

# Exploratory Data Analysis of the Walmart Sales

---

**Author:** Venu Gopal Rao Annamaraju

**Student Id:** 23031779

**GitHub Repository:** <https://github.com/Vgopi3055/venu-gopal-rao-annamaraju-statistics-and-trends-assignment>

## 1. Introduction to the Dataset

In this report, we will perform an exploratory data analysis (EDA) on the Walmart Sales dataset sourced from Kaggle. We will use Python libraries such as pandas, seaborn, and matplotlib. Our objective is to explore and analyze the dataset, understand the relationships between different variables, and get actionable analysis.

## 2. Dataset Overview

This dataset consists of the data of 45 stores of Walmart's weekly sales. In addition to that, this dataset also contains information about other attributes such as store type, unemployment, fuel price, and consumer price index.

## 3. Exploratory Data Analysis:

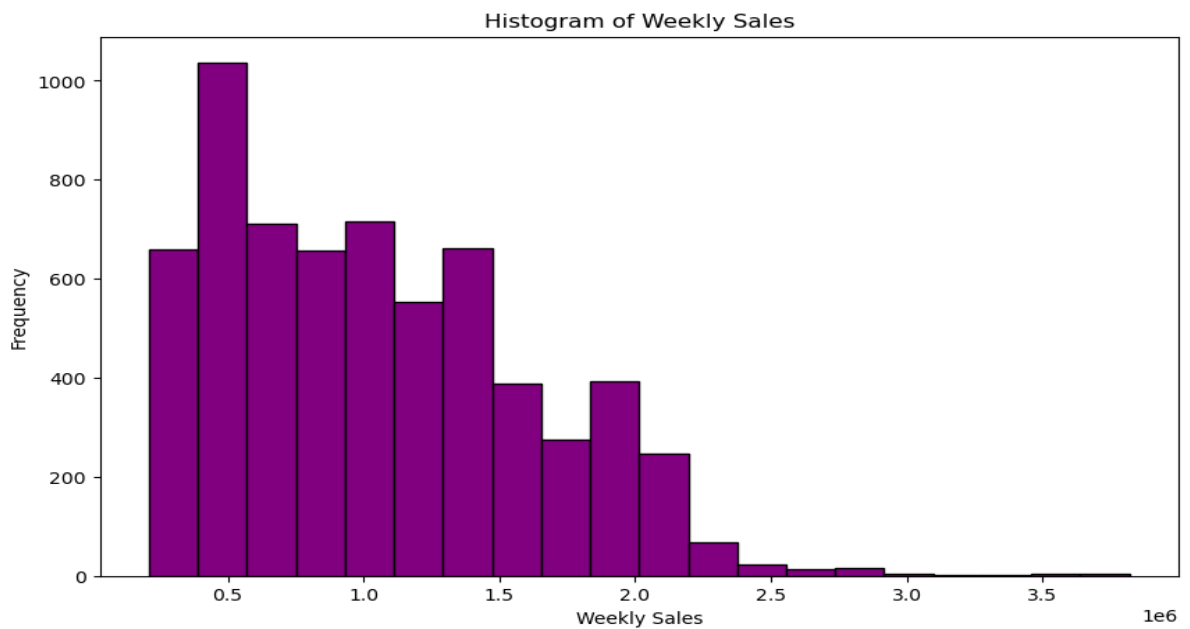
### Dataset Exploration:

By exploring the dataset, we come to know that the dataset consists of 8 columns and 6435 rows. The numerical data consists of columns store, weekly sales, holiday flag, temperature, fuel price, CPI, and unemployment. Whereas the categorical data consists of a Date column.

## 4. The Visualization of Graphs:

### 4.1 Analysis of Histogram:

A histogram is a type of graph that looks similar to a bar chart but has different purposes. It helps in summarizing the distribution of numerical data gathered from a dataset. We have denoted the x and y axes to weekly sales and frequency respectively. We can observe the visual depiction of the histogram of weekly sales of stores are comparatively the same.



*Fig 1: Histogram of Weekly Sales*

## 4.2 Analysis of Line Graph:

Now, let's analyze the line graph, before analyzing we have to convert the date column, which is in the object data type to an integer, so we have all numerical values to plot the graph.



*Fig 2: Line graph analysis of weekly sales data over time*

To simplify the data, I have grouped by the stores and year to eliminate overlapping of data while plotting the graph. As we observe, in 2010, sales of stores 1 to 35 were very high when compared to other years of sales. The year 2012 amounted to the least number of sales among the stores. We can notice hard correlations of sales of 45 stores ranging from 1.0 to 2.0.

### 4.3. Analysis of Heatmap:

A correlation matrix is a table that displays how different variables relate to each other. In this case, it shows how various attributes correlate with Walmart's weekly sales. The values in the correlation matrix range from -1 to 1 means negative and positive correlation respectively.

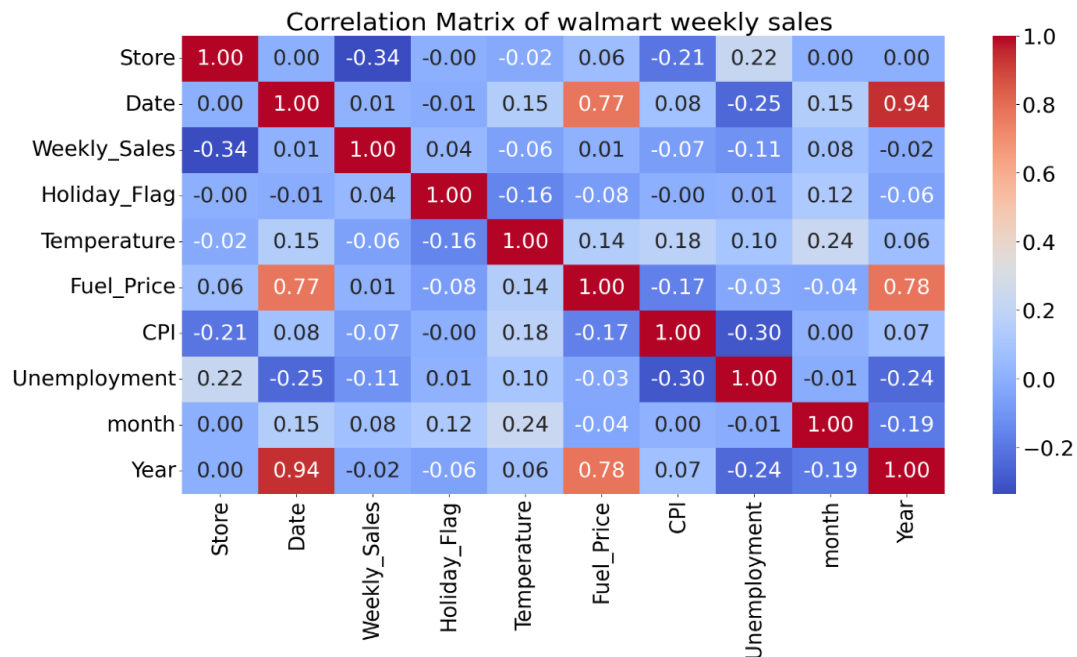


Fig 3: Heatmap of the correlation matrix of Walmart's weekly sales

As per observations when unemployment is higher sales tend to decrease, which showcases a negative correlation, and when the consumer price index is high stores see an increase in sales sign of a positive correlation. Because two things are correlated there is no prediction that one will cause another

## 5. Conclusion:

To conclude things, weekly sales of newer stores have been impacted not just because of unemployment, other things also factored in such as cpi and fuel prices as well. The stores had a very good run of sales in the first year and comparatively good in 2011, but they saw a very big dip in 2012 as depicted in the line graph. These visualizations provide valuable insights for understanding the sales patterns and relationships within the dataset, which can guide stores in terms of performance, and profitability and enhance decision-making processes quite easily for the person in command.