Final project
Retail Analysis with Walmart Data

Dataset Description

Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this assignment is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available.

This is the historical data that covers sales from 2010-02-05 to 2012-11-01, in which you will find the following fields:

- Store - the store number
- Date - the week of sales
- Weekly_Sales - sales for the given store
- Holiday_Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week
- Temperature - Temperature on the day of sale
- Fuel_Price - Cost of fuel in the region
- CPI – Prevailing consumer price index
- Unemployment - Prevailing unemployment rate

1. Perform exploratory data analysis
   a) ~~Import data~~
   b) ~~display data~~
   c) ~~visualize quantitive variables distributions~~
   d) ~~perform data cleaning~~

2. Answer the following questions:
    a) Which store has maximum sales?
    b) Which store has maximum standard deviation i.e., the sales vary a lot
    c) Some holidays have a negative impact on sales. Find out holidays that have higher sales than the mean sales in the non-holiday season for all stores together.
    d) Provide a monthly and semester view of sales in units and give insights.
    e) Plot the relations between weekly sales vs. other numeric features and give insights.

Dataset :
https://www.kaggle.com/datasets/varsharam/walmart-sales-dataset-of-45stores/download?datasetVersionNumber=1