ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Retail Sector Analysis and Forecasting using Seasonal - Trend Decomposition using Loess

Shwetha Chippa[1], Kiran Chavan[2], Isha Deshpande [3], Rashmi Jolhe [4]

U.G. Student, Department of Information Technology, Datta Meghe College of Engineering, Airoli, Mumbai, India[1]

U.G. Student, Department of Information Technology, Datta Meghe College of Engineering, Airoli, Mumbai, India[2]

U.G. Student, Department of Information Technology, Datta Meghe College of Engineering, Airoli, Mumbai, India[3]

Asst. Professor, Department of Information Technology, Datta Meghe College of Engineering, Airoli, Mumbai, India[4]

ABSTRACT:Most companies like to estimate the upcoming trades. A superior forecasting can avoid them from over-estimating or under-estimating the longer-term trades which results in an excellent harm to the businesses.Using reliable trade calculation, companies could assign their properties more sensibly and make improved revenues. But projecting the sales is rather complicated due to several internal external factors from the neighbouring atmosphere. Hence within the current marketplace, there's an urgent requirement for the progress of a sensible forecasting system which is fast, flexible and may provide high accuracy. We intend to put on various machine learning strategies to build and alter a business anticipating demonstrate and perform estimation on deals information to return over this necessity. We also shall provide an easy to use result with various visualization tools for the easiness of users. Our project also works on analysis of the various parameters which could affect the sales.

KEYWORDS:Season, Trend, Reminder, Decomposition, Walmart, Weekly_Sales.

I. INTRODUCTION

The aim of this project is to enable category managers of Walmart to check weekly sales of departments. Analysis includes the effect of the markdown on sales, and also the extent of effect on sales by fuel prices, temperature unemployment, CPI etc has been analysed using simple and multiple linear regression models.

One challenge of modelling retail data is that the got to make decisions supported limited history. If Christmas comes but once a year, so does the prospect to ascertain how strategic decisions impacted rock bottom line. In this project, Walmart company are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and we must project the sales for each department in each store. To add to the challenge, selected holiday markdown events are included within the dataset. These markdowns are known to affect sales, but it's challenging to predict which departments are affected and therefore the extent of the impact.

II. FLOW OF PROJECT

The work in this paper is divided in several stages.

1) Input and Output. 2) Analysis of Parameters. 3) Handling Difficulties for each forecasting object. 4) Forecasting through STL Algorithm. 4) Producing output.

Input and Output is for the explanation of every variables which are presented in the datasets. Analysis of Parameters is for understanding and explanation of conducting analysis process over the dataset's every column. Selecting the variables which actually shows correlation, covariance or dependency over the target variables. Then we could understand the variation and difficulties through we need to deal with. After, removing every issue occurs processing throughseason trend decomposition using loess algorithms. Producing its output.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

III. INPUT AND OUTPUT

Input given historical sales data for 45 Walmart stores located in several regions. Each store contains variety of departments, and We have tasked with predicting the department-wide sales for every store. Additionally, Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labour Day, Thanksgiving, and xmas. The weeks including these holidays are weighted five times higher within the evaluation than non-holiday weeks. a part of the challenge presented by this competition is modelling the consequences of markdowns on these holiday weeks within the absence of complete/ideal historical data.Stores.csv - This file contains anonymized information about the 45 stores, indicating the type and size of store.

1) Stores.csv - This file contains anonymized information about the 45 stores, indicating the kind and size of store.
2) Train.csv - This is often the historical training data, which covers to 2010-02-05 to 2012-11-01. Within this file you'll find the subsequent fields:
    - Store - the shop number
    - Dept - the department number
    - Date - the week
    - Weekly_Sales - sales for the given department within the given store
    - IsHoliday - whether the week may be a special holiday week
3) Test.csv - This file is just like train.csv, except we've withheld the weekly sales. you want to predict the sales for every triplet of store, department, and date during this file.
4) Features.csv - This file contains additional data associated with the shop , department, and regional activity for the given dates. It contains the subsequent fields:
    - Store - the shop number
    - Date - the week
    - Temperature - average temperature within the region
    - Fuel_Price - cost of fuel within the region
    - MarkDown1-5 - anonymized data associated with promotional markdowns that Walmart is

running. MarkDown data is merely available after Nov 2011, and isn't available for all stores all the time. Any missing value is marked with an NA.

- CPI - the buyer price level
- Unemployment - the percentage
- IsHoliday - whether the week may be a special holiday week.

For convenience, the four holidays fall within the subsequent weeks within the dataset (not all holidays are within the data):

Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13.
Labor Day: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13.
Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13.
Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13.

Output contains the overall forecasted sales graph with confidence intervals. With providing the accuracy and analysing the output using residuals. As we need to process each department of each store independently, we need to combine the forecasted sales of all the applied forecasting algorithms independently. Thus, we can provide a single output graph which shows the average forecasted sales of Walmart.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

IV. ANALYSIS OF PARAMETERS

For the Analysis of various parameters. Need to Understanding of various attributes present in every table. Also need to analysis of dependency of those parametes to the target variable i.e., Weekly_Sales variable. If any Variable found to be showing interest. Then choose the variables which could be useful for further process.
Now Coming toward the variables checking. Having a scatter plot with the dependent variable should give us an distribution of the data with the Weekly_Sales variable. So, for such process looking towards some variables.

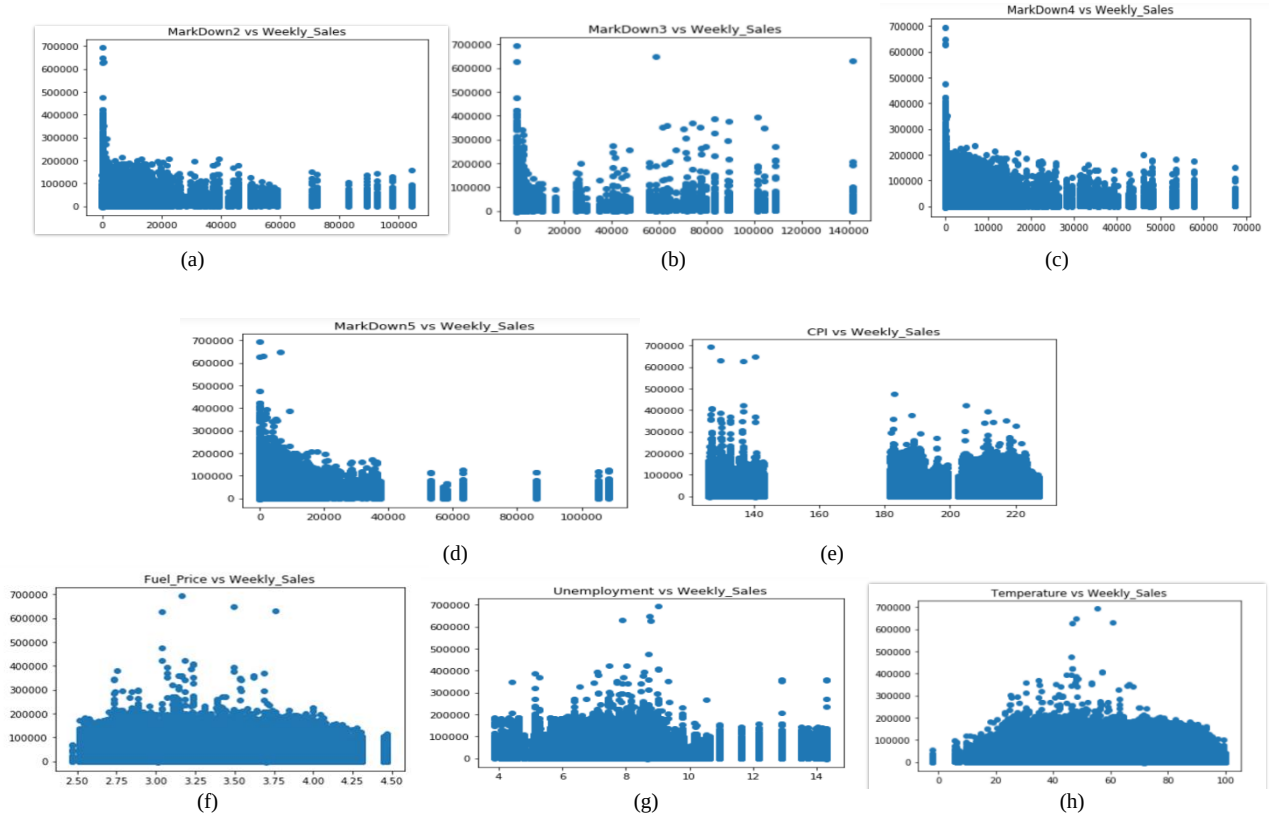


Fig.1.Analysis of various parameters (a) Markdown2 prices to weekly_sales(b) Markdown3 prices to weekly_sales (c) Markdown4 prices to weekly_sales (d) Markdown5 prices to weekly_sales (e) CPI index on every day to weekly_sales (f) fuel_price on stores to weekly_sales (g) unemployement of area to sales (h) temperature to every store

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Correlation of the variables and VIF values are shown as follows:

CPI	-0.020930
Dept	0.148077
Fuel_Price	-0.000124
IsHoliday	0.012762
MarkDown1	0.047161
MarkDown2	0.020703
MarkDown3	0.038554
MarkDown4	0.037457
MarkDown5	0.050452
Store	-0.085213
Temperature	-0.002318
Unemployment	-0.025838
Weekly_Sales	1.000000
store_b	-0.131224
store_c	-0.095393
store_size	0.243856

(a)

```
for  CPI  value of vif is: 1.28
for  Dept  value of vif is: 1.0
for  Fuel_Price  value of vif is: 1.17
for  IsHoliday  value of vif is: 1.16
for  MarkDown2  value of vif is: 1.11
for  MarkDown3  value of vif is: 1.09
for  MarkDown4  value of vif is: 1.14
for  MarkDown5  value of vif is: 1.19
for  Store  value of vif is: 1.39
for  Temperature  value of vif is: 1.21
for  Unemployment  value of vif is: 1.2
for  store_b  value of vif is: 2.46
for  store_c  value of vif is: 2.75
for  store_size  value of vif is: 3.1
```

(b)

Fig. 2. (a)Correlation on the dependent variable and (b)Variance Inflation Factor of each variables.

Correlation of the variables are clearing showing that almost all off variables don't seem to be correlated to the output variable. Also, they're also not correlated with one another also. Except MarkDown3 shows some correlation with MarkDown5. Second figure contains VIF values. VIF (Variance Inflation Factor) which might be commonly used for checking of the any variable whose variance is stricken by the other variable. Here as no covariance could even be obtained.

Thus, we couldn't extract any variable which might be use of for further processing. We could process with the STL algorithms. Which might be useful for the info which contains patterns of seasonal and trends.

.

V. SEASONAL TREND DECOMPOSITION USING LOESS

STL may be a versatile and robust method for decomposing statistic. STL is an acronym for "Seasonal and Trend decomposition using Loess", while Loess may be a method for estimating nonlinear relationships. The STL method was developed by Cleveland, Cleveland, McRae, & Terpenning (1990).

STL has several advantages over the classical, SEATS and X11 decomposition methods:

- Unlike SEATS and X11, STL will handle any type of seasonality, not only monthly and quarterly data.
- The seasonal component is allowed to change over time, and the rate of change can be controlled by the user.
- The smoothness of the trend-cycle can also be controlled by the user.
- It can be robust to outliers (i.e., the user can specify a robust decomposition), so that occasional unusual observations will not affect the estimates of the trend-cycle and seasonal components. They will, however, affect the remainder component.

A statistic decomposition are often accustomed measure the strength of trend and seasonality during a statistic (Wang, Smith, & Hyndman, 2006). Recall that the decomposition is written as:

$$Y_t=T_t+S_t+R_t$$

where T_t is the smoothed trend component, S_t is the seasonal component and R_t is a remainder component. For strongly trended data, the seasonally adjusted data should have far more variation than the rest component. Therefore

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

$Var(R_t)/Var(T_t + R_t)$ should be relatively small. But for data with little or no trend, the 2 variances should be approximately an equivalent. So, we define the strength of trend as:

$$F_t=\max(0,1-Var(R_t)/Var(T_t+R_t))$$

This will provide a measure of the strength of the trend between 0 and 1. Because the variance of the rest might occasionally be even larger than the variance of the seasonally adjusted data, we set the minimal possible value of F_t adequate to zero.

The strength of seasonality is defined similarly, but with reference to the detrended data instead of the seasonally adjusted data:

$$F_s=\max(0,1-Var(R_t)/Var(S_t+R_t))$$

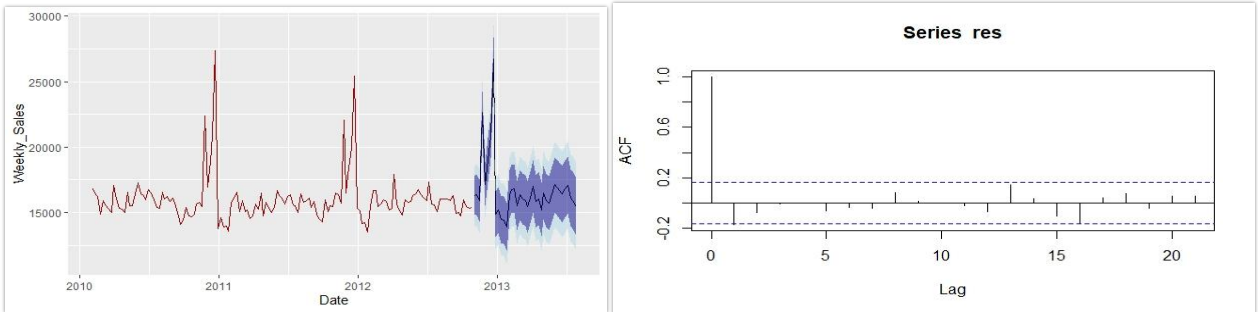
A series with seasonal strength F_s close to 0 exhibits almost no seasonality, while a series with strong seasonality will have Fs close to 1 because $Var(R_t)$ will be much smaller than $Var(S_t + R_t)$.

These measures are often useful, for instance, when there you've got an outsized collection of your time series, and you would like to seek out the series with the foremost trend or the most seasonality.

VI. FORECASTING THROUGH SEASONAL TREND DECOMPOSITION USING LOESS

To forecast a decomposed time series, we forecast the seasonal component S_t, and the seasonally adjusted component A_t, separately. It is usually assumed that the seasonal component is unchanging, or changing extremely slowly, so it is forecast by simply taking the last year of the estimated component. In other words, a seasonal naïve method is used for the seasonal component.

To forecast the seasonally adjusted component, any non-seasonal forecasting method may be used. For example, a random walk with drift model, or Holt's method, or a non-seasonal ARIMA model, may be used. Here, We are going to use ARIMA model. Which gives better accuracy then other models. The output of the Forecasted object obtained by combination of every iterative forecasting object run through each department of each stores.



(a) (b)

Fig. 2. (a)Forecasting object output graph and (b) Acf Residual plot.

Accuracy Measures: Here, Accuracy measure is taken by MAE (Mean Absolute Error).

MAE is calculated as :

$$T^{-1}\sum_{t=1}^{T}|y_t-\hat{y}_{t|t-1}|$$

The purpose of taking MAE over RMSE – We need to calculate the average accuracy of all the forecasting models iterated over each department of each store. Thus, information carried through MAE would not be changed by averaging them. Larger the MAE value lowers the accuracy of the model.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

The ACF Residual plot obtained of every algorithm suggest there any presence of information present in the residuals. Thus, checking the ACF plot with their given bounds we can have an intuition of information presence. In the above graphs we can have a clear vision that there would be no information presence in the residuals. Hence, we can get a clear that our model extracts whole information and it produces accurate results. Our MAE value produced is 760.665 which is quite better.

VII. CONCLUSION

The finish of our undertaking is to distinguish the effect on sales throughout various vital choices taken by the corporate. The examination is performed on authentic deals information across retail locations situated in various areas. Examination incorporates the impact of the markdown on deals , and furthermore the degree of impact on deals by fuel costs , temperature, joblessness , CPI and so on has been broke down utilizing basic and numerous straight relapse models Analytical apparatuses utilized in venture are RStudio. Predictive analytics is at the heart of supply chain process that helps Wal-Mart reduce overstock and stay properly stocked on the most in-demand products. Suppliers to retail company are required to use the real-time vendor inventory management system that helps them minimize the inventory for a particular product if there are no significant sales for it. This helps retailers to save funds to buy products that have greater demand and have increased probability for greater profits.

REFERENCES

[1] Forecasting: Principles and Practice second edition by Rob J Hyndman and George Athanasopoulos, University of Western Australia.
[2] STL : A Seasonal - Trend Decomposition procedure based on Loess by Robert B. Cleveland, William S. Cleveland, Jean E. Mcrae, Irma Terpenning.
[3] Forecast v8.11 package documentation by Rob J Hyndman, University of Western Australia.