|  |
| --- |
| Walmart Store Sales Forecasting |
| Data Mining Project Plan |
| http://www.cs.cmu.edu/~sjakkamr/images/cmu.png  Aakriti Bhargav  Aditya Padhe  Shubham Jaiswal  Smriti batra |
|  |



Table of Contents

[1. Plan Brief 2](#_Toc383727790)

[2. Business Background 2](#_Toc383727791)

[3. Data Background 2](#_Toc383727792)

[4. Data Structure 3](#_Toc383727793)

[Qualitative 3](#_Toc383727794)

[5. Objective 5](#_Toc383727795)

[6. Schedule of the Project 5](#_Toc383727796)

[7. Tools and Techniques 6](#_Toc383727797)

# 1. Plan Brief

The data we have chosen to analyse is up for competition on Kaggle. The competition is titled ‘Walmart Recruiting- Store Sales forecasting’ and provides information of 45 Walmart stores for analysing their sales.

As part of the competition, participants need to project the sales of each department of the 45 Walmart stores that are located in various regions. Along with the sales data of all the stores, the dataset includes details of a few holiday markdown events which affect sales as well.

# 2. Business Background

Walmart is an American multinational retail company that operates chains of more than 11,000 department stores in 27 different countries. Since Walmart operates on extremely low profit margins and serves a very large customer base, it relies heavily on consumer data analytics to project and improve sales.

# 3. Data Background

The data is a mix of numerical and nominal fields and is available in the form of 4 sub-datasets. They are:

* stores.csv

This file contains the type and size information of 45 different Walmart stores based in different locations.

* train.csv

This file contains the historical data from 2010-02-05 to 2012-11-01. This data contains the weekly sales data of different departments.

* test.csv

This is the test data that must be used for the testing and validation of results obtained from training data.

* features.csv

This file contains additional data related to markdowns and events which impact sales. It also contains other regional data that can have an effect on sales, such as unemployment rate, consumer price index etc.

# 4. Data Structure

**Train**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Data Type** | **Measurement** | **Description** |
| Store | Qualitative | 1 - 45 | Represents the Store Number and identifies it uniquely |
| Department | Qualitative | 1-99 | Represents the Department number |
| Date | Qualitative | MM/DD/YY | Represents the date the sales of a particular store was provided |
| Weekly\_Sales | Quantitative | Dollar | Represents sales for the given department in the given store |
| IsHoliday | Qualitative | TRUE (week is a special holiday week)  FALSE (week is not a special holiday week) | Provides information whether the week is a special holiday week or not i.e. True or False |

**Features**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Data Type** | **Measurement** | **Description** |
| Store | Qualitative | 1 – 45 | Represents the Store Number and identifies it uniquely |
| Date | Qualitative | MM/DD/YY | Represents the date the sales of a particular store was provided |
| Temperature | Quantitative | Fahrenheit | Represents the average temperature of the region |
| Fuel\_Price | Quantitative | Dollar/gallon | Represents the cost of fuel in the region |
| MarkDown1 - 5 | Quantitative | Dollar | Represents the anonymized data related to promotional markdowns that Walmart is running |
| CPI | Quantitative | Index | Represents the consumer price index |
| Unemployment | Quantitative | Percentage | Provides information about the unemployment rate |
| IsHoliday | Qualitative | TRUE (week is a special holiday week)  FALSE (week is not a special holiday week) | Provides information whether the week is a special holiday week or not i.e. True or False |

**Stores:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Data Type** | **Measurement** | **Description** |
| Store | Qualitative | 1 - 45 | Represents the Store Number and identifies it uniquely |
| Type | Qualitative | A, B or C | Provides information about the type of Store whether A. B or C |
| Size | Quantitative | 34875 - 219622 | Indicates the size of the store |

# 5. Objective

The objective of the project is to predict the weekly department wide sales for each of the 45 Walmart stores present in the test data set. We have to predict the sales for each week starting from 2nd November 2012 taking various factors into consideration.

The factors affecting the sales are as follows:

Store Level: Store size, Store type

Store locality features: Temperature, Fuel prices etc.

Promotions: Promotional Markdowns in that particular week.

# 6. Schedule of the Project

Data Preparation

The dataset is spread across multiple sub-datasets which will be combined over appropriate attributes and combined to make a dataset suitable enough for mining. We will also understand the various attributes of the data and comprehend their relevance in the context of the dataset. We will use Microsoft Excel to combine different datasets during the preprocessing stage.

Modelling

We will select the modelling technique after we have thoroughly understood the data. However, based on our preliminary understanding of data, we will use clustering to divide the data into chunks considering that the dataset is unsupervised. Further, we will use Association techniques to find relations between various attributes in the various clusters. To project the sales, we will use regression techniques.

Test Design

Training and Test datasets are available. The model will be built using the training set and evaluated on the test dataset. The error rate of various data models will be used as a measure to gauge their accuracy.

Evaluation and Recommendation

The model generated from the training data set will be used to project the sales of the various departments of the 45 Walmart stores based in different locations.

# 7. Tools and Techniques

The following tools and techniques will be considered for use in the project (subject to change):

* Data Mining Tool: Weka 3.6, Minitab and R
* Preprocessing Tool: Microsoft Excel
* Visualization Tool : Tableau

Reference

http://www.brooklynews.com/wp-content/uploads/2012/09/walmart-store2.jpg