

Project Report

INTRODUCTION

1. Project Overview

Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques like clustering and regression and apply them to more extensive datasets with the help of newer tools. Sales analytics refers to the technology and processes used to gather sales data and gauge sales performance. Sales leaders use these metrics to set goals, improve internal processes, and forecast future sales and revenue more accurately.

1.2 Purpose

Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques like clustering and regression and apply them to more extensive datasets with the help of newer tools. Sales analytics is a compass for your sales teams. It guides you towards growth using tools that track past data and current performance against goals. The analytics reports then help you plan targets in the future.” Sales involves many stages with their own set of metrics. Analytics has become an integral part of life, from finding the shortest route to work to forecasting stock market trends. Analyzing previous trends ensures that businesses always make the right decision. And as the scale of the decision and its impact magnifies, more robust analytics need to take over. The gut feeling cannot cut it anymore.

1. LITERATURE SURVEY

PAPER #1

Title:

Impact of big data analytics on sales performance in pharmaceutical organizations

Year:

April 2021

Briefing:

This study is related to the quantitative research method. This method emphasizes the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques.

Advantages:

Sales performance effectively and efficiently achieves the targets in the sales process by examining opportunities and improving closing rates

Disadvantages:

The data were collected from a developing country, and the results of this research may be different in developed countries.

PAPER #2

Title:

Effects of 3D Virtual “Try-On” on Online Sales and Customers’ Purchasing Experiences

Year:

September 2020

Briefing:

1) VTO technology consists of making a virtual body model from the customer's own body size, 3D garment modeling, 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 CAD patterns, we use 3D garment photos; to make a 3D garment mesh surface because 2D CAD patterns are always copyrighted and require professional CAD software, therefore our study proposes an efficient way.

Advantages:

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.

Disadvantages:

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.

PAPER #3**Title:**

Visual Analytics for Decision Support: A Supply Chain Perspective

Year:

June 2021

Briefing:

1. Visualization types suitable for particular analytical goals in each of the SC activities, i.e Visualization Techniques.
2. Analytical reasoning by analysts involved in the process of analysis, i.e., tactics.

Advantages:

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 identifying the state of the art in visualization techniques and tactics has been resolved.

Disadvantages:

First, the application of VA in supporting the source and make processes of the SC may be explored, such as cost modeling of different sources and production scheduling.

Second, visualizing the impact of external variables on different SC operations such as the effect of weather forecast on sales may be investigated more extensively.

PAPER #4

Title:

Sales Analytics and Big Data Developments Needed Now to Address Practitioner-Identified Emerging Biopharmaceutical Sales Force Strategic and Operational Issues

Year:

March 2021

Briefing:

The association of cancer drug costs and total cancer treatment costs per cancer site uses

two methods: Kaplan-Meier Sample Average (KMSA) method and an approach similar to the Cox proportional hazard model.

Generalized Propensity Score (GPS) based weighting with bootstrap standard errors can be used to estimate the marginal effect of detailing on drug utilization.

Advantages:

This paper provides a commercial/sales analytics, big data management, and organizational blueprint for companies on how to prepare and operate successfully in this evolving sales force pharma landscape.

Focuses largely on patient, payer, sales and marketing analytics.

Disadvantages:

Survey respondents from biopharmaceutical consulting companies also noted expertise focus is not in the areas identified here required for the industry to respond effectively to changing environmental trends.

References

1. Impact of big data analytics on sales performance in pharmaceutical organizations

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

2. Effects of 3D Virtual “Try-On” on Online Sales and Customers’ Purchasing Experiences

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

3. Visual Analytics for Decision Support: A Supply Chain Perspective

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

4. Sales Analytics and Big Data Developments Needed Now to Address Practitioner-identified

Emerging Biopharmaceutical Sales Force Strategic and Operational Issues

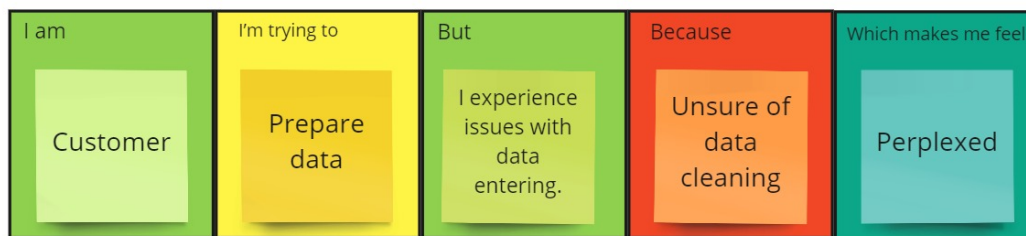
<https://www.pmsa.org/jpmsa-vol05-article01>

Problem Statement Definition

Customer Problem Statement:

Determine the crucial sales Key performance metrics you require, such as the win rate and average contract size. As leads move through your pipeline, use a solution (like Pipedrive's CRM) to track this information. Integrate this information in visual dashboards. Regularly compare the data to historical averages to track growth and problematic areas.

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Customer	Prepare data	I experience issues with data entering.	Unsure of data cleaning	Perplexed
PS-2	Customer	Know the region wise sales data	I'm unsure about the visualization	Sales data are inaccurate	Frustrated

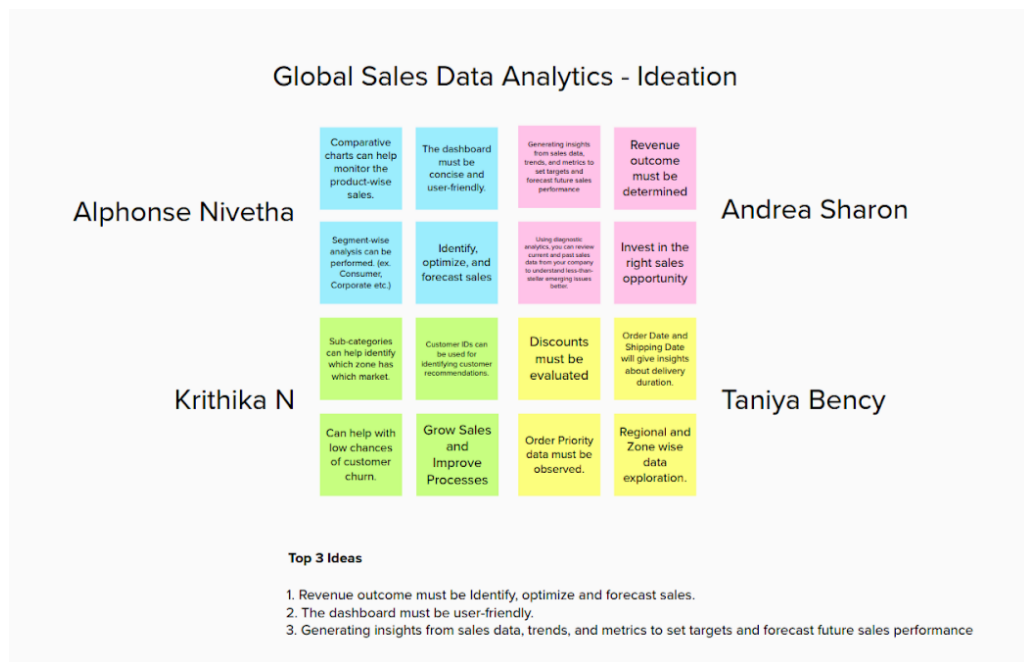


2. IDEATION & PROPOSED SOLUTION

Empathy Map Canvas



Ideation & Brainstorming



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)	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. Projects Aims T o Build an Analytical Dashboard For Global Sales Data Using IBM Cognos
2.	Idea / Solution description	Determine the crucial sales Key performance metrics you require, such as the win rate and average contract size. As leads move through your pipeline, use a solution (like Pipedrive's CRM) to track this information. Integrate this information in visual dashboards. Regularly compare the data to historical averages to track growth and problematic areas.
3.	Novelty / Uniqueness	Accurate Sales Forecasting Narrow and Refine Product Offerings Insights for improvement
4.	Social Impact / Customer Satisfaction	This idea is targeted towards the business people. The degree of interactive behaviour present in the dashboard can attract the customer into even reading tedious data. Plus, appealing visualizations can help them with better focus.
5.	Business Model (Revenue Model)	 <pre>graph LR; User((User)) -- 1 --> IBM[IBM COGNOS]; IBM -- 2 --> Data[(Data)];</pre>
6.	Scalability of the Solution	Can be useful in creating more visualizations thereby helping in gaining more insights in order to improve sales. (ex. Sales against the particular order priority)

Problem Solution fit

Project Title: Global Sales Data Analytics

Project Design Phase-I - Solution Fit

Team ID: PNT2022TMD27528

Define CS, J&P, BE and understand RC	<div>1. CUSTOMER SEGMENT(S)<div>CS</div><ul style="list-style-type: none">VendorsAdvertisersMarketing Department</div>	<div>6. CUSTOMER CONSTRAINTS<div>CC</div><ul style="list-style-type: none">Inaccuracy in terms of forecasting.Changing needs can be complex to update.Hesitation to change methodologies from the inferences.</div>	<div>5. AVAILABLE SOLUTIONS<div>AS</div><ul style="list-style-type: none">Region-wise Analysis of Sales data.Profits and Sales relationship on a sub-category basis.Insights as to where the sales succeed and for which category the reception is high.</div>	Explore AS, Differentiate
	<div>2. JOBS-TO-BE-DONE/PROBLEMS<div>J&P</div><ul style="list-style-type: none">Lack of understanding about the Dataset.Preparing and cleaning the dataset.Identifying the trends.</div>	<div>9. PROBLEM ROOT CAUSE<div>RC</div><ul style="list-style-type: none">Changing needs of people (needs an be based on the festival period, regional requirements etc.)Evolving styles and trends of products.</div>	<div>7. BEHAVIOUR<div>BE</div><ul style="list-style-type: none">Promotion of irrelevant product in an irrelevant region.Under/Over forecasting the sales-profit ratio in a particular country/region.</div>	
	<div>3. TRIGGERS<div>TR</div><ul style="list-style-type: none">Mistaken forecast leading to a loss in that area.Online sales profit over offline sales profit.</div> <div>4. EMOTIONS: BEFORE/AFTER<div>EM</div><ul style="list-style-type: none">Fear of having less knowledge about the people's need.Confidence if the expected profit is met in the anticipated region.</div>	<div>10. YOUR SOLUTION<div>SL</div><ul style="list-style-type: none">Visualizing the people's need in the region wise, season wise etc. and forecasting the sales of product which provide good profit.People's need is appropriately analyzed with their previous shopping.Usage of a dataset that handles all possible provision of insights.</div>	<div>8. CHANNELS OF BEHAVIOUR<div>CH</div><div>8.1 ONLINE</div><p>Online sales can be less challenging as the sales is global and the only challenge would be to deliver quality products within a stipulated time frame.</p><div>8.2 OFFLINE</div><p>Challenges in making a successful sale in that particular region.</p></div>	

3. REQUIREMENT ANALYSIS

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Proper medium to create the dashboard
FR-2	User uploading data (administrative)	To store the data set through the cloud
FR-3	End user benefits	Getting higher state of efficiency and also to know entire data analysis
FR-4	Site performance	For the dashboard to load the data module at a faster rate.
F4-5	Dashboard Creation	Visualization charts

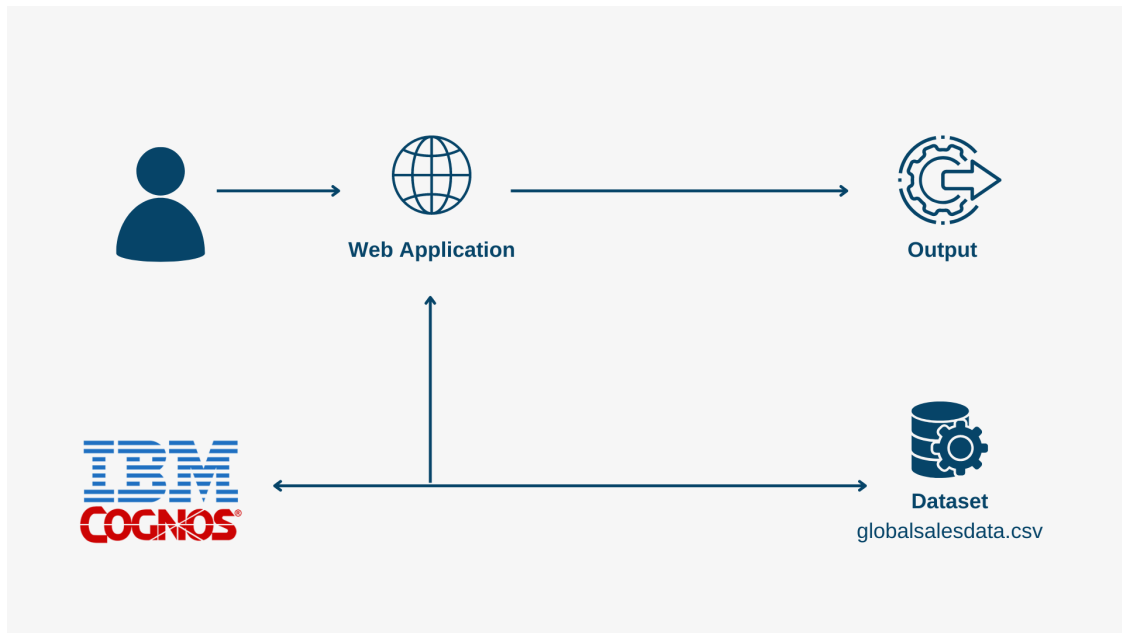
Non-functional Requirements:

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

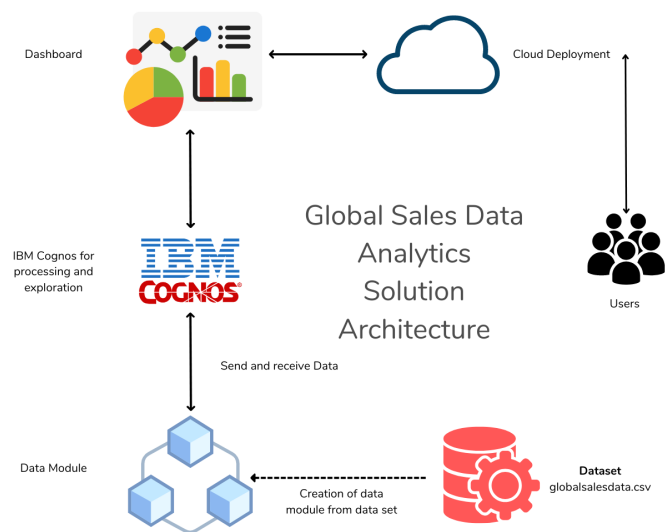
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Optimized resources and it can be used by everyone
NFR-2	Security	It has securable because it has end to end encryption
NFR-3	Reliability	It has high reliability based on development.
NFR-4	Performance	It has a high state of performance and efficiency.
NFR-5	Availability	It is available on all platforms and websites.
NFR-6	Scalability	The ability of a hardware and software parallel System to exploit increasing computing resources efficiency in the analysis of the (very) Large datasets

4. PROJECT DESIGN

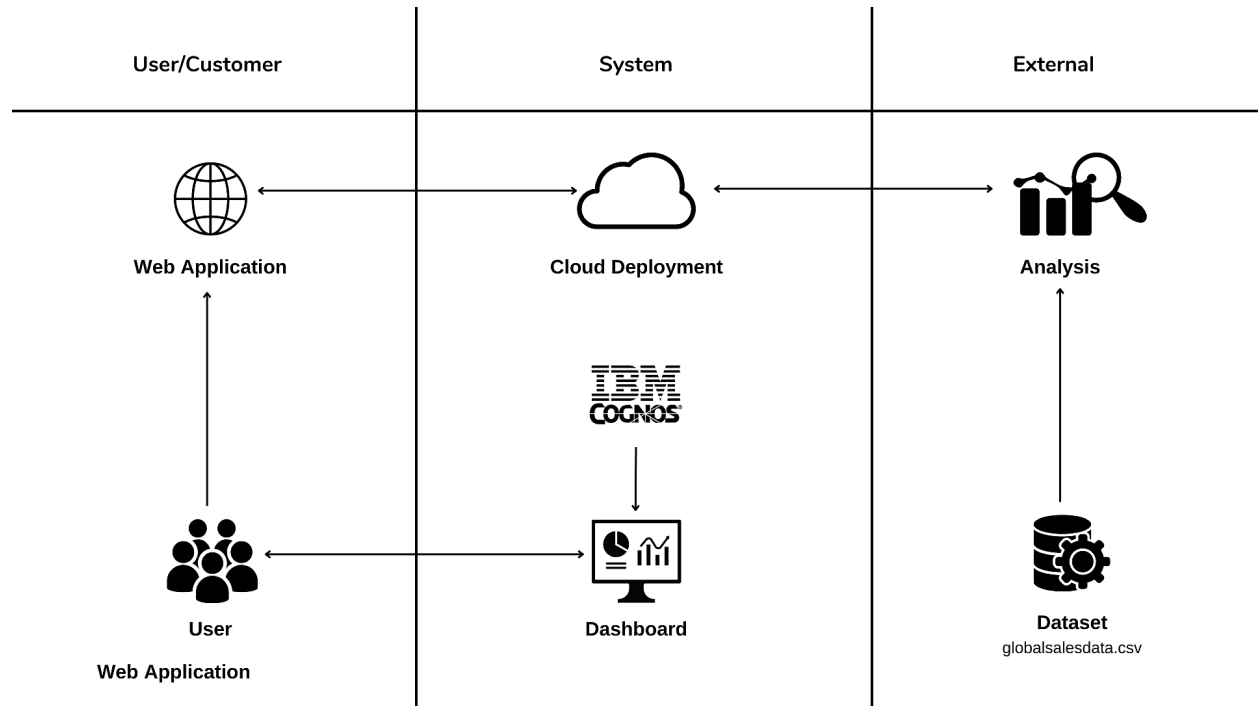
Data Flow Diagrams



Solution Architecture



Technical Architecture



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 I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1

	Login	USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
	Dashboard	USN-4	As a user, I can register for the application through Gmail	Access to dashboard is granted if my login is successful	Medium	Sprint-2
Customer Care Executive	Chat box	USN-5	As a user, I can log into the application by entering email & password	Access is always granted	High	Sprint-1
	Calling	USR-6	As a user, I can use calling option for any support or assistance help	Access is always granted only in working hours	High	Sprint-3
Administrator	Registration	USR-7	As an admin, I can register for the website by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-3
		USR-8	As an admin, I will receive confirmation email once I have registered for the website	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USR-9	As an admin, I can login for the website using my credentials	I can access my profile if credentials are correctly entered	High	Sprint-1
	Upload Dataset	USR-10	As an admin, I can upload the dataset on which the visualization has to be performed	Data should be clean and should not contain any duplicates or null values	High	Sprint-2

	Dashboard	USR-11	As an admin, I can customize dashboard if necessary	Changes in dashboard can be done only by the admin	High	Sprint-2
--	-----------	--------	---	--	------	----------

5. PROJECT PLANNING & SCHEDULING

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 need valid credentials to log in to my software.	5	High	Alphonse Nivetha, Andrea Sharon
Sprint 1	Data Preparation	USN-2	As a user, I need to upload my data to the software and the application must accept CSV/XLS files.	10	High	Andrea Sharon, Alphonse Nivetha, Taniya
Sprint 1	Data Preparation	USN-3	As a user, I must prepare my data in order to avoid discrepancies while visualizing.	5	High	Krithika, Taniya
Sprint 2	Dashboard	USN-4	As a user, I want filters so that I can view particular statistics of the worldwide sales data in relation to appropriate profitability.	7	Medium	Alphonse Nivetha, Taniya, Krithika
Sprint 2	Dashboard	USN-5	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.	7	Medium	Andrea Sharon, Krithika
Sprint 2	Dashboard	USN-6	As a user, I must be able to gain insights from the charts/graphs through a variety of relationships established in the dashboard.	6	Medium	Krithika, Taniya, Andrea, Alphonse Nivetha

Sprint 3	Integrating dashboard to webpage	US7	Integrating dashboard to webpage and deployment.	20	Low	Krithika
Sprint 4	Report & Story	-	Report & Story Preparation	20	Low	Alphonse Nivetha, Krithika

Sprint Delivery Schedule

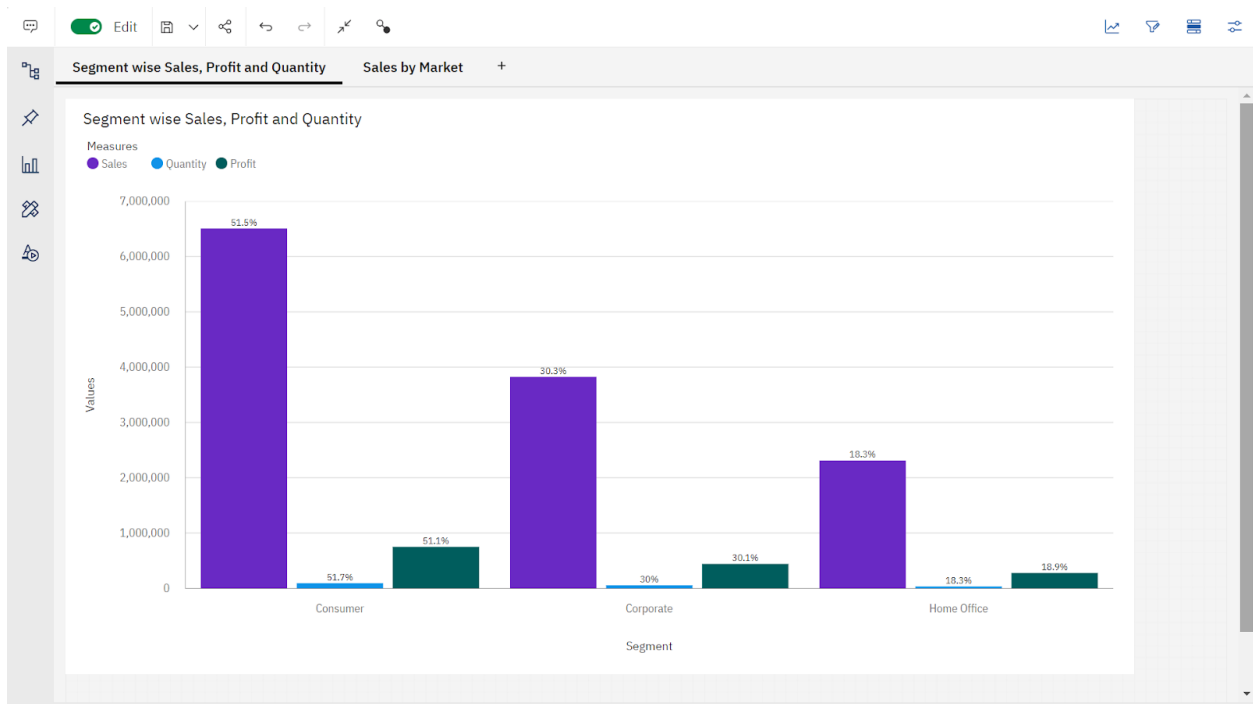
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29-10-2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	04-11-2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	15-11-2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	18-11-2022

6. CODING & SOLUTIONING

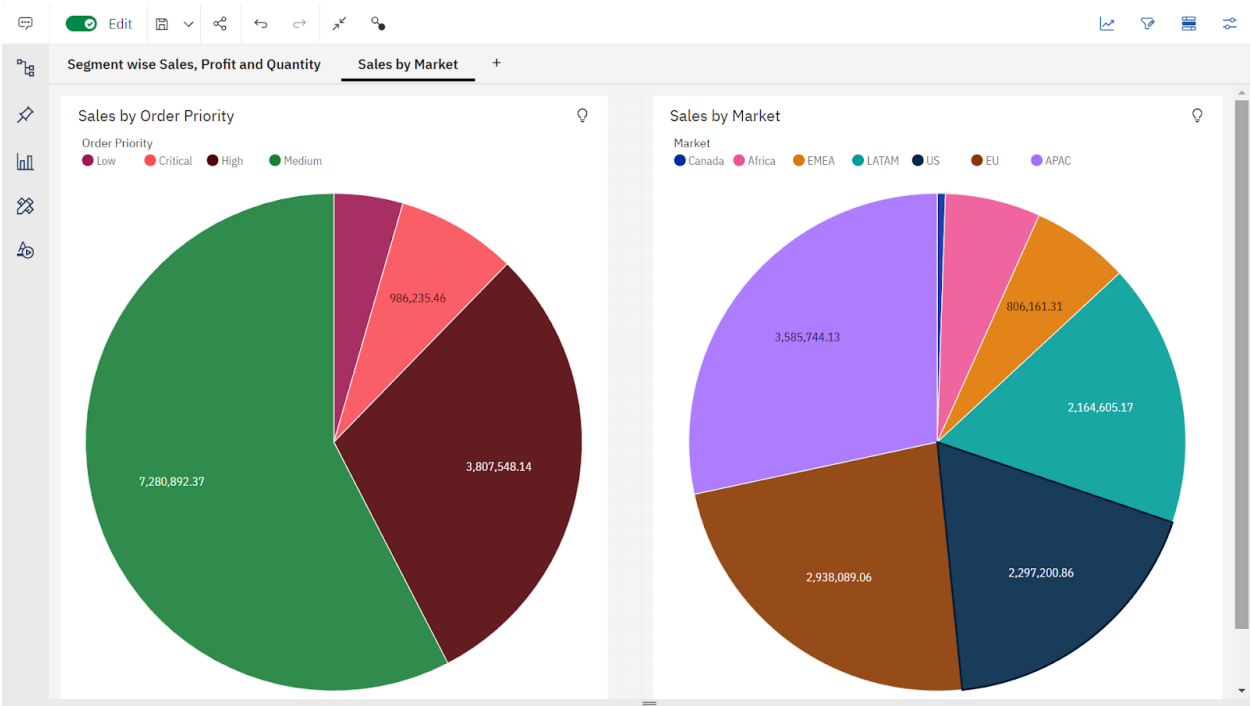
Dashboard with IBM Cognos

- Initially the dataset is loaded to IBM Cognos.
- A Data module is created and the data is explored.
- A Dashboard is created.
- Tabs covering several visualizations for several comparisons are created and enhanced.
- Inferences can be analyzed.

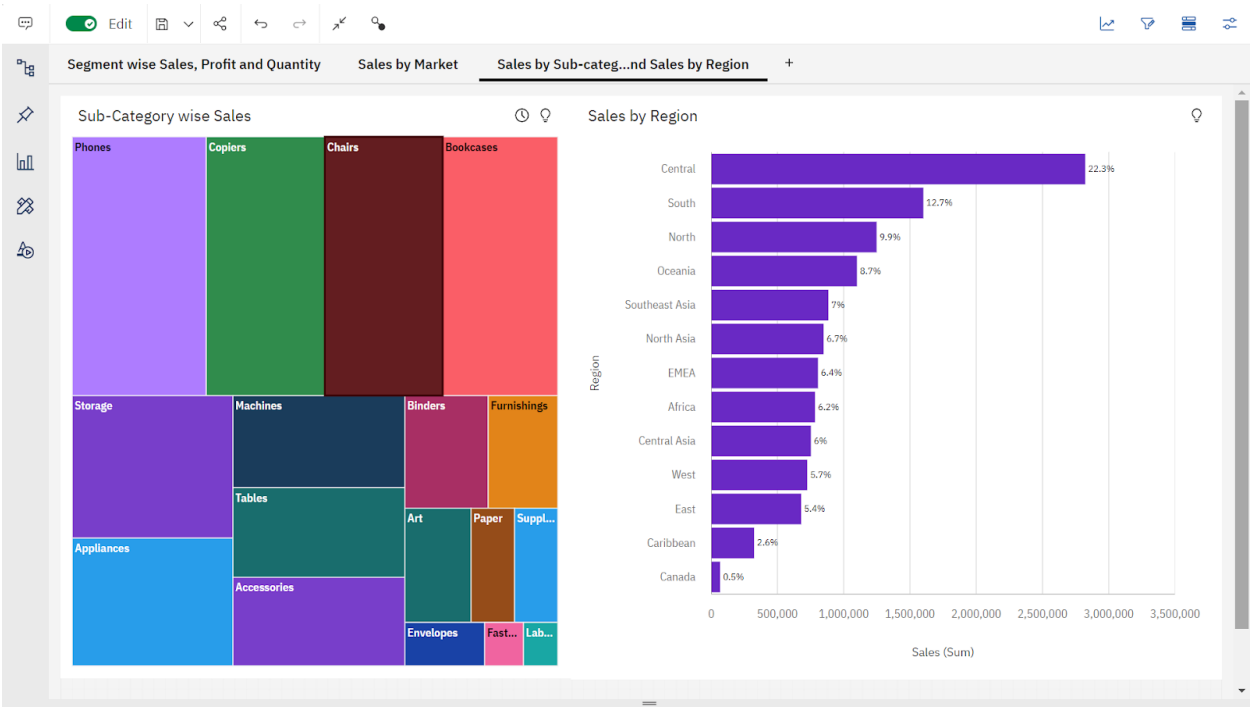
Segment wise Sales, Profit and Quantity



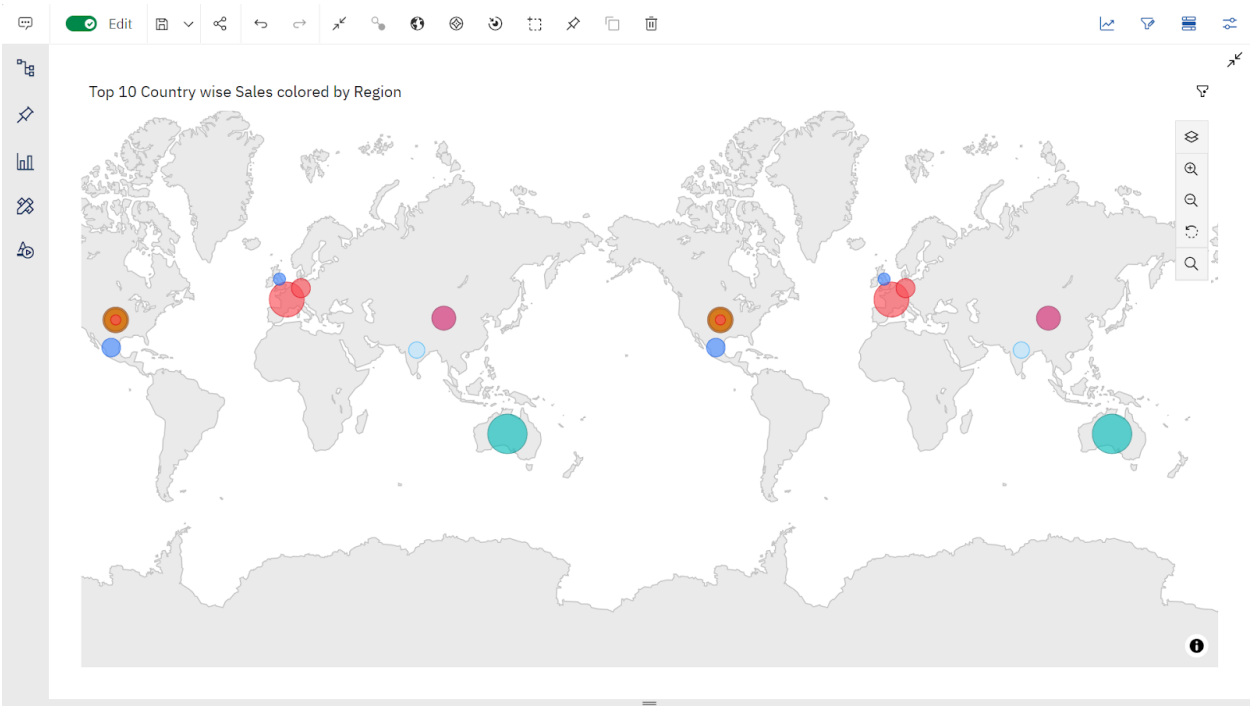
Sales by Market



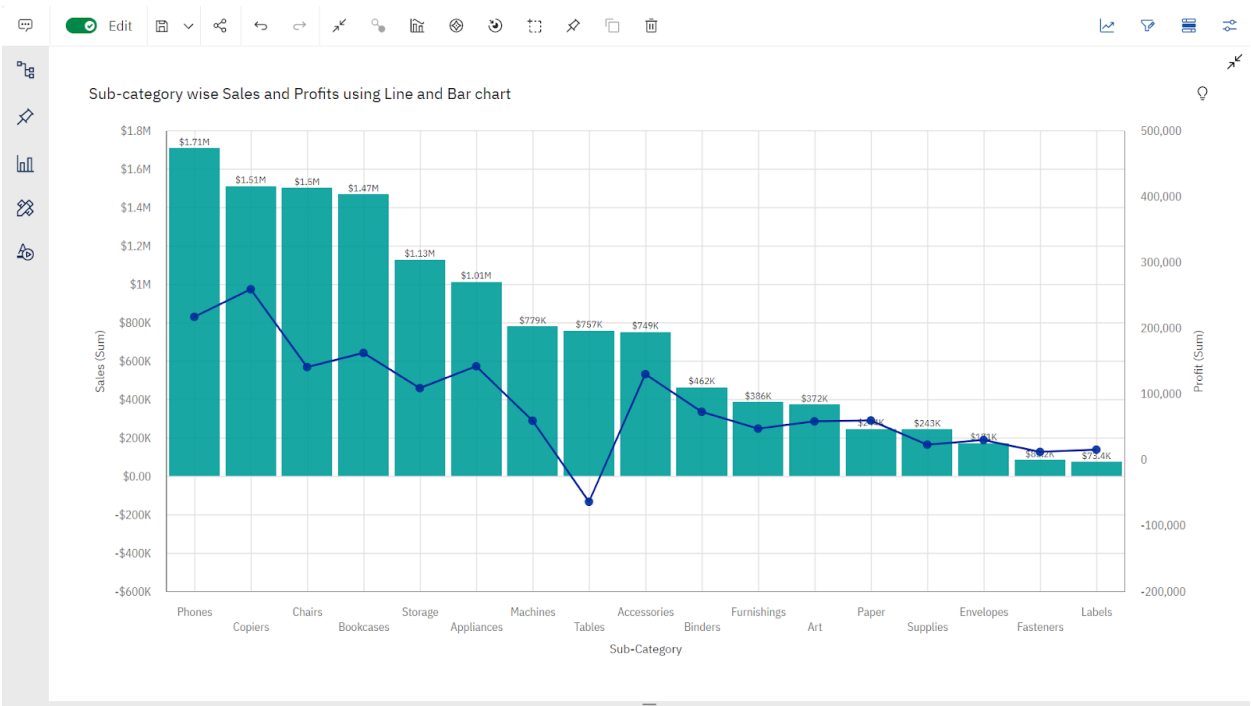
Sales by Sub-category and Sales by Region



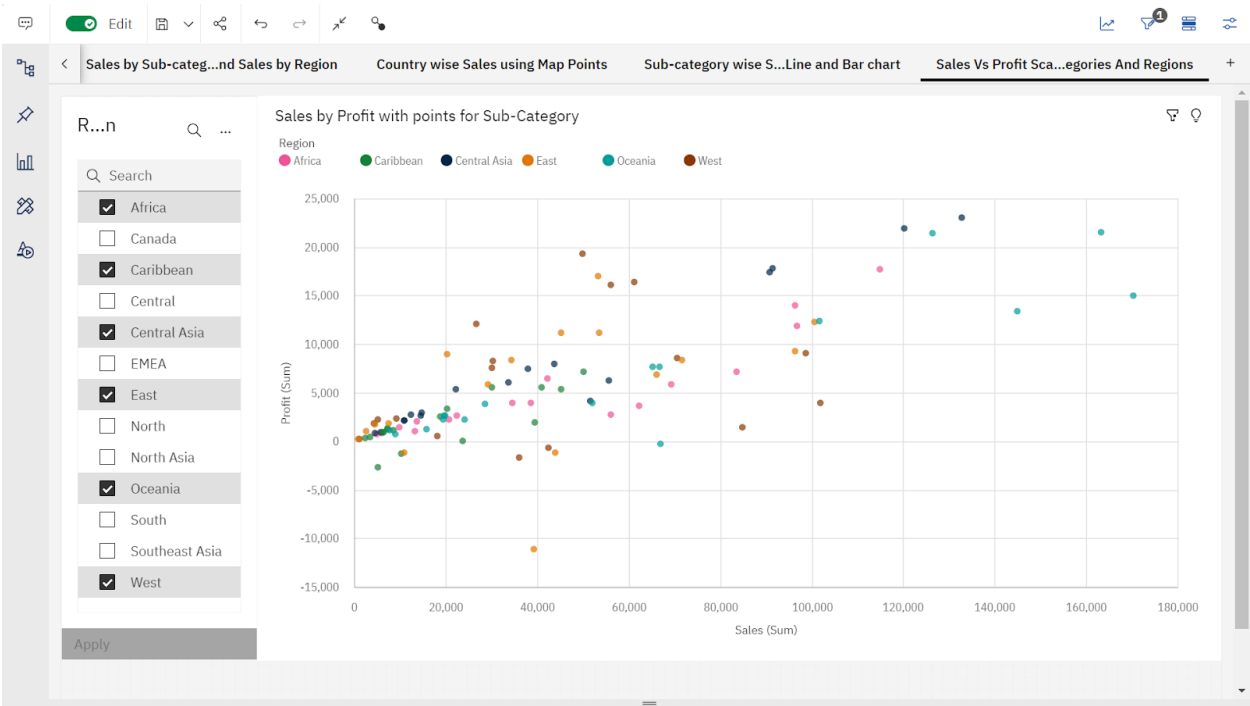
Country wise Sales using Map Points



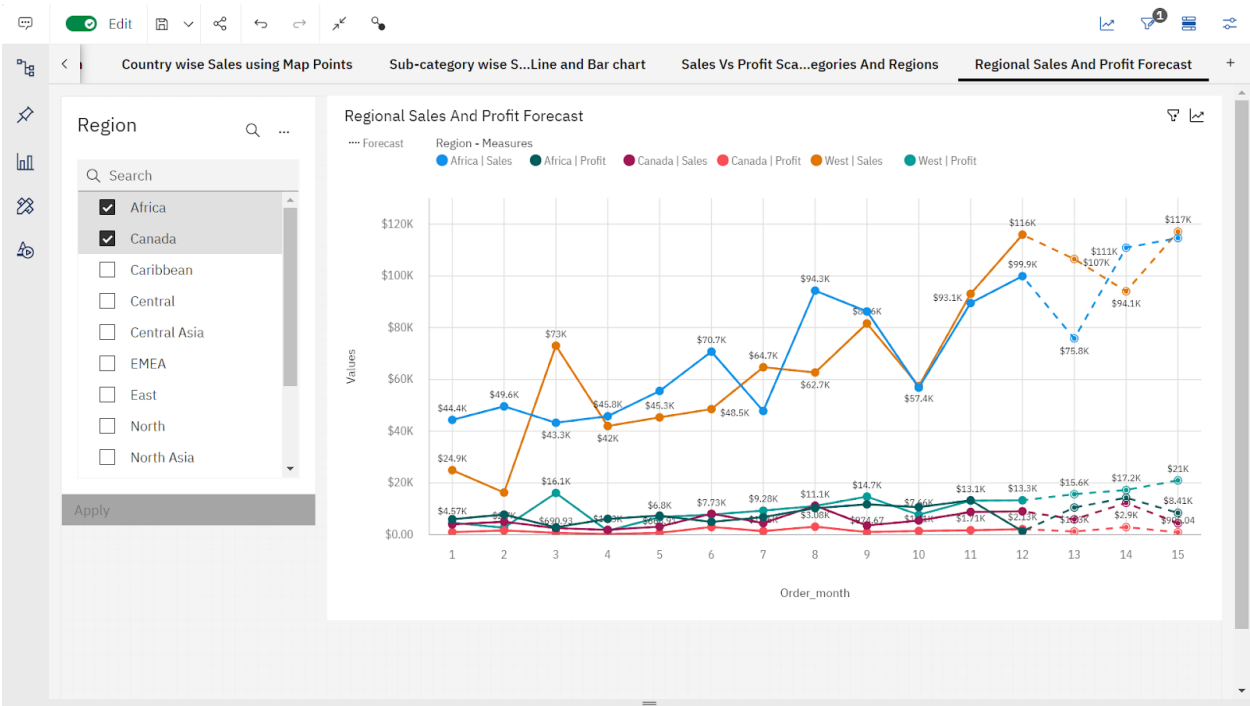
Sub-category wise Sales and Profits using Line and Bar chart



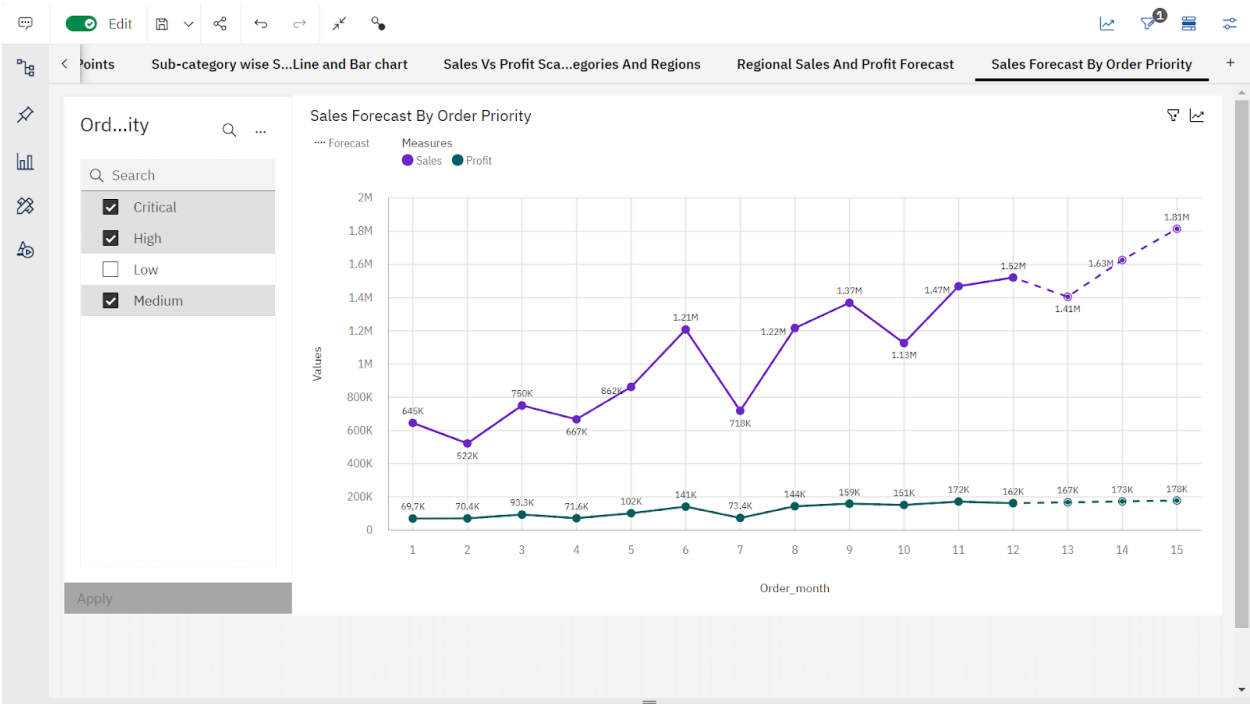
Sales Vs Profit Scatter Plot With Sub Categories And Regions



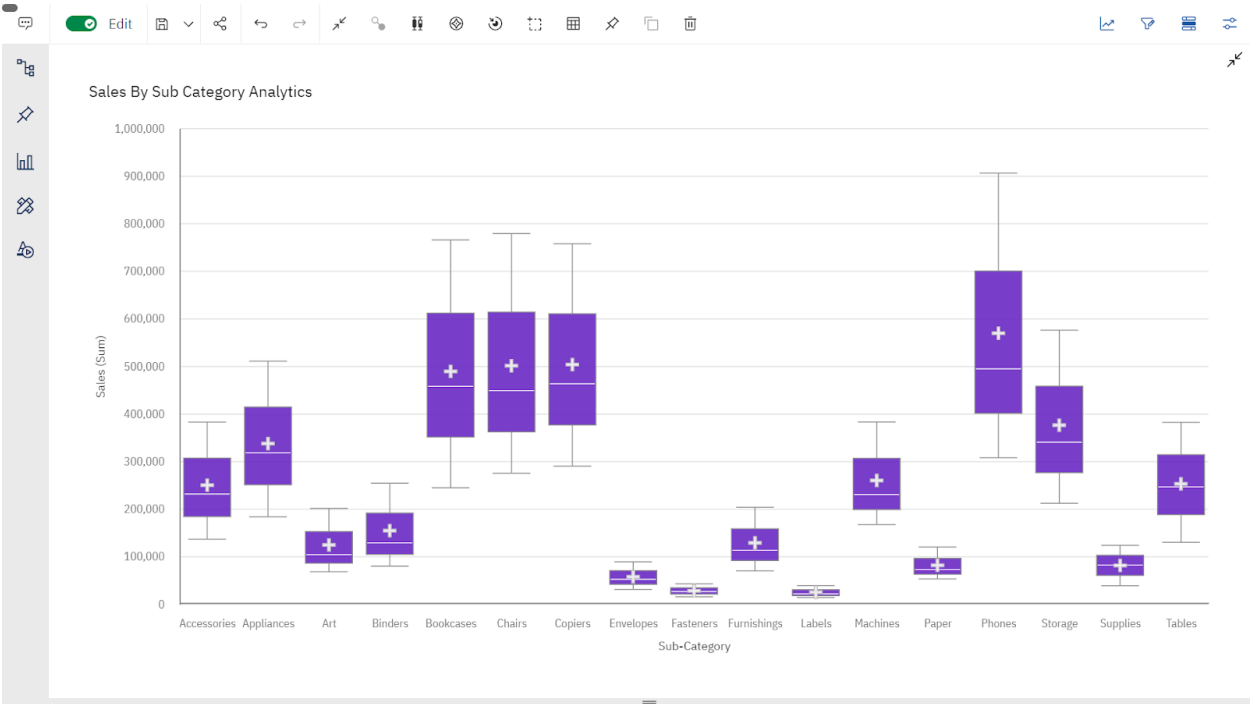
Regional Sales And Profit Forecast



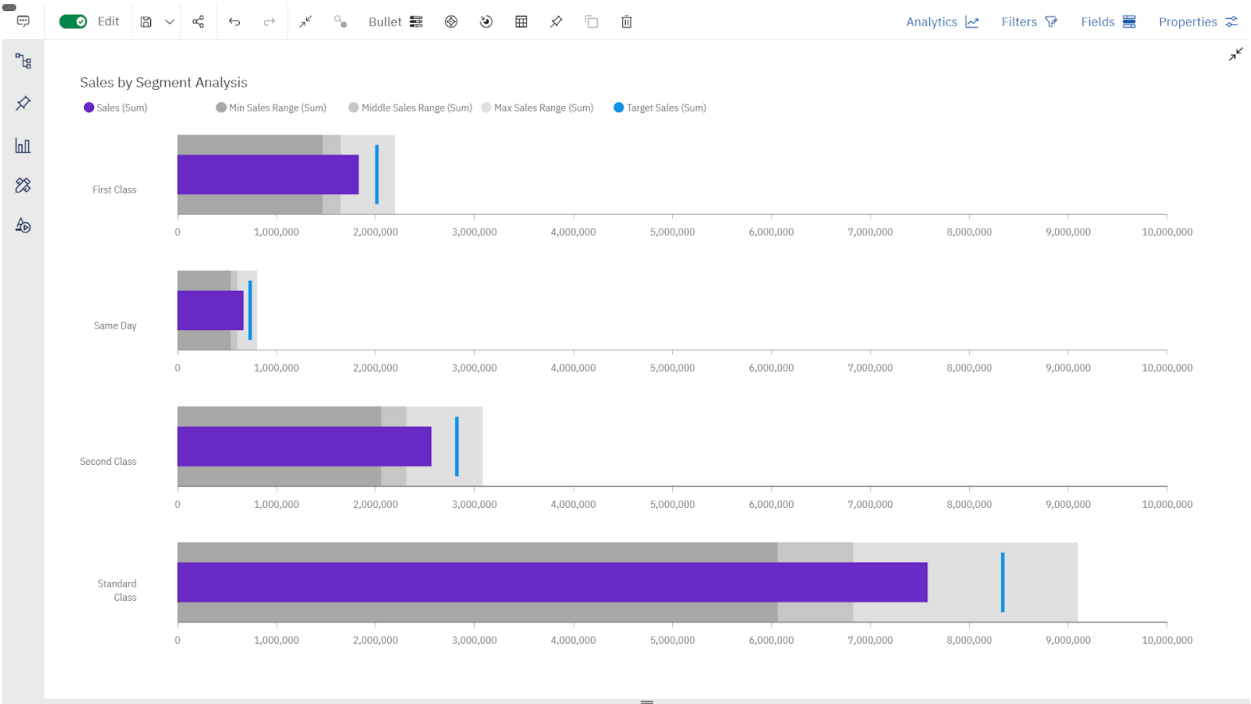
Sales Forecast By Order Priority



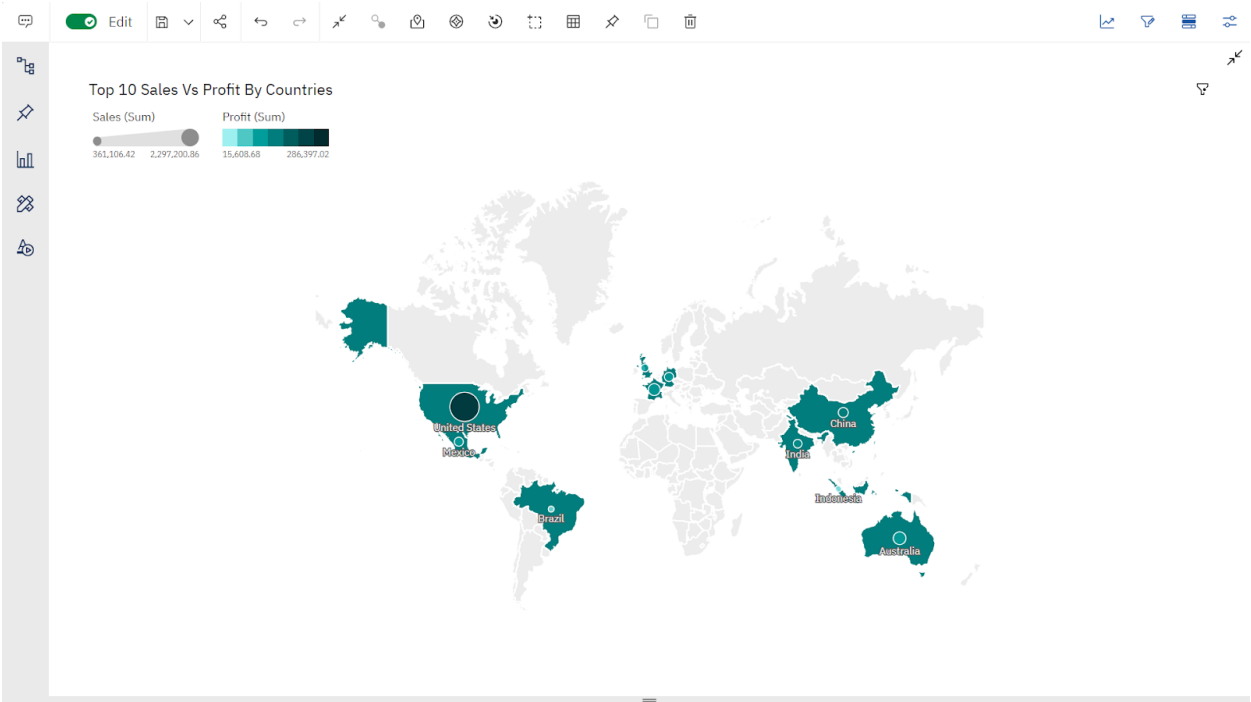
Sales By Sub Category Analytics



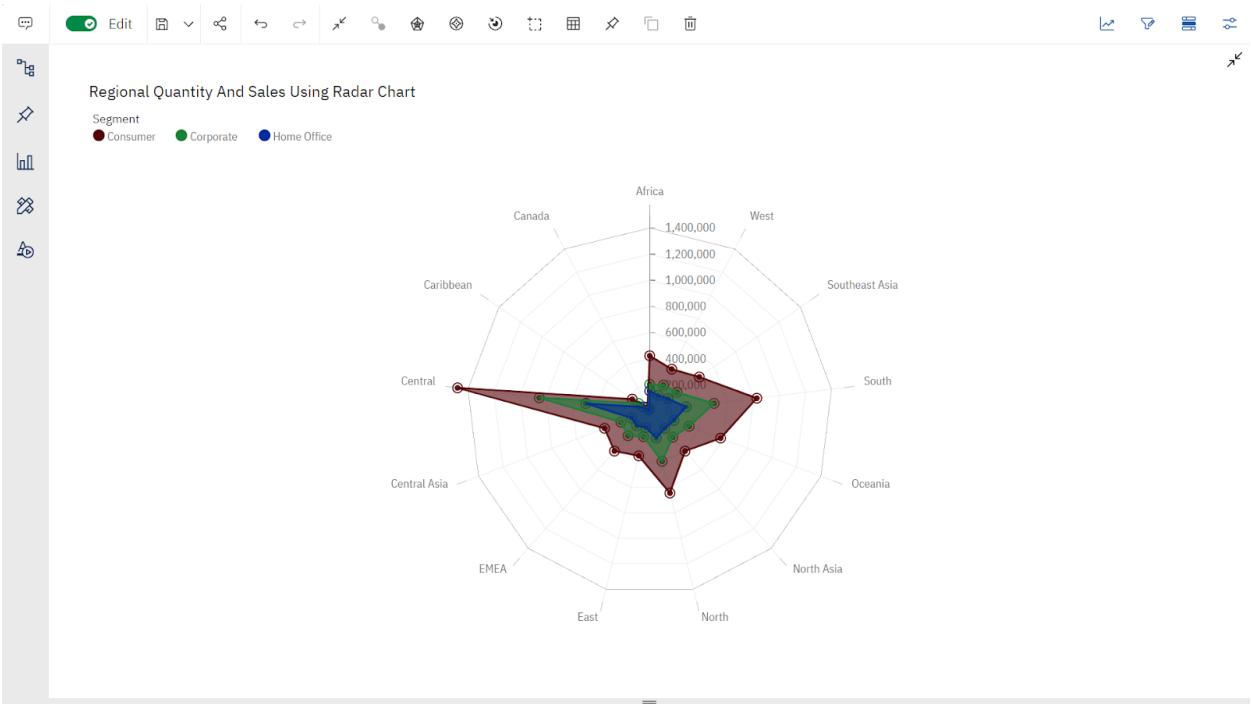
Sales by Segment Analysis



Sales Vs Profit By Countries



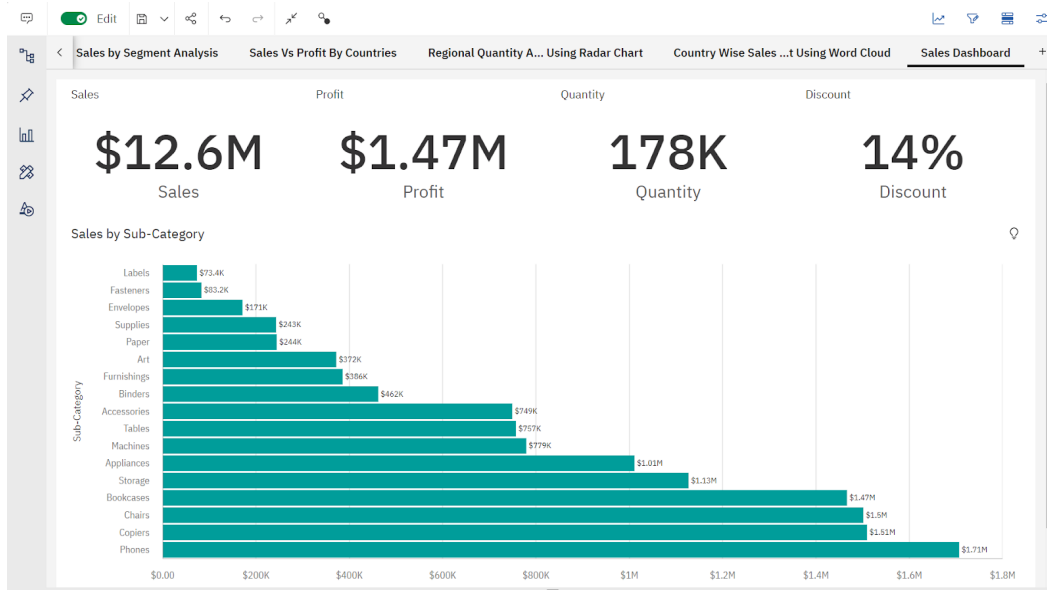
Regional Quantity And Sales Using Radar Chart



Country Wise Sales Vs Profit Using Word Cloud

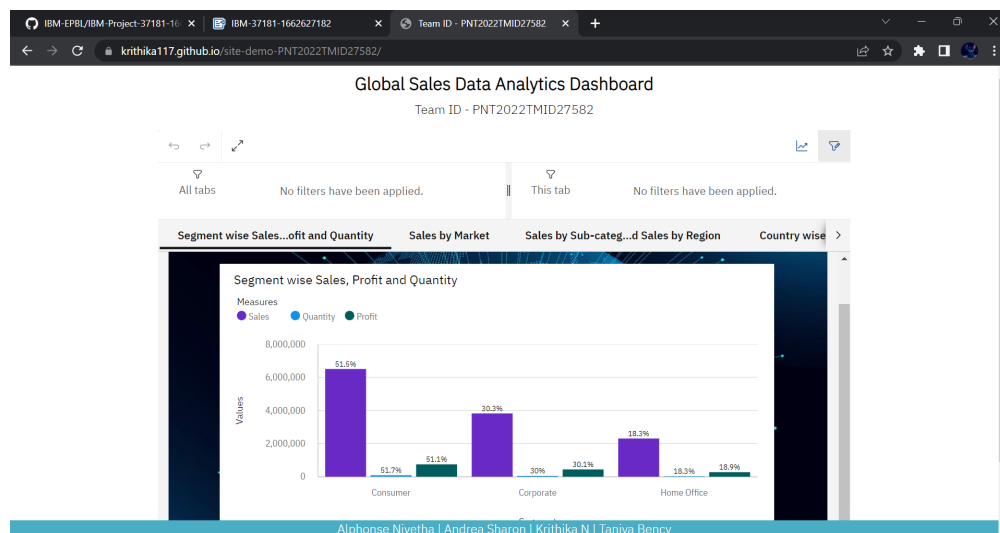


Sales Dashboard



Website Deployment for Public Access

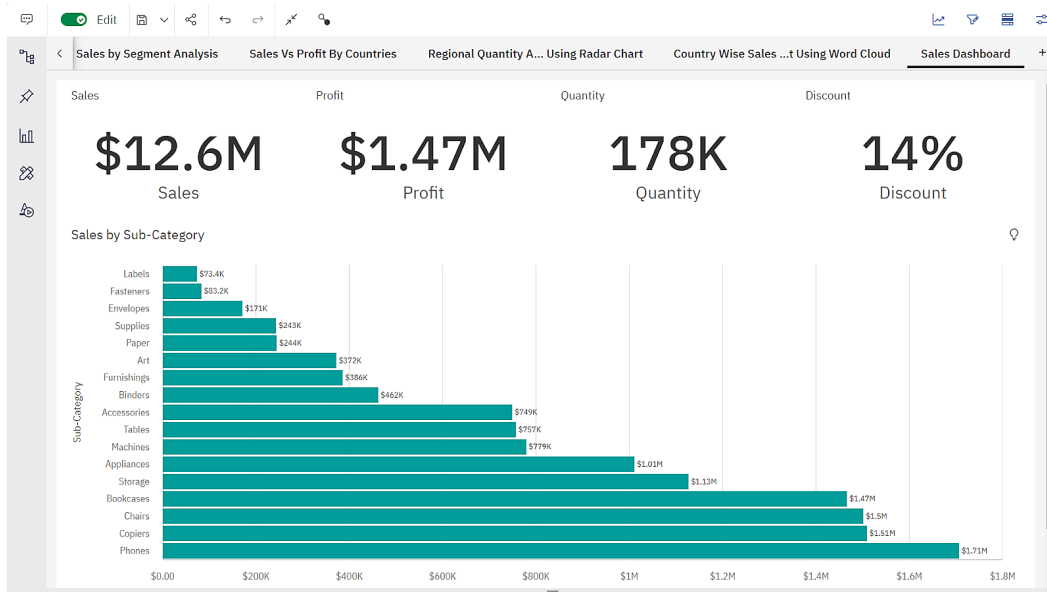
- The dashboard share link is published.
- A Project is created with required components using HTML and CSS.
- A frame medium is allotted for the dashboard.
- Site is deployed in GitHub Pages after being pushed. (Cloud Deployment)
- Site Deployment - <https://krithika117.github.io/site-demo-PNT2022TMID27582/>



7. TESTING

Testing just involved if the visualizations for the dashboard loaded correctly.

Sample Test



8. RESULTS

Performance Metrics

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, we've attempted to understand a few things like, Customer Analysis and Product Analysis of this Global Super Store.

The metrics that we've analyzed to be the key entities affecting the performance of the global sales are as follows.

Profit is based upon

- Product success in that region
- Regional Customer Mass
- Season of Sale
- Priority of the Product

9. **ADVANTAGES & DISADVANTAGES**

a. **Advantages**

- Clear insights about the Sales through several visualizations.
- Scalable Dashboard.
- IBM Cognos is helpful for easily editing charts.
- Data exploration in Cognos Analytics.
- Dashboard can be embedded.

b. **Disadvantages**

- The data module must first start, which delays the loading of the webpage.
- Anonymous Access isn't present in the system.

10. CONCLUSION

Every organisation generates a tonne of data every day, across all industries. This information is potent. Lessons gleaned from sales data may be potent enough to propel a company to achievement. We can fine-tune our company's operations and make each procedure as effective as possible if we harness the power of all this data. Things might become confusing and many questions are raised when there is a big ocean of data to sort through. The following guide outlines important data points to consider in order to address some of those fundamental questions and elaborate on some of the details of sales data analysis. The effectiveness of various marketing and sales initiatives your organisation does can be quantified with the help of sales analytics. They are a priceless asset as your business pushes for efficiency, excellence, and profitability consistently. There is a vast amount of data available that can improve your company.

11. FUTURE SCOPE

Speculation and physical labour are eliminated by data analytics. whether it be selecting the appropriate content, organising marketing initiatives, or creating products. Organizations can use the data analytics insights they uncover to make wise decisions. resulting in improved outcomes and satisfied clients. You may customise customer service to meet their demands using data analytics. Additionally, it offers customisation and strengthens connections with clients. Data analysis can disclose details about clients' preferences, issues, and more. It enables you to make suggestions for goods and services that are better. You can improve production, reduce costs, and streamline business processes with the use of data analytics. By having a better grasp of what your audience wants, you may spend less time producing ads and material that doesn't appeal to them. You can learn a lot about the effectiveness of your efforts via data analytics. This aids in perfecting them for the best results. You may also identify prospective clients who are most likely to engage with a campaign and turn into leads.

12. APPENDIX

Source Code

Site Deployment - <https://krithika117.github.io/site-demo-PNT2022TMID27582/>

index.html

```
<!DOCTYPE html>

<html lang="en">

  <head>

    <meta charset="UTF-8" />

    <meta http-equiv="X-UA-Compatible" content="IE=edge" />

    <meta name="viewport" content="width=device-width, initial-scale=1.0" />

    <title>Team ID - PNT2022TMID27582</title>

    <link rel="stylesheet" href="style.css" />

  </head>

  <body>

    <div class="header">Global Sales Data Analytics Dashboard</div>

    <div class="subheader">Team ID - PNT2022TMID27582</div>

    <div class="frame">

      <center>

        <iframe

          src="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FGlo
          bal%2BSales%2BData%2BAnalytics%2FVisualizations%2FGlobal%2BSuperstore%2BDashboard&
          amp;closeWindowOnLastView=true&ui_appbar=false&ui_navbar=false&shareMode=
          embedded&action=view&mode=dashboard"
```

```
width="100%"
height="100%"
frameborder="0"
gesture="media"
allow="encrypted-media"
allowfullscreen=""
></iframe>
</center>
</div>
<div class="footer">
    Alphonse Nivetha | Andrea Sharon | Krithika N | Taniya Bency
</div>
</body>
</html>
```

style.css

```
@import url('https://fonts.googleapis.com/css2?family=IBM+Plex+Sans:wght@300&display=swap');
* {
    margin: 0;
    padding: 0;
    box-sizing: border-box;
}
body {
    font-family: 'IBM Plex Sans', sans-serif;
```

```
}
```

```
.header {
```

```
    font-weight: bold;
```

```
    text-align: center;
```

```
    font-size: 1.4rem;
```

```
    margin: 0.5rem;
```

```
}
```

```
.subheader {
```

```
    text-align: center;
```

```
}
```

```
iframe {
```

```
    width: 70%;
```

```
    height: 600px;
```

```
    display: flex;
```

```
    align-items: center;
```

```
    justify-content: center;
```

```
    margin: 1rem
```

```
}
```

```
.footer {
```

```
    position: fixed;
```

```
    left: 0;
```

```
    bottom: 0;
```

```
    width: 100%;
```

```
    background-color: rgb(74, 173, 196);
```



```
color: white;

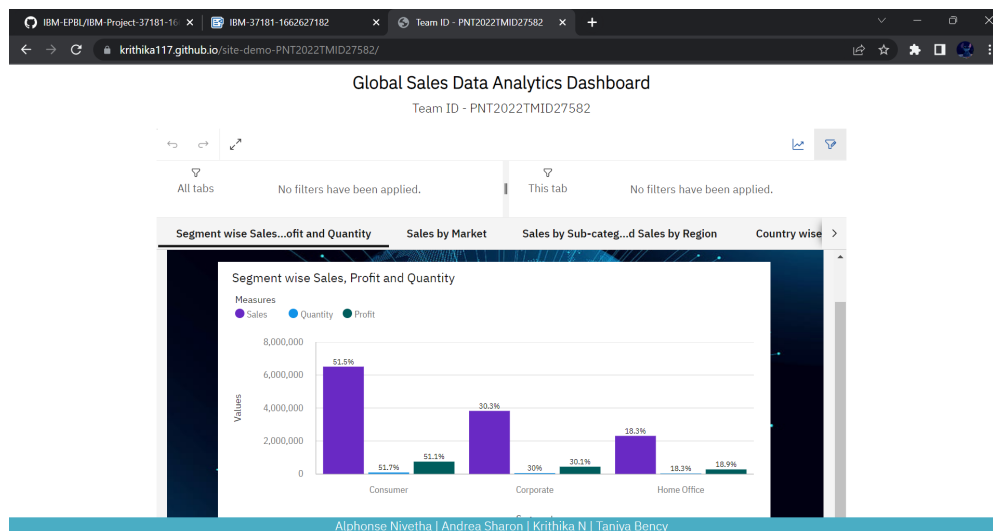
text-align: center;

}
```

GitHub & Project Demo Link

GitHub - <https://github.com/IBM-EPBL/IBM-Project-37181-1660301237>

Site Deployment - <https://krithika117.github.io/site-demo-PNT2022TMID27582/>



Video Demo Link

<https://drive.google.com/drive/folders/12V6vZo5vIAGuP2XUQoRHS-OztfnJ3GRJ?usp=sharing>