IBM NALAIYA THIRAN PROJECT

Global Sales Data Analytics

THIAGARAJAR COLLEGE OF ENGINEERING

TEAM ID: PNT2022TMID21264

TEAM MEMBERS:

1. NANDHAKUMAR RAJ S
2. NITISH KUMAR S
3. PRADHIKSHA S
4. VARSHA A

Table of Contents

1. INTRODUCTION

1.1 Project Overview

Data is being generated very rapidly due to an increase in data generation in day-to-day life. Huge amount of data gets accumulated from different organizations that is difficult to analyze. Processing, analyzing and communicating this data is 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 tough task to carry on. 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 relationship in the data, and visualizing the data and it will provide a conjecture and guesswork of events and will help to find answers that can be sufficiently disguised for a particular problem to come up with an optimal conclusion and a convincing solution.

1.2 Purpose

The purpose of sales data analytics is to boost sales and increase profits for the business owners. It helps business owners gain a clear understanding of their data. It is also used to forecast sales and profits of the business.

2. LITERATURE SURVEY

2.1 References

1. Singh, Manpreet; Ghutla, Bhawick; Lilo Jnr, Reuben; Mohammed, Aesaan F S; Rashid, Mahmood A (2017). [IEEE 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) - Mana Island, Fiji (2017.12.11-2017.12.13)] 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) - Walmart's Sales Data Analysis - A Big
   Data Analytics Perspective. , (), 114–119. doi:10.1109/apwconcse.2017.00028
2. Shahbaz M, Gao C, Zhai L, Shahzad F, Luqman A, Zahid R. Impact of big data analytics on sales performance in pharmaceutical organizations: The role of customer relationship management capabilities. PLoS One. 2021 Apr 28;16(4):e0250229. doi: 10.1371/journal.pone.0250229. PMID: 33909667; PMCID: PMC8081224.
3. Heli Hallikainen Emma SavimäkiTommi Laukkanen. .,Fostering B2B sales with customer big data analytics
4. Singh, Kiran; Wajgi, Rakhi (2016). [IEEE 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave) Coimbatore, India (2016.2.29-2016.3.1)] 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave) - Data analysis and visualization of sales data. , (), 1–6. doi:10.1109/STARTUP.2016.7583967

2.2 Problem Statement Definition

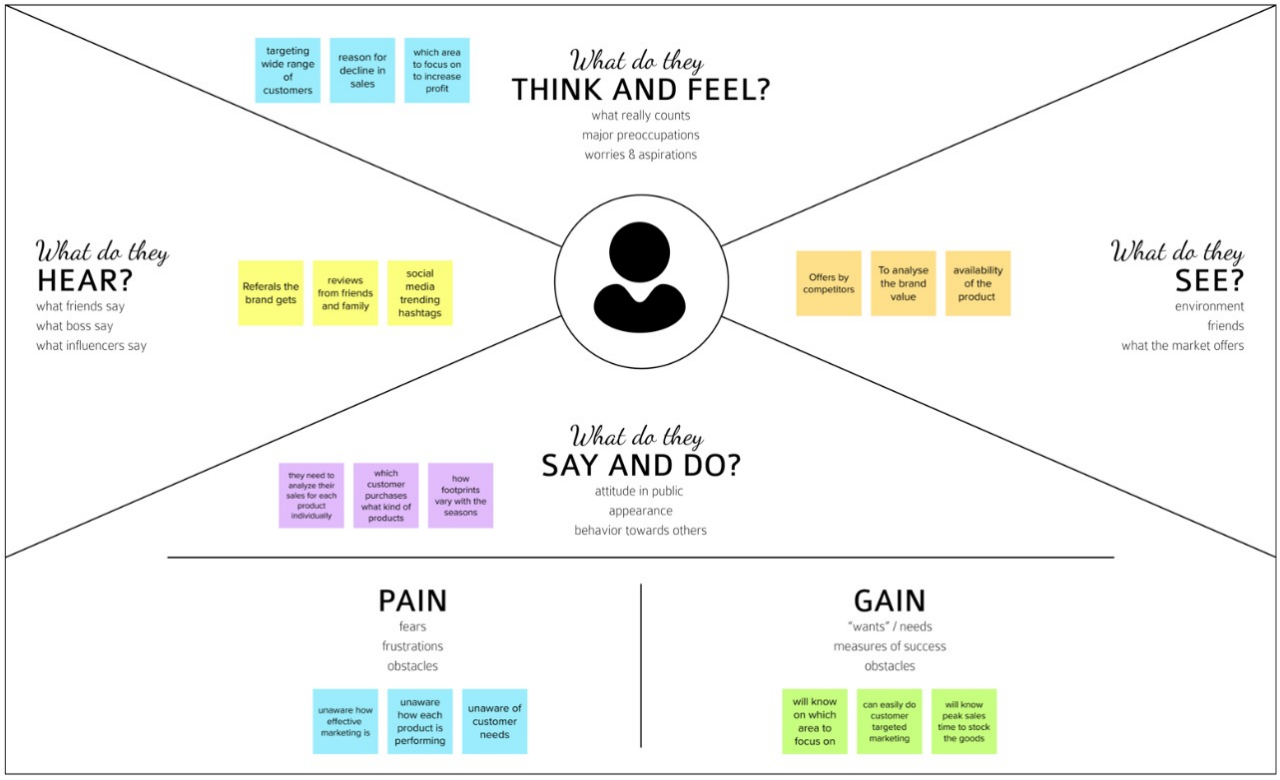
Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Customer	Prepare a dataset	I get errors during data entry	Of the huge amount of data	Frustrated
PS-2	Customer	Analyze the sales	The results are not accurate	Of the datasets which are from different sources	Dissatisfied

PS-1:

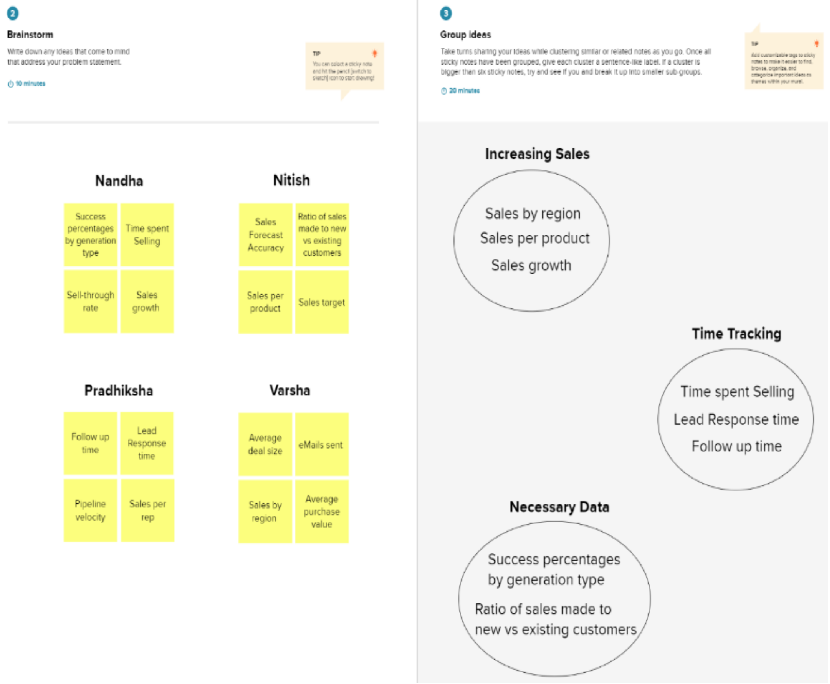
PS-2:

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Campus



3.2 Ideation and brainstorming



Idea Prioritization

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Businesses generate huge amounts of data which needs to be analysed to boost their sales and run their business efficiently. Through sales analytics, you can decide which products to focus on, where to sell and how best to reach customers.
2.	Idea / Solution description	We analyse the data using IBM Cognos in order to Identify new sales opportunities, Plan effective sales targets, Improve customer acquisition and Incentivise sales teams.

3.	Novelty / Uniqueness	1. Interactive and user-friendly dashboards 2. ML model to predict sales.
4.	Social Impact / Customer Satisfaction	1. To gain a competitive advantage. 2. Increase Revenues 3. increase customer retention
5.	Business Model (Revenue Model)	Two pricing model – Standard and Premium 1. Standard – Basic analysis to Identify new sales opportunities and plan effective sales targets. 2. Premium - ML model to predict sales along with advanced analysis to gain a competitive advantage.
6.	Scalability of the Solution	The base model can be used for start-ups and big corporations with just a few changes.

3.4 Proposed Solution Fit

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Business owners who are willing to understand their sales trends and improve it.

6. CUSTOMER CONSTRAINTS

- Availability of proper dataset to acquire the business insights they need.
- Standardized format of data.

5. AVAILABLE SOLUTIONS

Manually monitoring sales trends.
Generating graphs and other data visualization using simple tools like excel.

Explore AS, differentia

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS

- Process the dataset.
- Understand the customer requirements
- Perform suitable analysis to produce optimal results

9. PROBLEM ROOT CAUSE RC

Huge amounts of data generated by businesses which cannot be processed and analyzed manually.

7. BEHAVIOUR BE

Analysing data using basic analytic tools.
Wearisome way of analysing data requiring heavy manual labour.

Focus on J&P, tap into BE, understand RC

3. TRIGGERS TR

The necessity to make informed business decisions.
To be aware of how their business is performing.

4. EMOTIONS: BEFORE / AFTER EM

Before – Lost and confused

After – Clear and Confident

10. YOUR SOLUTION SL

Creating an interactive, simple and powerful dashboard and provide services based on their subscription model.

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE

Upload their dataset to receive the necessary data visualization and insights they require.

8.2 OFFLINE
Making business decisions based on the sales insights provided.

4.REQUIREMENT ANALYSIS

4.1 Functional Requirements

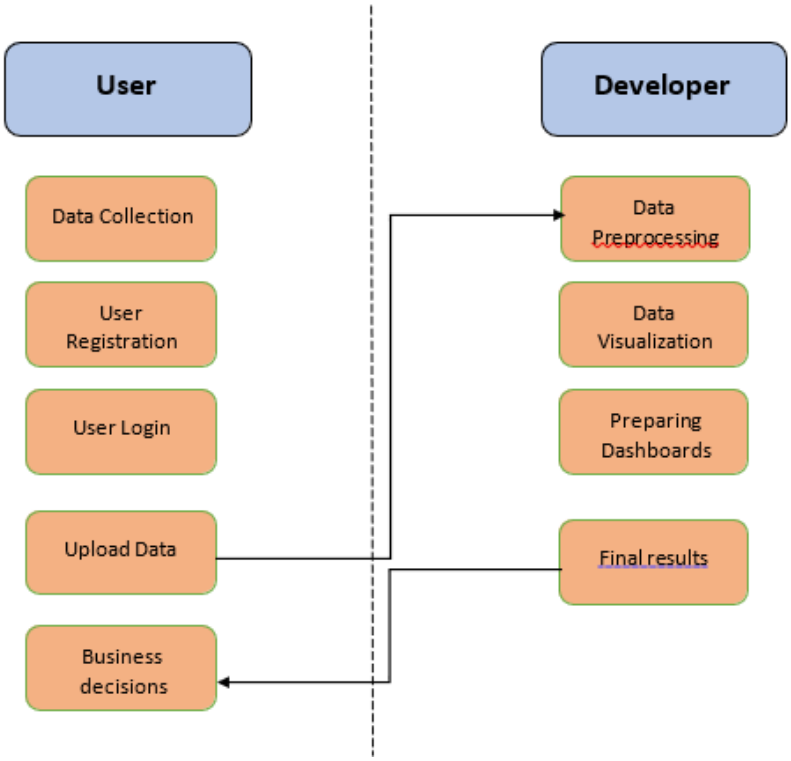
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Login	Login via Email and password
FR-4	User Input	Upload the dataset to be analysed
FR-5	Data pre-processing	Data cleaning
FR-6	Data Visualization	Meaningful charts and graphs to derive insights
FR-7	Business Decisions	Decisions are made according to the charts.

4.2 Non-Functional Requirements

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	User-friendly dashboard, easy to access and understand.
NFR-2	**Security**	Dashboards can only be accessed with valid login credentials.
NFR-3	**Reliability**	Dashboards must be accessible at all times even during traffic.
NFR-4	**Performance**	Must be able to handle huge volume of data Must support simultaneous access by multiple users.
NFR-5	**Availability**	Data must be available at all times and be fault tolerant.
NFR-6	**Scalability**	The dashboard must be accessible through all kinds of devices.

5.PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution and Technical Architecture

Technical Architecture

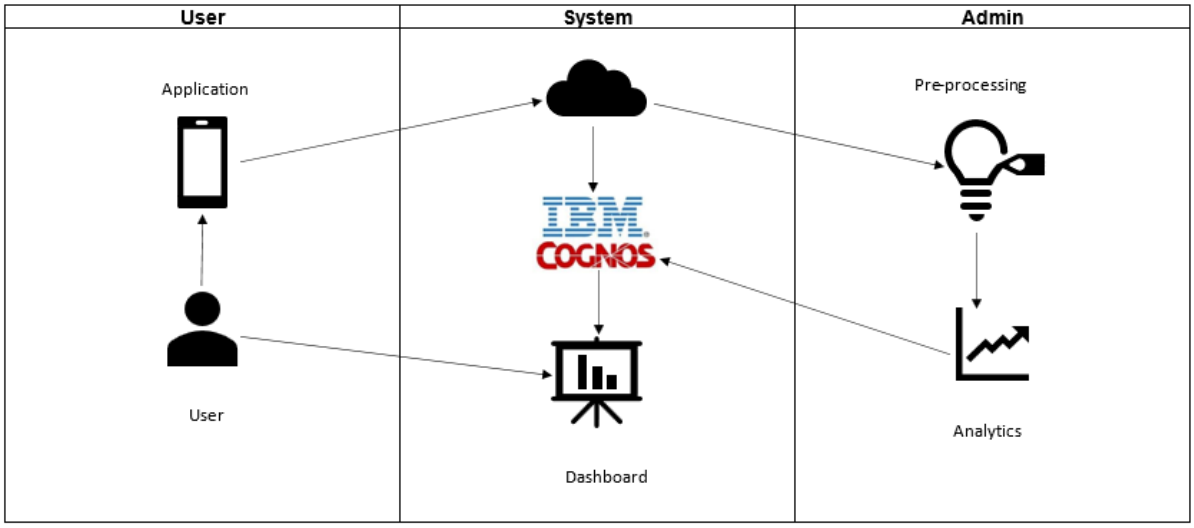


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	IBM Cognos
2.	Cloud	Data is uploaded in cloud through interface	IBM Cloud
3.	Data Pre-processing	Cleaning and processing Dataset	IBM Cognos, IBM Cloud
4.	Data Exploration	Uploaded dataset is explored to identify patterns	IBM Cognos
5.	Data Visualization	Visualization charts	IBM Cognos
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloud etc.
7.	Dashboard	User logs in to view the visualization	IBM Cognos Dashboard

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	IBM Cognos, IBM Cloud,Jupyter Notebook

2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	Active Directory
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	IBM Cloud
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Cognos business intelligence server
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	IBM Cognos, Performance management hub

5.3.USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user or web 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
	confirmation	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 login using my credentials	I can access my dashboard with right credentials	High	Sprint-2
	User input	USN-4	As a user, I can upload the dataset	Dataset in the right format will be accepted	Medium	Sprint-1
	Dashboard	USN-5	As a user, I can navigate through the dashboard and view visualizations	I can access my dashboard with right credentials	High	Sprint-1
Customer Care Executive	Chatbox	USN-6	As a user, I can use the chat box for assistance	It can be accessed at all times	Medium	Sprint-3
	Calling	USN-7	As a user, I can use the calling option for any support or assistance	It can be used in working hours	High	Sprint-3
	Mail	USN-8	As a user, I can use the mailing option for any support or assistance	It can be accessed at all times	Medium	Sprint-3
Administrator	Login	USN-9	As an admin, I can login using my credentials	I can login if my credentials are right	High	Sprint-1
	Pre-process	USN-10	As an admin, I can access the uploaded dataset and pre-process them for further use		High	Sprint-2
	Dashboard	USN-11	As an admin,I can customize dashboard	Dashboard can be edited by admin only	High	Sprint-2

6.PROJECT PLANNING AND SCHEDULING

6.1 Sprint Planning & Estimation

MILESTONES	TASKS
MILESTONE-1	Data collection through KAGGLE

MILESTONE-2	Inserting the necessary data into the platform (IBM COGNOS)
MILESTONE-3	Visualize and Explore the Data
MILESTONE-4	Interactive Dashboard is Created
MILESTONE-5	Insights are shown in the Dashboard
MILESTONE-6	Construct a standardized data set and use the needed data with the assistance of a python program
MILESTONE-7	Use of different algorithms with Google Colab to achieve the desired result with more accuracy
MILESTONE-8	Present them in the necessary format
MILESTONE-9	Deployed in the Github

6.2 Sprint Delivery 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	Medium	Pradhiksha S,Nandhakumar Raj S, Nitish Kumar S, Varsha A

Sprint-1	Login	USN-2	As a user, I will receive confirmation email once I have registered for the application	1	Low	Pradhiksha S,Nandhakumar Raj S
Sprint-1	Collecting Data	USN-3	As a user, I can upload my dataset	2	Medium	Nitish Kumar S, Varsha A
Sprint-2	Data pre-processing	USN-4	As a data analyst I should pre-process and clean the dataset.	2	Medium	Pradhiksha S,Varsha A
Sprint-3	Data Visualization	USN-5	Preparing meaningful visualization of provided sales data	3	High	Nandhakumar Raj S, Nitish Kumar S
Sprint 4	Create Dashboard		As a Data analyst I need to perform data visualization and create a dashboard using IBM Cognos	3	High	Varsha A, Nandhakumar Raj S

7.CODING

7.1 Sprint 1

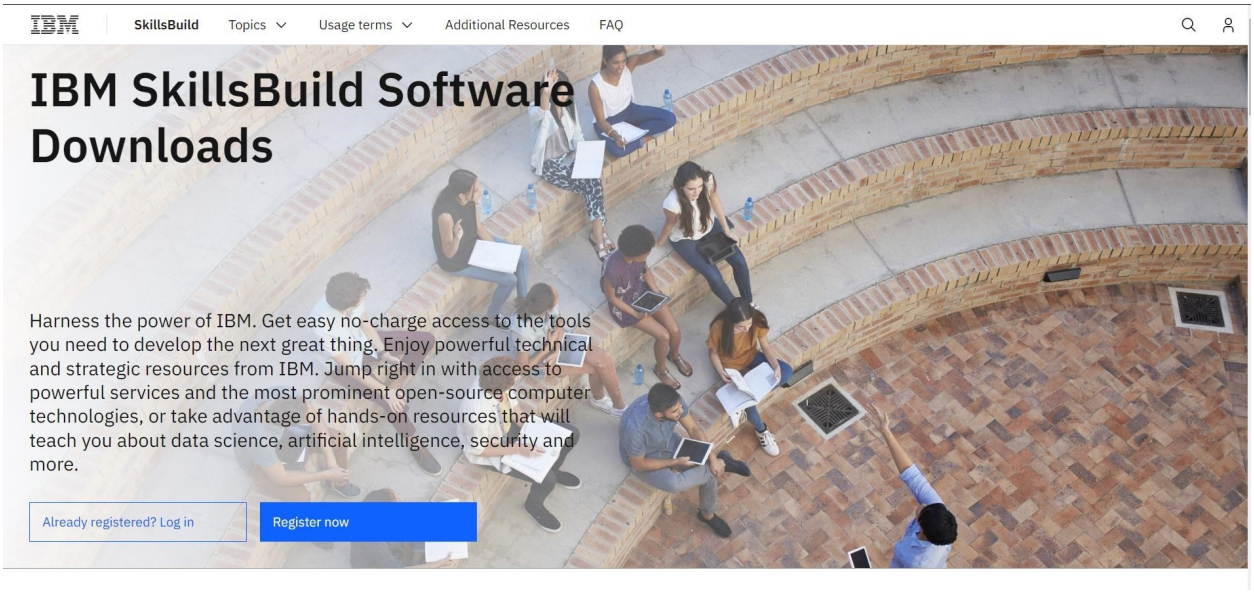
SPRINT 1 FUNCTIONAL REQUIREMENT

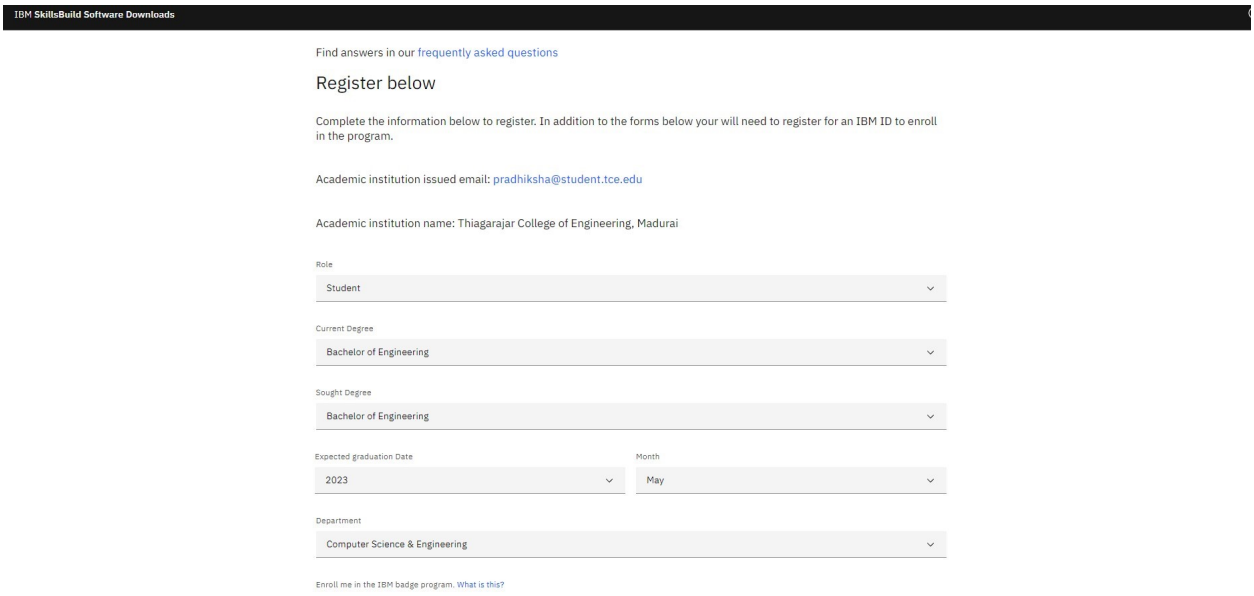
REGISTRATION

LOGIN

DATA COLLECTION

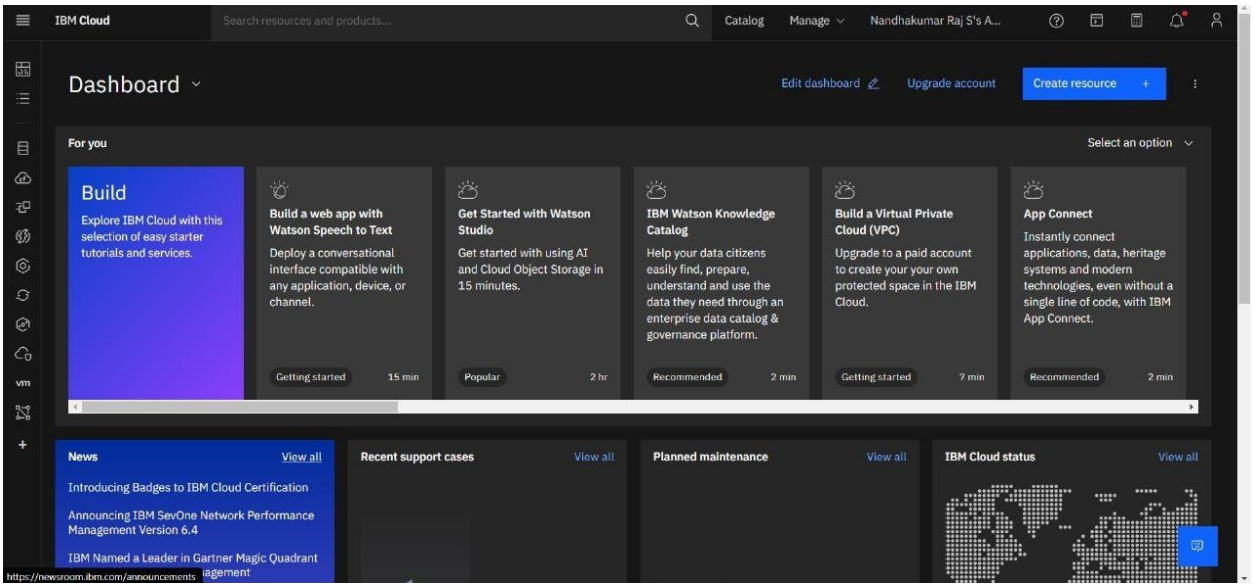
REGISTRATION



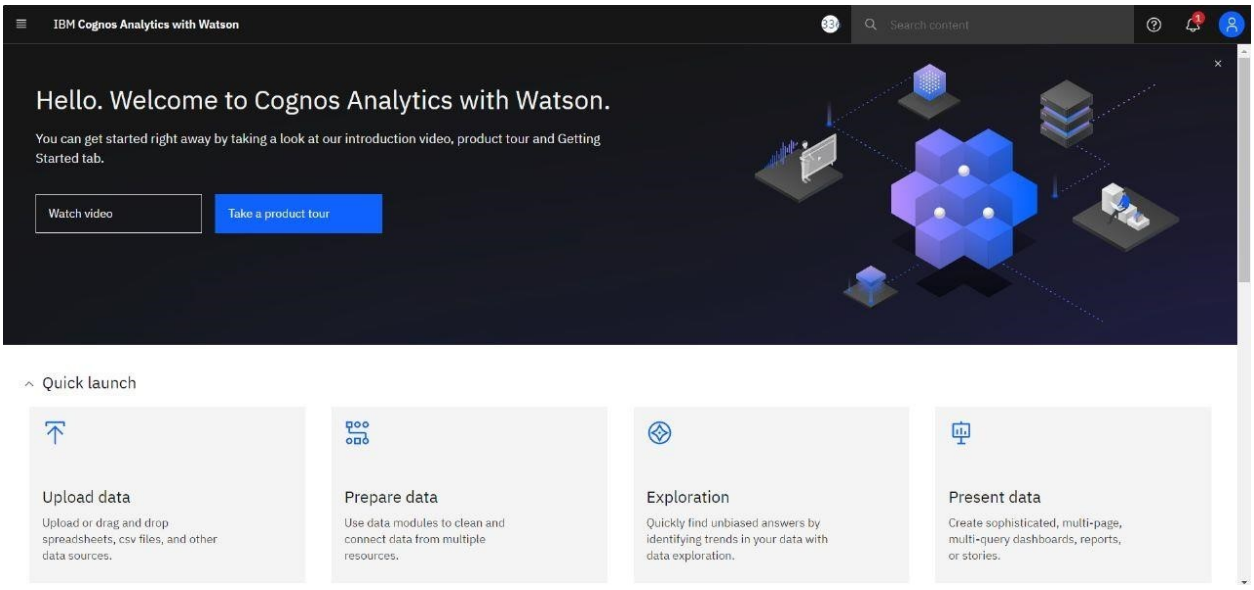


LOGIN

IBM COGNOS CLOUD LOGIN



IBM COGNOS ANALYTICS ACCOUNT LOGIN



7.2 Sprint 2

Sprint-2 Functional Requirement:

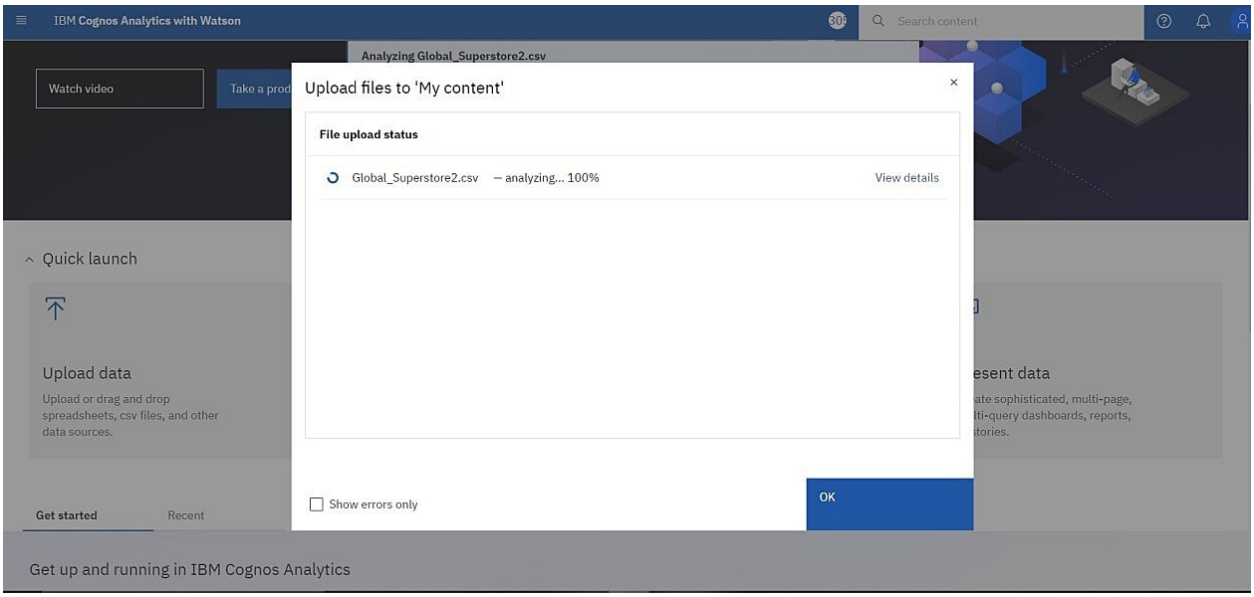
- Upload dataset
- Data Preparation

- Data Exploration

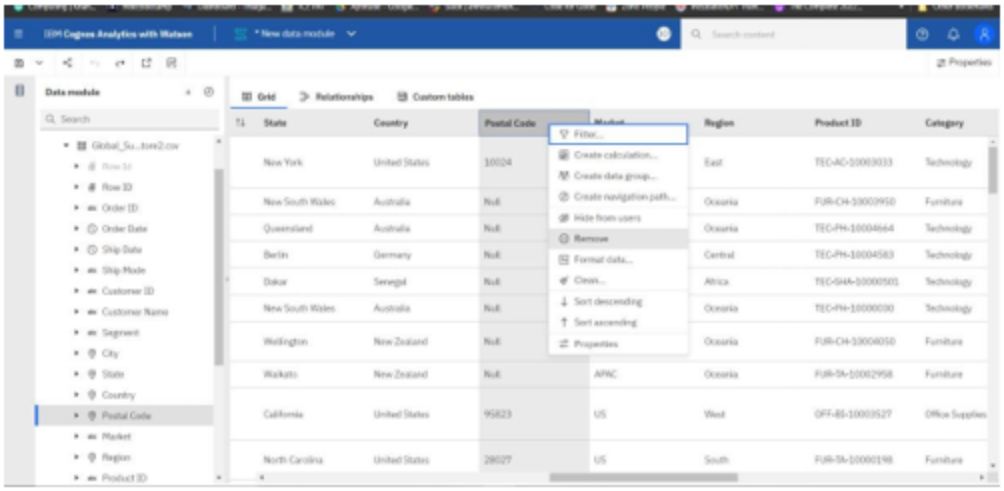
Download Dataset from:

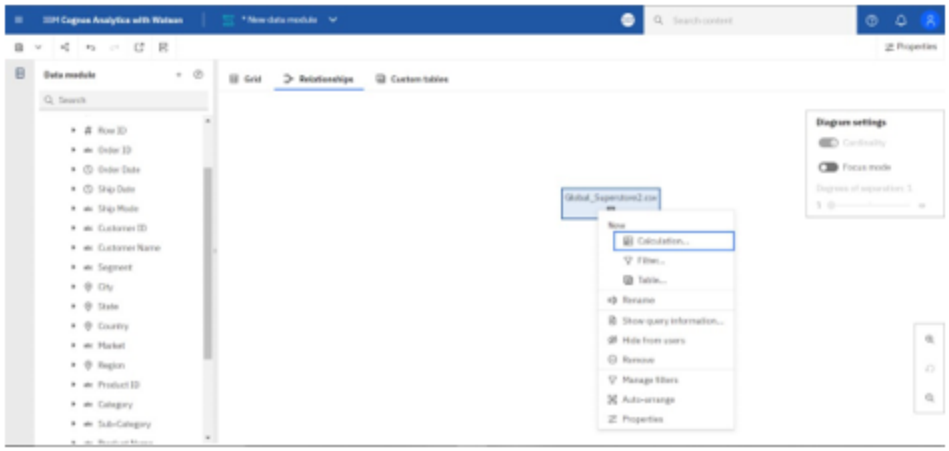
https://www.kaggle.com/apoorvaappz/global-super-store-dataset

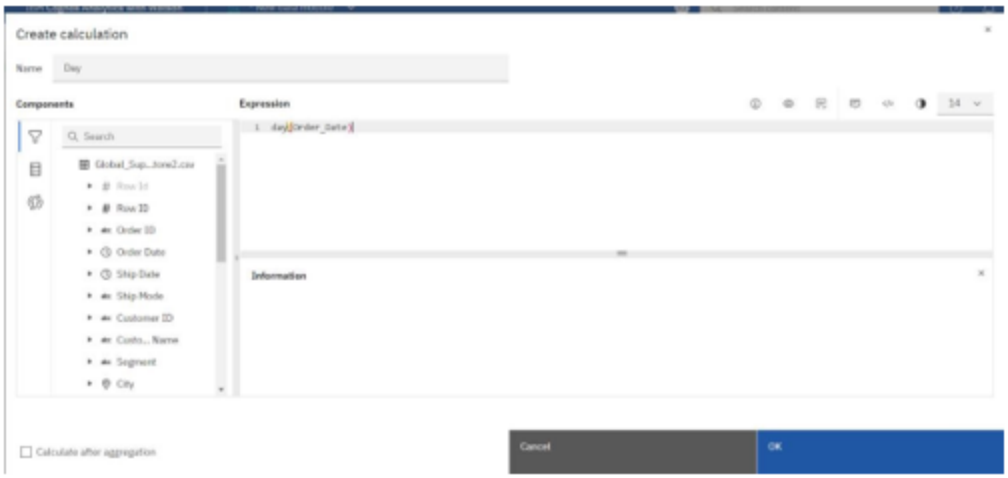
Upload dataset:

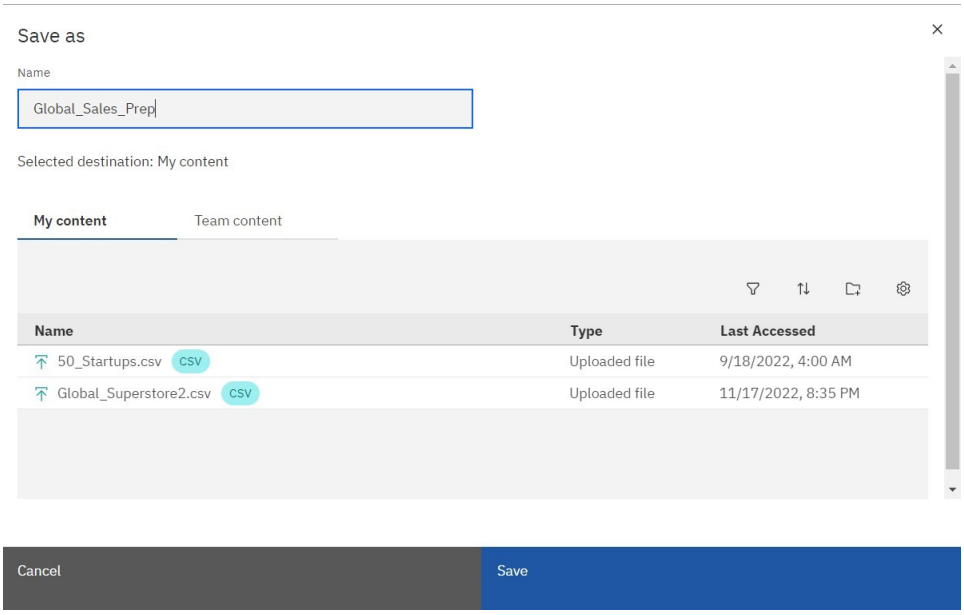


Data Preparation:

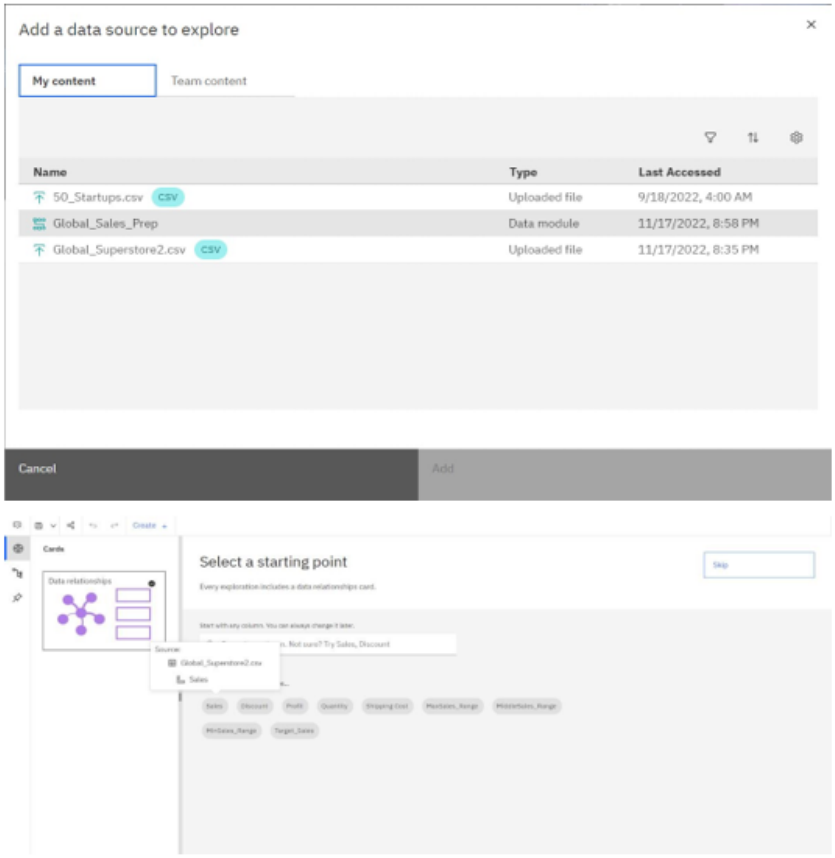








Data exploration

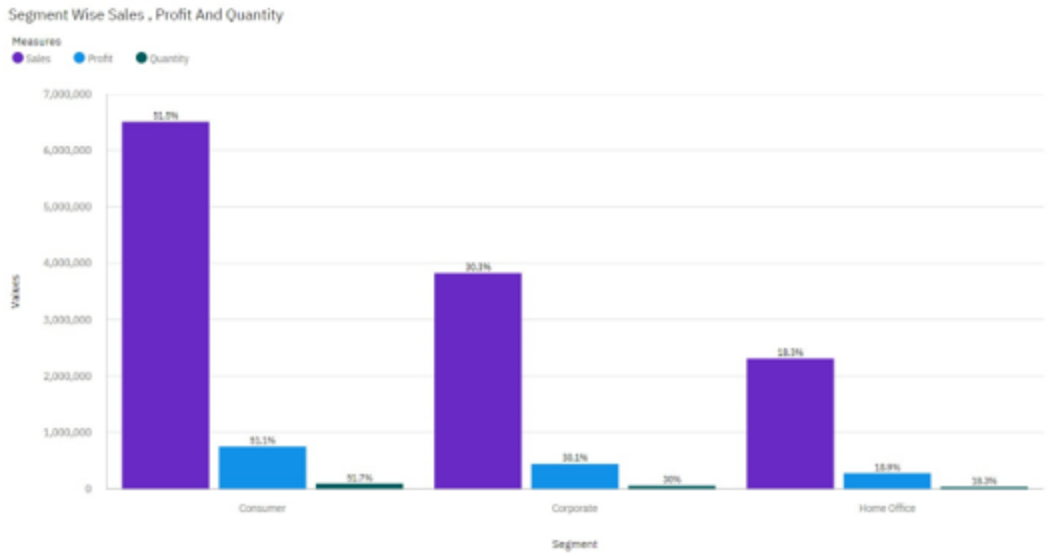


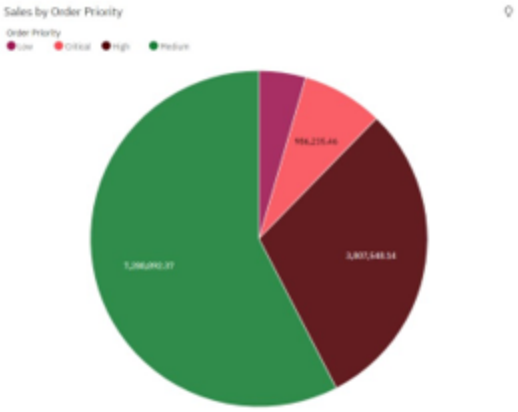
Data Exploration Link:

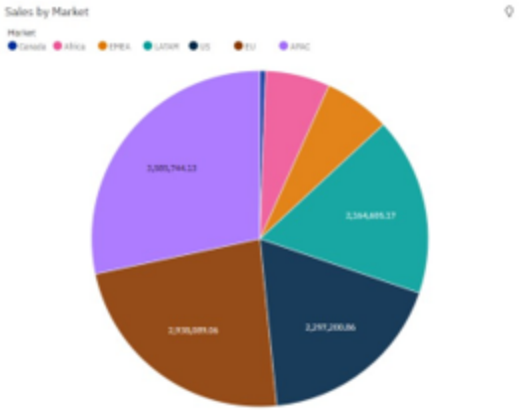
https://us1.ca.analytics.ibm.com/bi/?perspective=explore&pathRef=.my_folders%2FGlobal_Superstore%2BExploration&subView=model0000018488c590b2_00000004

7.3 Sprint 3

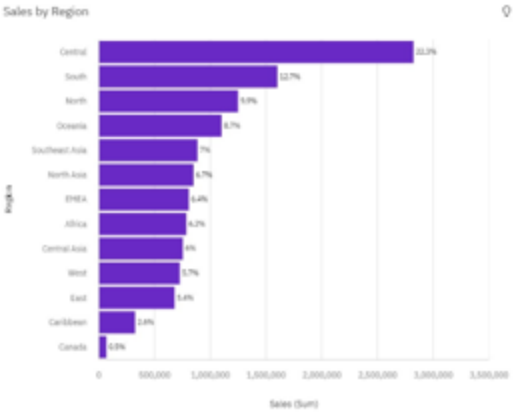
Data Visualization Charts:







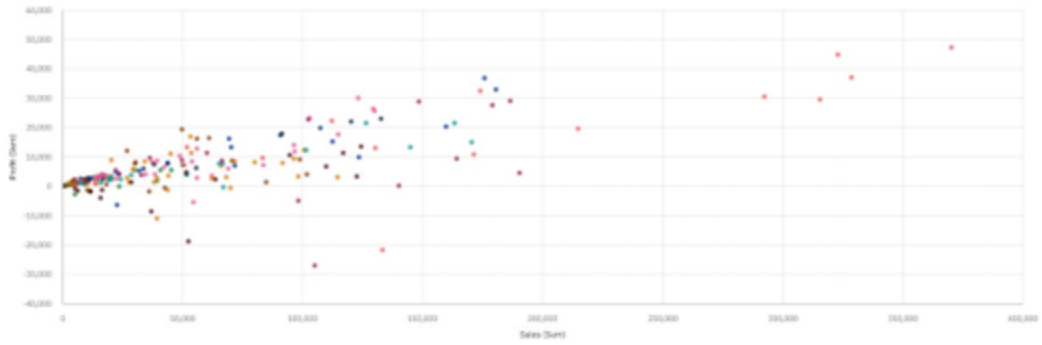

Sales by Sub-Category
Phones
Bookcases
Storage
Appliances
Copiers
Machines
Accessories
Binders
Furnishings
Chairs
Tables
Art
Supplies
Fast...
Paper
Envelopes
Lab...


Sales by Region
Central
South
North
Oceania
Southeast Asia
North Asia
EMEA
Africa
Central Asia
West
East
Caribbean
Canada
Region
0
500,000
1,000,000
1,500,000
2,000,000
2,500,000
3,000,000
3,500,000
Sales (Sum)

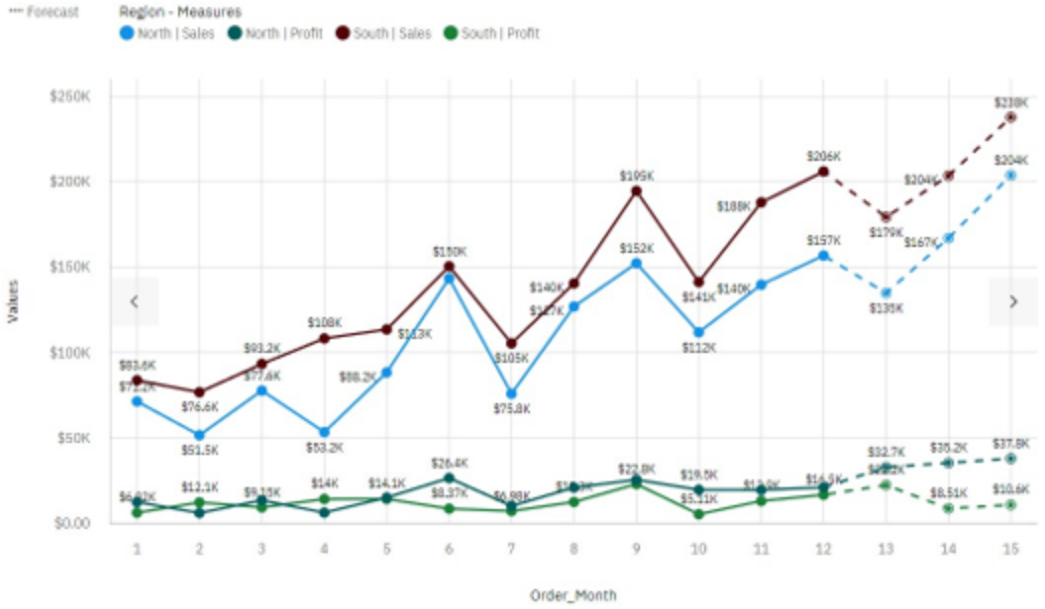

Country Wise Sales Using Map Points
Iceland
Sweden
Norway
Russia
Canada
United Kingdom
Belarus
France
Ukraine
Kazakhstan
Mongolia
Spain
Italy
Greece
Turkmenistan
China
Japan
Morocco
Tunisia
Iraq
Iran
Pakistan
Libya
Egypt
Oman
Bangladesh
Mauritania
Niger
Sudan
Yemen
Cambodia
Malaysia
Cameroon
Ethiopia
Maldives
Indonesia
Kenya
Papua New Guinea
Tanzania
Angola
Madagascar
Namibia
South Africa
Australia
United States
Mexico
Cuba
Colombia
Peru
Brazil
Bolivia
Chile
Uruguay
New Zealand
North Pacific Ocean
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
Indian Ocean


Sub Category Wise Sales And Profits
Sales (Sum)
Profit (Sum)
Phones
Copiers
Chairs
Bookcases
Storage
Appliances
Machines
Tables
Accessories
Binders
Furnishings
Art
Paper
Supplies
Envelopes
Fasteners
Labels
Sub-Category

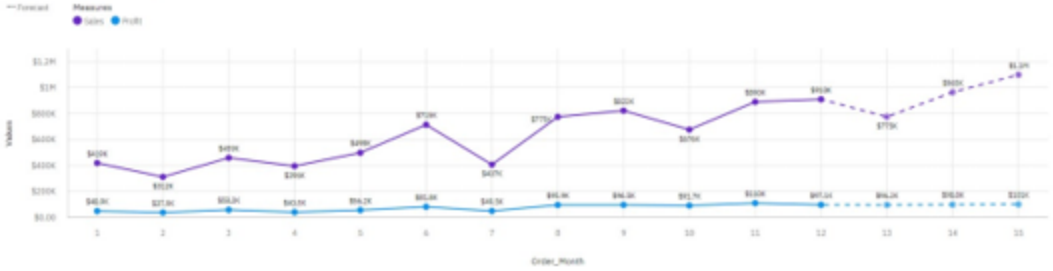
Sales Vs Profit Scatter Plot With Sub Categories And Regions



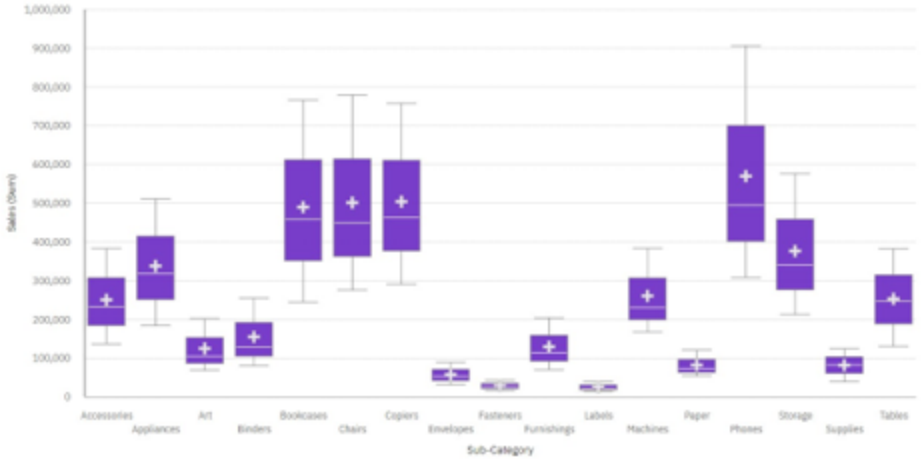
Regional Sales And Profit Forecast



Sales Forecast By Order Priority

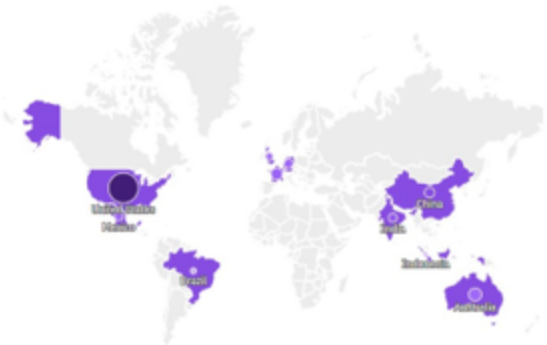


Sales By Sub Category Analytics

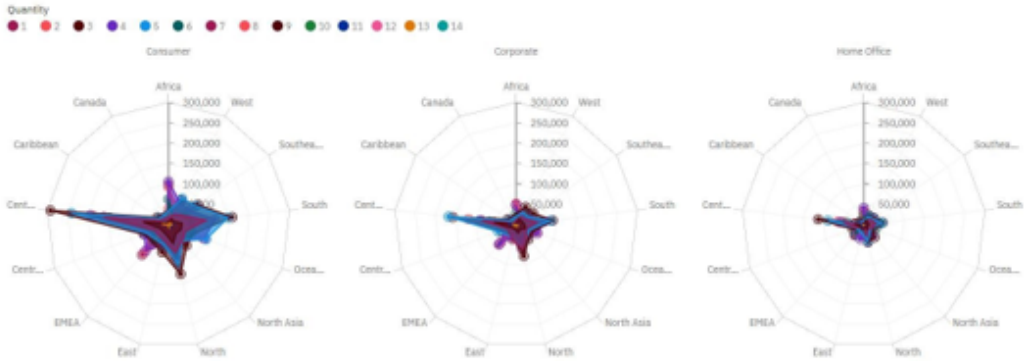


Sales By Segment Analysis

Sales Vs Profit By Countries



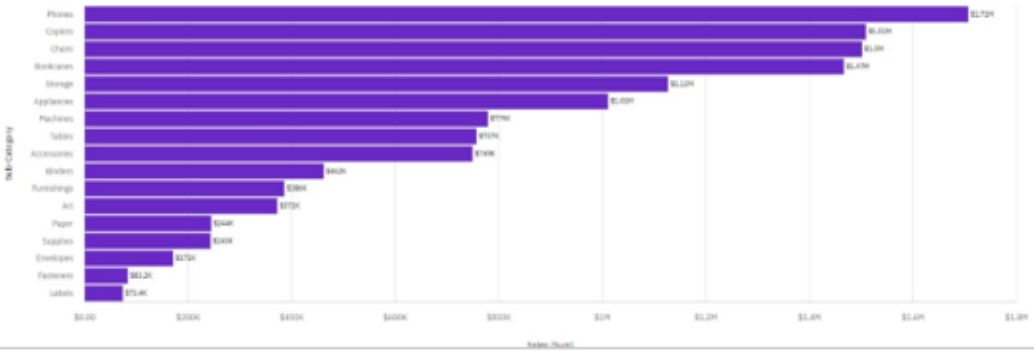
Regional Quantity And Sales Using Radar Chart For Various Segments



Country Wise Sales Vs Profit Using Word Cloud

Sales	Profit	Quantity	Discount
$12.6M	$1.47M	178K	14%
Sales	Profit	Quantity	Discount

Sales by Sub-Category



Dashboard Link:

https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FGlobal_Supersto re%2BDashboard&action=view&mode=dashboard&subView=model00000183f6968afd_00000000

7.4 Sprint 4

```
top5 = data.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
top5
```

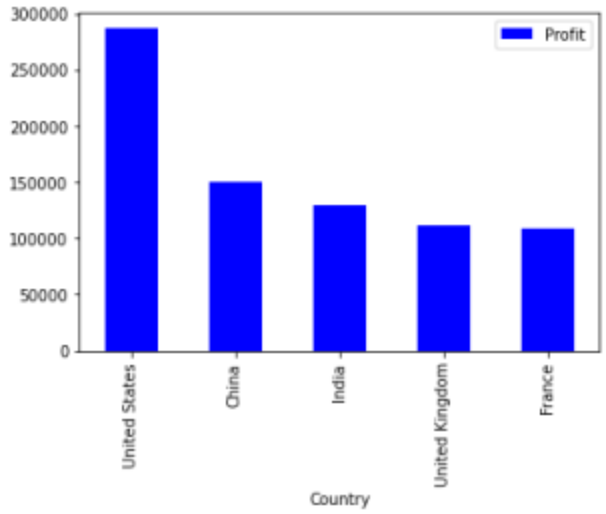
Country	Quantity
United States	37873
France	10804
Australia	10673
Mexico	10011
Germany	7745

```
topprof = data.groupby(['Product Name']).sum()[['Profit']].nlargest(n=5, columns=['Profit'])

topprof
```

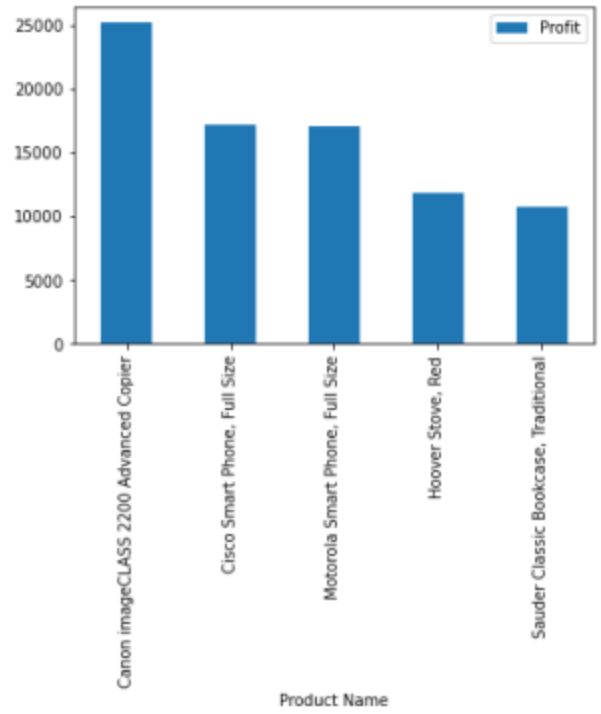
Product Name	Profit
Canon imageCLASS 2200 Advanced Copier	25199.9280
Cisco Smart Phone, Full Size	17238.5206
Motorola Smart Phone, Full Size	17027.1130
Hoover Stove, Red	11807.9690
Sauder Classic Bookcase, Traditional	10672.0730

```
data.groupby(['Country']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest
plt.show()
```




```
data.groupby(['Product Name']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nla
```

```
<AxesSubplot:xlabel='Product Name'>
```




```
data.groupby('Product Name')['Customer ID'].count().sort_values(ascending=True)
```

8. ADVANTAGES AND DISADVANTAGES

Advantages

Data analytics helps an organization make better decisions. Analytics can help with transforming the data that is available into valuable information for executives so that better decisions can be made. This can be a source of competitive advantage if fewer poor decisions are made since poor decisions can have a negative impact on a number of areas including company growth and profitability.The analytics keeps you updated of your customer behavioural changes. Customers tend to change their minds as they are continuously exposed to new information in this era of digitization. With vast amount of customer data, it is practically impossible for organizations to make senses of all the changes in customer perception data without using the power of analytics. Analytics gives you insights into how your target market thinks and if there is any change. Personalization of products and services. Customers crave products and services that can meet their individual needs. Analytics can help companies keep track of what kind of service, product, or content is preferred by the customer and then show the recommendations based on their preferences.

Disadvantages

Low quality of data

A top down approach is required where the business questions that need to be answered need to be known first and what data is required to answer these questions can then be determined. data may have been collected for historical reasons may not be suitable to answer the questions that we ask today. If the data quality is poor, the decision made by using this data is also going to be poor.

Privacy concerns

Sometimes, data collection might breach the privacy of the customers as their information such as purchases, online transactions, and subscriptions are available to companies whose services they are using. Some companies might exchange those datasets with other companies for mutual benefit. Certain

data collected can also be used against a person, country, or community.

9.CONCLUSION

Historically, sales success has relied on intuition and subjectivity. Sales reps conduct in-depth research on prospects and then chase the most suitable fits. This process relies on trial and error to figure out prospects' expectations and apply the rep's intuitions to understand prospects' pain points. Reps use sales data analysis to make critical decisions. Adopting a data-driven sales approach takes subjectivity out of the equation and makes the whole process of selling more predictable and efficient. sales data and proper sales data analysis tools can speed up your growth rapidly.

10.FUTURE SCOPE

Visualization is quickly growing in importance in the world of big data. Businesses can no longer afford to operate without graphs and dashboards in their arsenal since they're extremely useful for presenting and understanding complex information. All the technology surrounding us, and mobile devices, in particular, has turned into data sources that changed the way organizations collect and crunch numbers with the goal of generating actionable insights. The best way for humans to perceive complex data is by means of sight. Advanced visualization in the form of video makes the information consumption more entertaining and further improves understanding of the underlying message.he trend we're going to see is the emergence of hybrid services allowing organizations to connect their private clouds with public clouds to create a single framework.This will be done with the goal of having a secure environment where organizations will be able to process and analyze data, and then present it in a relevant visual format.

11.APPENDIX

Github Link:

https://github.com/IBM-EPBL/IBM-Project-26079-1660009734

Project Demo Link:

https://drive.google.com/file/d/13WCSOgIRDIPEenh9OatIDHvlyBzW_sXm/view?usp=share_link