**Ms in Data Science**

ITC6004A1 - Data Visualization - Winter Term 2024

**Final Project:**

*“Walmart Promotional Strategy Pitch”*

**Instructor:**

Dr. K. Mploutsos

**A logo with text on it

Description automatically generated**

**Submitted to:**

Deree – The American College of Greece

**Submitted by:**Kapsalis C. (275482)

Kounelis P. (260963)

Kritikakis N. (279114)

Stavrogiannis C. (225765)

**Submission Date:**

01/04/2023

**Word count:** 5123

**Table of Content**

[Table of Figures](#_Toc162876268)

[Abbreviations](#_Toc162876269)

[Executive Summary 1](#_Toc162876270)

[1. Analysis of the Case & the Dataset 2](#_Toc162876271)

[1.1. Scenario and Role 2](#_Toc162876272)

[1.2. The Goal 3](#_Toc162876273)

[1.2.1. C-Level Analysis 3](#_Toc162876274)

[1.2.2. Current Situation 4](#_Toc162876275)

[1.2.3. Socio-economic Level Analysis 4](#_Toc162876276)

[1.2.4. Store Level Analysis 5](#_Toc162876277)

[1.2.5. Correlations 6](#_Toc162876278)

[2. Data Analysis & Preprocessing 7](#_Toc162876279)

[3. Results & Analysis 13](#_Toc162876280)

[3.1. C-Level Analysis 13](#_Toc162876281)

[3.2. Current Situation 14](#_Toc162876282)

[3.3. Socio-Economic Analysis 18](#_Toc162876283)

[3.4. Store Level Analysis 21](#_Toc162876284)

[3.5. Correlation/Regression Analysis 22](#_Toc162876285)

[Conclusion 24](#_Toc162876286)

[References 25](#_Toc162876287)

# **Table of Figures**

[Figure 1:Boxplot displaying the distribution of the numerical variable(s). 9](#_Toc162876126)

[Figure 2:Date Reference structure. 10](#_Toc162876127)

[Figure 3:Overview of dataset star schema. 11](#_Toc162876128)

[Figure 4:Indicative additional columns in the "features" dataset. 11](#_Toc162876129)

[Figure 5:Indicative additional columns in the "sales" dataset. 12](#_Toc162876130)

[Figure 6:Indicative additional column holding the size-bin category for each store. 12](#_Toc162876131)

[Figure 7:Overview of store types and sales. 13](#_Toc162876132)

[Figure 8:Average promotional expenditures vs promotional cap. 14](#_Toc162876133)

[Figure 9:Average promotional expenditures vs promotional cap (high). 14](#_Toc162876134)

[Figure 10:Sliders used in current situation slide. 15](#_Toc162876135)

[Figure 11: Weekly sales over time. 15](#_Toc162876136)

[Figure 12: Major KPIs identified. 16](#_Toc162876137)

[Figure 13:weekly sales and promo for type A stores. 16](#_Toc162876138)

[Figure 14:Promotional expenditures and weekly sales for type B stores. 17](#_Toc162876139)

[Figure 15:Promotional expenditures and weekly sales for type C stores. 18](#_Toc162876140)

[Figure 16:Average unemployment and temperature by store type. 19](#_Toc162876141)

[Figure 17: Major KPIs for socio-economic features. 19](#_Toc162876142)

[Figure 18:Average unemployment and CPI through time. 20](#_Toc162876143)

[Figure 19: Average unemployment for each type of store and sales bin. 20](#_Toc162876144)

[Figure 20: General analysis of stores regarding their performance and promotional expenditures. 21](#_Toc162876145)

[Figure 21:Decomposition tree for average weekly sales. 22](#_Toc162876146)

[Figure 22:Intertemporal Promo Expenditures & Total Weekly Sales. 22](#_Toc162876147)

[Figure 23:Total Promo Expenditure & Total Weekly Sales. 23](#_Toc162876148)

# **Abbreviations**

BI……………………………………………………………………………………………….Business Intelligence

CPI…………………………………………………………………………………………….Consumer Price Index

KPIs……………………………………………………………………………………………Key Performance Indicators

USA……………………………………………………………………………………………United States of America

# **Executive Summary**

The realm of retail is a fast-evolving domain in which data analytics and their consequent visualization and most importantly their interpretation is of utmost significance while trying to understand and keep up with the ever-growing needs of the market. Consumer behavior is strongly linked with economic trends and sales numbers (Seetharaman et al., 2016).

Therefore, is safe to suggest that data serves as the cornerstone for informed business strategies and under this assumption successful retails strategies depend heavily on detailed analytics for ultimately providing better strategic decision-making and achieving favorable sales outcomes (Chandramana, 2017). Retail data analysis constitutes a complex effort, delving into various factors that intricately influence the financial outcomes of companies.

In this example we are going to delve into the analysis of such a retail related dataset (Kaggle, 2017). Within this context, the focus lies on a dataset covering 45 Walmart stores scattered across diverse regions of the United States.

Moreover, the dataset covers and displays a plethora of strategic decisions, such as marketing campaigns, undertaken by Walmart during the specified timeframe. These initiatives, aimed at enhancing the shopping experience and driving sales growth, present valuable insights into the relationship between business strategies and their consequence on consumer behavior.

Furthermore, major events and strategic decisions, essential to the operational landscape of retailers, add another layer of complexity to our analysis. These events carry the potential to significantly influence consumer behavior and financial performance. By examining the relationship between sales and changes in holiday schedules, major events, strategic initiatives, and promotions, we will create visualizations and provide a detailed analysis to accurately estimate sales dynamics and trends.

Finally, this analysis is instrumental in empowering stakeholders with actionable intelligence, in order to facilitate informed decision-making and supporting sustainable growth.

The above analysis will be presented in detail in the below chapters namely: a) Analysis of the case and dataset including our goal for this report file b) Data analysis and preprocessing c) Results and a detailed analysis explaining our findings.

# **Analysis of the Case & the Dataset**

The data presented here concerns the Walmart company and their retail for 45 stores scattered across the USA. Each store is anonymized and as such is identified solely by an assigned ID, withholding further specific information regarding their locations or operational departments.

Dominant among these variables are holidays, major events, and strategic initiatives, alongside the impactful realm of promotional activities, or markdowns, which directly impact sales figures.

The dataset highlights four prominent holidays observed in the United States: the Super Bowl, Labor Day, Thanksgiving, and Christmas. These occasions, deeply ingrained in American culture, not only mark significant cultural and sporting events but also wield considerable influence on consumer spending patterns. The dates of these holidays over a four-year period, from 2010 to 2012, are meticulously documented, reflecting their recurrent impact on retail sales.

Understanding the significance of holidays and promotional actions and their specific impacts on consumer spending patterns is crucial for devising effective retail strategies. For instance, the Super Bowl, taking place annually in early February, is renowned for its association with high levels of consumer spending on food, beverages, and electronics, as viewers gather to watch the championship game.

## **Scenario and Role**

**Scenario Overview**

Within the retail landscape, particularly for a giant like Walmart, understanding sales dynamics across multiple stores and departments is paramount. The challenge lies in leveraging historical sales data, store characteristics, and external factors such as seasonal trends, holidays, and promotional markdowns to predict future sales accurately. This is further compounded by the sporadic nature of markdowns and the amplified significance of holiday weeks. The dataset provided offers a rich source of information, spanning sales records, store attributes, promotional activities, and economic indicators. The goal is to forecast sales and to discern patterns, uncover insights, and formulate actionable strategies that capitalize on opportunities and mitigate risks. Furthermore, having as a specific focus to analyze sales on a per-store, per-year basis, especially during non-holiday weeks, the objective becomes clearer.

By concentrating on both non-holiday and holiday weeks, we aim to understand the baseline sales performance of each store throughout the year, discussing at the same time the heightened sales activity associated with holiday periods.

**Role Overview**

We have identified our role as an external contractor for Walmart offering data analysis and insights on data provided to us. Our primary role involves analyzing weekly sales data, with a strong focus on understanding the significance and purpose of markdowns, which encompass promotional events and expenditures. Our aim is to assess their impact on sales performance and uncover broader trends related to seasonality, such as holiday weeks, as well as other features specific to stores and the company, along with socio-economic factors.

This approach is crucial for addressing the ambiguity and anonymization in our dataset, particularly regarding missing specifics and details concerning markdowns or store types. In this multifaceted context, our analysis plays a critical role in providing clarity and insights.

Our analysis focuses on identifying the key drivers and inhibitors of weekly sales, aiming to optimize business outcomes beyond just seasonal fluctuations. We aim to uncover the underlying, longer-term trends and disparities within our nationwide results to understand the factors that consistently influence sales performance and their interrelationships.

## **The Goal**

Walmart's extensive network of large stores, offering a range of products tailored to the needs of middle to lower-income demographics across the nation. We'll examine dataset to reveal valuable insights for decision-making. Understanding the impact of promotional markdowns and holiday weeks on sales helps identify departments and stores vulnerable to demand fluctuations. This allows for targeted promotions, efficient resource allocation, and optimized inventory management. Secondly, correlating sales with external factors like temperature, fuel prices, CPI, and unemployment rates uncovers seasonal trends and consumer behavior, aiding proactive decision-making.

Below a breakdown of our analysis is presented divided into easier-to-comprehend sections.

### **C-Level Analysis**

Executive dashboards (C-Level) are customized to cover the needs of a particular audience. In a survey conducted by EY, it was found that 81% of businesses advocate for data to be the focal point of decision-making. However, the report reveals that merely 31% of companies have taken significant steps to effectively restructure their operations accordingly (Keshwani, 2023).

Therefore, C-Level dashboards should serve as reporting tools able to offer insights on crucial metrics and Key Performance Indicators (KPIs) while offering a high-level overview reflecting the operational activities within an organization (Pratt, 2023).

In our case, our goal is to analyze Walmart’s results in the given time period (5/Feb/2010 – 26/Oct/2012) in correspondence to its promotional strategies across stores. The information we deemed as most relevant for a high-level overview mainly included aggregations revolving around sales and promotional expenditure. This information, broken down by the main characteristics of Walmart’s stores and the capacity of implemented promotional campaigns, offers a simple introductory overview on what sales revenue and promotional expenditure has been registered so far across stores of different types and sizes, but also where the promotional priority has been concentrated.

### **Current Situation**

Our primary objective is to ascertain which types of stores yield optimal sales results and under what magnitude of promotional expenditures. By examining sales performance across various store types and sizes, we can identify KPIs and correlations that inform resource allocation strategies and immediate next steps. Additionally, we will explore how factors such as promotional events and holiday seasons impact sales performance. We will consider the previous in two ways:

First, we will investigate the intertemporal evolution of the values of our most valuable metrics, mainly weekly sales, and total promotional expenditure. Showing this in parallel to the evolution of other numeric features in the dataset (namely temperature, CPI, unemployment, and fuel price), we will get a hint at the potential existence of correlations between variables, which can inform resource allocation strategies and immediate next steps.

We will also examine the joint distribution of sales and promotional expenditure, in light of the specified weekly promotional spending cap, to uncover any room for improvement or decision-making shortfalls that have happened in the past. To facilitate this analysis, we have implemented preprocessing steps to categorize stores into bins based on their total proportional expenses and sizes. This approach allows us to analyze sales performance within specific store categories and sizes comprehensively.

This analysis, based on the use of a ‘magic sextant’ visual, will shed light on which store types are contributing more to our overall sales and under what expenditure, enabling us to prioritize strategic initiatives accordingly.

### **Socio-economic Level Analysis**

While most of our analysis will incorporate a comprehensive breakdown and examination of the correlations between sales and the features of our retail points our primary focus is understanding how two key (as identified initially) features (Holiday periods, markdown(s)) interacts and correlate with weekly sales, aiming to discern any significant correlations that can provide insights into consumer behavior and purchasing patterns.

However, examining the correlation between socio-economic factors such as unemployment rates, CPI and Temperature offers valuable insights into consumer behavior and purchasing patterns within different economic contexts. Our effort in this part is to examine in detail the secondary (as they deemed initially) features of our *features dataset*.

We believe that this will provide valuable insights able to identify patterns which may suggest a different approach on how to handle the current situation and may provide us with the ability to propose solutions and foresights for the future planning and actions of Walmart. Additionally, factors such as Walmart's competitive pricing strategies may mitigate the negative impact of unemployment on sales to some extent.

Therefore, we expect this socio-economic analysis to offer a better understanding on the non-detailed part of our dataset and sed some light on regional information and store details and understanding of our store dynamics and their importance for the weekly sales.

Moreover, it is anticipated that stores situated in regions with higher unemployment rates may experience lower sales volumes due to reduced disposable income and decreased consumer confidence. However, exceptions to this trend may arise, particularly considering Walmart's market positioning as a non-premium supermarket known for offering value-priced goods.

### **Store Level Analysis**

While marketing campaigns play a significant role in influencing consumer behavior and purchase decisions the physical stores are the points of interest. Therefore, we will examine and analyze their features.

We will examine the relationship between store types, size and other store characteristics and the weekly sales. We will identify which types are most profitable and in which types the promotional expenditures are higher.

Moreover, it is important to identify the impact of promotional expenditures by comparing both actual numbers and also proportional analysis.

Finally, we will provide insights into the importance of the size of each store and their correlation with sales and store type. Additionally, we will use our preprocessed dataset and additional features (i.e., bins created for store sizes) to clarify which store related factors and features play the most pivotal role on the weekly sales.

Expanding on our analysis, we also recognize the significance of considering regional variations and demographic factors in understanding store performance. In conclusion, our examination of store characteristics aims to provide actionable insights for enhancing sales performance and driving sustainable growth in Walmart's retail operations.

### **Correlations**

Our analysis will incorporate a comprehensive breakdown and examination of the correlation and regression between sales and the features of our retail points as outlined in our feature sheet. A primary focus will be on understanding how the key feature (markdown(s)) interacts and correlates with weekly sales, aiming to discern any significant insights into consumer behavior and purchasing patterns.

Markdowns play an important role in our analysis since they also constitute a standardized expenditure from our part that needs to revert back to improved sales. To gain a deeper understanding of their effectiveness while knowing that we do not have specific details on what each markdown exactly refers to, we have decided to enrich our dataset and aggregate the total markdown expenditure into a single promotional expenditure feature. This approach enables us to evaluate the success of our promotional efforts in each store type and assess the return on investment from these expenditures. In analyzing markdowns, we will consider both absolute values and proportional results to account for any discrepancies between total promotional expenditures and changes in weekly sales between different types of stores. This approach provides a comprehensive understanding of the effectiveness of our promotional strategies across different store types and regions and most importantly their correlation with weekly sales and finally which markdown and store type is more successful.

# **Data Analysis & Preprocessing**

The essence of this analytical endeavor lies in ensuring the integrity, reliability, and quality of the dataset (Wang & Strong, 1996).

To ensure a clean dataset devoid of null values and outliers that could compromise the analysis, meticulous preprocessing is conducted. This process aims to eliminate anomalies and inconsistencies, establishing a sturdy foundation for robust analysis in line with best practices and data quality assurance standards. Missing values, outliers, and duplicates are meticulously handled to bolster the dataset's integrity and mitigate inaccuracies and biases.

Our major features per Sheet as they appear in our dataset are:

|  |  |  |
| --- | --- | --- |
| **Sheet** | **Feature** | **Type** |
| **Store** | Size | Numeric |
| Type | Categorical |
| **Features** | Date | Datetime |
| Fuel\_price | Currency($) |
| IsHoliday | Boolean |
| Temperature | Temperature |
| Markdown (1-5) | Currency ($) |
| CPI | Currency ($) |
| Unemployment | Precentage |
| **Sales** | Weekly\_sales | Currency ($) |
| Date | Datetime |

The initial data files broke down weekly sales to single departments that are operating in various stores. However, we decided to drop that column of the ‘sales’ table, as the markdown-related data which represent Walmart’s promotional strategy so far are given on the store level, so no department-level insights could be drawn.

We have chosen to procced with the below data preprocessing and preparation in order not only have a more meaningful dataset but also to adhere with the common data quality guidelines which involves the below aspects:

* Completeness
* Uniqueness
* Timeliness
* Integrity
* Accuracy
* Conformity
* Consistency.

More specifically the data have been preprocessed in the below manner:

**Missing values:**

Calculating the advantages of common imputation techniques (focusing on numerical values) such as mean or median imputation it was decided to use the *multiple imputation technique using predictive mean matching* where “final imputed values are constrained to previously observed values in the dataset” (Shevchenko, 2021). This technique is highly used in datasets where distributions should be preserved. (Akmam et al., 2019). This careful approach ensures the integrity of the dataset and underpins the validity of subsequent analyses.

**Outlier detection and handling:**

Three files have been imported and ensured that all data is correctly represented in the appropriate data types. Additionally, two calculated columns have been created: sales-outlier, which identifies outliers in weekly sales for specific departments and stores within a given year, and features-outlier, which flags outliers in markdown values, Consumer Price Index (CPI), or unemployment rates for individual stores across three years. These boolean columns are applied as filters throughout the report, facilitating the exclusion of outlier values during regression analysis.

Although the distribution of store sizes appears non-normal, it exhibits a degree of symmetry with a concentration of stores around the median size. While there are outliers, the IQR method, which analyzes boxplot representations, doesn't detect any outliers for store sizes in the dataset.

Regarding the weekly sales dataset, it's recognized that sales performances vary between stores, potentially reflecting differences in magnitude and occasional extraordinary sales figures. However, such variations are inherent in business operations and shouldn't be indiscriminately treated as outliers.

Instead, the aim is to analyze sales on a per-store, per-year basis, particularly focusing on non-holiday weeks to provide a more granular perspective. This approach ensures that conclusions are grounded in an understanding of the underlying business dynamics.

**A graph with a blue rectangle

Description automatically generated**

Figure 1:Boxplot displaying the distribution of the numerical variable(s).

**Segmentation Analysis:**

We need to separate the data based on spatially driven information as they appear in our dataset (i.e., Temperature, Unemployment, Fuel\_Size, CPI). We will try to inspect the data and identify patterns that could be based on the location and similar socio-economic factors would be systematically different between segments and very similar within single segments. The 'Type' attribute within the stores dataset presents an opportunity for detailed segmentation analysis. This segmentation will allow stakeholders to tailor strategies and allocate resources effectively across store segments.

**Temporal Dynamics Analysis:**

Temporal analysis serves as a cornerstone for identifying seasonal trends and/or recurring patterns within the dataset (Meulmeester, 2022). This is important since it empowers stakeholders to predict and leverage seasonal fluctuations in consumer behavior (Sopha et al., 2018). In our case and considering the above in detail we identify notable holidays which are particularly prominent within the USA.

These include:

* Super Bowl on February 12th, 2010; February 11th, 2011; February 10th, 2012.
* Labor Day on September 10th, 2010; September 9th, 2011; September 7th, 2012.
* Thanksgiving falls on November 26th, 2010; November 25th, 2011.
* Christmas is celebrated on December 31st, 2010; December 30th, 2011.

Time-series analysis techniques such as seasonal decomposition, in conjunction with a dedicated *Date Reference* table (figure 2), enable comprehension of temporal dynamics. This can reveal intuitions related to the event(s) (i.e., weekly sales) frequency and consistency and subsequently allows stakeholders to anticipate and capitalize on seasonal fluctuations in consumer behavior.

A screenshot of a computer

Description automatically generated

Figure 2:Date Reference structure.

**Implementation of Star Schema:**

The adoption of a star schema architecture facilitates efficient data modeling and analysis, enabling seamless navigation through complex datasets (figure 3):

1. Fact Table - 'Sales':
   * At the heart of our analytical framework lies the 'Sales' fact table, which aggregates weekly sales data across departments and stores. This centralized *repository* of sales provides an all-inclusive view of sales performance.
2. Dimension Tables - 'Stores' and 'Features':
   * Dimension tables, including 'Stores' and 'Features', enrich our understanding by incorporating contextual attributes and socio-economic factors. These dimensions enhance the analysis performed, by providing additional layers of granularity.
3. Date Reference Table:
   * This serves as the cornerstone of temporal analysis, facilitating dynamic filtering, trend identification, and seasonality decomposition.

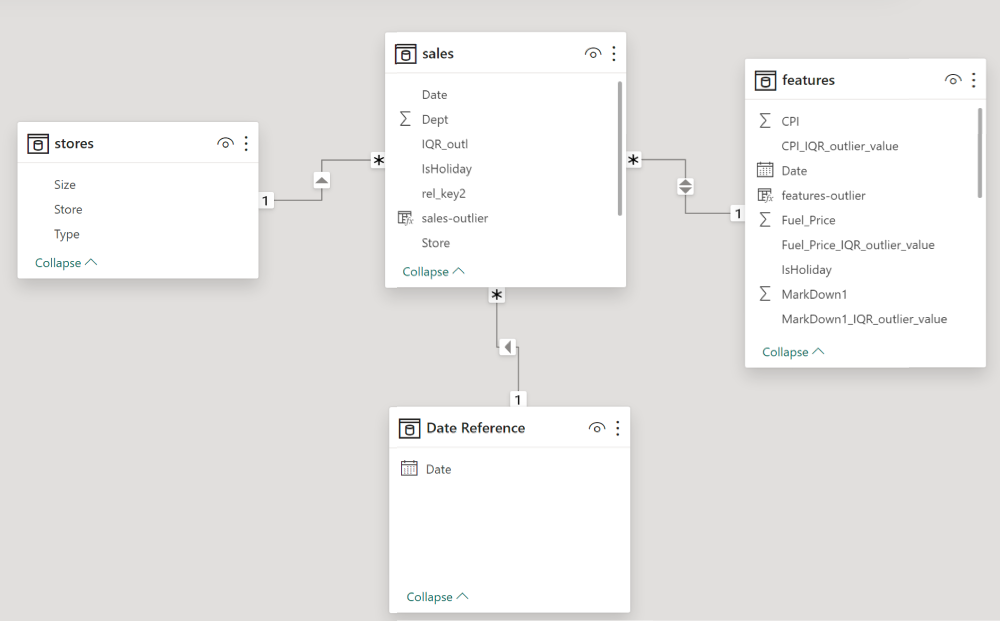


Figure 3:Overview of dataset star schema.

**Enriching the Dataset:**

To enhance the analytical depth and robustness of the dataset, newly derived columns are introduced for outlier detection and analysis refinement (figures 4, 5, 6).

1. Derived Columns for Outlier Detection:
   * Several newly calculated columns were added to the ‘sales’ and ‘features’ sheets enabling the identification and flagging of outliers within the dataset. These columns serve essentially as filters during regression analysis, ensuring the reliability and accuracy of predictive models by mitigating the impact of influential data points.
   * A screenshot of a computer

     Description automatically generatedPromotional Markdown, we added a new “promotional expenditures” column where we aggregated all markdown features (1-5). This clarifies somewhat our dataset since we don’t possess the details of each markdown and what exactly it signifies.

Figure 4:Indicative additional columns in the "features" dataset.

A white box with black text

Description automatically generated

Figure 5:Indicative additional columns in the "sales" dataset.

A screenshot of a computer

Description automatically generated

Figure 6:Indicative additional column holding the size-bin category for each store.

**Statistical Analysis and Model Building:**

We apply regression analysis to clarify the intricate relationships between sales performance and other exogenic factors:

1. Regression Analysis:
   * Using Power BI for regression analysis we try to uncover the drivers of sales variability. By evaluating departmental sensitivities to variables such as holidays, major events, and socio-economic indicators, this analysis informs strategic decision-making and resource allocation.

# **3. Results & Analysis**

The overall analysis and description of the results.

## **3.1. C-Level Analysis**

At this part of our analysis, we showcased some high-level metrics and KPIs so as to set the tone of what has happened so far and what Walmart’s priorities have been, so as to recognize the starting point of our analysis. Walmart officials have provided us with data on 45 stores, which are divided into three distinct types. We do not have any detail as of what those types pertain to, so in our analysis we tried to uncover systematic differences between them so as to come up with dedicated promotional approaches for each.

The total sales figure of Walmart in the period considered is mainly attributed to type A stores, followed by type B and then type C.

A close-up of a graph

Description automatically generated

Figure 7:Overview of store types and sales.

We also noted that, among stores of store types, Walmart was the least eager to make promotional expenditures on type C stores; this seems reasonable at a first glance, since those stores are its low earners, so that they may be located in disadvantageous areas with little prospect of improvement. The opposite is true for type B stores, which are closest to the high earners of type A and Walmart would like the two types to converge revenue-wise.

A blue and white graph

Description automatically generated

Figure 8:Average promotional expenditures vs promotional cap.

We noticed that the mean weekly promotional spend per Walmart store was much lower than the cap proposed by Walmart’s officials; This is also true even if we consider only the weeks on which it performed ‘high’ promotional spending.

A blue and black graph

Description automatically generated with medium confidence

Figure 9:Average promotional expenditures vs promotional cap (high).

This helped shape the approach and not concentrate our focus solely on achieving reductions in promotional spend, but also on figuring out opportunities for Walmart to re-allocate promotional spending and/or setting up new promotional campaigns in order to improve the standing of stores of type B and C, so as to help them converge to those of type C.

## **3.2. Current Situation**

This part includes a detailed analysis regarding the situation in Walmart’s sales and promotional activities in the time frame under study.

This page shows how sales evolved over time, together with other numeric features mentioned in our data. All information displayed are dynamically filtered based on store characteristics:

A screenshot of a computer

Description automatically generated

Figure 10:Sliders used in current situation slide.

We first show the intertemporal change of the total weekly sales, in parallel to some other features mentioned in our data, so as to uncover potential seasonalities. Regarding major holidays, it became clear to us that Thanksgiving is deeply linked to Walmart; our conjecture is that this happens because families tend to gather at homes and have dinner on that day, so grocery shopping surges. Then, the Christmas period is linked to a drawback in sales, which are then revamped prior to the Super Bowl, when once again gather at home to watch the game. A graph of a number of different colored lines

Description automatically generated

Figure 11: Weekly sales over time.

As for the other features displayed, temperature and fuel price fluctuated much more heavily than the others in the given data, so we kept them into consideration later into our analysis to uncover links and correlations with other features in our dataset.

We proceeded by presenting some descriptive statistics for weekly sales.

A close up of a price tag

Description automatically generated

Figure 12: Major KPIs identified.

The fact that the average weekly sales are higher than the median cues at the existence of positive skewness in their distribution, while their wide range is explained by Walmart’s operating stores of varying sizes.

Finally, we created a visualization that is used as magic sextant which displays how weekly sales are linked to weekly promotional expenditure. The weekly sales – total promotional expenditures space was divided into six sections by considering the spending cap imposed to us by Walmart’s administration ($50,000 per week & store), in parallel with the bins we created so as to characterize weekly sales as high, medium, and low.

A graph with blue and orange dots

Description automatically generated

Figure 13:weekly sales and promo for type A stores.

For stores of type A, we noticed that the situation is already promising, since most observations were located in the lower-right sextant, corresponding to weekly promotional expenditures less than the cap and high weekly sales. Low sales cases were rather scarce, and the medium sales cases were skewed toward the higher bin. From this we deduct that our approach to type A stores should be focused toward safeguarding their position as Walmart’s most established retail outlets, while trying to conserve money that we could spend for promotional efforts for other types of stores.

A graph of blue and orange dots

Description automatically generated

Figure 14:Promotional expenditures and weekly sales for type B stores.

When it comes to type B stores, we saw that there was some room for improvement. Most observations were concentrated in the medium sales area. Fortunately, Walmart did not spend more than the proposed cap in many of the cases shown. Our focus for this type of stores will be to drive their sales, since they are the closes type (sales-wise) to type A stores and budget could be re-allocated from the latter to them.

A graph of sales

Description automatically generated

Figure 15:Promotional expenditures and weekly sales for type C stores.

Type C stores mainly performed poorly on the weekly level, and that is possibly why it has not been a priority for Walmart to spend a lot of money on markdown. We have analyzed type C stores prospects and whether promotional spend should increase on them later in our analysis.

## **3.3. Socio-Economic Analysis**

We chose to focus on a socio-economic analysis to gain a better understanding of the customers of Walmart. This will provide valuable insights regarding the situation of our customers and their relationship with each type of store.

We initiate our analysis by segmenting stores by type and examining their respective temperatures and unemployment rates (figure 10).

We identify common differences regarding temperature and unemployment and as such we can assume that different types of stores belong, more often than none, to different regions.

Under the above assumption, we compare unemployment percentages and Consumer Price Index (CPI) values among these regions. Our focus extends to assessing unemployment rates and CPI across different regions for each store type, providing insight into regional economic trends. By comparing these socio-economic indicators for each store type and region, we gain valuable insights to guide our promotional expenditures and strategies tailored to specific regions and store types.

Initially, we prioritize Type B stores in our promotional timeline. However, we also recognize the need for remedial strategies for Type C stores experiencing lower weekly sales. We consider options such as either removing these stores or implementing strategies to address underlying economic factors affecting their sales performance.

Our analysis reveals a promising trend: a consistent decline in unemployment rates and an increase in CPI for Type C stores, indicating a potential rise in disposable income and steadily economic expansion on the area. Based on these findings, we recommend shifting focus towards a mid to long-term strategy for Type C stores. This strategic realignment aims to capitalize on the anticipated improvement in disposable income within the regions where Type C stores operate, ultimately boosting their sales performance over time.

A graph showing the average of temperature

Description automatically generated

Figure 16:Average unemployment and temperature by store type.

On the above figure (figure 16) we are offering an overview of the temperature conditions as well as the average unemployment percentage for each store type. This is targeted besides a generic overview of the situation to decrypt information regarding the region and our subsequent assumption under which we work under.

A screenshot of a phone

Description automatically generated

Figure 17: Major KPIs for socio-economic features.

A graph showing a line going up

Description automatically generated

Figure 18:Average unemployment and CPI through time.

We identified that Type C stores follow the same pattern as type A and B regarding the outcome and disposable income. This provides us with the opportunity to suggest a solution regarding promotional expenditures for such types of stores to Walmart and as such involve and affect their strategy as to what has to be done in lower sales stores.

In this effort we have also included the table shown below.

A blue squares with black text

Description automatically generated

Figure 19: Average unemployment for each type of store and sales bin.

There we see the current average unemployment for each type of store and also related to a high, low, or medium sales bin. This can help us identify the current unemployment and based on these as well as the probable projection of unemployment we can see that the unemployment of C will soon converge with the levels that required (regarding unemployment), meaning within the threshold of store types (A, B) which belong to medium or high sale bins.

Therefore, we can derive that there is a steady economic expansion for areas where type C stores are located and as such, we suggest to increase the promotional expenditure (long term) for these types as well since it is very probable that we will eventually increase our sales as well.

## **3.4. Store Level Analysis**

In this section we will try to answer with the help of our insights and diagrams the three questions below:

1. What type of stores give better results when it comes to the total sales?
2. How does the current promotional expenditure affect the sales per type of the stores?
3. How does the size of the bins per type affect the sales and promotional expenditure all in all?

This section provides a snapshot of sales data categorized by store and product type. The first question gets a clear answer from the two diagrams on the left side, which showcase the fact that stores of the type A are the sales leader, followed by type B, with type C completing the show with the distinction of the best-selling product taking into account the amount of money spent on promotion compared to its revenues. The answer to the second question is found right on the pie chart, which reveals the total promotional expenditure for each store segmented by product type (represented by colors). Store of the type A appear to be by far the costliest when it comes to the total promotion expenditure, however the bigger emphasis is given in the promotion of type B, proportionally and not in absolute numbers, as the bar chart we have created on the right shows clearly. Interestingly, Type C seems to be the biggest contributor to overall sales across all stores, followed by Type B and then Type A.

A blue pie chart with a bar and a bar

Description automatically generated

Figure 20: General analysis of stores regarding their performance and promotional expenditures.

Figure 11 offers insights into markdown and sales per store type. Type B takes the lead with the highest markdown divided by the sales figures, followed by Type A and then Type C. The table on the down right side show average sales amount by product type across different size bins. However, without understanding what these size bins represent (e.g., sales volume, customer segment), it's difficult to glean specific insights from this section.

In summary, the dashboard suggests that Store type A is the sales leader while Store type B might invest the most in promotions. A more comprehensive analysis would require additional context, such as the specific product types involved, and the meaning of the size bins in the table.

The below decomposition tree (figure 14) allows the analysis of the most important factors that affect the sales. Here we get the answer to our 3rd and final question we have set right from the start: the larger the bin size, the higher the sales and promotional expenditure, but also vice versa. We can say that with certainty when it comes to the type A and C, while the B type’s behavior remains somewhat unclear. Following, we will examine whether there is a strong correlation or not, which will provide clearer understanding of the situation.

A screenshot of a computer

Description automatically generated

Figure 21:Decomposition tree for average weekly sales.

## **3.5. Correlation/Regression Analysis**

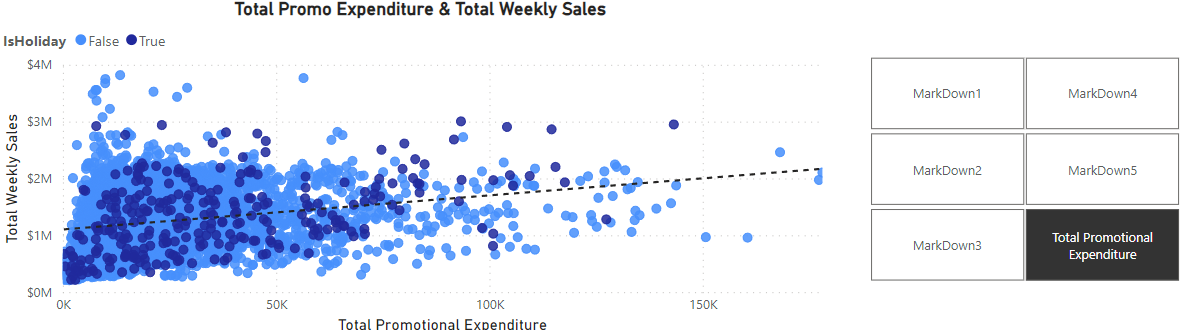
Correlation and regression analysis are indispensable tools in BI for understanding relationships between variables. Whether it's optimizing marketing strategies, forecasting sales trends, or enhancing customer engagement, these analytical techniques provide a robust framework for data-driven decision-making. By leveraging correlations and regression models within BI dashboards, businesses can gain a competitive edge in today's dynamic marketplace. In this section, we'll delve into the application of correlation and regression techniques within two distinct BI dashboards, highlighting their significance and utility.

A graph of a bar chart

Description automatically generated with medium confidence

Figure 22:Intertemporal Promo Expenditures & Total Weekly Sales.

Above, we can observe the quarterly promotional expenses spent for its markdown specifically as stacked bar plots compared to the quarterly sum of total weekly sales in the secondary Y axis as a line. It is apparent that these two values are correlated. Specifically, it looks like the line is following in a way the peaks of the bar plots, suggesting that increased promotional expenses lead to higher sales.

A screenshot of a cell phone

Description automatically generated

Figure 23:Total Promo Expenditure & Total Weekly Sales.

Total Promo Expenditure & Total Weekly Sales dashboard focuses on correlation coefficient, alpha and beta values for the total weekly sales (depended variable) versus the promotional expenses (independent variable). Moreover, by using our slider we can focus on the expenses for a specific markdown.

There are differences between the markdowns. In general, there is a weak positive linear correlation for all the markdowns but the markdown with the highest correlation to the total weekly sales is markdown 1 with correlation coefficient 0.22. However, if we are talking about regression, we can conclude the most successful markdown is markdown 5, with the highest beta value of 16.02. This means that for every 1$ that Walmart spends on promotion through markdown 5, it gets back 16.02$ in sales. Of course, we can also filter with type of store and understand which markdown is more successful per type so that we boost different promotional activities on each store. This will help Walmart maximize its sales by getting the most value for the dollars that it spends on promotions.

# **Conclusion**

In conclusion, the analysis presented herein underscores the pivotal role of data in shaping informed business strategies, particularly in the retail sector. Through a comprehensive examination of a dataset covering 45 Walmart stores across the United States, we have illuminated the complex relationship between various factors such as marketing campaigns, major events and their impact on weekly sales figures. By applying detailed analytics and visualization techniques, we have provided valuable insights into the dynamics and trends influencing Walmart's departmental sales. This analysis not only enhances our understanding of consumer behavior but also empowers stakeholders with actionable intelligence to drive sustainable growth in the retail landscape. As businesses navigate the ever-evolving market, informed decision-making supported by data-driven insights becomes increasingly indispensable, ensuring adaptability and competitiveness in today's dynamic environment.

# **References**

Akmam, E. F., Siswantining, T., Soemartojo, S. M., & Sarwinda, D. (2019). Multiple Imputation with Predictive Mean Matching Method for Numerical Missing Data. 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2019, 1–6. <https://doi.org/10.1109/icicos48119.2019.8982510>

Chandramana, S. B. (2017). RETAIL ANALYTICS: DRIVING SUCCESS IN RETAIL INDUSTRY WITH BUSINESS ANALYTICS. TIJ’s Research Journal of Social Science & Management - RJSSM, 7(4).https://www.theinternationaljournal.org/ojs/index.php?journal=tij&page=article&op=view&path%5B%5D=6121

Kaggle. (2017, September 1). Retail Data Analytics. <https://www.kaggle.com/datasets/manjeetsingh/retaildataset>

Keshwani, R. D. (2023, February 27). Leveraging good data for better decision making by C-level Executives! https://www.linkedin.com/pulse/leveraging-good-data-better-decision-making-c-level-drabla-keshwani

Meulmeester, A. (2022, January 7). Introduction to Temporal Dynamics and Change - Ayrton Sambo, Malik Rigot & Niels Meulmeester - medium. *Medium*. <https://n-n-w-meulmeester.medium.com/a-brief-introduction-to-temporal-dynamics-and-change-b99046fd3740>

Pratt, M. K. (2023, August 4). executive dashboard. CIO. https://www.techtarget.com/searchcio/definition/executive-dashboard

Seetharaman, A., Niranjan, I., Tandon, V., & Saravanan, A. (2016). Impact of big data on the retail industry. Corporate Ownership and Control, 14(1), 506–518. https://doi.org/10.22495/cocv14i1c3p11

Shevchenko, E. (2023, October 21). How to deal with outliers or missing values in a dataset. Medium. <https://medium.com/@eugenesh4work/how-to-deal-with-outliers-or-missing-values-in-a-dataset-b24e85ed66a7>

Sopha, B. M., Makasenda, F., & Asih, A. M. S. (2018). Temporal and spatial prediction of retailer growth. MATEC Web of Conferences, 204, 01001. <https://doi.org/10.1051/matecconf/201820401001>

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

# **Appendix**

*This form should be filled out by all team members after the completion of the group assignment. The team leader should be chosen upon agreement and is responsible for uploading the group assignment after its completion and deal with any technical and other issues that might arise during the submission process.*

Team Leader name:

Team Leader ID:

Team member name:

Team member ID:

***I herewith express my agreement with the submission of this final version of the group project by the team leader.***

Date: 02/03/2024

Team member Signature: \_\_\_\_Stavrogiannis Christos\_\_\_\_\_\_\_\_\_\_\_\_

Team member name:

Team member ID:

***I herewith express my agreement with the submission of this final version of the group project by the team leader.***

Date: 02/03/2024

Team member Signature: \_\_\_\_\_Christoforos Kapsalis\_\_\_\_\_\_\_\_\_\_\_\_\_

Team member name:

Team member ID:

***I herewith express my agreement with the submission of this final version of the group project by the team leader.***

Date: 02/03/2024

Team member Signature: \_\_\_\_Kounelis Panagiotis\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Team member name:

Team member ID:

***I herewith express my agreement with the submission of this final version of the group project by the team leader.***

Date: 02/03/2024

Team member Signature: \_\_\_\_\_Kritikakis Nikolaos\_\_\_\_\_\_\_\_\_\_\_\_\_