WALMART PROJECT DOCUMNETATION

DESCRIPTION

One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales data available for 45 stores of Walmart. The business is facing a challenge due to unforeseen demands and runs out of stock some times, due to the inappropriate machine learning algorithm. An

ideal ML algorithm will predict demand accurately and ingest factors like economic conditions including CPI, Unemployment Index, etc.

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 competition 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.

Dataset Description

This is the historical data which covers sales from 2010-02-05 to 2012-11-01, in the file Walmart_Store_sales. Within this file 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

Holiday Events

Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
Labour Day: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

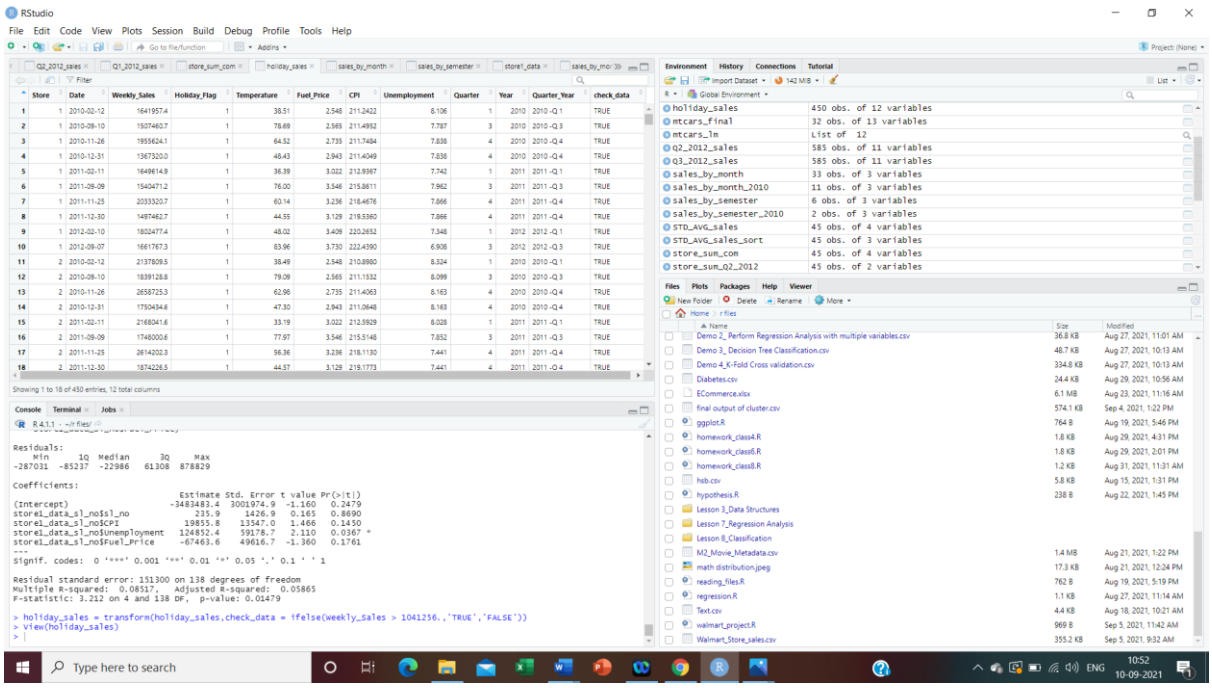
Analysis Tasks

Basic Statistics tasks

- Q1 -> Which store has maximum sales
- Ans ->The Store 20 has maximum Sales.(By using aggregate and Arrange function)

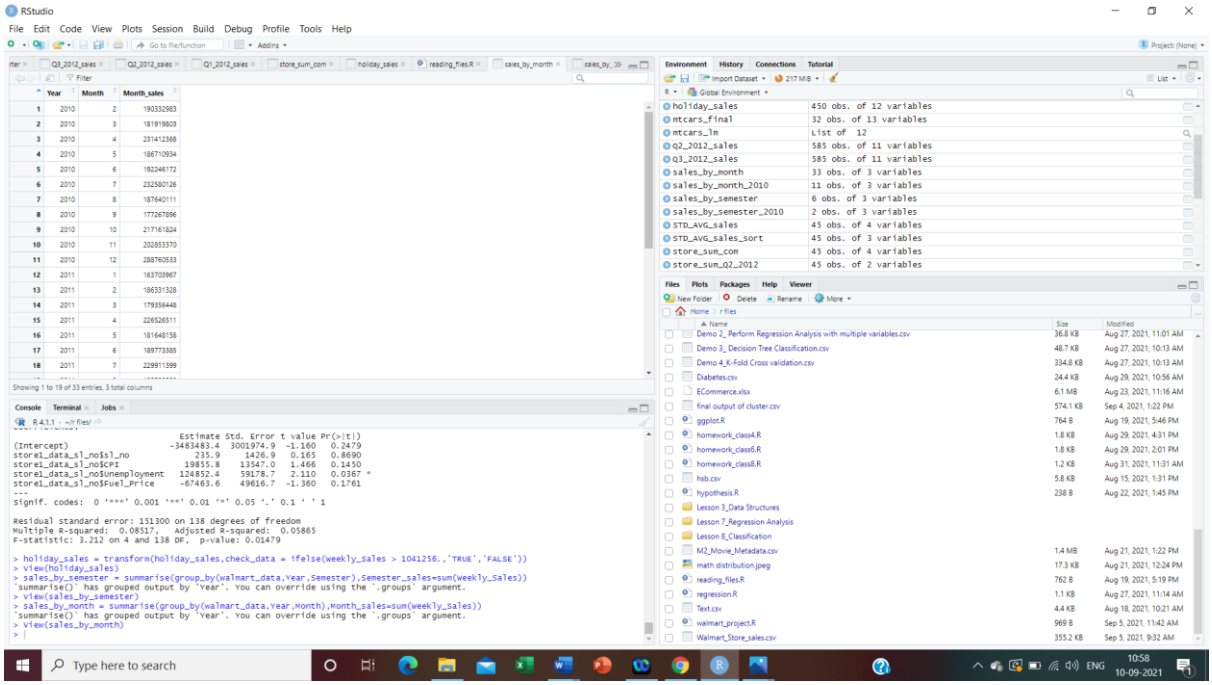
- Q2 -> Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation

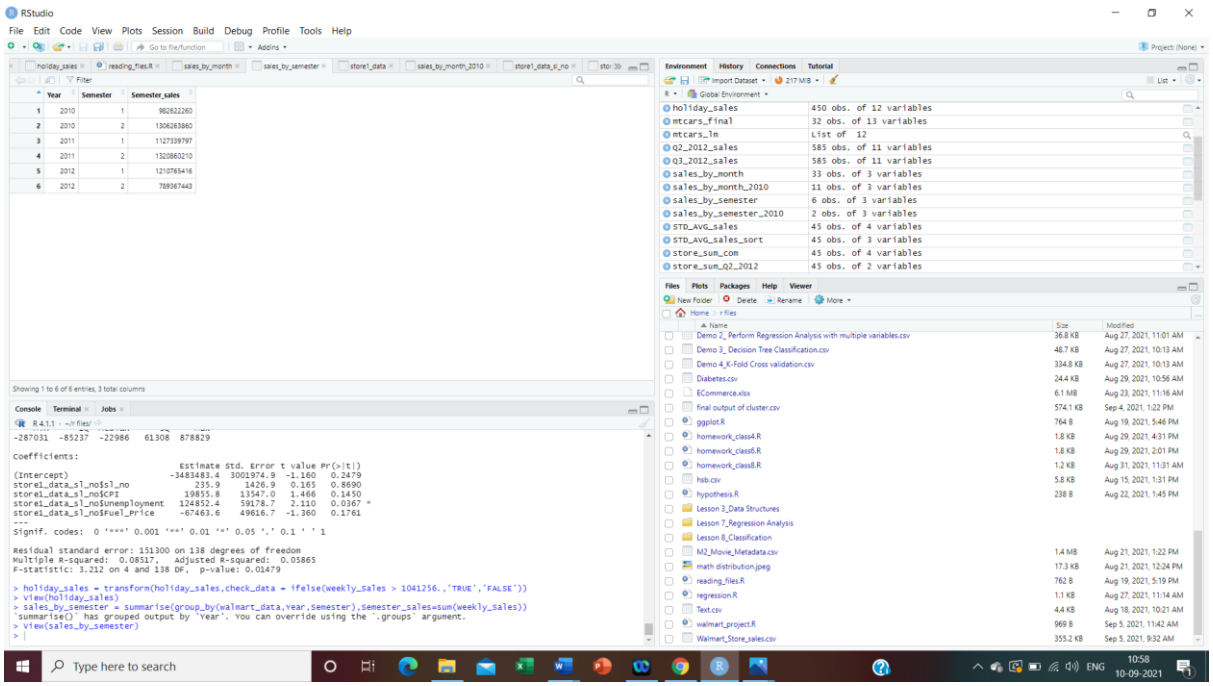
- Ans -> The Store with Maximum Standard Deviation is Store 14.(By using summarise with sd function and arrange function.)
- Q3 -> Which store/s has good quarterly growth rate in Q3’2012
- Ans -> The Store with Highest Quaterly Growth rate is Store 7.(By using filter function,summarise with sum function,Transform function and arrange function.)
- Q4 -> Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together



- Q5 -> Provide a monthly and semester view of sales in units and give insights

- Ans ->



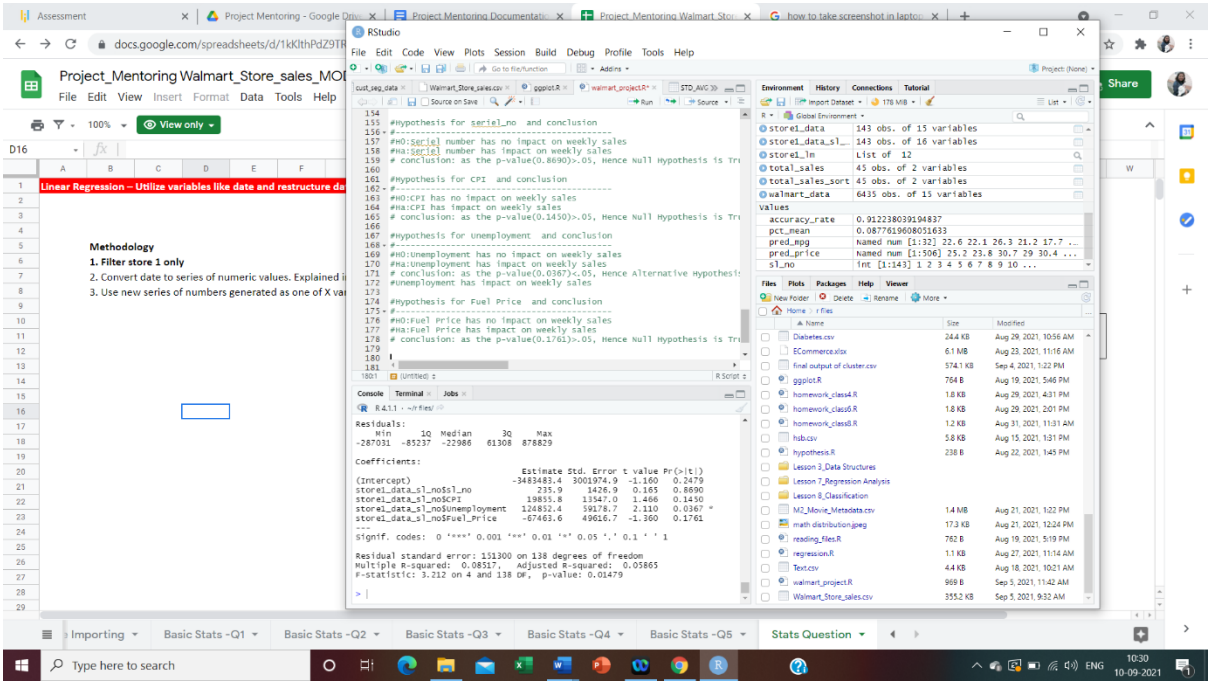


After data Visualization for Month sales it can be seen December month has the highest sales as compared to other Months and it's a outlier.

Form semester Sales , it can be seen that for year 2010 & 2011 the 2nd semester has better sales than the 1st semester. But for year 2012 the 1st semester has better sales than the 2nd semester.

Q6 -> Statistical Model

For Store 1 – Build prediction models to forecast demand

- Linear Regression – Utilize variables like date and restructure dates as 1 for 5 Feb 2010 (starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.
- Change dates into days by creating new variable.
- Ans ->
- Hypothesis for Seriel number,CPI,Unemployment,Fuel Price
- 

- The Accuracy rate is 93%
-