Sales Forecasting Challenge

Problem Statement

Predicting the sales values of SKUs is important for any business. It helps the business plan better, making it grow faster. This challenge is associated with a similar problem, where weekly forecast must be generated for ABC Retail Pvt. Ltd. This retail giant has multiple stores, and each store has multiple departments. The department head has the responsibility in each store to maintain the inventory, keep track of stock and make demands based on market condition. This challenge can be utilized by Data Scientists to create state-of-the-art solutions for this retailer, to help them predict the sales based on the given dataset.

Objective

To predict weekly sales for each department in each store. Training data consists of weekly level sales from Feb 2010-Apr 2012. The forecasts must be submitted for the time-period May 2012 – Oct 2012.

Data Description

This challenge will contain a train dataset and a test dataset.

train_sales_forecast.csv

This is the training dataset consisting of 10 stores, store 1-10, each store consisting of 99 departments and weekly sales of that department in the store.

Data Dictionary

- Store: Store ID (1-10)
- Dept: Department ID in that store (1-99)
- Weekly_Sales: Weekly sales for the department (Target Variable)
- IsHoliday: Does the week contains holiday or not (Binary variable – True/False)

Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	05-02-2010	24924.5	FALSE
1	1	12-02-2010	46039.49	TRUE
1	1	19-02-2010	41595.55	FALSE
1	1	26-02-2010	19403.54	FALSE
1	1	05-03-2010	21827.9	FALSE
1	1	12-03-2010	21043.39	FALSE
1	1	19-03-2010	22136.64	FALSE

test_sales_forecast.csv

This is the test data without the column 'Weekly_Sales'. You must compute the values to this column, and accordingly MAPE will be used to evaluate the results.

Store	Dept	Date	IsHoliday
1	1	04-05-2012	FALSE
1	1	11-05-2012	FALSE
1	1	18-05-2012	FALSE
1	1	25-05-2012	FALSE
1	1	01-06-2012	FALSE
1	1	08-06-2012	FALSE
1	1	15-06-2012	FALSE

Submission file

The submission file should contain all the features and records from 'test_sales_forecast.csv' file and an added column with the name 'Weekly_Sales'. This column will be used to evaluate with actual values for score calculation.

Metric

The metric used in this challenge is MAPE, computed as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where, $MAPE$ = Mean Absolute Percentage Error

n = Number of times the summation iteration happens

A_t = Actual Value

F_t = Forecast Value