Walmart Sales Data Report

**Project Title: -** Walmart Sales Data Analysis

**Project Link: -** [Github Repository](https://github.com/PriyankaPatel1543/Priyanka_DataVista-Sales-Data-Analysis-and-Visualization_Infosys_Internship_Oct2024/tree/main/Final%20Walmart%20%20full%20Project)

**Libraries Used:** - Pandas, NumPy, Matplotlib, Seaborn, Sklearn, Linear Regression

**Tools Used: -** Anaconda |Jupyter | Python | Numpy |Pandas | Statsmodels | Matplotlib | Seaborn |

Sklearn

**Skill Set Used: -** Data Science | Data Analytics | Artificial Intelligence | Machine Learning | Python |

Statistical Modelling | SQL | Exploratory Data Analysis (EDA) | Data Cleaning | Data

Wrangling | Data Mining | Data Visualization .

**Steps Involve in the Project: -**

**Step 1: - Business Problem Understanding**

* Let's figure out what factors influence its revenue.
* Can factors such as air temperature and fuel cost influence the success of a huge company along with the purchasing power index and seasonal discounts? And how does machine learning minimize costs and increase economic impact?

**Step 2: - Data Understanding**

* Collecting Data and Loading it.

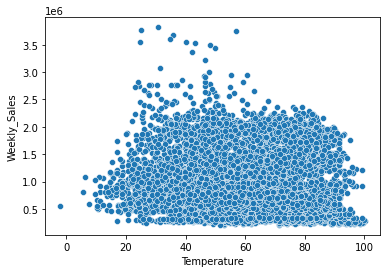
**The data contains the following columns:**

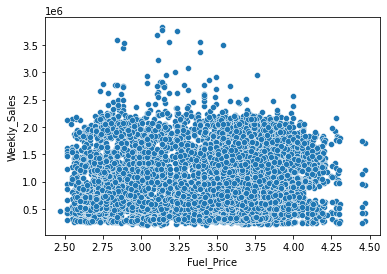
* **Date:** Sales week start date
* **Store:** Store number
* **Weekly\_Sales:** Sales
* **Holiday\_Flag:** Mark on the presence or absence of a holiday
* **Temperature:** Air temperature in the region
* **Fuel\_Price:** Fuel cost in the region
* **CPI:** Consumer price index
* **Unemployment:** Unemployment rate
* **Data Exploration**

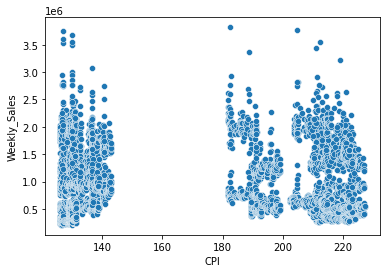
Functions Used: -

* df.shape - To find the Total Number of row and column
* df.info() - To get the number of non-null values and type of data type for each

column

* df[data].describe() - To get the statistical data like count, mean, std, min, max, etc.
* df[data].skew() – To know skewness in the data
* df[data].corr() – To find the Relation Between 2 variables.
* **Feature Transformation -** For the data which shows more skewness we need to apply feature transformation.
* **Plots**
* **Scatter Plot: -**

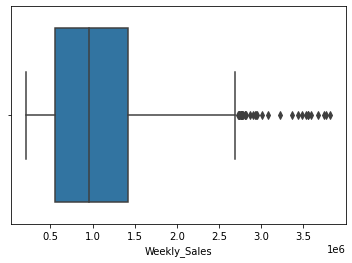
1. Temperature vs Weekly\_sales
2. Fuel\_price vs Weekly\_sales
3. CPI vs Weekly\_sales



**Step 3: - Data Preprocessing**

It includes data Cleaning and Processing of data.

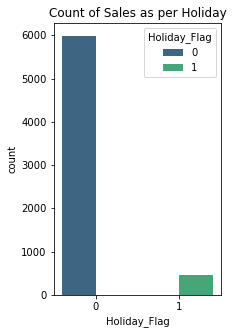
* df.isnull().sum() – It shows total number of null value present in the column.
* df.duplicated().sum() – It Shows the number of duplicate present in the data.
* Boxplot –



* Finding the quartile Value like Q1, Q2,Q3, IQR, Uppe4r limit and lower limit

87ui

* **Bar Plot :-**



* **Heat Graph:-**



* **Machine Learning: -**

Train Test Split

**Step 4: - Modeling**

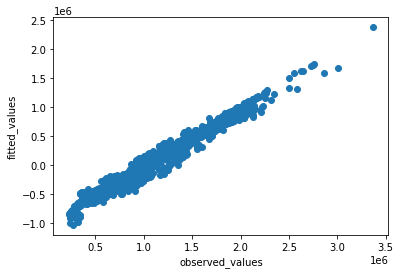
Applying the Linear Regression for finding the Coefficient and Intercept.

**Step 5: - Evaluation Metrics**

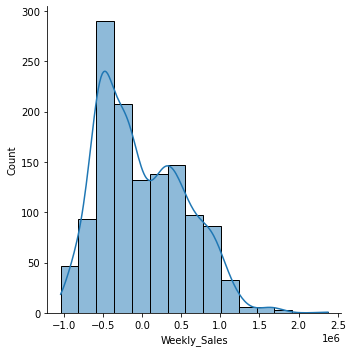
Checking whether the Actual Output and Predicted Output are same or not.

**Check for the Assumptions: -**

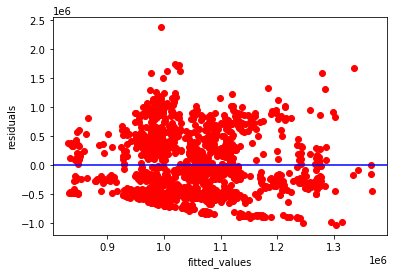
* **Linearity of Errors**



* **Normality of Error**



* **Equal Variance of Errors (Homoscadesicity)**



* **Variable Significance Hypothesis Testing for Variable**

Value of p>|t| should be less than 0.5.

R-Squared and Adj. R-Squared should be same.

**Conclusion: -**

* Weekly Sale and the revenue of Walmart is affected due to temperature and CPI. We can Predict the Weekly Sales using the Multi Linear Regression Model.
* With the help of Multi Linear Regression we can also increase the Revenue.