

Chocolate Sales Analysis and Business Insights Using Power BI and Python

P. Logaiyan ¹, D. Monica²

¹Associate Professor, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Puducherry - 605107, India

²PG Student, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Puducherry - 605107, India

Abstract

The project titled “Chocolate Sales Analysis and Business Insights Using Power BI and Python” is developed to study chocolate sales data and convert raw sales records into useful business information. Sales data collected from Excel/CSV files contains details such as product names, sales amount, profit, customer information, regional sales performance, and monthly sales records. This data is stored in MySQL, processed using Python, and visually represented through Power BI dashboards, while HTML and CSS are used for the frontend. The analysis helps in understanding sales trends, identifying best-selling products, comparing region-wise performance, and observing customer purchasing behavior. By presenting the information in a clear visual format, the project makes sales reporting easier to understand and reduces the effort involved in manual analysis. The insights generated from the dashboard help in identifying important patterns, improving reporting accuracy, and supporting better business decisions for future growth.

Keywords -- Chocolate sales, sales analysis, business insights, Power BI, Python, MySQL, data visualization, sales performance, profit analysis, dashboard reporting.

I. INTRODUCTION

In today’s business environment, sales data plays an important role in understanding how a business is performing. Every sales transaction carries useful information about customer preferences, product demand, profit levels, and market performance. When sales information remains only as raw records, it becomes difficult to identify patterns or understand the overall condition of the business. Proper analysis helps in converting this data into meaningful information that supports better planning and decision-making.

Chocolate products are sold across different regions, in different quantities, and during different periods of time. Because of this, chocolate sales data offers a useful way to study sales trends and business performance. By analyzing this data, it becomes easier to understand which products are selling more, which regions are generating better profit, and how sales change from month to month.

Sales records collected from Excel/CSV files can be processed and organized in a structured way using Python and MySQL, while Power BI helps in presenting the results through clear and interactive dashboards. This makes the sales information easier to understand and helps in identifying important business patterns. Through proper analysis and visualization, businesses can track performance more effectively, recognize weak areas, and make better decisions for future growth.

II. LITERATURE SURVEY

1. “Sales Performance Analysis for Retail Business Using Machine Learning and Power BI” (2025):

This study explains how sales data can be analyzed to understand business performance more clearly. It focuses on important factors such as sales trends, customer behavior, and seasonal changes. The findings are presented through Power BI dashboards, which make the information easier to understand. The study shows that data analysis improves reporting and supports better business planning. However, the work mainly focuses on general retail sales and does not specifically examine chocolate sales.

2. **“Applying Power BI for Improved Retail Business Analytics and Decision-Making” (2025):** This paper discusses how Power BI helps convert raw sales data into meaningful business insights. It explains that interactive dashboards make it easier to understand sales patterns, profit levels, customer trends, and overall business performance. The study highlights that visual reporting supports faster and more effective decision-making. However, the analysis focuses on retail business in a broad way and does not include detailed data processing using Python or database management using MySQL.

III. PROBLEM STATEMENT

Sales data contains important information about product demand, customer preferences, profit, and regional performance. When this data is stored only as raw records in Excel sheets or CSV files, understanding the overall sales performance becomes difficult. Manual checking of large amounts of sales data takes time and often makes it hard to identify useful patterns.

In chocolate sales, factors such as monthly sales variation, product demand, profit changes, and region-wise performance need proper analysis to understand business growth clearly. Without a structured method of analysis, it becomes difficult to identify best-selling products, low-performing regions, and changing sales trends.

In many cases, business decisions are made only by looking at basic sales figures without understanding the actual reasons behind performance changes. This may lead to missed opportunities and delay the identification of weak areas. Therefore, there is a need for a better system that can organize raw sales data, analyze it effectively, and present the results in a clear visual format for easier understanding and better decision-making.

IV. OBJECTIVES

The main objective of this project is to analyze chocolate sales data and convert raw sales records into meaningful business insights. Proper analysis helps in understanding how sales are performing across different products, regions, and time periods.

Another objective is to organize sales data collected from Excel/CSV files in a structured manner. Storing the data properly helps in reducing confusion and makes further analysis easier and more accurate.

The project also aims to study monthly sales trends in order to understand how sales performance changes over different periods. This helps in identifying growth patterns, seasonal changes, and periods of low sales.

One more objective is to identify best-selling products, compare region-wise sales performance, and examine profit variation. These factors help in understanding product demand and the contribution of different regions to overall sales performance.

The final objective is to present the analyzed data in a clear visual format using Power BI dashboards. Visual representation makes the results easier to understand, reduces manual reporting effort, and supports better business decision-making.

V. EXISTING SYSTEM

In many business environments, sales information is usually maintained in Excel sheets, CSV files, or basic manual records. These records contain useful details such as product names, sales amount, profit, customer information, and region-wise sales. Even though the data is available, it often remains in raw form without proper analysis.

When the amount of sales data increases, manual checking becomes difficult and time-consuming. Understanding sales trends, product performance, or monthly changes by looking only at tables and numbers is not easy. Important patterns may remain unnoticed, and generating reports also takes more effort.

The existing method mainly focuses on storing data rather than extracting meaningful insights from it. As a result, identifying best-selling products, low-performing regions, and profit variations becomes more difficult. This

shows the limitation of traditional sales reporting methods and creates the need for a more effective analytical approach.

VI. PROPOSED SYSTEM

The proposed system is designed to analyze chocolate sales data in a more organized and meaningful way. Sales records collected from Excel or CSV files are first arranged in a structured format so that the data becomes easier to manage and process.

The collected data is stored in MySQL, where sales-related details such as product information, sales amount, profit, customer details, regional performance, and monthly sales records are maintained in an organized manner. This structured storage helps in improving data handling and reduces the difficulties faced in manual analysis.

Python is used to process and analyze the stored data. During this stage, the data is cleaned, arranged, and examined to identify useful patterns such as monthly sales trends, product demand, region-wise sales performance, and profit variation.

After the analysis is completed, Power BI is used to create interactive dashboards and reports. These dashboards present the sales information in graphical form, making it easier to understand the performance of products, regions, and overall sales activity.

The proposed system makes sales reporting faster, more accurate, and easier to understand. By converting raw sales data into meaningful business insights, it helps in identifying strong areas, weak areas, and important sales patterns that support better decision-making.

VII. DATASET DESCRIPTION

The dataset used for this analysis is collected from Kaggle and is available in Excel/CSV format containing chocolate sales records. It includes important sales-related information that helps in understanding overall business performance across different products, regions, and time periods.

The dataset contains details such as product names, sales amount, profit, quantity sold, customer information, sales date, and region-wise sales records. Each record represents a sales transaction and provides useful information about how different chocolate products are performing in the market.

The sales date field helps in studying monthly sales trends and identifying changes in demand over different periods. Product details make it possible to identify best-selling chocolates and products with lower sales performance.

Region-related information helps in comparing sales performance across different locations. Profit values in the dataset also make it easier to understand which products or regions contribute more to overall business growth.

Since the dataset contains both numerical and categorical information, it becomes suitable for analysis using Python and for creating clear visual dashboards in Power BI. The dataset forms the foundation for identifying meaningful sales patterns and business insights.

VIII. METHODOLOGY

The methodology begins with collecting the chocolate sales dataset from Kaggle in Excel/CSV format. The dataset contains sales-related information such as product names, sales amount, profit, quantity sold, customer details, sales date, and regional sales records. Before starting the analysis, the data is checked carefully to remove missing values, duplicate entries, and unwanted errors so that the dataset becomes suitable for processing.

After the cleaning process, the data is stored in MySQL in a structured format. Organizing the sales records in a database makes data handling easier and helps maintain accuracy when working with large amounts of information.

Python is used to process and analyze the stored data. During this stage, different sales factors such as monthly sales trends, best-selling products, profit variation, and region-wise sales performance are examined to understand the overall business performance.

Once the analysis is completed, Power BI is used to create dashboards and visual reports. Charts, graphs, and other visual elements present the analyzed information in a simple and clear format, making the results easier to understand.

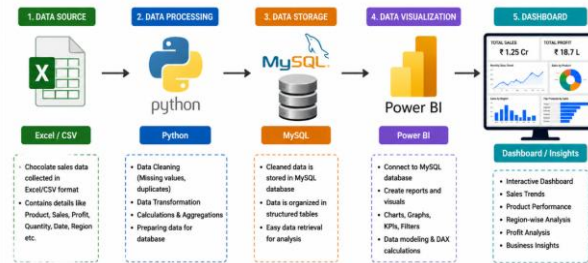


Figure 1: Workflow

Through this step-by-step process, raw chocolate sales data is transformed into meaningful business insights that help in identifying patterns, understanding performance, and supporting better decision-making.

IX. ANALYSIS AND INTERPRETATION

After processing the chocolate sales data, different calculations are used to understand the overall business performance more clearly. The first step is calculating the total sales and total profit from the dataset. These values give a general idea of the total revenue generated and the profit earned during the selected period.

Total Sales = Sum of Sales Amount

Total Profit = Sum of Profit

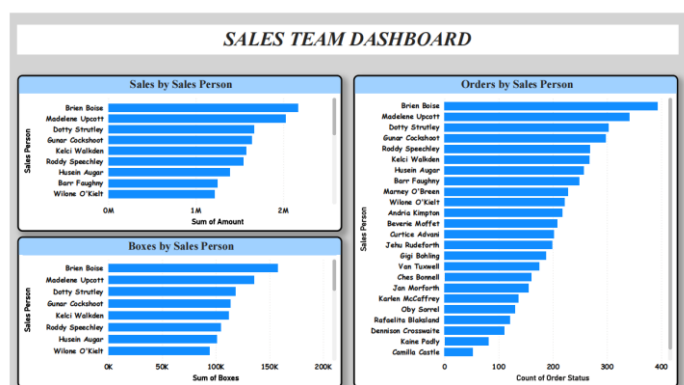


Figure 2: Sales Team Dashboard

Average sales can be calculated to understand the average value of each sales transaction. This helps in knowing whether the sales value per transaction is high or low.

Average Sales = Total Sales / Number of Sales Records

Average profit can also be calculated to understand the average profit earned from each transaction.

Average Profit = Total Profit / Number of Sales Records

Profit percentage is another important measure in sales analysis. It shows how much profit is earned from the total sales amount.

$$\text{Profit Percentage} = (\text{Total Profit} / \text{Total Sales}) \times 100$$

Monthly sales comparison helps in understanding whether sales increased or decreased over time. This makes it easier to identify seasonal changes and demand variation.

$$\text{Monthly Growth Percentage} = ((\text{Current Month Sales} - \text{Previous Month Sales}) / \text{Previous Month Sales}) \times 100$$



Figure 3: Chocolate Sales Dashboard

Product-wise analysis helps in identifying best-selling chocolates. The total sales and total profit of each product are calculated separately. This shows which products contribute more to overall business growth.

Product Sales = Sum of Sales Amount for Each Product

Product Profit = Sum of Profit for Each Product

Quantity sold is also calculated to understand customer demand more clearly.

Total Quantity Sold = Sum of Quantity

Region-wise calculations are useful for comparing sales performance across different locations. This helps in identifying strong-performing regions and regions with lower sales.

Region-wise Sales = Sum of Sales Amount for Each Region

Region-wise Profit = Sum of Profit for Each Region

Another useful calculation is contribution percentage. It shows how much a particular product or region contributes to the overall sales.

$$\text{Contribution Percentage} = (\text{Particular Product Sales} / \text{Total Sales}) \times 100$$

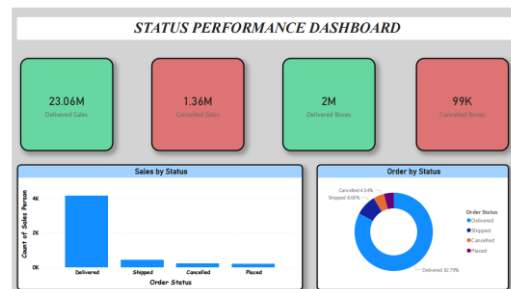


Figure 4: Status Performance Dashboard

Sales variance can also be used to compare actual sales with average sales. This helps in understanding whether sales are performing above or below the overall average.

$$\text{Sales Variance} = \text{Actual Sales} - \text{Average Sales}$$

By combining these calculations with Power BI dashboards, the sales data becomes easier to interpret. The analysis clearly shows sales trends, product performance, profit variation, regional contribution, and demand patterns, making business understanding more meaningful.

X. IMPLEMENTATION

The implementation begins with collecting the chocolate sales dataset from Kaggle in Excel/CSV format. The dataset contains sales-related information such as product names, sales amount, profit, quantity sold, customer details, sales date, and regional sales records. The raw data is first checked carefully to remove missing values, duplicate entries, and unwanted errors before starting the analysis.

Python is used as the backend tool for data cleaning, processing, and analysis. During this stage, the dataset is organized properly and basic calculations such as total sales, total profit, average sales, product-wise sales, and monthly sales comparison are performed. This helps in preparing the data for further interpretation.

After processing, the cleaned data is stored in MySQL in a structured format. Using a database helps in maintaining the sales records properly and makes data retrieval easier when handling large amounts of information.

Power BI is used as the visualization tool to create dashboards and reports. Different charts, graphs, and summary visuals are created to represent monthly sales trends, best-selling products, profit variation, and region-wise performance. These dashboards make the sales information easier to understand and interpret.

HTML and CSS are used for the frontend part of the project. They help in creating a simple interface for presenting project-related information in an organized manner. By combining all these tools, the raw chocolate sales data is transformed into meaningful business insights in a structured and efficient way.

IX. FUTURE WORK

The current analysis focuses mainly on understanding sales trends, product performance, profit variation, and regional sales using the available chocolate sales dataset. In the future, the system can be improved by using larger and more recent sales data so that the analysis becomes more detailed and accurate.

More advanced techniques can also be added using Python to predict future sales trends based on past performance. This can help in estimating product demand and planning sales strategies more effectively.

The dashboard can also be expanded by including customer segmentation, seasonal demand analysis, and more detailed product-level comparisons. Such improvements can provide deeper business insights and support better decision-making in the future.

X. CONCLUSION

The analysis of chocolate sales data helps in converting raw sales records into meaningful business insights. By studying sales amount, profit, product performance, monthly trends, and regional sales, a clearer understanding of overall business performance is obtained. The use of Python, MySQL, and Power BI makes data handling, analysis, and visualization more organized and efficient. The insights gained from the analysis help in identifying best-selling products, strong-performing regions, profit variation, and changes in customer demand. Overall, the project shows how sales data can be transformed into useful information that supports better reporting, clearer business understanding, and improved decision-making for future growth.

XI. REFERENCES

[1] Kaggle Dataset – Chocolate Sales Data (CSV/Excel format)
<https://www.kaggle.com>

[2] Microsoft Power BI Documentation – Data Visualization and Dashboard Creation
<https://learn.microsoft.com/power-bi/>

[3] Python Official Documentation – Data Analysis and Processing
<https://docs.python.org/3/>

[4] MySQL Official Documentation – Database Management System
<https://dev.mysql.com/doc/>

[5] Business Intelligence and Data Analytics study materials used for understanding sales analysis and visualization techniques.