



Odette School of Business

University of Windsor

BSMM-8720: Data Analytics & Project Management

Section 1

Master of Management

Summer 2023

Submitted to

Dr. Shashi Shahi

Submitted By

<u>Members</u>	<u>Student ID</u>
Binz Baby Joseph	110094887
Maliha Qayyum	110100391

Table of Contents

Table of Contents

<i>Executive Summary.....</i>	<i>3</i>
<i>Business problem</i>	<i>3</i>
<i>Goal</i>	<i>3</i>
<i>Objectives</i>	<i>3</i>
<i>Methodology</i>	<i>3</i>
<i>Implementation plan.....</i>	<i>3</i>
<i>Dataset Overview.....</i>	<i>4</i>
<i>Understanding the Dataset</i>	<i>4</i>
<i>Removing The Outliers.....</i>	<i>5</i>
<i>Correlation With Weekly Sales</i>	<i>6</i>
<i>Previous Dashboard</i>	<i>6</i>
<i>Sales Forecasting</i>	<i>7</i>
<i>Regression Model.....</i>	<i>7</i>
<i>Time Series Forecasting</i>	<i>7</i>
<i>Analytic Hierarchical Process (AHP) Analysis.....</i>	<i>9</i>
<i>Implementation and Recommendations</i>	<i>11</i>
<i>Conclusion.....</i>	<i>12</i>
<i>References.....</i>	<i>13</i>

Executive Summary

Walmart Inc is an American multinational retail corporation that operates a chain of supercenters, discount department stores, and grocery stores. Its headquarters is located in the United States of Bentonville, Arkansas.^[10] The company was founded by brothers Sam and James Walton in 1962. It also owns and operates Sam's Club retail warehouses.

A comprehensive market analysis was conducted to identify key locations with the highest retail sales using data from years 2010 to 2012. Additionally certain factors like holiday dates, temperature, unemployment rate was considered for sales forecasting. For sales forecasting we used various techniques like time series analysis, ANOVA and we also used Analytic Hierarchy Process (AHP) for making the final decision.

Business problem

Walmart is one of the world's largest company by revenue, With its widespread global presence Walmart wants to optimize the performance of its various stores, to boost the overall sales performance for maintaining a competitive edge in the industry.

As the project manager responsible for optimizing Walmart's sales performance, they leverage insights derived from the available historical data from 2010, 2011 and 2012. This will be used for sales forecasting for the next two years, indicating a steady increase in sales. However, as a proactive measure to further enhance sales, the project manager needs to develop and prioritize different strategies for implementation in one of the three zones, consisting of the 45 stores. The strategies will be chosen based on the insights from the sales forecasting, market analysis, and AHP analysis, with the aim of identifying the zone that will have the most significant impact on overall sales performance when the selected strategies are executed.

Goal

To improve Walmart's sales performance by leveraging insights derived from sales forecasting, market analysis, and AHP analytics. The intent is to focus on one zone at a time out of those 3 zones that would impact the overall sales performance the most.

Objectives

To boost the overall sales performance of the company and maintain a competitive edge in the industry.

Methodology

- Use time series forecasting functions to forecast future sales by using advanced machine learning algorithms like Exponential Triple Smoothing.
- Use Regression for sales based on attributes like Temperature, Fuel Price, CPI, and Unemployment
- Compare both the models.
- Use Analytic Hierarchical Process (AHP) to identify the most suitable zone to select.

Implementation plan

- Analyze the year-wise trends in sales to identify factors that contribute to lower sales.
- Implement strategies to address location-specific factors affecting Walmart's sales.
- Identify successful strategies implemented in high-performing years and replicate them.
- Use Insights from Sales Forecasting and AHP.

- Introduce new products/services to attract customers.
- Optimize product offering & inventory levels based on estimated demand during holiday weeks.

Dataset Overview

The data considered here is from our Dashboard Visualization project which we have further extending to forecast sales. Our Dataset of the weekly sales consists of 6436 records and 8 columns of data extracted from Kaggle. To measure the sales performance, we used variables like store number, date, weekly sales, holiday flag(whether the week had a holiday or not) were considered to understand the sales performance among certain factors including holiday category, temperature, fuel price, unemployment rate, and CPI. The holiday events on the holiday week included the Super Bowl, Labor Day, Thanksgiving and Christmas. Before analyzing the data we performed some basic data cleaning checking for true data types and changing the datatypes in case of discrepancy. There were no duplicates or missing value found in the data.

Understanding the Dataset

From the above figure we get an overview of the spread of the data and how our data is distributed.

From the histogram and box plot we can observe the following:

- The dataset for Temperature is somewhat normally distributed with the Distribution for Unemployment which is skewed to the right. The weekly sales data is slightly skewed to the right as well.
- CPI and Fuel price have somewhat of a bimodal data distribution, suggesting that we are looking at two groups of data and the measure of central tendency should not be our first choice.
- From the box plots we can see that weekly sales and unemployment rate have outliers which needs to be removed going forward.

Weekly_Sales		Temperature	
Mean	1046964.878	Mean	60.66378244
Standard Error	7035.371661	Standard Error	0.229933793
Median	960746.04	Median	62.67
Mode	-	Mode	50.43
Standard Deviation	564366.6221	Standard Deviation	18.44493288
Sample Variance	3.1851E+11	Sample Variance	340.2155488
Kurtosis	0.053140927	Kurtosis	-0.612800959
Skewness	0.668361797	Skewness	-0.336767601
Range	3608700.2	Range	102.2
Minimum	209986.25	Minimum	-2.06
Maximum	3818686.45	Maximum	100.14

Fuel_Price		CPI		Unemployment	
Mean	3.358606838	Mean	171.5783938	Mean	7.999151049
Standard Error	0.005722121	Standard Error	0.490619196	Standard Error	0.023384704
Median	3.445	Median	182.6165205	Median	7.874
Mode	3.638	Mode	126.4420645	Mode	8.099
Standard Deviation	0.459019707	Standard Deviation	39.3567123	Standard Deviation	1.875884782
Sample Variance	0.210699092	Sample Variance	1548.950803	Sample Variance	3.518943715
Kurtosis	-1.177377796	Kurtosis	-1.839813342	Kurtosis	2.639711784
Skewness	-0.0961583	Skewness	0.06349185	Skewness	1.188143933
Range	1.996	Range	101.1688068	Range	10.434
Minimum	2.472	Minimum	126.064	Minimum	3.879
Maximum	4.468	Maximum	227.2328068	Maximum	14.313

The above two figures give us the descriptive statistics for the dataset with the measure of central tendency and the measure of dispersion. The following are some of the key insights:

- Kurtosis measures the heaviness of the tails (outliers) of a distribution compared to the normal distribution. Weekly Sales has a positive kurtosis, indicating that its distribution has heavier tails and more outliers than a normal distribution. While, Temperature has a negative kurtosis, suggesting that its distribution has lighter tails and fewer outliers.
- Skewness measures the asymmetry of a distribution. Weekly Sales has positive skewness, indicating that it is skewed to the right. While CPI has nearly zero skewness suggesting it is approximately symmetric.

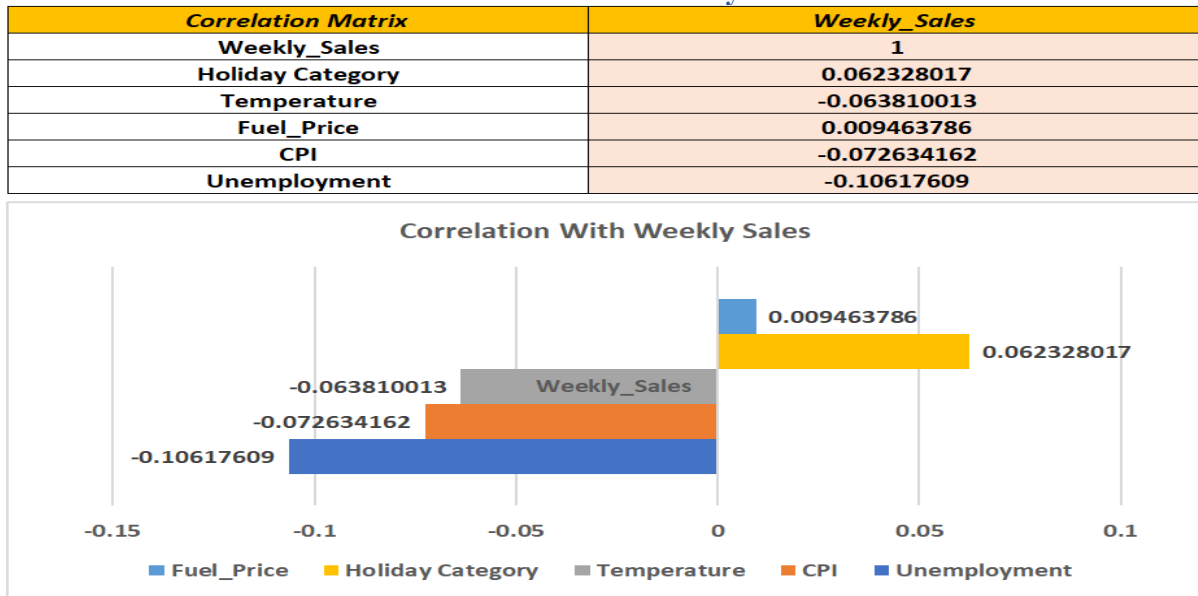
Removing The Outliers

From the above insights we discovered that there are some outliers present in Weekly Sales column and Unemployment column. To remove those outliers, we used the quartile method (dividing the dataset into four quarters) using which we calculated the lower limit and the upper limit based on the interquartile range. Any value greater than the upper limit or less than the lower limit will be our outlier.

Unemployment		Weekly_Sales	
QTL 1	6.891	QTL 1	553350.105
QTL 3	8.622	QTL 3	1420158.66
IQR	1.731	IQR	866808.555
Lower Limit	4.2945	Lower Limit	-746862.7275
Upper Limit	11.2185	Upper Limit	2720371.493

- In case of weekly sales, the outliers is less than 0.5% of the data so we can remove them
- While in case of Unemployment the outliers is around 7.5% of the data so we adjust the data by normalizing it. Thus, not skewing our analysis and not losing the data.

Correlation With Weekly Sales



The above figure shows that there is very less correlation between factors with weekly sales. However, comparing all Unemployment rate has the most correlation with Weekly Sales having a correlation value of 0.106 with a negative sign which implies that as Weekly Unemployment rate increases the Weekly Sales decreases and vice versa. This correlation will be further used in our comparison matrix weights for the AHP.

Previous Dashboard



Key insights from Dashboard

- Yearly average sales: in terms of average sales for Walmart we observed that the average sales of 2011 were substantially lower than the year 2010. By contrast the sales performance had gotten better for the year 2012.

- Bottom 5 stores' performance: The stores with the lowest sales performance have higher variations in sales which may be due to the geographical location and customer demand.
- Top 5 stores' performance: amongst all stores Store 20 appears to be the top performer with an average sales performance of 4.47% followed by store#4 at 4.45%. The top 5 stores collectively represent 21.55% of all of Walmart's total weekly sales. More so, the top 5 stores are almost neck to neck with each other with relatively little variation in sales throughout the years.
- Holiday sales: Thanksgiving is the most popular of all holidays with 32.54% of the total holiday sales for 3 years. The average sales for the other holidays are equally close to each other by 1% - 1.5%.

Sales Forecasting

Regression Model

Here we used Multilinear regression model to predict our dependent variable weekly sales based on the independent variables - Holidays, Temperature, Fuel price, CPI, and Unemployment rate. The following are the results of the model.

Regression Statistics	
Multiple R	0.152703268
R Square	0.023318288
Adjusted R Square	0.022554659
Standard Error	539012.6353
Observations	6401

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	4.4359E+13	8.87181E+12	30.53614078	8.49724E-31
Residual	6395	1.85797E+15	2.90535E+11		
Total	6400	1.90233E+15			

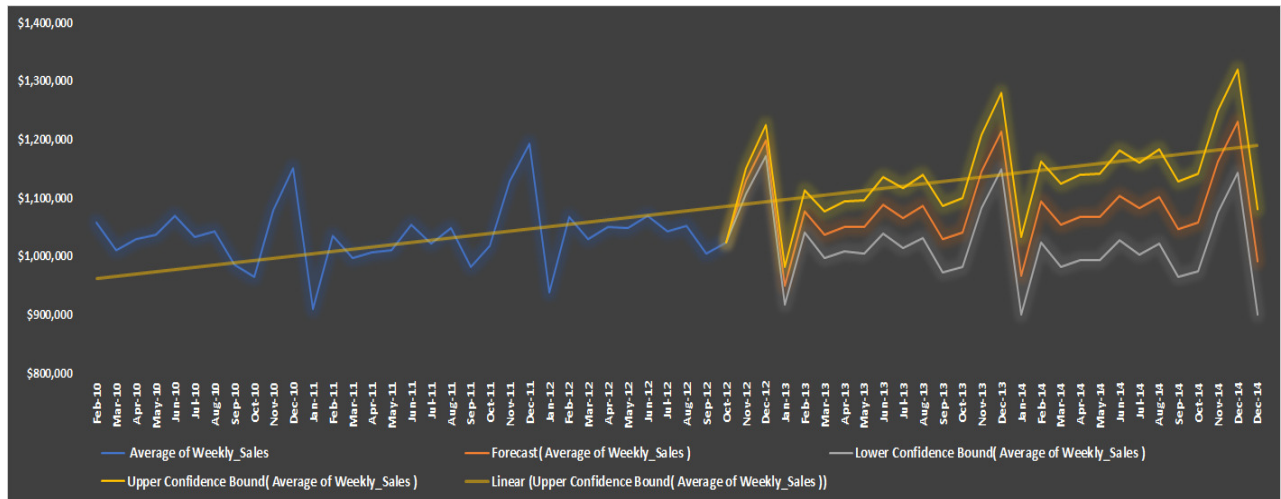
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1693542.945	80232.80139	21.10786257	1.15984E-95	1536259.776	1850826.115	1536259.776	1850826.115
Holiday_Flag	54974.12612	26984.00731	2.037285474	0.04166253	2076.431868	107871.8204	2076.431868	107871.8204
Temperature	-226.5472388	388.0425338	-0.583820636	0.559361584	-987.2406036	534.146126	-987.2406036	534.146126
Fuel_Price	-4585.220769	15275.51013	-0.300168095	0.764058697	-34530.33809	25359.89655	-34530.33809	25359.89655
CPI	-1515.596541	188.5219669	-8.039363087	1.06742E-15	-1885.162753	-1146.030329	-1885.162753	-1146.030329
Nw_Unemployment	-47004.03798	4480.207814	-10.49148609	1.52274E-25	-55786.74621	-38221.32975	-55786.74621	-38221.32975

In conclusion, while the model shows some statistical significance, its practical usefulness in predicting weekly sales is limited due to the low R-squared value. However, this model provides us a useful insight of how significant impact these attributes are having on the weekly sales. It is observed that unemployment has the most negative impact on weekly sales which further proves our correlation matrix.

Time Series Forecasting

Here we used time series forecasting functions to predict future values based on historical data. These functions use advanced machine learning algorithms, such as Exponential Triple Smoothing (ETS).

Statistic	Value
Alpha	0.75
Beta	0.00
Gamma	0.25
MASE	0.19
SMAPE	0.01
MAE	9,815.58
RMSE	13,210.33



For calculating the accuracy of the model we use the following:

Accuracy Percentage (MASE) = $100\% - \text{MASE} * 100$ Accuracy Percentage (MASE) = $100\% - 0.19 * 100$
 Accuracy Percentage (MASE) $\approx 99.81\%$

Accuracy Percentage (SMAPE) = $100\% - \text{SMAPE} * 100$ Accuracy Percentage (SMAPE) = $100\% - 0.01 * 100$
 Accuracy Percentage (SMAPE) $\approx 99.99\%$

Accuracy Percentage (RMSE) = $100\% - (\text{RMSE} / (\text{Maximum} - \text{Minimum})) * 100$ Accuracy Percentage (RMSE) = $100\% - (13,210.33 / (2,685,351.81 - 209,986.25)) * 100$
 Accuracy Percentage (RMSE) $\approx 99.47\%$

Accuracy Percentage (MAE) = $100\% - (\text{MAE} / (\text{Maximum} - \text{Minimum})) * 100$ Accuracy Percentage (MAE) = $100\% - (9,815.58 / (2,685,351.81 - 209,986.25)) * 100$
 Accuracy Percentage (MAE) $\approx 99.603\%$

These values indicate high accuracy of the model which is highly close to 100%.

Here even though the sales is increasing we can further boost our sales by focusing our strategies on one of the zones out of the three zones in which the 45 stores is distributed.

Analytic Hierarchical Process (AHP) Analysis

To understand which location to prioritize, AHP applications have been used for site selection. The quantitative data for the analysis was sales, temperature, CPI and unemployment rate. The qualitative data includes the holiday events, and landscape.

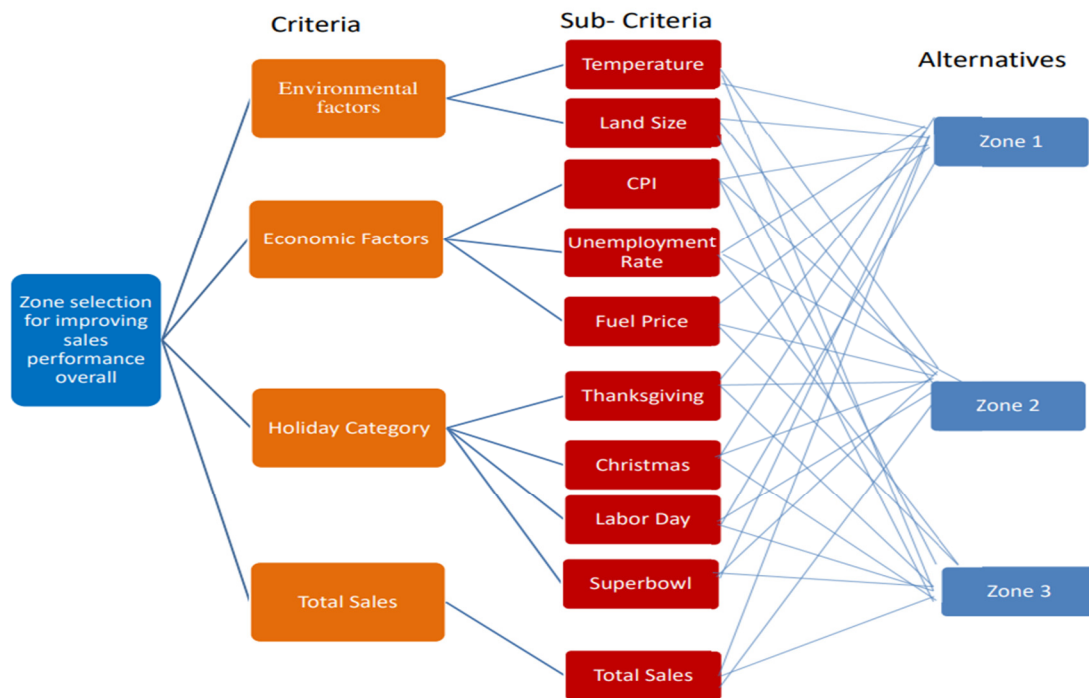
Goal - To improve Walmart's sales performance across the 45 stores divided into 3 segments located in various areas.

Alternatives:

Zone 1

Zone 2

Zone 3



1	Criteria Wise Comparison Matrix				
	Criteria	Environment	Economic Factors	Total Sales	Holiday Category
	Environment	1	1/5	1/7	1/3
	Economic Factors	5	1	1/3	3
	Sales	7	3	1	5
	Holiday Category	3	1/3	1/5	1
	Sum	16.00	4.53	1.68	9.33

2	Normalized Criteria Matrix					
	Criteria	Environment	Economic Factors	Total Sales	Holiday Category	Priority Eigen Vector
	Environment	0.06	0.04	0.09	0.04	0.06
	Economic Factors	0.31	0.22	0.20	0.32	0.26
	Sales	0.44	0.66	0.60	0.54	0.56
	Holiday Category	0.19	0.07	0.12	0.11	0.12

			n	RI
3	λ_{max}	4.177	4	0.9
	Consistency Index, CI	$(\lambda_{max} - n)/(n - 1)$	0.06	
	Consistency Ratio, CR	CI/RI	6.54%	Since 6.54% < 10% the matrix is consistent

	Final Ranking		
	Alternative Candidate	Overall Priority	Rank Based on Overall Priority
Zone 1	L1	0.408	1
Zone 2	L2	0.266	2
Zone 3	L3	0.137	3

Explanation For Weights

- Environment is given the least priority since it is not that of an impact on sales compared to other factors. (Temperature and Land Size)
- Economic Factors stand second due to the reason being the subcriterias.
- Unemployment, CPI and Fuel Price has the highest impact on sales which is gained from our correlation matrix and the regression analysis.
- Sales has the highest priority since our main goal is to improve overall sales performance.
- Holiday category seems to also impact sales based on zone to zone with our 4 holiday categories being Christmas, Thanksgiving, Labour day, and Super bowl which is almost equally weighed.

Here from the AHP analysis Zone 1 has the highest rank, and we need to prioritize our strategies on Zone 1 first for improving our sales performance.

Implementation and Recommendations

Walmart can improve its sales performance by implementing the following recommendations:

1. **Expansion and Location Optimization:** Walmart should consider relocating underperforming stores to more strategic locations and investigate the financial aspects of opening new stores where the Walmart stores are performing well.
2. **Targeted GPS marketing:** When customers are in the vicinity of a Walmart, Walmart can use GPS data to place online ads to customers in that specific geographical area to encourage them to visit their nearest Walmart store. This location-Specific Ads can also include localized promotions, offers and events that resonate with the community.
3. **Offering seasonal discounts and promotions during the holiday season -** Walmart may benefit from the higher sales volume seen around Thanksgiving and Christmas by offering a variety of seasonal discounts and promotions. Customers are encouraged to buy more by using this tactic, which could boost revenue and sales. Furthermore, Walmart can attract more customers and create a sense of urgency to shop at their stores during the festive periods by strategically timing and promoting these offers.
4. **Creating product bundles with attractive pricing options -** Walmart can encourage customers to buy complementary items together by developing product bundles based on recent trends and seasons. This will help create further needs for customers and increase their purchase and attention intent. This strategy of offering products bundles at attractive pricing options will allow Walmart to drive sales of related items and leverage the popularity of exclusive products, eventually boosting revenue.
5. **Analyzing customer purchases for targeted marketing campaigns:** Walmart can gain insights into customers' specific preferences and needs by analyzing customer purchase patterns by collecting this information, it will enable Walmart to implement targeted marketing campaigns to consumers by providing them with exclusive deals and offers tailored to their interests.
6. **Implementing community engagement programs –** to develop its reputation and build stronger connections with customers Walmart should engage with the local community. Not only will this create goodwill but also increase brand visibility. The Walmarts located located in zone 3 can actively form partnerships with local organizations, supporting local causes and participating in community events. Such community engagement programs may include 1% of the customers total bill will be donated to a support a local cause like orphanages.

These actions will contribute to Walmart's goal of driving future growth and making informed decisions in the competitive retail industry.

Conclusion

According to our analysis we have chosen to prioritize zone 1 over the other two zones because the stores located within the zone 1 impacts the overall performance and profitability of Walmart, playing a crucial role in driving the company's success. We have identified and quantified the importance of each criterion and sub-criterion through the use of the AHP model, in the decision-making process. Walmart can maintain a competitive edge in the market and continue its upward trajectory by using the quantitative and qualitative data collected from the analysis to influence decisions about future expansion, marketing plans, and resource allocation.

To predict weekly sales, Regression and time series forecasting models are utilized based on factors like temperature, fuel price, holidays, CPI, and unemployment rate. Analyzing these factors' will aid the project managers to prioritize strategies that address location-specific factors and replicate successful initiatives from high-performing years.

By focusing on Zone 1, Walmart can effectively increase its sales performance and uphold its dominant position in the retail industry. With a data-driven and strategic approach, Walmart will be able to ensure sustained success in the market by effectively navigating the challenges of the ever-changing retail landscape.

References

1. Dataset: <https://www.kaggle.com/datasets/yasserh/walmart-dataset>
2. Excel Toolpak: <https://support.microsoft.com/en-gb/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4>
3. https://www.researchgate.net/figure/Saaty-fundamental-scale_tbl1_313258932
4. <https://core.ac.uk/download/pdf/230936818.pdf>
5. <https://en.wikipedia.org/wiki/Walmart>
6. <https://cedcommerce.com/blog/awesome-ways-to-increase-sales-walmart/>