DMH DATA INSIGHTS - TRACK 1 (CLIENT ONLY DATA) IDEATION CHALLENGE

With dataset provided in the csv format , few columns were removed these columns were (Macaddr,Ipaddr,Block,Serialnbr,ROOMGROUP). Mac address and Ip address is not considered however mail address was considered as the unique ID. Room name was considered instead of ROOMGROUP as Roomname was more specific. On considering geography , city , country code and continent was considered.

Using the data available, tableau was used to visualize where the sales of the items was dominant in the world. This helped to get insights about which item was in demand in which region. Using the drop down one can see which item is in demand in which region(top 5 items based on sales have been showed in tableau story published on tableau public).

Next, the promocodes were many used in United States , so only items sold using promocode in United states were considered and visualized.

Below is the link for tableau public where one can view the insights drawn:

https://public.tableau.com/shared/X22SHXYTP?:display_count=yes

R was used as a statistical tool , to forecast the future sales.

Firstly the data was converted into time series and decomposed. It was checked for any stationarity (repetitions that don not allow to forecast properly , so any stationarity has to be removed).

Have added the source code in the github and can be accesses easily through the below link:

https://github.com/gdn7/PyR/blob/master/README.md

The code is built for reusability ,

If similar data is added with time , there is no need to make any major changes except in line:

sales=ts(DMH$mailaddr, start = c(2014,7), end=c(2016,9),frequency = 12)

where **start =c()** and **end=c()** are start and end time period that is considered. i.e in this case 7/2014 to 9/2016 time duration is considered. **Frequency=12** indicated that all 12 months are considered. Can be changed to any desired time based on requirement.

A time series ARIMA model was built to forecast sales. A auto ARIMA function in R was considered and AIC and BIC criteria was considered for model evaluation.

forecasting using R yielded the following results:

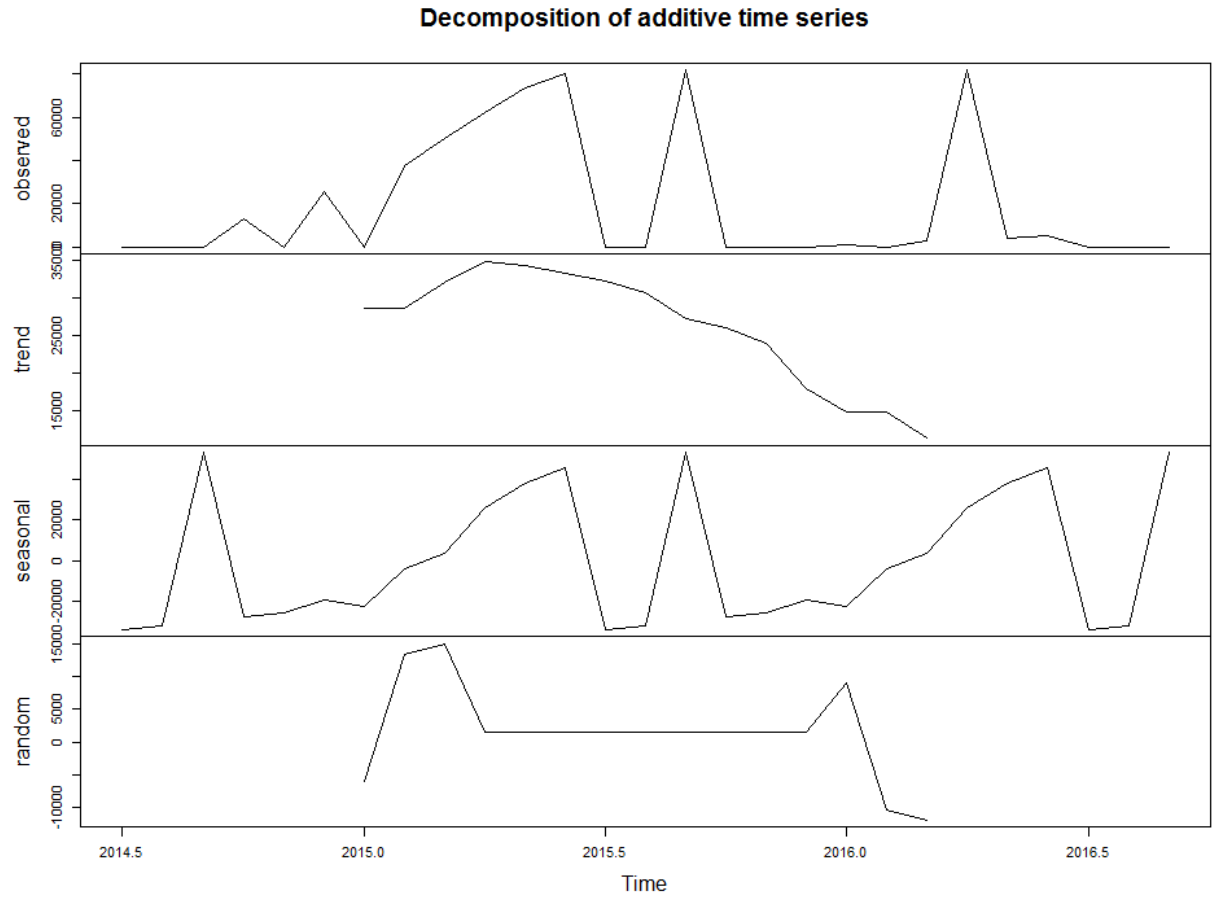

Figure1.

From figure1. It can be seen that sales peaks up at certain month during an year(usually in June/July , March/April). There is a downward trend and needs to looked at very carefully.

Figure2.

From Figure2. It is to be seen that in 2017 the will be decrease in the sales based on the fitted model.

Conclusions:

1.Most of the items are purchased are in North America and in Europe.

2.As a marketing strategy if promocode is used , it is mainly is used in United States than by any other country.

3.Kansas in one state in US where there are lot of demand for items.

4. East and west coast OF North America is prime place for sales of items.

5.Netherland , Denmark , Sweden and Great Britain are the countries in Europe where the sales are more.

6.Eastern Australia and Netherland have also great demand for items sold.

7.Items , HEOS1BK ,HEOS3 ,HEOS5 ,HEOSAMP and HEOSLINK these are the top 5 items sold.