

Walmart Sales Prediction

Lachandra Ash

November 11, 2022

Topic

Walmart is a multinational corporation that manages a chain of supermarkets, chain stores, department stores, bargain warehouses, and small grocery stores (GlobalData, 2022). Food and commodities, health and fitness technology, workplace, and recreation, hardlines, fashion and housing, and much more are all available at affordable costs in the company's retail locations (GlobalData, 2022).

In addition, the company runs warehouse clubs under the Sam's Club brand. Equate, Bonobos Fielder, Mainstays, George, Onn, Parent's Choice, Time and Tru, Wonder Nation, and No Boundaries are just few of the private labels and licensed brands that Walmart sells (GlobalData, 2022).

Business Problem

Walmart needs to increase their profit through their product sales to each consumer. Walmart wants to increase their sales during the super bowl event, labor day, thanksgiving, and Christmas.

Background | History

First Walmart store opened in Rogers, Arkansas on July 2, 1962 (Walmart, 2022). Sam Walton founded the company. The Walton family has 24 retail outlets, grossing around \$12,700,000 (Walmart, 2022). Wal-Mart Stores, Inc. is the legal name of the corporation. Mr. Sam launches his plan to expand Walmart across the country in the 1970's (Walmart, 2022).. Eventually, Walmart will be listed on public stock markets. The first stock offering was priced at \$16.50 for every unit (Walmart, 2022).

Data Explanation

I obtained the Walmart dataset from the Kaggle website. The data involves Walmart sales during the holiday sales event. The dataset features are:

- Store: The store number
- Weekly_Sales: Sales for the given store
- Holiday_Flag: Whether the weeks is a special holiday week. 1=Holiday week 0=Non-holiday week
- Temperature: Temperature on the day of sale
- Fuel Price: Cost of fuel in the region.
- CPI: Consumer Price Index.
- Unemployment: Unemployment rate
- Month: The month of sales
- Year: The year of sales (Yasser, n.d.).

Methods

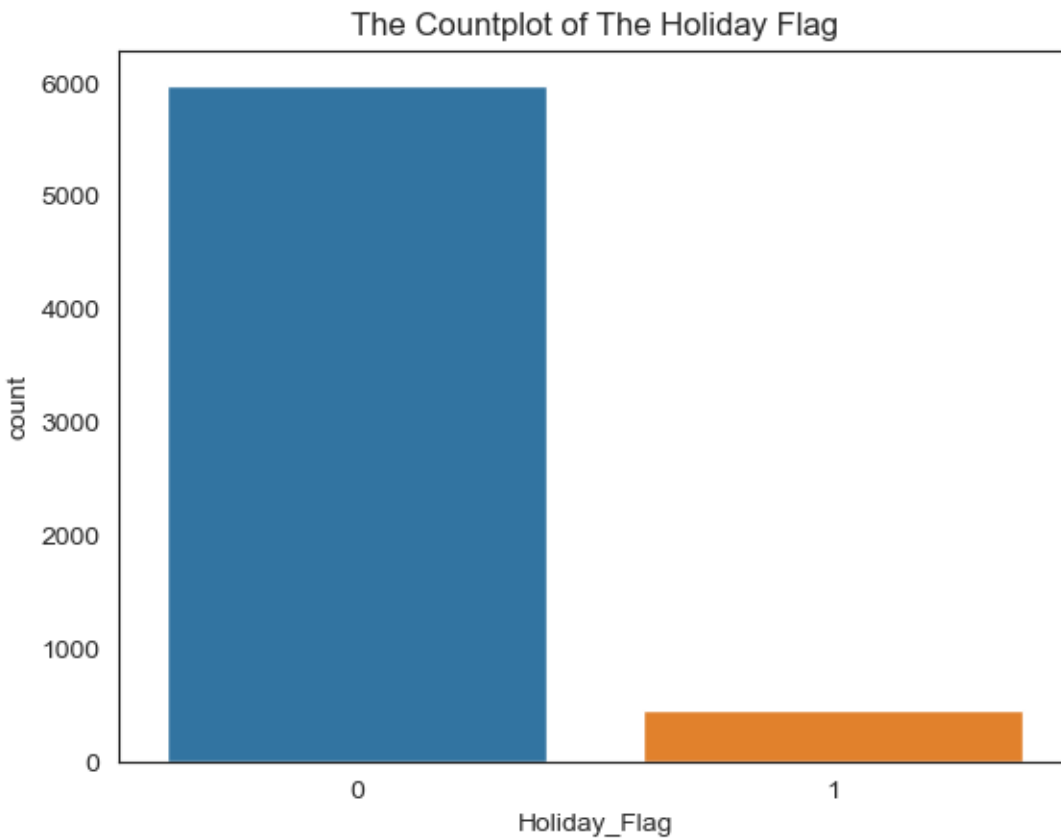
I evaluated the dataset with Jupyter Python Notebook 3. I preprocessed the data by importing the libraries and modules, and Walmart dataset. After I imported the dataset into the dataframe, I viewed the first five records. I used isna to check the dataframe for missing values, and there

were zero missing values. I checked the dataframe for duplicates and there were no duplicates. I dropped the date column and created year and month columns.

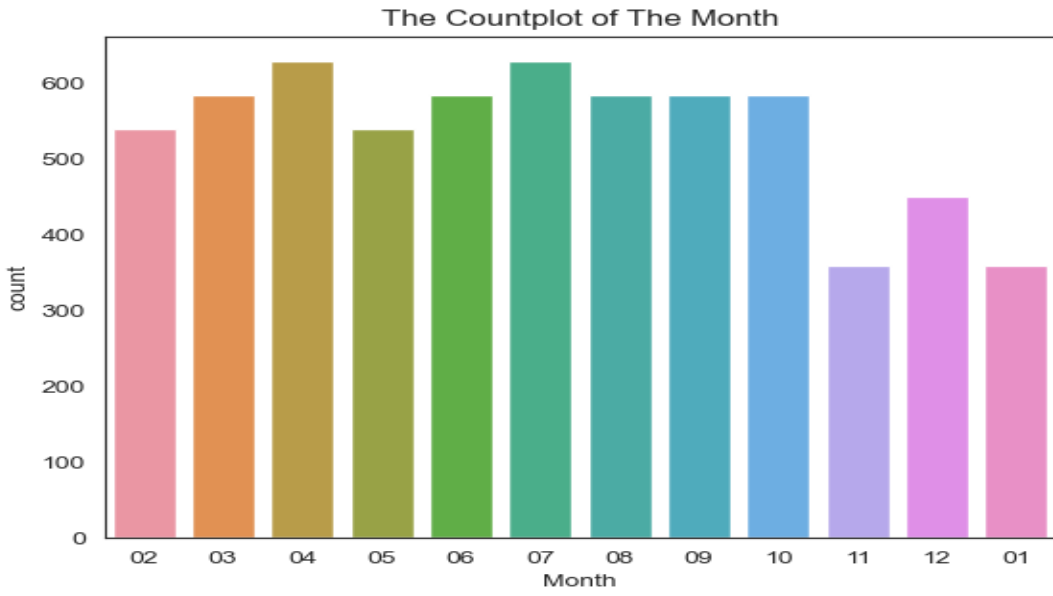
I explored and analyzed the dataframe by examining the index, dtypes, information, shape, description, sum, corr, cov, count, and unique values.

Analysis

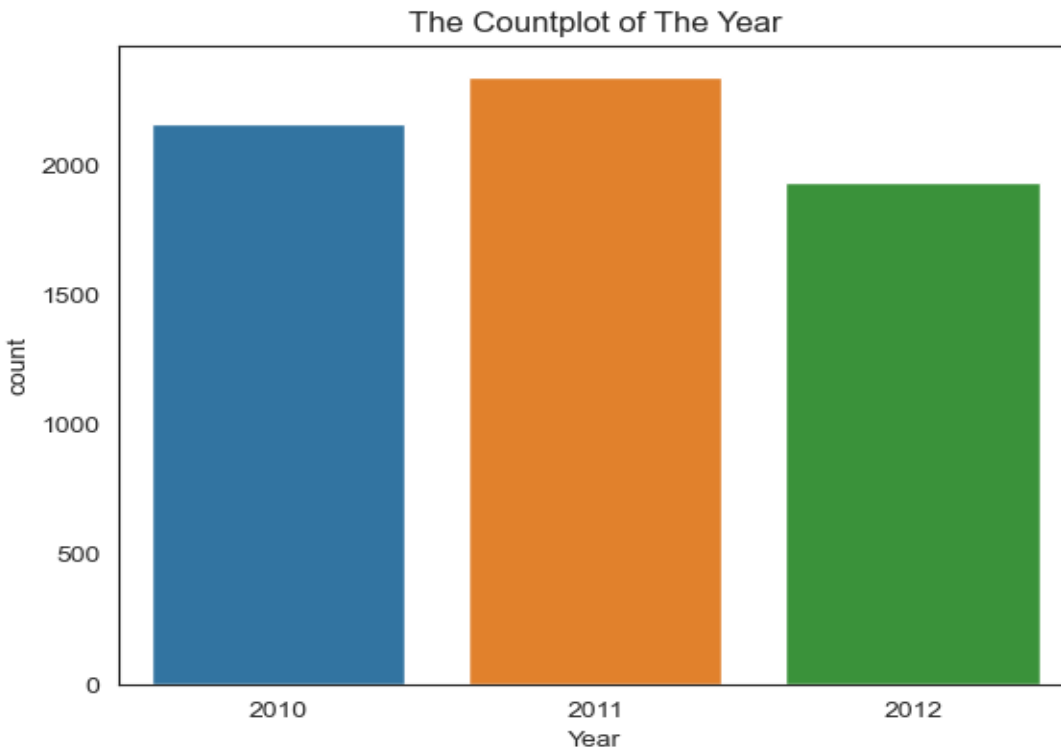
1).Holiday_Flag: There are more non-holiday weeks versus holiday week.



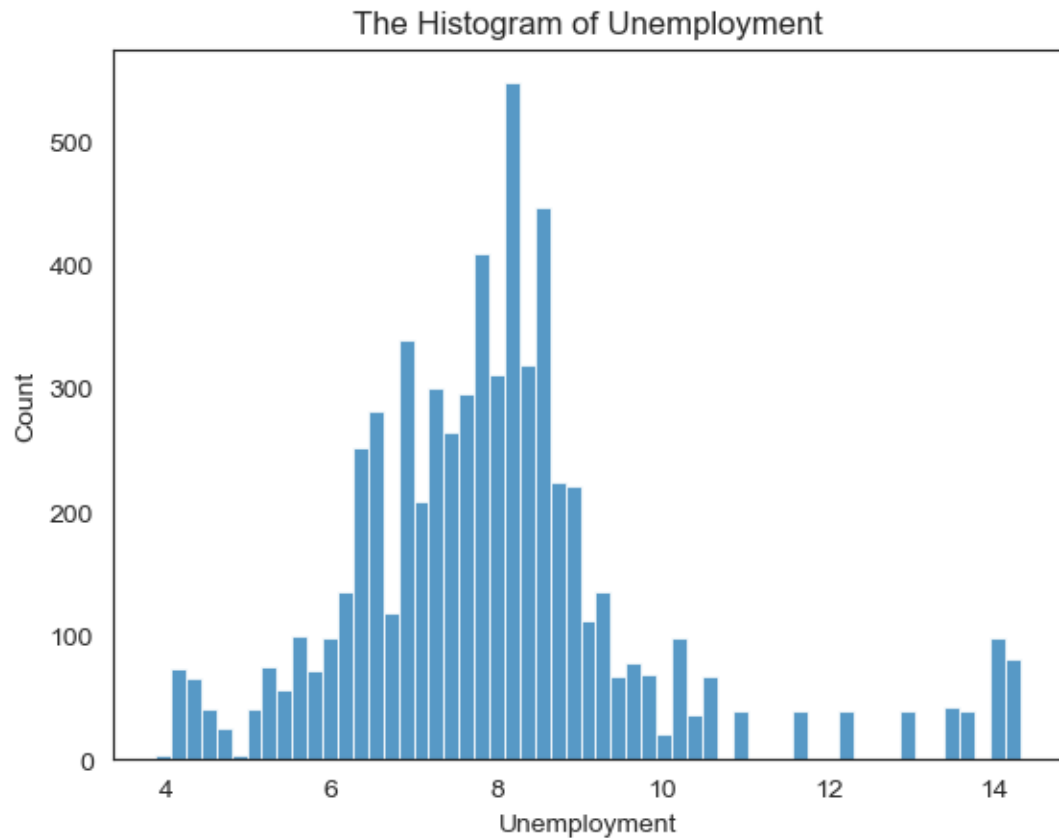
2) Month: Month 04 and Month 07 had the highest number of sales. Months 02, 03, 05, 06, 08, 09, and 10 had the second highest number of sales and the months sales were near the same number of sales. Months 11, 12, and 01 sold the least number of sales.



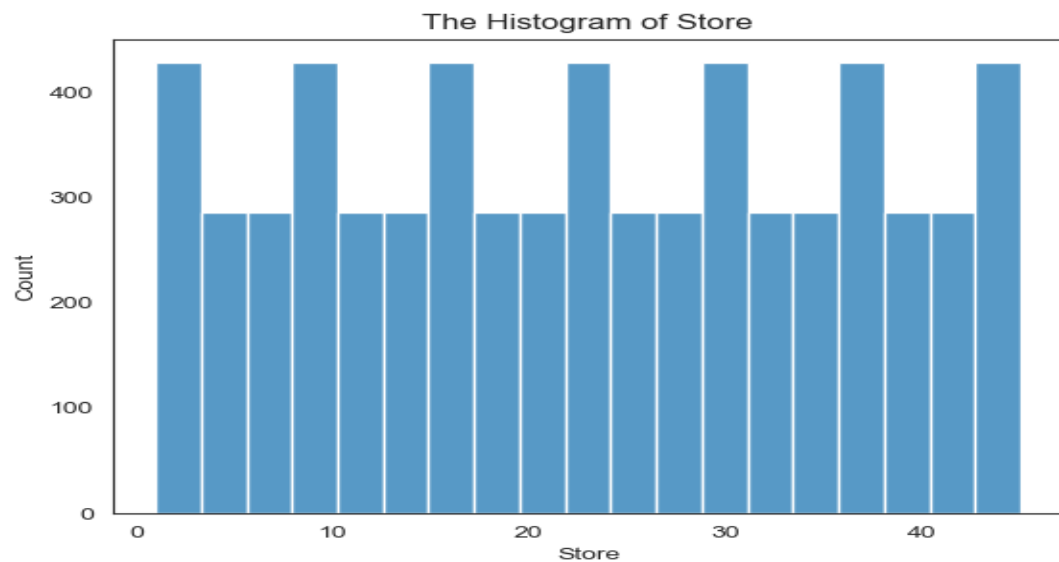
3)Year: The 2011 year has the highest number of sales, 2010 has the second highest number of sales, and 2012 has the least number of sales.



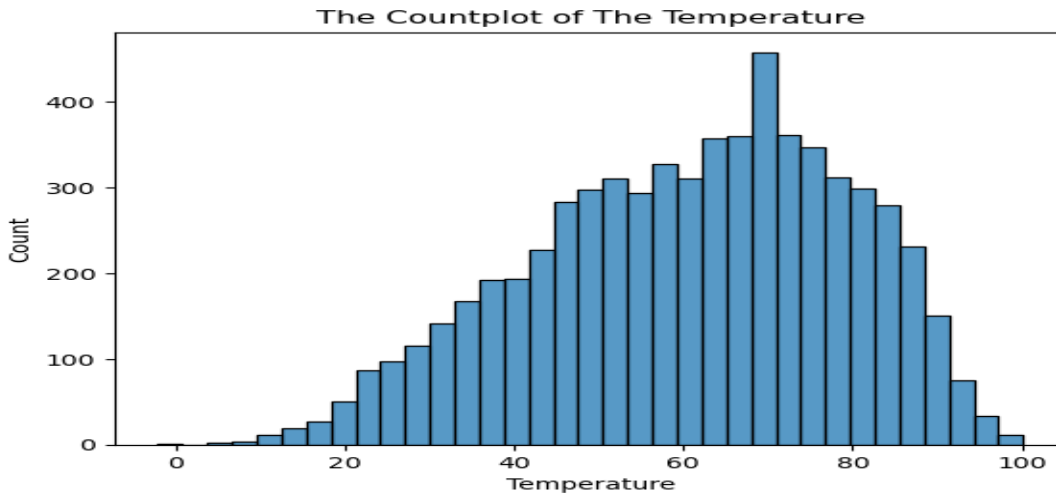
4)Unemployment: The employment rate of 8 had the highest number of employment rates. The least number of employment rates are between 1-6 and 9-10.



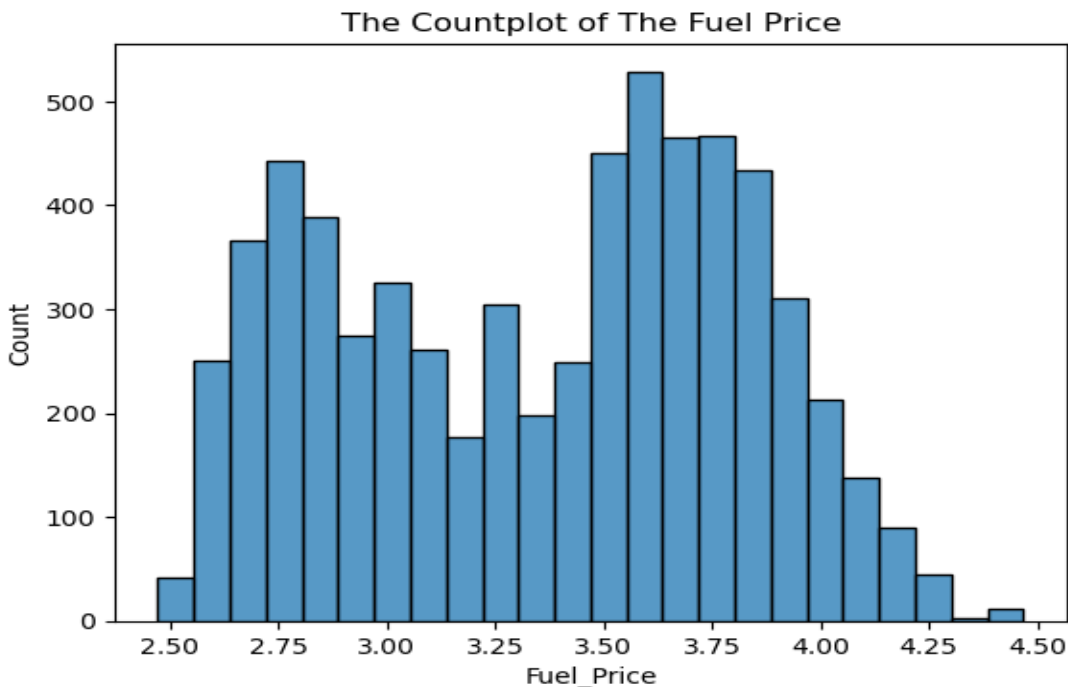
5)Store: There are forty-five stores within the dataframe.



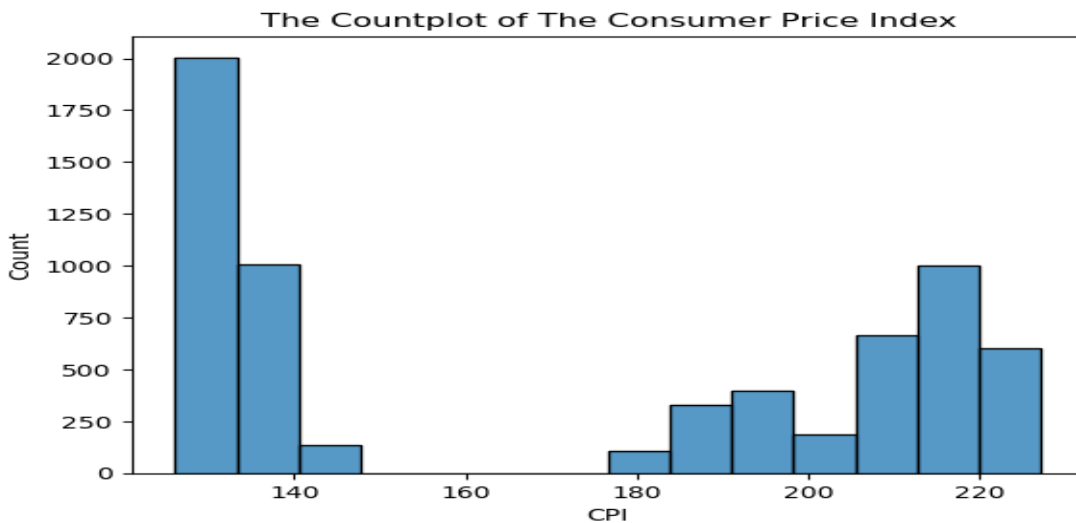
- 6) Weekly Sales: The seventy-degree temperature had the highest number of days with sales. As the temperature raised from 0 to 70 degrees, the number of sales began to increase. The warmer the temperature after 70 degrees, the sales began to drop.



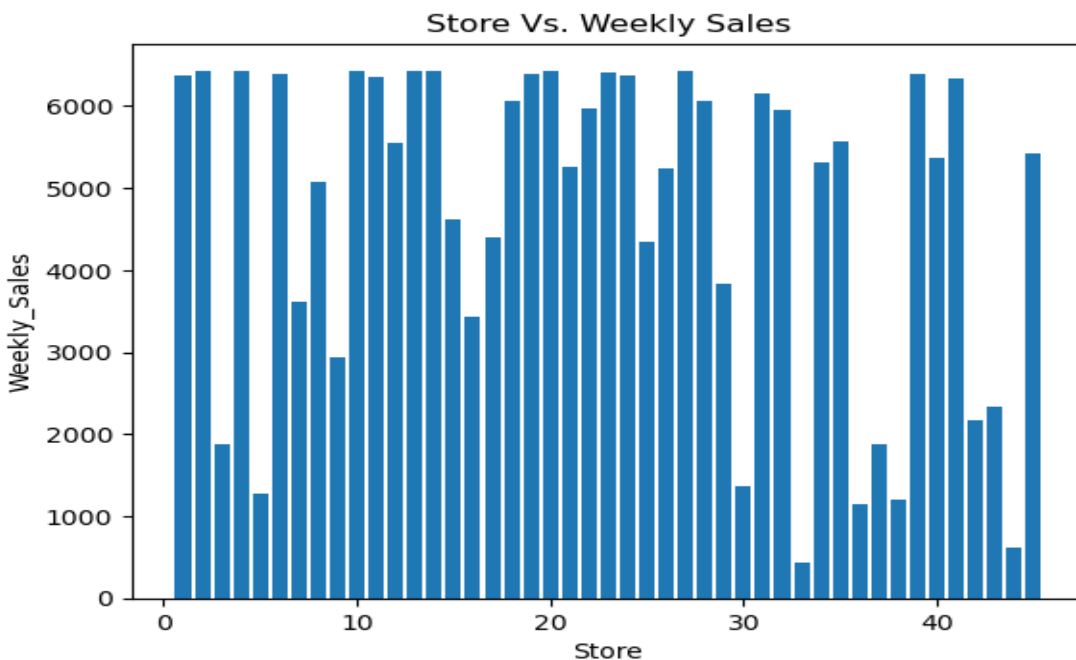
7) Fuel Price: The fuel prices with the highest count of consumers are approximately between \$3.50 and \$3.75. The more expensive the fuel prices became, the number of consumers who brought the fuel began to decrease. The fuel prices with the second highest count of consumers are approximately between \$2.60-\$3.10.



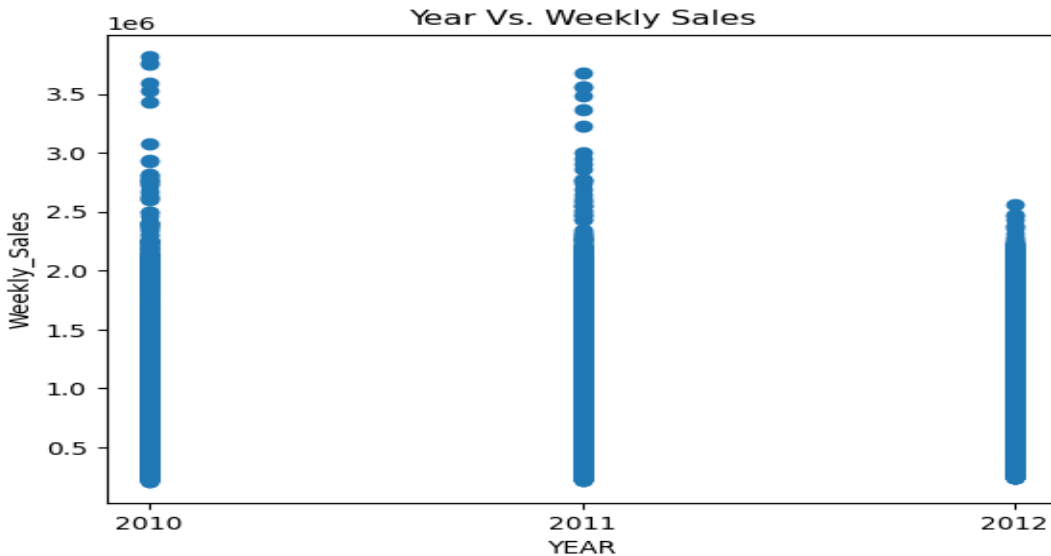
8)Consumer Price Index: The consumer price index ranges from 140 to approximately 120 has the highest number of consumers. the highest count of consumers is below the number of 140. The price index range with the lowest number of consumers is approximately between 140-205.



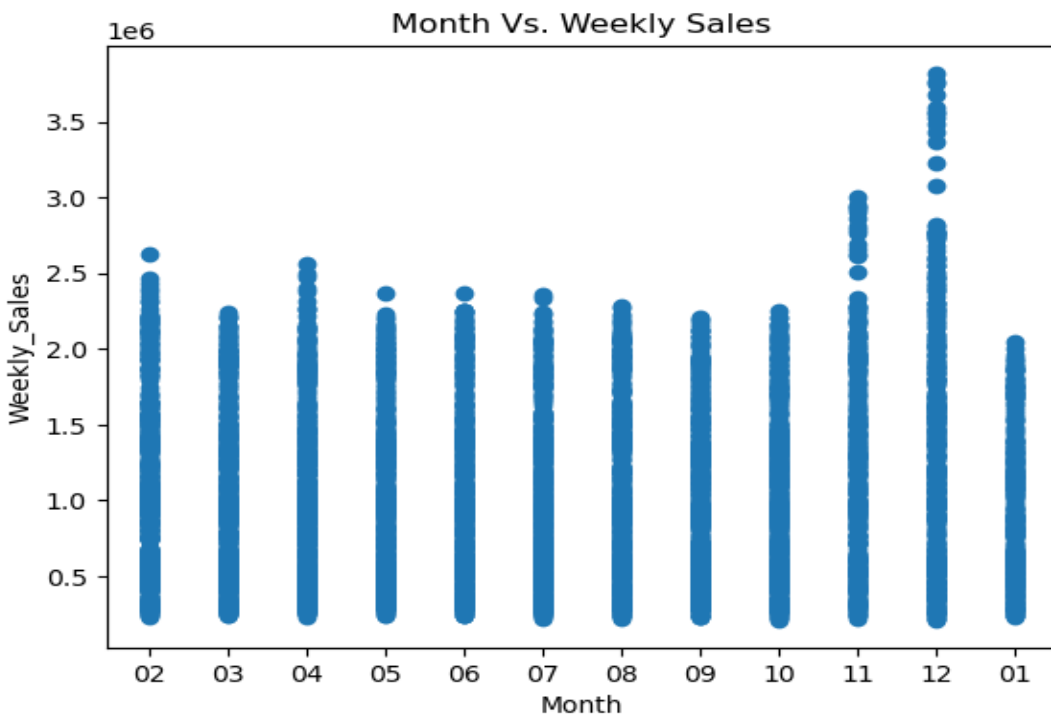
9).Stores versus weekly sales: Stores 1,2,4,6,10,11,13,14,18,19,20,21, 23, 24, 25, 28, 29, 32, 33, 40, and 45 had the highest number of weekly sales.



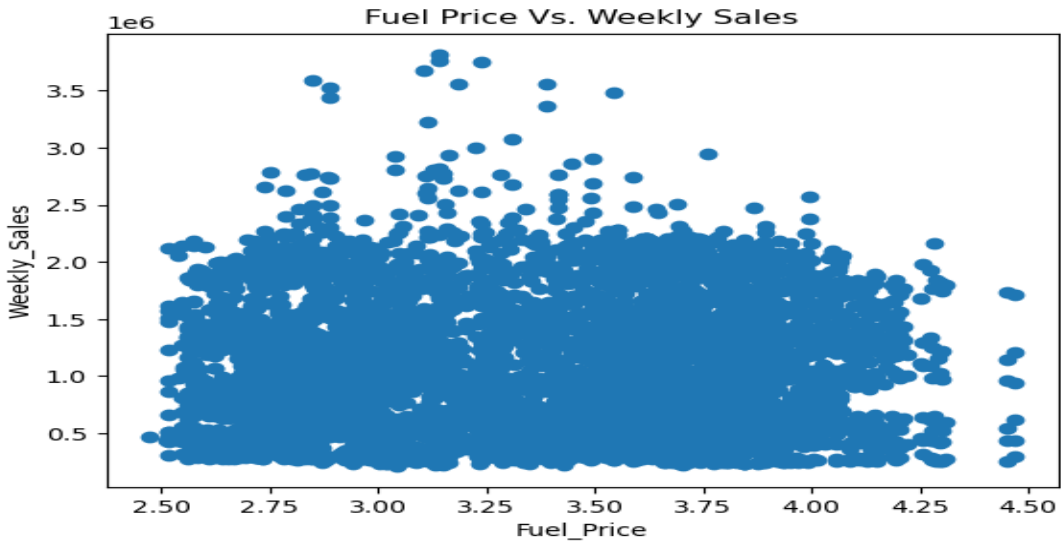
10) Year versus weekly sales: Year 2010 had the highest number of sales and year 2012 had the least number of sales. Year 2011 had the second highest number of sales.



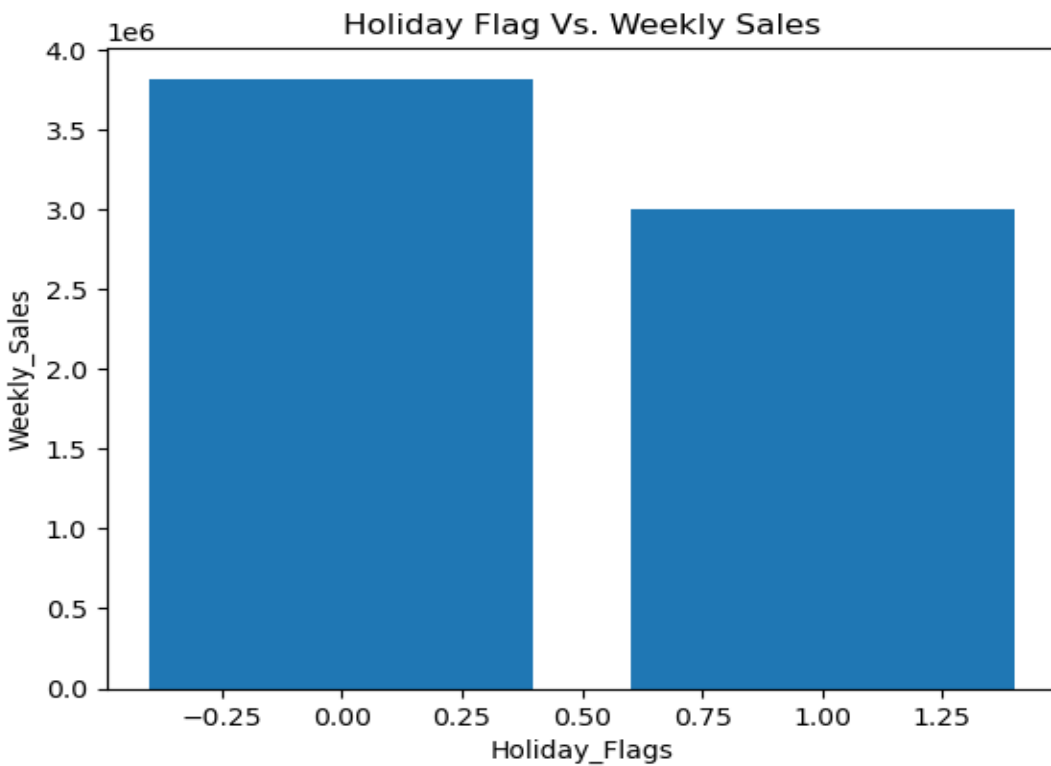
11) Month versus weekly sales: Months 12 and 11 had the highest number of weekly sales. Months 03 and 01 had the least number of weekly sales. Month 02, 04, and 05-10 had approximately same number of weekly sales.



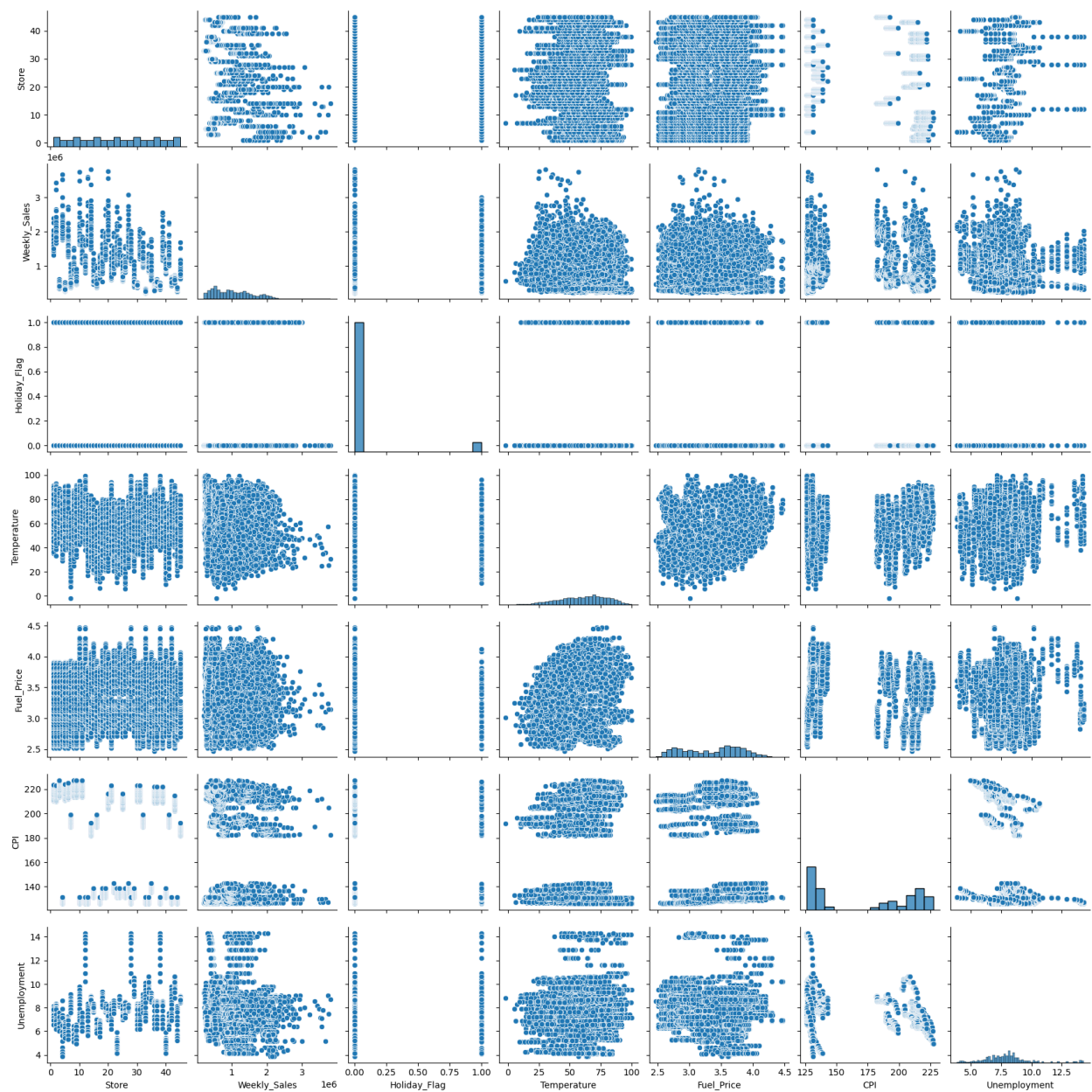
12)Fuel prices versus weekly sales: Fuel Prices that are approximately between the ranges of \$2.75- \$4.00 have near the same number of weekly sales.



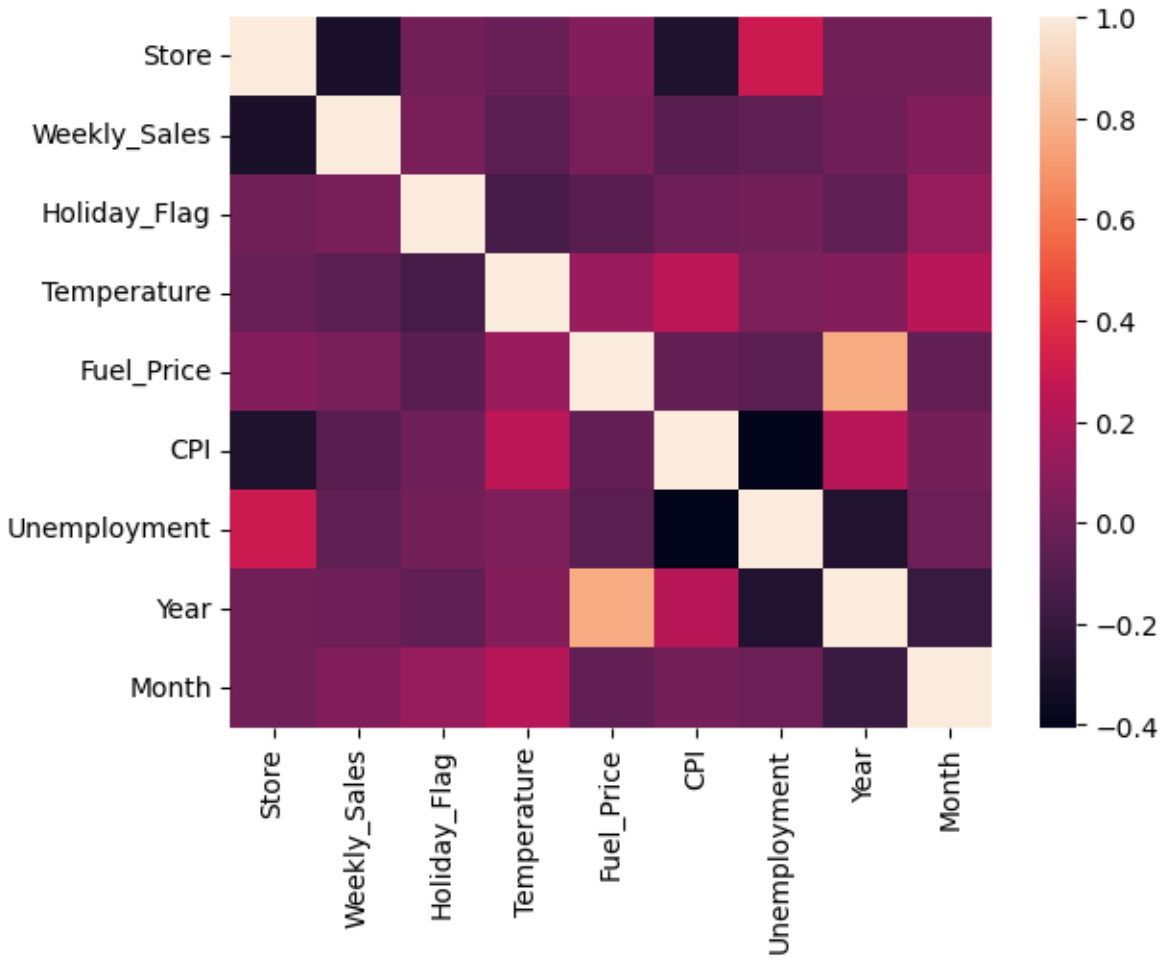
13)Holiday flag versus weekly sales: The holiday weeks have the least number of weekly sales versus the non-holiday weeks.



14)Pairplot: Pairplot of the dataframe's features.



15)Heatmap: Heatmap correlation of the dataframe.



The Models

The three models that were used for the accuracy score are DecisionTreeClassifier, Logistic Regression, and Support Vector Machine Accuracy. The holiday flag column was dropped before fitting the model. The data was trained, tested, and split.

View Appendix: Table 1: Model Accuracy Scores

Each model was trained, and the Decision Tree accuracy score was 94.8%, Logistic Regression accuracy score was 93.3%, and the Support Vector Machine accuracy score was 93.3%.

Conclusion

There was more non-holiday versus the holiday weeks. The non-holiday weeks had more weekly sales versus the holiday weeks. More consumers purchase items from Walmart during the spring and summer months. The least number of consumer purchases took place during the

Thanksgiving and Christmas holidays. Walmart can increase their number of weekly sales during the holiday weeks by running ads and discounting products.

Since there are less holiday weeks, the Walmart stores will always experience the least number of weekly sales versus the non-weekly sales. The three models produced near the same accuracy score.

References

GlobalData (n.d.), Walmart Inc: Overview. Retrieved from [Walmart Inc Company Profile - Overview - GlobalData](#), on November 11, 2022.

Walmart (n.d.). From Humble Beginnings. To Redefining Retail. Retrieved from [Walmart History](#), on November 11, 2022.

Yasser, M. (n.d.). Walmart Dataset. Retrieved from [Walmart Dataset | Kaggle](#), on November 11, 2022.

Appendix:

Table 1: Model Accuracy Scores

```
In [77]: print("Support Vector Machine Accuracy:", svm_model.score(X_test, y_test))
print("Decision Tree Accuracy:", dec_model.score(X_test, y_test))
print("Logistic Regression Accuracy:", log_model.score(X_test, y_test))
```

```
Support Vector Machine Accuracy: 0.9326773692387365
Decision Tree Accuracy: 0.9487312273433454
Logistic Regression Accuracy: 0.9326773692387365
```
