

E-COMMERCE SALES DATA ANALYSIS

Virendra Chaurasiya

Virendra Chaurasiya Sunder deep Engineering College Ghaziabad, India

Ms. Vandana Sharma Dept. Of Computer Science Sunder deep Engineering College Ghaziabad ,Dept. Of

Computer Science Sunder deep Engineering College, Uttar Pradesh, India

Abstract - The research shows that organizations in today digital business environment create their sales data through customer purchases product sales and daily customer interactions. Organizations depend on data analysis for assessing their business results which helps them make better decisions. The research team used Microsoft Excel SQL and Power BI to study sales data in order to find important business insights. The dataset contains essential information which includes product categories and sales amount and customer details and order dates and payment methods. The research team applied data preprocessing methods which include cleaning and filtering and removing inconsistencies to achieve accurate and reliable results.

Key Words: E-Commerce, Data Analysis, Data Visualization, Power BI, SQL, Customer Behavior, Profit Analysis, etc

1. INTRODUCTION

The current digital economy experiences its most rapid growth through e-commerce which generates extensive daily transactional data. Online retail businesses keep gathering customer order data together with product category information and sales totals and payment method details and regional sales data. Unprocessed sales data presents challenges for analysis because it requires specific analytical methods to extract meaningful insights from its unstructured state. The project analyzes sales data from structured CSV datasets which include Orders and Details to create valuable business insights through data analytics and visualization techniques. The dataset contains essential information which includes customer names and order dates and product sub-categories and quantities and profit and payment methods

1.1 Data Preprocessing

The business performance assessment establishes a complete operational view through data integration and processing activities. The data preprocessing process used data cleaning procedures to remove duplicate records and address missing data and format data through Microsoft Excel to achieve accurate and uniform results. The research identified critical patterns through analysis which examined how sales distributed among customers and profit fluctuated monthly and different product sub categories performed in the market

1.2 Data Visualization

Power BI was used to develop interactive dashboards which enable effective data visualization through their interactive dashboard system. The dashboard displays essential performance metrics which include total sales and quantity sold and average order value (AOV) and total profit. The visual elements of the presentation include bar charts and pie charts and donut charts which show profit distribution by sub-category and payment mode distribution and monthly profit trends and customer-wise sales contribution.

2. SALES TRENDS AND CUSTOMER CONTRIBUTION ANALYSIS

The analysis shows that sales and profit exhibit different patterns throughout the year because of seasonal variations. Certain product categories contribute more to overall revenue, while specific customers generate higher sales volumes.

The customer purchasing preferences are revealed through the distribution of payment modes which includes EMI cash on delivery credit card and debit card payments. The study demonstrates that Excel data preprocessing combined with Power BI visualization enables organizations to convert their raw sales data into useful business insights.

The data analysis process enables businesses to gain insights about their customers and their sales results and their overall operational efficiency.



Chart -1: Sales Trends

The techniques enable the processing of data through four distinct operations which include handling missing values and removing duplicate records and correcting inconsistencies and organizing the data into a structured format. The research team conducted data analysis after they completed the data cleaning process to identify patterns and relationships in the dataset through various analytical methods which included high-performing product identification and customer purchase behavior analysis and regional sales performance examination. The paper continues with Section 2 which examines existing research that relates to leaf disease detection. The proposed methodology in Section 3 presents all its components through a detailed step by-step explanation. The experiments are explained in Section 4 through the presentation of results and evaluation methods which included specific equations. The research findings are examined in Section 5 through an assessment of the positive and negative aspects that were identified during the study. Section 6 provides a work summary while assessing upcoming research opportunities.

3. RELATED WORK

The business intelligence field now depends on sales data analysis to support its decision-making processes which have become business-critical functions. Many researchers and professionals have explored different tools and techniques to transform raw sales data into meaningful insights that can improve organizational performance. One of the most commonly used tools in sales data analysis is Microsoft Excel. According to Winston (2016), Excel provides powerful features such as pivot tables, charts, and conditional formatting that allow users to clean, organize, and visualize data effectively. People use this tool for basic data analysis because it offers an easy-to-use interface that anyone can access. Structured Query Language (SQL) has also been highlighted as an essential tool for managing and analyzing large datasets. The data extraction process becomes easier through SQL which Linoff (2015) described as a powerful tool for locating specific data from relational databases while handling extensive structured sales information. SQL serves as an essential component required to convert raw data into formats suitable for subsequent analytical procedures and data representation activities. Power BI has become a powerful business intelligence solution that enables users to build interactive dashboards and display real-time data. Clark (2020) described Power BI as a reporting tool that enables organizations to view KPIs and monitor sales patterns while producing interactive business intelligence reports. The system provides effective sales analysis capabilities because it can connect with different data sources. People now choose Python-based data analysis tools because they provide users with flexible yet powerful tools which allow advanced data analysis work. Data manipulation and visualization tasks depend on the Pandas and Matplotlib libraries which data scientists widely employ. Researchers have shown that Python enables efficient handling of large datasets and supports advanced analytical techniques, including predictive analysis and machine learning. Research studies have demonstrated that people use data visualization techniques to comprehend complex datasets because these techniques help them interpret complex information. Sales data visualization techniques, including bar charts, line graphs, and dashboards, assist users in tracking sales data through its various trends and patterns while identifying any outlier sales activities

4. PROPOSED METHODOLOGY

The proposed methodology of this study focuses on transforming raw sales data into meaningful insights using a structured data analysis approach. The process involves multiple stages including data collection, preprocessing, analysis, and visualization using tools such as Microsoft Excel and Power BI. The proposed methodology ensures a systematic approach to analyzing sales data by combining data preprocessing, analysis, and visualization techniques to generate meaningful and actionable insights.

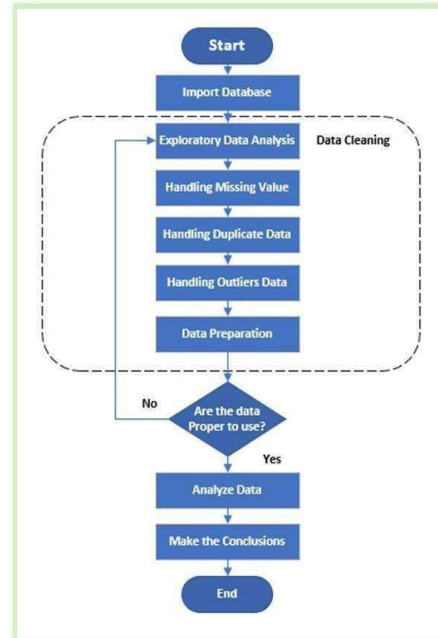


Figure 1. Sales Data Analysis Flowchart

4.1 Data Collection

The data used in this project is obtained from structured CSV files, namely Orders and Details. These datasets contain important information such as order dates, customer names, product categories and sub-categories, quantity sold, sales amount, profit, and payment modes. The datasets are integrated to form a unified data model for analysis. The Orders dataset includes attributes such as Order ID, Order Date, Customer Name, State, and Payment Mode. This dataset primarily focuses on customer-related and transactional information. On the other hand, the Details dataset contains product-specific information such as Product Category, Sub-Category, Quantity, Sales Amount, and Profit. Together, these datasets provide a complete view of sales activities from both customer and product perspectives. The collected data spans multiple transactions and includes various product categories such as clothing, electronics, and furniture. It also reflects different customer segments, regional sales distribution, and multiple payment methods including Cash on Delivery (COD),

EMI, credit card, and debit card. Overall, the collected dataset provides a reliable foundation for performing data preprocessing, analysis, and visualization. It enables the identification of sales trends, customer purchasing behavior, and product performance, which are essential for making informed business294. and product performance, which are essential for making informed business. The study used data which was obtained from structured CSV files containing two files named Orders and Details which contain complete details about sales transactions. The datasets offer a real-world e-commerce sales setting which enables researchers to study customer behavior and product performance and business operations. The two datasets get connected through a shared key which identifies all items in a customer's order through their unique Order ID. The relevant attributes need to be linked together for this step because this connection enables efficient combined analysis. The methodology requires data integration to create a single data structure which combines multiple datasets so that analysts can conduct their research. The study integrated two datasets' Orders and Details to achieve complete sales transaction data. The Orders dataset contains customer and transaction-level information such as Order ID, Order Date, Customer Name, State, and Payment Mode. The Details dataset provides product-level information through its Category, Sub Category, Quantity, Sales Amount, and Profit fields. The two datasets exist as separate entities which show limited information but their combination delivers a comprehensive view of customer behavior together with product performance. Data integration improved dataset quality through its enhancement process which enabled advanced analysis and valuable insights to support business decision making.

4.2 Data Augmentation

The researchers used data augmentation in their study to improve the existing sales dataset which resulted in better data quality and deeper data analysis capabilities. The original datasets which included Orders and Details contained fundamental transactional and product information but researchers developed extra derived attributes to enable more useful analysis and visualization. The existing dataset was used to create new features which served as the basis for augmentation instead of acquiring additional external data. The Order Date field enabled the extraction of additional time-based attributes which included month, quarter, and year. The derived attributes enabled researchers to study seasonal patterns and evaluate monthly sales results.

4.3 Data Preprocessing

Before analysis, the data is cleaned and prepared using Microsoft Excel. The process involves handling missing data and eliminating duplicate records and correcting entry errors and proper column formatting. The process of data filtering and sorting improves the quality and structural organization of the data. The team evaluated both outliers and data points that showed inconsistent behavior. The research team examined extreme values which had the potential to disrupt their analysis process. The research team handled the extreme values with caution to ensure their analysis results remained unaffected. Data preprocessing serves as an essential fundamental process because it establishes clean and consistent data standards which enable subsequent analytical work. The study team used Microsoft Excel to preprocess the Orders and Details datasets before they started data integration and visualization work in Power BI.

The initial process involved checking the datasets to find missing data and duplicate entries and data inconsistencies. The analysis process required missing data to be managed through two methods, which involved filling missing data with appropriate values or discarding incomplete records. The process helped to maintain data accuracy while ensuring the dataset remained trustworthy. The process of duplicate entry detection and deletion was performed to eliminate duplicate records, which would have led to incorrect analysis results. The calculation of sales and profit and quantity metrics needs every transaction to be treated as a distinct event to achieve precise results.

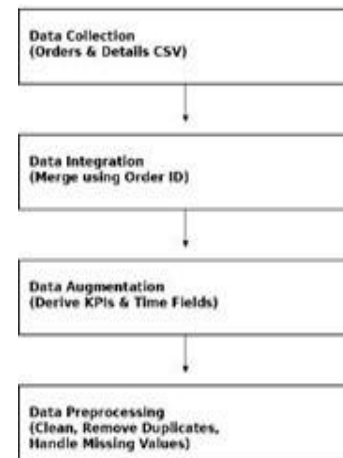


Figure 2. Workflow of data analysis

4.4 Data Analysis

The process of EDA begins after the completion of data preprocessing since it helps to discover both patterns and trends. The calculation process determines key metrics which include total sales and total profit and quantity sold and average order value AOV. The research analysis includes the following areas of study: Profit by sub-category Sales distribution across customers Monthly sales and profit trends Category-wise quantity distribution Payment mode preferences The initial evaluation of business performance used key performance indicators KPIs which included total sales and total profit and total quantity sold and average order value AOV. The dataset metrics provided a summary which helped to identify the complete sales pattern. The analysis of payment modes examined how customers use different payment methods by studying their usage of Cash on Delivery COD and EMI and credit card and debit card payment methods. The data analysis process changed raw sales information into useful business insights which enabled better marketing decisions and product performance improvements and data-led decision support.

4.5 Data Visualization

This research study depends on data visualization because it uses graphical representation to make difficult-to-understand and large amounts of sales data accessible. The analysis results were shown through various Power BI visualization techniques which interactive presentation method to display the obtained insights. The main goal of data visualization in this project exists to make sales data easier to understand while showing critical patterns

and existing trends and their associated connections. Business performance becomes easier to understand through visual elements that include charts and dashboards because these tools enable users to see the information without needing to analyze the actual numerical data. The "Fashion Hub Sales Dashboard" interactive dashboard was created to display essential business information. The dashboard displays vital key performance indicators (KPIs) which track total sales and total profit and total quantity sold and average order value (AOV) metrics. The indicators deliver a rapid assessment of total business performance.

4.6 Insight Generation

The visual analysis process creates business insights which help organizations identify their best-selling products and seasonal sales trends and their most important customers and their most popular payment methods. The resulting insights from this process help guide marketing efforts and inventory management and business strategy development.

Credit Card, and Debit Card. The donut chart shows how much of the total transactions were completed using each payment method through its different segments. The data shows that customers use various payment methods because no payment method has become completely dominant in customer preferences. The customer base prefers payment methods that enable them to make payments at their own schedule because EMI and COD account for a large portion of transactions. Customers use EMI to buy expensive items, while COD serves as a payment method for people who want to pay after they receive their products because they have trust issues with the product. The usage of credit card and debit card payments has grown because these payment methods now make up a large part of all transactions. People prefer these methods for their fast and easy and safe online transaction capabilities. The study of payment modes gives businesses important information about how customers buy products and choose their payment methods. Businesses can use this information to create better payment systems by detecting payment patterns that will help customers find their preferred payment methods. The information helps organizations develop their promotional campaigns by choosing suitable EMI offers and digital payment discounts and online



Fig.3 Power BI-Based Dashboard

4.7 Decision support

The process ends with researchers using their created insights to support business managers and decision-makers who need to make data-driven choices. The method provides organizations with a functional framework which they can use to enhance their performance to achieve successful business growth. The proposed methodology establishes a systematic sales data analysis process which combines data preprocessing, analysis, and visualization methods to produce valuable insights for decision-making, and Managers can then use it to adapt to market changes quickly via on time data-supported strategic decisions.

5. ANALYSIS SETUP AND RESULTS

Your data training extends until the month of October in the year 2023. The "Payment by Category" visualization shows how customers use different payment methods to make their purchases. The chart shows that customers use four main payment methods, which include EMI, Cash on Delivery (COD),

transaction rewards.

The "Profit by Month" visualization shows how profit changes between different months, which enables users to detect business performance patterns based on time. The chart shows that profit levels throughout the period display substantial variations because they do not maintain a constant value. The analysis shows that February produces the highest profit figure when compared to January and March because sales performance reaches its highest point during that month. The analysis provides cash flow insights which enable better financial planning because it shows cash flow patterns that occur during various periods. Businesses need to identify these trends because it helps them create better operational plans.



Fig. 4. Distribution of Payment Modes by Category



Fig. 5 Profit Trend Across Months

The profit by sub-category visualization demonstrates how different product subcategories contribute to total profit. The analysis helps identify products that perform well along with products that perform poorly within the company. The chart shows that phones and bookcases generate much more profit than other product subcategories. These products serve as essential revenue sources which drive business financial success. The subcategories of accessories and t-shirts with shirts generate less profit which demonstrates their performance level. Sub-category profit differences arise from customer demand patterns and pricing strategies and product popularity within the market. Businesses should promote their high-performing products while keeping them in stock and they should evaluate their low-performing products through new marketing techniques and pricing modifications. The analysis provides executives with essential support for making decisions about product management and pricing strategies and resource distribution. Businesses that concentrate their efforts on profitable sub-categories will achieve maximum profitability which leads to their sustainable growth.



Fig. 6 - Sub-Category Performance Based on Profit

DISCUSSION

The research results demonstrate that data analysis and data visualization techniques serve as essential tools for comprehending sales performance and customer behavior. The integrated sales dataset analysis uncovered multiple significant patterns which companies can use to make better business decisions. The monthly profit analysis shows that business performance patterns change throughout the year because of seasonal fluctuations. Certain months, such as February, show higher profitability, while others experience relatively lower performance. The customer demand fluctuations stem from three main factors which include promotional activities, market trends, and seasonal preferences. Businesses can enhance their marketing strategy development and resource distribution through understanding these variations. The discussion proves that organizations can use data-driven insights from proper analysis and visualization to achieve their strategic goals while their operational efficiency and business growth will become more sustainable.

6. CONCLUSION AND FUTURE SCOPE

The research work uses Microsoft Excel and Power BI tools to analyze and visualize e-commerce sales data. The main goal of the study required transforming the direct transactional data into business insights which would help organizations make better decisions. The research discovered essential patterns and trends from the data through sequential operations which included data collection, data integration, data preprocessing and data analysis and data visualization. The analysis showed that sales and profit numbers vary between months because seasonal demand patterns exist. The research identified multiple product sub-categories which contribute to total profit while others showed lower profit results. The customer analysis showed that a small number of customers account for most of the total sales which

makes customer retention strategies crucial for business success. The study showed that customers use various payment methods which create a need for businesses to offer different transaction methods for better customer service. Through Power BI dashboards, users can apply data visualization techniques to obtain actionable insights from intricate projects which they can follow through to their end results. The study shows that researchers can use Excel data preprocessing methods together with Power BI visualization tools to conduct effective sales data analysis. Researchers can enhance the dataset by adding customer demographic data and geographic data and market data from outside their organization. The marketing strategies of organizations can benefit from these insights while their product performance will increase and their business growth will proceed to the next level. The research study plans to use modern technologies which will enable data expansion to achieve better business intelligence results which will assist with decision-making and sustainable growth activities. Organizations can use advanced visualization tools and techniques to build interactive dashboards which users find easier to navigate and interact with.

7. REFERENCES

1. Clark D 2020 Beginning Microsoft Power BI A Practical Guide to Self Service Data Analytics Apress.
2. McKinney W 2022 Python for Data Analysis 3rd ed O'Reilly Media.
3. Provost F Fawcett T 2023 Data Science for Business 2nd ed O'Reilly Media.
4. Shmueli G Bruce P Gedeck P 2023 Data Mining for Business Analytics 4th ed Wiley.
5. Microsoft Corporation 2024 Power BI Documentation Available at <https://learn.microsoft.com>.
6. Microsoft Corporation 2024 Excel Data Analysis Tools Guide Available at <https://learn.microsoft.com>.
7. Oracle Corporation 2023 SQL Language Reference Available at <https://docs.oracle.com>.
8. Stack Overflow 2024 Developer Discussions on SQL and Power BI Available at <https://stackoverflow.com>.
9. Google Cloud 2023 Data Analytics and Visualization Tools Available at <https://cloud.google.com>.
10. Chen H Chiang R H Storey V C 2022 Business Intelligence and Analytics From Big Data to Big Impact MIS Quarterly.
11. Davenport T H Harris J G 2021 Competing on Analytics Updated Edition Harvard Business Review Press.
12. Marr B 2023 Data Strategy How to Profit from a World of Big Data Kogan Page.
13. Knaflic C N 2020 Storytelling with Data Wiley.
14. O'Neil C 2021 Weapons of Math Destruction Crown Publishing.
15. Aggarwal C C 2022 Data Mining The Textbook Springer.
16. IBM Data Science Team 2024 Introduction to Data Analytics IBM Learning Resources.
17. Microsoft Learn 2025 Power BI Data Visualization and Dashboard Design Available at <https://learn.microsoft.com>.
18. Google Analytics Academy 2023 Advanced Data Analysis.