Using packages is a two step process

First thing is we did was install all the relevant packages

Then load them into the session, there are comments next to each to show what they are used for.

—----------------------------------------------------------------------------------------------------------

Now to load the data into R itself

After downloading the data set from kaggle, I made sure I knew where I have saved the data as I need the file path in order for R to find it

I set the file path as a variable as it alot of characters to keep repeating

Use the read csv function to pull the data into R

(When inspecting the data within excel half of the dates changed to american date format, when I inspected the csv in notepad all the dates were of the same format)

Then used the head function to view the first few rows of the data

—----------------------------------------------------------------------------------------------------------

Using the df summary function it generated a summary view that can be seen here on the right. It gives statistical analysis for each of the columns

The double colon is to explicitly specify which “view()” as there is another view within another package.

The str(data) output provides an overview of the dataset's structure, including the number of rows and columns, data types of each column, and a preview of the data in each column.

Finally the last part of the code checks for missing values. As the entire data set show False there are no missing values

—----------------------------------------------------------------------------------------------------------

The first 3 lines of the code generate a the mean,median and standard deviation of weekly sales

The using the cat function prints them in a more readable way seen here in the purple

The last lines of code generate the correlation matrix between all fields   
A positive correlation between one variable and any of the other variables indicates that as one increases, so does the other.

A negative correlation indicates an inverse relationship.

A correlation value close to 0 suggests no linear relationship.

—----------------------------------------------------------------------------------------------------------

Anyone not familiar with scientific notation, 1e^6 is the same as 1 with 6 0s which in this case is 1 million

As you can see from this time series graph there is a peak at christmas every year followed by a dip in January

### **Conclusions and Recommendations:**

1. **Sales Trends:**
   * **Seasonality:** If the time series plot shows clear seasonal trends, Walmart could plan inventory, staffing, and promotions to align with high and low sales periods.
   * **Promotions:** Spikes in sales might correlate with promotional activities. Understanding these can help Walmart optimise the timing and scale of future promotions.