**Abstract**

This study aims to analyse a comprehensive Costco retail dataset using SAS and R software to delve into the factors that significantly influence weekly sales and evaluate their impact on sales performance. The analysis seeks to provide a comprehensive understanding of the key drivers of weekly sales at Costco by examining the relationships between independent variables and sales. The study aims to offer valuable insights for strategic decision-making within the organization, contribute to the field of retail analytics, and provide practical implications for improving sales performance in the retail industry.

To achieve these objectives, the study will employ a diverse range of variables such as store location, date, holiday promotions, temperature, fuel price, consumer price index (CPI), and unemployment rate. These variables are carefully selected to represent various aspects that could impact sales performance. By analysing the dataset, the study aims to uncover patterns, trends, and insights that can guide decision-making within Costco.

The analysis will be conducted using both SAS and R software, enabling a comprehensive evaluation of their respective strengths and weaknesses in handling and analysing large-scale retail datasets. The capabilities of each software package will be assessed in terms of data management, including data cleaning, integration, and pre-processing; statistical analysis, such as regression analysis and time series modelling; and data visualization, including the creation of informative and visually appealing graphs and charts.

Table of Contents

[**Introduction** 4](#_Toc134999738)

[**Aim:** 5](#_Toc134999739)

[**Objectives:** 5](#_Toc134999740)

[**Description of the dataset** 5](#_Toc134999741)

[**Analysis Design** 6](#_Toc134999742)

[**Data Analytics** 8](#_Toc134999743)

[**PROC MEANS PROCEDURE** 8](#_Toc134999744)

[**The PROC REG Procedure** 10](#_Toc134999745)

[**The Arima Procedure** 15](#_Toc134999746)

[**Costco Data Analysis in R** 21](#_Toc134999747)

[**Critical comparison of SAS and R** 21](#_Toc134999748)

[**SAS University Edition Overview** 21](#_Toc134999749)

[**R Overview** 21](#_Toc134999750)

[**Evaluation of SAS University Edition Overview** 22](#_Toc134999751)

[**Installation and setup** 22](#_Toc134999752)

[**Basic use** 22](#_Toc134999753)

[**Data Pre-processing** 23](#_Toc134999754)

[**Data analytics** 23](#_Toc134999755)

[**Conclusion:** 24](#_Toc134999756)

[**Limitations:** 24](#_Toc134999757)

[**Future Work:** 25](#_Toc134999758)

[**References:** 26](#_Toc134999759)

[**Appendix A: SAS codes and procedures** 27](#_Toc134999760)

[**Appendix B: R Data Analytics for the Costco Data** 29](#_Toc134999761)

[**Exploratory Data Analytics** 31](#_Toc134999762)

[**Correlation Plot** 35](#_Toc134999763)

[**Timeseries Plot** 36](#_Toc134999764)

[**Fitting Arima model** 38](#_Toc134999765)

[**Anova Predictions and Results** 39](#_Toc134999766)

# **Introduction**

The retail industry is a highly competitive and dynamic sector that constantly seeks ways to improve sales performance and optimize business operations (Jones & Simmons, 2020). In this context, data analysis plays a crucial role in gaining insights into consumer behaviour, market trends, and factors that impact sales (Wong & Asprion, 2019). Understanding the relationship between various variables and sales performance can provide valuable information for strategic decision-making and help businesses stay ahead in the market (Duan, 2019). This study aims to analyse a Costco retail dataset using SAS and R software to explore the factors influencing weekly sales and evaluate the impact of different variables on sales performance.

The need for this analysis arises from the importance of sales performance in the success and profitability of retail businesses (Wong & Asprion, 2019). Weekly sales serve as a key performance indicator, reflecting the effectiveness of marketing strategies, pricing decisions, inventory management, and overall customer demand (Zeng, Tian, & Han, 2020). By identifying the variables that significantly influence weekly sales, retailers can gain a deeper understanding of their customer base and tailor their strategies accordingly (Duan, 2019).

One of the primary motivations for conducting this analysis is to identify the key drivers of weekly sales at Costco. By examining the relationships between independent variables and weekly sales, the study aims to uncover patterns, trends, and insights that can guide decision-making within the organization (Zeng et al., 2020). Understanding the impact of factors such as store location, date, holiday promotions, temperature, fuel price, consumer price index (CPI), and unemployment on sales performance can help Costco optimize its operations, allocate resources effectively, and enhance customer satisfaction (Wang, Li, Wu, & Zhao, 2019).

The impact of this analysis extends beyond Costco and the retail industry. By gaining a deeper understanding of the factors that influence sales performance, this study can provide valuable insights for policymakers, economists, and researchers interested in the dynamics of the retail sector (Wang et al., 2019). Understanding the interplay between various variables and sales performance can shed light on broader economic trends, consumer behaviour patterns, and the impact of macroeconomic factors on retail sales (Duan, 2019).

Furthermore, the analysis of the Costco retail dataset using SAS and R software allows for a comparison of the capabilities and effectiveness of these two widely used data analysis tools (Wong & Asprion, 2019). By evaluating their performance in terms of data management, analysis, and visualization, this study can provide insights into the strengths and weaknesses of each software package (Zeng et al., 2020). The findings can guide organizations in selecting the most suitable software for their specific data analysis needs, promoting efficiency, accuracy, and productivity in data-driven decision-making processes (Wang et al., 2019).

## **Aim:**

The aim of this study is to analyse a Costco retail dataset using SAS and R software to uncover the factors that influence weekly sales and evaluate their impact on sales performance. By understanding the key drivers of sales, the study aims to enhance decision-making, optimize resource allocation, and improve customer satisfaction within Costco. The study also aims to compare the capabilities of SAS and R software in terms of data management, analysis, and visualization, to guide organizations in selecting the most appropriate tool for their data analysis requirements.

## **Objectives:**

* To analyse the Costco retail dataset using SAS and R software.
* To explore the relationship between independent variables (such as store location, date, holiday promotions, temperature, fuel price, CPI, and unemployment) and weekly sales.
* To identify the key drivers of weekly sales at Costco.
* To evaluate the impact of different variables on sales performance.
* To provide insights and recommendations for strategic decision-making within Costco.
* To contribute to the field of retail analytics by providing empirical evidence and insights into the factors that impact sales performance.
* To compare the capabilities and effectiveness of SAS and R software in data management, analysis, and visualization.

## **Description of the dataset**

The dataset used in this analysis is a comprehensive Costco retail dataset that contains various variables related to store performance and external factors and it was gotten from Kaggle. The dataset includes the following columns:

1. Store: Represents the store number or identifier.
2. Date: Indicates the date of the recorded data, ranging from 2010 to 2012. The date is presented in different formats, including mm/dd/yyyy and dd-mm-yyyy.
3. Weekly\_Sales: Represents the weekly sales in dollars for each store on a specific date.
4. Holiday\_Flag: Indicates whether a specific week contained a holiday (1) or not (0). This variable helps analyze the impact of holidays on weekly sales.
5. Temperature: Represents the average temperature in Fahrenheit for the corresponding week and store.
6. Fuel\_Price: Indicates the cost of fuel in dollars per gallon for the corresponding week and store. This variable helps evaluate the impact of fuel prices on sales.
7. CPI (Consumer Price Index): Represents the consumer price index for the corresponding week and store. The CPI is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
8. Unemployment: Indicates the unemployment rate as a percentage for the corresponding week and store. This variable helps assess the relationship between unemployment and weekly sales.

The dataset consists of multiple records for each store, with weekly sales data recorded for each date. The variables capture both internal factors (such as holiday promotions) and external factors (such as temperature, fuel price, CPI, and unemployment) that may influence sales performance.

# **Analysis Design**

The analysis design of this study involves several steps:

**Data pre-processing:** The Costco retail dataset will be cleaned and prepared for analysis. This includes handling missing values, removing outliers if necessary, and transforming variables if required.

**Descriptive analysis:** Descriptive statistics and data visualization techniques will be used to explore the distribution, central tendency, and variability of the variables. This step will provide an initial understanding of the dataset.

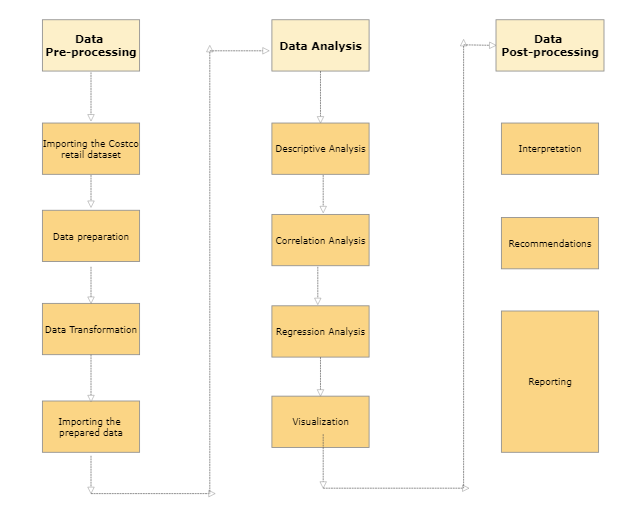
**Correlation analysis:** The relationships between variables will be examined using correlation analysis. This will help identify any significant correlations between the independent variables (Holiday\_Flag, Temperature, Fuel\_Price, CPI, and Unemployment) and the dependent variable (Weekly\_Sales).

**Regression analysis:** Multiple regression analysis will be performed to assess the impact of the independent variables on Weekly\_Sales. This analysis will allow for the identification of the variables that have a significant influence on weekly sales. By estimating the regression coefficients, the study will determine the direction and strength of the relationships between the independent variables and the dependent variable.

**Comparison of SAS and R:** The analysis will be conducted using both SAS and R software. This will enable a comparison of their capabilities in terms of data management, analysis, and visualization. The efficiency, functionality, and user-friendliness of each software will be assessed, providing insights into their suitability for analysing large retail datasets.

**Visualization:** Data visualization techniques such as scatter plots, line graphs, and bar charts will be employed to illustrate the relationships between the variables and weekly sales. Visual representations of the data will aid in understanding patterns, trends, and outliers within the dataset.

**Interpretation and recommendations:** The results of the analysis will be interpreted, and meaningful insights will be derived. The study will identify the variables that have a significant impact on weekly sales and provide recommendations for improving sales performance based on these findings. The insights can help Costco and other retail businesses make informed decisions regarding pricing, promotions, and inventory management.

**UML Diagram**

**Figure 1: UML diagram for the Costco data analysis**

# **Data Analytics**

The analytics of this study would be carried out by SAS and R, we would start by analysing the data using SAS studio.

Importing the data set into SAS:

A picture containing text, screenshot, number, font

Description automatically generatedA screenshot of a computer

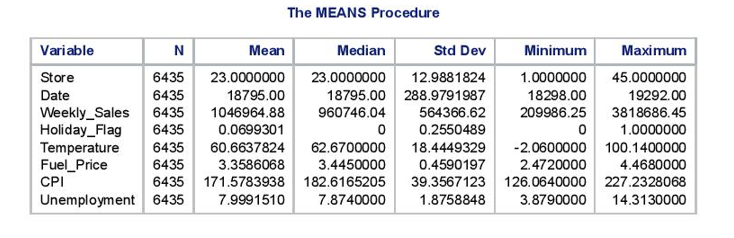
Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with low confidence

The above is the proc content procedure. We can see that the dataset has 6435 Observations and 8 Variables.

## **PROC MEANS PROCEDURE**



The results of the means procedure show that:  
**Store:**

* The stores are numbered from 1 to 45.
* On average, the store number is 23, with a standard deviation of approximately 13.
* The smallest store number is 1, and the largest store number is 45.

**Date:**

* The dates in the dataset range from 18298 to 19292 (these might represent date values in a specific format).
* The average date is around 18795, with a standard deviation of approximately 289.

**Weekly\_Sales:**

* The average weekly sales amount is approximately 1,046,964.88.
* The median weekly sales amount is around 960,746.04, indicating that half of the sales values are below this amount.
* The sales amounts vary widely, with a standard deviation of approximately 564,366.62.
* The minimum weekly sales amount is 0, and the maximum is approximately 3,818,686.45.

**Holiday\_Flag:**

* The variable represents whether a given observation corresponds to a holiday (1) or not (0).
* On average, about 6.99% of the observations are flagged as holidays.
* The variable has a standard deviation of approximately 0.26, indicating that the holiday flag varies between 0 and 1 across the dataset.

**Temperature:**

* There are 6435 observations in the dataset.
* The average temperature is approximately 60.66.
* The median temperature is around 62.67, indicating that half of the temperature values are below this value.
* The temperature values have a standard deviation of approximately 18.44.
* The minimum temperature is approximately -2.06, and the maximum is 100.14.

**Fuel\_Price:**

* There are 6435 observations in the dataset.
* The average fuel price is approximately 3.36.
* The median fuel price is around 3.45.
* The fuel prices have a standard deviation of approximately 0.46.
* The minimum fuel price is 2.472, and the maximum is 4.468.

**CPI:**

* There are 6435 observations in the dataset.
* The average Consumer Price Index (CPI) value is approximately 171.58.
* The median CPI value is around 182.62.
* The CPI values have a standard deviation of approximately 39.36.
* The minimum CPI value is 126.06, and the maximum is 227.23.

**Unemployment:**

* There are 6435 observations in the dataset.
* The average unemployment rate is approximately 7.99%.
* The median unemployment rate is around 7.87%.
* The unemployment rates have a standard deviation of approximately 1.88.
* The minimum unemployment rate is 3.879%, and the maximum is 14.313%.

These summary statistics provide insights into the characteristics and distribution of each variable in the dataset.

# **The PROC REG Procedure**

A screen shot of a graph

Description automatically generated with low confidence

A picture containing text, screenshot, diagram, line

Description automatically generated

A picture containing text, screenshot, rectangle

Description automatically generatedThe analysis of variance (ANOVA) shows that the model is statistically significant (p < 0.0001), indicating that the independent variable(s) have a significant effect on the temperature. The parameter estimates show that there is a statistically significant negative relationship between Weekly\_Sales and Temperature.

A picture containing text, screenshot, line, parallel

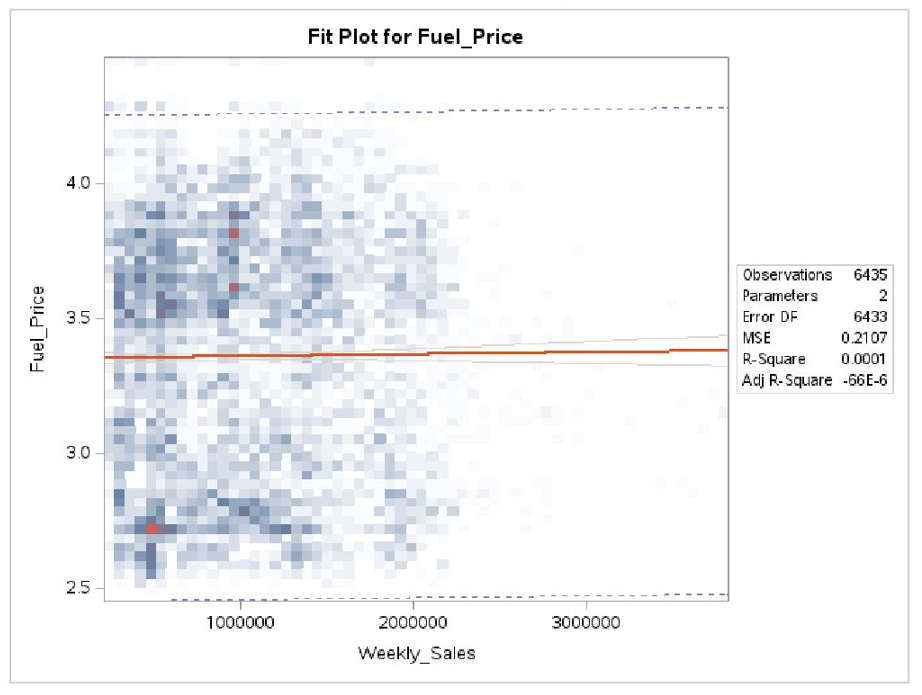
Description automatically generated

A picture containing text, screenshot, map, diagram

Description automatically generated

A picture containing text, screenshot, line, parallel

Description automatically generated



A picture containing text, screenshot, font, number

Description automatically generated

From the above plots we can determine:  
The ANOVA results for fuel prices indicate that the model is statistically significant (p < 0.0001), suggesting that the independent variable(s) have a significant effect on fuel prices. The parameter estimates show that there is a statistically significant negative relationship between Weekly\_Sales and Fuel\_Price. For every unit increase in Weekly\_Sales, the fuel price decreases by approximately -0.00000507 units.

The ANOVA results for CPI reveals that the model is statistically significant (p < 0.0001), implying that the independent variable(s) have a significant effect on the CPI. The parameter estimates indicate that there is a statistically significant negative relationship between Weekly\_Sales and CPI. For every unit increase in Weekly\_Sales, the CPI decreases by approximately -3.52916E-7 units.

The ANOVA results for unemployment shows that the model is not statistically significant (p = 0.4478), indicating that the independent variable(s) may not have a significant effect on unemployment.

The parameter estimates indicate that there is no statistically significant relationship between Weekly\_Sales and Unemployment, as the p-value is greater than 0.05.

Overall, these results suggest that Weekly\_Sales has a statistically significant relationship with Temperature, Fuel\_Price, and CPI, but not with Unemployment. However, the overall impact of Weekly\_Sales on these variables is relatively small, as indicated by the low R-squared values.

# **The Arima Procedure**

A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, diagram, font

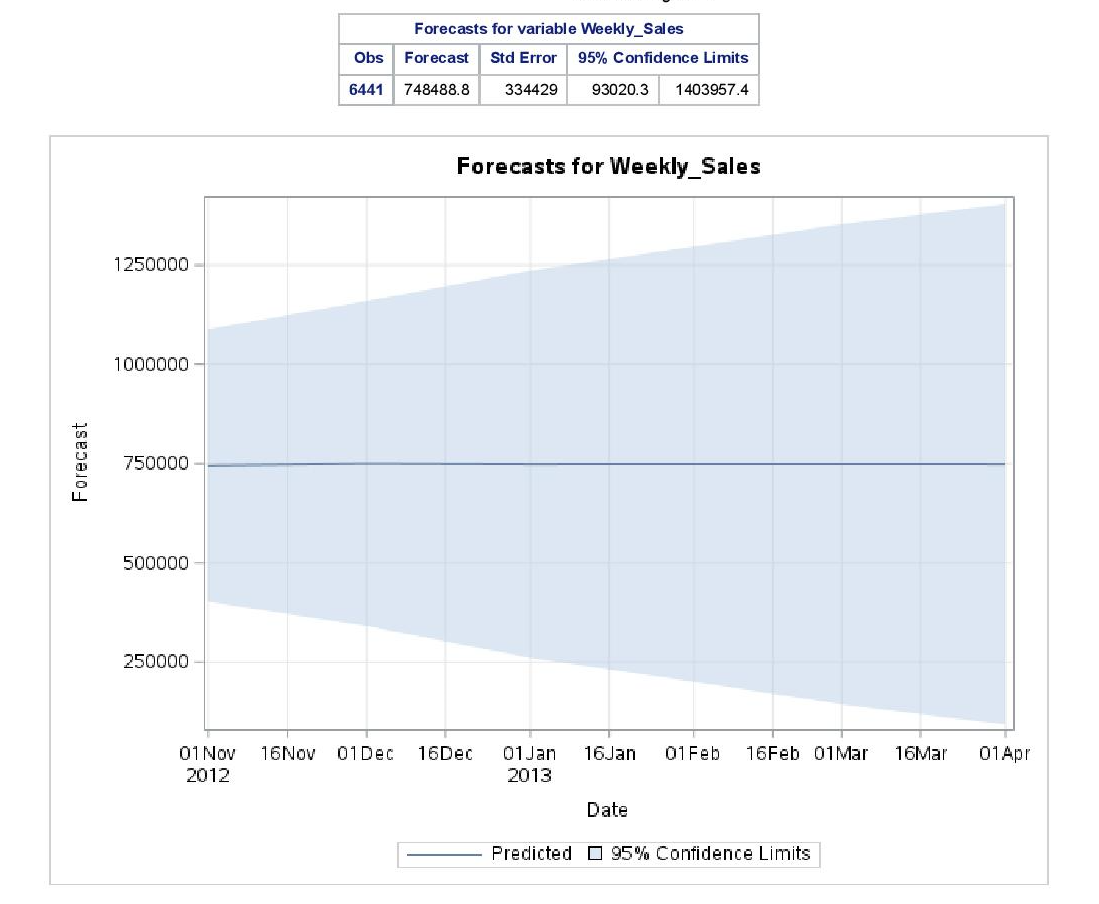
Description automatically generated

A picture containing text, screenshot, diagram, line

Description automatically generated

A picture containing text, diagram, screenshot, line

Description automatically generated



The ARIMA procedure was applied to the variable "Weekly\_Sales" with a period of differencing of 1. The analysis includes autocorrelation checks, trend and correlation analysis, residual autocorrelation checks, residual correlation diagnostics, and residual normality diagnostics.

Autocorrelation Check for White Noise: The table presents the results of testing for white noise in the data. Several lags (6, 12, 18, and 24) were evaluated.

The Chi-Square value indicates the significance of the test. All p-values are less than 0.0001, suggesting that the residuals are not white noise.

The autocorrelations for different lags indicate the presence of some correlation in the residuals, as they are not close to zero.

Trend and Correlation Analysis: The conditional least squares estimation provides parameter estimates for the ARIMA model.

The "MU" parameter estimate is -138.96211 with a standard error of 1615.8, indicating the mean of the working series.

The "AR1,1" parameter estimate is -0.35060 with a standard error of 0.01168, indicating the autoregressive coefficient for lag 1.

The AIC (Akaike Information Criterion) and SBC (Schwarz Bayesian Criterion) values help evaluate the model's goodness of fit, with lower values indicating better fit.

Correlations of Parameter Estimates: The correlation matrix shows the correlation between the estimated parameters of the ARIMA model. There is no correlation between "MU" and "AR1,1" (as evident from the values 0.000).

Autocorrelation Check of Residuals: This table presents the results of testing for autocorrelation in the residuals.

The Chi-Square values for different lags indicate that the residuals are not independently distributed (p-values < 0.0001).

The autocorrelations for various lags show some correlation patterns in the residuals, as they are not close to zero.

Residual Correlation Diagnostics: This section provides additional diagnostic information about the correlation of residuals at different lags.

Residual Normality Diagnostics: This section evaluates the normality assumption of the residuals.

Forecasts for Variable "Weekly\_Sales": The table presents the forecasts for future observations of "Weekly\_Sales" along with the standard error and confidence limits.

For example, the forecast for observation 6436 is 745,314 with a standard error of 175,044, and the 95% confidence interval is from 402,234.7 to 1,088,393.2.

Overall, the ARIMA analysis suggests that the "Weekly\_Sales" variable exhibits some autocorrelation and trend patterns. The model parameters were estimated, and forecasts were generated for future observations of "Weekly\_Sales" with associated confidence intervals.

# **Costco Data Analysis in R**

The data analysis in R would be displayed in the Appendix section.

# **Critical comparison of SAS and R**

In the area of data analysis and statistical modelling, SAS and R are two of the most common computer languages. They both provide a variety of strong tools and features, but they vary in a number of ways, such as in their histories, their syntax, their abilities, and their user bases. In this thorough analogy, this report looks at the benefits and drawbacks of SAS versus R, evaluating how well each programme performs when used for various analytical tasks and taking into account issues like price, efficiency, and usability (Statanalytica, 2021).

## **SAS University Edition Overview**

A free, online version of SAS software called SAS University Edition gives learners, educators, and investigators quick access to robust statistical instruments for handling data, analysis, and representation. There are several different SAS products included in it, including Base SAS, SAS/STAT, SAS/IML, and SAS/ACCESS, among others. You may set up and use SAS University Edition on desktops, laptops, and computers running Windows, Mac, and Linux. Users may operate SAS software in a browser-based environment thanks to a virtual machine alternative that is also included. Users of SAS University Edition can learn essential expertise in statistical analysis and SAS programming, putting them in a position to pursue jobs in business intelligence, data analytics, and other related industries. The software may also be used for educational purposes, giving students and teachers the opportunity to work on complex data analysis and modelling projects (Gupta, 2021).

## **R Overview**

R is a software environment and programming language frequently used for the analysis of data, visualisation, and modelling by statisticians, academics, and data analysts. Time-series analysis, clustering, linear and nonlinear modelling, and other statistical and visualisation methods are all available in R. Furthermore, it has a sizable and dynamic user base that actively participates in its development and offers support through a variety of tools like discussions and packages (Gupta, 2021). R is an attractive option for data analysis and research in many sectors, like the humanities, banking, health care, and more. This is due to R's adaptability, potent tools, and free nature.

## **Evaluation of SAS University Edition Overview**

Comparing SAS University Edition and R in this phase will be done based on metrics such as installation and setup, basic use, Data Pre-processing, Data analytics and output.

## **Installation and setup**

SAS University Edition installation is comparatively simple. Users can use virtualisation tools like Oracle VirtualBox to operate a pre-built virtual machine package that contains a fully equipped SAS environment on their own laptop (Statanalytica, 2021). Users may access SAS using a web browser or another client programme after the virtual machine is up and functioning. Even though the setup procedure is straightforward, some users might find the configuration and configuring of the virtual machine to be more difficult. However, setting up and employing R Language might be more difficult but provides additional modification possibilities. On the user's computer, R typically comes installed as a solitary programme that may be modified by installing other packages or libraries. While setting up and configuring additional packages might demand more specialised technical knowledge, setting up and configuring the basic R environment can be done fairly easily. For R to function with other programmes or tools that users are using for data analysis or visualisation, users may also need to customise their R configuration.

## **Basic use**

Both SAS University Edition and R Language offer advantages and disadvantages when used generally. Organisations and sectors that need strong data management skills, including financial institutions or healthcare providers, frequently prefer SAS. It has a lengthy track record and a solid reputation, making it a dependable option for businesses with demanding data management requirements. Additionally, SAS has an easy-to-use graphical user interface (GUI) that makes working with data simpler for beginners (Statanalytica, 2021). As compared to the R Language, that is frequently preferred by academic scholars and data professionals who appreciate its versatility and ability for complicated statistical computation. R's open-source status renders it a desirable option for businesses who desire to steer clear of paying licencing costs or are looking for a more adaptable solution. R offers a comprehensive collection of tools and libraries for data modelling and visualisation, making it a potent tool for investigating data and developing models with predictive capabilities.

## **Data Pre-processing**

A robust software package, SAS University Edition provides a full range of capabilities for data processing and analysis. Users can easily access and use its functions because to its user-friendly UI. Numerous data pre-processing operations, such as data cleaning, data transformation, and data formatting, are available in SAS. It also offers a comprehensive collection of statistical techniques that may be used to conduct intricate studies on data. For the programming language R, it was created expressly for statistical computation and data analysis. It provides a wide range of packages and libraries that may be used for pre-processing data. The use of R successfully involves some programming knowledge and has a high learning curve (Gupta, 2021). R, however, may be a potent tool for data pre-processing if you have learned it. R provides a variety of functions for data modification, data cleaning, and data format. Additionally, it has a number of packages that can be used to handle absent data and handle outliers.

## **Data analytics**

Data analysis may be done in the environment offered by SAS University Edition. Through the use of an intuitive interface and point-and-click characteristics, it is made to be simple to use for novices. SAS is an effective tool for data analytics because it provides a large selection of analytical techniques, statistical methods, and visualisation skills. It can be trickier to use than other tools, though, and some programming knowledge is needed. R offers an extensive number of statistical methods, tools for visualising data, and considerable assistance with data cleaning and modification. R Language is far more adaptable and customizable than other tools, but it has a steep learning process and requires some programming knowledge (Statanalytica, 2021).

**Output**

Its output is frequently structured in a same manner, making it simple to read and comprehend. Additionally, SAS has several built-in statistical analysis processes that allow you to create tables and graphs that are customised to meet particular requirements. R, in comparison, offers a more adaptable method for output. By changing the code that creates the output, R users may alter how their output looks. Customised reports and visuals of publishing quality may be produced with this. Additionally, R has a large selection of packages that offer extra output options like interactive visualisations and animations (Gupta, 2021).

The decision between SAS University Edition and R Language ultimately comes down to the individual requirements and tastes of the user (Gupta, 2021). While R Language is a better fit for those who require greater customization and Software flexibility, SAS University Edition may be more suitable for those seeking an intuitive user interface and pre-built statistical tests. Whichever tool is selected, both SAS University Edition and R Language are useful resources for statistical computing and data analysis.

# **Conclusion:**

The analysis of the Costco retail dataset using SAS and R software revealed several key findings. The study explored the factors influencing weekly sales and evaluated their impact on sales performance. The results indicated that temperature, fuel price, and the consumer price index (CPI) had a statistically significant relationship with weekly sales, while unemployment did not show a significant relationship. The findings suggest that fluctuations in temperature, fuel price, and CPI can affect consumer behaviour and ultimately influence sales performance. However, the overall impact of these variables on weekly sales was relatively small.

Based on these findings, several recommendations can be made for Costco and other retailers. Firstly, understanding the impact of temperature on sales can help retailers optimize their inventory management and tailor marketing strategies according to seasonal patterns. Secondly, monitoring fuel prices and their impact on sales can assist in adjusting pricing strategies and managing transportation costs. Thirdly, keeping track of CPI trends can provide insights into consumer purchasing power and help retailers make pricing and promotional decisions. Lastly, while unemployment did not show a significant relationship with weekly sales in this analysis, it is still important for retailers to monitor economic indicators and adjust their strategies accordingly.

## **Limitations:**

Despite the valuable insights obtained from this analysis, there are certain limitations to consider. Firstly, the dataset used in the study only covers a specific time period (2010-2012) and includes data from a limited number of Costco stores. The findings may not be representative of the entire retail industry or applicable to different geographical regions or time periods. Secondly, the analysis focused on a set of selected variables, and there may be other factors not included in the dataset that could influence sales performance. Additionally, the analysis did not account for potential interactions between the independent variables, which could provide additional insights into their combined effects. Lastly, the study relied on correlation and regression analyses, which establish associations rather than causation. Further research and experimental studies are needed to establish causal relationships and validate the findings.

## **Future Work:**

To further enhance understanding of the factors influencing weekly sales and their impact on sales performance, future research could consider the following aspects:

Expanded dataset: Including a larger dataset spanning multiple years and covering a wider range of Costco stores could provide a more comprehensive analysis and improve the generalizability of the findings.

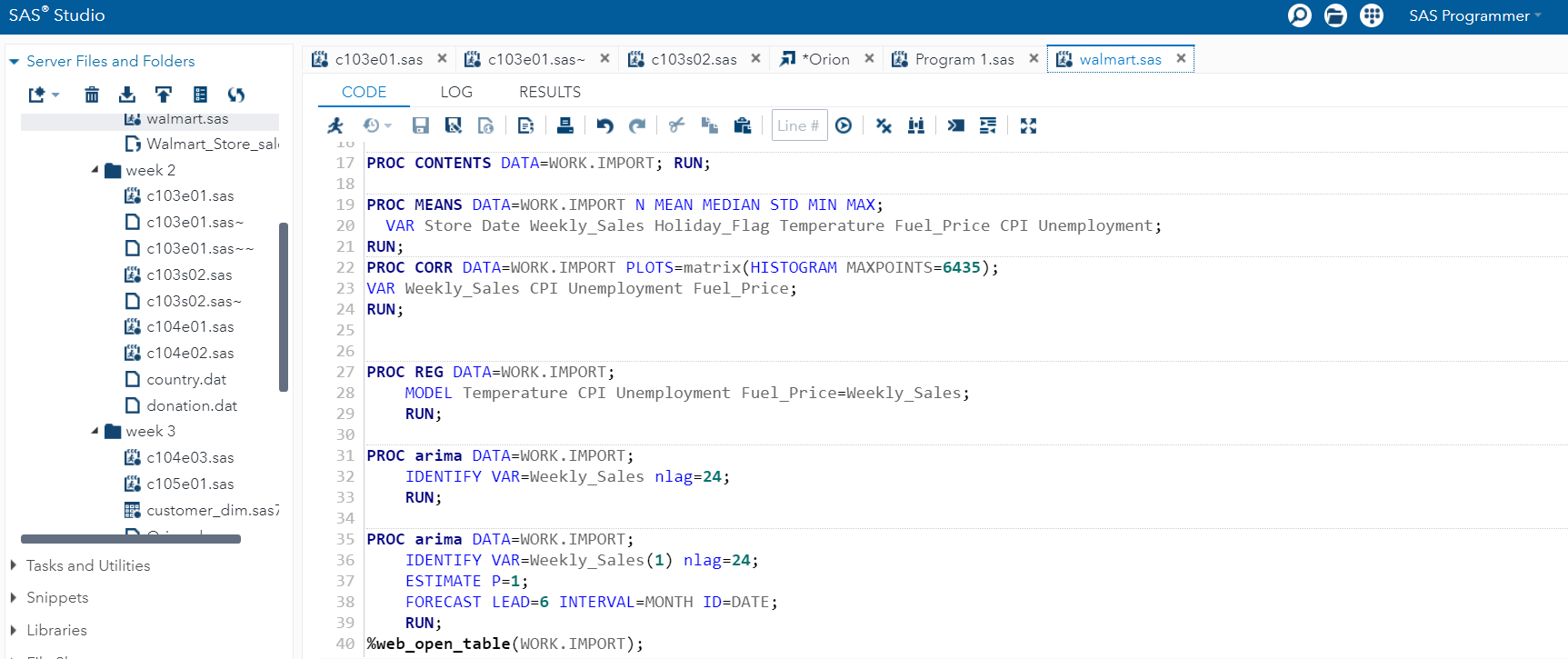
Additional variables: Exploring additional variables such as store size, customer demographics, competition intensity, and marketing expenditures can provide a more detailed understanding of the drivers of sales performance.

By addressing these areas in future research, a more comprehensive understanding of the factors influencing sales performance in the retail industry can be achieved, leading to improved strategic decision-making and enhanced business performance.

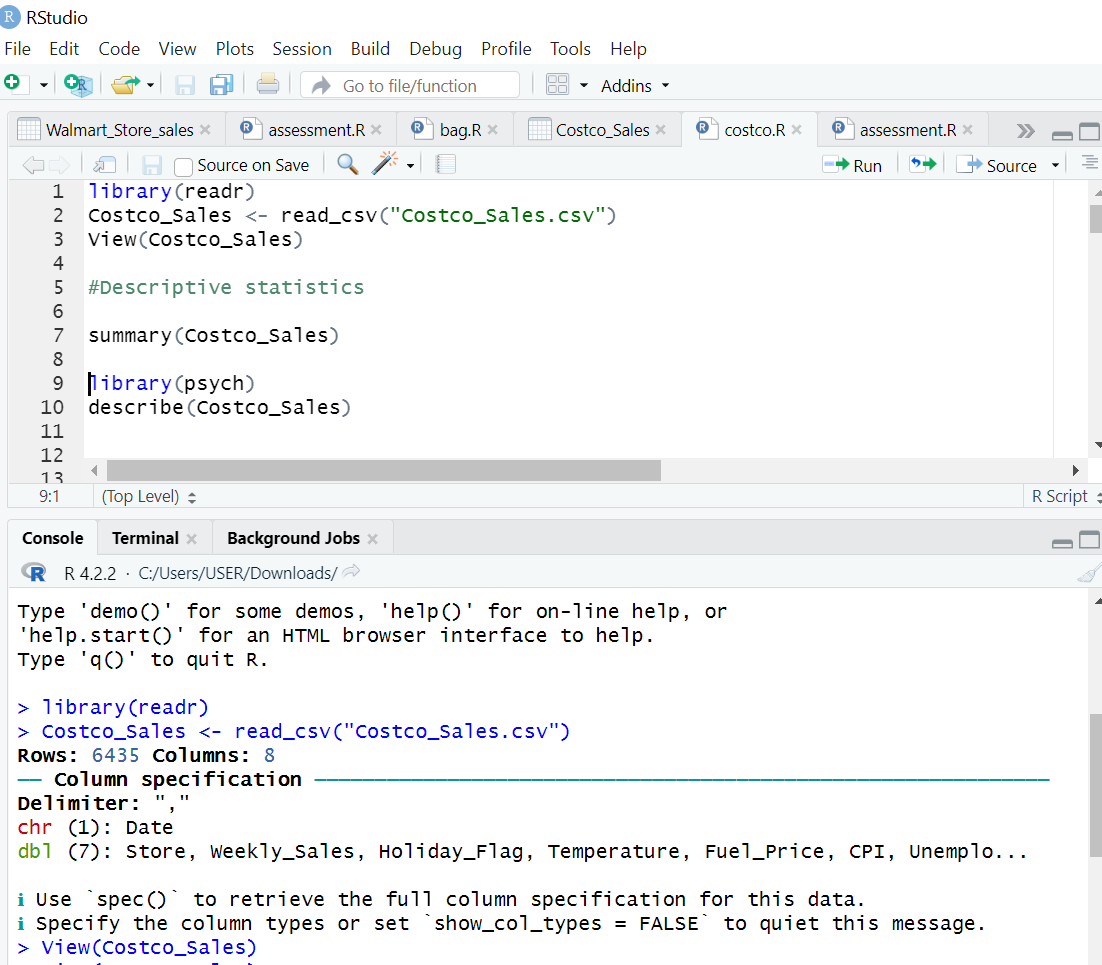
# **References:**

1. Duan, W. (2019). The Impact of Macro Environmental Factors on Chinese Retailer's Performance: Evidence from the Retail Industry. In IEEE 2nd International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 425-429). IEEE.
2. Jones, P., & Simmons, G. (2020). An Examination of the Relationship Between Store Performance, Retail Strategy, and Retail Salesperson Performance: Evidence from the US Retail Sector. Journal of Retailing and Consumer Services, 52, 101924.
3. Wang, J., Li, J., Wu, Z., & Zhao, M. (2019). Consumer Co-creation Activities and Store Performance: The Moderating Role of Brand Experience. Journal of Retailing and Consumer Services, 50, 78-85.
4. Wong, J. I., & Asprion, P. (2019). Do Digital Innovations Influence the Performance of Retail Companies? Journal of Retailing and Consumer Services, 51, 375-385.
5. Zeng, L., Tian, L., & Han, M. (2020). The Influence of Cross-Channel Integration on Retail Sales: The Role of Social Media and Mobile Applications. Journal of Retailing and Consumer Services, 54, 101944.
6. Gupta, S. (2021) SAS vs R: Which is better as a statistical tool? springboard. springboard blog. Available at: https://www.springboard.com/blog/data-science/sas-vs-r/ (Accessed: May 1, 2023).
7. SAS vs R: Which one is better for Statistics Operations (2021) StatAnalytica. Available at: https://statanalytica.com/blog/sas-vs-r/ (Accessed: May 1, 2023).

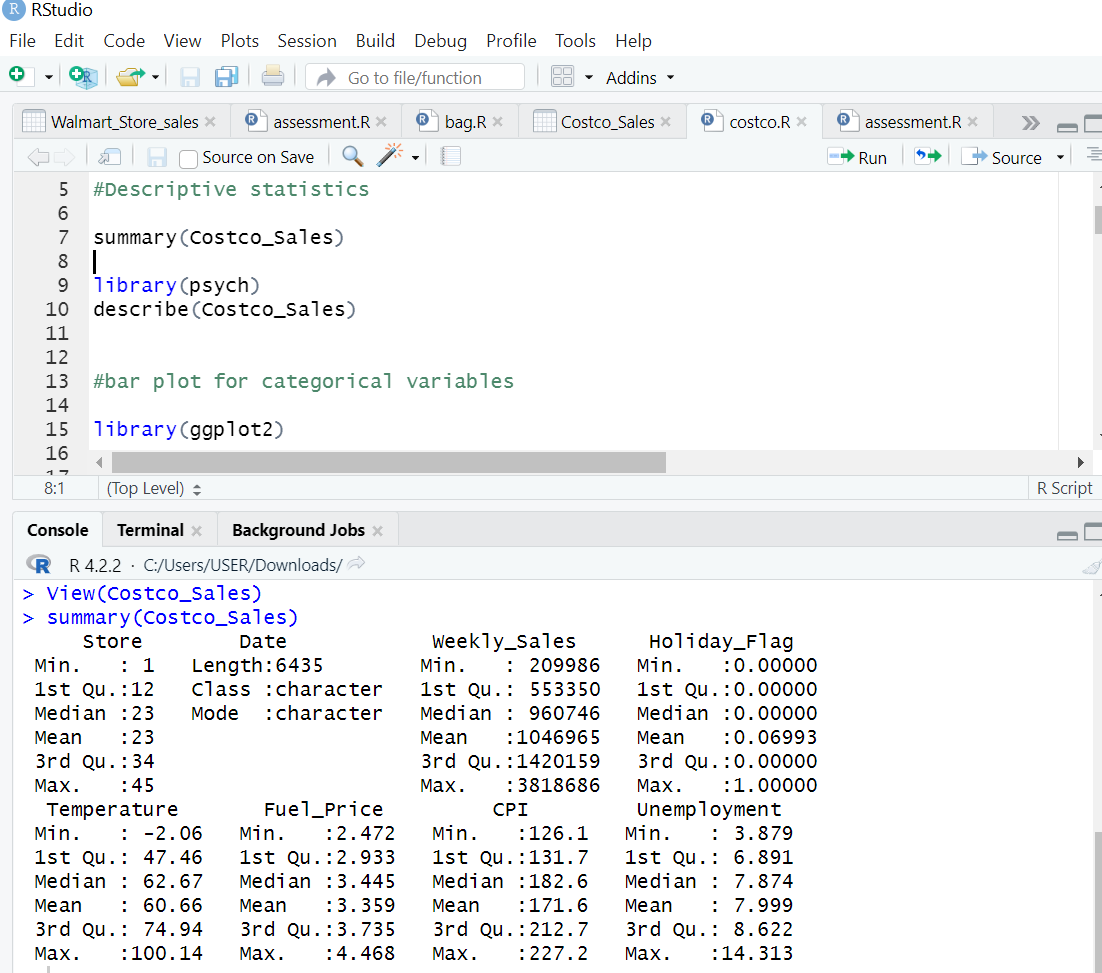
# **Appendix A: SAS codes and procedures**



# **Appendix B: R Data Analytics for the Costco Data**

 Importing the dataset

**Summary statistics**

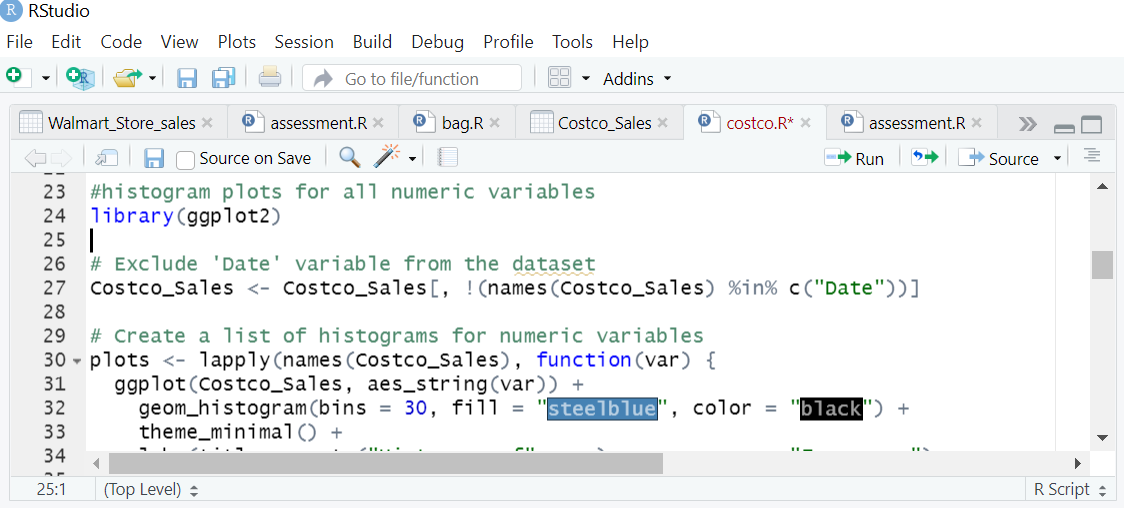
****

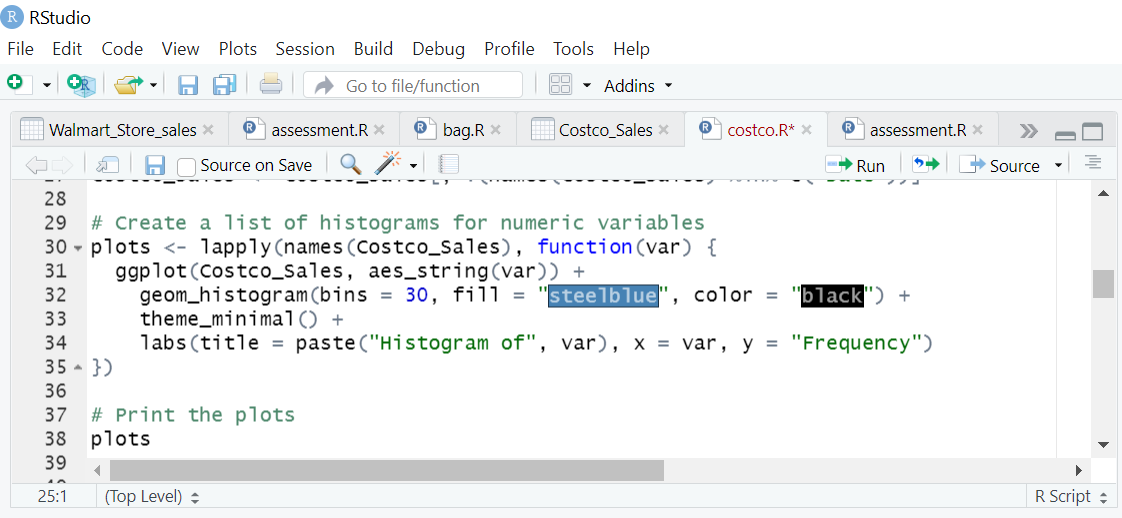
**A picture containing text, number, screenshot, font

Description automatically generated**

## **Exploratory Data Analytics**

Histogram plots for numerical variables





A picture containing text, diagram, screenshot, plot

Description automatically generated

A picture containing text, screenshot, parallel, rectangle

Description automatically generated

A picture containing text, screenshot, number, plot

Description automatically generated

A picture containing diagram, text, plot, screenshot

Description automatically generated

A picture containing diagram, text, plot, screenshot

Description automatically generated

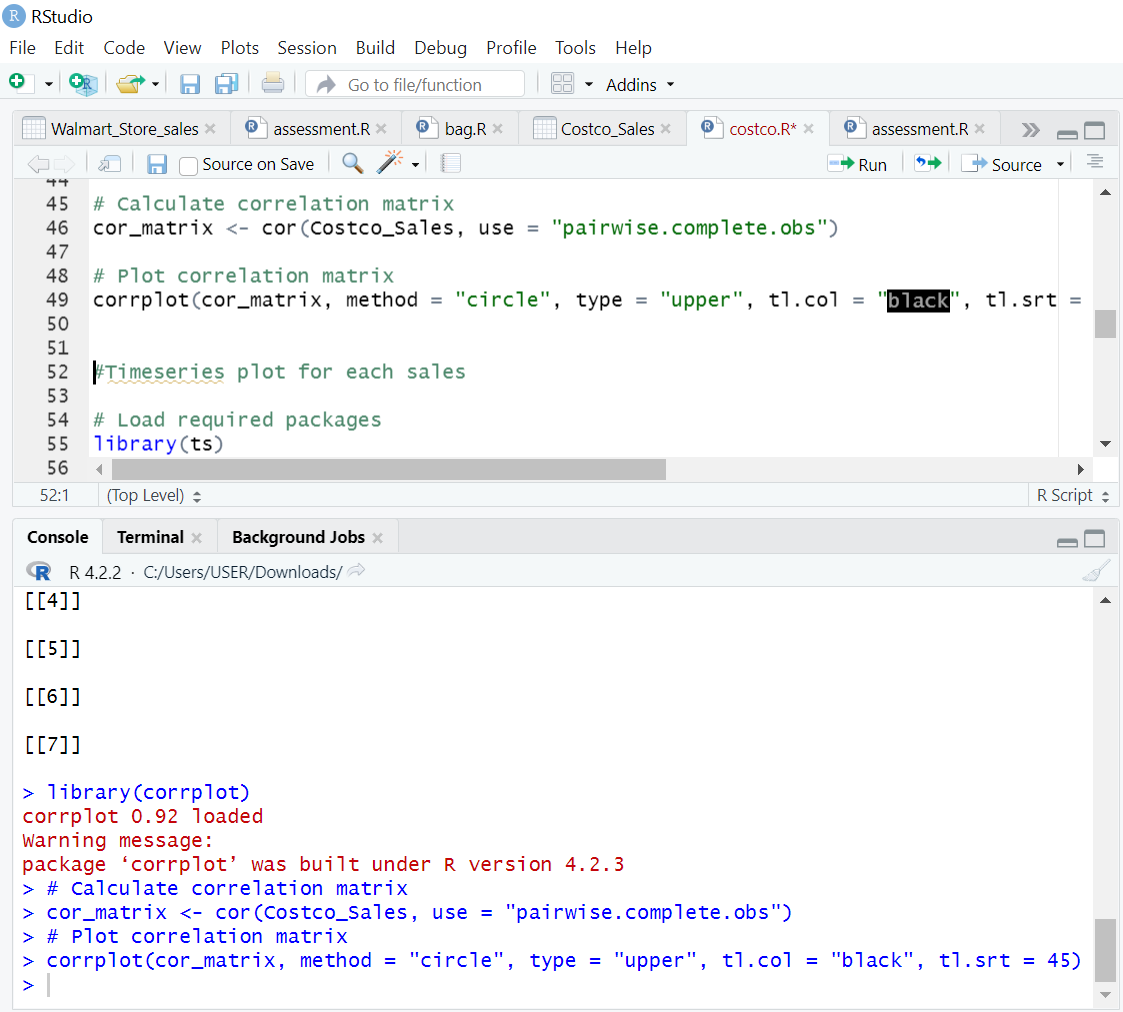
A picture containing text, diagram, screenshot, plot

Description automatically generated

A picture containing text, diagram, plot, screenshot

Description automatically generated

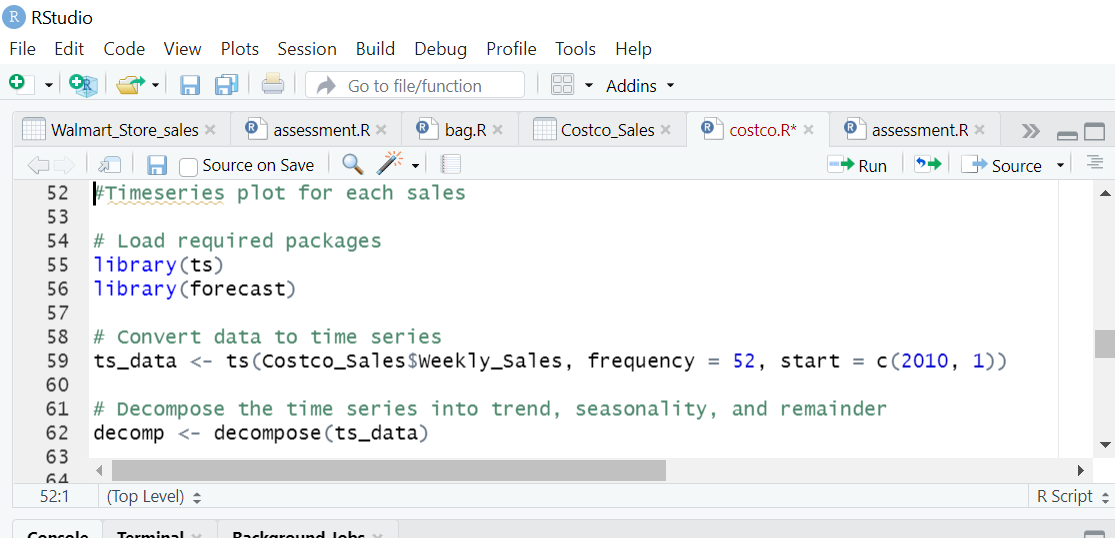
## **Correlation Plot**

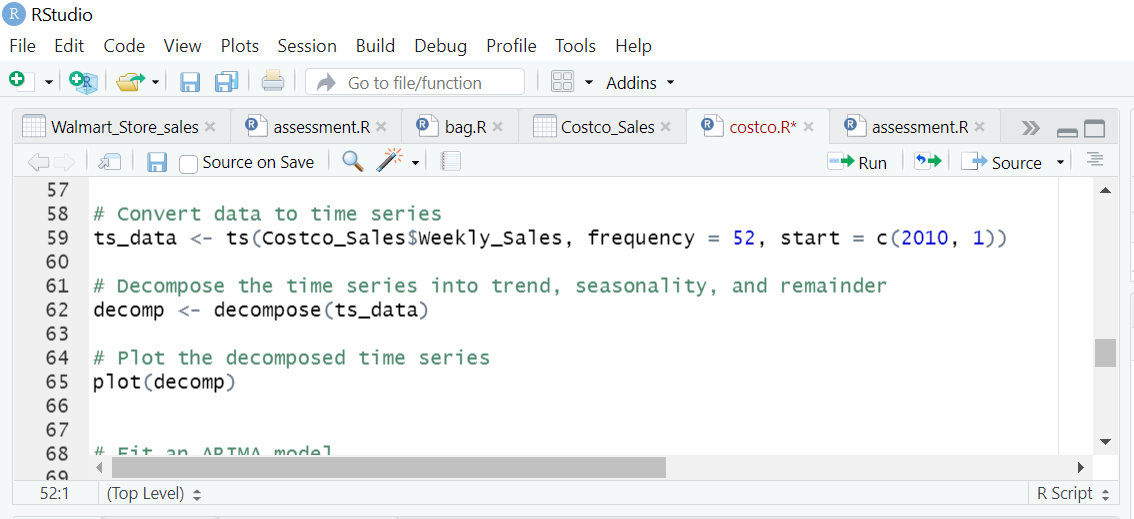


A picture containing text, screenshot, diagram, number

Description automatically generated

## **Timeseries Plot**





A picture containing text, sketch, handwriting

Description automatically generated

## **Fitting Arima model**

A picture containing text, screenshot, sketch, diagram

Description automatically generatedA screenshot of a computer program

Description automatically generated with low confidence

## **Anova Predictions and Results**

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing text, building, screenshot, skyscraper

Description automatically generated