**Walmart Retail analysis:**

**Intro :**

This project focuses on conducting a comprehensive analysis of store sales data and building a predictive model for demand forecasting for all stores. The project encompasses various statistical tasks to gain insights into sales trends and factors impacting sales. It also involves developing a linear regression model to forecast future demand based on variables like date, CPI, unemployment, and fuel price.

**Which store has maximum sales ?**

*walmart\_data\_sales = walmart\_data.groupby('Store')['Weekly\_Sales'].sum()*

*print("Store Number {} has maximum Sales. Sum of Total Sales {}".format(walmart\_data\_sales.idxmax*

*(),walmart\_data\_sales.max()))*

This code calculates the total weekly sales for each store and then finds the store with the maximum sales using the idxmax() function. The variable store\_with\_max\_sales will contain the store number with the maximum sales.

***Store Number 20 has maximum Sales. Sum of Total Sales 301397792.46***

**Which store has maximum standard deviation & Also, find out the coefficient of mean to standard deviation**

by Calculating the standard deviation of sales for each store and identify the store with the maximum standard deviation. Additionally, calculating the coefficient of variation (mean to standard deviation ratio) for each store.

walmart\_data\_std = walmart\_data.groupby('Store').agg({'Weekly\_Sales':'std'})

print("Store Number {} has maximum Standard Deviation. STD {}”.format(walmart\_data\_std[‘Weekly\_Sales'].idxmax(),walmart\_data\_std['Weekly\_Sales'].max()))

***Store Number 14 has maximum Standard Deviation. STD 317569.9494755081***

*walmart\_data\_std\_coeff = walmart\_data.groupby(‘Store').agg({'Weekly\_Sales':['mean','std']})*

This code calculates and tells us the coefficient of mean to standard deviation for each store

**Stores with good quarterly growth rate in Q3’2012:**

By Calculating the quarterly growth rate for each store in Q3’2012 and identifying stores with significant positive growth.

# Convert the 'Date' column to datetime format

*walmart\_data['Date'] = pd.to\_datetime(walmart\_data['Date'])*

# Filter the data for Q3'2012 (July to September)

*Q3\_data = walmart\_data[(walmart\_data['Date'] >= '2012-07-01') & (walmart\_data['Date'] <= '2012-09-30')]*

# Calculate the quarterly growth rate for each store

*Q3\_growth = Q3\_data.groupby('Store')['Weekly\_Sales'].sum()*

# Find the store with the highest growth rate in Q3'2012

*best\_store = Q3\_growth.idxmax()*

*highest\_growth = Q3\_growth.max()*

print(f"Store with the highest growth rate in Q3'2012 is Store {best\_store} with growth rate: {highest\_growth:.2f}")

by running this code we get,

***Store with the highest growth rate in Q3'2012 is Store 4 with growth rate: 25652119.35***

**Find Holidays with higher sales than the mean sales in non-holiday season ?**

By Calculating the mean sales during non-holiday periods and identify holidays where sales exceeded this mean for all stores together.

By running the relevant codes , we get the answer as :

*Total Super Bowl Sales: $145682278.34 is higher than the mean non-holiday sales of $1041256.38*

*Total Labor Day Sales: $140727684.68 is higher than the mean non-holiday sales of $1041256.38*

*Total Thanksgiving Sales: $132414608.50 is higher than the mean non-holiday sales of $1041256.38*

*Total Christmas Sales: $86474980.04 is higher than the mean non-holiday sales of $1041256.38*

**Monthly and Semester view of sales:**

* To analyze the sales data with a monthly and semester view,
* Prepare the Data: Ensure your dataset includes sales units and dates.
* Group by Month: Group the data by month and sum sales units for each month.
* Calculate Semester Data: Sum sales units for two consecutive quarters to create a semester view.
* Visualize the Data: we used scatter and bar plots to show month of each year from 2010 to 2013 and as well as just each w.r.t weekly sales.
* Gain Insights: Analyze the plots to identify patterns, seasonality, and changes over semesters.
* This analysis helps in understanding sales patterns, seasonality, and long-term trends to make informed business decisions.These insights can help in making informed decisions, such as adjusting inventory, marketing efforts, or sales strategies based on sales trends.

**Creating Model :**

In this project, we aim to create a demand forecasting model for all Stores , using various statistical techniques. The primary goal is to predict sales accurately and understand the factors that impact sales.

**Linear Regression Model:**

Split the dataset into a training and testing set to evaluate model performance.

Select relevant features, such as day, CPI, unemployment, and fuel price, hypothesizing their impact on sales.

We Build a linear regression model to predict sales, using these features.

**Model Comparison:**

Then comparing the linear regression model's accuracy with other regression model Random Forest Regression

By comparing we found out that Random Forest Regression gives 94.37% Accuracy

***Random Forest Regressor:***

***Accuracy: 94.37298247904533***

***Mean Absolute Error: 75699.15892412013***

***Mean Squared Error: 18647639415.610523***

***Root Mean Squared Error: 136556.35985046806***