**PREDICTING SALES FOR MULTIPLE WAL-MART STORES**

- David Gill *(Data 606)*

**GOAL:**

The goal of the project is to predict the sales for each triplet of the store, department and the date of its occurrence – in the test.csv file. Part of the challenge would be building a model to cater for markdowns on the given holidays, where ideal/complete historical data is absent.

**DATA SOURCE:**

Historical sales data for 45 Walmart stores located in different regions, containing a number of departments has been downloaded from [*https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data*](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data). The data set consists of 4 csv files –

1. **Stores.csv** – This file contains anonymized information about the 45 stores. Attributes include number, type and size of store.
2. **Features.csv** – This file contains additional data pertaining to the stores. Attributes include store number, date(week), temperature *(average temp. in the region)*, fuel price in the region, CPI *(consumer price index)*, unemployment rate, markdown1-5 *(promotional markdowns values)*,is holiday week *(whether a week is a special holiday week or not).*
3. **Train.csv –** This file contains historical training data, from Feb'2010 to Nov'2012. Attributes include store number, dept. number, date(week), weekly sales *(sales for a given department in a given store)*, is a holiday week *(if the week has a special holiday or not)*.
4. **Test.csv** - This file is identical to train.csv, excluding the weekly sales.

In addition, Walmart runs several promotional markdown events throughout the year, which precede prominent holidays, the four largest being - the Super Bowl, Labor Day, Thanksgiving, and Christmas. Holiday weeks are weighted five times higher in the evaluation than non-holiday weeks. This information has also been provided.

**METHODOLOGY:**

To fulfil the goal the methodology I have planned to adopt is to use -

1. Csv files for storage on the local machine
2. R/Python to perform data cleansing and exploratory data analysis
3. R/Python to build a Model for training and testing purposes
4. R/Python to predict the sales
5. PowerBI/Tableau to present the data in meaningful visuals

**CONSEQUENCES/USES:**

Walmart is one of the oldest shopping marts in the US and now has international branches. It provides employment and services to a huge population. Being able to predict sales for the stores would be highly beneficial and would have the following consequences –

1. Scrupulous management of revenue.
2. Budget forecasting would be more accurate.
3. Stock take/ inventory control management would become more efficient, knowing the commodity being sold in which stores.
4. Growth of the stores could be planned in and effective manner.
5. Intelligent business decisions could be made ensuring profit for the stores.