**Project Requirement:-**

|  |
| --- |
| **Context** |
| The data is for retail client where customers buy products at a price. The delivery of ordered products may not be on same day. |
| The client wants to plan better inventory so as to avoid stock build-up and at the same time ensure he meets demands. |
| The client thinks, he can plan inventory if he knows what the sale would be in future |
| As a Data Scientist deployed in client location, you are provided with this historical data. How would you go about consulting? |

**Author:** Arijit Pal (aryanarijit@gmail.com),

Abhishek Biswas (abby33459@gmail.com)

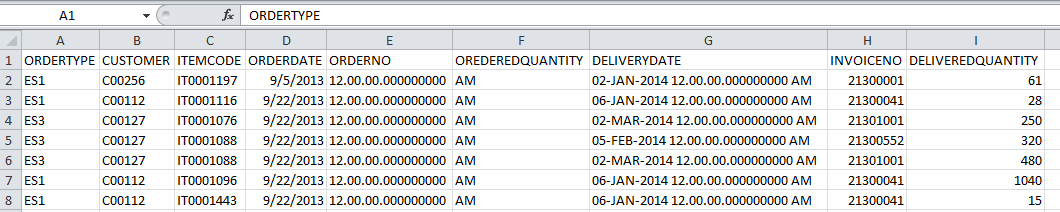
**Introduction :-**

This project is focused on the top 10 items which is an output of Market Basket Analysis using Association rule.

We are doing the coding the time series forecasting for one item and this can be applicable for any item.

**Data set : Totalcsv**.csv

A glimpse of raw data set

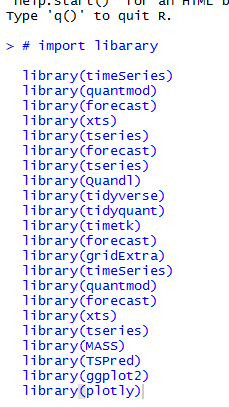


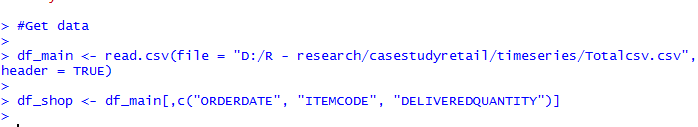
**Basic Analysis and data preparation:-**

We need to understand the data before start the analysis and we need to consider only two column (Item code and Deliverydate).

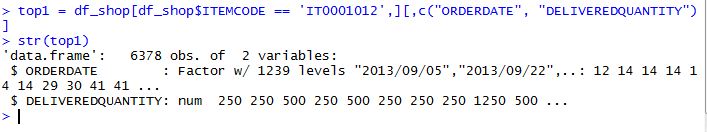
We need to take an extract of data of only two column and need to get aggregation of each data for each month and year.

Before start of the code, we need to import the required library.



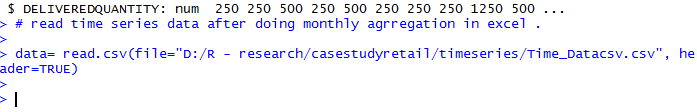


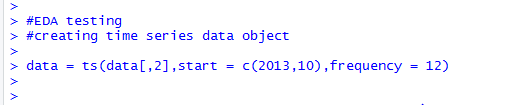
# after performing apriori , we have listed top 10 products. here we are analyzing the time series for item no -IT0001012.



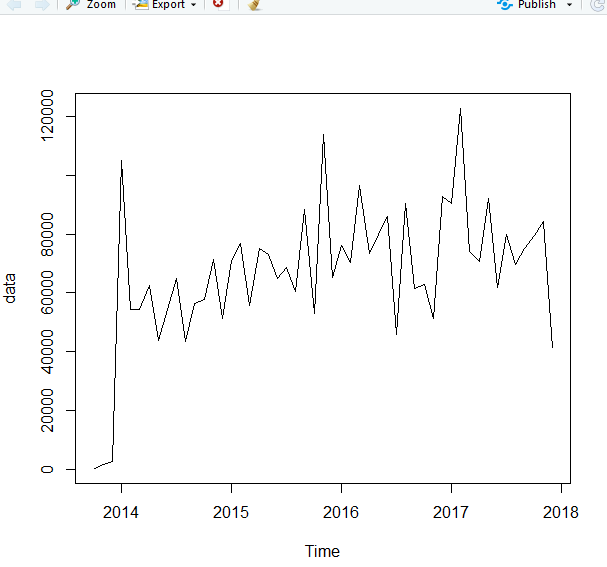
Export the data into local machine for perform the aggregation .

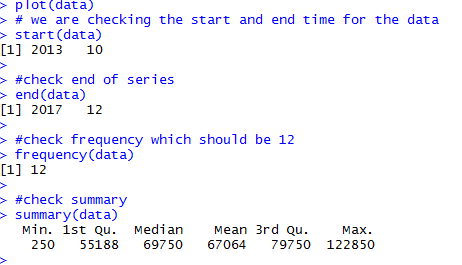
write.csv(top1,"top1.csv", quote = FALSE, row.names = TRUE)

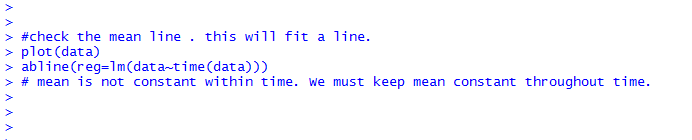


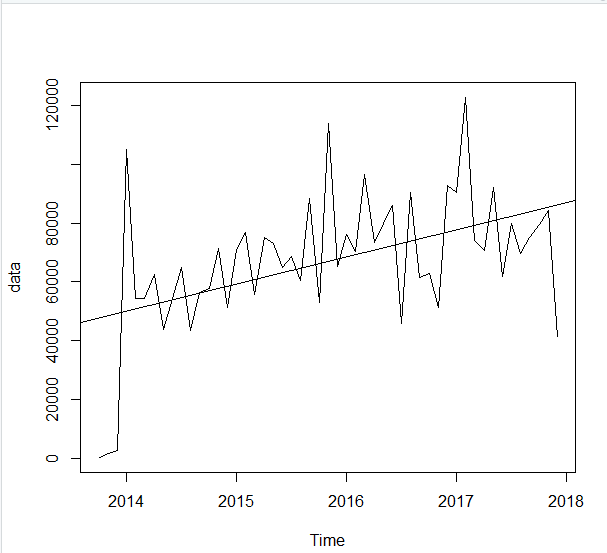


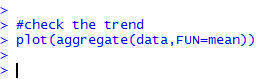


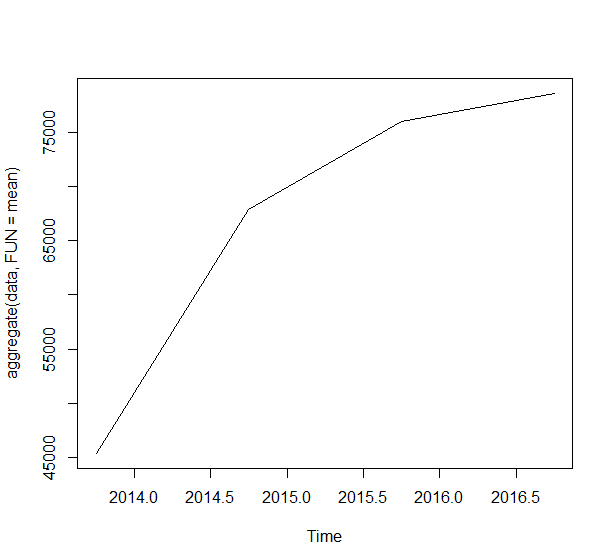


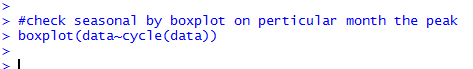


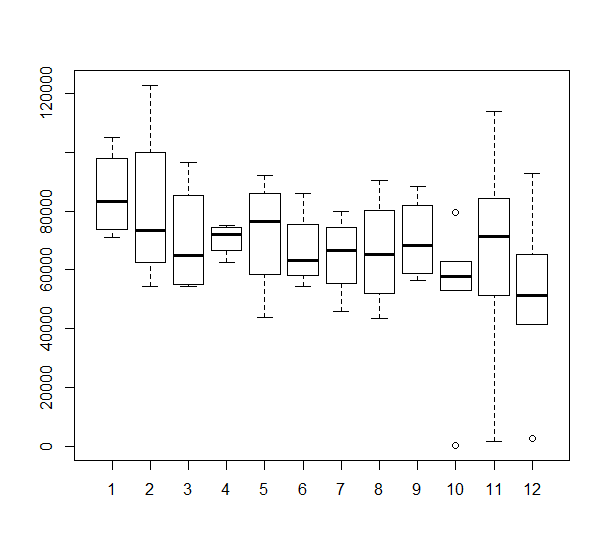










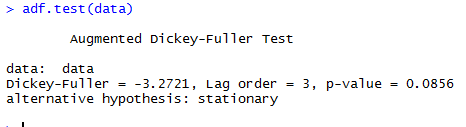


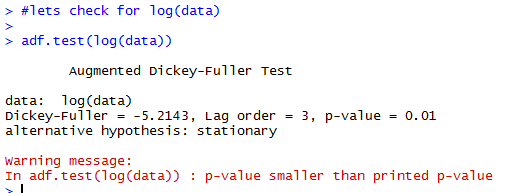
We can see that the seasonal trend is high on January.

#checking for stationarity.

#this is Augmented Dickey fuller test for stationarity. The null hypothesis states that large p value indicate non stationarity.

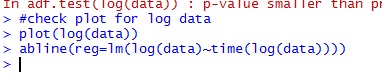
#and smaller p value indicate stationarity ( thresold :0.05)

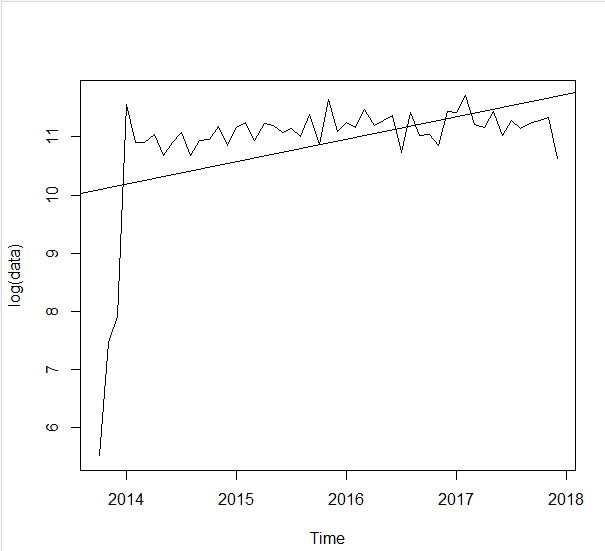


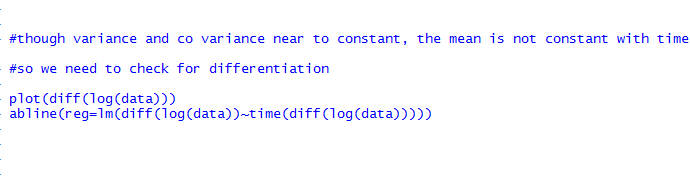


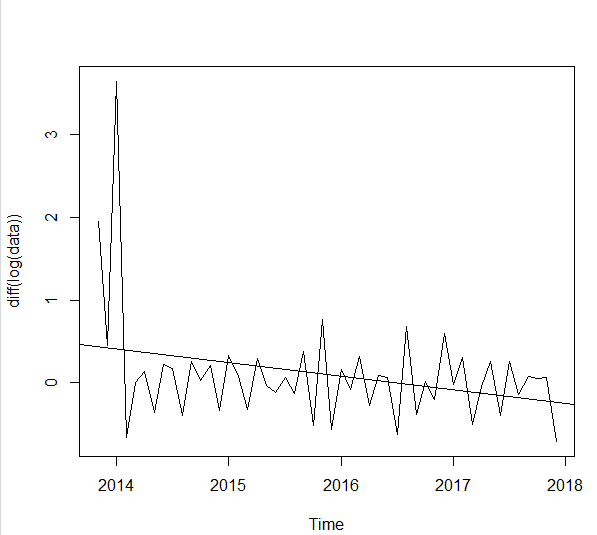
#p-value =0.01

#now data is stationary



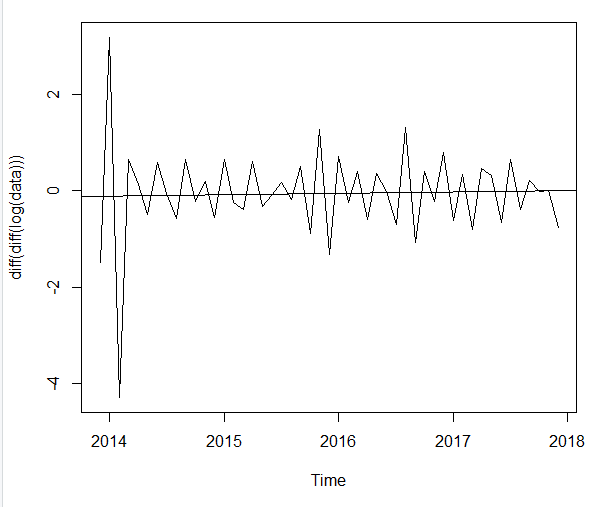


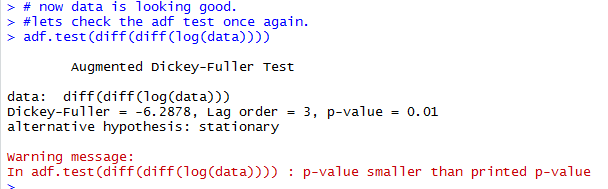


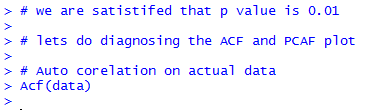


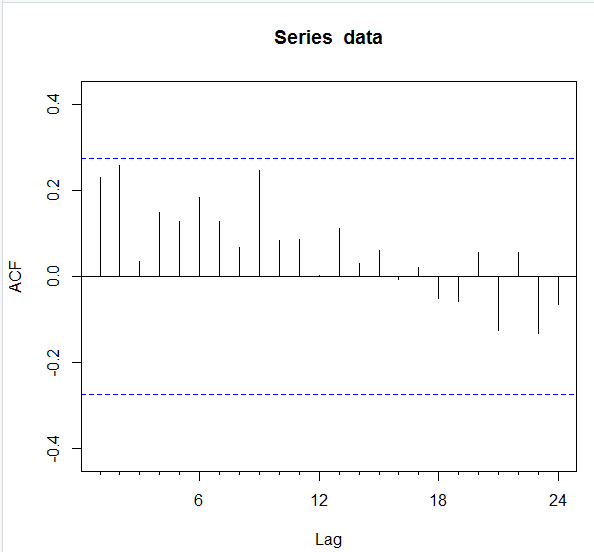
Let’s do 2nd deferentiation :-

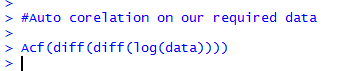


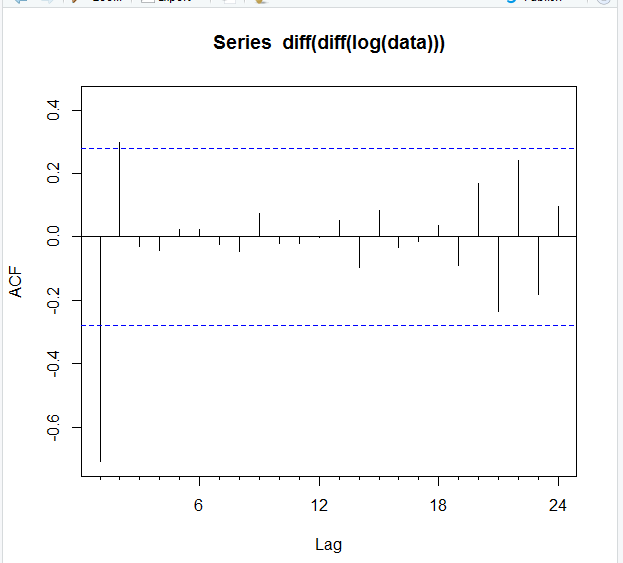












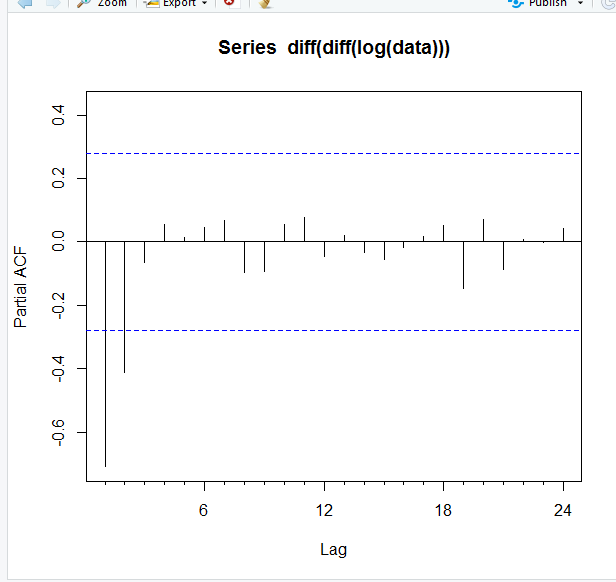
# The first line is way far out of blue region. So we are ignoring.

Lets take the next line. And the immediate line which got inverted is line no : 2 . So we have to take that value for q (MA) as 1 .

# we already decided to take value of d is 2

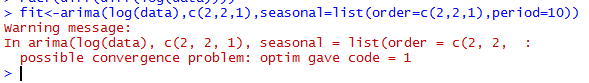
# lets check the value for p by Pacf





# The immediate line which got inverted is line no : 3 . So we have to take that value for p (AR) as 2

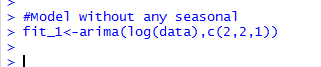
**Model Building Process:-**

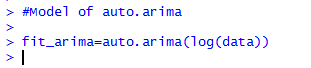


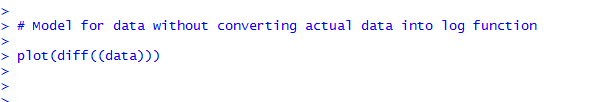
# ignore the warning

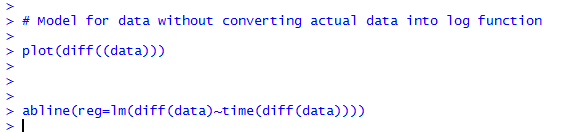
# for less period

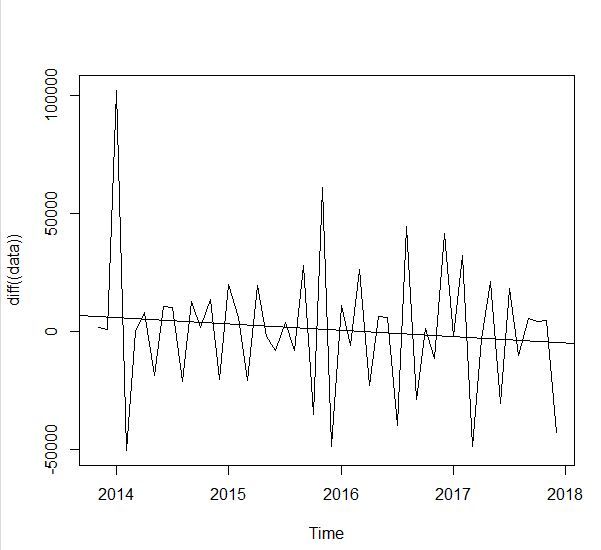




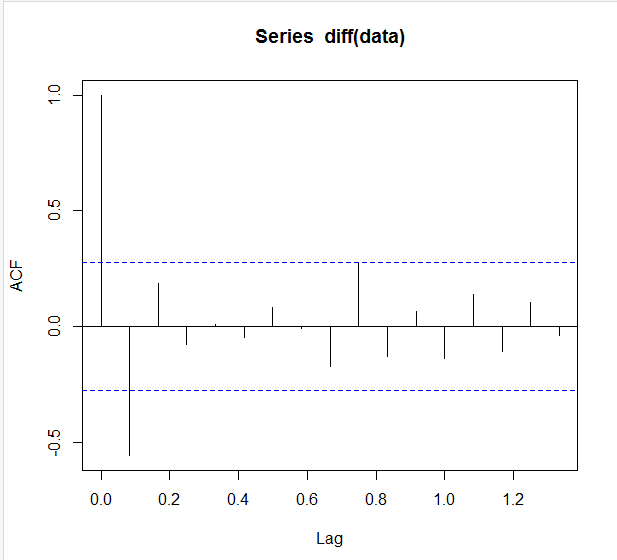






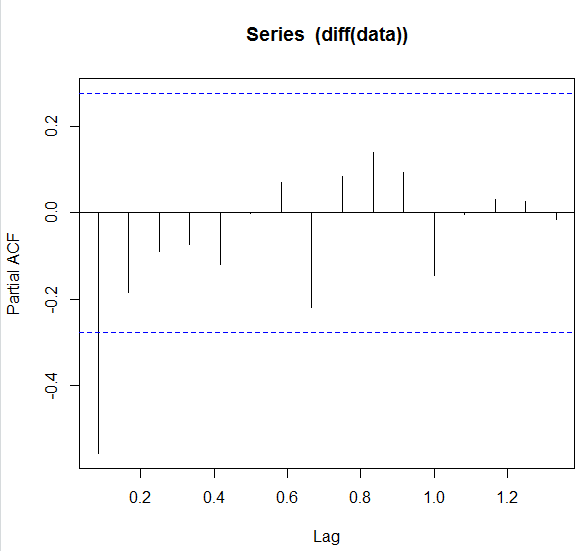






# q is 0

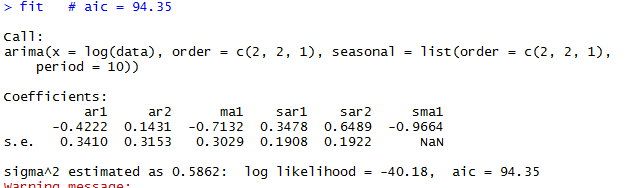


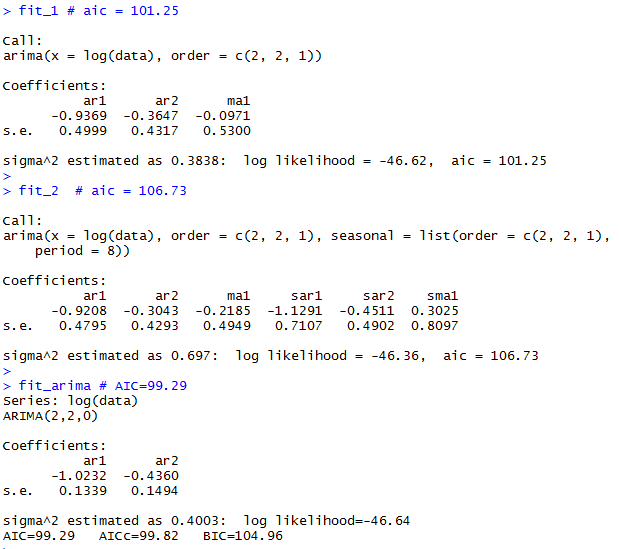


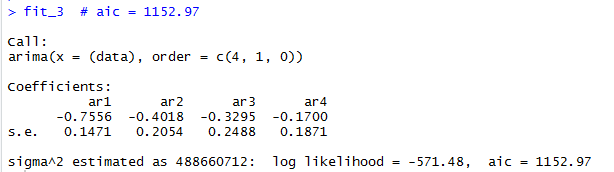
# p is 4



# let’s check the AIC value for 3 models

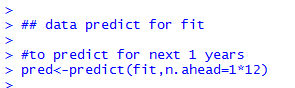


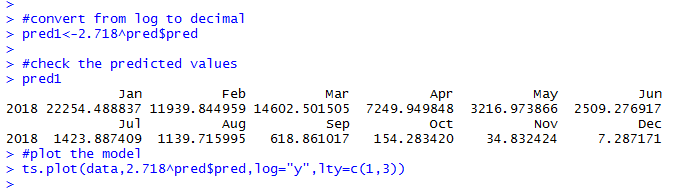


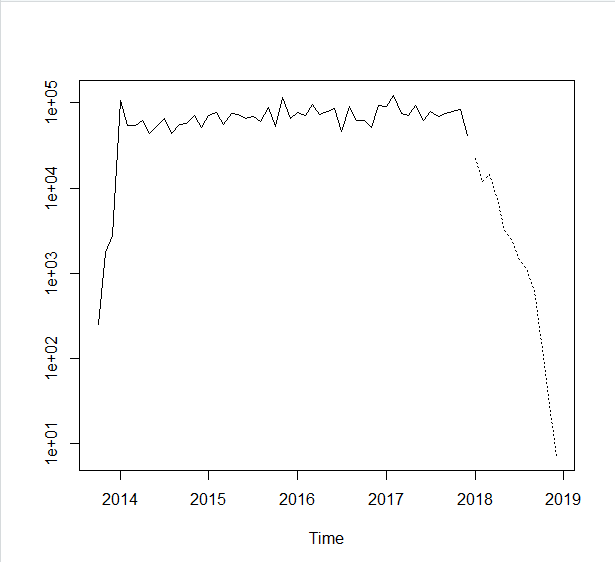


## if AIC is less the model seems to good. Let’s check on predicted data and we shall decide the best model based on predicted data and accuracy.

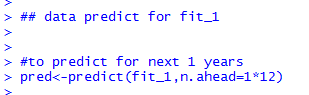
**### Forecast and side by side test**

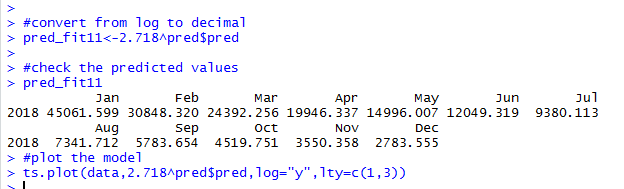


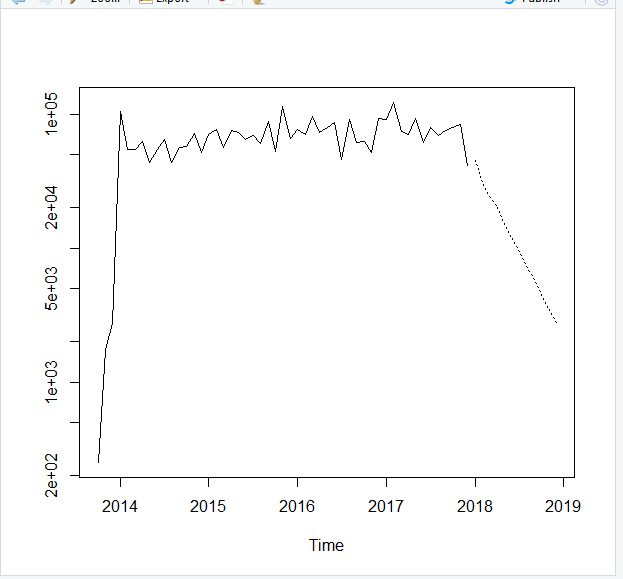




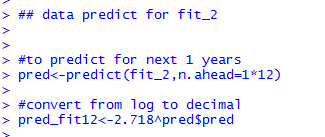
## data is drastically going down which does not look good at all.

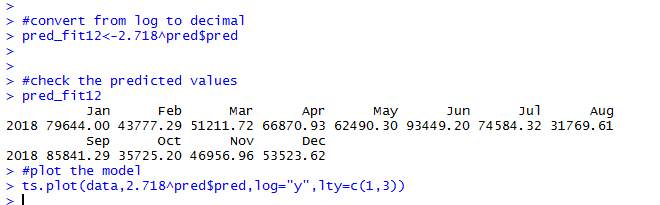


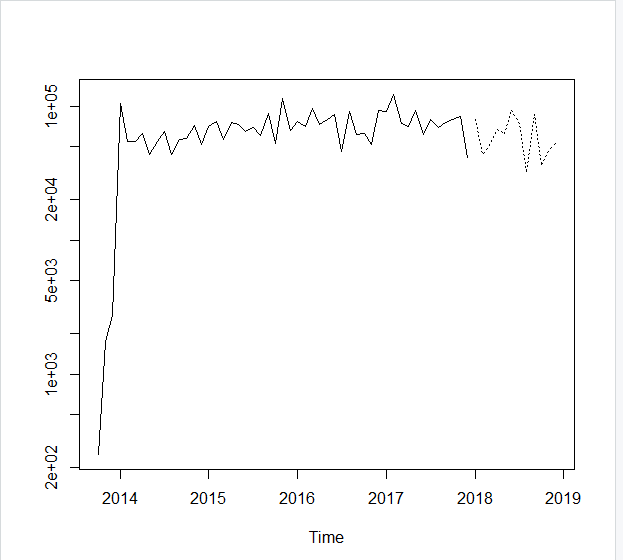




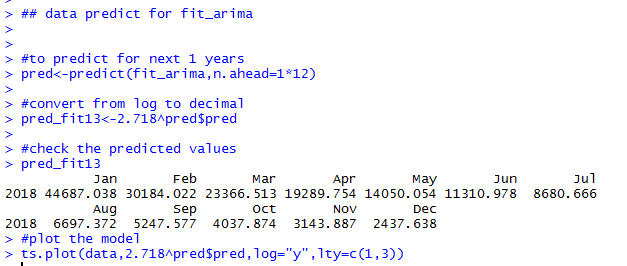
## data is drastically going down which does not look good at all.

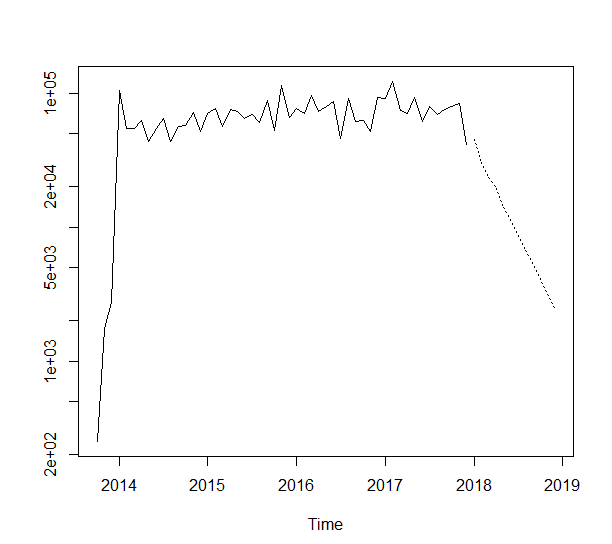




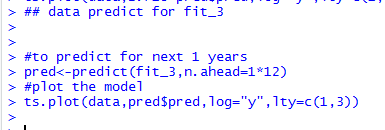


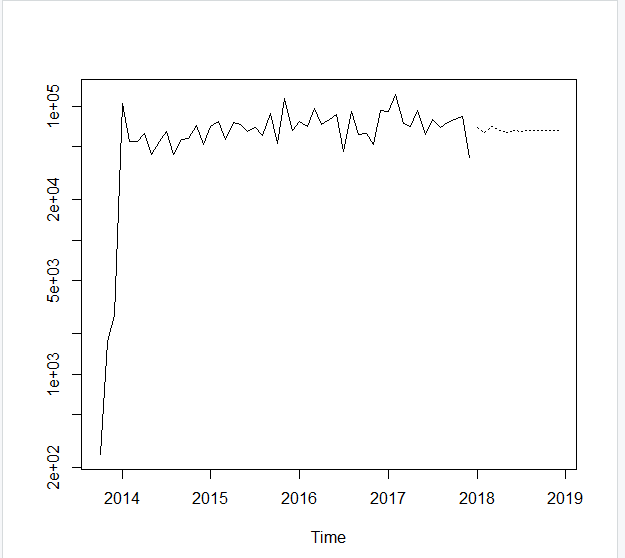
## data prediction is zigzag and maintain the consistency . We can consider fit\_2 for prediction.





## data is drastically going down which does not look good at all.

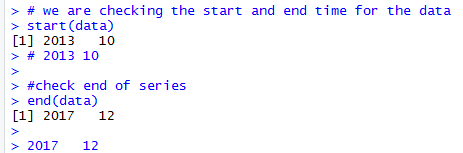




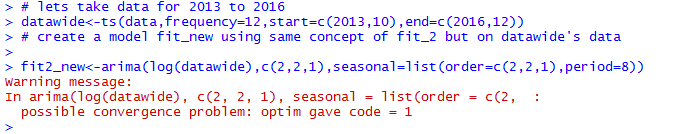
## data is near to straight line but not going down drastically. So we can consider it.

**Model Comparision**

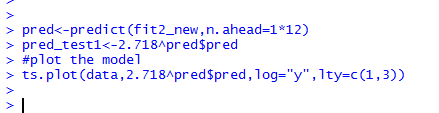
##### So finally we shall take fit\_2 and fit\_3 for prediction.

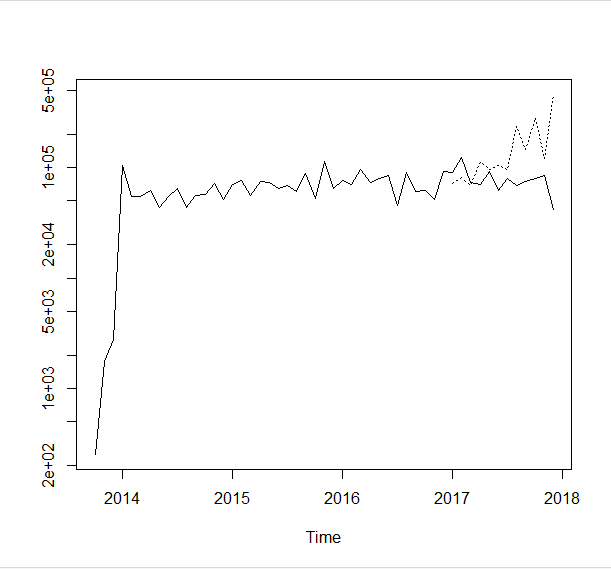






## ignore the warning





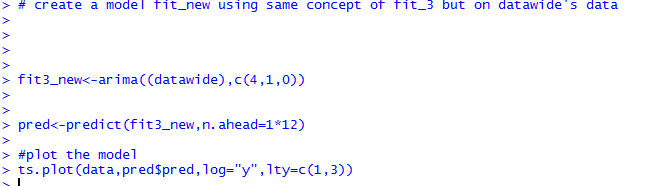
# data diverted from actual data very highly in graph. Let’s check accuracy

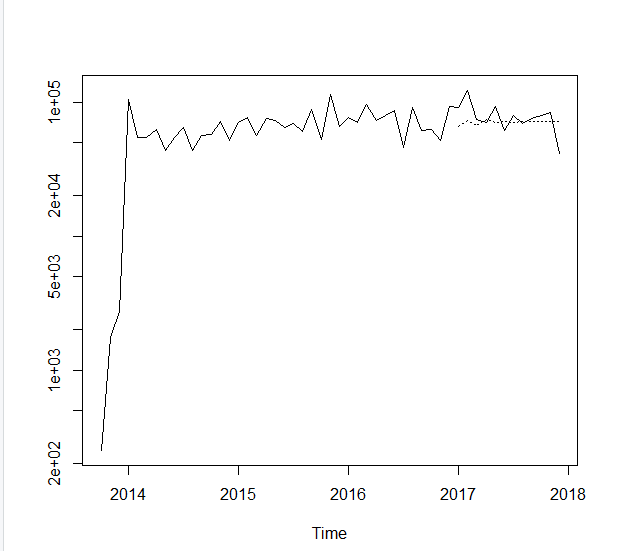
> round(accuracy(pred$pred,data),3)

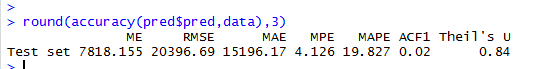
ME RMSE MAE MPE MAPE ACF1 Theil's U

Test set 78519.05 80691.71 78519.05 99.984 99.984 -0.045 3.63

# create a model fit\_new using same concept of fit\_3 but on data wide's data







# we shall consider lowest MAE and MAPE as best model

So **we shall consider fit\_3 as best model**