1] WALMART

Dataset Description

This is the historical data that 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

Which store has maximum sales

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

Which store/s has good quarterly growth rate in Q3’2012

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

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

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

2] SCHOOLS

This dataset is especially useful for starters in their Data Science journey, also whoever is interested in knowing how the education in India is progressing over the years. You can start your exploration by answering the below questions:

* Which states have the highest Dropout Ratio?
* Do boys and girls have enough water and toilet facilities available?
* How is the Gross Enrollment ratio for boys and girls in various levels of schooling life?
* Which level of school life is there less number of enrollment? Is it dropping or increasing?
* Any other ideas

STEPS:

1. DATA BLENDING / JOINS – CUSTUM SQL
2. PARAMETER
3. Context FILTER
4. ACTION FILTER
5. GROUPS
6. DYNAMIC SETS
7. COMBINED SETS
8. TABLE CALCULATIONS
9. MOVING AVG
10. TRENDS
11. FORECASTING WITH R
12. ANNIMATION
13. LOD
14. YTD, MTD, QTD
15. CUSTOM CHARTS
16. BAR-IN-BAR CHART
17. MATRIX CHART – CORRELATION
18. SWAP CHARTS
19. MAPS
20. CONDITION FORMATING
21. TABLEAU PREP – DATA PREP
22. DATE CODES = DATE DIFF, DATEADD, MIN,MAX,
23. DATE AND TIME FUNCTIONS
24. CASE
25. UP-DOWN ARROWS
26. NESTED IF , ELSEIF
27. INDEX, LAST, RANK
28. CLUSTERS
29. KIRIL COAL PROJECT
30. USERFORM