Author: Joti Khatri

Retail Sales Forecasting: Walmart Case Study

A data analytics project following the Google Data Analytics framework.

Introduction

Retailers face constant challenges in balancing inventory, pricing, and promotions to meet customer demand. Walmart, one of the largest retail chains in the world, collects extensive sales and store data that provides opportunities for data-driven decision making.

This case study applies the **data analysis process (Ask → Prepare → Process → Analyze → Share → Act)** to Walmart's historical sales dataset. The dataset contains weekly sales records across multiple stores and departments, enriched with features such as store size, type, holiday events, markdowns, fuel prices, consumer price index (CPI), and unemployment rates.

The primary goal is to answer the business question:

"How can Walmart optimize promotions and inventory based on seasonal demand patterns and external factors?"

By exploring trends, seasonal fluctuations, and the influence of holidays and economic indicators, this study aims to generate actionable insights for:

- Store managers (inventory allocation).
- Marketing teams (promotion planning).
- Business strategists (long-term demand forecasting).

The outcome of this case study will be a set of clear recommendations, supported by data visualizations and analysis, that demonstrate how advanced analytics can guide real-world retail decision making.

Step 1 – Ask

- **Business question:** How can a retail chain optimize promotions and inventory based on seasonal demand patterns and external factors?
- **Stakeholders:** Store managers (inventory planning), Marketing (promotion timing and targeting), Supply Chain (stock allocation), and Academic reviewers (evaluation of analysis methodology).
- **Objective:** Identify seasonal patterns, holiday effects, department/store differences, and produce actionable recommendations for promotion timing and inventory allocation.

Step 2 – Prepare

- **Dataset:** "Walmart Recruiting — Store Sales Forecasting" (Kaggle). Combined files used: train.csv (weekly sales per Store & Dept), features.csv (Temperature, Fuel_Price, CPI, Unemployment, MarkDown1–5, IsHoliday), and stores.csv (Store, Type, Size).
- **Time span & granularity:** Weekly sales over multiple years across ~45 stores and many departments (exact years covered are shown in the dataset).
- **Limitations:** Markdown columns contain many missing values; there is no customer-level demographic data; external features are macro indicators, not causal proofs. These limitations are noted and handled in the Process section.

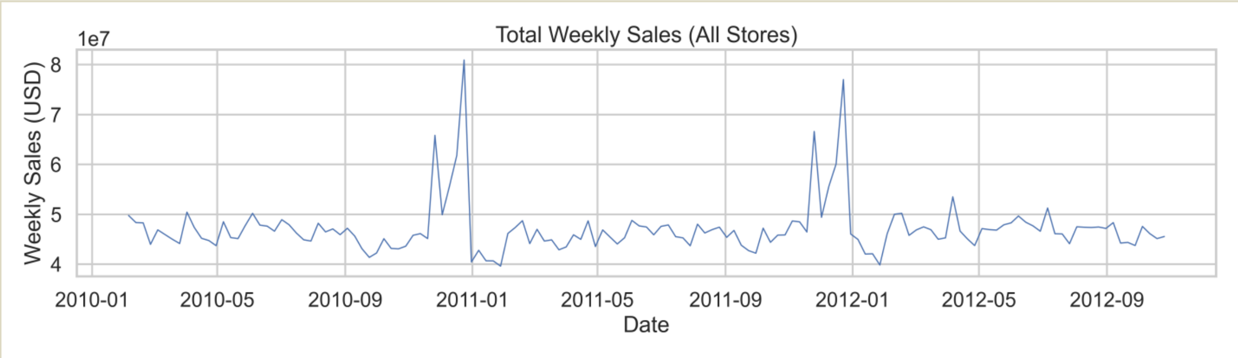
Step 3 – Process

- Combined the three main datasets (`train.csv`, `features.csv`, `stores.csv`) into one master dataframe using common keys (`Store`, `Date`).
- **Handled missing values:**
  - Used forward-fill for continuous columns like `CPI`, `Unemployment`, and `Fuel_Price`.
  - For Markdown features (`MarkDown1-5`), missing values were retained for analysis transparency since imputation could distort seasonal effects.
- Converted date columns into `datetime` objects and extracted **Year**, **Month**, and **Week** features.
- Removed duplicate columns (merged `IsHoliday_x` and `IsHoliday_y` into one `IsHoliday` column).
- Verified data consistency — each record represents one store–department–week combination.

Step 4 – Analyze

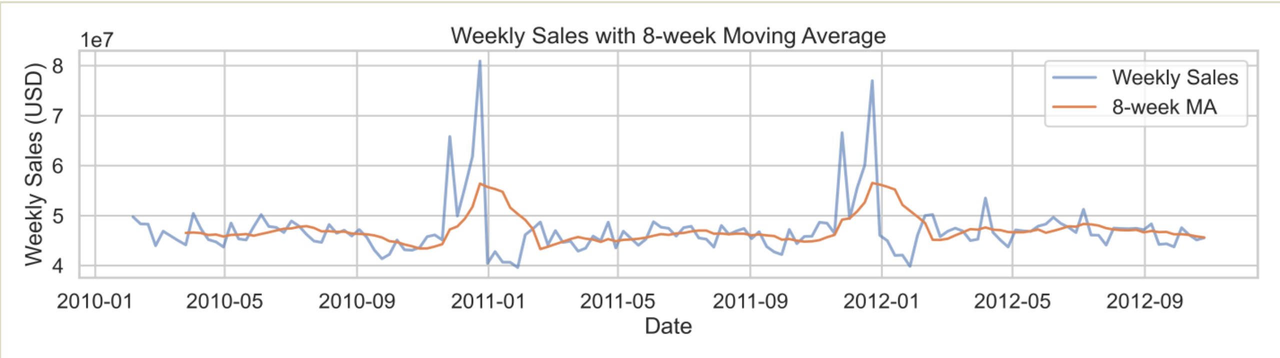
4.1 Total Weekly Sales Over Time

- **Visualization:** Line plot showing total sales aggregated across all stores per week.
- **Insight:** Overall company-wide sales follow seasonal patterns; volatility may be influenced by promotions, weather, or economic conditions.



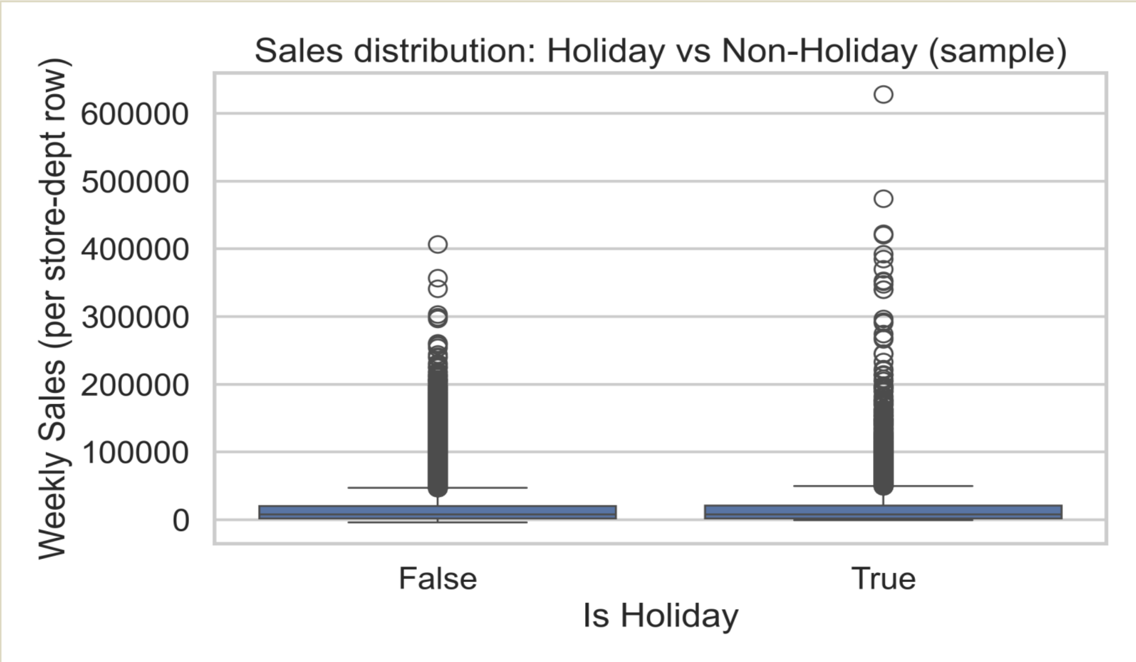
4.2 Weekly Moving Average of Sales

- **Visualization:** Line plot showing the 4-week moving average of weekly sales.
- **Insight:** Sales trends show periodic increases around major holidays, with noticeable dips post-holiday — reflecting clear seasonal cycles.



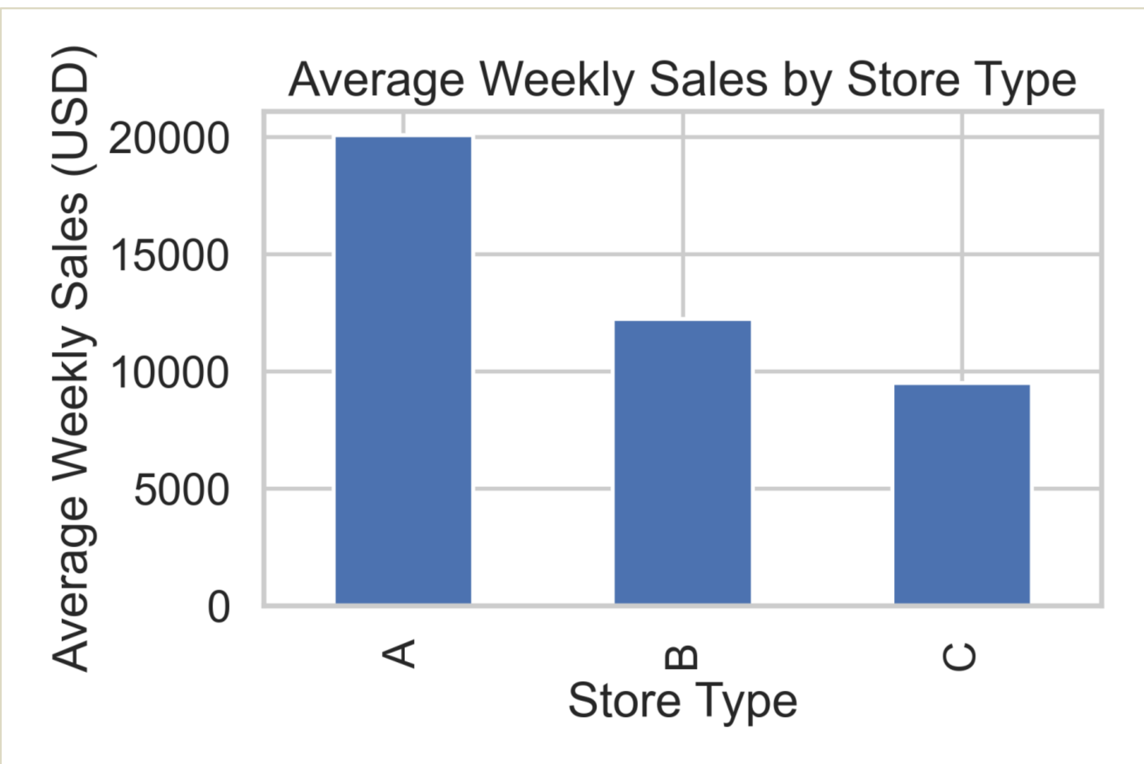
4.3 Weekly Sales by Holiday vs. Non-Holiday

- **Visualization:** Boxplot comparing sales during holiday and non-holiday weeks.
- **Insight:** Average weekly sales show noticeable spikes during holiday weeks, confirming that promotions and stock allocations should anticipate increased demand.



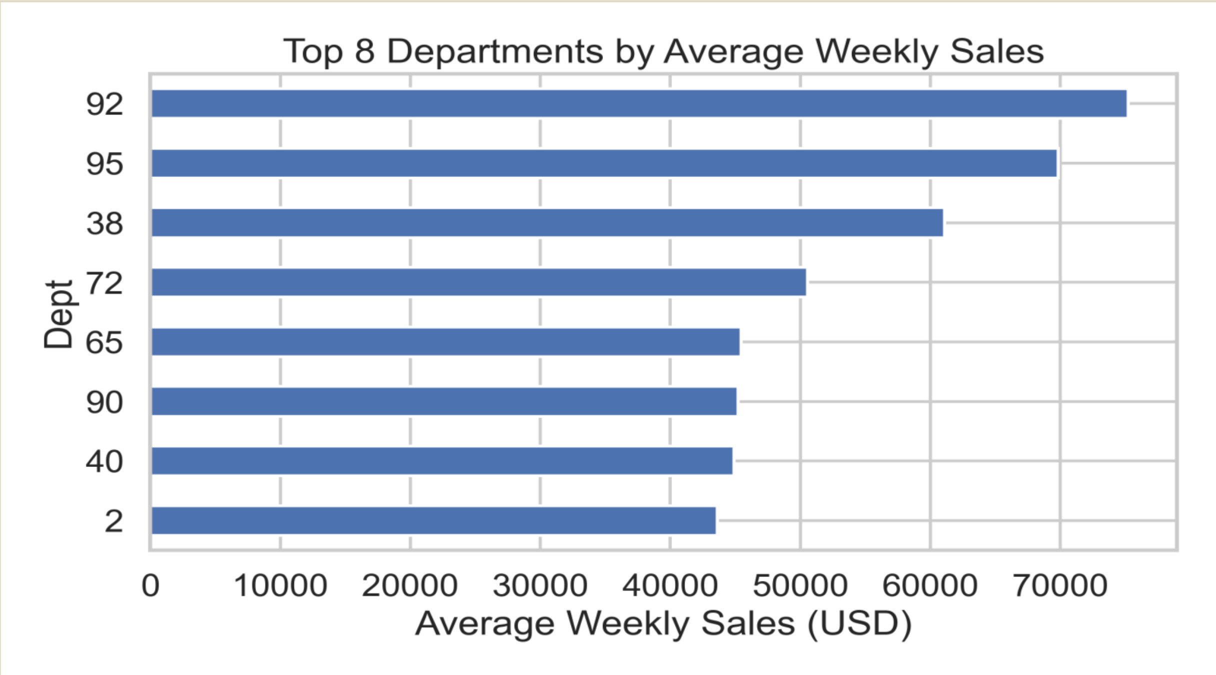
4.4 Average Sales by Store Type

- **Visualization:** Bar chart comparing average weekly sales across store types (A, B, C).
- **Insight:** Type A stores (largest) consistently outperform others, likely due to higher customer traffic and product variety.



4.5 Top Performing Departments

- **Visualization:** Bar chart of top departments by average weekly sales.
- **Insight:** A few departments dominate total revenue, suggesting that promotions and inventory focus should prioritize these high-performing areas.



Step 5 – Share

Summary of insights:

The analysis revealed clear seasonality and structural performance patterns across Walmart stores. Weekly sales peak around major holidays and gradually decline afterward, highlighting predictable cycles. Type A stores drive the majority of revenue, and a few departments consistently outperform others—making them prime candidates for targeted promotions and optimized stock levels.

Communication approach:

Insights could be shared through a **dashboard** (built in Tableau, Power BI, or Python dashboards) and a **concise executive report** summarizing:

- Weekly and moving-average sales trends
- Holiday vs. non-holiday performance
- Store-type and department comparisons
- Key recommendations for marketing and inventory teams

For stakeholders:

- **Store managers** → view store-level dashboards for inventory alignment.
- **Marketing teams** → focus on timing and depth of discounts.
- **Supply-chain planners** → adjust stock allocations to match predicted demand surges.
- **Academic reviewers** → evaluate methodology and reproducibility through the shared notebook.

Step 6 – Act

Business recommendations

1. **Holiday readiness:** Increase inventory and promotional intensity 2–3 weeks before major holidays to capture seasonal spikes.
2. **Department prioritization:** Focus on top-performing departments for markdown campaigns; monitor underperforming ones for potential rationalization or assortment refresh.
3. **Store-type strategy:** Allocate resources proportionally—Type A stores can handle bulk promotions, while smaller types may benefit from niche targeting.
4. **Weather & economic factors:** Track variables such as temperature, fuel price, and CPI regularly; integrate them into future demand-forecasting models.
5. **Automation & dashboards:** Deploy a live dashboard to monitor weekly sales KPIs, update automatically from source files, and trigger alerts for anomalies.

Next steps for improvement

- Introduce **predictive modeling** (e.g., regression or machine learning) to forecast future weekly sales.
- Incorporate **external datasets**—such as regional events or consumer sentiment—to enhance accuracy.
- Conduct **A/B testing** on promotions to quantify impact.
- Transition from retrospective analysis to **real-time analytics** for proactive decision-making.

Developed in Python (Pandas, Matplotlib, Seaborn). Full notebook available on GitHub.