

# Kaggle Competition: Walmart Weekly Sales Forecasting

William West

April 26, 2014

## 1 Introduction

Kaggle is a website that hosts data mining competitions. Typically, companies will organize a competition on Kaggle by offering a sample dataset, a well-defined problem, and some incentive for competing, whether it be some monetary prize, internship, or job. Data Scientists can then compete with one another for the best performance, sometimes collaborating with one another and learning new things in the process.

The Walmart Weekly Sales Forecasting competition asks competitors to predict the weekly sales for a set of store/department pairs in 2013, given the weekly sales from 2010-2012.

## 2 Dataset Description

The dataset consists of four data files: *features.csv*, *train.csv*, *test.csv*, and *stores.csv*.

- The *train.csv* and *test.csv* files contain the weekly sales for each (store, department, date) triple. Note that the test file contains null entries for the weekly sales column, as expected
- The *features.csv* file contains the temperature, fuel price, markdowns, CPI, unemployment rate, and a holiday indicator for each (store, department, date) triple
- The *stores.csv* file contains the size and type of each store

We combine all files into two distinct datasets—the training set and test set. Each set contains all features and the weekly sales for each (store, department, date) triple. We will assume that these files are combined for the remainder of this report.

### 3 Data Exploration

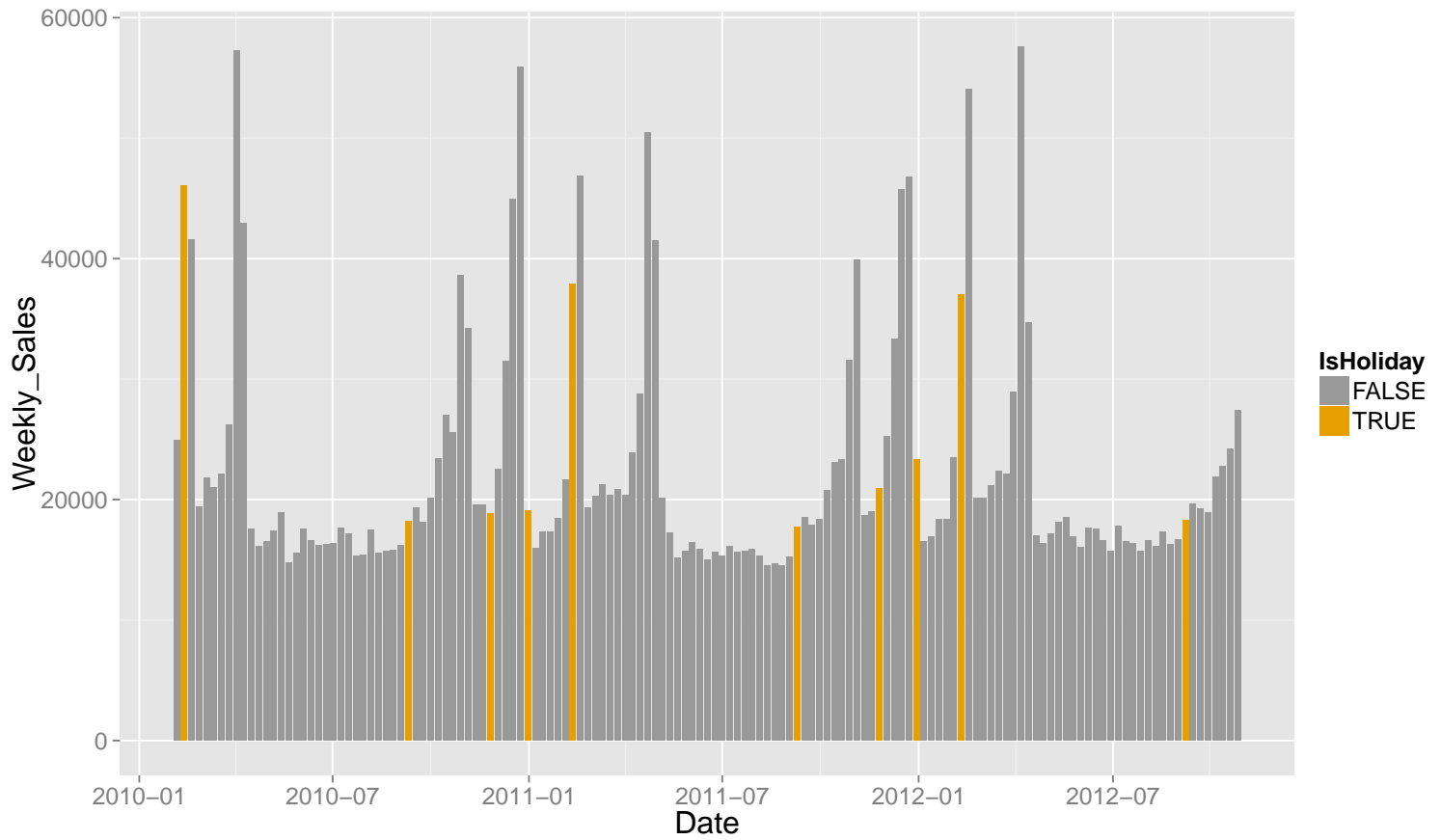


Figure 1: Weekly Sales of store.1, highlighting weeks marked as holidays

### 4 Feature Extraction

### 5 Modeling

### 6 Evaluation

### 7 Conclusion

### References