E COMMERCE(WALMART) SALES ANALYSIS

Project Submitted to the

APSSDC

Bachelor of Technology

In

Computer Science and Engineering

and

Information Technology

Submitted By

Grandhi Sujaina 322506402128

GANDI CHANDRA SEKHAR 322506402107

GANGISETTI L V S S SUBRAHMANYAM 322506402110

Karra kanaka kalpavalli Sravani 322506402169

Akhila Kumari Bandi 322107311058

Under the guidance of

K. Meenakshi Mam,

M. Ruthumma Mam

ABSTRACT

This study presents a comprehensive analysis of Walmart's e-commerce sales data using Python’s robust data handling library, Pandas, along with insightful data visualization tools such as Matplotlib and Seaborn. The dataset employed in this analysis contains detailed transaction records from Walmart's online retail operations. It includes multiple variables such as order dates, product categories, regional sales performance, customer segments, and promotional discounts.

The primary goal of this analysis is to uncover trends, seasonal patterns, and key factors influencing sales performance and customer purchasing behavior. Through data preprocessing, exploratory data analysis (EDA), and visualization, the study highlights important insights such as top-performing categories, sales fluctuations over time, and regional demand variations.

Advanced metrics and KPIs such as total revenue, average order value, and customer retention patterns are analyzed to support business strategy. The insights derived from this project are valuable for optimizing marketing efforts, inventory planning, and enhancing customer engagement. By leveraging the power of data analytics, Walmart can make informed, data-driven decisions to maintain competitiveness and improve overall e-commerce performance.

INTRODUCTION

IN THE ERA OF DIGITAL TRANSFORMATION, E-COMMERCE HAS EMERGED AS A CRITICAL DRIVER OF RETAIL SUCCESS. AMONG LEADING GLOBAL RETAILERS, WALMART HAS SUCCESSFULLY EXPANDED

ITS TRADITIONAL BRICK-AND-MORTAR BUSINESS INTO THE ONLINE DOMAIN, SERVING MILLIONS OF CUSTOMERS THROUGH ITS E-COMMERCE PLATFORMS. THIS PROJECT, "E-COMMERCE WALMART SALES ANALYSIS," FOCUSES

ON ANALYZING HISTORICAL SALES DATA FROM WALMART'S ONLINE STORES TO UNCOVER KEY TRENDS, PERFORMANCE PATTERNS, AND ACTIONABLE BUSINESS INSIGHTS

THE PRIMARY OBJECTIVE OF THIS PROJECT IS TO EXPLORE HOW VARIOUS FACTORS—SUCH AS STORE TYPE, LOCATION,

SEASONAL EVENTS (LIKE HOLIDAYS), AND PROMOTIONAL STRATEGIES—IMPACT SALES PERFORMANCE ACROSS WALMART'S E-COMMERCE CHANNELS. THROUGH DATA VISUALIZATION AND STATISTICAL ANALYSIS, THE PROJECT

AIMS TO IDENTIFY OPPORTUNITIES FOR IMPROVING SALES FORECASTING, INVENTORY PLANNING, AND CUSTOMER TARGETING.

BY LEVERAGING TOOLS SUCH AS PYTHON, PANDAS, MATPLOTLIB, AND MACHINE LEARNING TECHNIQUES,

THIS ANALYSIS PROVIDES A COMPREHENSIVE OVERVIEW OF WALMART'S E-COMMERCE PERFORMANCE.

IT SUPPORTS DATA-DRIVEN DECISION-MAKING AND OFFERS RECOMMENDATIONS TO ENHANCE OPERATIONAL EFFICIENCY AND BOOST REVENUE IN A HIGHLY COMPETITIVE ONLINE RETAIL ENVIRONMENT.

SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENT:

- **OPERATING SYSTEM:**

  The analysis can be performed on Windows.

- **Python:**

  Python 3.x is required for running the analysis. Make sure you have the latest stable version of Python installed.

- **Libraries:**

  Pandas: Install the Pandas library using pip, a package manager for Python.

  pip install pandas

  Matplotlib: Install the Matplotlib library using pip

  pip install matplotlib

  Seaborn: Install the seaborn library using pip command

  pip install seaborn

HARDWARE REQUIREMNTS:

- **IDE** – Jupyter Notebook, Google Collaboratory
- **Storage Space** – free storage space enough for running on machine

ARCHITECTURE

1.DATA SOURCE

The analysis begins with obtaining the Walmart_sales dataset. Data can be collected from various sources, such as a comapany's database, publicly available datasets, or through web scraping from company's websites and platforms.

Once the data is acquired, it is typically stored in a structured format like CSV, Excel, or a database.

2.Data Ingestion

The raw data is imported into the analytical environment. Using tools like Python's Pandas, data is loaded through functions such as read_csv(). This ensures that the raw input is ready for further processing**.**

3.Data Preprocessing:

Data preprocessing is a crucial step to ensure data quality and consistency.

Using Python's Pandas library, the data is loaded into a DataFrame, allowing for easy data manipulation and analysis.

This step involves handling missing values, handling duplicates, converting data types, and addressing any data quality issues.

4.Exploratory Data Analysis (EDA):

EDA involves exploring the data to gain insights into its structure, distributions, and relationships between variables.

Pandas functions are used to perform summary statistics, groupings, and aggregations to understand key trends and patterns.

Data visualization with Matplotlib helps to create plots, histograms, scatter plots, and other visualizations to visualize patterns and correlations effectively.

5.Data Visualization:

Matplotlib is a powerful library for creating various types of visualizations, enabling the presentation of complex data in an intuitive and informative manner.

Visualizations are used to communicate key findings, such as booking patterns over time, customer segmentation, room preferences, and revenue trends.

6.Insights and Decision Making:

The final step is to draw meaningful insights from the analysis and make data-driven decisions to optimize hotel operations, marketing strategies, and customer experiences.

The analysis results can be presented in reports, dashboards, or interactive visualizations to aid stakeholders in understanding the key takeaways effectively

THE ARCHITECTURE OF THE E-COMMERCE [WALMART] SALES ANALYSIS USING PYTHON, MATPLOTLIB,AND PANDAS INVOLVES SEVERAL KEY STEPS THAT FORM A COHESIVE WORKFLOW..

This architecture helps you process, store, and analyze large volumes of sales data effectively:

IN CODING ARCHITECTURE WE HAVE DONE THE FOLLOWING STEPS:

1. IMPORTING DATASET

2.APPLYING BASIC FUNCTIONS

3 .DATA CLEANING

4.DATA FILTERING

5.GROUPING

6. SORTING

7. AGGREGATE QUERIES

8. VISUALIZATION

USES OF DATA ANALYSIS LIBRARY:

Pandas, Matplotlib, and Seaborn play crucial roles in the hotel booking analysis using Python, enabling a comprehensive and data-driven approach to understand booking patterns, customer preferences, and revenue trends. Here's a detailed short note on their uses in this analysis**:**

Pandas:

- Data Manipulation: Pandas provides powerful data manipulation capabilities, enabling easy loading, cleaning, and preprocessing of the hotel booking dataset. It allows filtering, grouping, and aggregating data to derive meaningful insights.

- Data Exploration: Pandas facilitates the exploration of booking patterns over time, customer segmentation based on demographics, and analysis of cancellation reasons, room preferences, and revenue metrics**.**

- Handling Missing Data: Pandas' functions handle missing data points effectively, ensuring data quality and preventing biases in the analysis.

- Data Transformation: It aids in transforming data into a format suitable for analysis, such as converting data types and applying mathematical operations.
- Joining and Merging: Pandas is used to combine datasets when additional information, such as customer reviews or hotel amenities, is available separately.

Matplotlib:

- Data Visualization: Matplotlib allows the creation of various visualizations like line plots, bar charts, and scatter plots to depict booking trends, revenue fluctuations, and customer distribution.

- Time Series Analysis: With Matplotlib, time series plots can illustrate booking patterns over specific time intervals, highlighting seasonal variations and trends.

- Geospatial Analysis: Matplotlib can generate geographic maps that visualize customer distribution, providing insights into high-demand regions and popular hotel locations**.**

Seaborn:

- Enhanced Data Visualization: Seaborn is built on top of Matplotlib and offers more aesthetically pleasing and informative visualizations. It simplifies the creation of complex plots like heatmaps, pair plots, and violin plots.

- Statistical Insights: Seaborn provides built-in statistical functions that allow us to easily visualize relationships between variables, such as correlation matrices or regression plots for revenue analysis**.**

- Categorical Data Visualization: It excels at visualizing categorical data, such as room types or customer nationalities, using bar plots or box plots, which aids in understanding preferences and trends.

ADVANTAGES

1.VERSATILITY AND FLEXIBILITY

- multi-category support: the system can handle diverse product categories like electronics, clothing, groceries, etc., making it adaptable to walmart's wide product range.
- customizable filters and reports: users can filter data by time period, product type, store location, etc., allowing for targeted insights.

2.RICH DATA ANALYSIS LIBRARIES

- pandas, matplotlib, and seaborn provide a comprehensive suite of data manipulation, visualization, and statistical functions. their extensive capabilities empower analysts to perform complex data analysis and generate informative visualizations with relative ease.

3.DATA VISUALIZATION

- matplotlib and seaborn excel at creating visually appealing and informative plots, charts, and graphs. these visualizations enhance the presentation of complex data, making it easier for stakeholders to understand trends, patterns, and correlations**.**

4.DATA-DRIVEN DECISION MAKING

- by analyzing booking patterns, customer preferences, and revenue trends, hotel management can make data-driven decisions to optimize pricing, marketing strategies, and operational efficiency. this leads to better resource allocation and improved customer experiences.
- Converts raw sales data into actionable business insights.
- Supports data-driven decisions in pricing, store performance, and customer targeting**.**

5.Identifies Sales Trends Over Time

- Analyse weekly, monthly, and seasonal sales trends.
- Helps Walmart plan inventory, logistics, and promotions based on high or low-demand periods**.**

6.Improves Inventory Management

- Highlights fast-moving vs. slow-moving products.
- Prevents overstocking or understocking by using sales history to predict future demand.

7.Optimizes Marketing and Promotions

- Evaluates impact of holidays and promotional events on sales.
- Helps target the right product categories and regions during marketing campaigns.

8.Assesses Store Performance

- Compares sales performance across stores to identify high-performing and underperforming locations.
- Helps focus managerial efforts and resources on weaker performing stores.

9.Customer Behavior Insights

- While the dataset is transactional, analysis can help infer patterns in purchase timing, category preferences, and sensitivity to discounts or holidays**.**

CONCLUSION

This project undertook a comprehensive analysis of Walmart's e-commerce sales data using Python, with a focus on discovering actionable insights to support data-driven business decisions. Using the walmart_sales.csv dataset, the analysis covered various dimensions of sales performance, such as store-wise sales trends, the effect of holidays, and the influence of economic indicators like CPI, unemployment rate, and fuel prices.

The dataset was first cleaned and prepared using Pandas, where missing values were handled, date columns were formatted, and additional features like month, week, and holiday flags were extracted. This was followed by exploratory data analysis (EDA) and visualization using libraries such as Matplotlib and Seaborn, which allowed for meaningful interpretation of sales data.

Several valuable insights were discovered:

- Store-wise Performance: Certain stores consistently outperformed others, suggesting the presence of high-traffic locations or better regional marketing strategies. This insight can help allocate resources more efficiently.
- Holiday vs. Non-Holiday Sales: Holiday periods significantly affected weekly sales, both positively and negatively, depending on the type of holiday. This finding indicates that marketing and stocking strategies must be tailored for each holiday season.
- Seasonal Trends: Sales spikes were observed during specific months, especially around the end-of-year holidays. These trends are essential for demand forecasting and inventory planning.
- Impact of Economic Indicators: A moderate correlation was identified between unemployment rates, CPI, and weekly sales. This implies that Walmart's online sales are sensitive to broader economic factors and must be monitored for planning future campaigns.

Visualizations such as time-series plots, heatmaps, and box plots helped convey these findings clearly and supported better understanding of complex relationships within the data.

Implications, Limitations, and Future Work

The implications of this analysis are substantial. Walmart and similar retailers can use these insights to:

- Optimize inventory by predicting high-demand periods.
- Customize marketing strategies by aligning them with seasonal behavior and holiday impacts.
- Improve operational planning by comparing underperforming and overperforming stores.
- Monitor macro-economic trends and their effects on customer spending patterns.

Despite the meaningful insights generated, this project is not without limitations:

- The dataset used (walmart_sales.csv) is limited in scope. It lacks detailed customer information (like demographics, loyalty, or cart size), which would allow for deeper customer segmentation and behavioral analysis.
- The analysis did not incorporate real-time data or external data sources such as competitor pricing, social media sentiment, or promotions beyond the holiday flag.
- Machine learning models for sales forecasting were not implemented, although the data trends clearly support such applications.

Future Scope of Work

This project lays the groundwork for more advanced analytics and business intelligence capabilities:

- Sales Forecasting Models: Implementing models like ARIMA, Prophet, or LSTM could help predict future sales at store and category levels.
- Interactive Dashboards: Using tools like Tableau, Power BI, or Streamlit to create dynamic dashboards for real-time monitoring.
- Customer Behavior Modeling: With enriched data, further studies can explore churn prediction, recommendation systems, or basket analysis.
- A/B Testing: Evaluate the impact of different promotions or campaigns through statistical testing.

In conclusion, the Walmart E-Commerce Sales Analysis project demonstrates how structured data analysis can transform historical sales data into valuable insights. It highlights the power of data in driving smarter business strategies, boosting profitability, and adapting to evolving market demands. With further refinement and scale, this kind of analysis can form the core of any data-driven retail decision-making system.

Reference Links and websites:

https://www.google.com/

Dataset:

https://www.kaggle.com/datasets/yasserh/walmart-dataset