



Santa Clara University

Walmart Stores Data Analysis

Aashna Rungta, Sneha Batchu, Sony Kumari, Udisha Madnani

Marketing Analytics, MKTG 2505

Dr. Sujata Ramnarayan

March 7th, 2022

2. Table of contents

- [1. Cover page](#)
- [2. Table of contents](#)
- [3. Introduction to Company Overview](#)
- [4. Situation analysis](#)
- [5. Identification of problem](#)
- [6. Approach](#)
- [7. Model and Insights](#)
- [8. Conclusion](#)
- [9. Recommendation](#)
- [9. References](#)

3. Introduction to Company Overview

Walmart Inc (Walmart) is a retailer that operates grocery stores, supermarkets, hypermarkets, department and discount stores, and neighborhood markets. As of January 31, 2022, Walmart has 10,593 stores and clubs in 24 countries, operating under 48 different names. Walmart is the world's largest company by revenue, with US \$548.743 billion, according to the Fortune Global 500 list in 2020. It is also the largest private employer in the world with 2.2 million employees.

4. Situation analysis

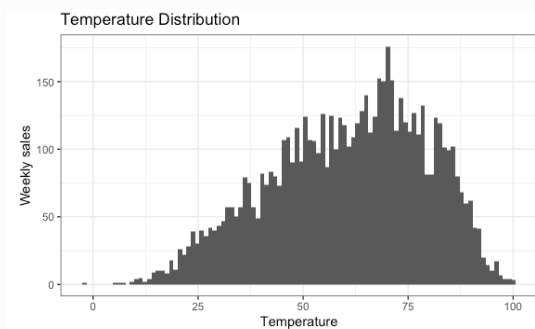
We want to study the effect of various factors on the weekly sales at Walmart, so that we can cater towards increasing revenue. Therefore, we want to explore the data with few features, focusing on the weekly sale numbers.

Fig1:



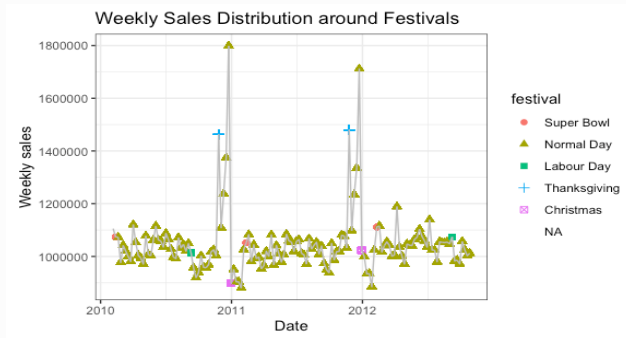
- Insight : The weekly sales are different in different stores, depending on the location.

Fig2:



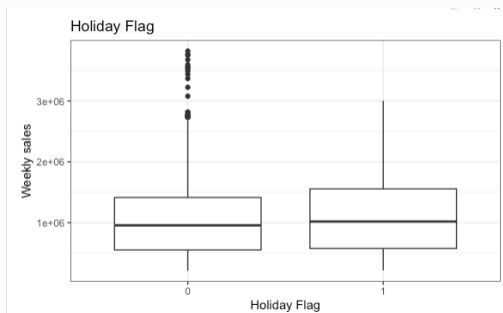
- Insight: More sales were observed when the temperature was around 75 degrees.

Fig3:



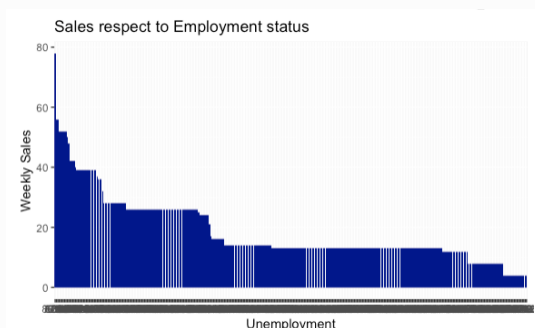
- Insight: Spikes in sales were observed in weeks close to holidays. Our understanding is, this is because sales start a couple of weeks earlier and people want to shop before the holiday.

Fig4:



- Insight: Extreme values were observed for non-holiday weeks (above explanation justifies this trend). On average, there isn't a huge difference between the average sale for holidays as compared to non-holiday weeks.

Fig5:



- Insight: Our understanding of this observation is, the employed section of the society would prefer shopping from high end markets. Walmart is targeting mass over class.

Other insights:

- CPI and fuel do not have an observable effect on the weekly sales data

5. Identification of problem

The problem we are trying to solve is understanding the consumer behavior for a retail supermarket. We want to identify the factors that influence the sales data.

6. Approach

We have used Multi-linear Predictive Regression Analysis, in order to remove possible omitted variable bias, and use all the variables in the dataset and come up with a more accurate model. We felt using the multi - linear regression model was appropriate because we wanted to analyze the causal effect of different significant variables(Temperature, festival, CPI, Unemployment, and Fuel Prices) on average weekly sales of Walmart. Also, to magnify the significance of few variables, we have added interaction terms in the regression model. Linear regression model was found as a suitable technique to study impact of various numeric independent variables on a numeric dependent variable, which is our variable of interest.

7. Model and Insights

In our model, we regressed multiple independent variables on weekly sales, namely store number, temperature, festival, inflation rate, unemployment rate, and two interaction terms. This model allowed for us to use all the data at hand in order to predict any variability in the weekly sales. In particular, we wanted to use two interaction terms - CPI*unemployment and CPI*fuel price because those are variables that are closely related to each other - i.e. unemployment affects inflation rate and vice versa, and the fuel price is a large determinant of the inflation rate. The interaction terms ensured that we had dealt with the multicollinearity of the variables appropriately.

From our model, we found that the RMSE of the training set was 619,378.7 sales, as compared to the RMSE of the test set, which was 510,856.7 sales. While this may seem very high, we have to look at it relative to the average weekly sales, which is 1,082,071. Both the RMSE values are over 50% of the average weekly sales, indicating that the model is not very accurate in predicting weekly sales.

Additionally, the mean absolute error was 408376.8, which again is very high, especially compared to the average weekly sales (MAE is 37% of the average weekly sales). Again, this shows that the model could do better in predicting sales. However, we wanted to standardize the numbers and use a measure that better represents the fit. In order to do this, we calculated the R2, which was 0.2028. An R2 value of 0.2 shows that the model is not necessarily the best at predicting the weekly sales. One reason for this could be that the store number does not have any mathematical meaning, and that those numbers are randomly assigned. With all these three measurements in mind, it seems like the model gives the following results. However, we need to be mindful that the model is not highly accurate in predicting sales, and therefore consider obtaining additional data in order to make the data more accurate.

Fig6:

```

Residuals:
    Min       1Q   Median       3Q      Max
-1125108 -364955  -54652   365877  2751969

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6648859.09  285789.84   23.265 < 2e-16 ***
Store        -19107.76    535.53   -35.680 < 2e-16 ***
Temperature   -63.23     378.54    -0.167  0.8673
festivalNormal Day -44903.22  44981.75   -0.998  0.3182
festivalLabour Day -19440.73  70148.87   -0.277  0.7817
festivalThanksgiving 389482.68  68709.82    5.669 1.50e-08 ***
festivalChristmas -124767.01  68532.58   -1.821  0.0687 .
CPI           -33508.77   1750.49   -19.142 < 2e-16 ***
Unemployment  -386993.07   18206.12  -21.256 < 2e-16 ***
Fuel_Price    -521081.11  65004.01   -8.016 1.29e-15 ***
CPI:Unemployment    2606.00    127.18    20.490 < 2e-16 ***
CPI:Fuel_Price     3412.62    377.87     9.031 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 503600 on 6378 degrees of freedom
(45 observations deleted due to missingness)
Multiple R-squared:  0.2042,    Adjusted R-squared:  0.2028
F-statistic: 148.8 on 11 and 6378 DF,  p-value: < 2.2e-16

```

The table above shows that during Thanksgiving, sales skyrocketed as seen in the coefficient of the Thanksgiving festival variable. It seems that other festivals did not have this effect on weekly sales.

Unemployment has, predictably, a negative correlation with the weekly sales. This is seen in the

negative coefficient of the unemployment variable, keeping all other variables constant. The two interaction terms do not have a significant effect on the sales either, with low coefficients. This might be due to the slow changing effect of unemployment, fuel price, and inflation rate - these are factors that do not have a very high variability on a week-to-week basis.

8. Conclusion

- During festivities, we always observe that there is a huge rush at Walmart and our research from the data also backs the same hypothesis - Thanksgiving actually has extremely high weekly sales.
- Our research shows that the sales reach a peak just a week before Christmas.
- Festivals like Labor day and the Super Bowl do not cause extremely high sales.
- Unemployment rate negatively affects the weekly sales - the lower the unemployment rate, the higher will be the weekly sales.
- Certain variables have little or no effect at all on the weekly sales, like - CPI and fuel price.

9. Recommendations

There has been a significant drop in the average weekly sales from 2011 to 2012 as per the data analysis. Thus, as per our primary and secondary research, we come to the following recommendations for Walmart -

- Cater to supply chain disruption issues like, transportation delays, etc and kick in automation for activities like inventory management, etc.
- Stores should restock shelves fast enough in order to avoid undue shrinkage and out-of-stocks.
- Digital services should be integrated more with the physical stores.

9. References

1. R for Fundamental Data Analysis in Market Research by Sujata Ramnarayan
2. Essentials of Marketing Analytics by Joseph Hair, Dana E. Harrison, Haya Ajjan
3. <https://talkbusiness.net/2021/09/walmart-ceo-sees-supply-chain-challenges-inventory-shortfalls-into-2022/>
4. <https://www.cnbc.com/2015/05/19/wal-mart-knows-it-has-problems-how-it-plans-to-fix-them.html>

10. Appendix

More charts and figures from our analysis are as follows:

