technical Architecture

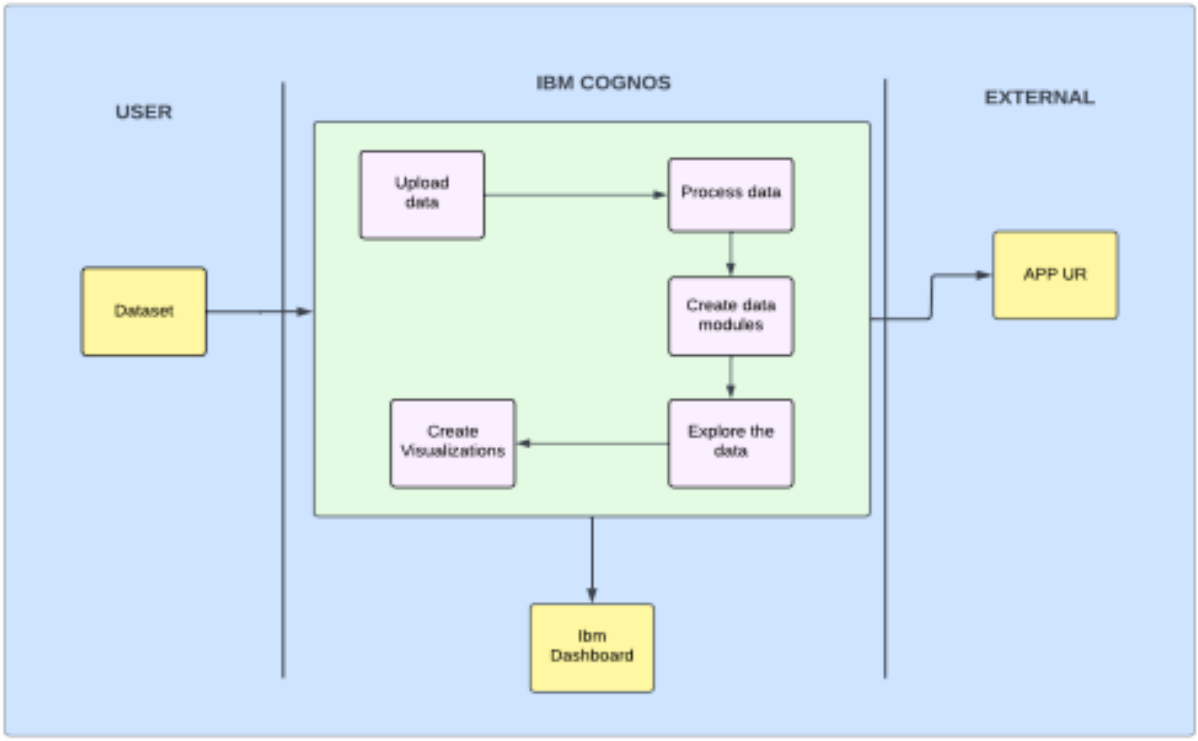


Table-1: Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	User interacts with application or Web	IBM Cognos analytics
2.	Dataset	Global sales dataset is uploaded	IBM Cognos
3.	Working with the dataset	Uploading, cleaning and processing the dataset	IBM Cognos analytics with Watson
4.	Data Exploration	Uploaded data is explored to identify trends	IBM Cognos
5.	Data Visualization	Various data are represented in charts, graphs according to the needs of the customers.	IBM Cognos

6.	Cloud database	Database service on Cloud	IBM DB2, IBM Cloudant
7.	Viewing data	User logins to the application to view the visualizations.	IBM Cognos Dashboard
8.	File Storage	File storage requirements	IBM Block storage or other storage services or local file system

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 different data sizes.	IBM Cloud
4.	Availability	Multiple reports are viewed irrespective of platform and device specifications	IBM Cognos
5.	Performance	withstand large amounts of data and process it without crashing.	IBM Cognos analytics with Watson

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-4	As a user , I can login to the application to view the dashboard	I can look into the insight	High	Sprint-2
Customer (Web user)	Registration	USN-5	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
	Login	USN-6	As a user, I can log into the application by entering email & password	I can add my profile photo and edit my account information.	High	Sprint-1
	Dashboard	USN-7	As a user, I can enter my sales data to clean and prepare it for analysis.		High	Sprint-2
		USN-8	As a user, I can identify trends in data and make visualizations.	I can edit and look into the insights	High	Sprint-2

Customer Care Executive	Communication	USN-9	As a customer care executive, I can answer user queries.	I can maintain good relationship with clients.	High	Sprint-3
Administrator	Chief Executive	USN-10	As an admin, I can take business driven decisions to improve the growth of the company.	Add or remove products	High	Sprint-2

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

TITLE	DESCRIPTION	COMPLETED DATE
Literature Survey & Information Gathering	Prepare Literature survey for the selected project & gathering information	22 OCTOBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	2 NOVEMBER 2022
Ideation	List them by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	22 OCTOBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	23 OCTOBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	23 OCTOBER 2022
Solution Architecture	Prepare a solution architecture document.	24 OCTOBER 2022
Customer Journey map	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	2 NOVEMBER 2022
Functional Requirement	Prepare the functional requirement document.	25 OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	25 OCTOBER 2022
Technology Architecture	Prepare the technology architecture diagram.	5 NOVEMBER 2022

Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	3 NOVEMBER 2022
Sprint delivery plan	Prepare the sprint delivery plan of the project	3 NOVEMBER 2022
Project Development - Delivery of Sprint-1	Develop & submit the developed code by testing it.	WORK IN PROGRESS.
Project Development - Delivery of Sprint-2	Develop & submit the developed code by testing it.	WORK IN PROGRESS
Project Development - Delivery of Sprint-3	Develop & submit the developed code by testing it.	WORK IN PROGRESS
Project Development - Delivery of Sprint-4	Develop & submit the developed code by testing it.	WORK IN PROGRESS

6.2 Sprint Delivery Schedule

Product Backlog, Sprint Schedule, and Estimation.

Use the below template to create product backlog and sprint schedule.

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	Abinayaa Shree Raghav Sarath Kumar Tanmayee
		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	Abinayaa Shree Raghav Sarath Kumar Tanmayee
		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	Abinayaa Shree Raghav Sarath Kumar Tanmayee
		USN-7	As a user, I can analyse the data by performing calculations and executing several visualization charts.	2	High	
		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	Abinayaa Shree Raghav Sarath Kumar Tanmayee
		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	

Project Tracker, Velocity & Burndown Chart:

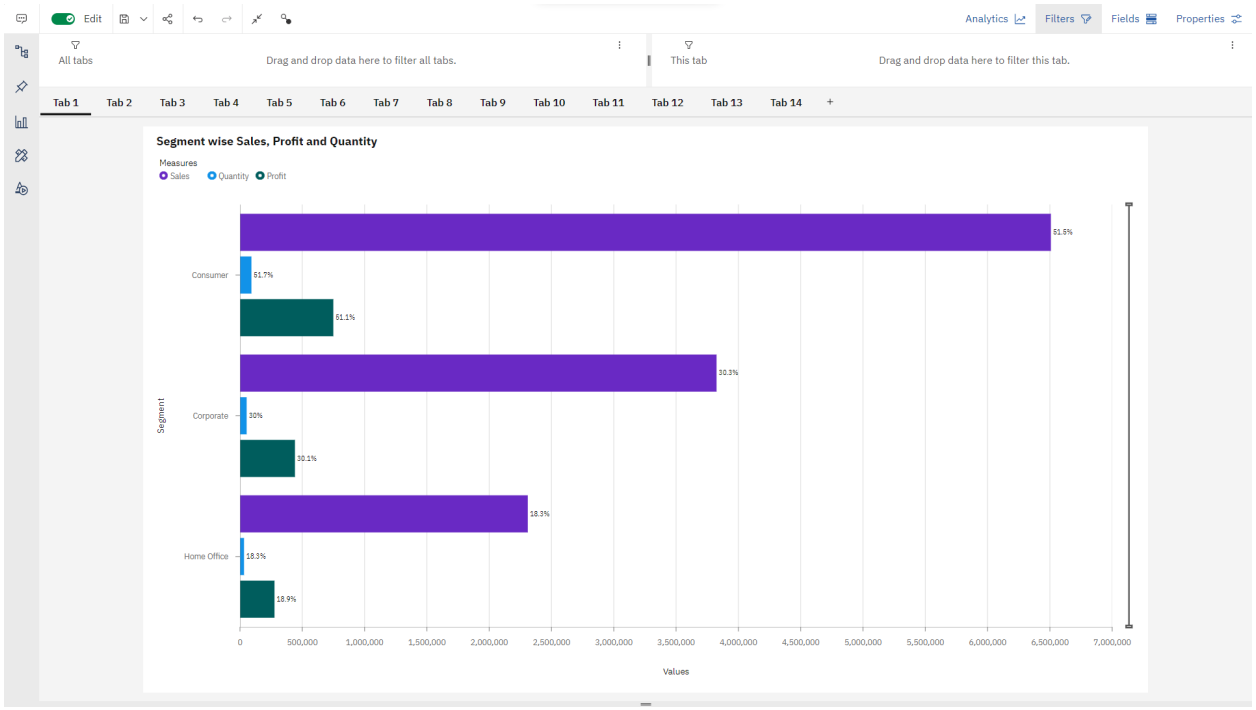
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	4	6 Days	04 Nov 2022	10 Nov 2022	4	10 Nov 2022
Sprint-2	4	6 Days	05 Nov 2022	11 Nov 2022	4	11 Nov 2022
Sprint-3	6	6 Days	06 Nov 2022	12 Nov 2022	6	12 Nov 2022
Sprint-4	4	6 Days	07 Nov 2022	13 Nov 2022	4	13 Nov 2022

7. CODING & SOLUTIONING

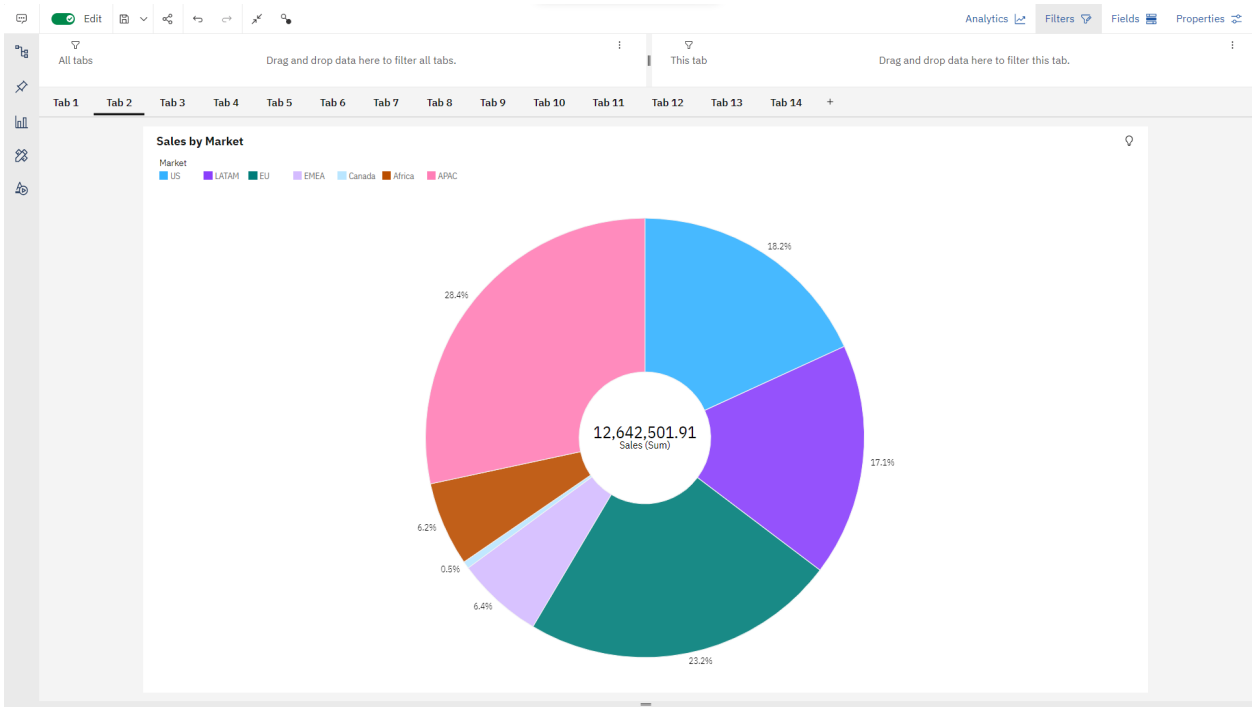
7.1 Feature 1

The dashboard is completed by performing the following visualization charts.

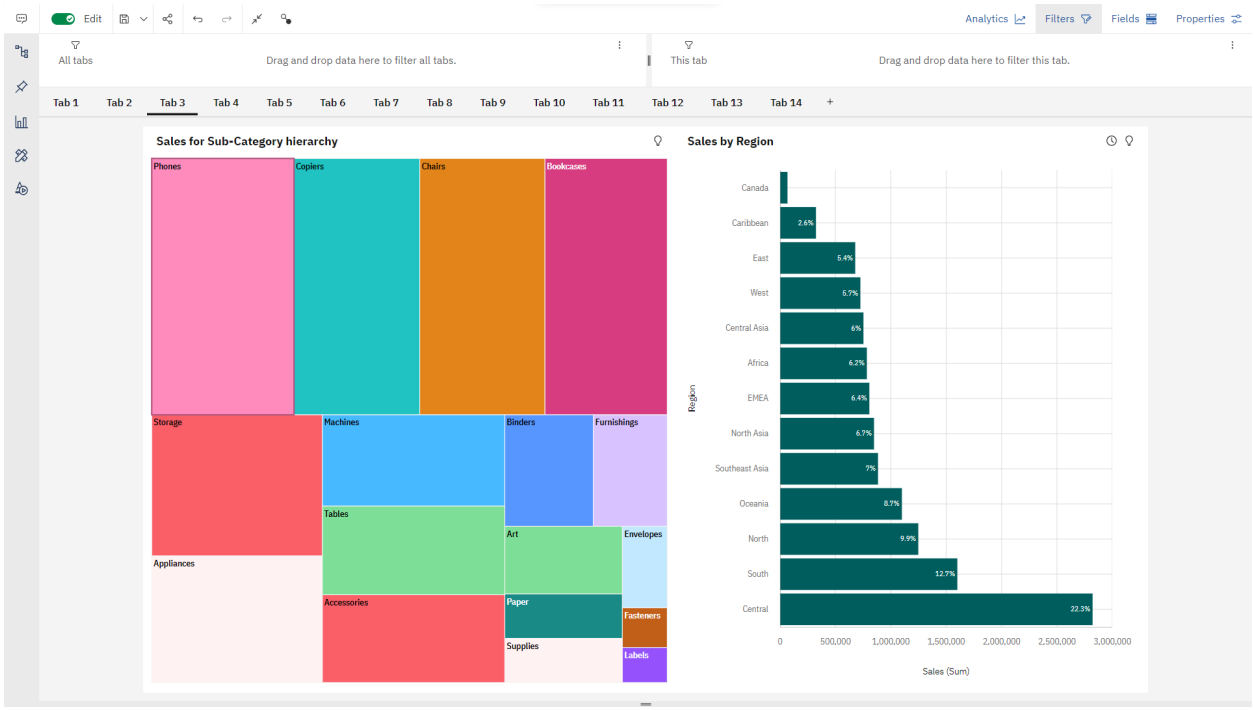
1. Segment wise Sales, Profit and Quantity using Bar Chart



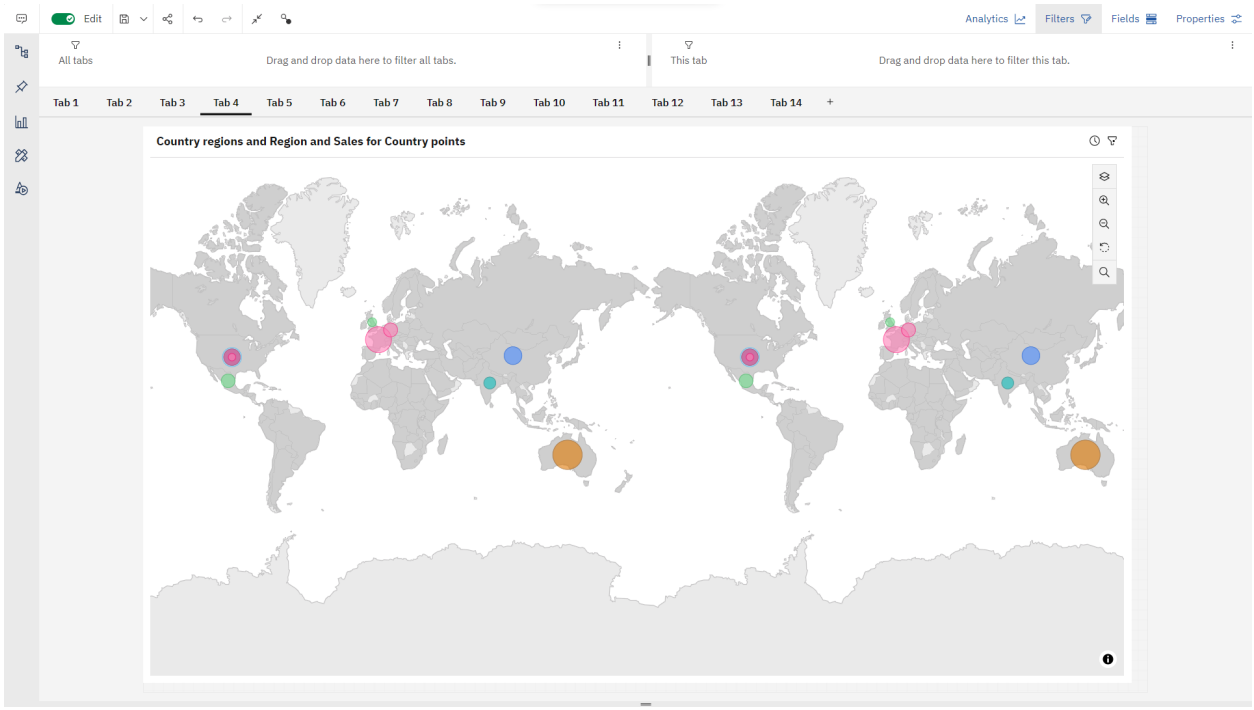
2. Sales by Market using Pie Chart



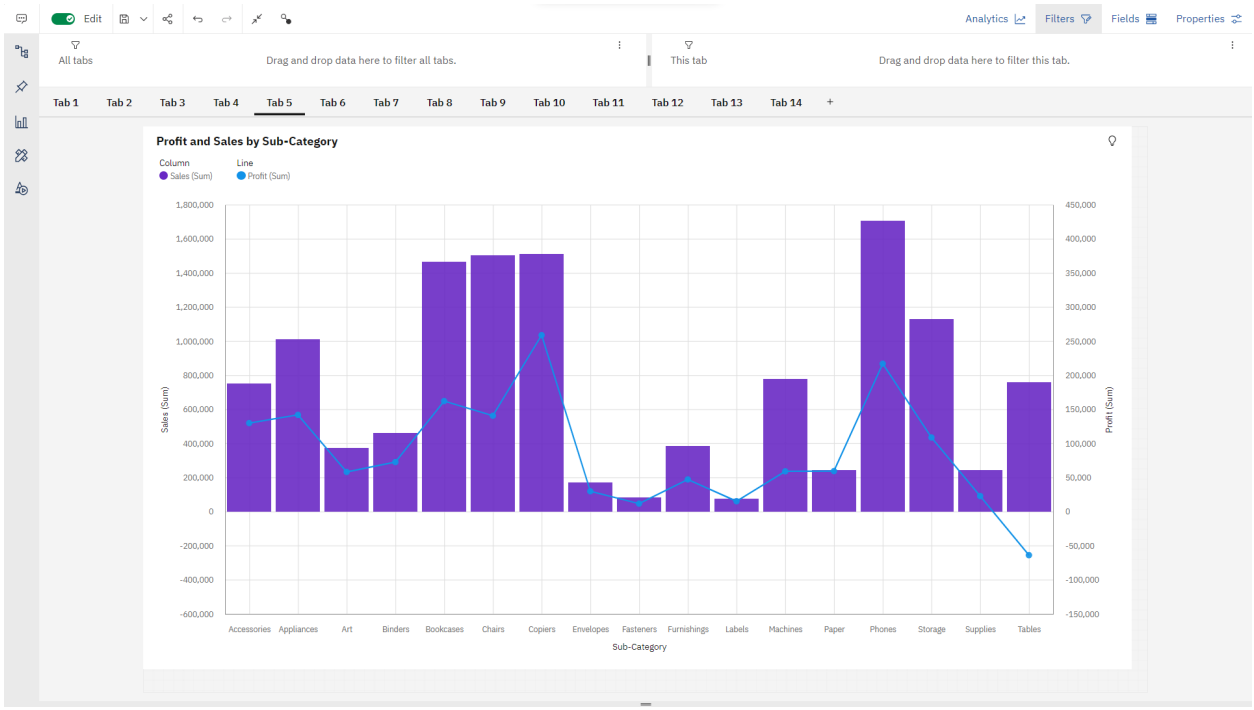
3. Sales by Sub-Category using Tree Map and Sales by Region using Bar/Column Chart



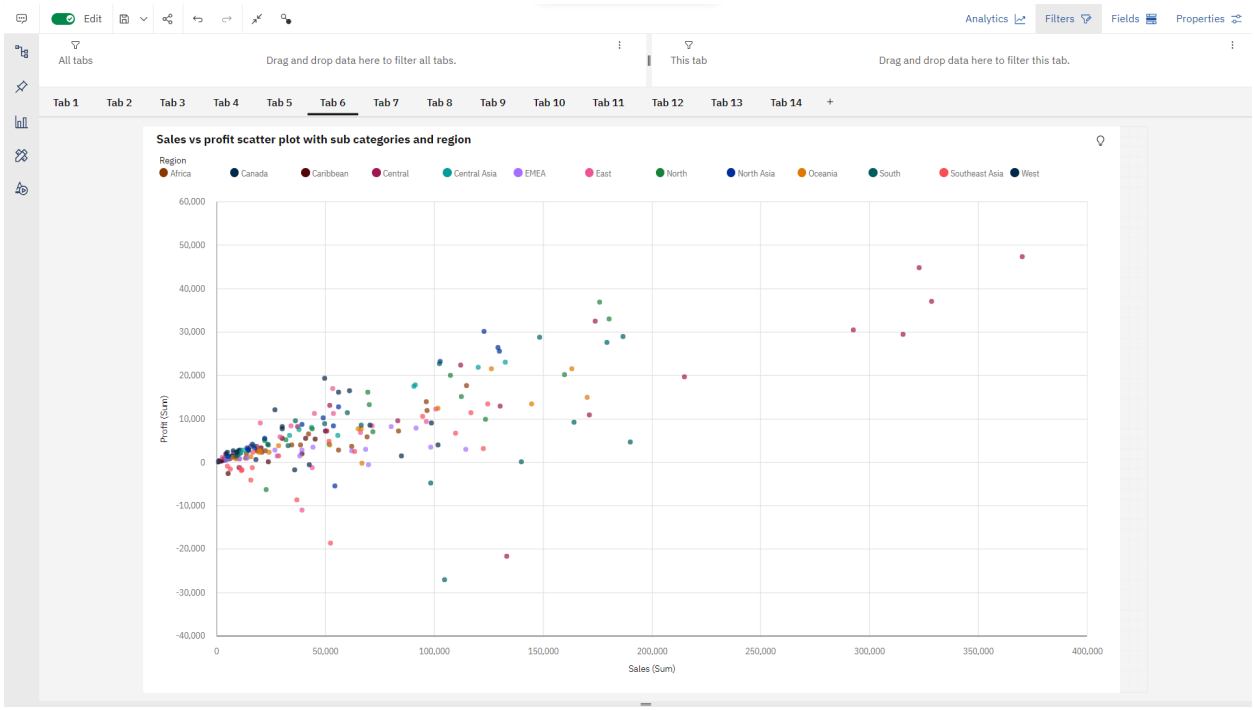
4. Top 10 Countries wise Sales using Map points



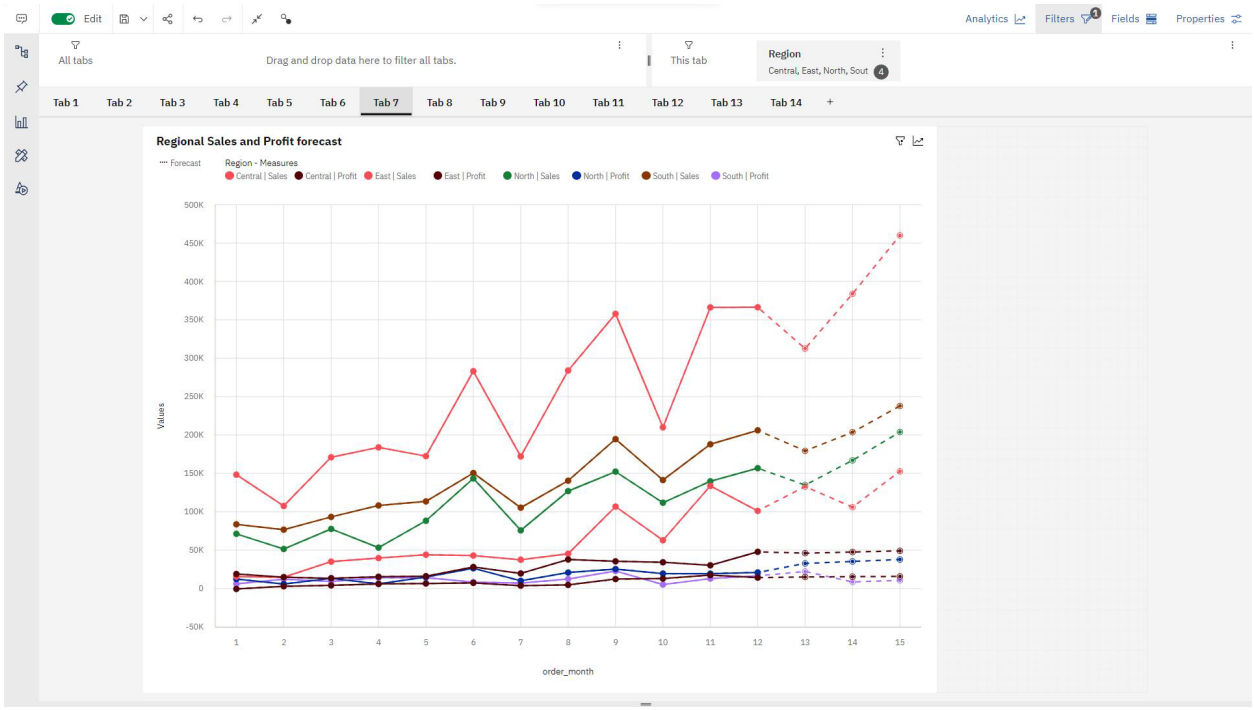
5. Sub-Category wise Sales and Profits using Line and Bar Chart



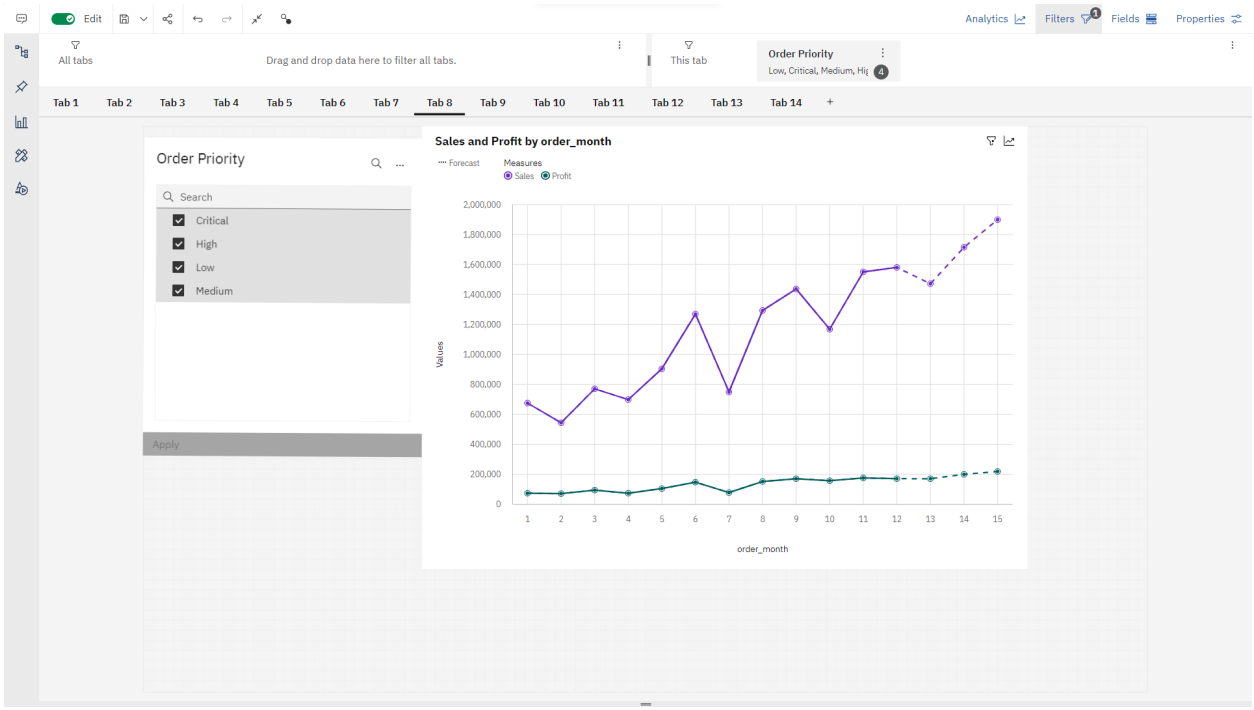
6. Sales vs. Profit Scatter plot with Sub-Categories and Regions



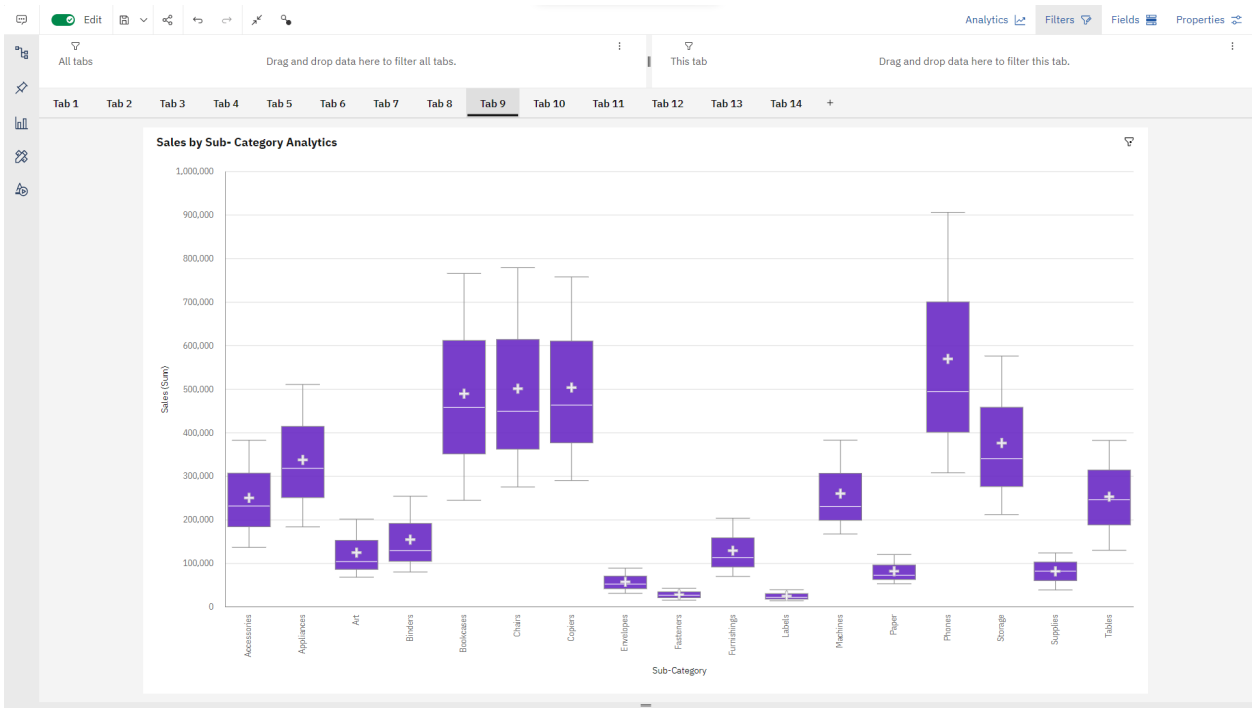
7. Regional Sales and Profit Forecast using Line Chart



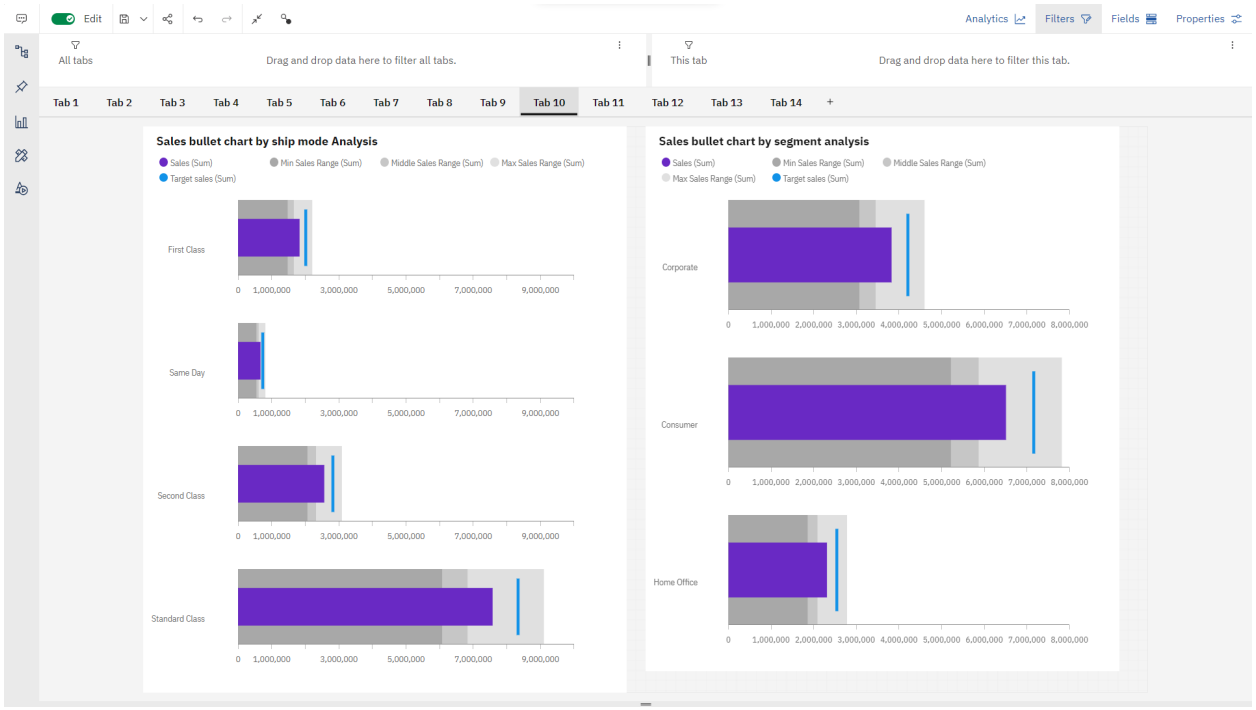
8. Sales Forecast by Order Priority using Line Chart



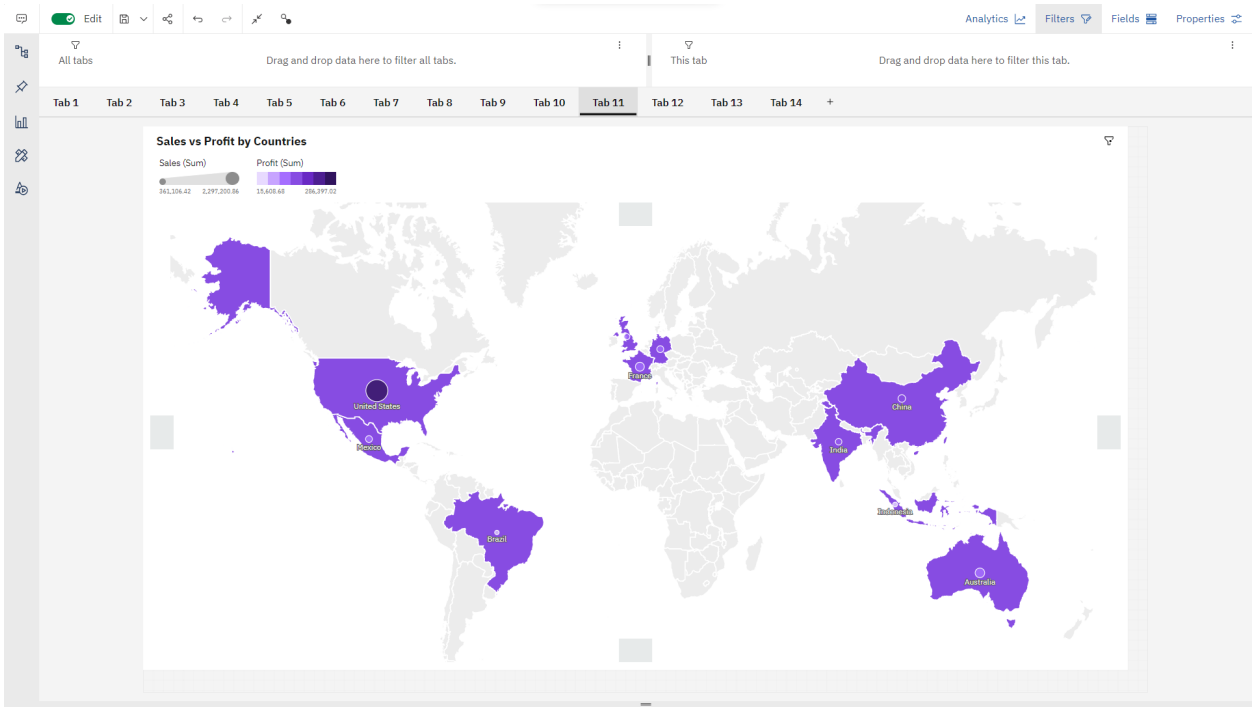
9. Sales by Sub-Category Analytics using Box Plot



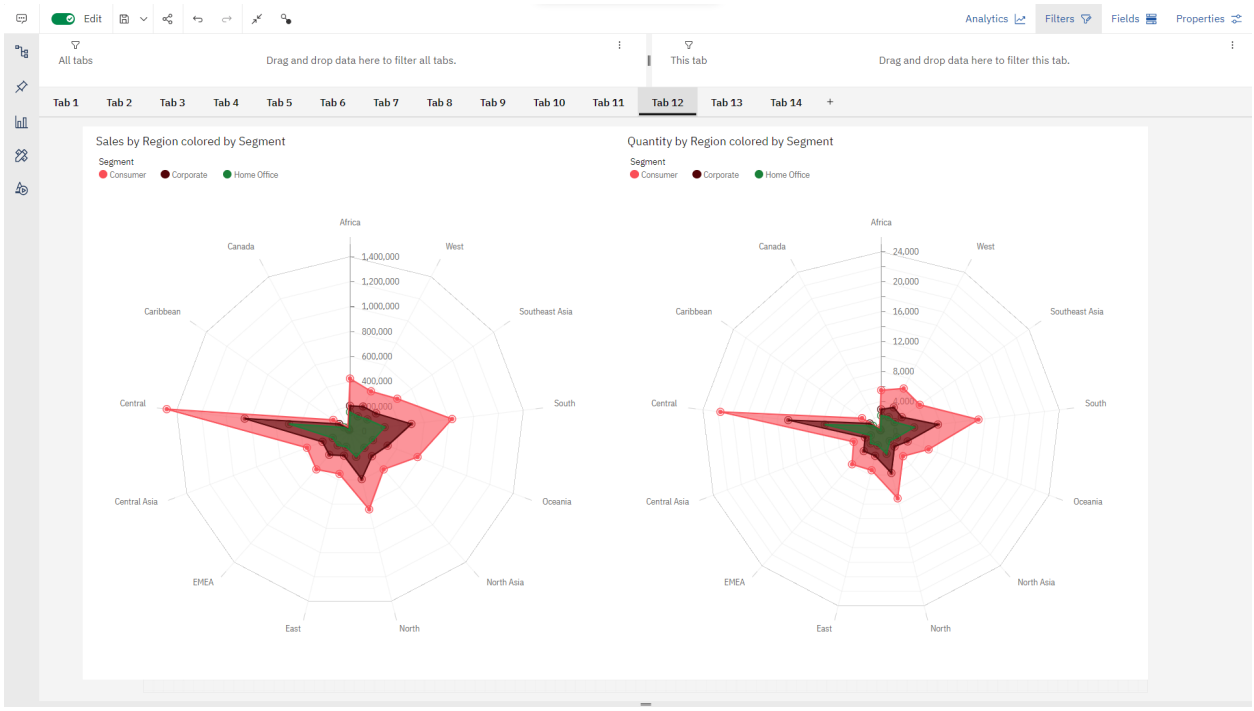
10. Sales by Segment Analysis using Bullet Chart



11. Sales vs. Profit by Countries using Legacy Map

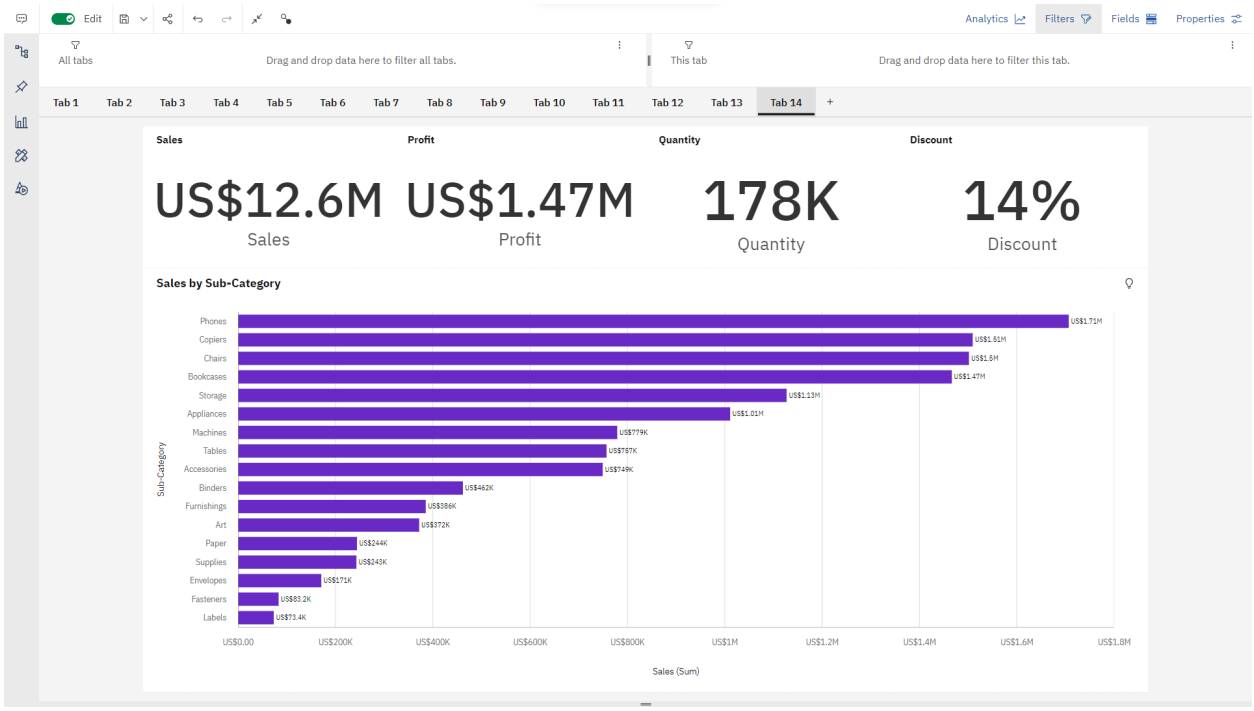


12. Regional Quantity and Sales using Radar Chart



13. Country wise Sales vs. Profit using Word Cloud

14. Sales Dashboard



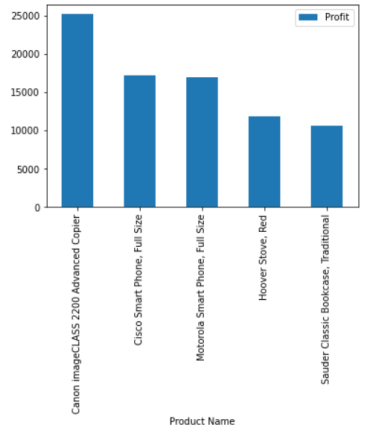
8. TESTING

8.1 Test cases:

```
#TOP 5 PRODUCT BY TOTAL PROFIT
```

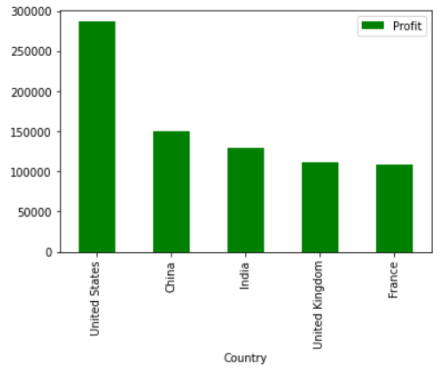
```
[ ] df.groupby(['Product Name']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.bar()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f856d2d0>
```



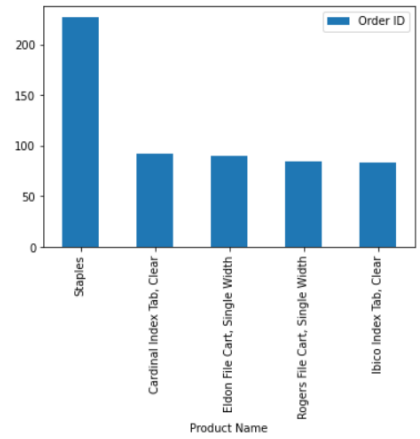
#TOP 5 COUNTRY BY TOTAL PROFIT

```
[ ] df.groupby(['Country']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.bar(color="green")
    plt.show()
```



#TOP 5 PRODUCT BY TOTAL ORDER

```
df.groupby(['Product Name']).count()[['Order ID']].sort_values(by="Order ID",ascending=False).nlargest(n=5, columns=['Order ID']).plot.bar()
plt.show()
```

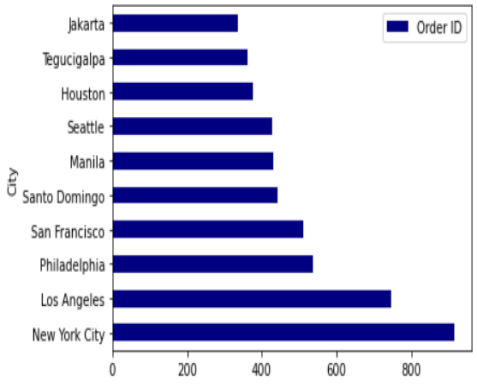


```
[ ]  df.groupby(['Product Name']).count()[['Order ID']].nlargest(n=5, columns=['Order ID'])
```

Product Name	Order ID
Staples	227
Cardinal Index Tab, Clear	92
Eldon File Cart, Single Width	90
Rogers File Cart, Single Width	84
Ibico Index Tab, Clear	83

#TOP 10 CITY BY TOTAL ORDER

```
[ ]  df.groupby(['City']).count()[['Order ID']].sort_values(by="Order ID",ascending=True).nlargest(n=10, columns=['Order ID']).plot.barh(color='navy')
     plt.show()
```



9. Result

9.1. Performance Metrics

The confusion matrix below shows the performance metrics of the machine learning model.

10. ADVANTAGES & DISADVANTAGES

Advantages:

- Data analytics helps an organization make better decisions
- Increase the efficiency of the work
- The analytics keeps you updated of your customer behavioral changes
- Personalization of products and services
- Improving quality of products and services

Disadvantages:

- Lack of alignment within teams
- Lack of commitment and patience
- Low quality of data
- Privacy concerns
- Complexity & Bias

11. CONCLUSION

Global sales analysis is a difficult task to analyze so our sales Dashboard helps the users visualize the Sales data, which is helpful for efficient decision-making and data analysis. Additionally, tracking which products are selling on a day-to-day basis helps our reps successfully cross-sell and upsell.

12. FUTURE SCOPE

- Due to fast growing population E-commerce business is also increasing
- Showing dominant insights
- Predict Trends in future sales
- Vital role in digital world

Appendix

Github : https://github.com/IBM-EPBL/IBM-Project-18820-1659690539

Project demo link :

https://drive.google.com/file/d/1FoyPn_9uo4i8i69q7Dwt6YSAdU9cUjCW/view?usp=share_link