



Deccan Education Society's
**NAVINCHANDRA MEHTA INSTITUTE OF
TECHNOLOGY AND DEVELOPMENT**
NAAC Accredited "B++"

DESIGNING A DASHBOARD FOR SALES ANALYSIS

SUBMITTED BY

SIDDHARTH GANPAT JAMBHAVDEKAR

C22055

UNDER THE GUIDANCE OF

Dr. Swapnali Mahadik

[2023-2024]

Submitted to University of Mumbai
in partial fulfilment of the requirements for qualifying
MASTER OF COMPUTER APPLICATION
Examination

Deccan Education Society's
NAVINCHANDRA MEHTA INSTITUTE OF TECHNOLOGY AND DEVELOPMENT

PROJECT CERTIFICATE

This is to certify that the Project done at **Data Talks Lab** by Mr. **Sidharth Ganpat Jambhavdekar** (Seat No. **C22055**) in partial fulfilment for MCA Degree Examination has been found satisfactory. This report had not been submitted for any other examination and does not form part of any other course undergone by the candidate.

Dr. Swapnali Mahdik
Internal Guide

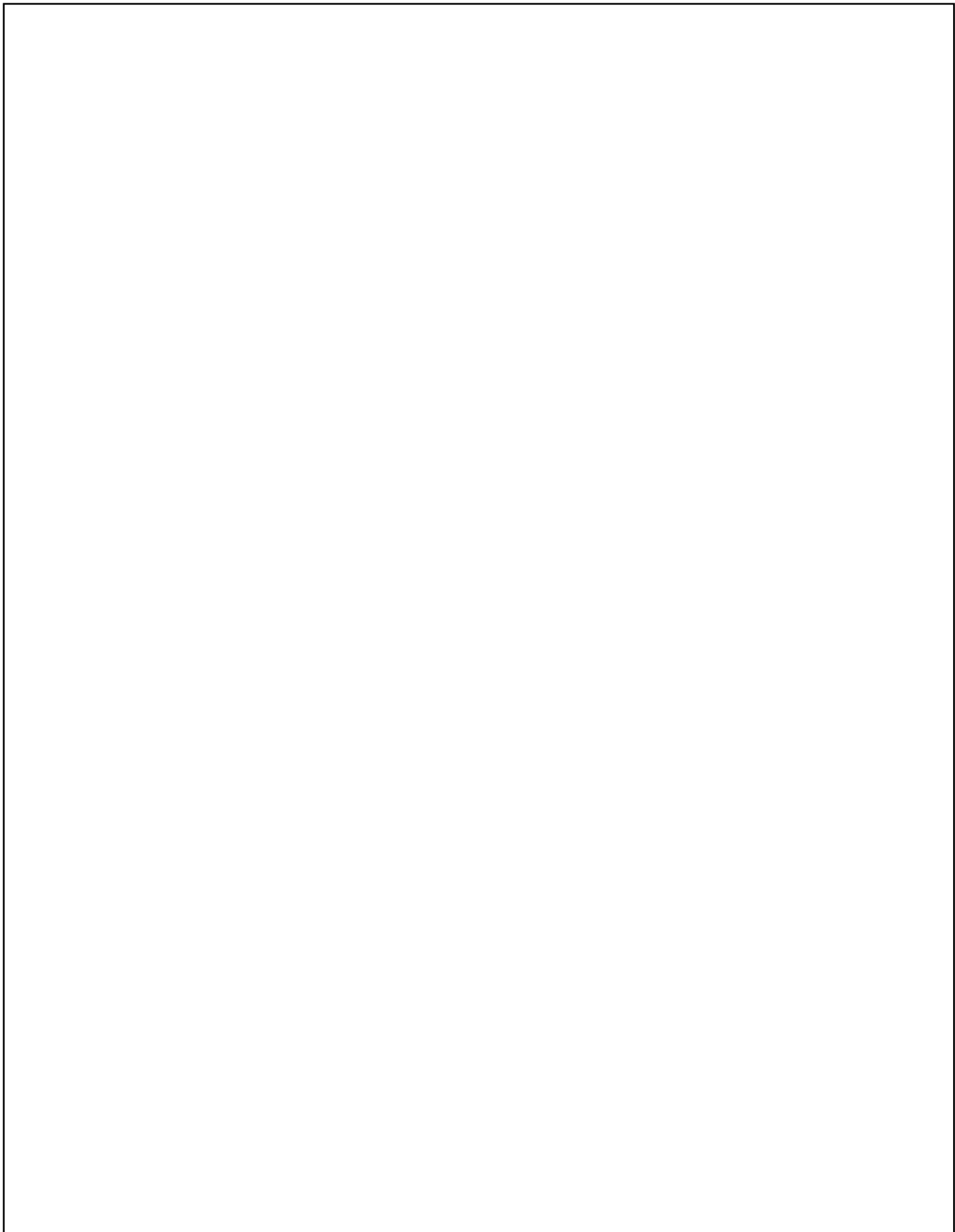
Dr. Rasika Mallya
I/C Director

EXAMINED BY

EXTERNAL EXAMINER

DATE:

College Stamp



Acknowledgement

I take this opportunity to express my profound gratitude to our **Director, Dr. Rasika Mallya**, our **HOD Dr. Sulakshana Vispute** whose wisdom as well as foresight, I continually benefit from. I am indebted to them for extending me all the cooperation and patronage they render towards us. They have been wonderful supporters and continuous motivators for me which helped me in turning my ordinary project into an excellent one.

I also take this opportunity to express a deep sense of gratitude to Company project guide **Karan Thakur, Data Talks Lab** for his cordial support, valuable information and guidance, which helped me in completing this task through various stages.

I am obliged to staff members of Data Talks Lab for the valuable information provided by them in their respective fields. I am grateful for their cooperation during the period of my internship. It is my privilege to express my profound gratitude and deep regards to my project guide **Dr. Swapnali Mahdik** for her exemplary guidance, monitoring and constant encouragement throughout the course of this summer internship project. The blessing, help and guidance given by her/him time to time shall carry me a long way in the journey of life on which I am about to embark.

Lastly, I thank almighty, my parents, brother, sisters and friends for their constant encouragement without which this summer internship project would not be possible.

Abstract

The retail markets have become extensively competitive and each player in the retail market is striving for the ability to optimize the marketing serving processes, while satisfying the customer expectations. Therefore, it is important for any business/firm to manage and channelize the data to achieve the customer's delight as well as in generating healthy profits. With regard to big retail players internationally as well as in the USA, data mining or rather big data analytics is now being at every single stage of the retail market/business process, i.e., tracking customer order placements and predicting the forecast sales of the particular product, optimizing the product sales and the offers based on consumer preferences, tracking the emerging products in the market, forecasting and predicting the sales and future demand based on the predictive simulation tools. In parallel to this, recognizing the customers' expectations and interest in specific product types based on their previous purchase actions, and working out the best technique to approach them through targeted marketing efforts and ultimately what to sell them next in what configures the core of data analytics. This project is the outcome of a descriptive research on the past, present, and future of retail industry and the application of business analytics in shaping appropriate marketing strategies with data sources, data structures, and DAX query language through dashboard in Power BI. The project aims to show how we can use the Power BI with business-oriented retail analytics data using DAX query language and its performance on presenting the dashboard to the end users. So, in this project, I have created an analytical dashboard to know historic trends and business performance, and also to know which products are sold mostly, which are the top regions and managers/market performance. Additionally, I have created what if analysis for future planning on the basis of historic trend - this dashboard is created for stake holders to know business growth trend and functional areas and with the visualizations. This project represents the large dataset into visualization form to quickly see the performances of all the commodities.

INDEX

Data Analytics

Chapter		Contents	Page No.
I		Introduction	1-7
	1.1	Company Profile	
	1.2	Background	
	1.2.1	Power BI	
	1.2.2	DAX Queries	
	1.2.3	Retail Analytics	
	1.2.4	Accessibility in Power BI	
	1.2.5	Universal Design	
	1.2.6	To access these dashboards	
	1.2.7	GUI	
II		Proposed System	8-12
	2.1	Scope of Work	
	2.2	Operating environment	
	2.4	Proposed System Overview	
	2.5	System Overview(flowchart)	
	2.6	Limitations of the System (legacy system)	
III		Gantt Charts	13-14
IV		Methodology	15-19
	4.1	Data Source	
	4.2	Data Flow Process	
	4.3	How to upload raw data into Power BI	
	4.4	Data Analysis Expressions (DAX)	
V		Analysis and Design	20-25
	4.1	Data Preprocessing specification	
	4.2	System Flow Chart	
	4.3	Use Case Diagram	
	4.4	Sequence Diagram	
	4.5	Activity Diagram	
	4.6	Component Diagram	
	4.7	Data Flow Diagram	
V		Implementation and Evaluation	26-30
VI		Conclusion	31-32
VII		References	33-34

List of Figures, if any:

Sr. No.	Figure Number	Title	Page Number
1.	Figure 1	Figure showing overall accessibility of the project	6
2.	Figure 2	Figure showing a filter option available to get only particular region data	6
3.	Figure 3	Figure showing how to export data from the dashboard	7
4.	Figure 4	Gantt Chart	14
5.	Figure 5	Excel raw data	17
6.	Figure 6	Power BI flow process	17
7.	Figure 7	Connections between multiple tables for calculations	18
8.	Figure 8	Flow chart	23
9.	Figure 9	Use Case Diagram	23
10.	Figure 10	Sequence Diagram	24
11.	Figure 11	Activity Diagram	24
12.	Figure 12	Component Diagram	25
13.	Figure 13	Data Flow Diagram	25
14.	Figure 14	Main Dashboard	27
15.	Figure 15	Forecast Dashboard	27
16.	Figure 16	Product Details Dashboard	28
17.	Figure 17	Product Analysis Dashboard	28
18.	Figure 18	Shipping Details Dashboard	29
19.	Figure 19	Top 5 Customers Dashboard	29
20.	Figure 20	Bottom 5 Customers Dashboard	30

List of Tables, if any:

Sr. No.	Table Number	Title	Page Number
1.	Table 1	Accessibility Checklist	4-5

List of Symbols / acronyms, if any:

Sr. No.	Symbol / Acronym	Meaning / Name if any
1.	Σ	Summation

I. INTRODUCTION

1.1 COMPANY PROFILE

- Data Talks Lab, founded in January 2019 by Karan Thakur, who serves as both the Founder and CTO, is dedicated to helping everyday individuals build the skills they need to launch successful careers in analytics and Full Stack Web Development. Karan Thakur, a Microsoft Certified Trainer, leverages his expertise to guide students through custom learning paths designed to meet their unique needs.
- The company offers 1-on-1 coaching and world-class courses in Excel, Power BI, SQL, Tableau, and Machine Learning. This personalized approach has transformed the careers of over 1500 students, providing them with the knowledge and confidence to excel in their chosen fields.
- Through these tailored educational experiences, Data Talks Lab is committed to empowering individuals to achieve their professional goals and thrive in the competitive tech industry.

1.2 BACKGROUND

1.2.1 Power BI

Power BI is a business analytics service provided by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities to create their own reports and dashboards for the end users. Our data may be in an Excel spreadsheet, or a collection of a cloud based .pbix file which is designed for to use with Power BI desktop. Power BI was initially released in 2014, operating system: Microsoft windows. Power BI provides cloud- based services, along with desktop interface, called Power BI Desktop. The key components of the Power BI are (I have mentioned only 3 which I have used in this project):

- Power BI desktop: designing and publishing reports and dashboards to the service.
- Power BI service: software as a service.
- Power BI mobile apps: iOS, android, and windows phones.

The objective of my project is to analyze the retail sales data of the items sold across multiple countries and stores. I have used visualization and fields tools to create the data sale analysis charts shown in this report and project.

Benefits of Retail analysis: with the help of Power BI, a retailer can even analyze if there are any issues in its products, transportation, services. Transactions, supply chain, buying behaviors, multi-channel sales, trends and global coverage can also be tracked with this tool to maximize profit. The sales data stored yearly could be used to improvise the


business. To store the yearly sales data, products data, and services data, the retailers usually use the excel sheet or SAP tools.

Instead of presenting each and every data in different layouts, I have used Power BI tools and presented all the charts information in five layouts with the complete information for the year 2014 and all the years.

By looking at this retail analysis layout, the audience who are seeing the layout will come to an understanding what profit they have made in multiple countries. For all the charts in the top corner there are 3 symbols i.e., FILTER, HIGHLIGHT, and NONE. We can choose what option we need to demonstrate the charts regarding our requirements. Filtering is like filtering the icons to what to present in all the charts and what not to show. Highlight is just like highlighting the requirements mentioned. None is like not to show which is not needed. For visual interaction there is an option called edit interact, using this option we can directly visualize and edit the data. I have created the Power BI tool for the visualization of data for the years 2011-2014. Kindly note that I am only presenting the data of the year 2014 for a few layouts.

1.2.2 DAX Queries

Syntax Formula:

- ' ' single quotes indicates to the related table [] related to the table name.
- Measures for raw data
- Column's calculations are identified by summation symbol Σ in the fields' Box with cross lines and calculator symbol is calculated columns.
- Measure is identified through calculator symbol  from the fields.

All queries are in DAX language with calculated column and measures using formulas and syntax and measure as syntax consumes less space it is visualize, whereas measures contain lot of space and it is not visualized in raw data.

1.2.3 Retail Analytics

As we are all aware, the retail business sales output is the fulcrum of a company's performance. To achieve the best output sales, there are many parameters to be mastered.

- Detailed analysis of the particular product in the area where it must be marketed. The product demand analysis must be done practically by interacting with different segment people of that area.
- Analysis of competitive brands. Evaluation of positives and negatives of our brand product to be marketed. Plan to make improvisation of product and give more emphasis for acceptability of the product by all segment people of society.

- Quality of product, which is a very important parameter of product sales. To achieve quality, one should look for quality raw materials at the most reasonable rate, which will help to sell products at best marketing competitive price.
- Marketing of finished product. Selection of a marketing team is very challenging. The key is to recruit the marketing team. The team efforts, right from sales manager to field salesman gets the best of product sales.
- Production of product in relative, with demand in market is more important. If demand is more and production is less, because of shortage of raw material or production staff, will give room for competitive brands to dilute the sales curve.
- After establishing the best sales of product in the market with outstanding co-ordination of above parameters. The company management must give more emphasis for the consistency product sales by research and sales development. The company which has better research and development will perform best in the market.

Table 1:

Accessibility checklist

Standard	Y/N	Comments
High Contrast	Y	Available on web service
Focus Mode	Y	Available in each visualization
Titles & Labels	Y	Available in each visualization
Legends	Y	Available in each visualization
Filter / Slicer	Y	Adds filters for more drill down
Cards	Y	Few cards created for KPI
Tabs/Pages	Y	Created different pages/tabs for more clarity of visualization
Data Flow	Y	Data visualize in proper way: 1 st page – overview of business; 2 nd page: product wise 3 rd page: market wise; 4 th page – geographic wise data
Font size & color	Y	We can change font size and color
Icons for visualization	Y	Uploaded category wise or graph wise icons to understand easily
Wallpaper	Y	Uploaded background with the help of PPT
Page Navigation	Y	Created page navigation with the help of button and applied to tabs
Theme	Y	Created universal them for all page but layout is different
Show Data	Y	we can check raw data

Model	Y	we can check and create connection between multiple tables with primary key
Sort Option	Y	Sort option is available in each visualization
View and Hide	Y	we can hide any visual anytime
Measures	Y	Multiple measures created for calculations, we can check by clicking on measures
Transform	Y	we can Transform data and modify as per required
Format painter	Y	want to apply same format for other visual them by using this we can do
Drill down	Y	we can drill down with down arrow in all visualization
Disabilities	Y	The above options are accessible to disable people through key board
Color contrast	Y	With the options published on web, using low, high, yellow, blue colors

1.2.4 Accessibility in Power BI

While working with Power BI, I considered the different types of users who may interact with these reports. I created reports that are easily navigated and understood by keywords and button / tabs / filters.

1.2.5 Universal Design

I created Universal design for all pages with background design and colors to make it simple and attractive and user friendly. Created a few tabs to separate all pages and applied navigate function to go directly on that page.

1.2.6 To access these dashboards

To make it accessible I must publish this report from desktop to web service and mentioned users email id so they can access these reports. We can also set limitations.

In all pages I have created multiple tabs. If you click on any tabs, you will go directly to that page. A retail analytics dashboard is created to quickly see the performance of all the products that a company sell. This layout relationship calculations are done with DAX query language and then visualizations are created through fields and filters. This dashboard contains 5 layouts that are overview, product wise, market wise, geographic wise and what if analysis. The overview of my dashboard/report includes how the sales are done regarding different products in multiple countries and regions and how many customers are buying the products how it is transported how many managers the best profits and comparisons between quartered and month regarding sales. What if analysis to improve the future business in retail analytics. Our cerebrum attracts more to the colors and visualizations rather than written documents. Accessibility and unit testing

performance is done with my dashboard. For accessibility When spotlight option is selected the other are not seen and there are many options like drill up, drill down, colors, transparency, focus mode, show as a table etc.



Figure 1:
Figure showing overall accessibility of the project

Africa	Caribbean	Central Asia	EMEA	North Asia	South	West
Canada	Central	East	North	Oceania	Southeast Asia	

Figure 2:
Figure showing a filter option available to get only particular region data

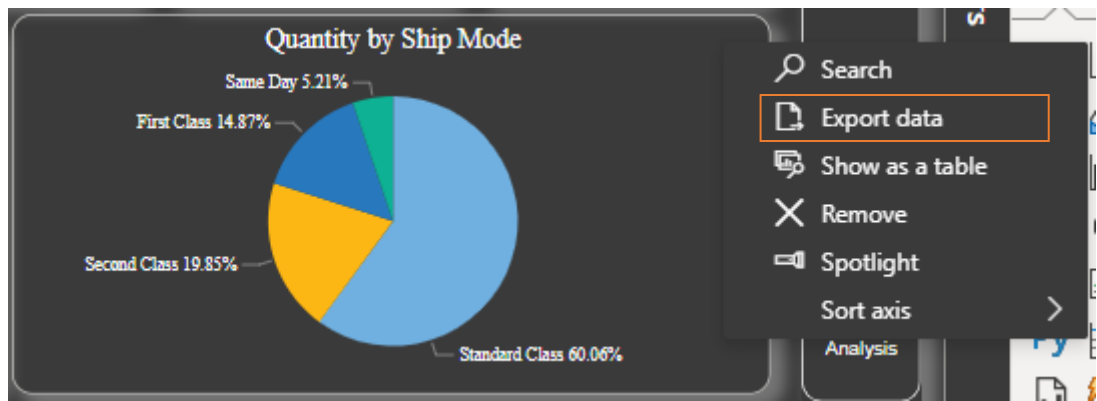


Figure 3:

Figure showing how to export data from the dashboard

1.2.7 GUI

The GUI I have worked on is the DAX query Language that is backend program of the application. Power BI interface includes data source, theme, query language, visualizations, we can manipulate as per our requirement.

Power BI have 2 types of applications:

- Desktop - in Desktop version you can create / modify reports
- Web - you can publish your reports from Desktop to web so everyone can access your report or dashboard

II. PROPOSED SYSTEM

2.1 Scope of work

Objective

The objective of this project is to design and develop a comprehensive sales analysis dashboard using Power BI, leveraging sales data from the years 2011 to 2014. This dashboard aims to provide actionable insights, enhance decision-making, and improve overall sales performance by visualizing key metrics and trends. The specific objectives include:

1. Centralized Data Visualization:

- Consolidate sales data from various sources into a single, interactive dashboard.
- Ensure stakeholders have a unified view of sales performance across the specified years (2011-2014).

2. Enhanced Sales Performance Tracking:

- Track and monitor key sales metrics such as total sales, sales growth, revenue by product, and sales by region.
- Identify top-customers and bottom-customers.
- Detect underperforming areas that require attention and improvement.

3. Trend Analysis:

- Analyze historical sales data to identify trends and patterns over the four-year period.
- Provide visual representations of sales trends, seasonal patterns, and cyclical behavior.
- Enable forecasting based on historical data trends.

4. Improved Decision-Making:

- Provide stakeholders with easy-to-understand visualizations to support data-driven decision-making.
- Highlight key insights that can drive strategic decisions, such as market expansion, product development, and sales strategies.
- Enable quick identification of opportunities and risks through real-time data analysis.

5. Interactive and User-Friendly Interface:

- Develop an intuitive and interactive dashboard interface that allows users to explore data dynamically.
- Implement features such as filters, slicers, and drill-down capabilities for personalized data exploration.
- Ensure the dashboard is user-friendly and accessible to users with varying levels of technical expertise.

6. Operational Efficiency:

- Reduce the time and effort required to generate sales reports and analyses.
- Automate data updates and report generation to provide timely insights.

- Eliminate manual data processing errors through automated data integration and visualization.

7. Continuous Improvement:

- Establish a foundation for continuous improvement in sales analytics and reporting.
- Collect feedback from users to iteratively enhance the dashboard.
- Keep the dashboard updated with new features, data sources, and analytical capabilities as the organization's needs evolve.

Data Preparation

- 1. Data Collection:** Gather sales data from relevant sources for the years 2011 to 2014.
- 2. Data Cleaning:** Clean and preprocess the data to ensure accuracy and consistency.
- 3. Data Transformation:** Transform the data into a format suitable for analysis in Power BI.
- 4. Data Integration:** Integrate data from various sources (e.g., CRM systems, ERP systems, spreadsheets).

2.2 Operating environment

2.2.1 Hardware

- Processor: Intel(R) Core (TM) i7-8650U CPU 1.90GHz 2.11 GHz.
- RAM: Minimum 16 GB or Greater.
- Hard disk: HDD, SSD 20 GB (Free Space).

2.2.2 Software

- Software : Power Bi
- Operation System: 64-bit operating system, Windows 10 or higher.

2.2.3 Project Overview

- Front End : Power Bi Desktop
- Back End : Excel, Azure and DAX query language.

2.3 Proposed System Overview

1. Data Collection and Integration

- **Data Sources:** Integrate data from multiple sources such as CRM systems, ERP systems, sales databases, and spreadsheets. This includes historical sales data from 2011 to 2014.

- **ETL Process:** Implement an Extract, Transform, Load (ETL) process to gather, clean, and integrate data into a centralized data warehouse. Tools like SQL Server Integration Services (SSIS), Azure Data Factory, or other ETL tools can be used.
- **Data Storage:** Store the integrated data in a centralized, secure, and scalable data warehouse, such as Azure SQL Database or another cloud-based storage solution.

2. Data Modeling

- **Data Schema:** Design a data schema that accurately represents the relationships between different data tables (Customers, Orders, Products, Category, Sub_Category, Calendar, Territory, and Superstore_Dataset2011-2014).
- **Data Relationships:** Define and enforce relationships between tables using primary and foreign keys to ensure data integrity and enable efficient querying.
- **Data Refresh:** Implement a data refresh schedule to keep the data up-to-date, ensuring that the dashboard reflects the latest information.

3. Power BI Dashboard Development

- **Data Import:** Import the cleaned and modeled data into Power BI using the built-in connectors.
- **Data Visualization:** Design and develop interactive visualizations to represent key sales metrics, trends, and insights. These can include charts, graphs, tables, and maps.
- **Dashboard Layout:** Create an intuitive and user-friendly dashboard layout that allows users to easily navigate and explore the data. Include interactive features such as filters, slicers, and drill-down capabilities.
- **Custom Calculations:** Use DAX (Data Analysis Expressions) to create custom measures and calculated columns for advanced data analysis.

4. Key Features and Functionalities

- **Sales Performance Metrics:** Display key metrics such as total sales, sales growth, revenue by product, sales by region, and profit margins.
- **Trend Analysis:** Visualize historical sales trends and patterns over the specified period (2011-2014).
- **Segmentation Analysis:** Provide insights into customer segments, product categories, and geographical regions.
- **Interactive Filters:** Allow users to filter data by different dimensions such as time period, region, product category, and customer segment.
- **Drill-Down Capabilities:** Enable users to drill down into detailed views for more granular analysis.
- **Forecasting:** Implement basic forecasting features to predict future sales based on

- historical data.

5. User Access and Security

- **User Roles and Permissions:** Define user roles and permissions to control access to different parts of the dashboard based on user responsibilities.
- **Data Security:** Implement data security measures to protect sensitive sales information. Use Power BI's security features to manage data access and sharing.

6. Deployment and Maintenance

- **Deployment:** Deploy the Power BI dashboard to the Power BI service, making it accessible to authorized users through web and mobile interfaces.
- **Training and Support:** Provide training sessions and documentation to ensure users are proficient in using the dashboard. Offer ongoing support to address any issues or enhancements.
- **Continuous Improvement:** Collect user feedback and continuously improve the dashboard based on evolving business needs and user requirements. Regularly update the dashboard with new features and data sources.

III. Gantt Chart

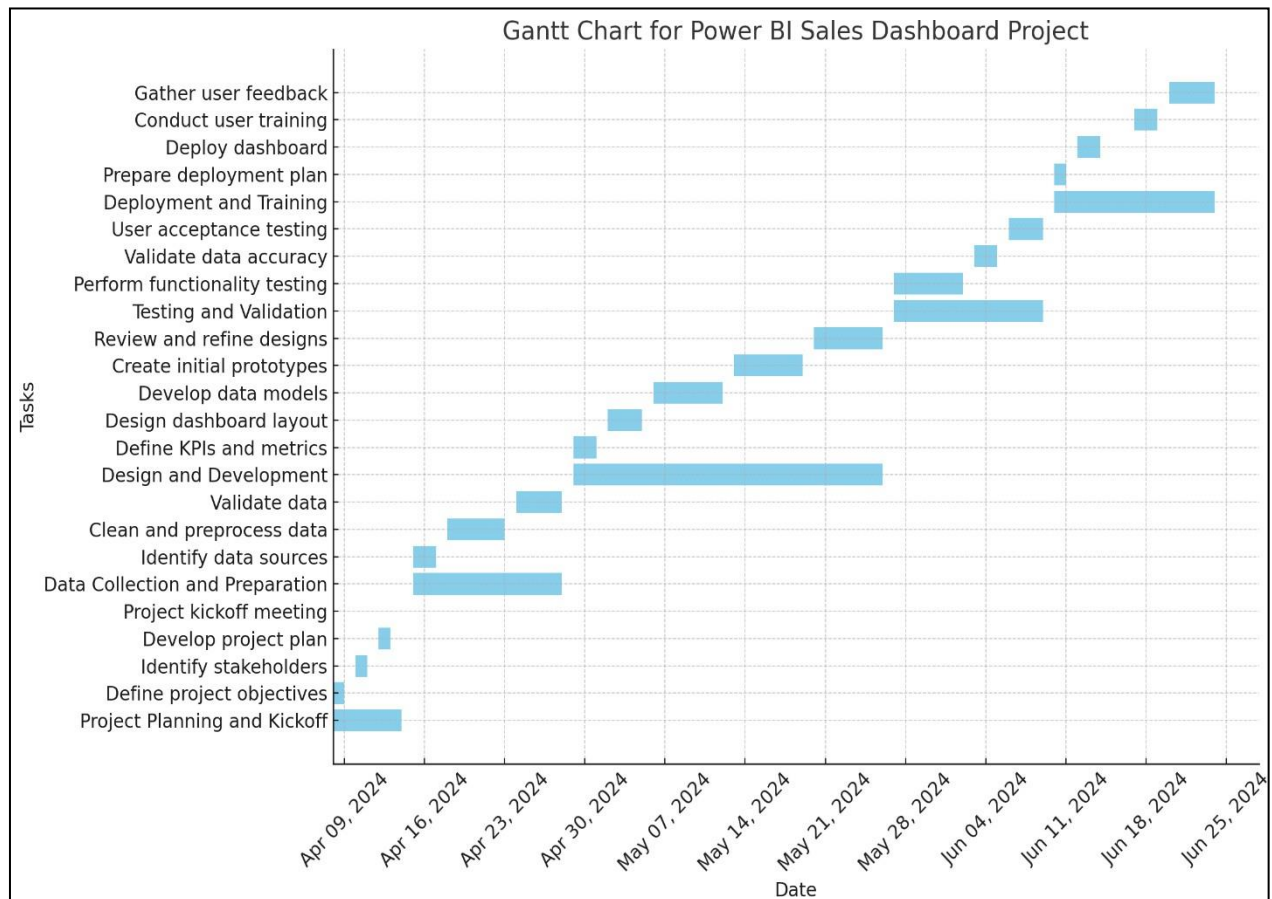


Figure 4:
Gantt Chart

IV. METHODOLOGY

The main objective in making this dashboard is that the reader can directly analyze what we had in the last financial year. Which areas are not making sales, or which is our best product, top 5 customers, etc. Readers can take decisions for business growth, and it is easy to understand business performance.

4.1 Data Source

Retail raw data taken from [kaggle.com](https://www.kaggle.com)

- Row ID
- Order ID
- Order Date
- Ship Date
- Ship Mode
- Customer ID
- Customer Name
- Segment City
- State
- Country Postal Code
- Market
- Region
- Product ID
- Category
- Sub-Category
- Product Name
- Sales
- Quantity
- Discount
- Profit
- Shipping Cost
- Order Priority

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount
1	MX-2014-02-10-2014	06-10-2014	Standard (SC-20575	Sonia Co	Consumer	Mexico	Distrito Fe	Mexico			LATAM	North	OFF-LA-1C	Office Sup Labels	Hon File F	13.08		3	
2	MX-2012-15-10-2012	20-10-2012	Standard (KW-16570	Kelly Willi	Consumer	Dos Queb	Risaralda	Colombia			LATAM	South	FUR-FU-1C	Furniture	Furnishin	Tenex Clo	252.16	8	
3	MX-2012-15-10-2012	20-10-2012	Standard (KW-16570	Kelly Willi	Consumer	Dos Queb	Risaralda	Colombia			LATAM	South	FUR-BO-1C	Furniture	Bookcase	Ikea 3-She	193.28	2	
4	MX-2012-15-10-2012	20-10-2012	Standard (KW-16570	Kelly Willi	Consumer	Dos Queb	Risaralda	Colombia			LATAM	South	OFF-BI-10	Office Sup Binders	Cardinal B	35.44	4		
5	MX-2012-15-10-2012	20-10-2012	Standard (KW-16570	Kelly Willi	Consumer	Dos Queb	Risaralda	Colombia			LATAM	South	OFF-AR-1C	Office Sup Art	Sanford C	71.6	2		
6	MX-2012-15-10-2012	20-10-2012	Standard (KW-16570	Kelly Willi	Consumer	Dos Queb	Risaralda	Colombia			LATAM	South	OFF-EN-1C	Office Sup Envelope	GlobeWei	56.12	2		
7	MX-2013-27-09-2013	01-10-2013	Standard (DP-13000	Darren Po	Consumer	São Paulo	São Paulo	Brazil			LATAM	South	OFF-EN-1C	Office Sup Envelope	GlobeWei	56.12	2		
8	MX-2013-27-09-2013	01-10-2013	Standard (DP-13000	Darren Po	Consumer	São Paulo	São Paulo	Brazil			LATAM	South	TEC-MA-1C	Technology Machines	Konica Cai	344.64	3		
9	MX-2013-27-09-2013	01-10-2013	Standard (DP-13000	Darren Po	Consumer	São Paulo	São Paulo	Brazil			LATAM	South	OFF-SU-1C	Office Sup Supplies	Elite Box C	97.36	4		
10	MX-2013-27-09-2013	01-10-2013	Standard (DP-13000	Darren Po	Consumer	São Paulo	São Paulo	Brazil			LATAM	South	TEC-AC-1C	Technology Accessori	Enermax F	341.52	2		
11	MX-2013-27-09-2013	01-10-2013	Standard (DP-13000	Darren Po	Consumer	São Paulo	São Paulo	Brazil			LATAM	South	OFF-BI-10	Office Sup Binders	Wilson Joi	12.06	3		
12	MX-2013-05-03-2013	12-03-2013	Standard (TB-21250	Tim Brock	Consumer	Managua	Managua	Nicaragua			LATAM	Central	OFF-FA-1C	Office Sup Fasteners	Accos Stag	20.76	3		
13	MX-2013-05-03-2013	12-03-2013	Standard (TB-21250	Tim Brock	Consumer	Managua	Managua	Nicaragua			LATAM	Central	FUR-CH-1C	Furniture	Chairs	Hon Steel	210.64	4	
14	MX-2014-24-09-2014	01-10-2014	Standard (PK-18910	Paul Knut	Home Off	Managua	Managua	Nicaragua			LATAM	Central	OFF-EN-1C	Office Sup Envelope	Kraft Mail	80.1	3		
15	MX-2014-18-10-2014	23-10-2014	Standard (JK-15625	Jim Karlss	Consumer	Brasilia	Federal Di	Brazil			LATAM	South	OFF-AR-1C	Office Sup Art	Stanley C	132.64	4		
16	MX-2014-18-10-2014	23-10-2014	Standard (JK-15625	Jim Karlss	Consumer	Brasilia	Federal Di	Brazil			LATAM	South	OFF-FA-1C	Office Sup Fasteners	Advantus	12.94	1		
17	US-2013-1-26-06-2013	01-07-2013	Standard (HE-14800	Harold Enj	Corporate	San Migue	Panama	Panama			LATAM	Central	OFF-EN-1C	Office Sup Envelope	Jiffy Peel	18.84	2		
18	US-2013-1-26-06-2013	01-07-2013	Standard (HE-14800	Harold Enj	Corporate	San Migue	Panama	Panama			LATAM	Central	FUR-FU-1C	Furniture	Furnishin	Deflect-O	308.28	7	
19	US-2013-1-26-06-2013	01-07-2013	Standard (HE-14800	Harold Enj	Corporate	San Migue	Panama	Panama			LATAM	Central	OFF-EN-1C	Office Sup Envelope	GlobeWei	40.176	2		
20	US-2013-1-26-06-2013	01-07-2013	Standard (HE-14800	Harold Enj	Corporate	San Migue	Panama	Panama			LATAM	Central	OFF-EN-1C	Office Sup Envelope	Jiffy Clasp	8.784	3		
21	MX-2013-19-12-2013	25-12-2013	Standard (JE-15610	Jim Epp	Corporate	San Cristó	Chiapas	Mexico			LATAM	North	FUR-CH-1C	Furniture	Chairs	Hon Rocki	273.472	4	
22	MX-2013-29-10-2013	02-11-2013	Second CI TB-21055	Ted Butte	Consumer	Santiago	Santiago	Chile			LATAM	South	TEC-AC-1C	Technology Accessori	Enermax F	27	1		
23	MX-2013-29-10-2013	02-11-2013	Second CI TB-21055	Ted Butte	Consumer	Santiago	Santiago	Chile			LATAM	South	OFF-SU-1C	Office Sup Supplies	Acme Box	207	9		
24	US-2014-03-10-2014	08-10-2014	Standard (DW-13195	David Wie	Corporate	Panama C	Panama	Panama			LATAM	Central	OFF-BI-10	Office Sup Binders	Acco Bind	60.66	3		

Figure 5:

Excel raw data

4.2 Data Flow Process

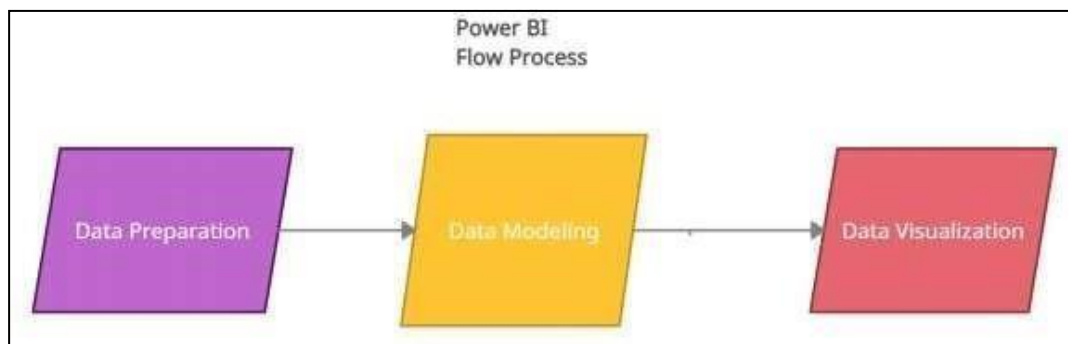


Figure 6:

Power BI flow process

1. **Data preparation** – organizing the data in a flat table.
2. **Data modeling** – It is used to connect multiple data sources in BI tool using a relationship.
3. **Data Visualization** – It is the process of transforming huge data sets into graphs charts and other visualizations.

4.3 How to upload raw data into Power BI

To Create Power BI, visualize dashboard, you need to download Power bi desktop and install. Once you have installed, then we need to follow below process (Microsoft Power BI documentation, 2021):

Click on Get Data => Excel workbook => browse your raw file => select required sheets and then click on transform => Power Query editor window will be open => then from each table you have to check all data types, then you have to click on close & apply.

Now your base data is ready.

Date Table:

I have created date table for measures and apply into charts. Below is DAX query:

Date = CALENDAR(MIN(Sales[Order Date]),MAX(Sales[Order Date]))

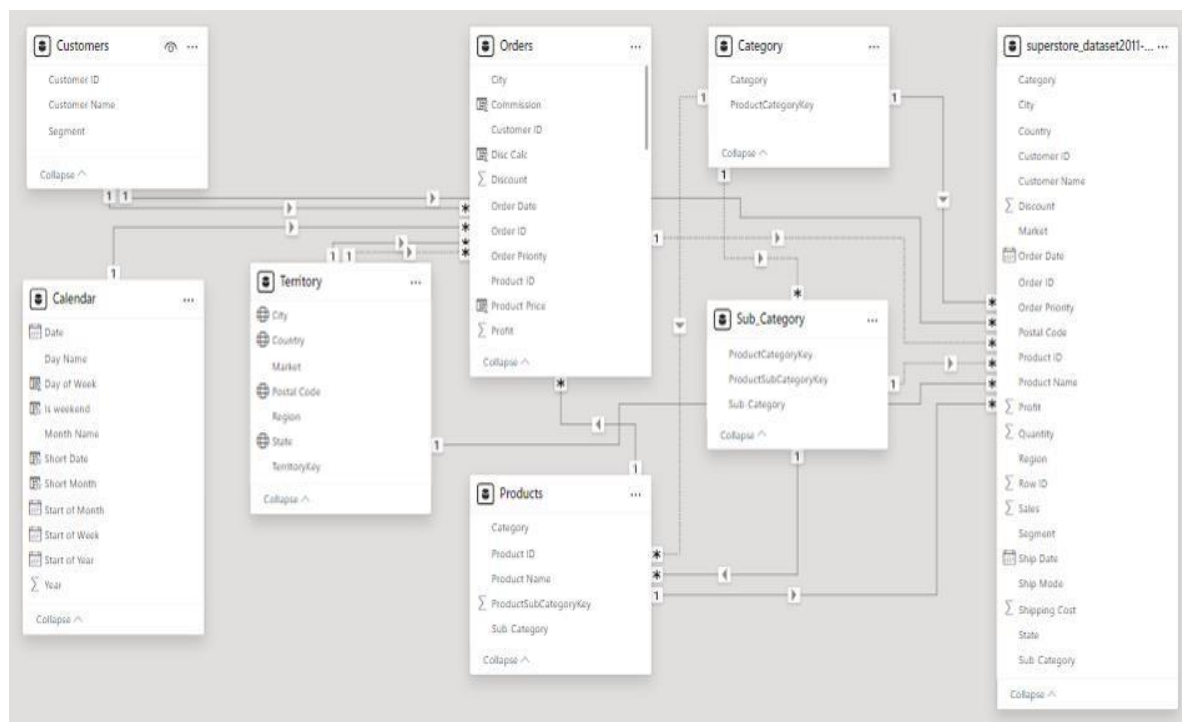


Figure 7:

Connections between multiple tables for calculations

4.4 Data Analysis Expressions (DAX)

Data Analysis Expressions (DAX) is a query programming language that is used throughout Microsoft Power BI for creating calculated columns, measures, and custom tables. It is a collection of functions, operators, and constants that can be used in a formula, or expression, to calculate and return one or more values.

DAX Queries:

- Total Orders = DISTINCTCOUNT(Orders[Order ID])
- Total Profit = [Total Revenue] - [T_Cost]
- Total Revenue = SUM(Orders[Sales])
- Total Sales = SUM(Orders[Sales])
- Weekend_Order = CALCULATE([Total Orders], 'Calendar'[Is weekend] = "Yes")
- YTD Revenue = CALCULATE(Orders[Total Revenue], DATESYTD('Calendar'[Date]))
- Quantity_Sold = SUM(Orders[Quantity])
- Previous_Month_Revenue = CALCULATE([Total Revenue], DATEADD('Calendar'[Date], -1, MONTH))
- Previous_Month_Order = CALCULATE([Total Orders], DATEADD('Calendar'[Date], -1, MONTH))

- High_Ticket_Order = CALCULATE([Total Orders], FILTER(Orders, Orders[Product Price] > [Overall Average Price]))

V. ANALYSIS AND DESIGN

5.1 Data Preprocessing Specification

1. Data Cleaning

- **Handling Missing Values:**

- Identify missing values in all columns.
- Replace missing values with appropriate imputation methods (e.g., mean, median, mode) or use domain-specific rules.
- If necessary, remove records with excessive missing values.

- **Removing Duplicates:**

- Identify and remove duplicate records to ensure data uniqueness.
- Verify duplicates using unique identifiers such as Order ID, Customer ID, etc.

- **Correcting Inconsistencies:**

- Standardize text data to ensure consistency (e.g., consistent use of upper/lowercase, correcting typos).
- Ensure consistency in date formats across all date fields.

- **Outlier Detection and Treatment:**

- Identify outliers in numerical data fields (e.g., sales, profit, discount).
- Treat outliers based on business rules, either by capping extreme values or investigating and correcting erroneous data entries.

2. Data Transformation

- **Normalization and Scaling:**

- Normalize or scale numerical features as needed for analysis (e.g., min-max scaling, z-score normalization).

- **Encoding Categorical Variables:**

- Convert categorical variables into numerical formats using techniques such as one-hot encoding or label encoding.
- Ensure categorical encoding is consistent across training and deployment datasets.

- **Date and Time Transformation:**

- Extract relevant date parts (e.g., year, month, day, quarter) from date fields for time-series analysis.
- Create additional time-based features such as IsWeekend, Weekday, WeekNumber.

- **Aggregating Data:**

- Aggregate sales data to required granularity (e.g., monthly sales, quarterly sales).
- Summarize key metrics (e.g., total sales, average discount, total profit) at different levels (e.g., by region, by product category).

- **Calculating Derived Metrics:**

- Calculate additional metrics such as Sales Growth, Profit Margin, Customer Lifetime Value.
- Ensure derived metrics are accurately calculated and stored for analysis.

3. Data Integration

- **Merging Data from Multiple Sources:**

- Merge data from different tables (e.g., Orders, Customers, Products) using appropriate join keys.
- Ensure the integrity of joins by verifying key relationships.

- **Data Consistency Checks:**

- Verify consistency across merged data (e.g., ensure product IDs match between Orders and Products tables).
- Resolve any data discrepancies identified during the merging process.

- **Creating a Unified Dataset:**

- Combine cleaned and transformed data into a single unified dataset.
- Ensure the unified dataset includes all relevant fields required for the dashboard analysis.

4. Validation and Verification

- **Data Quality Checks:**

- Perform data quality checks to ensure the accuracy and completeness of the processed data.
- Validate key metrics against known values or benchmarks.

- **Business Rule Validation:**

- Ensure the processed data adheres to business rules and logic (e.g., no negative sales amounts, valid customer segments).

- **Documentation:**

- Document all data preprocessing steps, including methods used for handling missing values, encoding, normalization, and derived metrics.
- Maintain a data dictionary describing all fields and transformations applied.

5.2 System Flow Chart

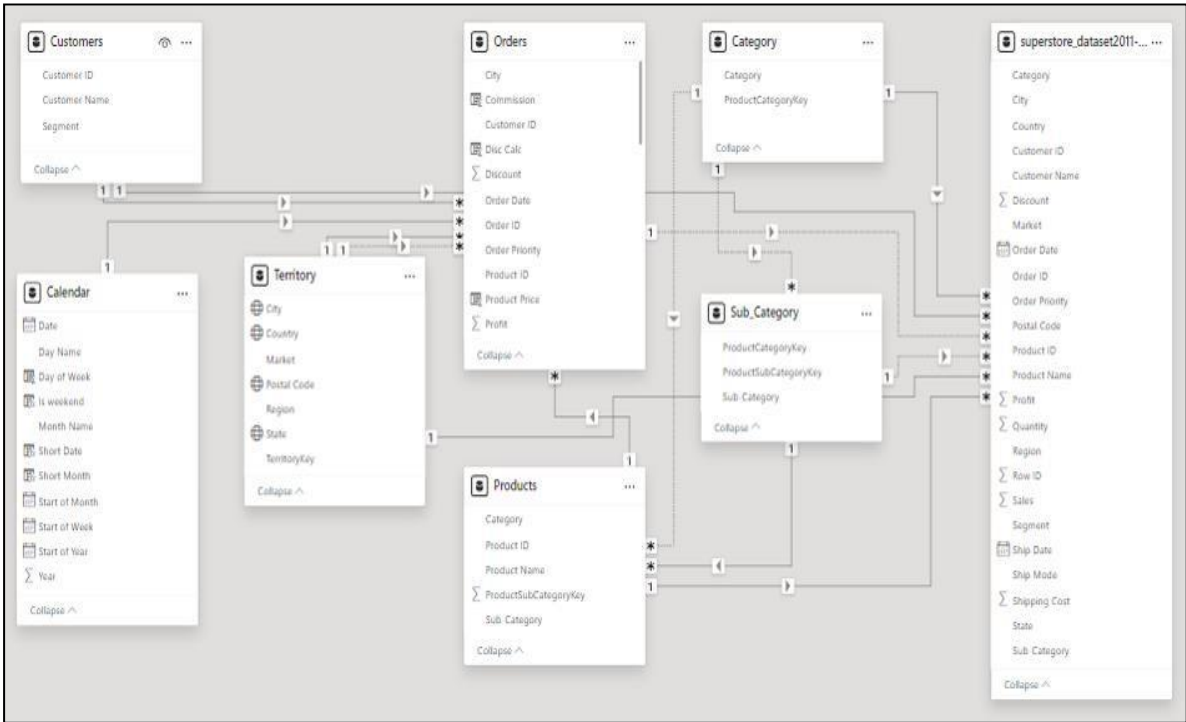


Figure 8:

Flowchart

5.3 Use Case Diagram

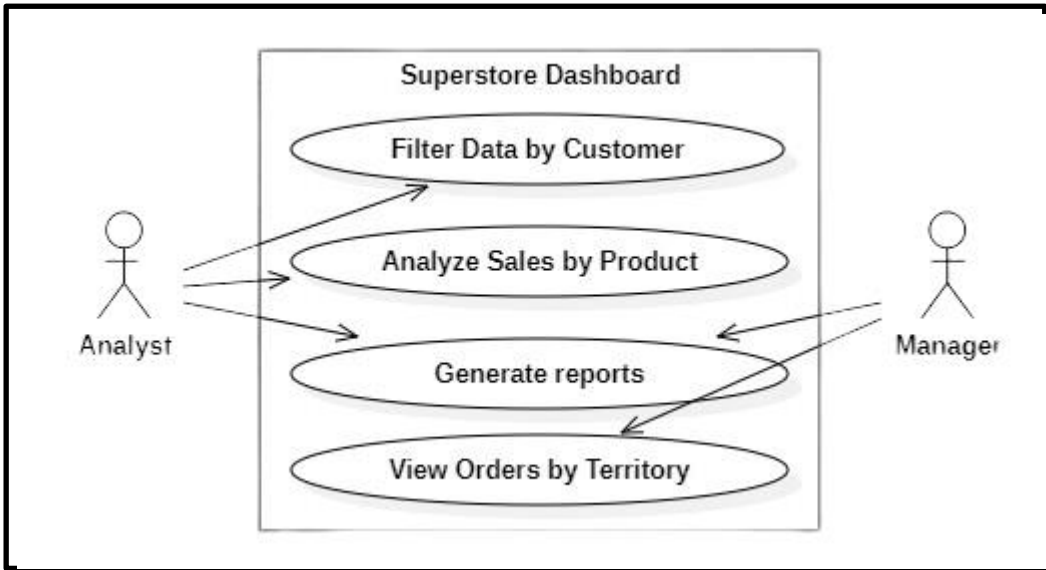


Figure 9:

Use Case Diagram

5.4 Sequence Diagram

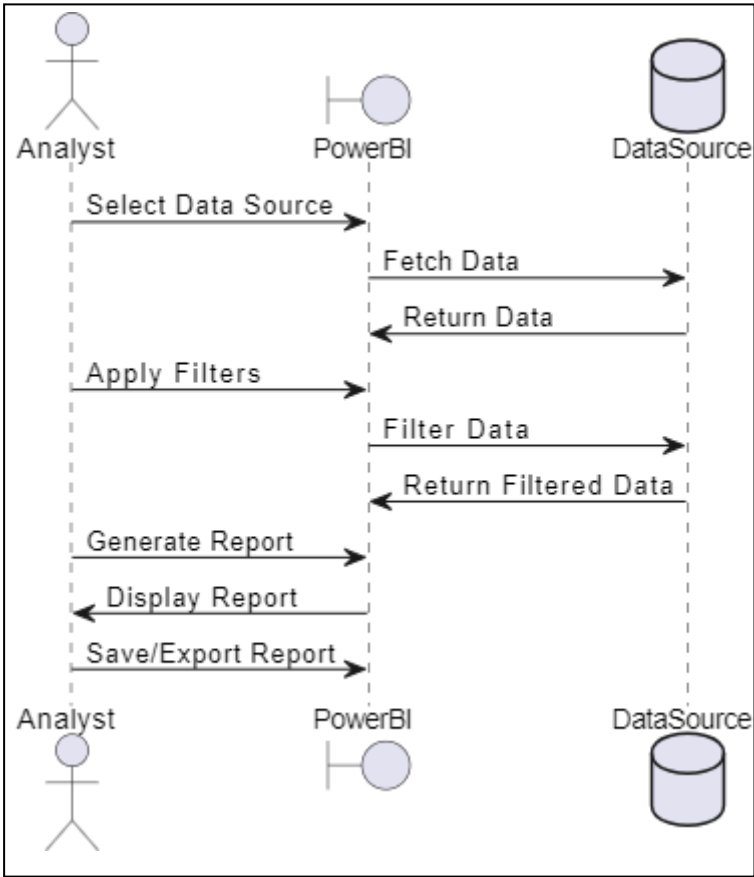


Figure 10:
Sequence Diagram

5.5 Activity Diagram



Figure 11:
Activity Diagram

5.6 Component Diagram

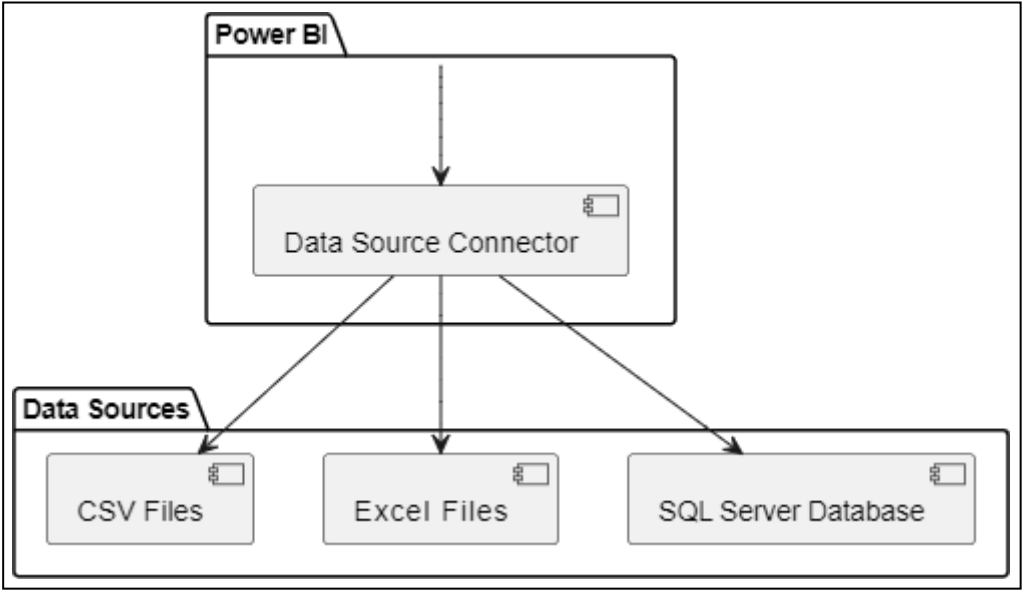


Figure 12:
Component Diagram

5.7 Dataflow Diagram

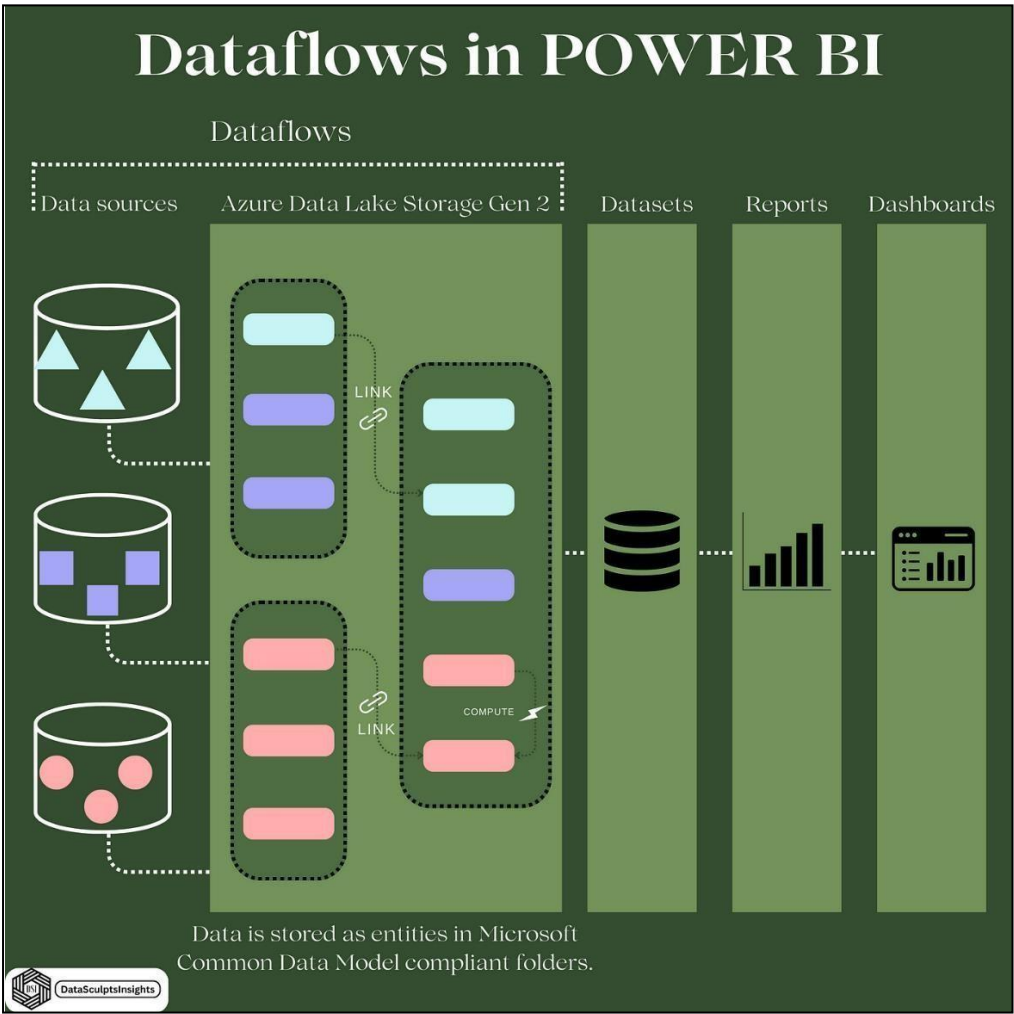


Figure 13:
Data Flow Diagram

VI. IMPLEMENTATION AND EVALUATION

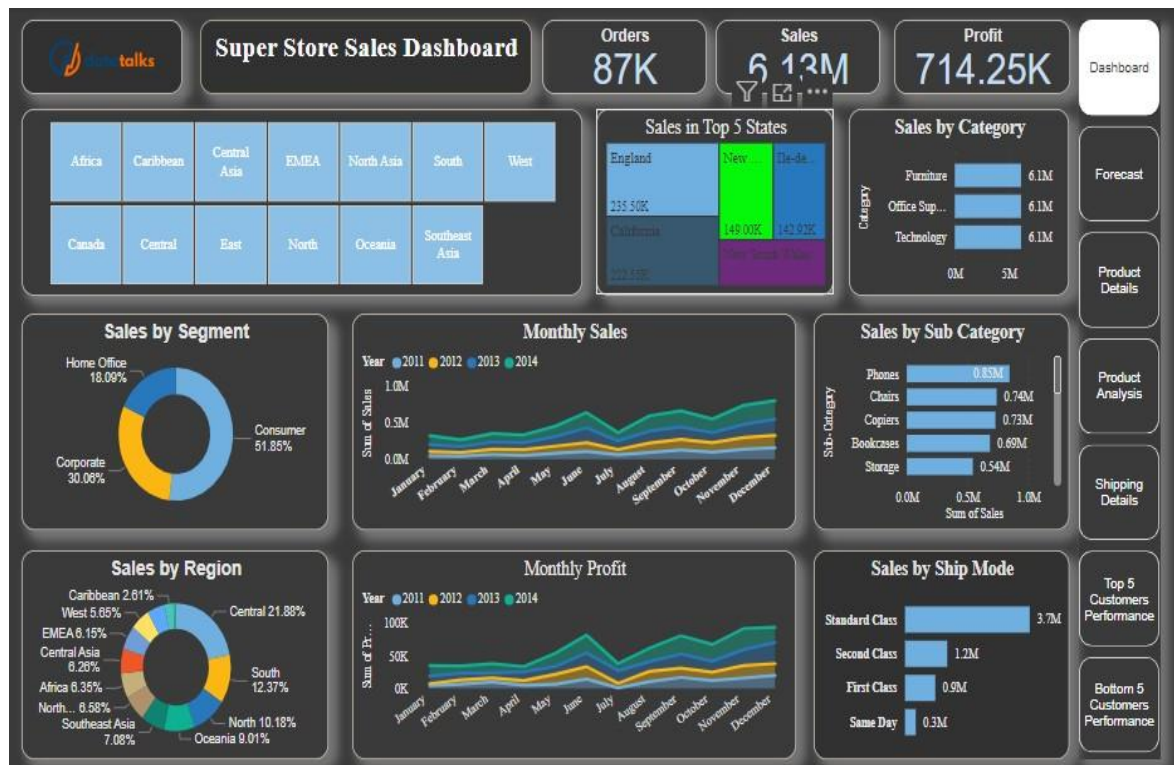


Figure 14:
Main Dashboard

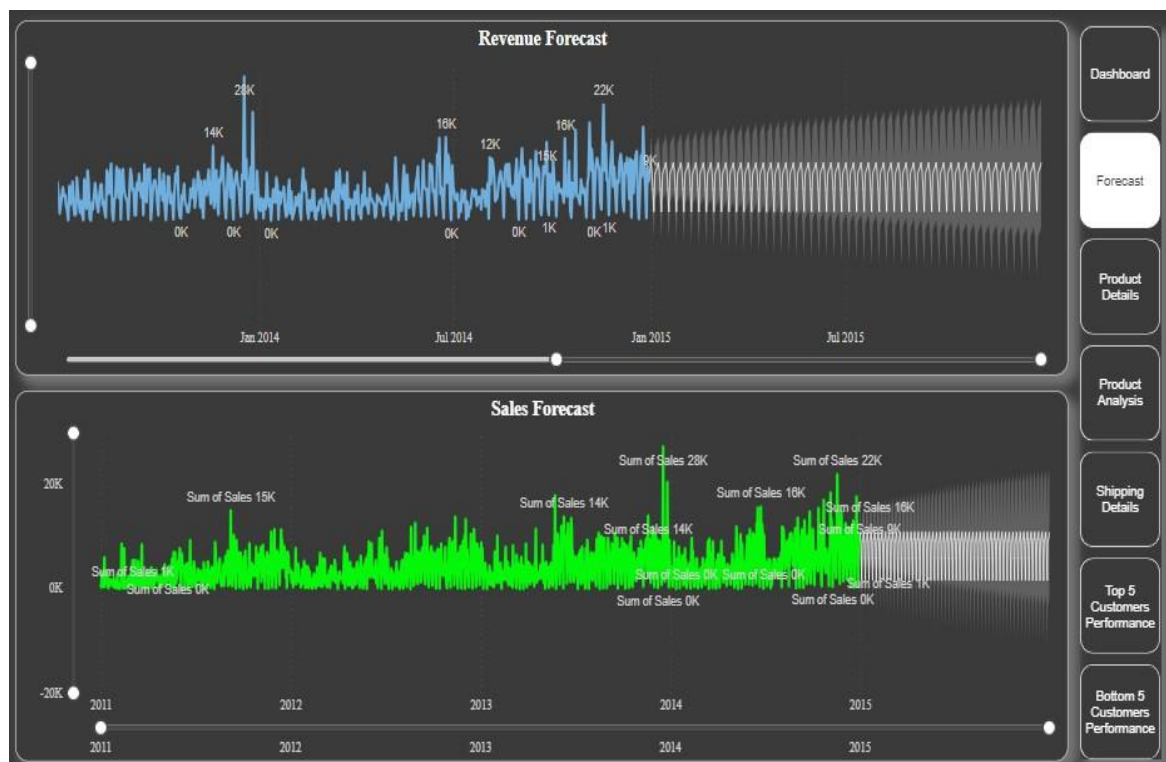


Figure 15:
Forecast dashboard

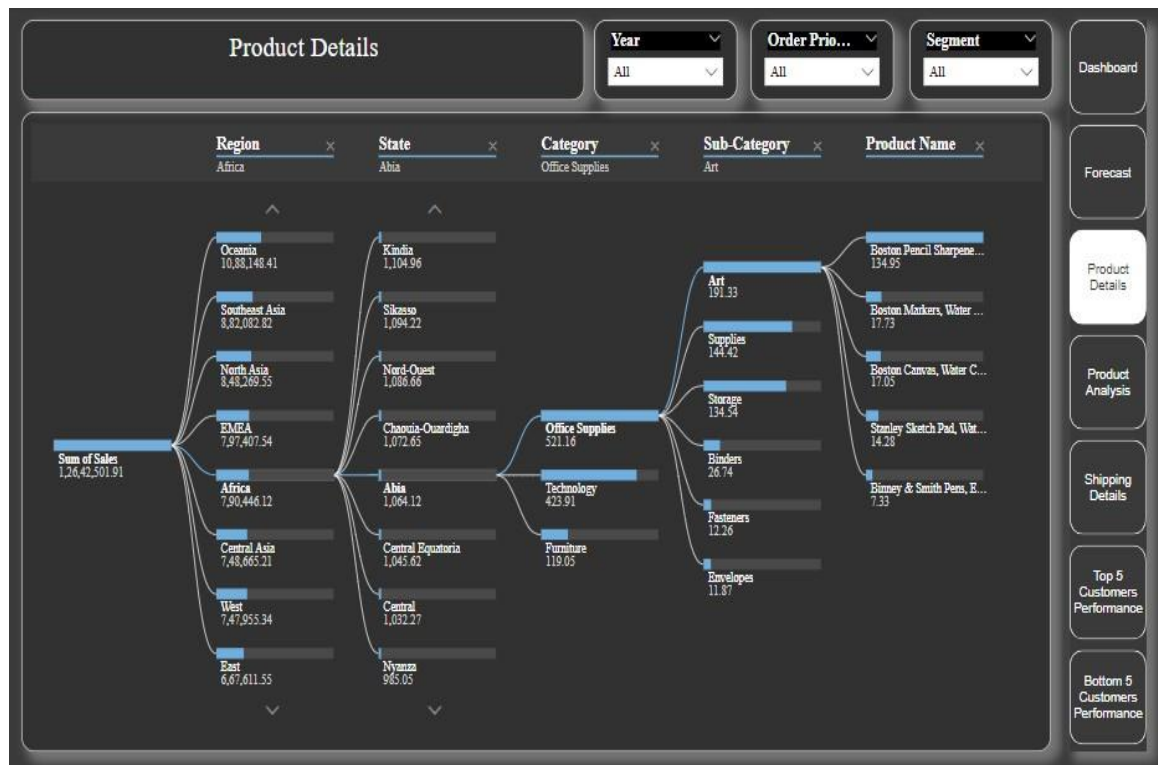


Figure 16:
Product Details Dashboard



Figure 17:
Product Analysis Dashboard

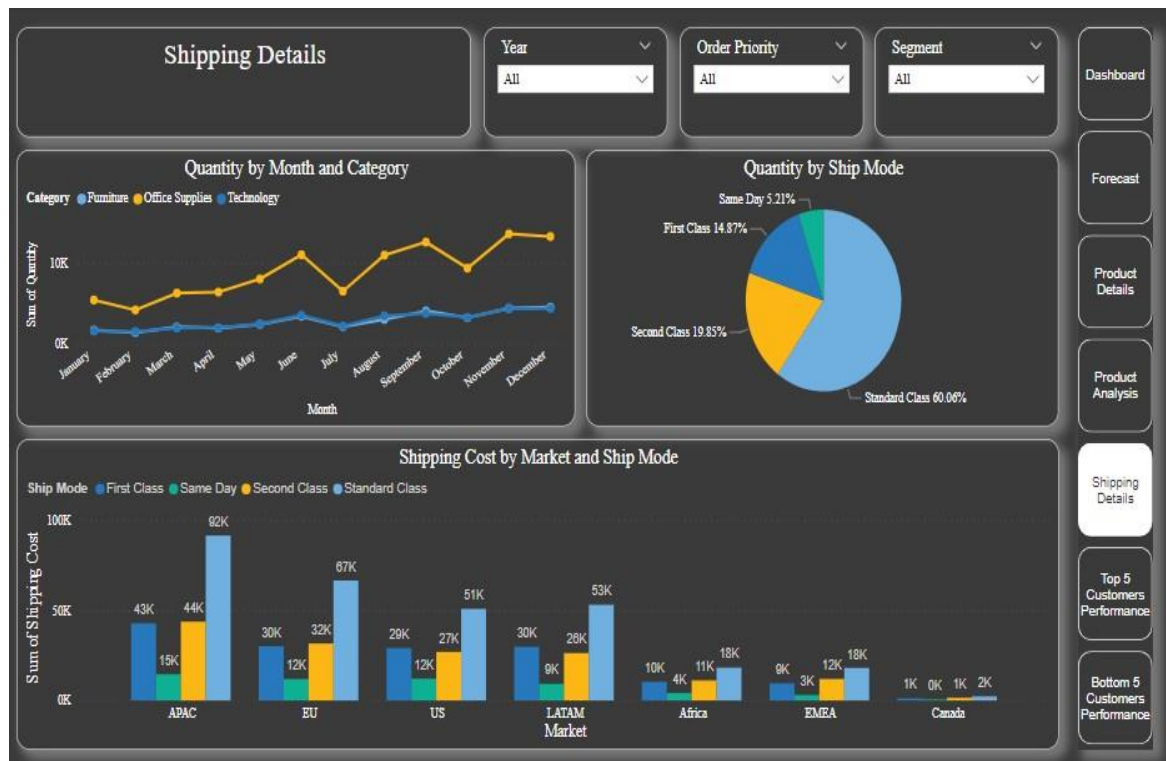


Figure 18:
Shipping Details Dashboard

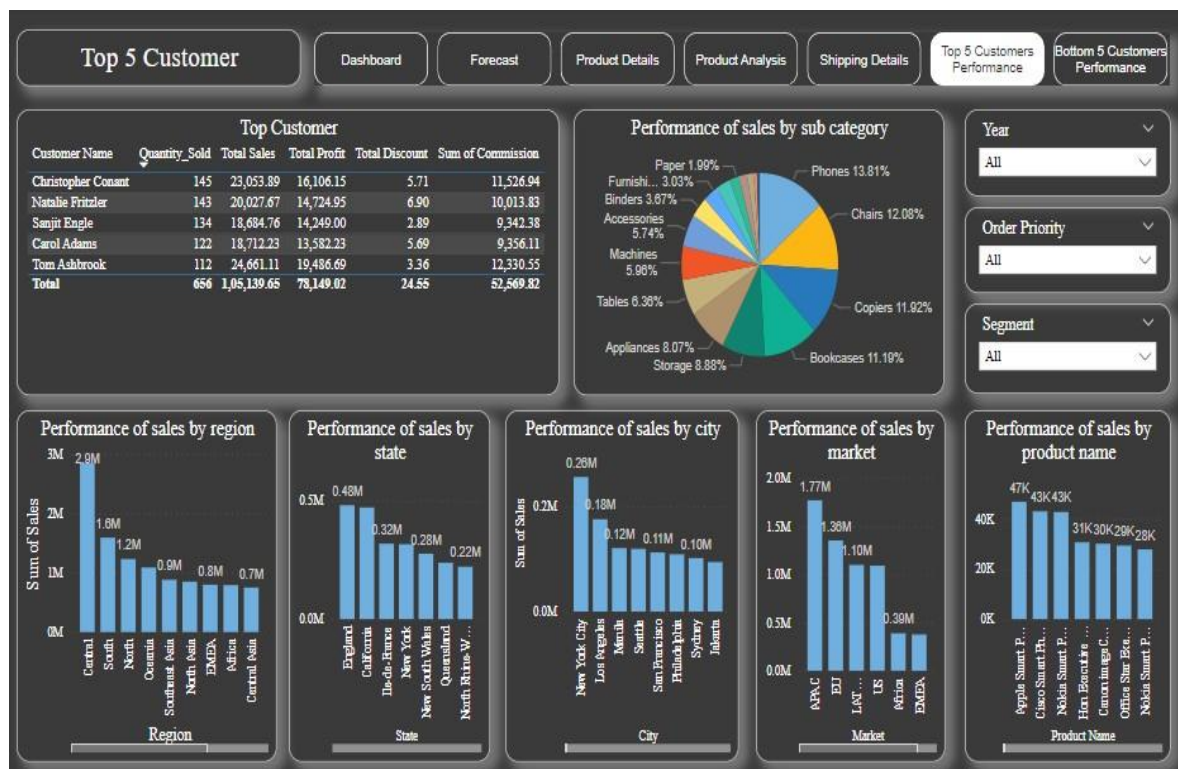


Figure 19:
Top 5 Customers Dashboard

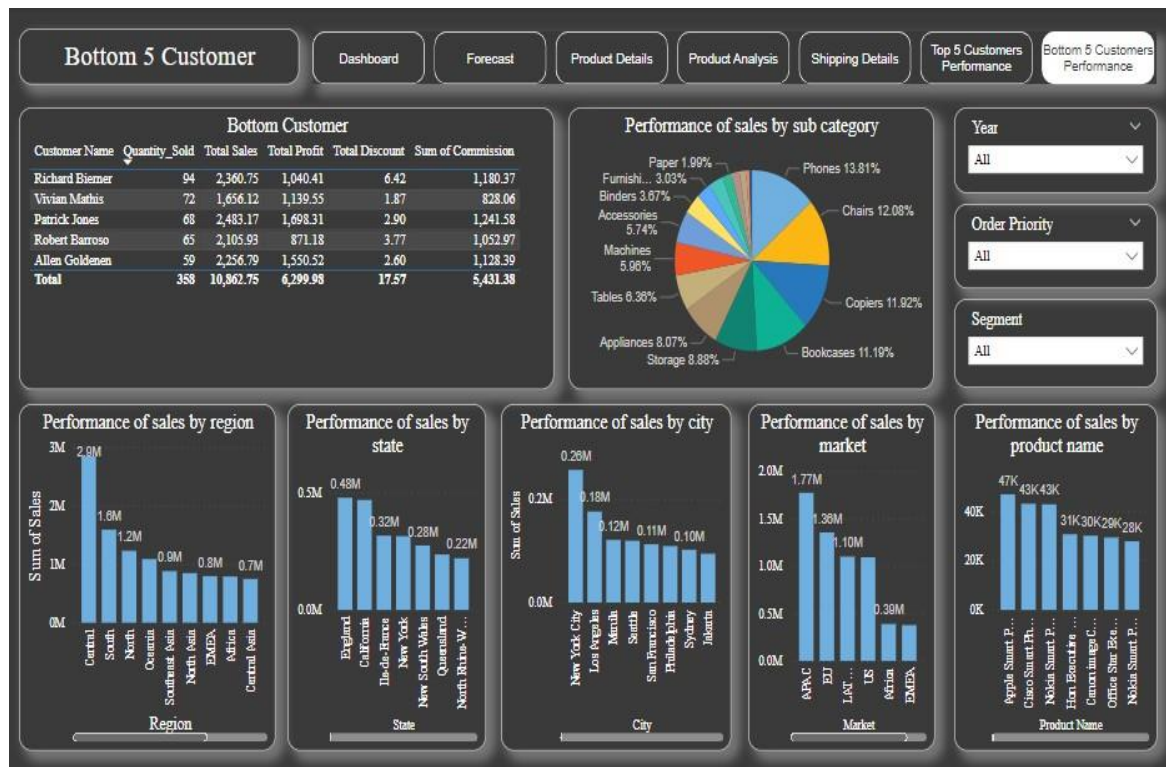


Figure 20:
Bottom 5 Customers Dashboard

VII. CONCLUSION

This project gave me an opportunity to learn about the new platform, Power BI. I have done data visualization project previously. In this report, I have done accessibility and unit testing of the DAX query language as I have only reported the data for the year 2014. The project data and the Power BI tool through which we can visualize the data for the years 2011, 2012, and 2013. The Power BI platform not only gave me an opportunity to analyze the data and present it in meaningful way, but also gave me an opportunity to think about the key indicators that are essential for the success of the sales in various countries. I am planning on pursuing my career in Power BI.

VIII. REFERENCES

1. Microsoft Power BI documentation. 2021. Create a report from an Excel file in the Power BI service. <https://docs.microsoft.com/en-us/power-bi/create-reports/service-report-create-new> (Accessed on September 20, 2021).
2. Microsoft Power BI documentation. 2021. Manage data storage in Power BI workspaces. <https://docs.microsoft.com/en-us/power-bi/admin/service-admin-manage-your-data-storage-in-power-bi> (Accessed on September 20, 2021).
3. Microsoft Power BI documentation. 2021. Apply DAX basics in Power BI desktop. <https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-quickstart-learn-dax-basics> (Accessed on September 15, 2021).
4. Microsoft Power BI documentation. 2021. Design Power BI reports for accessibility. <https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-accessibility-creating-reports> (Accessed on November 07, 2021).
5. <https://hevo.com/learn/10-best-power-bi-reports-examples/>
6. <https://powerbi.microsoft.com/en-us/>
7. <https://rigorousthemes.com/blog/best-power-bi-dashboard-examples/>
8. <https://inblog.in/Power-BI-Projects-HdlkEZeaX/>
9. Gonzalo, A., Harreis, H., Altable C.S., and Villepelet, C. 2020. Fashion's digital transformation: now or never. <https://www.mckinsey.com/industries/retail/our-insights/fashions-digital-transformation-now-or-never> (Accessed on September 05, 2021).
10. Hamilton, T. 2021. Business intelligence (BI) testing: sample test cases. <https://www.guru99.com/business-intelligence-testing-sample-test-cases.html> (Accessed on November 07, 2021).
11. Microsoft Power BI documentation. 2021. Create a report from an Excel file in the Power BI service. <https://docs.microsoft.com/en-us/power-bi/create-reports/service-report-create-new> (Accessed on September 20, 2021).
12. Microsoft Power BI documentation. 2021. Manage data storage in Power BI workspaces. <https://docs.microsoft.com/en-us/power-bi/admin/service-admin-manage-your-data-storage-in-power-bi> (Accessed on September 20, 2021).
13. Microsoft Power BI documentation. 2021. Apply DAX basics in Power BI desktop. <https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-quickstart-learn-dax-basics> (Accessed on September 15, 2021).
14. Microsoft Power BI documentation. 2021. Design Power BI reports for accessibility. <https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-accessibility-creating-reports> (Accessed on November 07, 2021).