Student Names: Team AJPWY
Peiyang (Leo) Li, Yanjia(Bridget) Qian, Winnie Huang, Yile (Arin) Wang, and Junchen Xiong
Professor: Dr. N. Orkun Baycik

**BA 472 Course Project**

**WALMART RETAIL ANALYSIS**

# I. Introduction

Walmart is one of the largest multinational retail corporations that operates a chain of supercenters, discount department stores, and grocery stores from the United States, headquartered in Bentonville, Arkansas. The company operates over a multitude of product categories and is facing various challenges due to unforeseen customer demands, fierce competition in the retail industry, and the vacillation of national economies, which all lead to fluctuating sales and unstable inventory levels.

While numerous companies learn to analyze their own individual business operation and attempt to find which factors impact the desired dependent variable the most, such as revenue, or which variable can be controlled to achieve business objectives, most companies ignore the impact of the macro environment. We consider that the macro environment would have a direct impact on the business. Although factors from the macro environment are out of the business`s control, it is still necessary to run experiments to evaluate the specific impact of metrics that measure the current economic situation on the business's revenue since it helps businesses to adapt better and take prescriptive actions to face the changes of social economies. Furthermore, Walmart needs to know more about its customers' preferences to provide better service.

Holding the project objectivity to study the impact of different variables on business sales in a more comprehensive way while facing the lack of available datasets for the company, we employed three different datasets and chose the variables from the macro-environment and variables that the company can control.

**Dataset and Variables Explanation**

FIRST DATASET - Walmart Sales Data

Recording the weekly sales and other related variables of 45 Walmart stores located in different regions

- **Store:** this is the store number and ranges from 1 to 45
- **Date:** the week of sales
- **Weekly_Sales:** the sales of the given **week** in 'Date'

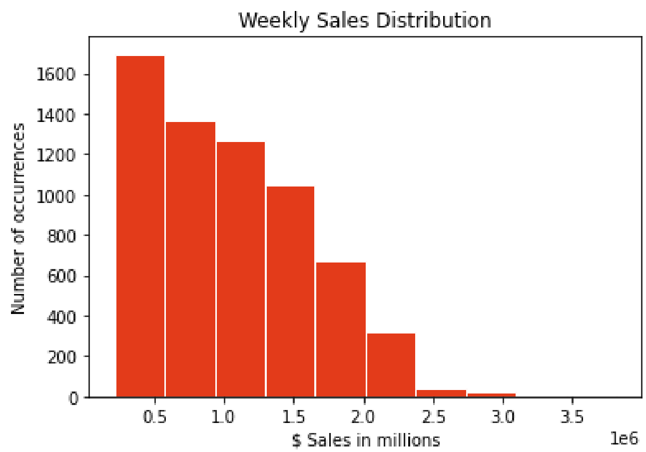


Figure 1. Weekly Sales Distribution Graph

- **Holiday_Flag:** whether the week contains a special holiday (Super Bowl, Labor day, Thanksgiving, Christmas)
    - 1 if yes, 0 otherwise
- **Temperature:** Average temperature of the week. From the table below, we can observe that temperatures are normally distributed with a right skew.

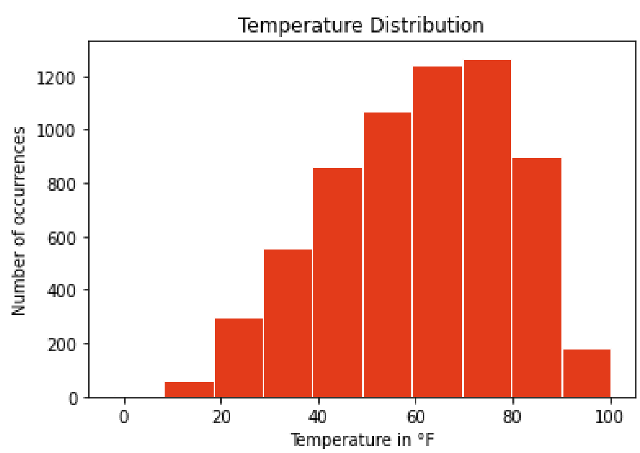


Figure 2. Temperature Distribution Graph

- **Fuel_Price:** the cost of fuel in the region
- **CPI:** Consumer Price Index
    - A measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services ( a measure of inflation & deflation)
    - If CPI increases by 5%, it means that an average consumer has to pay 5% more of the same basket of goods
- **Unemployment:** prevailing unemployment rate

SECOND DATASET - Checkout preference survey data

The survey result asking customers whether they prefer self-checkout or cashier-assisted checkout

- **Retired:** whether the responder is retired
    - 1 being yes, 0 being no
- **Preference:** whether the customer prefers self-checkout or cashier-assisted checkout
- **Age**: age of customer that filled out the survey. From the table below, we can observe that age of customers is evenly distributed in the survey.

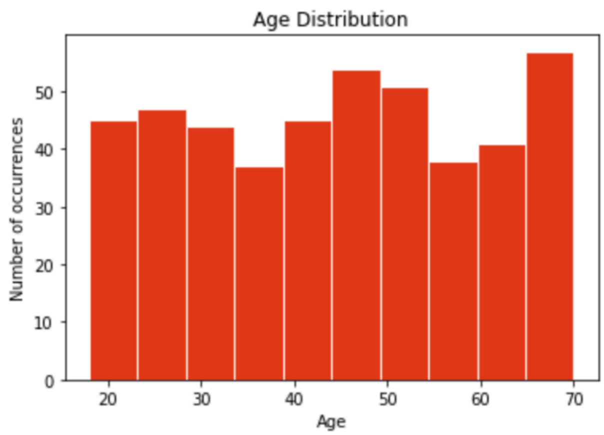


Figure 3. Age Distribution Graph

THIRD DATASET - Customer reward program data

Walmart started a loyalty program in the first quarter of 2021 that rewards customers as they make purchases. Our team believes that analyzing the change brought by introducing the customer loyalty program would give us useful insights, helping Walmart make better decisions.

- **Date**: Records the date of the purchase of the customer. Please note that the reward program is implemented after 2021-01-01, so all purchases before the date do not include the reward program.
- **Age**: the age of the customer that completes the purchase.
- **Shop_Online**: If the purchase is made at Walmart's online store or if it is made at the retail store.
- **Annual Income**: customer's annual income
- **Amount_Spent**: records the dollar amount that the customer purchased. We can observe that most sales made at Walmart are relatively low dollar amounts.

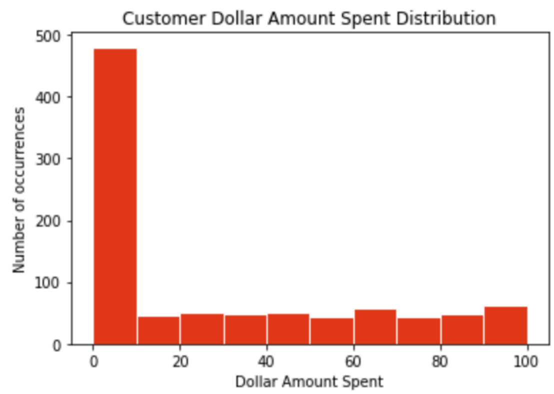


Figure 4. Amount_spent Distribution Graph

## II. Research Questions

We are interested in how Walmart can leverage its data to cope with different economic situations, as well as different approaches that the company can utilize to stimulate revenue and purchase from customers in various operating environments. From Walmart's perspective, its revenue is largely impacted by the overall economic trends of the country, considering the scale of Walmart's business.

The first part of our research question is about how the change in environment and economics will affect the overall sales level of stores. More specifically, how would factors including temperature, CPI, and fuel prices impact the revenue of each individual store? Therefore, we designed the research questions below:

- How would a change in temperature affect its sales?
  - We will use two variables in this experiment: Temperature and Weekly_Sales in the t-test. Each sale is correlated with the specific week's temperature. With this research question, we aim to explore if the change in temperature affects sales, i.e., if the change in seasonality, which is related to temperature, will influence the revenue. Our follow-up experiment will be testing which specific temperature range will lead to higher sales, performed by a one-tailed t-test.
- How would economic indicators, including CPI and Fuel Price, impact sales level?
  - We will use three variables in this two-way ANOVA test: CPI level, Fuel Price level, and Weekly_Sales. Because of the numerical nature of CPI and Fuel Price, we categorize them into four respective levels for the experiment. This experiment looks for sales' relationship with important economic indicators and analyzes if sales are related to them. Considering Walmart's business scale, its revenue will likely be affected by the country's

economic performance, and this experiment will provide Walmart with meaningful indicators.

By answering these questions, Walmart can predict its revenue ahead of time when facing either a change in seasonality or a decrease in CPI, which indicates poor economic performance that applies to the entire country. This can facilitate Walmart's inventory planning and forecasting, as well as its preparation for financial reports.

In the event of a decrease in sales, it is important for Walmart to incorporate strategies that have been statistically proven effective in stimulating sales. The research questions above lead to the second part of our research questions, which is the approaches and methods Walmart can utilize to increase sales. We designed the following research questions:

- In terms of switching checkout lanes, will different customer groups react differently between the cashier-assistance model and self-checkout kiosks?
    - We understand that customer perception is very subjective. Even though Walmart had publicly claimed that installing self-checkout kiosks would save the company roughly $12 million[1] per year in cashier wages for every second in average transaction time, to fully understand the dynamics of customer perception, our team will conduct a heterogeneity test to support the Walmart management team make a reasonable business decision regarding checkout lanes.
- Will a reward program lead to an increase in sales? If so, is there a specific range of customers` annual incomes that are more likely to be influenced by the program?
    - We attempt to perform a regression analysis on the relationship between the reward program and sales. We will use two variables in the initial experiment: reward(yes or no) and related sales. This is an attempt to test if the current reward program will stimulate revenue, and if so, Walmart can adopt it when necessary. We will also perform the experiment with age as a covariate.
    - We will use DoWhy analysis to assess whether there is a causal relationship between reward and sales, providing further support for the previous conclusion from regression analysis.

[1] Wal-Mart Boosts Self-Checkout, But Its Claimed Cost Savings Don't Add Up. (n.d.). FierceRetail. https://www.fierceretail.com/operations/wal-mart-boosts-self-checkout-but-its-claimed-cost-savings-don-t-add-up

By setting up experiments that address the above questions, Walmart will obtain valuable information on methods that can increase sales, directly and indirectly, and take actions to adjust and improve the operation process. This will further help Walmart make informed business decisions that can avoid losses of revenue if they predict that sales will be decreased for a specific period of time.

# III. Experiment and Analyses

## T-test/Assumption check: Temperature's effect on Sales

To answer the first research question we proposed, which is "how would a change in temperature affect sales?" we picked the Walmart retail dataset that contains weekly sales of 45 stores and its respective average temperature as below:

Table 1. Walmart Retail Dataset

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment |
|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 05-02-2010 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 |
| **1** | 1 | 12-02-2010 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 |
| **2** | 1 | 19-02-2010 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 |
| **3** | 1 | 26-02-2010 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 |
| **4** | 1 | 05-03-2010 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 |

Our team stated that the temperature change would affect the population's weekly sales, and the mean weekly sales in lower temperatures are different from the mean in higher temperatures. To determine whether the statement should be rejected, we needed to set up a two-sample hypothesis testing the desired dependent variable: temperature and the response variable: weekly sales. Because we could only obtain sample data, meaning that we were trying to do inferential statistic testing on the population's weekly sales, the t-test was used. Before initiating the analysis, we checked the assumptions of the t-test.

### T-test assumption check

Table 2. Summary of T-test assumption check for Temperature

| **Temperature** | Desired Outcome | Actual Output | Conclusion |
|---|---|---|---|
| the central limit theorem | >= 30 | 6435 | Satisfied |
| Shapiro-Wilk Test | Reject the null hypothesis | 3.811e-43 | Drawn from a normally distributed population |

| | | | |
|---|---|---|---|
| | | | (p-value may not be accurate for N > 5000) |
| Anderson-Darling Test | Test statistic is greater than critical values | 32.89 > [0.576, 0.656, 0.787, 0.917, 1.091] | Is not drawn from a normally distributed population |
| Skewness | Near 0 | -0.336768 | Negatively skewed |

Table 3. Summary of T-test assumption check for Weekly_Sales

| **Weekly_Sales** | Desired Outcome | Actual Output | Conclusion |
|---|---|---|---|
| the central limit theorem | >= 30 | 6435 | Satisfied |
| Anderson-Darling Test | Test statistic is less than critical values | 86.88 > [0.576, 0.656, 0.787, 0.917, 1.091] | Is not drawn from a normally distributed population |
| Skewness | Near 0 | 0.668362 | Positively Skewed |

From the tables above, we observe that Anderson_Darling Test for both variables is not drawn from a normally distributed population. The descriptive histograms from the introduction section can also confirm our observation since temperature displays a moderate negative skew while the weekly_sales column is highly skewed on the right side.

In addition, we generated probability plots of temperature and weekly sales against the normal distribution in the figure below to further test our observation that the two variables are not drawn from a normally distributed population.

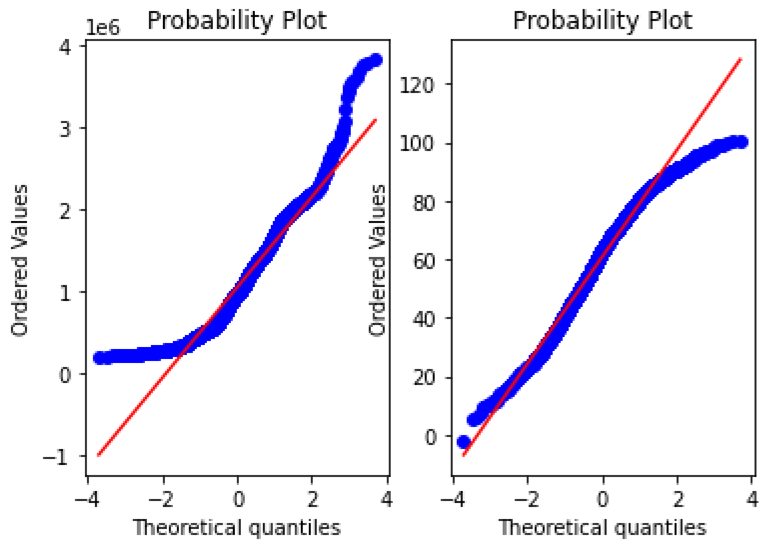


Figure 5. Weekly_Sales/Temperature Probability Plot Against the Normal Distribution

From the above two plots, as well as the statistics, our team argued that both weekly sales and temperature sample data do not conform to a normal distribution, and we will perform the necessary transformations after we partition them into our desired samples. One corresponds to the weekly sales of lower temperatures, and the other matches sales of higher temperatures. The threshold value for temperature is 60.66 Fahrenheit, which is the average temperature for all seasons. Since we partitioned the dataset into two groups, it was necessary to check if skewness was still present.

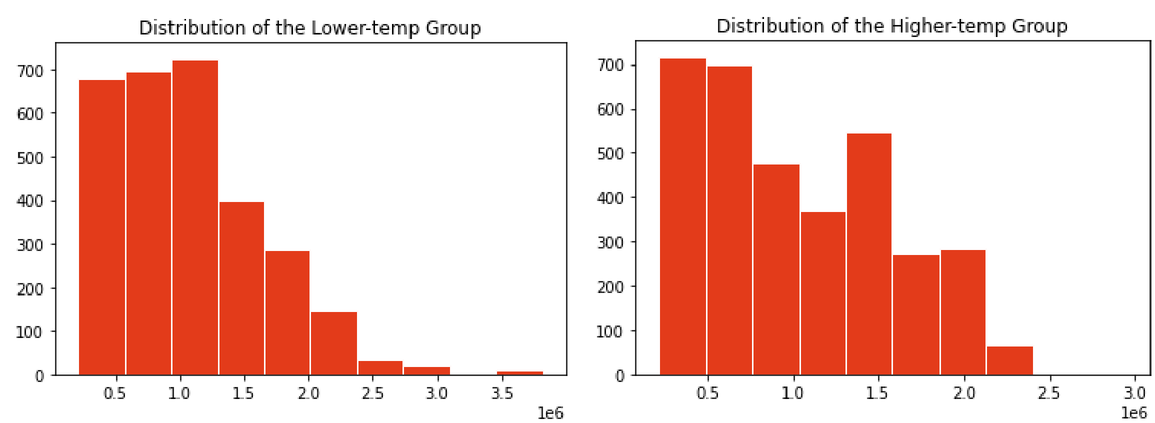


Figure 6. Descriptive Summary of Two Samples

From the histograms above, we can clearly see that both groups still display a relatively high positive skew pattern. With the skewness level of 0.931(lower temperatures) & 0.409 (higher temperatures), we can confirm that the two groups also do not conform to the normal distribution with the following probability plots.

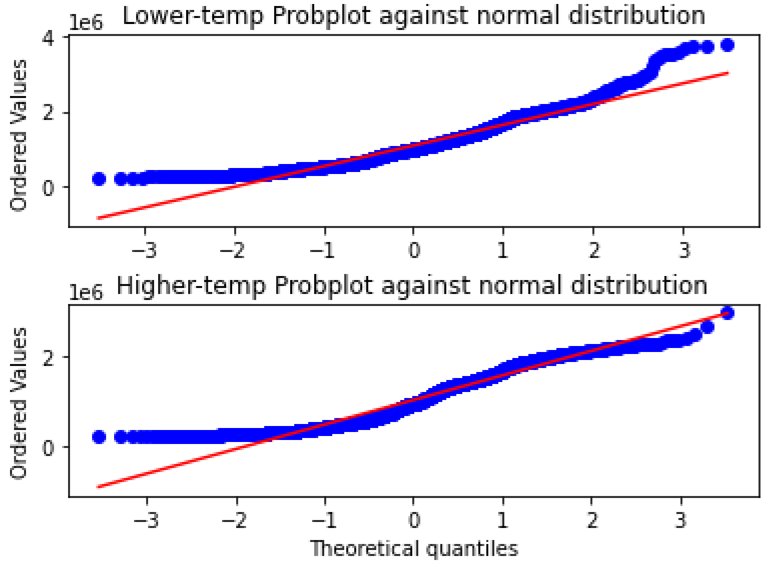


Figure 7. Two Samples' Probability Plots Before Normalization

Therefore, our group performed square root transformation on both groups to normalize the data. The following figure summarizes our probability plots after the normalization.

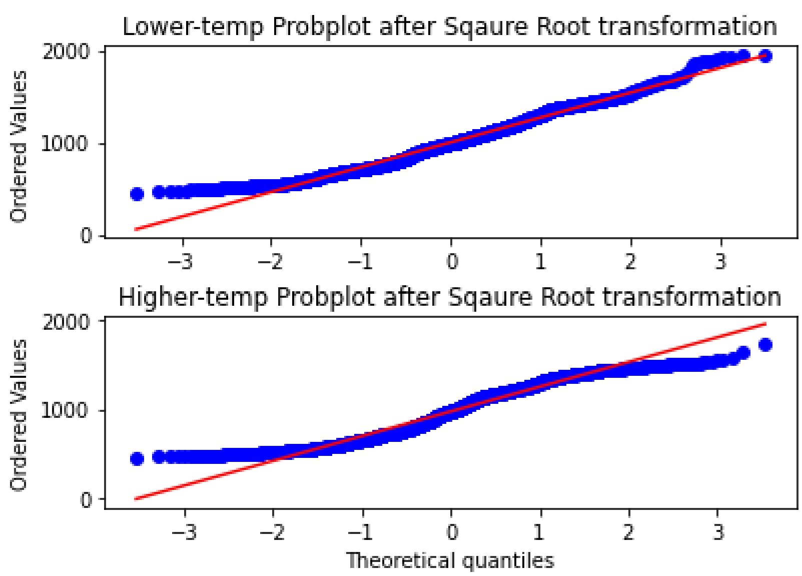


Figure 8. Probplots After Square Root Transformation

We now see that both distributions are less skewed. From the respective skewness: 0.316 (lower temperatures) & 0.059 (higher temperatures), we can also conclude that the transformations successfully normalize the data distributions. But we still have to test for the desired sample sizes of the two groups.

Effect size

Mean1 (lower temperature group) = 1074571.13199

Mean2 (higher temperature group)= 1022854.6112

Meandiff = -51716.52079

Std1 (lower temperature group) = 573840.21989

Std2 (higher temperature group) = 554921.77767

$Effect\ size\ = \frac{1074571.13 - 1022854.61}{\sqrt{\frac{573840.22^2 + 554921.78^2}{2}}}$ =0.3539039615. According to Cohen's D value, we had a small effect size and using the tt_solve_power with power 0.80 and an alpha value of 0.05; the desired sample size is 64.62. Even when raised the desired power to 0.95, the sample size needed was 105.69. Since our dataset has 6435 entries, we conclude that the company collected sufficient data for the effect sizes we set.

### Two-sided Two-sample T-test on Temperature

Then, according to our assumption statement, our team came up with the following hypotheses:

- H0: the mean weekly sale in lower temperatures is the same as the mean in higher temperatures.
- H$a$: the mean weekly sale in lower temperatures is different from the mean in higher temperatures.

We also needed to check if the population variances were equal in comparing the two samples. To do so, we used the Levene test with the null hypothesis: all input samples are from populations with equal variances. The p-value from the Levene test is 3.588e-10, which is less than the significance level of 0.05 (See Table 4 below). Therefore, our group would reject the null hypothesis and conclude that the lower and higher temperature groups do not have equal variance. With that information in mind, we proceeded to the two-tailed t-test with two samples. The p-value of the t-test is 2.98e-05, which is less than the significance level of 0.05, and our group concluded that a temperature change would affect the weekly sales for Walmart. It is informative to our research question because now we know that temperatures would affect the overall sales and different seasons have distinct temperature ranges; we could now begin the next step of the experiment to see if higher or lower temperatures will result in higher weekly sales.

Table 4. Summary of Equal Variance Check & Two-Sample t-test

| | P-value | Conclusion |
|---|---|---|
| The Levene test (Variances Check) | 3.588e-10 | Reject the null hypothesis |
| Two-sided two-sample t-test | 2.98e-05 | Reject the null hypothesis |
| One-sided two-sample t-test | 1.55e-05 | Reject the null hypothesis |

### One-sided Two-sample t-test on Temperature

During this step, we proposed that**:**

- H0: the mean weekly sale in lower temperatures is less than or equal to the mean in higher temperatures.
- H*a*: the mean weekly sale in lower temperatures is larger than the mean in higher temperatures.

After conducting the experiment, we reject the null hypothesis and conclude that the average weekly sales in lower temperatures are larger than in higher temperatures, given the p-value of 1.55e-05. There are several potential explanations for this result: firstly, during the months with lower temperatures, there are more occasions with huge discounts brought by several largest national holidays, including Halloween, Thanksgiving (Black Friday), and Christmas. The special discounts stimulate public consumption. In addition, people have high and stable demands to prepare different gifts during holidays. Moreover, the necessity in winter is generally more expensive than the product that is used in summer. For example, people are willing to spend a higher price to purchase coats, down coats, and so forth. Therefore, for Walmart, it is necessary to increase the stock during the lower temperature, timely adjust the discount policy following the coming of several large holidays, and further improve the return service or policy to prepare for the surge of consumption.

### Two-way ANOVA/Tukey's Test: Fuel Price and CPI Level with Weekly Sales

As a business with such a large scale that is fundamentally affected by the economics of the nation, it is important for Walmart to be able to predict the sales level ahead of the economic trends, and potentially develop strategies that can mitigate the loss of revenue due to downward economic trends and stimulate purchase from customers.

With that aim in mind, we adopt the Walmart retail dataset that records the weekly sales of 45 stores at different times of the year and the respective economic indicators. We picked two variables in our dataset that are the most accurate representation of holistic economic trends: CPI and Fuel Price. CPI, Consumer Price Index, refers to a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services[2]. Although fuel price only takes less than 6%[3] in the CPI, the volatility of fuel price is higher than the changes of CPI, becoming the main source driving the movement of CPI. As a large retail corporation providing a broad choice of consumer goods, Walmart has the demand to keep track of CPI and understand the impact of the interaction between CPI and fuel price on sales revenue.

CPI and fuel price have inherently different scales and ranges. To avoid bias because of the scale of data, we divide fuel price and CPI into four different levels using the 25%, 50%, and 75% percentile. In our dataset, we categorized the 4 levels of Fuel price as 1 to 4 and the 4 levels of CPI as a to d. Fuel level 1 refers to fuel price below $2.933 per gallon, level 2 refers to $2.933 - $3.445 per gallon, level 3 refers to $3.445 - $3.735 per gallon, and level 4 refers to all above $3.735 per gallon. CPI level a refers to CPI below $131.735, level b refers to CPI from $131.735 to $182.616, level c refers to CPI from $182.616 to $212.743, and level d refers to all above $212.743. We then append the average weekly sales to each row with the corresponding fuel price and CPI, and the result is as follows:

Table 5. Fuel level, CPI level, and Weekly Sales Dataset

| | Fuel_Level | CPI_Level | Weekly_Sales |
|---|---|---|---|
| **6430** | 4.0 | c | 713173.95 |
| **6431** | 4.0 | c | 733455.07 |
| **6432** | 4.0 | c | 734464.36 |
| **6433** | 4.0 | c | 718125.53 |
| **6434** | 4.0 | c | 760281.43 |

After fitting the data into the Two-way Anova model, we have the summary as the following:

[2] U.S. Bureau of Labor Statistics. (2022). CPI Home : U.S. Bureau of Labor Statistics. Bls.gov. https://www.bls.gov/cpi/

[3]Crawford, M., & Reed, S. (n.d.). PRICESSPENDING Measures of gasoline price change. https://www.bls.gov/opub/btn/volume-2/pdf/measures-of-gasoline-price-change.pdf

Table 6. Two-way ANOVA Output

```
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x7f71453b1f70>
                                 sum_sq      df          F        PR(>F)
C(Fuel_Level)              1.515813e+12     3.0   1.615221  1.835666e-01
C(CPI_Level)               1.931783e+13     3.0  20.584711  2.860658e-13
C(Fuel_Level):C(CPI_Level) 2.121444e+13     9.0   7.535236  4.701553e-11
Residual                   2.007981e+15  6419.0        NaN           NaN
```

From the table above, we conclude that there is no significant relationship between fuel price and weekly sales level given the p-value of 1.83566e-01. There is a significant relationship between CPI level and weekly sales, given the p-value of 2.860658e-13. We also conclude that the effect of changing CPI level depends on the level of fuel price, meaning there is an interaction between the two factors. This is also confirmed by the interaction plot below, in which the lines are not roughly paralleled, indicating the existence of interaction between CPI and Weekly Sales. Therefore, we identified CPI as the factor that has a significant relationship with the weekly sales of Walmart.

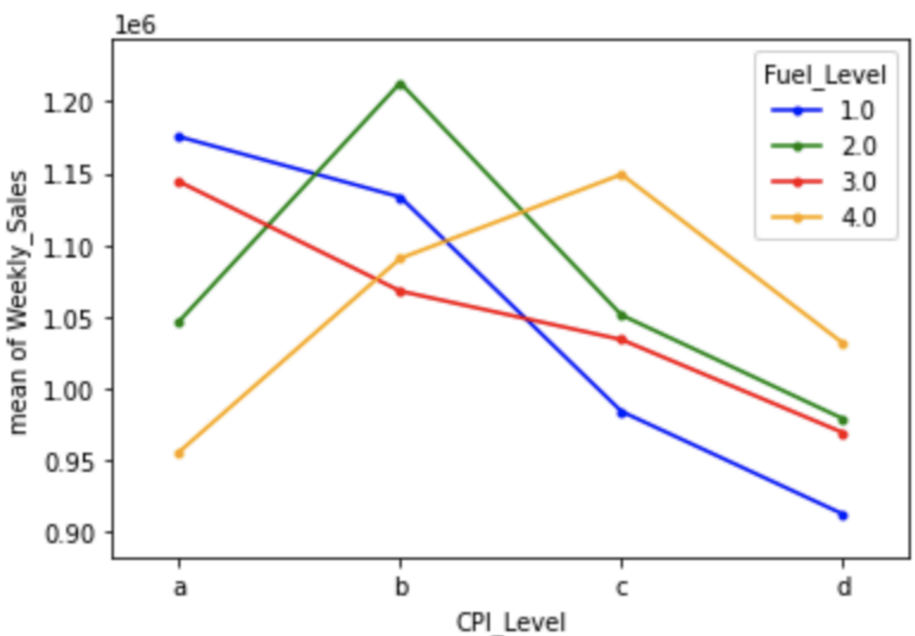


Figure 9. Interaction Plot of Two-way ANOVA

From the interaction plot above, we observed that sales achieve the highest in level 2 fuel price and level b CPI, which translates to 2.933 - 3.445 dollars per gallon fuel price and \$131.735 to \$182.616 CPI. This is a helpful indicator for Walmart for their financial predictions, in that revenue is going to increase if CPI and fuel price fall into these ranges. We also observed that average weekly sales start a decreasing trend for fuel levels 1, 2, and 3 when CPI level increases from b to d. This indicates that when CPI is above

$182.616 while the fuel price is $3.735/gallon, the weekly sales level decreases as CPI increases. We interpret this result as when people have higher purchasing power, which correlates to a higher CPI, they prefer shopping in higher-end stores rather than Walmart. In such situations, Walmart should adopt effective strategies to stimulate sales.

## Tukey's test for Two-way ANOVA

To have a more precise and detailed idea of which levels of CPI and Fuel price lead to the highest weekly sales, we performed a follow-up Tukey's test after the Two-way ANOVA test. Firstly, we examined whether different levels of Fuel price impact sales. The result is the following table:

Table 7. Fuel Price and Sales Tukey Output

| | group1 | group2 | Diff | Lower | Upper | q-value | p-value |
|---|---|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 26254.803659 | -24416.893125 | 76926.500443 | 1.882955 | 0.537837 |
| **1** | 1.0 | 3.0 | 1194.029070 | -49406.891416 | 51794.949556 | 0.085754 | 0.900000 |
| **2** | 1.0 | 4.0 | 10015.921850 | -40687.454279 | 60719.297979 | 0.717878 | 0.900000 |
| **3** | 2.0 | 3.0 | 27448.832729 | -23199.370736 | 78097.036194 | 1.969502 | 0.503489 |
| **4** | 2.0 | 4.0 | 16238.881809 | -34511.681844 | 66989.445462 | 1.162818 | 0.823640 |
| **5** | 3.0 | 4.0 | 11209.950920 | -39469.946575 | 61889.848415 | 0.803831 | 0.900000 |

As we can see from the table above, none of the p-values is smaller or equal to 0.05, which means that we don't have enough evidence to conclude that the weekly sales mean for different fuel price levels are different. This result further confirms our findings in the Two-way ANOVA of no significant relationship between fuel price and weekly sales.

Next, we investigated whether CPI impacts weekly sales. The results are shown in the table below:

Table 8. CPI and Sales Tukey Output

| | group1 | group2 | Diff | Lower | Upper | q-value | p-value |
|---|---|---|---|---|---|---|---|
| **0** | c | d | 54147.757428 | 3484.017595 | 104811.497261 | 3.884005 | 0.030758 |
| **1** | c | a | 34414.378516 | -16217.932176 | 85046.689209 | 2.470067 | 0.299839 |
| **2** | c | b | 93962.598405 | 43251.457654 | 144673.739157 | 6.733614 | 0.001000 |
| **3** | d | a | 88562.135945 | 37921.951495 | 139202.320394 | 6.355495 | 0.001000 |
| **4** | d | b | 148110.355834 | 97391.353563 | 198829.358105 | 10.612344 | 0.001000 |
| **5** | a | b | 59548.219889 | 8860.612492 | 110235.827286 | 4.269368 | 0.013580 |

The p-values smaller than 0.05 tell us that different levels of CPI impact weekly sales, which is in accordance with the result from the Two-way ANOVA test. By interpreting the results, we can see that the CPI level b outperforms all other levels, followed by level a, level d, and lastly, level c. More specifically, CPI level between \$131.735 and \$182.616 (level b) results in about \$59.5K higher weekly sales than CPI level a (below \$131.735); CPI level of below \$131.735 results in \$88.5K higher weekly sales than CPI level d (above \$212.743); CPI level d (above \$212.743) results in \$54.1K higher weekly sales than CPI level c (from \$182.616 to \$212.743).

We also conducted Tukey's test for the interaction term. As additional support for the previous conclusion from Two-way Anova, Table 9 records more detailed information about how much each fuel price and CPI level affects weekly sales. Since there are 33 significant combinations, we selected three that are interesting to us to analyze, which are shown below:

Table 9. Tukey Output with CPI and Fuel Price

| group1 | group2 | Diff | Lower | Upper | q-value | p-value |
|---|---|---|---|---|---|---|
| (1.0, d) | (1.0, a) | 263059.533937 | 91274.477063 | 434844.59081 | 7.422859 | 0.001 |
| (2.0, d) | (2.0, b) | 234119.489948 | 99609.237416 | 368629.74248 | 8.436936 | 0.001 |
| (3.0, d) | (3.0, a) | 174949.629129 | 47345.381608 | 302553.876649 | 6.645849 | 0.001 |

The first row indicates that at fuel price below \$2.933 per gallon, CPI level a (below \$131.735) increases sales by \$263K compared to CPI level d (above \$212.743). The second row shows that at fuel price level 2 (between \$2.933 and \$3.445 per gallon), CPI level b (between \$131.735 and \$182.616) increases sales by roughly \$234K than CPI level d (above \$212.743). The last row demonstrates that at fuel prices between \$3.445 and \$3.735 per gallon, a CPI level below \$131.735 increases sales by about \$175K than CPI level d (above \$212.743). We chose these three specific rows because they represent the comparison between the highest and lowest sales of a given fuel price level.

### Heterogeneity: Customer's Preference for Check-out Method

After analyzing the impact of seasonal and economic factors on sales, we then turned to answer the rest of the research questions by analyzing controllable factors. Our team was curious to know how types of check-out lanes would affect Walmart's sales of customer volume. But to answer this question, our team sought out supplemental datasets and found one relating to the customer survey. In the survey, customers were asked which type of checkout they would prefer, and their responses were recorded along with their respective ages and employment status. But before we jumped into experiments, we did realize that given the highly subjective customer perception, the costs of switching check-out methods might outweigh the benefits[4]. In order to provide more informed suggestions, we needed to carefully study if the change of checkout method (the treatment) worked the same way for different groups of customers. More specifically, we wish to determine if there is a difference between the attitude regarding self-checkout and cashiers for people younger than 60 and older than or equal to 60. The age threshold is 60 since it is roughly when retirement benefits start (62, according to the social security office.)

In the checkout preference survey dataset, we employed variables of interest: Retired and Preference. To check whether the two groups would react differently to the change in the checkout method, we used the Chi-Square test. Before the testing, we first checked if the test's assumptions were satisfied:

- Variables are categorical:

  Yes, the variables are categorical since the responses are check-out option preferences. It is either cashier or self-checkout, representing a category, not quantity.
- Data come from independent random samples:

  According to the survey data presented in the dataset, we can assume that the data points are collected & recorded independently:
- All expected cell counts should be five or greater.

As Table 10 shows, all expected cell counts are larger than 5.

[4] Wal-Mart Boosts Self-Checkout, But Its Claimed Cost Savings Don't Add Up. (n.d.). FierceRetail. https://www.fierceretail.com/operations/wal-mart-boosts-self-checkout-but-its-claimed-cost-savings-don-t-add-up

Table 10. Expected cell counts for heterogeneity

| | Cashier | Self-checkout |
|---|---|---|
| Younger than 60 | 178.68 | 176.32 |
| Older than or equal to 60 | 48.32 | 47.68 |

Since all the requirements for the Chi-squared test are met, and our team proceeded to run the test based on the following hypotheses:

- H0: The populations are homogeneous with respect to the variable.
- Ha: The populations are not homogeneous with respect to the variable.

After running the chi-squared test, we obtained a p-value of 0.68. With a significance level of 0.1, we failed to reject the null hypothesis. It was good news for the future operation of Walmart since now our group could assure Walmart management that customers will have similar behaviors in response to a treatment, whatever it might be. In our case, using the survey results on preferences regarding checkout lanes, Walmart could proceed to switch checkout methods and evaluate the efficacy of the treatment without worrying about the inherent disparity between control and treatment customer groups.

### Regression analysis: Reward program's effect on Sales

For the regression analysis, we try to test how the reward program will impact consumer spending amount. We used the dataset that records every purchase of a customer, the transaction amount, as well as other descriptive data of the customer, including age and annual income, as below:

Table 11. Walmart Reward Program Dataset

| | Date | Customer ID | Age | Shop_Online | Annual_Income | Amount_Spent | Reward | Online |
|---|---|---|---|---|---|---|---|---|
| 0 | 5/29/21 | 1 | 33 | Yes | 124 | 22 | 1 | 1 |
| 1 | 8/19/20 | 2 | 48 | No | 115 | 0 | 0 | 0 |
| 2 | 10/7/19 | 3 | 50 | Yes | 115 | 95 | 0 | 1 |
| 3 | 10/29/20 | 4 | 37 | No | 93 | 39 | 0 | 0 |
| 4 | 6/11/20 | 5 | 67 | No | 137 | 0 | 0 | 0 |

As shown in the table above, we created a separate column named "Reward" with values 0 and 1, corresponding to no reward and reward. Our Y variable is amount_spent, which represents the dollar value of the corresponding purchase. After running the regression model, the result summarize into the following table as follow:

Table 12. Regression Output with Reward as Independent Variable

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           Amount_Spent   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                  0.005365
Date:                Sat, 03 Dec 2022   Prob (F-statistic):              0.942
Time:                        19:58:16   Log-Likelihood:                -4597.5
No. Observations:                 933   AIC:                             9199.
Df Residuals:                     931   BIC:                             9209.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         27.0063      1.326     20.365      0.000      24.404      29.609
Reward         0.1722      2.350      0.073      0.942      -4.440       4.785
==============================================================================
Omnibus:                      165.855   Durbin-Watson:                   2.040
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              130.317
Skew:                           0.818   Prob(JB):                     5.03e-29
Kurtosis:                       2.179   Cond. No.                         2.42
==============================================================================
```

From our regression output, we could see that the constant value (the intercept) is 27.00, which indicates the average amount spent without the reward program. And the ATE for our experiment is 0.172, meaning that introducing a reward program can potentially have a positive effect on the customer's amount spent in Walmart. From those two results, the regression equation is:

$Y = 27.006 + 0.172x$, where Y is the average amount spent for each customer and x is the indicator of whether a reward program is introduced or not. But there are still some pitfalls of our regression analysis that we should watch out for.

The p-value for the regression model is 0.942, which is larger than the significant value of 0.05; we conclude that the model is not statistically significant. Several reasons may lead to this result. First of all, the reward might not be sensitive enough to stimulate consumers to spend more money because of Walmart's "Everyday low pricing" strategy. Customers are already getting good deals from the merchant, so adding reward benefits won't be as appealing as it may sound. Moreover, the factor of offering rewards to customers has no strong correlation with customer consumption because households normally treat Walmart as their go-to grocery shopping location. They may have their fixed routines on what to purchase during a visit to one of the Walmart supercenters, meaning a consistent consumption level for each of their Walmart visits. A reward program might not be enough to push households to add additional items to their routine shopping cart.

In addition, the R-squared statistics for our regression model is only 0.000, meaning that the model cannot account for any variability in the outcome of our model. Our team decided to add a covariate to try to improve the regression model and help Walmart make better business decisions.

Table 13. Summary of regression output with Reward and the covariate Annual_Income

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           Amount_Spent   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.7877
Date:                Sat, 03 Dec 2022   Prob (F-statistic):              0.455
Time:                        20:03:21   Log-Likelihood:                -4596.8
No. Observations:                 933   AIC:                             9200.
Df Residuals:                     930   BIC:                             9214.
Df Model:                           2
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const            31.1642      3.573      8.721      0.000      24.152      38.177
Annual_Income    -0.0418      0.033     -1.253      0.211      -0.107       0.024
Reward            0.1744      2.350      0.074      0.941      -4.437       4.786
==============================================================================
Omnibus:                      161.059   Durbin-Watson:                   2.033
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              129.817
Skew:                           0.819   Prob(JB):                     6.47e-29
Kurtosis:                       2.192   Cond. No.                         348.
==============================================================================
```

After adding the covariate, the annual income of customers, the p-value of the model decreases from 0.942 to 0.455, which is still larger than the significant value. Annual income is a good control for this experiment, considering that customers' income will not be changed because of the reward program(see Figure 10). One useful insight from our second regression model is that customers' annual incomes might have a negative impact on their consumption in Walmart, meaning that people with relatively low incomes might be more inclined to shop in Walmart when a reward program is present. Another possible explanation is that income level does not impact customers' purchases in Walmart, since people with higher income might be purchasing similar items from Walmart compared to people with lower income, because Walmart is selling necessities which are presumably not affected by income level. Nevertheless, Walmart needs to carefully evaluate the feasibility of targeting the low-incomed customer group since our model is not statistically significant.

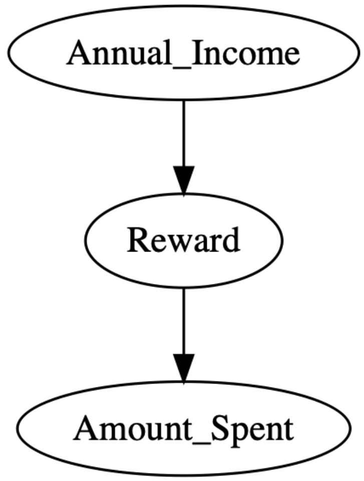


Figure 10. Causal diagram of Regression with Covariate

**DoWhy Analysis: Causal Relationship between Reward and Sales**

In order to further study the causal relationship between reward and the amount of consumption, we chose to employ DoWhy. To begin with, we made the initial assumption shown below:

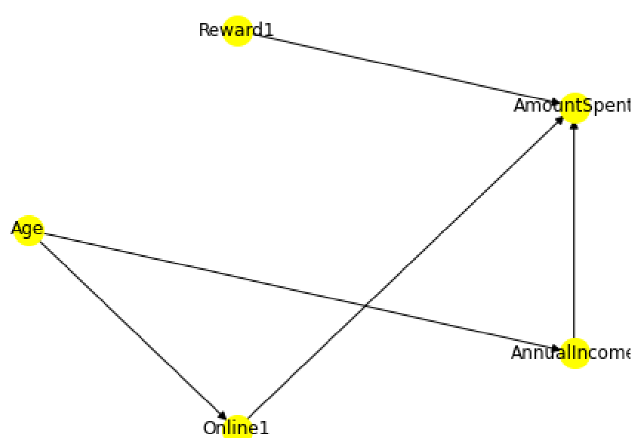


Figure 11. Causal diagram of DoWhy analysis

Based on our causal diagram, we do not have a confounder. Therefore, we used linear regression to estimate the causal relationship rather than propensity-score matching. After running the model, DoWhy provides the estimated effect of 0.5001488, meaning that giving rewards leads to an increase of approximately $0.5 on the amount spent.

In order to confirm the validity of our assumption, we conducted two refutation tests: placebo treatment and data subsets validation.

Table 14. Summary of validation check using linear regression

| Estimated Effect:0.500 | | |
|---|---|---|
| | Placebo Treatment | Data Subsets |
| New Effect | 0.273 | 0.448 |
| P value | 0.447 | 0.484 |

For placebo treatment, through replacing our treatment with a randomly generated variable, we have the new effect of 0.273. Since our original estimate is already very close to zero, this value doesn't necessarily mean that our assumption is valid. To interpret the result more accurately, we compared our p-value with a significant value of 0.05. Since the p-value is larger, we can conclude that there is no problem with the estimate. For the data subsets validation, with a randomly selected subset, we have the new effect value of 0.447, which is close to our estimated effect. Same with the placebo treatment, we look at the p-value to further confirm. Given the higher p-value of 0.484, we can conclude that this refutation test proves that our estimate has no problem and that our assumption is valid.

To sum up, for Walmart, although both regression models are not significant regarding the relationship between reward and the amount of consumption, from DoWhy, we can conclude that having the reward does help increase the spending amount. Walmart can increase the amount of spending of customers by promoting more rewards.

## IV. Conclusion

With the mentioned analysis of Walmart's sales, we conclude that the overall sales of Walmart stores are related to the overall economic trend: sales are likely to decrease as people have increasing purchasing power. Within every year, we conclude that revenue is higher when the weather is relatively cold. Regarding the impact of reward programs on customers' consumption, though the result of regression models is not statistically significant, our DoWhy analysis confirms the causal relationship between rewards and the amount spent and suggests that having rewards will increase customers` spending by a small amount. This conclusion is consistent with the results from our two-way ANOVA analysis that as CPI increases, people actually purchase less from Walmart because they would prefer to buy higher-end products. Therefore, money incentives making products cheaper might not be very effective for Walmart customers. In addition, it reflects that most goods from Walmart are necessities, meaning that customers form the habit of visiting retail stores periodically to stock up whether there is a reward or not. Therefore, in order to utilize the reward policy as a marketing strategy, it is necessary for the company to study the impacts further. In addition, with heterogeneity analysis, we are able to assure Walmart's management that they can proceed with any future experiments and successfully evaluate the correct efficacy of treatments without worrying about the different behaviors of different experimental groups.

However, we understand that there are limitations to our research, and we believe this is where future research can take into consideration and make corresponding adjustments.

1. Walmart can collect data on a larger scale. With a larger number of sample data, the result of each experiment can be more representative, yielding a more reliable conclusion.
2. We did not conduct competitor analysis and took the information in our analysis. In order to develop sustainably, it is significant for the company to survive, maintain a stable cash flow and grow in an industry that is filled with fierce competition. From a larger picture perspective, it is necessary to analyze the business strategy of the rivals and how they react facing different situations. More meaningful research questions can be studied and answered by combining the available data from competitors. For example, why certain strategy has different effects on each company`s sales differently.

For future research, limitations and suggestions as listed above can be taken into consideration.
However, despite those foreseeable limitations, our group still believes that with data collected from multiple dimensions and across time series, our group is confident that Walmart can easily leverage data from all sources and experiment with an economic cost-performance ratio due to its established economy of scale & scope. In addition, other approaches to stimulate sales could be considered and experimented with that are more suitable for Walmart's positioning in the market. Reward programs might not be the most fitting approach for Walmart's supercenters with large-scale and all-encompassing commodities.

## V. References


*U.S. Bureau of Labor Statistics. (2022). CPI Home : U.S. Bureau of Labor Statistics.* Bls.gov. https://www.bls.gov/cpi/

*Crawford, M., & Reed, S. (n.d.). P R I C E S A N D S P E N D I N G Measures of gasoline price change.* https://www.bls.gov/opub/btn/volume-2/pdf/measures-of-gasoline-price-change.pdf

*Wal-Mart boosts self-checkout, but its claimed cost savings don't add up*. FierceRetail. (2012, March 8). Retrieved December 8, 2022, from https://www.fierceretail.com/operations/wal-mart-boosts-self-checkout-but-its-claimed-cost-savings-don-t-add-up