# WALMART SALES FORECASTING

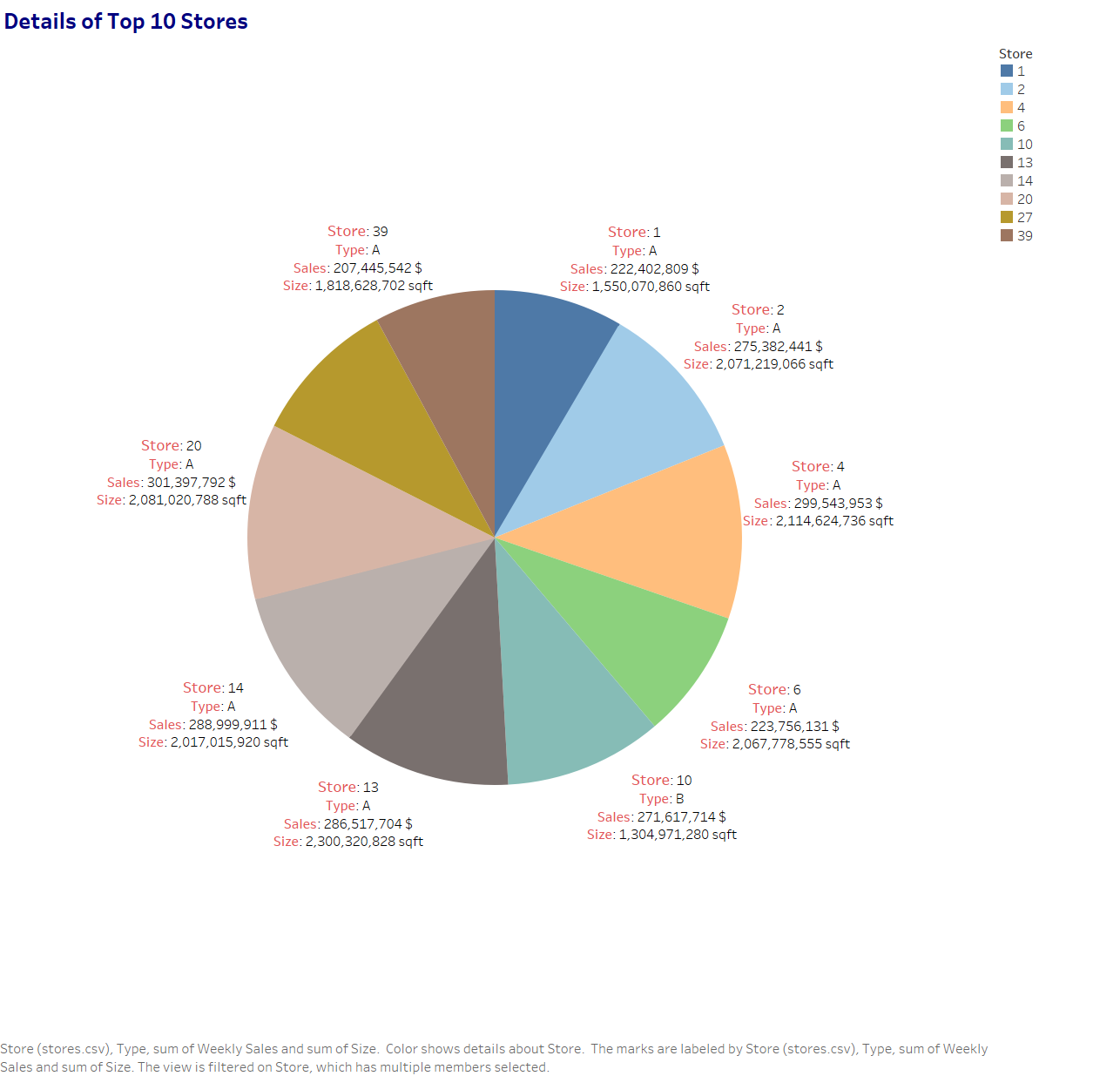
## Source for dataset and problem statement:

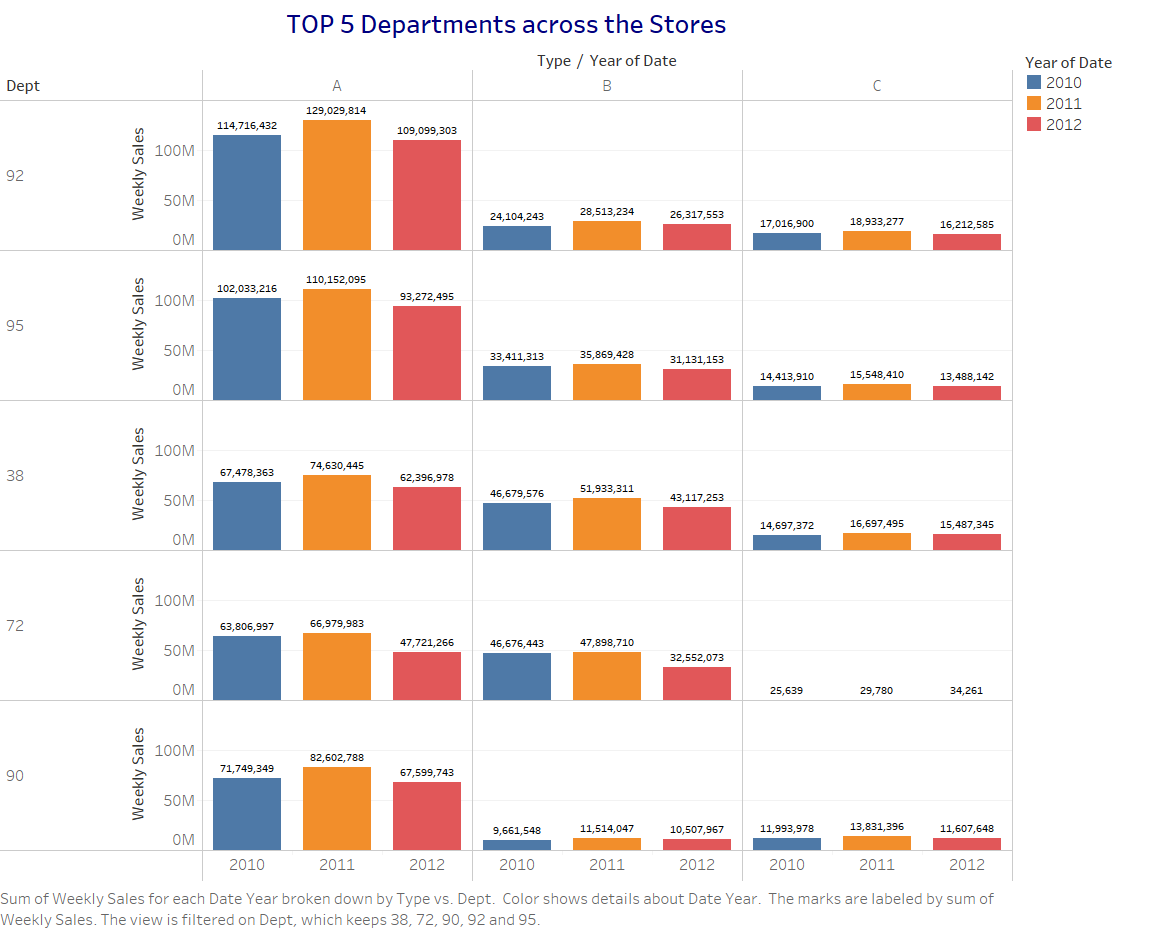
<https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data>

## Introduction

* Super store, Walmart has a huge impact on their sales during markdown events like the Super Bowl,President’s day and other holidays like Black Friday, Christmas,etc.
* The reason being there are several promotional offers preceding these holidays.
* Therefore, there is a necessity to analyze the sales data set of the MegaStore Walmart, during markdown events in order to plan more efficient strategies for improving sales

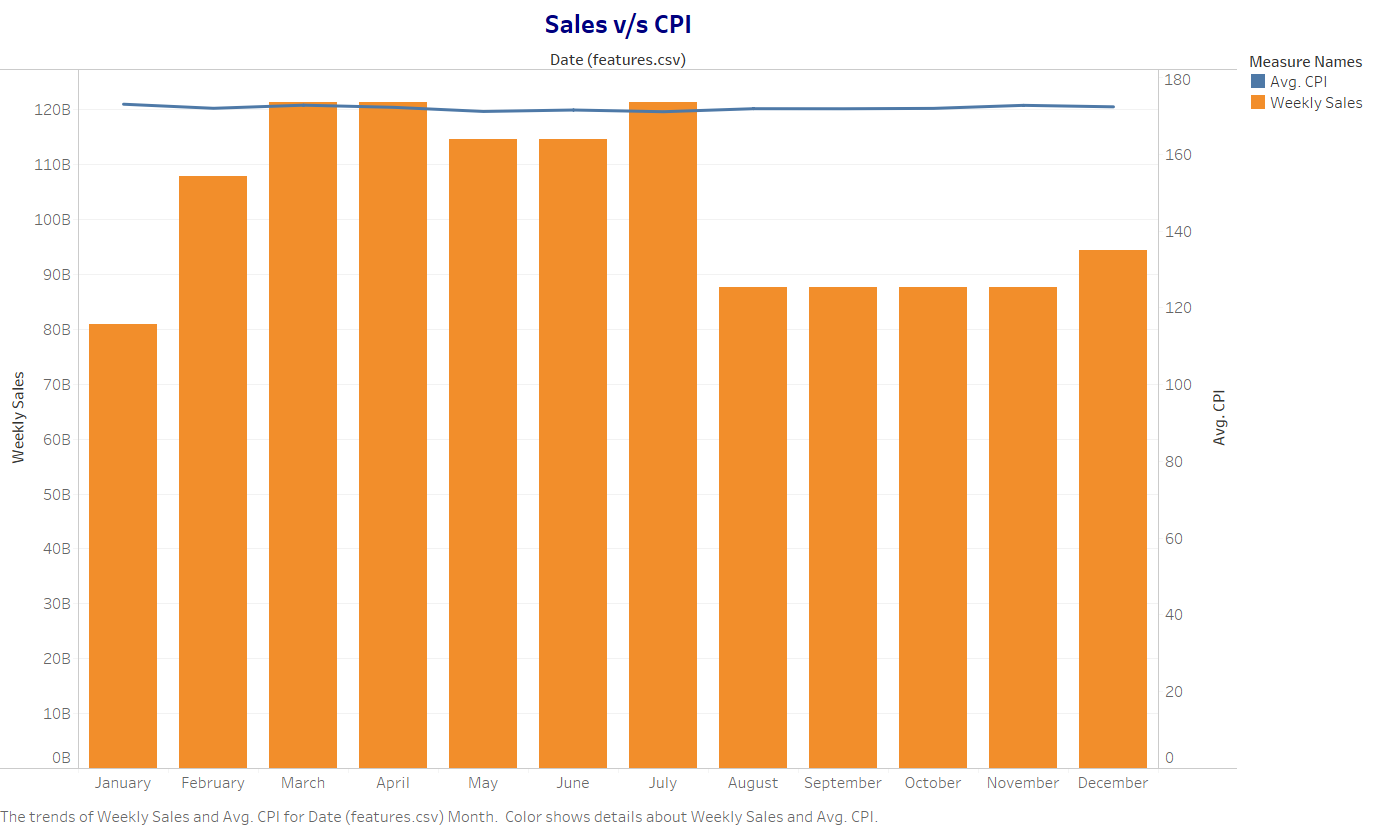
## Exploratory Analysis

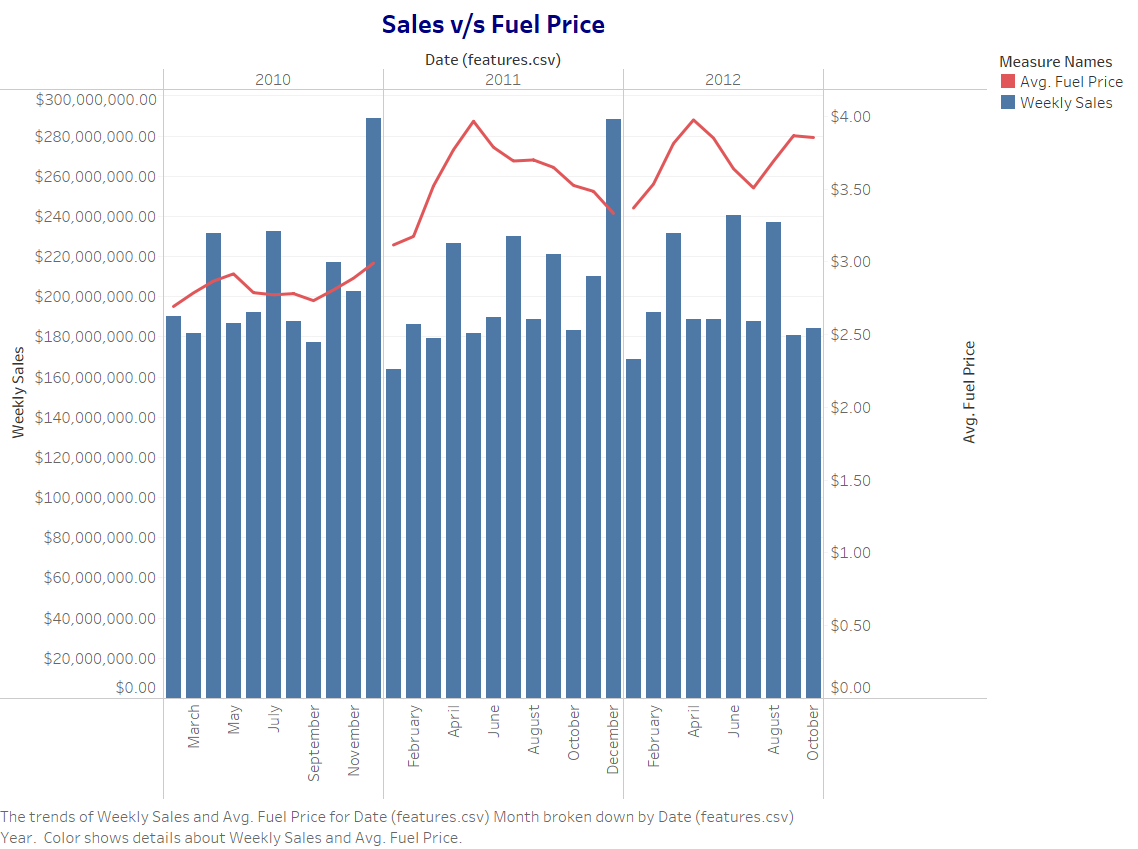




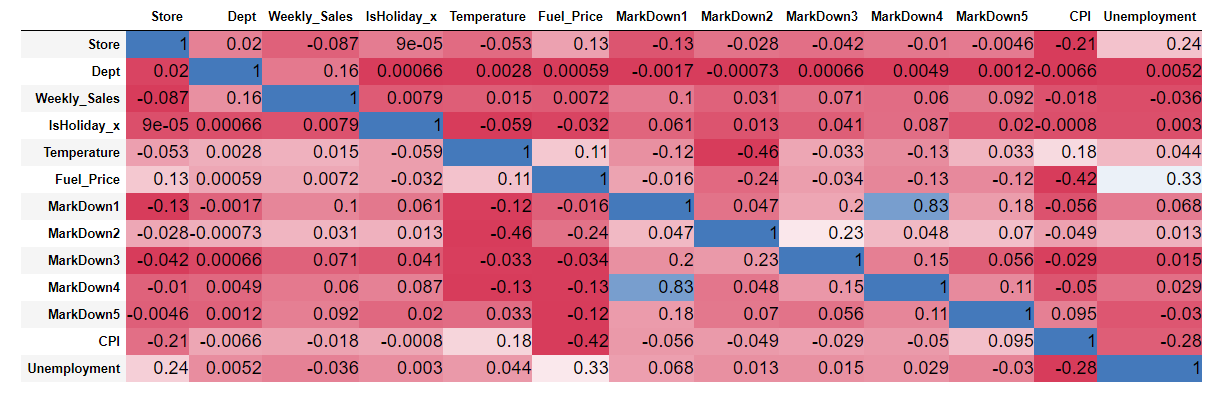
|  |  |
| --- | --- |
| **Store Number** | **Cumulative Weekly Sales($)** |
| 92 | 483,943,342 |
| 95 | 449,320,163 |
| 38 | 393,118,137 |
| 72 | 305,725,152 |
| 90 | 291,068,469 |

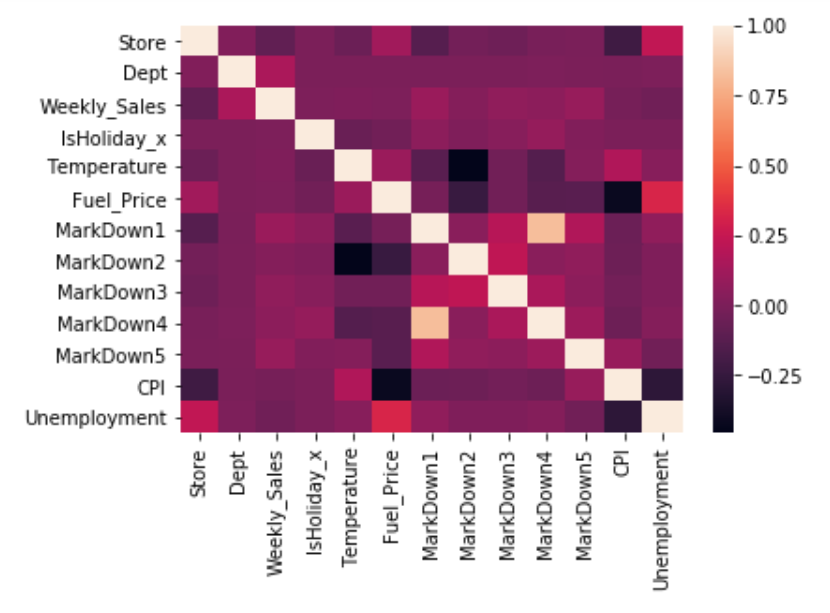
* No strong relationships were clear from visualizing the weekly sales data with respect to the CPI and the fuel price during that week.





## Autocorrelation Heat Map





## Process:

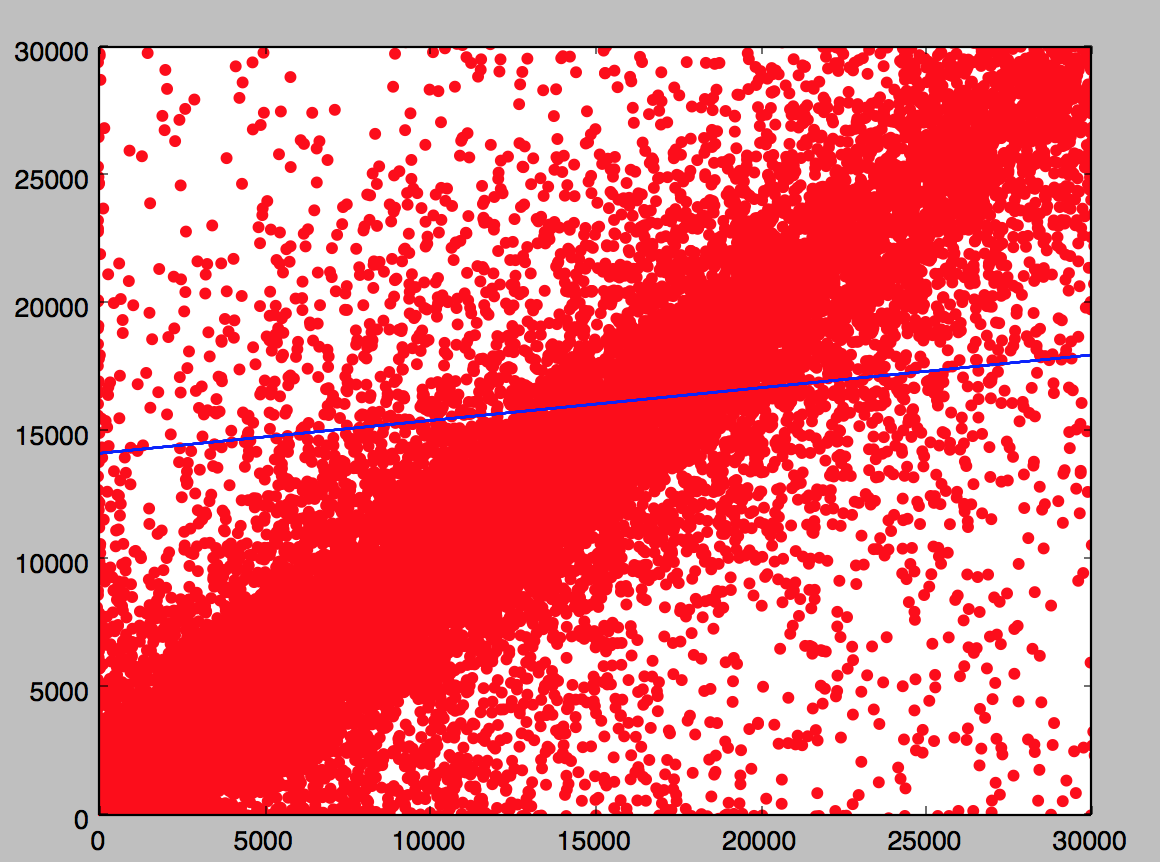
### Data Preparation: Merging, Cleaning, and Transforming the Data

* We are merging train.csv and features.csv based on Store numbers and Date.
* From the Merged Data, we are removing records dated from Jan 2011 to November 2011, since the markdowns(1-5) in features.csv for this period is unavailable.
* For the remainder of Merged Data, we found missing values in Markdown(1-5) columns. We are filling these missing values with mean grouped by store and department.
* Test.csv consists of Store, Department and Date columns dated from November 2012 to July 2013.
* Only for the computation of accuracy score, we are dividing the train data in 70:30 ratio.

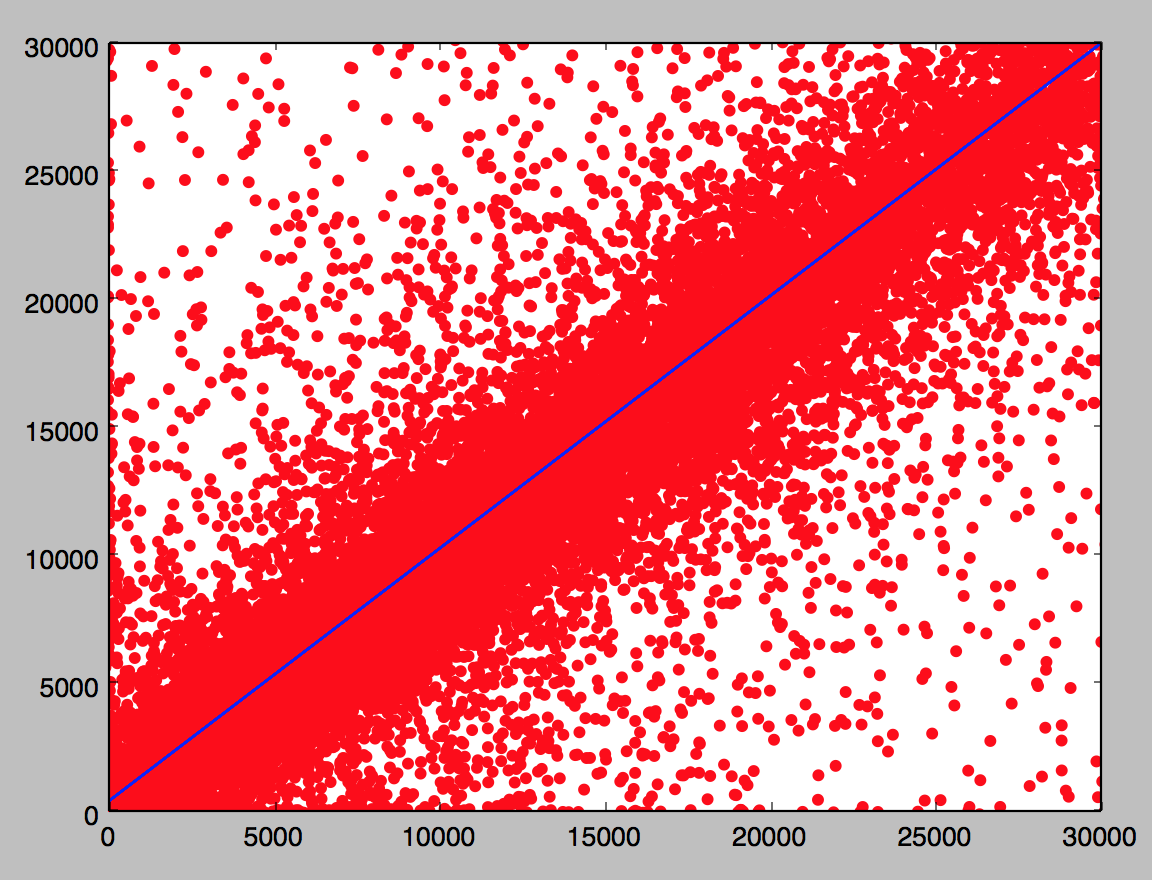
## Supervised Modelling

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Linear Regression** | **Random Forest Regressors** | **Extra Trees Regressors.** |
| R2 | -5.43 | 0.86 | 0.87 |
| Mean Log Square error | 0.91 | 0.67 | 0.65 |

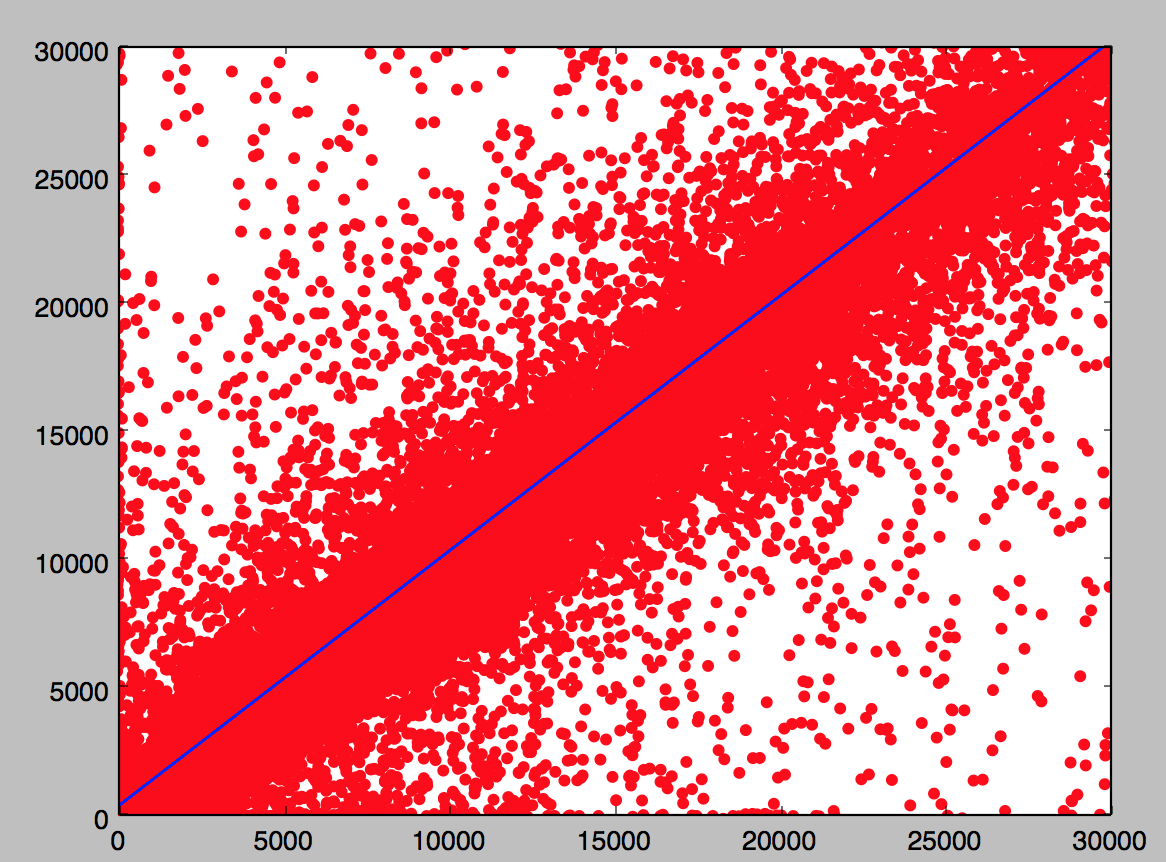
* **Linear Regression**



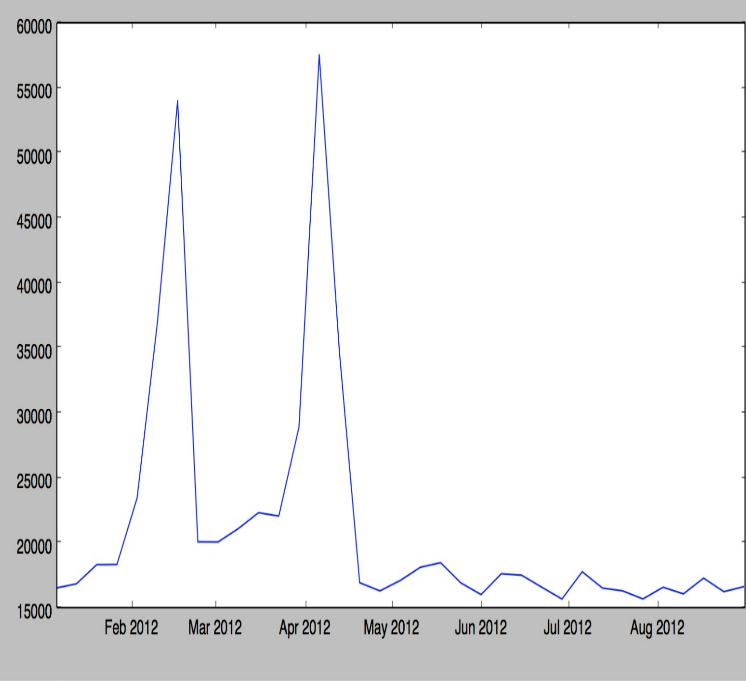
**Random Forest Regressors**

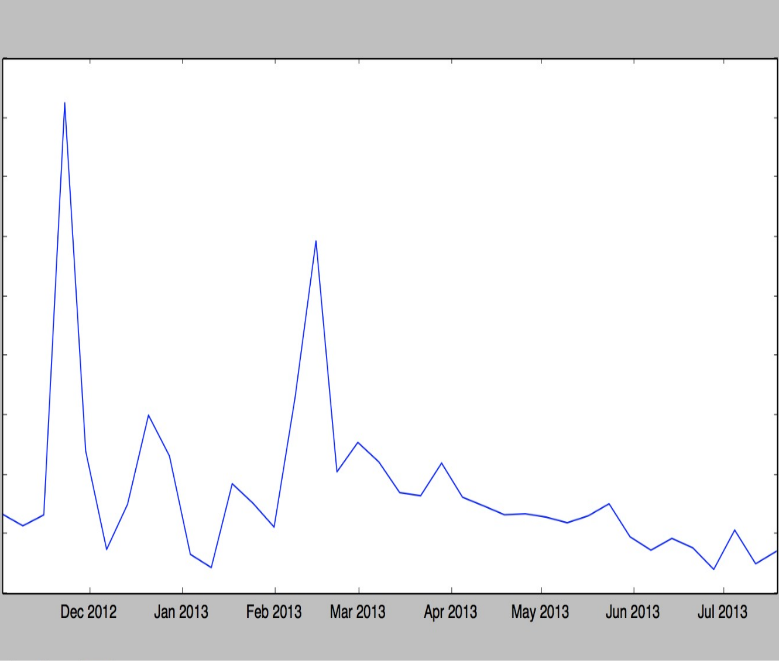


**Extra Trees Regressors**



## Sales Trend Analysis on Modeled Data





## Conclusion and Improvements

* Performance : Random Forest Regressors, Random Forest Regressors > Linear Regression
* Sales :Sales increase during the months February, April, November and December which is in accordance with our Trend Analysis.