**Problem:** The problem that I identified was based off recent news of closing Walmart locations. The problem was how can Walmart easily view the metrics of their stores as well as consider the specific stores, the economic climate, and the time of year. This code processes the data and then generates visualizations to analyze sales data that help to identify sales trends by specific store, CPI (Consumer Price Index), and date (Shows the time of year, if holiday season, climate, etc.).

**Data:** The data is stored in a CSV file entitled Walmart Data Analysis and Forecasting.csv which includes information about weekly sales, stores, dates, and the CPI that week. This .csv file was located and downloaded from: https://www.kaggle.com/datasets/asahu40/walmart-data-analysis-and-forcasting?select=Walmart+Data+Analysis+and+Forcasting.csv

**Functions:**

1. process\_data(df): This function drops unused columns, converts the 'Date' column to datetime format, creates additional columns for year, month, and day, and converts the 'Store' column to a category type. These actions clean and process the data in Python. The input is a df object and the output is the processed df object.

2. read\_data(file\_path): This function reads the data from the provided .CSV file and returns a df object. If the .CSV file isn’t found, it will return none. The input is the access to the file and the output is a df object.

3. plot\_sales\_by\_store(df): This function plots the total sales by store. The input is the df object from the read\_file function and the output is a bar chart which shows the total sales in millions of dollars. The three bars representing the stores with the lowest sales are red, and the three bars representing the stores with the most sales are green. All others are dark grey.

4. plot\_sales\_by\_cpi(df): This function plots the total sales by CPI during that week. The input is the df object from the read\_file function and the output is a scatter chart that shows the total sales in millions of dollars. The x-axis represents the CPI and the y-axis represents the sales in millions of dollars. This scatter chart shows the relationship of the CPI to the amount of sales.

5. plot\_sales\_by\_date(df): This final function plots the total sales by date. The input is the df object from the read\_file function and the output is a line chart that shows the total sales by date. The x-axis shows the date, and the y-axis represents the amount of sales. The values on the y-axis are formatted in millions of dollars, and the x-axis is supposed to show the dates in the format of 'day-month-year'. After tireless scouring of the internet, I could not make those dates appear properly on the x-axis lable.

*\*\*If points could be lost for the x-axis on chart not working, could the previous 4 just count as my 4 required functions?\*\**

**Packages used:**

1. pandas – for Data Analysis
2. matplotlib – for Plotting of charts