S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Postlab

Aim: To implement Power BI as a Big Data analysis tool to visualize, analyze, and interpret large-scale datasets effectively, and to explore its integration with Big Data sources for generating interactive dashboards and insightful business intelligence reports.

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: 7th / 4th

Academic Session: 2025-2026

Date of Performance: _____

Date of Submission: _____

AIM: To implement Power BI as a Big Data analysis tool to visualize, analyze, and interpret large-scale datasets effectively, and to explore its integration with Big Data sources for generating interactive dashboards and insightful business intelligence reports.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- To understand the role of **Power BI** in Big Data analytics and visualization.
- To connect **Power BI** with various Big Data sources such as Hadoop, Azure, or SQL-based systems.
- To create and design **interactive dashboards** for effective data interpretation and decision-making.
- To perform **data transformation, cleaning, and modeling** using Power BI tools.
- To analyze and visualize **large-scale datasets** to derive meaningful insights and trends.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement:

- 64-bit operating system (Linux, CentOS, or Ubuntu)
- At least 8 GB of RAM for each node and sufficient storage

Software Requirement:

- Power BI

THEORY:

Power BI is an advanced Business Intelligence (BI) and data visualization tool developed by Microsoft. It helps users connect to multiple data sources, transform large datasets, and create interactive dashboards for insightful analysis. In Big Data analytics, Power BI acts as a visualization and reporting layer that makes large-scale data understandable and actionable.

It can seamlessly integrate with Big Data platforms such as Hadoop, Spark, Google BigQuery, Azure Data Lake, Amazon Redshift, and SQL Server. Power BI uses different connection modes like Import, DirectQuery, and Live Connection to handle vast amounts of data efficiently. With the help of Power Query Editor, users can perform data cleaning, transformation, and integration before visualization.

Power BI supports Data Analysis Expressions (DAX) for creating custom measures and performing advanced data calculations. It also allows real-time data monitoring and AI-driven analytics, enhancing decision-making capabilities. Through Power BI Service (cloud) and Power BI Desktop, reports can be created, published, and shared securely across teams or organizations.

Features of Power BI:

1. **Data Connectivity:** Connects to multiple data sources—SQL, Excel, Azure, Hadoop, APIs, etc.
2. **Interactive Dashboards:** Provides dynamic and visually rich dashboards for better insights.

3. **Real-Time Analytics:** Supports live dashboards for monitoring business performance instantly.
4. **Data Transformation Tools:** Power Query allows easy data cleaning, merging, and transformation.
5. **DAX Language:** Enables advanced calculations and expressions for complex analytics.
6. **AI and Machine Learning Integration:** Built-in AI visuals for predictive and sentiment analysis.
7. **Cloud and Mobile Accessibility:** Reports and dashboards can be accessed anywhere via the cloud.
8. **Security:** Offers role-based access control and data encryption.

Advantages of Power BI:

1. **User-Friendly Interface:** Easy drag-and-drop features make visualization simple.
2. **Integration with Microsoft Ecosystem:** Works seamlessly with Excel, Azure, and Office 365.
3. **Scalability:** Handles large datasets from multiple Big Data platforms.
4. **Cost-Effective:** Offers a free desktop version and affordable Pro/Enterprise plans.
5. **Real-Time Data Updates:** Dashboards update automatically when data changes.
6. **Custom Visuals:** Users can import or create visuals tailored to their analysis needs.

Disadvantages of Power BI:

1. **Data Volume Limitations:** Import mode may struggle with extremely large datasets.
2. **Complex DAX Formulas:** Advanced calculations require learning DAX syntax.
3. **Performance Issues:** Dashboards may lag when connected to very large or complex data models.
4. **Limited Customization in Free Version:** Some advanced features require Pro or Premium licenses.
5. **Dependency on Internet:** Cloud-based reports need a stable internet connection.

INPUT / OUTPUT (SCREENSHOTS):

DATASET USED – Sample SuperStore Dataset

Link -
https://www.kaggle.com/datasets/bravehart101/sample-supermarket-dataset

Data Analysis

1. **Sales Performance Analysis:**

- **Total Sales:** Analyze overall sales performance over time and identify trends.
- **Sales by Category/Sub-category:** Evaluate sales performance across different product categories and sub-categories to identify top-performing segments.
- **Sales by Region:** Explore sales distribution across different regions to identify geographical areas with high or low sales.

- **Sales by Customer Segment:** Analyze sales by customer segment to understand which segments contribute the most to overall revenue.

2. **Product Analysis:**

- **Product Performance:** Identify top-selling products and analyze their contribution to overall sales.
- **Product Profitability:** Evaluate the profitability of each product by analyzing profit margins.
- **Product Sales Trend:** Examine sales trends for individual products over time to identify seasonal variations or product lifecycle patterns.

3. **Customer Analysis:**

- **Customer Segmentation:** Segment customers based on purchasing behavior, such as frequency of purchases, total spending, or product preferences.
- **Customer Lifetime Value (CLV):** Calculate CLV to identify high-value customers and tailor marketing strategies accordingly.
- **Customer Geography Analysis:** Analyze customer distribution by region to understand geographic customer preferences and target specific regions for marketing efforts.

4. **Order Analysis:**

- **Order Frequency and Size:** Analyze the frequency of orders and average order size to identify purchasing patterns.
- **Shipping Mode Analysis:** Evaluate the popularity and profitability of different shipping modes.
- **Discount Analysis:** Assess the impact of discounts on sales revenue and profitability.

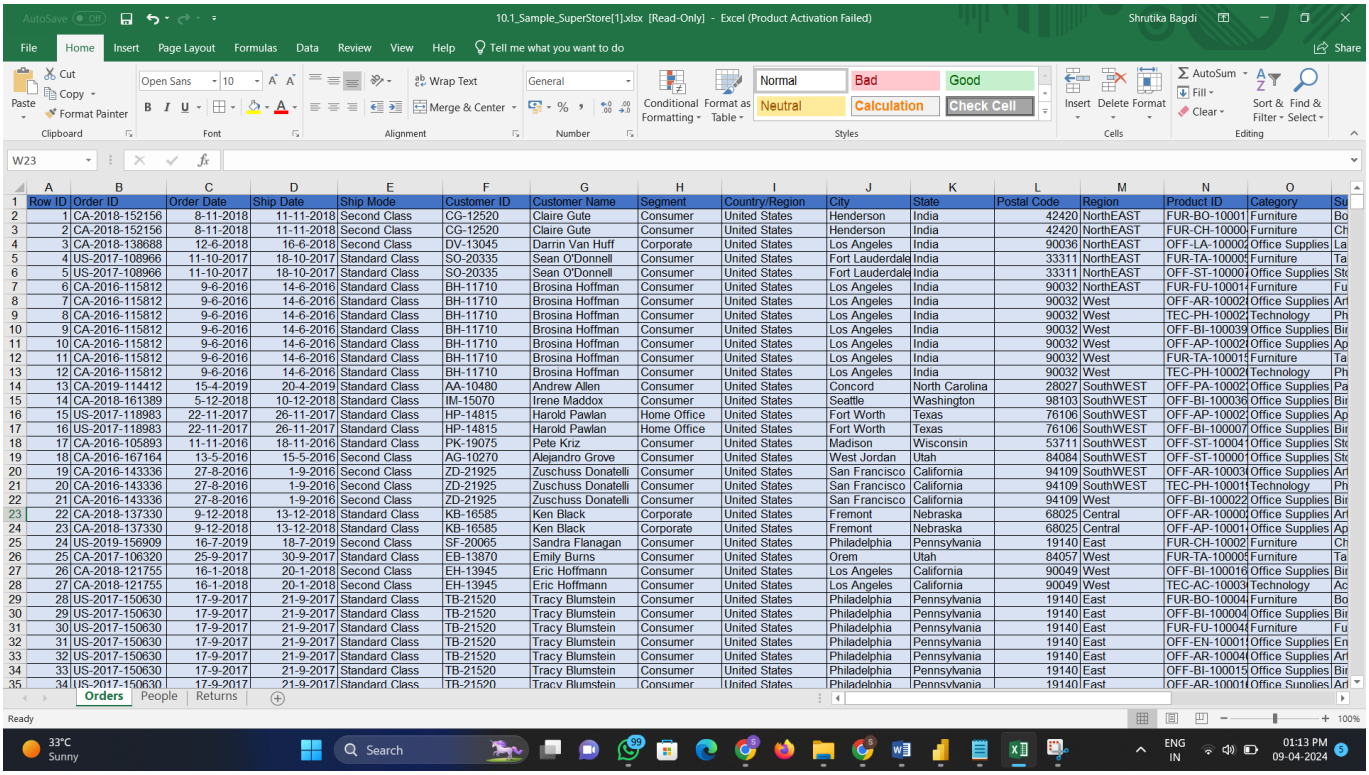
5. **Profitability Analysis:**

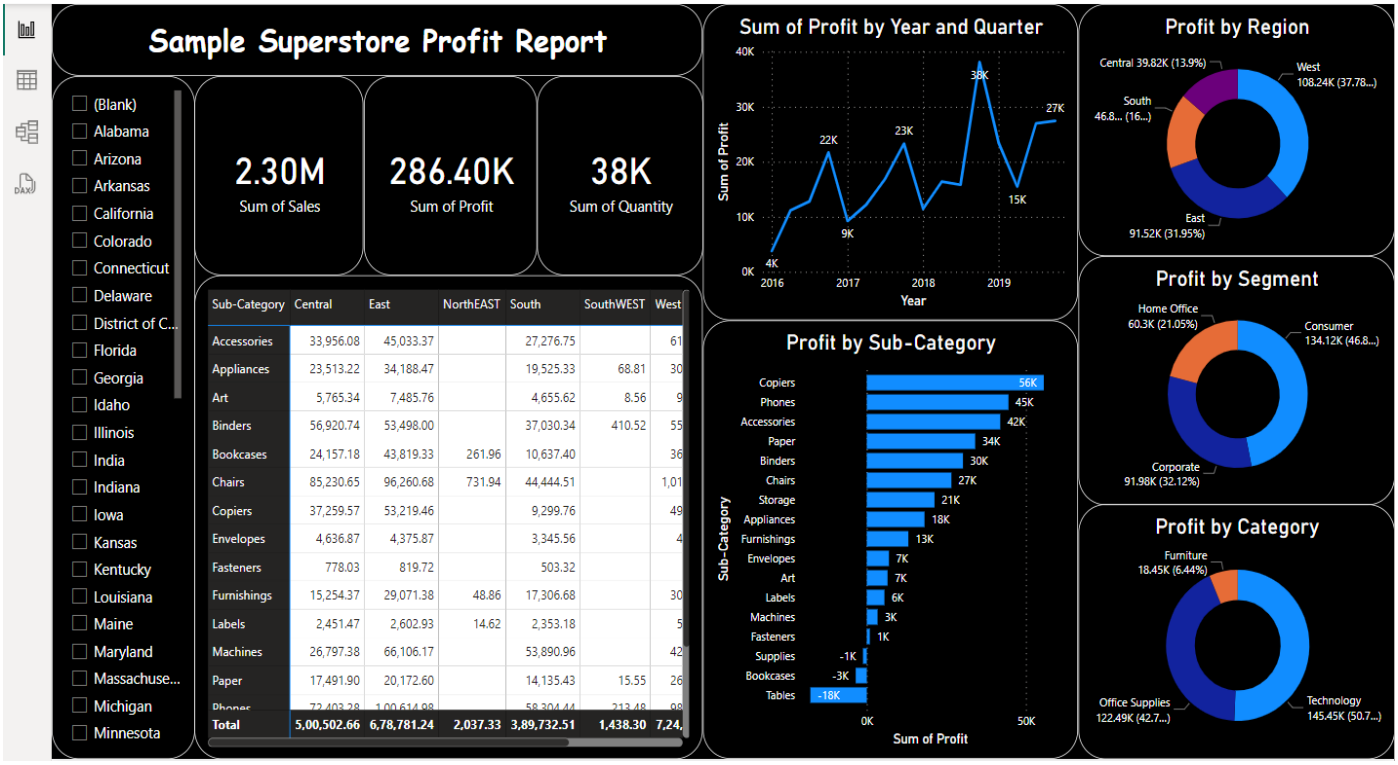
- **Overall Profitability:** Calculate overall profitability by analyzing profit margins and identifying areas for cost optimization.
- **Profitability by Region/Segment:** Evaluate profitability by region and customer segment to identify areas with potential for improvement.
- **Profitability Trends:** Analyze profitability trends over time to identify factors influencing profitability and implement strategies to improve margins.

6. Forecasting and Predictive Analysis:

- **Time Series Forecasting:** Use historical sales data to forecast future sales trends and identify potential opportunities or challenges.
- **Predictive Modeling:** Utilize machine learning algorithms to predict future sales based on various factors such as seasonality, promotions, and market trends.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region
43	CA-2018-101343	17 July 2018	43303	Standard Class	RA-19885	Ruben Ausman	Corporate	United States	Los Angeles	California	90049	West
514	CA-2019-163405	21 December 2019	43824	Standard Class	BN-11515	Bradley Nguyen	Consumer	United States	Los Angeles	California	90049	West
515	CA-2019-163405	21 December 2019	43824	Standard Class	BN-11515	Bradley Nguyen	Consumer	United States	Los Angeles	California	90049	West
1626	CA-2017-157084	19 December 2017	43093	Standard Class	JG-15160	James Galang	Consumer	United States	Los Angeles	California	90049	West
2060	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2061	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2063	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2065	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2067	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2068	CA-2016-106439	31 October 2016	42678	Standard Class	GG-14650	Greg Guthrie	Corporate	United States	Los Angeles	California	90049	West
2191	CA-2018-118913	25 June 2018	43280	Standard Class	AS-10240	Alan Shonely	Consumer	United States	Los Angeles	California	90049	West
2709	CA-2017-121797	30 January 2017	42772	Standard Class	CC-12145	Charles Crestani	Consumer	United States	Los Angeles	California	90049	West
2944	CA-2019-126242	19 November 2019	43793	Standard Class	MC-18100	Mick Crebagga	Consumer	United States	Los Angeles	California	90049	West
2945	CA-2019-126242	19 November 2019	43793	Standard Class	MC-18100	Mick Crebagga	Consumer	United States	Los Angeles	California	90049	West
3534	CA-2016-110849	18 April 2016	42483	Standard Class	JL-15835	John Lee	Consumer	United States	Los Angeles	California	90049	West
3535	CA-2016-110849	18 April 2016	42483	Standard Class	JL-15835	John Lee	Consumer	United States	Los Angeles	California	90049	West
3536	CA-2016-110849	18 April 2016	42483	Standard Class	JL-15835	John Lee	Consumer	United States	Los Angeles	California	90049	West
3759	CA-2017-167745	18 September 2017	43001	Standard Class	GB-14530	George Bell	Corporate	United States	Los Angeles	California	90049	West
3760	CA-2017-167745	18 September 2017	43001	Standard Class	GB-14530	George Bell	Corporate	United States	Los Angeles	California	90049	West
3761	CA-2017-167745	18 September 2017	43001	Standard Class	GB-14530	George Bell	Corporate	United States	Los Angeles	California	90049	West
4338	CA-2017-125234	27 November 2017	43070	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Los Angeles	California	90049	West
4339	CA-2017-125234	27 November 2017	43070	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Los Angeles	California	90049	West
4481	CA-2018-109365	03 November 2018	43412	Standard Class	XP-21865	Xylona Preis	Consumer	United States	Los Angeles	California	90049	West
4482	CA-2018-109365	03 November 2018	43412	Standard Class	XP-21865	Xylona Preis	Consumer	United States	Los Angeles	California	90049	West
4484	CA-2018-109365	03 November 2018	43412	Standard Class	XP-21865	Xylona Preis	Consumer	United States	Los Angeles	California	90049	West
4486	CA-2018-109365	03 November 2018	43412	Standard Class	XP-21865	Xylona Preis	Consumer	United States	Los Angeles	California	90049	West
4487	CA-2018-109365	03 November 2018	43412	Standard Class	XP-21865	Xylona Preis	Consumer	United States	Los Angeles	California	90049	West





CONCLUSION:

Power BI offers an efficient and interactive platform for Big Data visualization and analysis. It transforms large datasets into meaningful insights, enabling quick data-driven decisions. Despite minor limitations with very large data, its strong integration and visualization features make it a leading Big Data analytics tool.

REFERENCE:

- *https://www.datacamp.com/blog/all-about-power-bi*
- *https://en.wikipedia.org/wiki/Microsoft_Power_BI*
- *https://www.godaddy.com/en-in/help/what-is-microsoft-power-bi-42272*

Observation book: (3)	Viva-Voce (3)	Quality of Submission and timely Evaluation (4)
Total:		**Sign with date:**