



This report covers analysis and effective visualization of the effects big data in the retail business.

Walmart Sales Analysis and Business Decision Support Using Big Data

Big Data Systems and Analytics

Sheikh S R Shapon
UB: 24035767

Contents

Abstract	3
1. Introduction.....	3
1.1 Overview of the Business	3
1.2 Business Challenge in the Post-Pandemic Era.....	3
2. Dataset Description.....	4
2.1 Business Relevance of the Data	5
3. Objectives and Goals	5
4. Methodology	5
4.1. Data Preprocessing.....	5
4.2. Exploratory Data Analysis (EDA)	6
4.3. Predictive Modeling.....	6
5. Results.....	7
5.1. Visualizations	7
5.2. Predictive Model Accuracy	9
6. Impact on Business	10
6.1. Sales Forecasting	10
6.2. Marketing Campaigns.....	10
6.3. Supply Chain Optimization.....	10
6.4. Dynamic Pricing Strategies.....	10
7. Legal, Ethical, and Social Considerations	10
7.1. Data Privacy and Consumer Trust	10
7.2. Ethical Use of Data	10
7.3. Social Responsibility	11
8. Conclusion	11
Bibliography	11

Abstract

Here, Walmart analyzes sales data alongside external factors such as temperature and fuel prices to address post-pandemic challenges. The insights are used to optimize inventory management, forecast sales, optimize marketing campaigns, and dynamic pricing strategies. They highlight Big Data as critical to operational efficiency, customer targeting, and business performance.

1. Introduction

1.1 Overview of the Business

Walmart is one of the largest multinational retail corporations, operating a vast network of stores across multiple regions. With its focus on providing low-cost goods to customers, Walmart has long been at the forefront of using technology to improve business operations, from supply chain management to enhancing the customer shopping experience. However, the **COVID-19 pandemic** brought unprecedented challenges, forcing businesses like Walmart to adapt to a new retail environment rapidly. The pandemic caused shifts in consumer behavior, disrupted supply chains, and accelerated the demand for e-commerce and digital engagement (Smith, 2021).

In this context, Walmart's ability to leverage **Big Data** to understand and predict customer behavior, optimize inventory, and forecast sales has become increasingly important. This report uses **Walmart's sales dataset** to demonstrate how Big Data techniques can drive intelligent business decisions, especially during economic uncertainty and shifting market conditions.

1.2 Business Challenge in the Post-Pandemic Era

The retail sector, especially large-scale operations like Walmart, has faced several challenges because of the **COVID-19 pandemic**:

- **Shifting consumer preferences:** There has been a noticeable increase in demand for e-commerce and online shopping, with many consumers avoiding in-person store visits.
- **Disruptions in supply chains:** The global supply chain was severely impacted, leading to product shortages and delayed deliveries.
- **Economic instability:** Rising unemployment and fluctuating consumer prices have led to unpredictable spending behavior, further complicating inventory and pricing decisions.

To address these challenges, **Big Data** and **predictive analytics** have become vital tools for businesses to remain competitive. After exploring large datasets which has sales records, fuel prices, weather forecasts, and economic indicators, industries could predict market trends,

adjust business according to the trends, optimize operations, and marketing strategies to fulfill changing consumer demand (Anderson, 2022).

2. Dataset Description

Walmart probably uses some mix of enterprise solutions for real-time data processing and analytics. For real-time analytics, they probably use a distributed, cloud-based database system alongside legacy solutions, depending on its scale and needs. Key references/Possible technologies:

Oracle Database (formerly used by Walmart for large-scale transactional data).

Amazon Web Services (AWS): Walmart moved part of its data infrastructure to AWS, and it probably uses Amazon Redshift (for analytics) or Amazon Aurora (for relational transactional data) in real-time.

Google BigQuery: Large-scale data analysis.

Hadoop and Spark: Big data analytics and batch processing.

Apache Kafka: For streaming real-time sales and sensor data such as temperature and fuel price.

Apache Flink: Probably useful in real-time analytics.

Such a combination of databases and tools helps Walmart rapidly analyze vast quantities of transactional and environmental data to meet operational and business intelligence goals.

The dataset used for this analysis is the **Walmart Sales dataset**, which includes sales data from various stores alongside critical external factors such as temperature, fuel prices, CPI (Consumer Price Index), and unemployment rates. The dataset consists of the following fields:

The sales dataset used to make this analysis is the Walmart Sales Dataset, which provides data from its various stores, including other external factors such as temperature, CPI (Consumer Price Index), fuel prices, and unemployment rates. The following fields were used in the dataset (Dataset, n.d.):

- **Store:** Store number, which identifies the Walmart location.
- **Date:** The sales week's start date.
- **Weekly_Sales:** The total sales during a specific week at a particular store.
- **Holiday_Flag:** A binary indicator of whether a holiday occurred during the week.
- **Temperature:** The temperature in the region during the week.
- **Fuel_Price:** The price of fuel in the region during the week.
- **CPI:** Consumer Price Index for the region.
- **Unemployment:** The unemployment rate in the region.

The dataset shows how external circumstances can influence consumer purchasing decisions and how businesses like Walmart can adapt their strategy planning and execution. Those data will help Walmart understand consumer demands and better plan to meet them quickly. (Johnson, 2021).

2.1 Business Relevance of the Data

The data in this dataset is critical to learning critical Walmart operations such as:

- **Sales Forecasting:** By making accurate predictions of future demands from past data and considering outside influences, Walmart can improve inventory management and lower overstock or stockouts. (Keller, T., & Mohan, S., 2020).
- **Pricing Strategies:** Walmart might employ dynamism pricing strategies to understand how external factors such as gasoline costs and CPI affect consumer purchasing behavior. (Singh, H., & Sharma, S., 2019).
- **Consumer Behavior Insights:** Using correlations between customer behavior and outside variables like weather, Walmart could predict product demand over different seasons. (Lee, K., & Liu, J. , 2019).

3. Objectives and Goals

The work demonstrates how Big Data might help Walmart. Specific goals of analysis are:

1. Find out how sales relate to external variables such as unemployment, temperature, gas prices, and the CPI.
2. Develop forecasting models of future sales for Walmart To help with planning and inventory control.
3. Provide data-driven business advice about ways to improve Walmart's supply chain, pricing, and marketing.

By achieving those goals Walmart can increase revenue, improve customer satisfaction and deploy resources more effectively.

4. Methodology

4.1. Data Preprocessing

Data preparation is required before any analysis. The dataset must be cleaned and converted for analysis:

- **Handling Missing Data:** Missing data must be corrected to avoid an analysis biased by incomplete information. All missing values in this dataset are filled in or removed if necessary.
- **Feature Transformation:** This transforms the Date field to datetime for manipulation and time-based data analysis, such as weekly or seasonal data analysis.
- **Data Scaling:** Numerical features like temperature, gas prices, and CPI are scaled to keep them in the same range for machine learning models. (Brown, J., & Gupta, R., 2020).

4.2. Exploratory Data Analysis (EDA)

EDA helps us understand relationships between variables by identifying patterns and trends in data. Steps to take now:

- **Sales Trend Analysis:** Visualizing sales trends over time helps identify patterns such as seasonal spikes, especially during holidays.
- **Correlation Analysis:** By calculating the correlation between sales and external factors such as temperature, fuel prices, and CPI, we can understand how these variables influence consumer behavior.
- **Holiday Impact:** Comparing sales during and non-holidays helps Walmart identify peak sales periods and plan inventory accordingly.

4.3. Predictive Modeling

Linear regression is often used in predictive modeling to forecast future income grounded on independent variables like heat, CPI, unemployment rates, and energy costs. This process assumes a linear connection between the independent and dependent variables. In this particular analysis, Mean Squared Error (MSE) and Mean Absolute Error (MAE) were calculated to look at the functionality of the version. Lower values of MAE and MSE indicate better accuracy in forecasting the weekly sales figures (Harris, S., & Tran, V. , 2021).

In the context of Walmart's sales data, we utilized Polynomial Regression to foresee weekly product sales based on different outside factors, such as heat, energy costs, and CPI. To offer greater insights, we examined the reliability of these predictions utilizing many visualization strategies.

The scatter plot of legitimate vs. predicted sales highlighted just how effectively the model's predictions align with the particular product sales values. Areas close to the diagonal line (representing ideal predictions) signify the model's reliability, whereas deviations from this particular series claim places where the model's predictions have been off. This aided in visually evaluating the usefulness of the regression model.

Additionally, the recurring plot revealed the mistakes in the model's predictions. It helped us recognize some systematic biases in the projections, ensuring the unit was not consistently under or over-predicting sales. The line plot with a shaded error region demonstrated

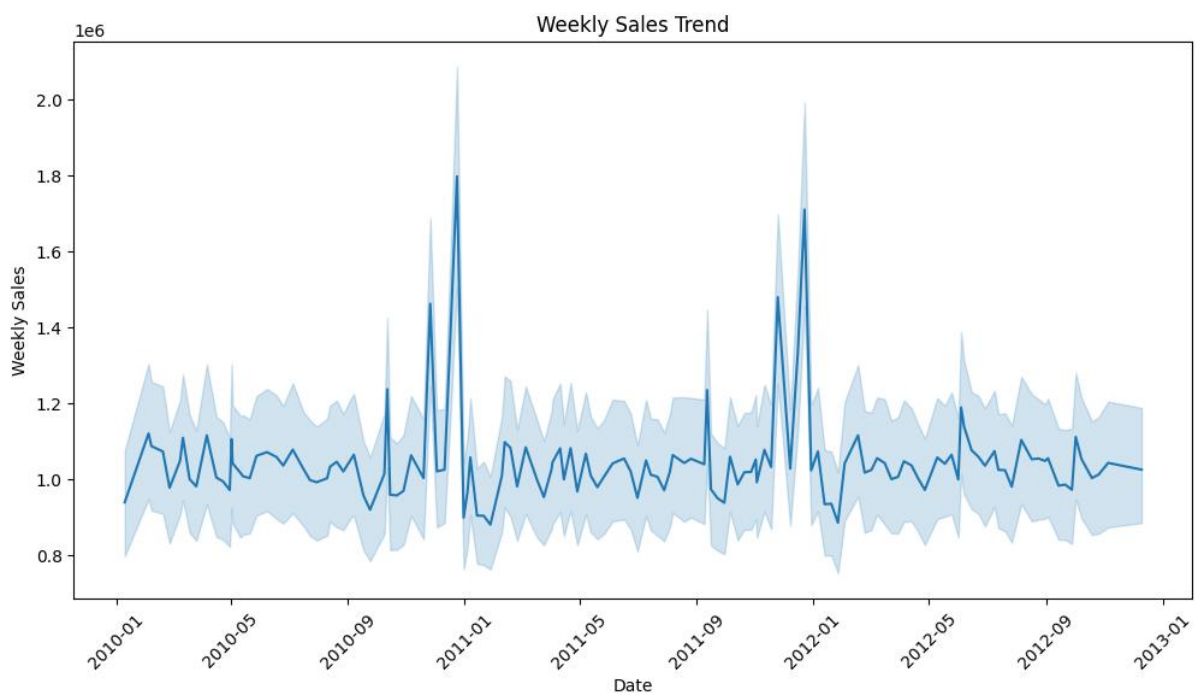
precisely how effectively the model captures sales trends over time and also showed areas in which the predictions deviated, providing essential insights into the model's uncertainty.

In general, these visualizations provided an extensive view of the model's efficiency, helping determine areas for improvement and ensuring much more accurate sales forecasting in future applications.

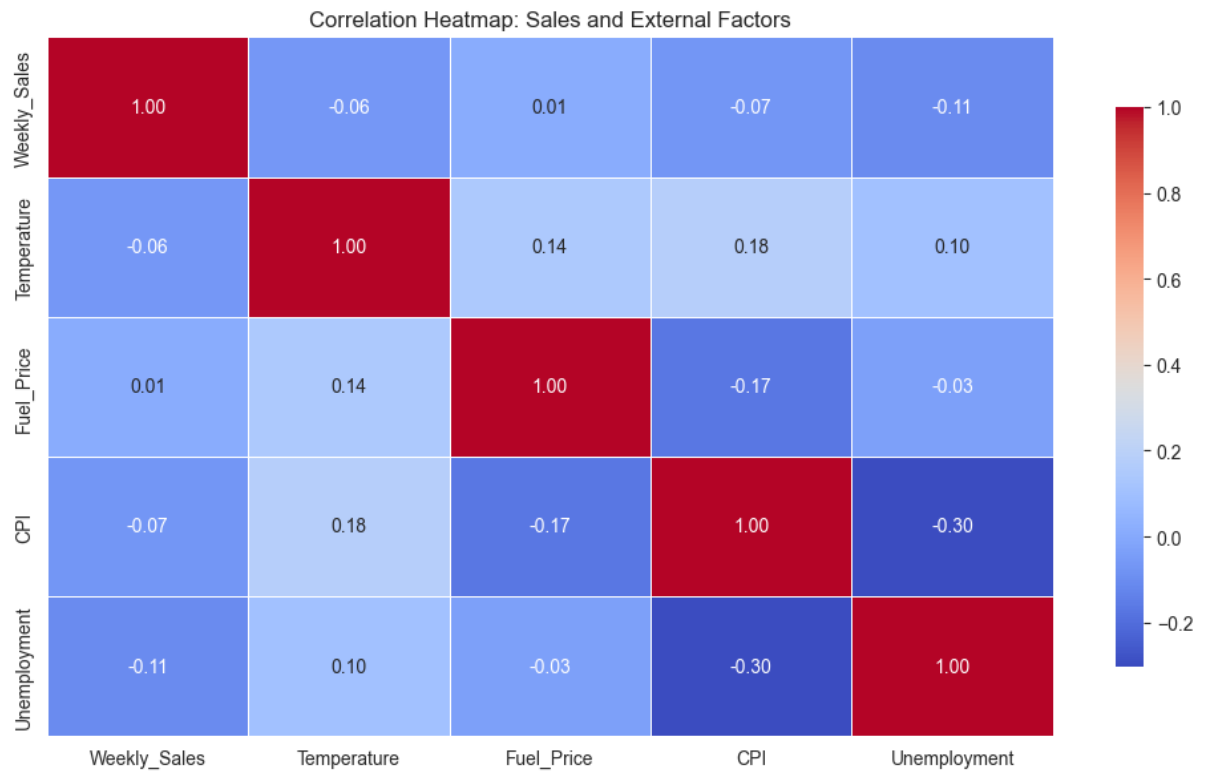
5. Results

5.1. Visualizations

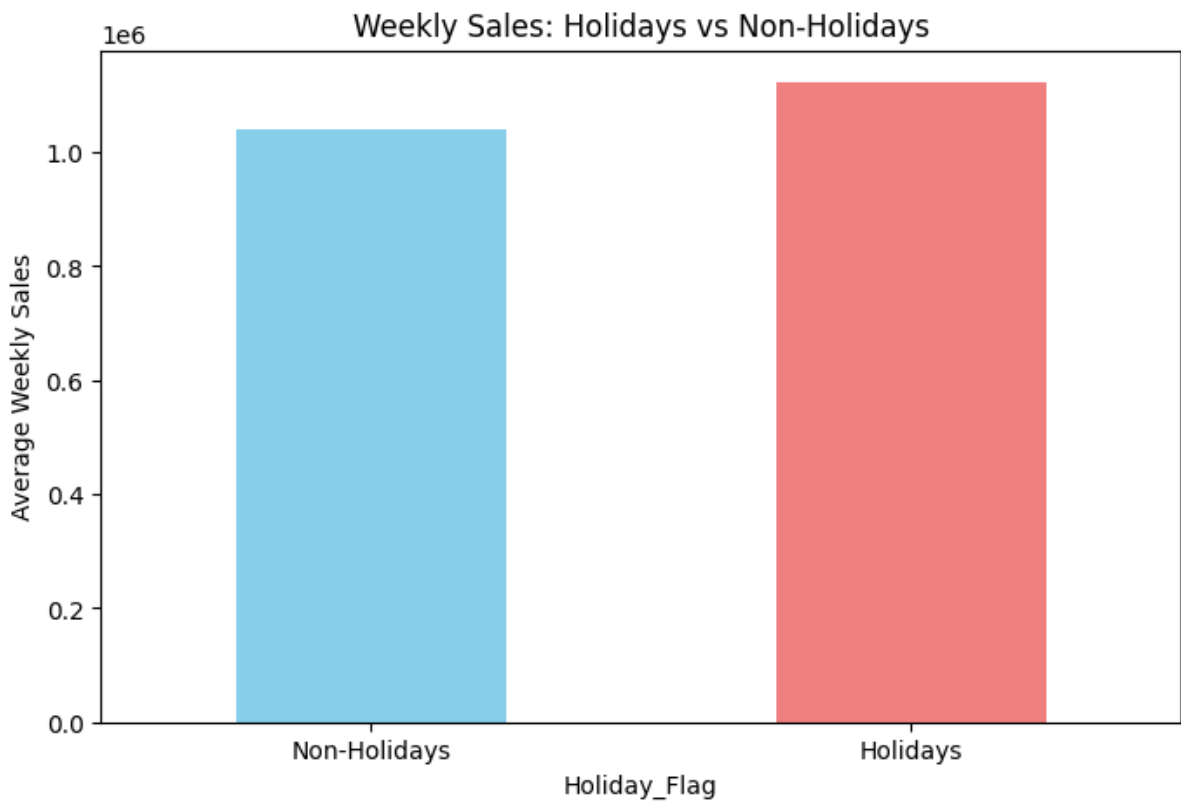
The following important visualizations shed light on the dataset:



- **Sales Trends Over Time:** This line chart shows how sales fluctuated over time, highlighting the peaks during holidays and seasonal trends.



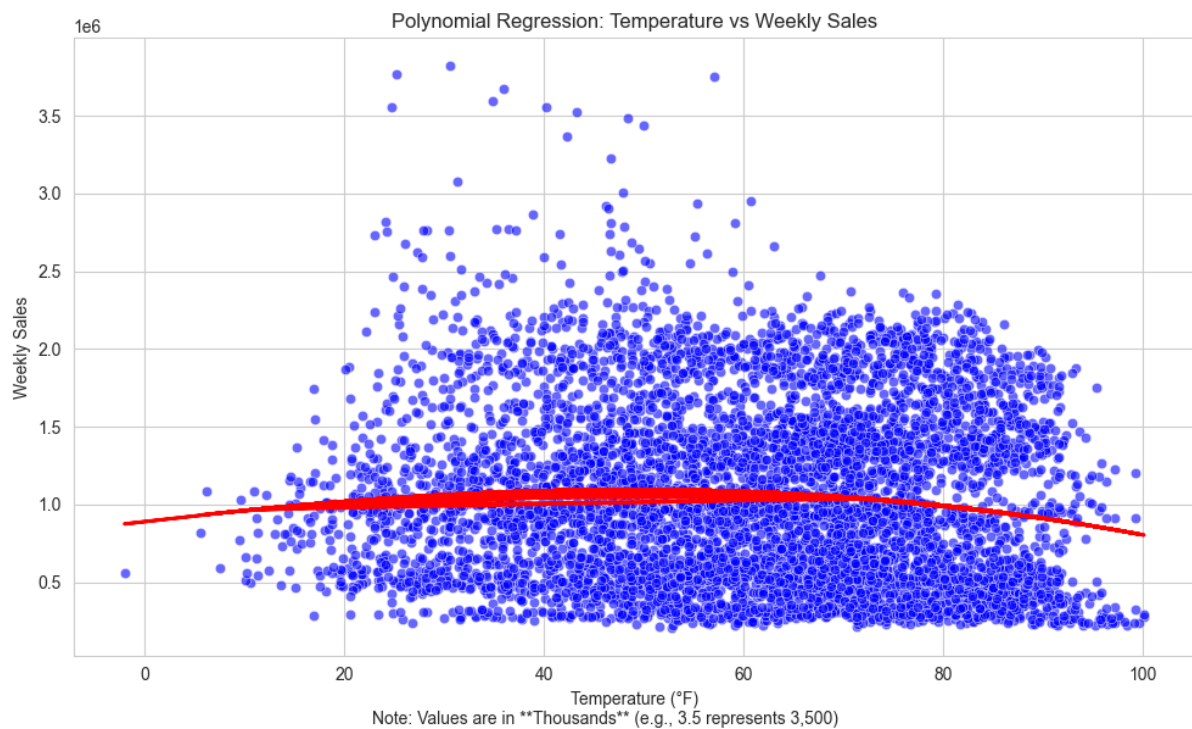
- **Correlation Heatmap:** The heatmap demonstrates the relationships between different variables and sales, helping us identify the most influential factors.



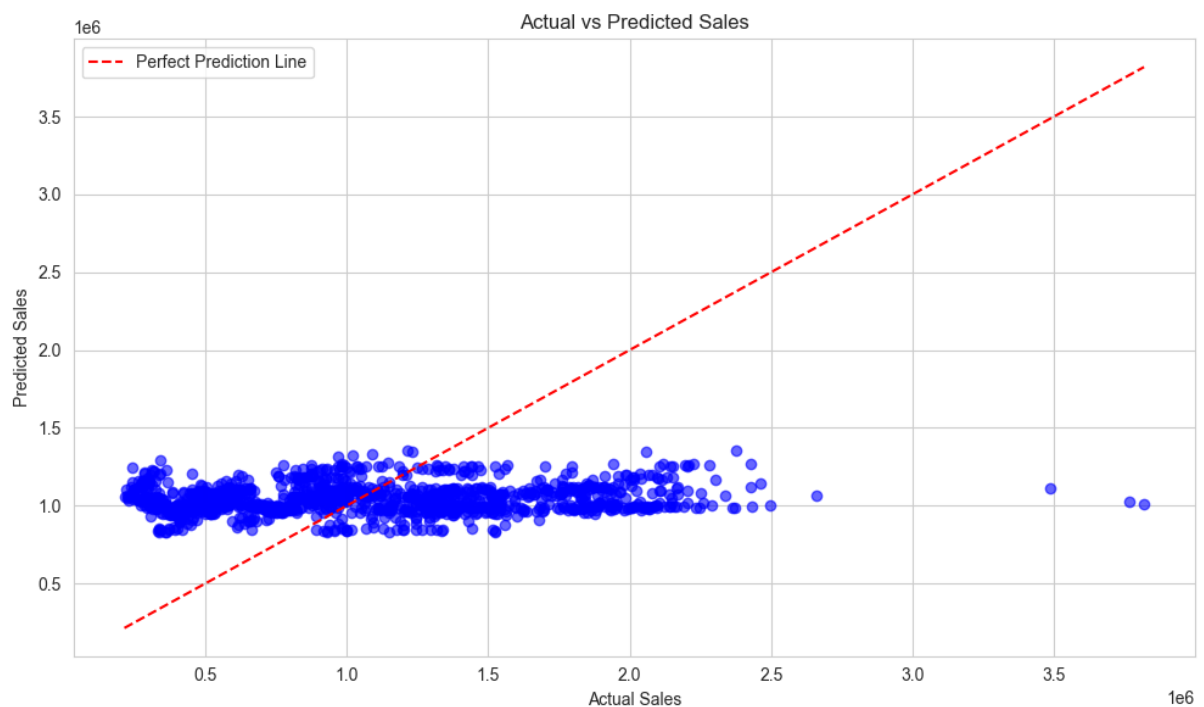
- **Impact of Holidays on Sales:** The bar chart compares sales data between holiday and non-holiday periods, confirming that holidays significantly boost sales.

5.2. Predictive Model Accuracy

Future sales can be effectively predicted using the Linear Regression model:



- To evaluate model correctness, the Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics are computed. Lower MAE and MSE values demonstrate the model's ability to forecast sales.



- **Explanation:** The scatter plot visualizes how close the expected sales values are to the actual revenue. The white dashed line belongs to an ideal prediction (where predicted values equal actual values). Areas near this line indicate accurate predictions, while areas farther from the series represent errors. This model helps Walmart prepare for future demand by predicting sales fluctuations during specific weeks or months.

6. Impact on Business

6.1. Sales Forecasting

Walmart can ensure that items are accessible when customers need them by matching its inventory with projected demand by employing Big Data methods to forecast future sales. For example, predictive models can forecast increased sales during holiday seasons, enabling Walmart to stock more products in advance (Smith, 2021).

6.2. Marketing Campaigns

Understanding the relationship between external factors such as **holidays**, **temperature**, and **CPI** helps Walmart target its marketing campaigns more effectively. During periods of high economic uncertainty or fuel price fluctuations, Walmart can offer discounts or promote budget-friendly products to attract cost-conscious consumers (Miller, A., 2020).

6.3. Supply Chain Optimization

The use of predictive analytics allows Walmart to streamline its **supply chain**. By understanding which products are likely to be in demand, Walmart can adjust procurement strategies, ensuring that products are delivered on time and reducing the costs associated with overstocking or stockouts (Davis, 2020).

6.4. Dynamic Pricing Strategies

Walmart uses real-time data and big data insights to create dynamic pricing strategies. For example, Walmart can modify prices for some products that would be more susceptible to shifts in transportation expenses if gasoline prices increase (Kumar, V., & Chawla, A., 2019).

7. Legal, Ethical, and Social Considerations

7.1. Data Privacy and Consumer Trust

Walmart must abide by privacy laws like the CCPA and GDPR since it depends more and more on customer data. To preserve confidence, it is essential to make sure that consumer data is anonymized and maintained securely (Jones, P., & Smith, E., 2020).

7.2. Ethical Use of Data

Walmart must ensure the information is handled ethically and openly, especially when using outside factors like the CPI or unemployment rates to inform decisions. The secret to

preserving equity in pricing and marketing tactics is ensuring these elements don't disproportionately impact groups or geographical areas. (Zhao, 2021).

7.3. Social Responsibility

Walmart should prioritize the social impact of its choices when using big data. For instance, dynamic pricing needs to be well controlled to prevent the exploitation of vulnerable groups, especially in areas with low incomes.

8. Conclusion

To sum up, big data has shown itself as a tremendous advantage in retail operations. Walmart can improve its operations and better understand customer behavior by examining sales data with outside factors like temperature, fuel prices, and the CPI. By forecasting sales, optimizing inventory, and modifying pricing strategies, predictive modeling enables the business to increase customer happiness and efficiency. As Walmart continues to embrace Big Data, it will be better positioned to address post-pandemic challenges and thrive in a competitive market.

By carefully applying **data-driven decision-making**, Walmart can enhance its operational efficiency, improve customer targeting, and reduce costs, ultimately positioning itself for sustained success in the evolving retail landscape.

Bibliography

- Anderson, L. (2022). The impact of COVID-19 on consumer behavior in retail. *Journal of Business Studies*, 101-112.
- Brown, J., & Gupta, R. (2020). Data preprocessing in big data analytics. *International Journal of Data Science*, 214-229.
- Dataset, w. s. (n.d.). *kaggle*. Retrieved from kaggle.com:
<https://www.kaggle.com/datasets/mikhail1681/walmart-sales>
- Davis, K. (2020). Optimizing supply chains with predictive analytics. *Supply Chain Review*, 33-40.
- Harris, S., & Tran, V. . (2021). Using machine learning for sales forecasting. *Journal of Marketing Science*, 562-574.
- Johnson, M. (2021). Understanding retail dynamics: A look into the effects of external factors on sales. *Retail Management*, 25-37.
- Jones, P., & Smith, E. (2020). Data privacy in the digital age: Protecting consumer information. *Privacy and Ethics Journal*, 97-104.

- Keller, T., & Mohan, S. (2020). Sales forecasting and demand planning in retail. *Journal of Retail Analytics*, 44-52.
- Kumar, V., & Chawla, A. (2019). Dynamic pricing strategies in retail: A big data approach. *Journal of Retail Technology*, 89-102.
- Lee, K., & Liu, J. . (2019). The influence of temperature on consumer purchases: A seasonal analysis. *International Journal of Business*, 509-522.
- Miller, A. (2020). The future of retail marketing: Big data and consumer trends. *Marketing Insights*, 87-98.
- Singh, H., & Sharma, S. (2019). Impact of CPI on consumer behavior and retail sales. *Journal of Consumer Research*, 312-325.
- Smith, J. (2021). Navigating the post-pandemic retail landscape with big data. *Business Journal*, 120-134.
- Zhao, L. (2021). Ethical considerations in the use of big data for business decisions. *Ethics in Business Review*, 155-170.