

## RETAIL ANALYSIS WITH WALMART DATA

**In this project we are given a data-set consisting of 45 stores of Walmart. Following are the attributes given:**

1. Store - the store number
2. Date - the week of sales
3. Weekly\_Sales - sales for the given store
4. Holiday\_Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week
5. Temperature - Temperature on the day of sale
6. Fuel\_Price - Cost of fuel in the region
7. CPI – Prevailing consumer price index
8. 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

### Basic Statistics tasks

1. Which store has maximum sales

Sol: We first calculate the total weekly sales of each store and then find out the maximum of the total weekly sales and the corresponding store number.

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

Sol: We first calculate the standard deviation of each store and then find out the maximum standard deviation and the corresponding store number. Then we find out the mean of total weekly sales of every store and store it along with their corresponding standard deviation.

3. Which store/s has good quarterly growth rate in Q3'2012

Sol: We calculate the growth of each store for Q3'2012 and then find the highest growth and the corresponding store number.

4. 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

Sol: We first calculate the mean sales of all the stores together during non-holiday season, then we calculate the sales of on each holiday week for all the stores and we compare them. We print those values of holiday weeks which have sales value greater than the mean sales.

5. Provide a monthly and semester view of sales in units and give insights

Sol: For monthly view we calculate the sales for each month for all stores for the year 2010, 2011 and 2012. Whereas for semester view we calculate the sales for each 6 months for all the stores for the year 2010, 2011 and 2012.

### Statistical Model

For Store 1 – Build prediction models to forecast demand

1. 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.

Sol: Firstly, we restructure the dates given for store 1 in the form of discrete number starting from 1 to 143. Then we form training and testing set consisting of

1. All the attributes
2. Removing CPI
3. Removing unemployment
4. Removing Fuel price
5. Removing CPI, unemployment, Fuel Price

We deploy a linear regression model and calculate the R<sup>2</sup> score as well as the MSE for each case and check which attributes are affecting the model more.

2. Change dates into days by creating new variable.

Sol: We add a new column called days and find out the corresponding day for each date in the date column.

3. Select the model which gives best accuracy.

Sol: We deploy Linear Regression as well as KN Regression models. Then we check the R2 score for both the models. The one with the appropriate score is selected.