

Project Report 2

Table of Contents

Project 2 – Walmart Sales Forecast	2
Introduction	2
Analysis – Part 1	2
Conclusion – Part 1	9
Analysis – Part 2	9
Department 92	10
Conclusion – Part 2	15
Future Work	15
References	15



Date 1/8/2023

Author: Jaimish Mehta

Reviewer: Aung Khant Nyar

Project 2 – Walmart Sales Forecast

Introduction

We are provided with sales data of 45 Walmart stores located in different regions and each store contains several departments.

In addition, Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas.

Based on the data provided, we are required to identify the strongest factor/s that are driving the sales at these stores and forecast sales for the following 2 years.

To perform necessary calculations and graphs for the analysis and presentation, I have used MS Excel & Tableau.

This project requires skills obtained from a variety of Management Courses from Purdue University: Business Analytics, Supply Chain Management, Industrial Applications of Statistics, and Operations Management.

Analysis – Part 1

I have been given 3 main datasets that gives us weekly sales per department at each store, value of markdowns, average weekly temperature, average weekly fuel price, average weekly unemployment rate, CPI, and store sizes.

To understand the datasets and identify trends in them, I perform a series of descriptive analysis as shown below.



Figure 1: Overall Sales for 45 Walmart Stores

The figure 1 shows that quarterly sales across the 2-year time period, the sales remained above average with maximum sales happening during the early Q4 of all the years. To understand the effect of holidays further and identify which holidays bring the maximum sales across all the stores, I filtered the sales data to show sales only during the holidays. As seen in the bottom graph of Figure 1, the sales around Thanksgiving Week are the maximum.

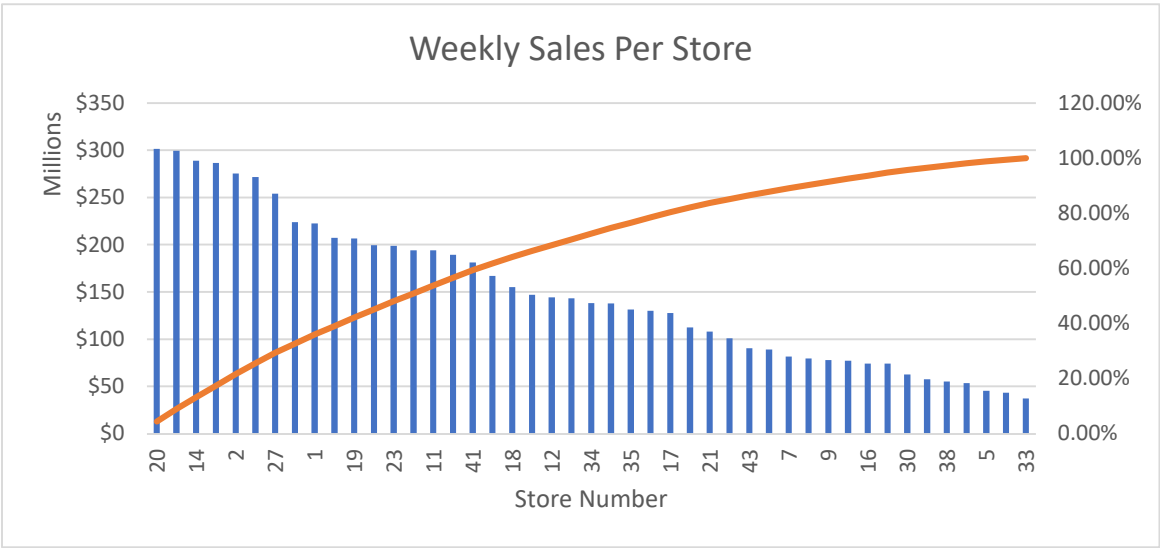


Figure 2: Weekly Sales Per Store

The figure 2 shows the total sales per store and the Stores 20, 4, 14, 13, 2 contribute to 20% of total sales of Walmart.

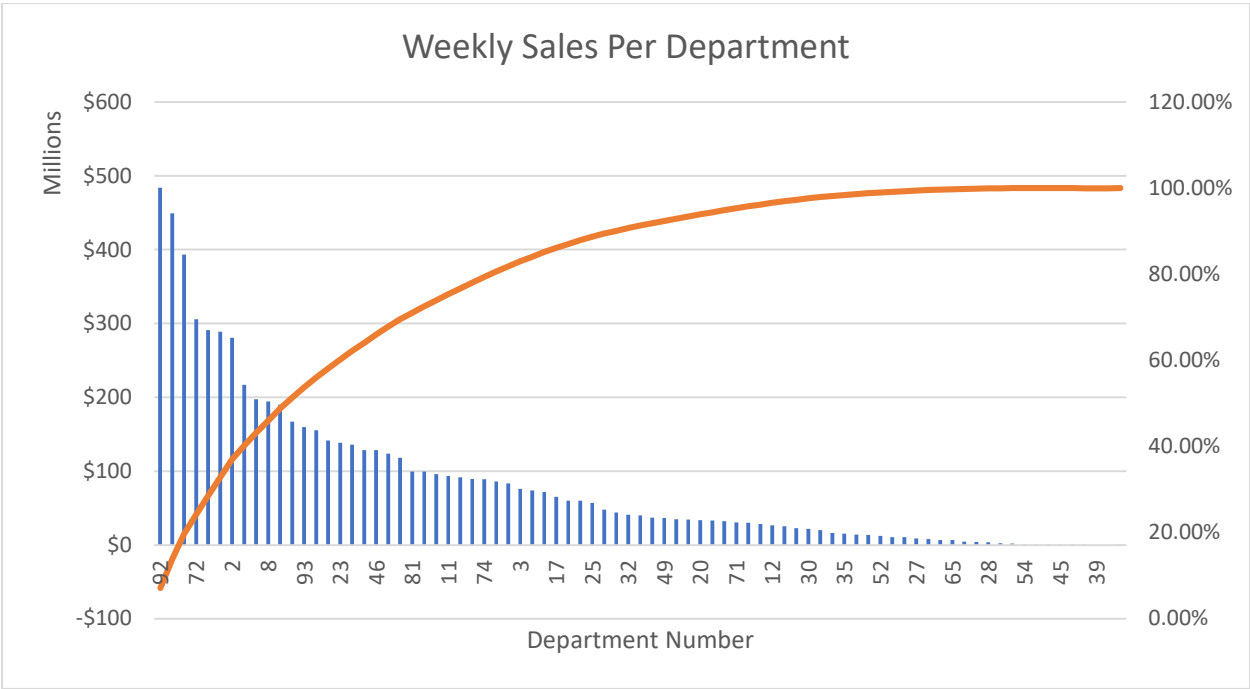


Figure 3: Weekly Sales Per Department

The departments that makes the largest contribution to the overall sales are 92,95, 38 making 20% of the total sales. Based on the % contribution of each department, we can see that Department 92 can affect 7% of sales (maximum overall). To ensure we are focusing on the department that makes significant impact, we will focus on the top 17 department that can affect 2% or more sales as shown in Figure 4.

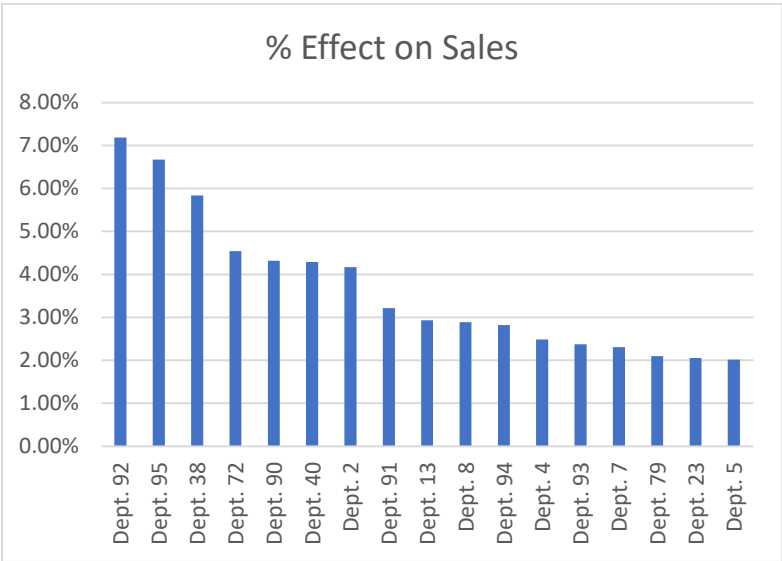


Figure 4: Sales Effect Per Department

To understand the if Fuel Price, CPI, Temperature or Unemployment rate affects the sales at Walmart, the following chart was created that p-value was obtained between Temperature and Weekly sales to see if the correlation is significant.

Based on the below analysis, the correlation is significant and sales are affected by temperature positively.

P-value: < 0.0001

Equation: Weekly Sales = $1.07837e+06 \times \text{Temperature} + 2.26271e+08$

Sales Vs. Temperature

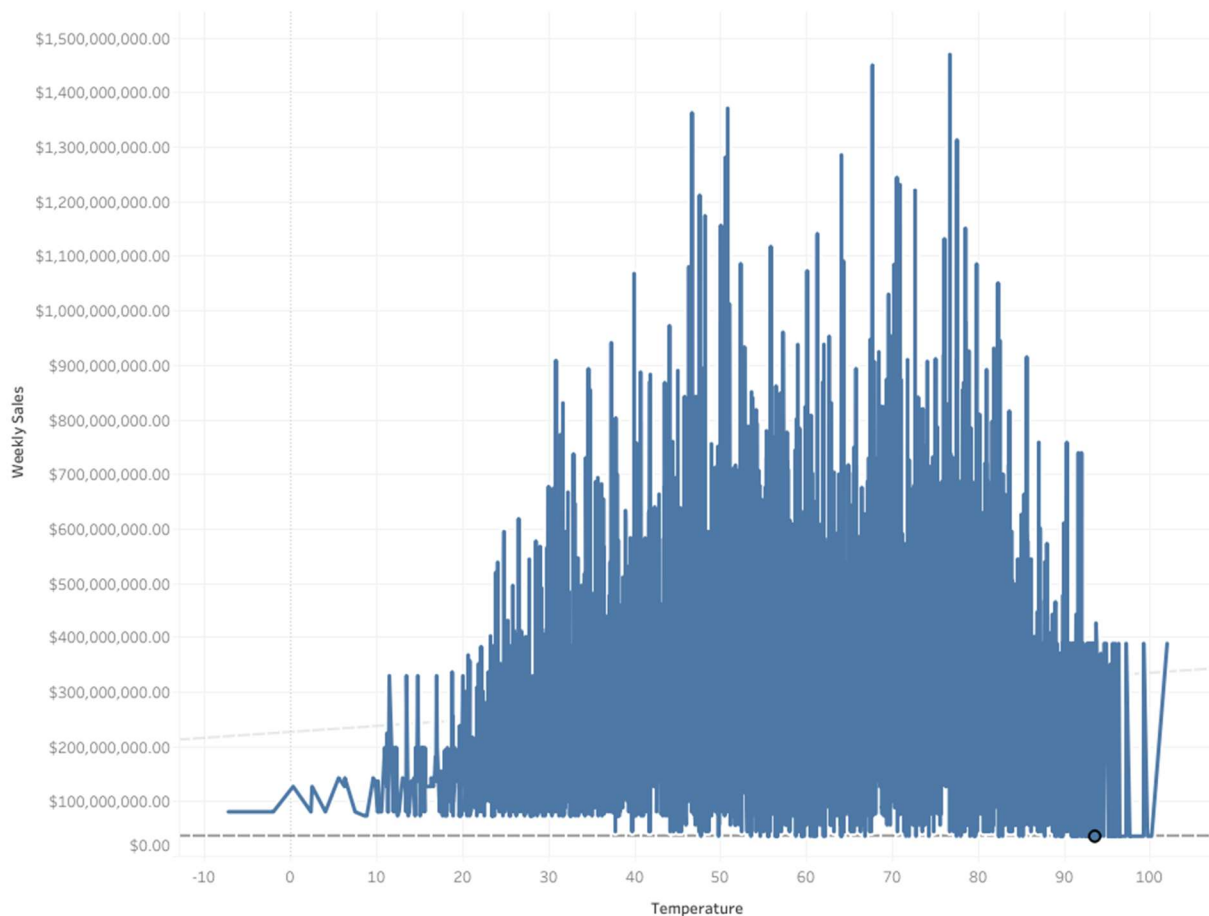


Figure 5: Sales Vs. Temperature

All the Walmart stores offer markdowns (discounts) on different products throughout the year starting from November 2011. These markdowns are intended to increase sales of a particular product in that store. Based on the data obtained for markdowns and their dollar value, the \$7.3M in 2013 during a single month. As you can see in the below chart, the markdowns are available throughout the year and peak during February and August of each year.

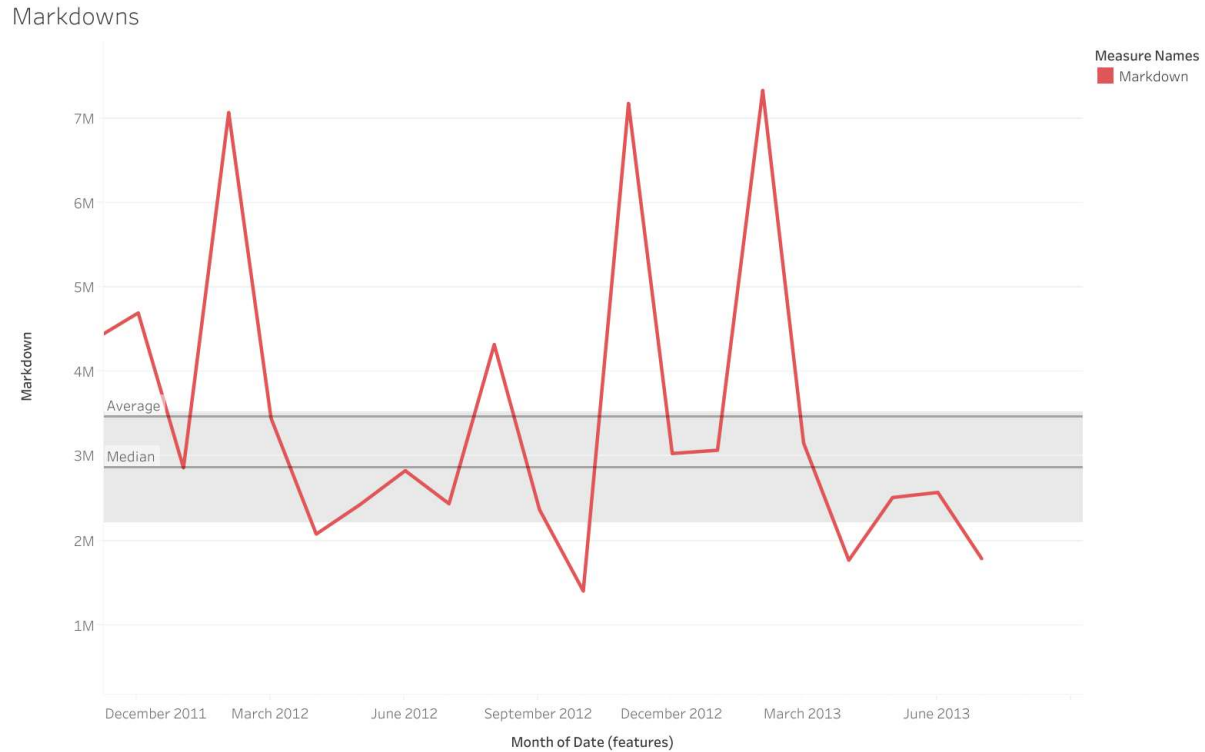


Figure 6: Markdowns throughout the year

While comparing the graph of Figure 6 against the yearly sales in Figure 1, the markdowns do not correlate with sales and hence they do not influence the overall sales. To ensure that this analysis is applicable to store wise sales, we split the sales data per store and the markdown for a side-by-side comparison as seen in Figure 7 below.

Markdowns Vs. Sales Per Store

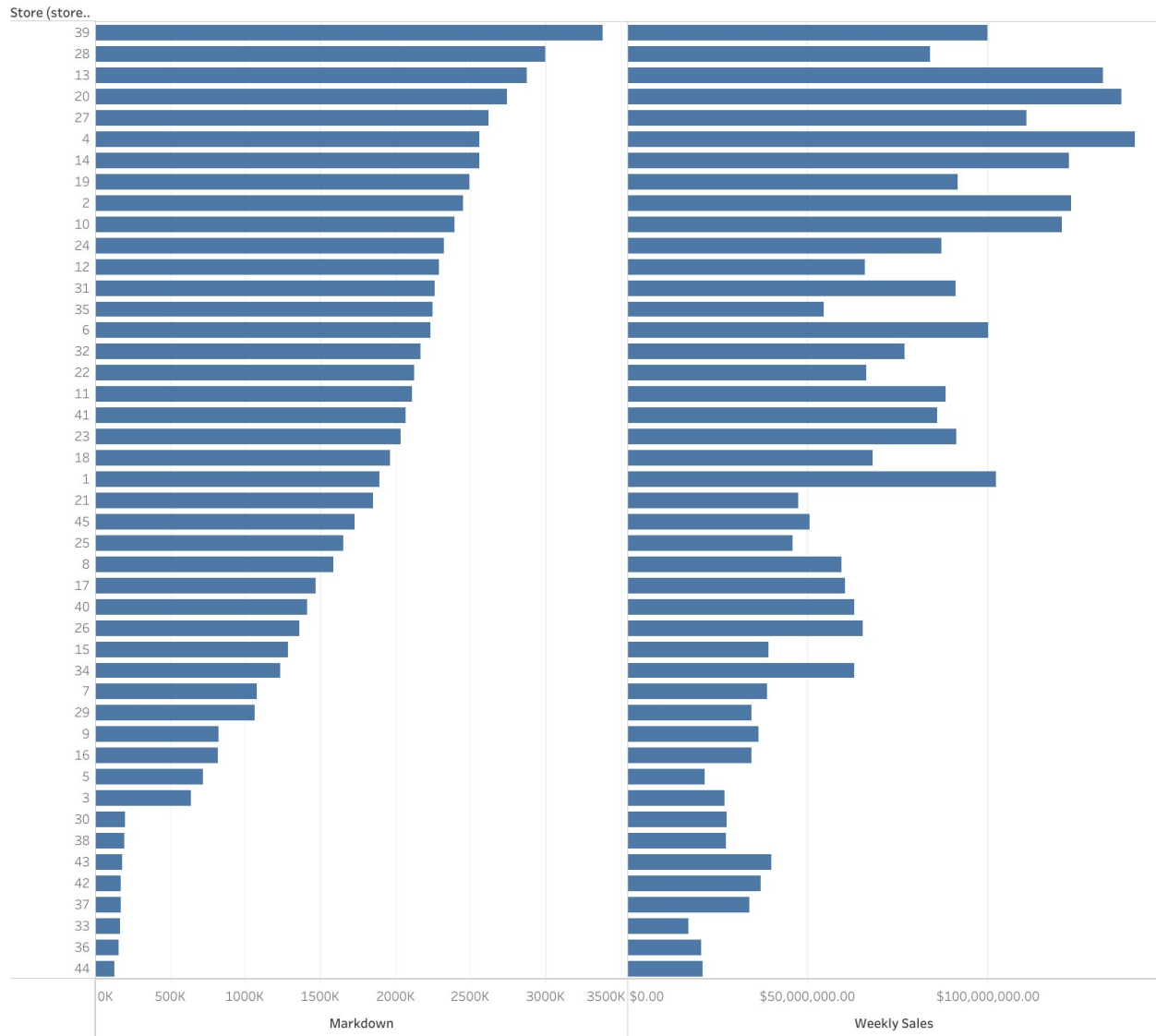


Figure 7: Store wise Markdowns Vs. Sales

As you can see in Figure 7, the markdowns per store are arranged in a descending order with the highest markdown at store 39 with \$3.5M but the total sales for that store is 10th from the top selling Walmart Store. The figure 7 clearly depicts that the markdowns do not bring more sales. To further compare the sales of a store before and after markdowns, the store 39 was selected to compare its sales before and after the introduction of markdowns. As seen in Table 0-1, the sales for store 39 did marginally increased with \$1.8M of markdowns in 2012.

Store 39	2010	2012
Sales	\$65.78M	\$65.88M
Markdowns	\$0	\$1.8M

Table 0-1: Store 39 Sales & Markdowns

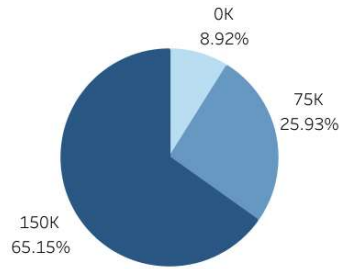


Figure 8: Walmart Store Size

The last factor that may influence the sales in a Walmart is its size. Since the data does not differentiate between a regular Walmart store and a Supercenter, the stores have been split into 3 distinct segments.

Small Stores -> Less than 75K Sq.ft

Medium Stores -> 75K – 150K Sq.ft

Large Stores -> 150K + Sq.ft

As seen in the Figure 8, 65% of 45 stores are Large Stores followed by 26% of medium stores. To understand which stores bring the highest amount of sales, the Weekly sales data was overlapped with store size categorization. The figure 9 shows that the store size directly correlates with the sales which is intuitive as the bigger stores will have more products in stock and that would attract more shoppers.

Store Size Vs. Sales

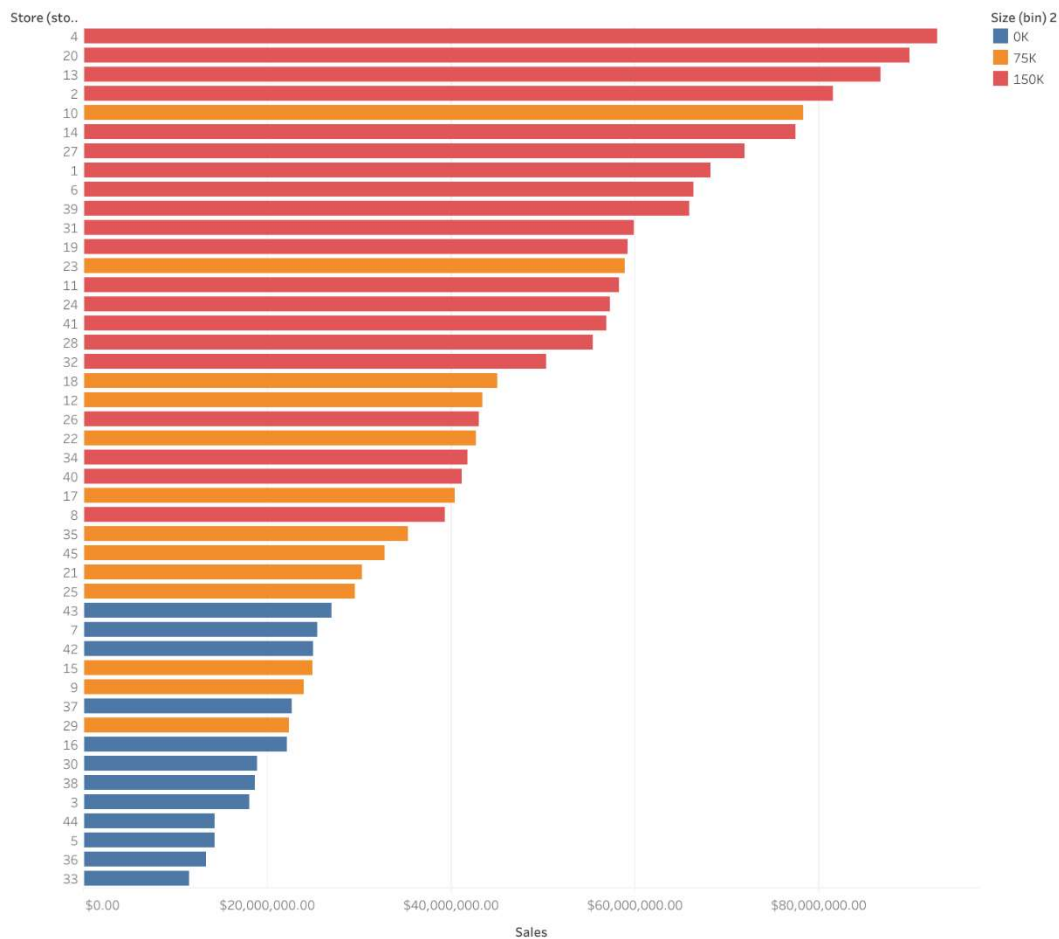


Figure 9: Sales per Store Size

Conclusion – Part 1

Based on the Analysis of part one, the most influential factors affecting the sales based on the given data are temperature of the day and store sizes. The markdowns, unemployment rate, and fuel prices do not affect the sales of Walmart to a significant extent.

Analysis – Part 2

The part 2 of this project requires a model that can be used by the Store Managers to predict sales of their store/department that can help them stock up items to ensure there are no lost sales.

To create a model, we will be doing a Linear Regression analysis to obtain an equation of sales based on the historical sales of department in the stores.

The Linear regression equation is in the form of

$$y = bx + a$$

Where:

y = sales of department

x = time period

a & b are constants from linear regression analysis for top 5 stores based on historic sales

For the quick turnaround of the project, we will focus on department that brought in the most sales at these stores.

The top 5 departments stores and their highest revenue generating departments are:

1. Department 92 (\$483.9M, 7.18% of total revenue)
2. Department 95 (\$449.3M, 6.67% of total revenue)
3. Department 38 (\$393.12M, 5.84% of total revenue)
4. Department 72 (\$305.72M, 4.54% of total revenue)
5. Department 90 (\$291.06M, 4.32% of total revenue)

Due to immense fluctuations of sales in each month, it is not possible to have a singular significant model that can be used to predict sales for the entire store or for the entire department. To overcome this, we will be focusing on predicting sales on a monthly basis. For the purpose of this project, we will be focusing on Department 92 of Walmart and will perform Linear Regression analysis to calculate an equation that can help the Store Manager predict sales.

Department 92

The weekly sales for Store 4’s department 92 is as seen in the below graph

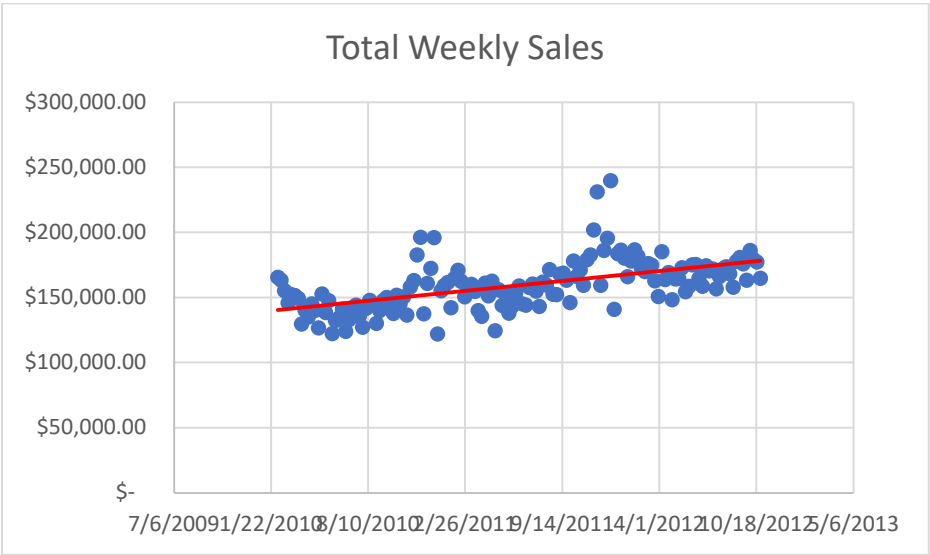
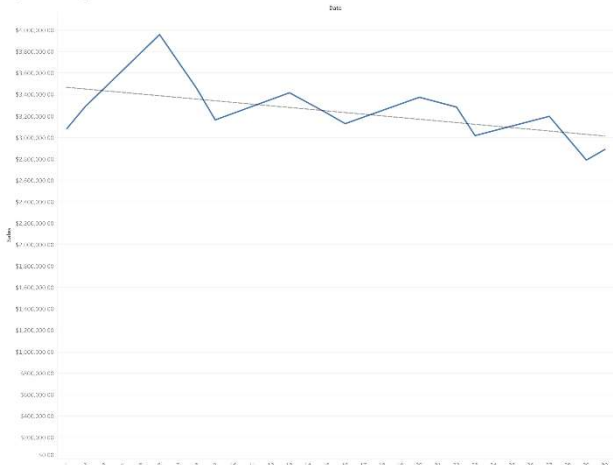
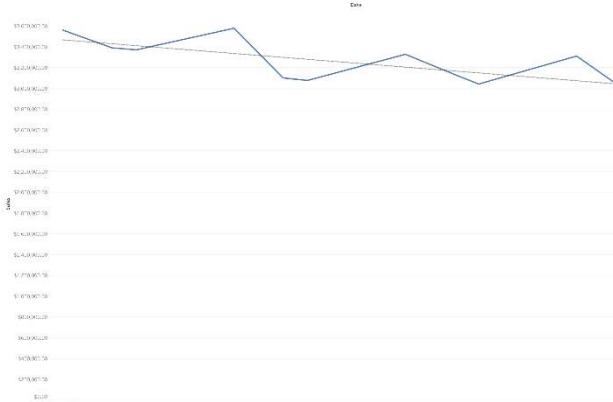
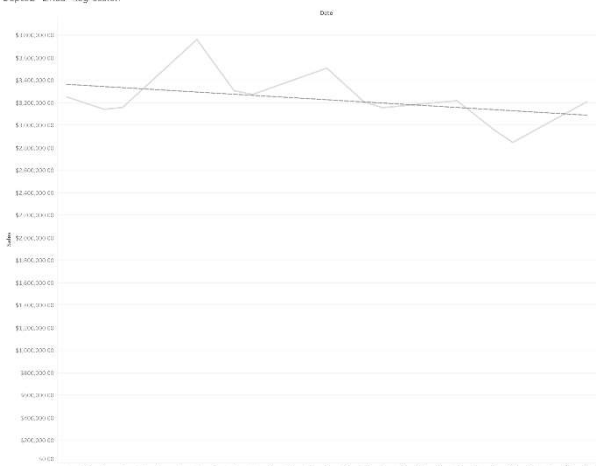
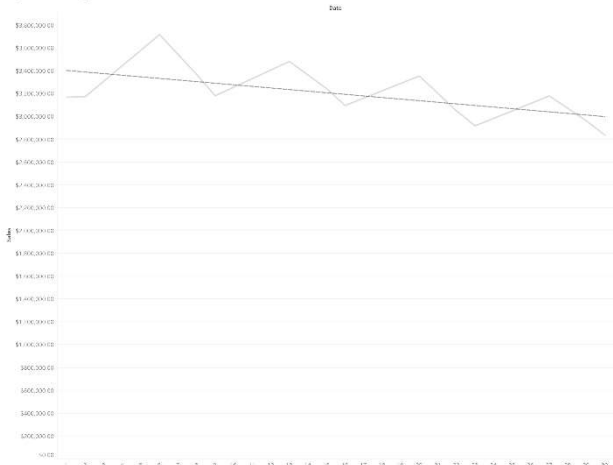
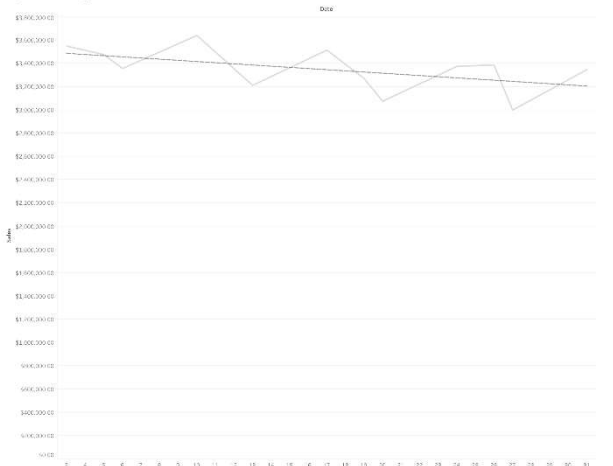
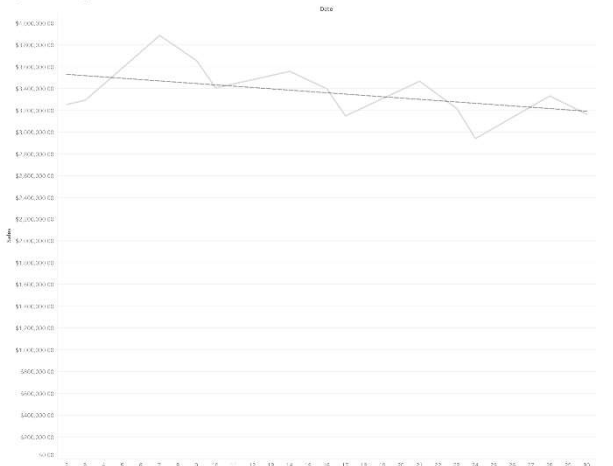


Figure 10: Total Weekly Sales for Store 4 – Department 92

Linear Regression Analysis for Department 92 of all stores as seen in shows the table below.

Month	Graph	Linear Regression Equation
January		Sales = -25644.2*Day of Date + 3.91493e+06 <i>R-Squared: 0.646589</i> <i>P-value: 0.0161401</i>
February		Sales = -24269.6*Day of Date + 3.86057e+06 <i>R-Squared: 0.647265</i> <i>P-value: 0.0016021</i>
March		Sales = -20389.9*Day of Date + 3.65778e+06 <i>R-Squared: 0.566959</i> <i>P-value: 0.0029691</i>

April	<p>Dept92 Linear Regression</p>  <p>Sales = $-15670.8 \times \text{Day of Date}$ + $3.48271e+06$ <i>R-Squared: 0.290177</i> <i>P-value: 0.0468689</i></p>
May	<p>Dept92 Linear Regression</p>  <p>Sales = $-18731.1 \times \text{Day of Date}$ + $3.54162e+06$ <i>R-Squared: 0.523527</i> <i>P-value: 0.0078165</i></p>
June	<p>Dept92 Linear Regression</p>  <p>Sales = $-9776.29 \times \text{Day of Date}$ + $3.37208e+06$ <i>R-Squared: 0.154953</i> <i>P-value: 0.18326</i></p>

July	<p>Dept92 Linear Regression</p> 	<p>Sales = -13940.8*Day of Date + 3.41593e+06 <i>R-Squared: 0.338743</i> <i>P-value: 0.0289918</i></p>
August	<p>Dept92 Linear Regression</p> 	<p>Sales = -10040.1*Day of Date + 3.51542e+06 <i>R-Squared: 0.252952</i> <i>P-value: 0.079796</i></p>
September	<p>Dept92 Linear Regression</p> 	<p>Sales = -12076.1*Day of Date + 3.55344e+06 <i>R-Squared: 0.206769</i> <i>P-value: 0.118487</i></p>

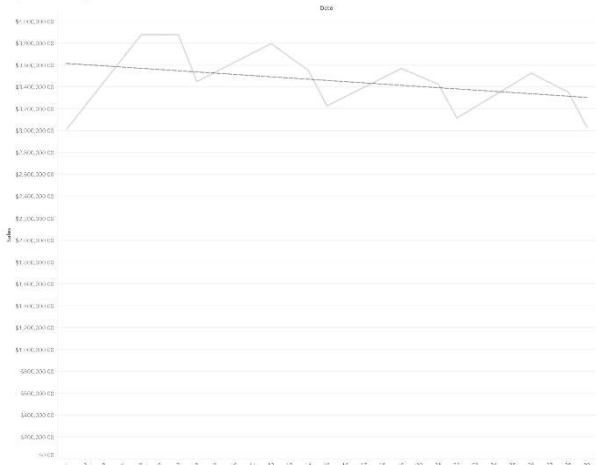
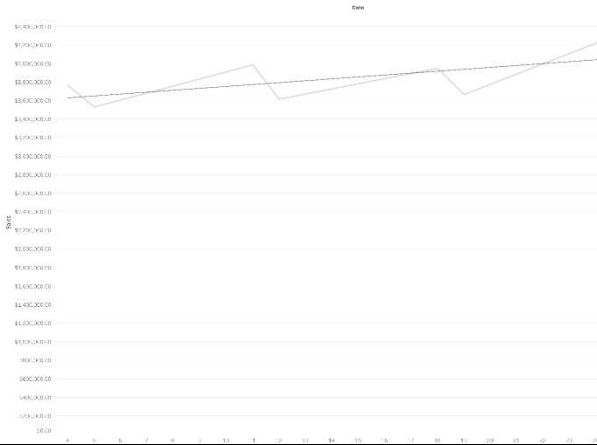
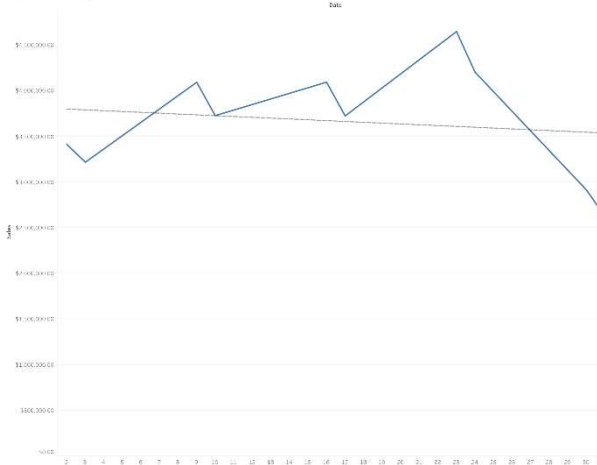
October	<p>Dept92 Linear Regression</p> 	<p>Sales = -11085.3*Day of Date + 3.62618e+06 <i>R-Squared: 0.116375</i> <i>P-value: 0.253996</i></p>
November	<p>Dept92 Linear Regression</p> 	<p>Sales = 20547.8*Day of Date + 3.54489e+06 <i>R-Squared: 0.437729</i> <i>P-value: 0.0739488</i></p>
December	<p>Dept92 Linear Regression</p> 	<p>Sales = -9055.68*Day of Date + 3.81535e+06 <i>R-Squared: 0.0227506</i> <i>P-value: 0.677457</i></p>

Table 0-2: Linear Regression For Dept. 92

For each month, the sales equation in the rightmost column determines the forecasted sales based on the date of the month. The accuracy of the forecasted sales can be determined by the value of R-Square and P-value. Typically, a P-value of less than 0.08 (92%) is determined to be significant and the R-square determines the percentage of data on the line.

Conclusion – Part 2

Based on the observations from for all 12 months, it can be noted that the model is not accurate for June, September, October, and December. This can be due to customers anticipating discounts in the upcoming weeks of those months.

Future Work

For future work, an Excel Macro can be created that can ask the Store Manager to enter Department number, month of the year, and date which can calculate the forecasted sales and its accuracy.

References

Dataset: <https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data?select=stores.csv>