Final Project Report

Title: VASTU (Visual Analytics and Sales Tracking Utilities)

Sector: Retail

Dataset: Walmart Sales Forecast (2010-2012).

Institute: Newton School of Technology

Faculty Mentor: Satyaki Das

Team Members: Bhavay Goyal, Manpreet Singh, Udata Varun, Ritesh, Kevish Sewliya, Siddarth Dangi

Submission Year: 2026

1. Executive Summary

This report evaluates Walmart's sales performance from 2010 to 2012 through a structured, KPI-driven dashboard built in Google Sheets. The dashboard integrates sales data with store attributes, promotional markdown information, and economic indicators to provide a consolidated view of business performance.

Interactive charts, trend analysis, and key performance indicators are used to track annual sales movement, compare store types, measure promotional effectiveness, and identify seasonal patterns. The dashboard enables quick identification of high-performing and underperforming stores, revenue concentration risks, and periods of sales volatility.

The objective of this dashboard-based analysis is to transform raw data into intuitive visual insights that support data-driven decision-making in retail operations, including promotion planning, store optimization, and strategic performance monitoring.

2. Sector & Business Context

The retail sector operates in a highly competitive and data-intensive environment, where sales performance is influenced by consumer demand, pricing strategies, promotional activities, and economic conditions. Large retailers like Walmart manage diverse store formats, making performance monitoring across locations essential.

During the 2010–2012 period, the industry experienced economic fluctuations and shifts in consumer spending, increasing the importance of promotions and seasonal planning. In this context, dashboard-based analysis helps retailers track performance, compare

store types, identify risks, and support informed operational and strategic decision-making.

3. Problem Statement & Objectives

This project represents a real-world retail analytics challenge commonly faced by large retail chains like Walmart. Managing multiple stores across regions requires accurate weekly sales forecasting to support inventory management, staffing, promotions, and overall operational efficiency.

Retail sales are influenced by factors such as seasonality, holidays, economic indicators (CPI, unemployment), fuel prices, and regional demand variations. Without data-driven analysis, retailers risk inefficiencies such as overstocking, stockouts, and lost revenue opportunities.

Objectives

The objectives of this project are to:

- Analyze historical weekly sales trends across stores and time periods
- Identify key drivers affecting store-level sales performance
- Evaluate the impact of holidays and economic factors
- Develop KPIs and dashboards to support forecasting and strategic decision-making

4. Data Description

- **Source:** Walmart Sales Dataset (Kaggle)
- **Time Period:** 2010 – 2012
- **Granularity:** Weekly, store-level and department-level data
- **Structure:** Multiple structured CSV files containing temporal, categorical, and numeric dimensions

- **Core Datasets & Columns:**
    - **Store** – Unique store identifier
    - **Dept** – Department identifier within each store
    - **Date** – Weekly sales date
    - **Weekly_Sales** – Sales revenue for the given store and department
    - **IsHoliday** – Indicator for holiday weeks
    - **Store Type** – Store classification (A, B, C)
    - **Store Size** – Physical size of the store
    - **MarkDown1–5** – Promotional markdown variables
    - **Temperature** – Regional weekly temperature
    - **Fuel_Price** – Fuel cost during the week
    - **CPI** – Consumer Price Index
    - **Unemployment** – Regional unemployment rate
- **Data Size:**
    - 45 stores
    - 80+ departments
    - ~420,000 weekly records
- **Strategic Consideration:**
  The dataset was provided in a partially uncleaned form, including missing values, inconsistent markdown data, and holiday flags. This design enables realistic data preparation and validation workflows before establishing a reliable analytical dataset for KPI design, dashboarding, and business insights.

5. Data Cleaning & Preparation

To ensure reliable analysis and accurate dashboard insights, a structured data cleaning process was applied before visualization and KPI development. The dataset contained formatting inconsistencies and missing values that required systematic treatment to establish a consistent and analysis-ready dataset.

5.1 Date Standardization

The original **Date** column was stored in text format, which limited its usability for time-series analysis and dashboard charting. To resolve this:

- A new column named **Cleaned Date** was inserted next to the original Date column.

  The `DATEVALUE()` function was applied to convert text-based dates into standard date values:

  ```
  =DATEVALUE(B2)
  ```

- This formula transformed text dates into serial date values, which were then formatted as standard dates using **Format → Number → Date**.

- The cleaned date column was applied across all rows to ensure consistency.

This step enabled accurate time-based analysis, trend visualization, and proper integration with dashboard filters.

5.2 Handling Missing Values

Several columns contained missing or null values that could distort analysis if left untreated. The following strategies were applied:

- **Promotional Markdowns (MarkDown1–MarkDown5):**
  Missing values were replaced with **0**, as the absence of markdown data indicates that no promotion was applied during that period.

- **Economic Indicators (CPI & Unemployment):**
  Missing values were imputed using the **column mean (average)**. This approach

preserves overall distribution trends while avoiding bias introduced by extreme values.

These methods ensured numerical stability and prevented gaps in KPI calculations and visualizations.

5.3 Data Validation

After cleaning:

- All date fields were verified for consistency
- No null values remained in critical analytical columns
- Numeric columns were validated for correct formatting

The final cleaned dataset is fully prepared for exploratory analysis, KPI creation, and dashboard-based performance monitoring.

6. KPI & Metric Framework

7.1 Business Objectives

The objectives of this analysis are to:

- Evaluate overall revenue growth (2010–2012)
- Identify high-performing stores and departments
- Analyze seasonal and yearly sales variations
- Assess the impact of promotional markdowns on sales
- Understand the influence of external economic factors such as fuel price

---

7.2 Key Performance Indicators (KPIs)

1. Total Sales

- **Definition:** Total revenue across all stores and departments
- **Formula:** ∑(Weekly_Sales)
- **Significance:** Measures overall business performance
- **Insight:** Sales dipped in 2011 and recovered strongly in 2012

2. Average Weekly Sales

- **Definition:** Average revenue generated per week
- **Formula:** ∑(Weekly_Sales) / Total Transactions
- **Significance:** Indicates sales consistency
- **Insight:** Higher averages in 2012 show improved performance

3. Total Markdown

- **Definition:** Total promotional discount offered
- **Formula:** ∑(Total_MarkDown)
- **Significance:** Measures promotional intensity
- **Insight:** Higher markdowns align with increased sales

4. Total Transactions

- **Definition:** Total number of weekly sales records
- **Formula:** COUNT(Entries)
- **Significance:** Represents operational scale

5. Average Fuel Price *(Supporting KPI)*

- **Definition:** Average fuel price over selected period
- **Formula:** AVG(Fuel_Price)
- **Significance:** Reflects macroeconomic influence

7. Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to understand sales behavior, identify patterns, and uncover relationships between sales and influencing factors before deriving insights and KPIs. Visual analysis and summary statistics were used to explore trends across time, store types, promotions, and economic conditions.

7.1 Sales Trend Analysis

- Annual sales analysis shows strong performance in 2010, a noticeable decline in 2011, and a clear recovery in 2012.
- This pattern indicates temporary business or economic pressure followed by operational stabilization and growth.

7.2 Store & Store Type Analysis

- Store Type **A** consistently generates the highest sales, contributing the majority of total revenue.
- Store Types **B** and **C** show significantly lower sales volumes.
- A small number of stores contribute disproportionately high revenue, highlighting revenue concentration risk.

7.3 Promotional Impact Analysis

- Weeks with higher promotional markdowns show increased average weekly sales.
- Medium promotions deliver moderate uplift, while low or no promotions result in comparatively lower sales.
- This confirms that promotions play a strong role in driving demand.

7.4 Seasonal & Holiday Patterns

- Sales demonstrate clear seasonal behavior, with strong mid-year performance and spikes during holiday weeks.
- Certain months consistently underperform, indicating opportunities for targeted promotional planning.

7.5 Economic Indicator Analysis

- CPI and unemployment trends show limited short-term correlation with weekly sales.
- Despite economic fluctuations, sales recovery in 2012 suggests operational resilience.

EDA Summary

The EDA confirms that sales performance is primarily influenced by store type, promotional activity, and seasonal demand patterns, while macroeconomic indicators have a secondary impact. These findings provide the foundation for KPI selection, dashboard design, and strategic business insights.

8. Advanced Analysis

Advanced analysis was performed to move beyond descriptive trends and identify deeper performance drivers and risk patterns within the retail sales data. This phase focuses on comparative evaluation, segmentation, and relationship analysis to support strategic decision-making.

8.1 Store Performance Segmentation

Stores were segmented based on total sales and average weekly sales to classify them into high-performing, mid-performing, and underperforming groups.

- A small subset of stores consistently outperforms others, contributing a significant share of total revenue.
- Several stores remain persistently underperforming, indicating potential issues related to location, store size, or demand.

8.2 Revenue Concentration Analysis

Contribution analysis reveals that a limited number of stores and Store Type A drive a large portion of overall sales.

- This highlights revenue concentration risk, where business performance becomes overly dependent on a small set of stores.
- Any disruption in these key stores could materially impact total revenue.

8.3 Promotion Sensitivity Analysis

Sales response was evaluated across different levels of promotional markdowns.

- High promotion periods demonstrate a clear uplift in sales.
- However, the analysis indicates diminishing returns beyond certain promotion levels, suggesting the need for optimized, targeted discounting rather than blanket promotions.

8.4 Time-Based & Seasonal Decomposition

Sales were analyzed across months and weeks to isolate seasonal effects.

- Strong recurring peaks were observed during specific holiday periods.
- Consistent low-sales periods indicate opportunities for inventory adjustment and demand-stimulation strategies.

9. Dashboard Design

- Designed using a layered executive-analytics structure, starting with high-level KPIs (Total Sales, Avg Weekly Sales, Total Markdown, Transactions, Avg Fuel Price) to provide an instant snapshot of overall business performance.

- Integrated interactive slicers (Year, Store Type, Holiday, Markdown Level) enabling dynamic filtering across all charts for multi-dimensional analysis and decision-making.

- Incorporated diverse visualization types including bar charts (annual performance), line charts (monthly seasonality), pie/donut charts (store type & promotional impact), bubble chart (fuel price influence), and ranking charts (top stores & departments) to capture trend, distribution, impact, and contribution analysis.

- Structured layout follows clear visual hierarchy principles, ensuring KPIs are most prominent, trend charts are centrally placed, and contribution charts are positioned for drill-down insights.

- Applied consistent color theme, clean alignment, proper currency formatting, and balanced spacing to maintain professional BI standards and executive readability while minimizing visual clutter.

10. Insights Summary

- **Sales Trend Recovery:**
  Sales performed strongly in 2010, declined in 2011, and showed a clear recovery in 2012, indicating business resilience after a temporary slowdown.

- **Store Type Dependency:**
  Store Type **A** contributes the majority of total revenue, highlighting strong performance but also over-dependence on a single store category.

- **Revenue Concentration Risk:**
  A small number of stores generate a disproportionately high share of total sales, increasing exposure to localized operational risks.

- **Promotion Effectiveness:**
  Promotional markdowns significantly increase average weekly sales, confirming their effectiveness in driving customer demand.

- **Seasonal Demand Patterns:**
  Sales exhibit strong seasonality, with noticeable spikes during holiday periods and consistent low-sales months.

- **Economic Impact:**
  CPI and unemployment show limited short-term influence on weekly sales, suggesting that internal operational factors play a larger role in performance.

- **Underperforming Stores:**
  Several stores consistently underperform relative to peers, indicating opportunities for operational improvement and targeted intervention.

11. Recommendations

Reduce Revenue Concentration Risk:

- Analyze best practices from top-performing stores
- Replicate successful product mix, layout, and promotion strategies in mid-performing stores

Optimize Promotion Strategy:

- Focus promotions on high-impact periods and categories
- Avoid blanket discounting to protect margins
- Use data-driven insights to time and target promotions

Strengthen Underperforming Stores:

- Improve inventory assortment based on local demand
- Implement localized marketing and pricing strategies
- Review store operations and layout for efficiency

Improve Seasonal Planning:

- Increase inventory and staffing ahead of peak seasons
- Plan targeted campaigns during historically low-sales periods

Monitor Economic Indicators:

- Track CPI and unemployment trends to anticipate demand changes
- Adjust pricing and inventory planning during economic fluctuations

12. Impact Estimation

The implementation of the proposed recommendations is expected to deliver measurable operational and financial benefits across Walmart's retail operations.

- **Improved Forecast Accuracy:**
  Better use of historical trends, seasonality, and promotional insights can enhance sales forecasting accuracy, leading to improved inventory planning and reduced stockouts or overstocking.

- **Revenue Growth Optimization:**
  Targeted promotions and improved performance of mid- and underperforming stores can increase overall sales without excessive reliance on discounts.

- **Cost Efficiency Gains:**
  More accurate demand planning can reduce inventory holding costs, minimize wastage, and improve staffing efficiency.

- **Risk Reduction:**
  Diversifying revenue contribution across stores and store types can lower exposure to revenue concentration risk.

- **Decision-Making Efficiency:**
  Dashboard-driven KPIs enable faster and more informed strategic and operational decisions.

13. Limitations

Historical Data Dependency:
The analysis is based on historical sales data from 2010–2012 and may not fully reflect current market conditions or recent changes in consumer behavior.

Limited External Factors:
While key economic indicators such as CPI and unemployment are included, other influencing factors (competitor pricing, marketing spend, online sales impact) are not captured.

Promotion Profitability Not Measured:
The analysis evaluates the impact of promotions on sales volume but does not assess profit margins or cost implications of markdown strategies.

Data Quality Constraints:
Certain fields, such as promotional markdowns, contained missing values that required imputation, which may slightly affect accuracy.

No Predictive Modeling:
The project focuses on descriptive and diagnostic analysis using dashboards rather than advanced predictive or forecasting models.

14. Future Scope

- **Advanced Sales Forecasting:**
  Implement time-series and machine learning models to improve weekly sales prediction accuracy at store and department levels.

- **Profitability Analysis:**
  Extend the analysis to include cost and margin data to evaluate the true profitability of promotional strategies.

- **Granular Store Segmentation:**
  Segment stores based on regional demand, customer behavior, and performance

trends for more targeted decision-making.

- **Real-Time Dashboard Integration:**
  Enhance the dashboard by integrating real-time or near real-time data for continuous performance monitoring.

- **Inclusion of Additional External Factors:**
  Incorporate variables such as competitor pricing, online sales influence, and local events to improve analytical depth.

15. Conclusion

This project demonstrates the value of data-driven analysis in understanding and improving retail sales performance. By analyzing Walmart's sales data from 2010 to 2012, the study identifies key trends related to store performance, promotions, seasonality, and revenue distribution.

Through data cleaning, exploratory analysis, KPI development, and dashboard visualization, raw data was converted into actionable business insights. The findings highlight the importance of targeted promotions, effective seasonal planning, and balanced store performance to reduce operational risks.

Overall, the project illustrates how analytical dashboards can support informed decision-making, improve operational efficiency, and enable sustainable growth in large-scale retail environments.

16. Appendix

A. Dataset Overview

- **Dataset Name:** Walmart Weekly Sales Dataset
- **Source:** Kaggle
- **Time Period:** 2010 – 2012

- **Granularity:** Weekly, store-level and department-level

B. Key Variables Used

- **Store:** Unique store identifier
- **Dept:** Department identifier
- **Date / Cleaned Date:** Weekly sales date
- **Weekly_Sales:** Sales revenue per store and department
- **IsHoliday:** Indicator for holiday weeks
- **Store Type:** Classification of stores (A, B, C)
- **Store Size:** Physical size of the store
- **MarkDown1–5:** Promotional markdown values
- **Temperature:** Regional weekly temperature
- **Fuel_Price:** Fuel price
- **CPI:** Consumer Price Index
- **Unemployment:** Regional unemployment rate

C. Data Cleaning Summary

- Text-based dates converted to standard date format using `DATEVALUE()`
- Missing markdown values replaced with zero
- Missing CPI and unemployment values imputed using column averages
- Data validated for consistency and numeric formatting

D. Tools & Techniques

- **Data Cleaning & Analysis:** Google Sheets
- **Visualization & Dashboard:** Google Sheets charts, filters, KPI cards
- **Techniques Used:**
    - Exploratory Data Analysis (EDA)
    - KPI-based performance evaluation
    - Trend and comparative analysis

E. Assumptions

- Missing markdown values indicate no promotional activity
- Historical sales patterns are representative for trend analysis
- Economic indicators have indirect influence on short-term sales

17. Contribution Matrix

Team member name	Data Collection	Data Cleaning & Preparation	EDA & Analysis	KPI & Dashboard	Report Writing	Overall Role
Bhavay Goyal	✓	✓		✓	✓	**Project Lead / Data Analyst**
Udata Varun	✓	✓	✓			**Data Preparation Lead**
Manpreet Singh	✓	✓	✓	✓		**Dashboard & Visualization Lead**
Ritesh	✓		✓		✓	**Data Support & Documentation**
Kevish Sewliya	✓				✓	**Analysis Support**
Siddarth Dangi	✓	✓				**Insights & Reporting Lead**