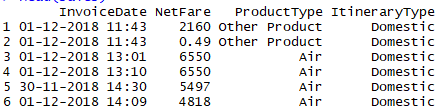
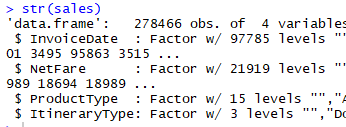
Forecasting Sales of a Travel Agency

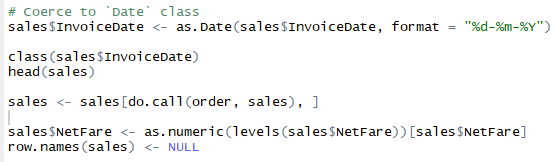
1. Exploratory Data Analysis
2. Manipulating Given Data



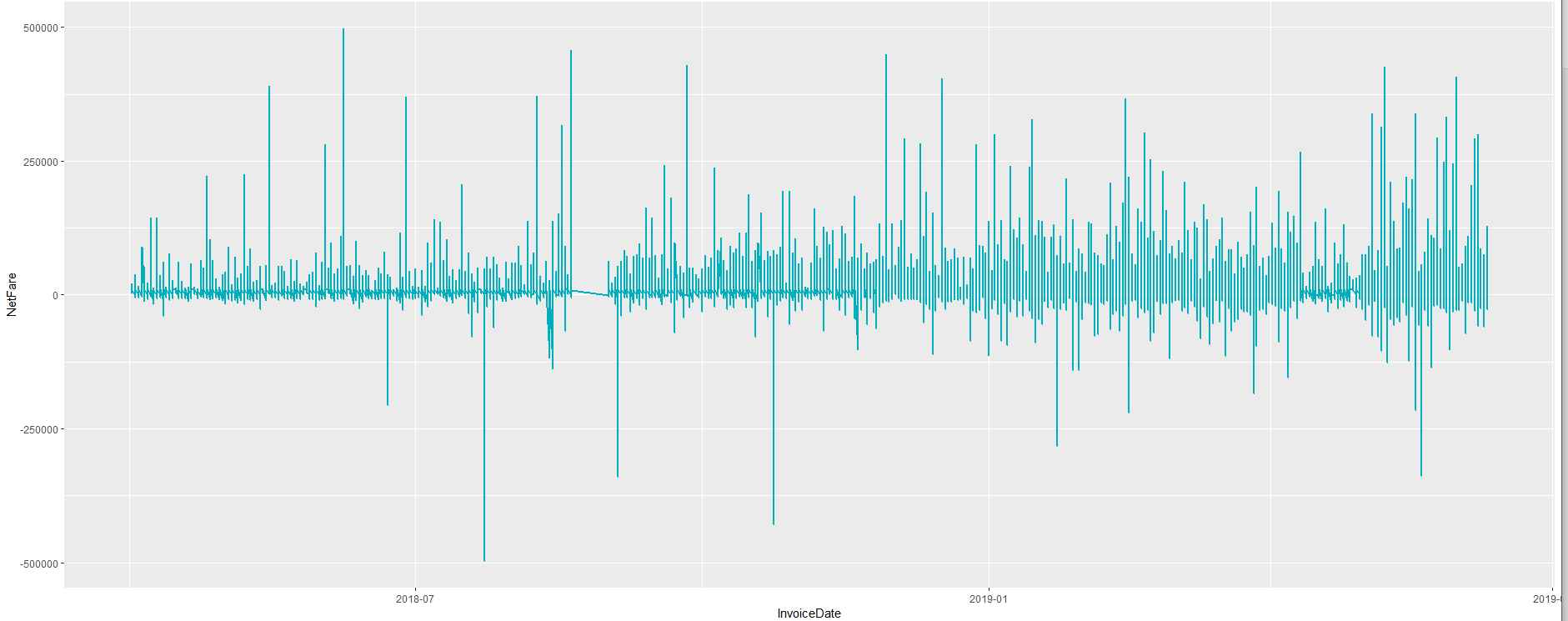
By looking at the data we can say the data is not sequenced and it has a time series with two categories and a value. Let’s look at the class of the data.



The InvoiceDate should be Date class and the Netfare should be Numeric class and the data ordered.



Plotting the above manipulated data, we get



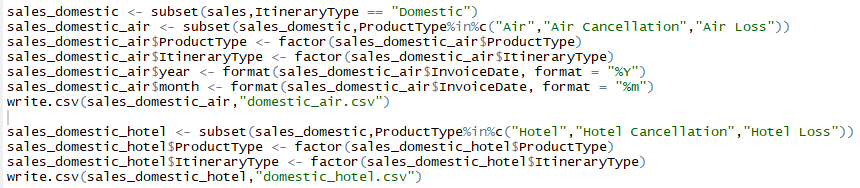
By looking at the graph we can infer that the data is showing stationary which is because it has some negative values which infer us the cancelled tickets or bookings by the result of which we can’t see the actual trend, seasonality of the time series also we can’t even decompose and recompose the time series.

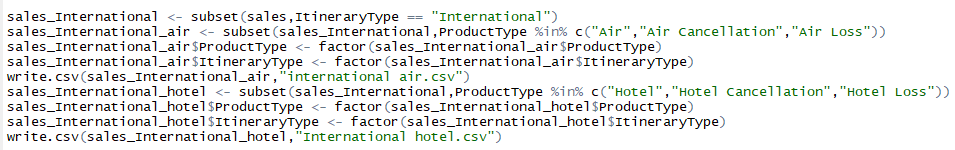
Another reason is that the data set is very huge which is like 2.7 lakh transactions of about one and a half year and also data has lot of NA values due to the dataset not provided the values which costumer paid or cancelled online via credit or debit card so we have simply remove those data with data cleansing operations because imputing nearly 60,000 values doesn’t makes sense as we already nearly 2 lakh values to work with to build our model.

1. Segregation of data for better accuracy and model building

As we know data has two categories “ProductType” and “ IterationType” which further has subcategories, so we are segregating the data first using the subcategories of IterationType which are Domestic and international and then two main subcategories of ProductType which are Air and Hotel in which we include three major subcategories which has values under Air and Hotel.

And save the results of segregation as csv files for further use for modelling.

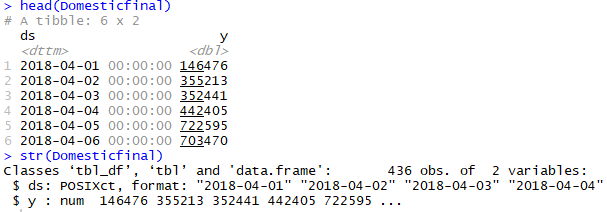




After segmenting the data, we have to convert each data in excel using sum function for getting daily and weekly data to nullify the effect of negative values influencing our analysis.

After segregation we have the data like

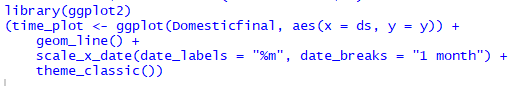
1. Domestic Air(daily)
2. Domestic Hotel(weekly)
3. International Air (weekly)
4. International Hotel(weekly)
5. Building Model for Forecasting based on Data Structure Outcome
6. Domestic Air
7. Exploratory Data Analysis
8. Data Operations

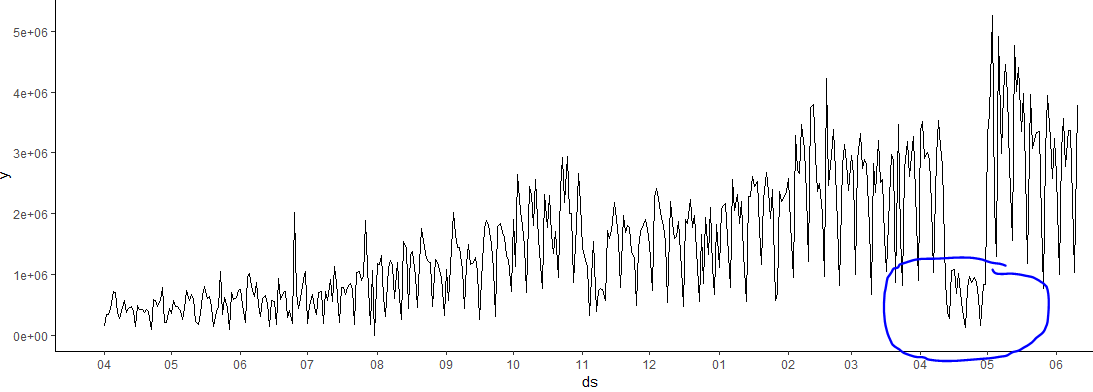


If we look at the structure of a data it is in POSIXct format, but we want it in date format



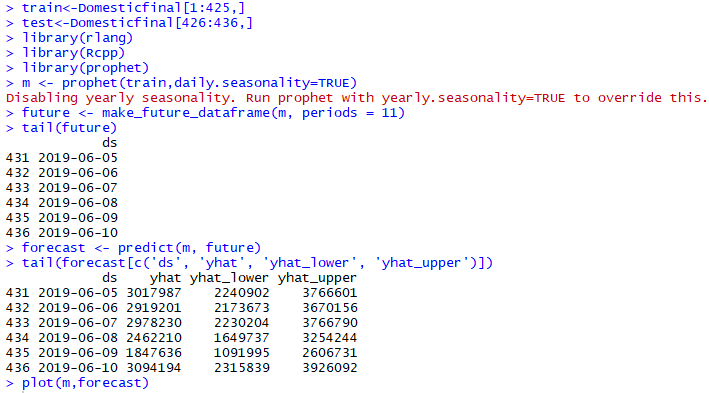
1. Plotting data in time series format

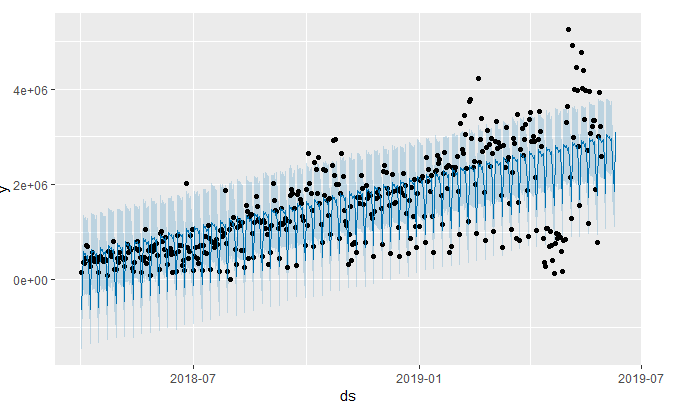


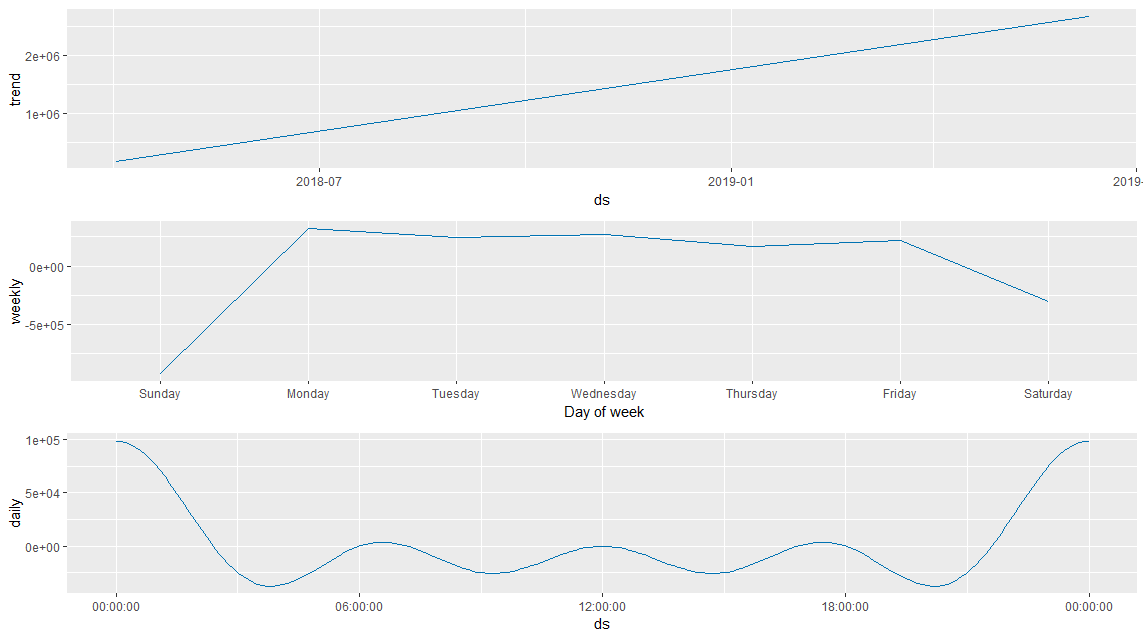


By looking at the time plot we can say that there is a fire drill in data which may affect our forecasting result.

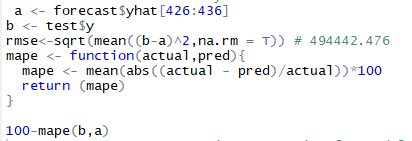
1. Building the Model with Fire Trail with Facebook’s Prophet Library







1. Calculating RSME and Accuracy

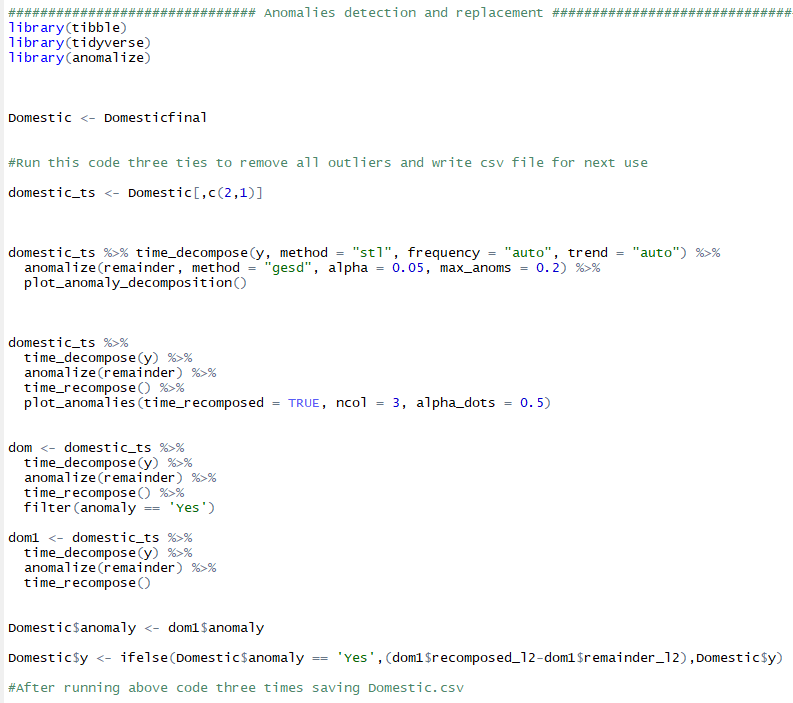


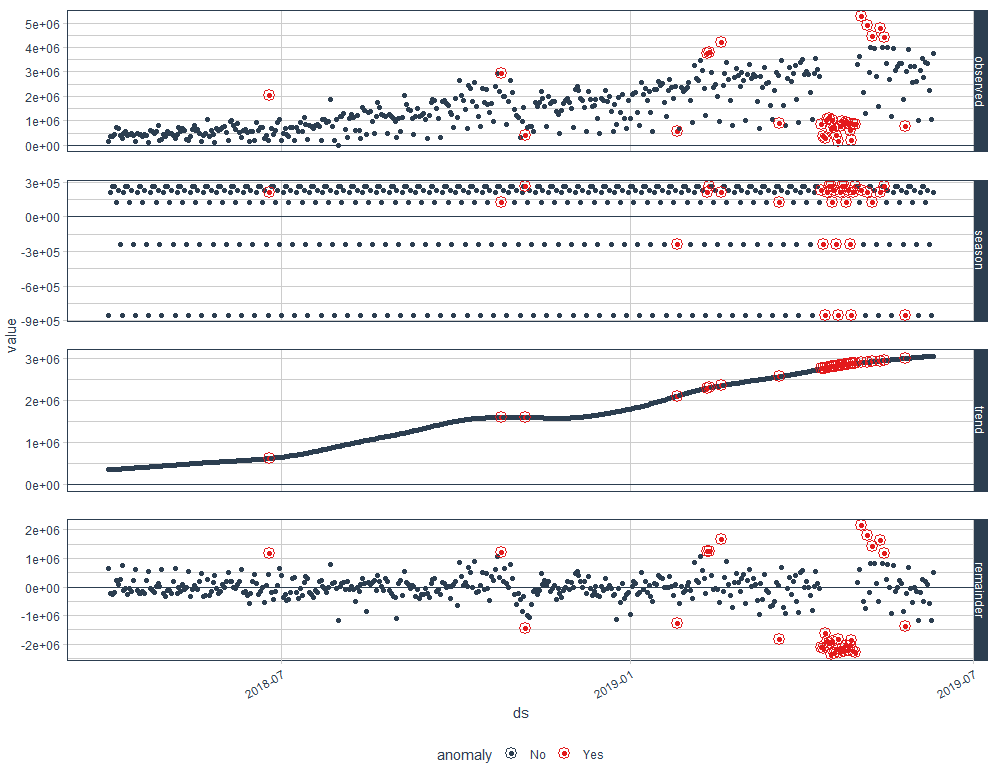
RSME = 494442.476 which is nearly equal to 4.9 lakh which is nearly 1/3 of the daily revenue generated.

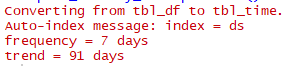
Accuracy = 76.75 %

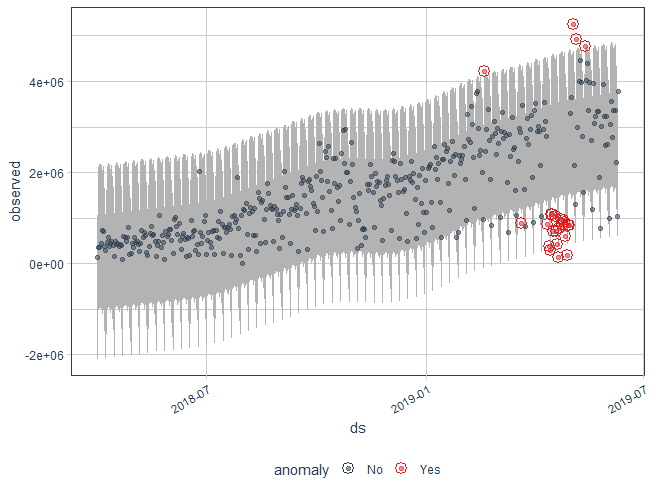
1. Eliminating the fire trail and replacing it with suitable imputation

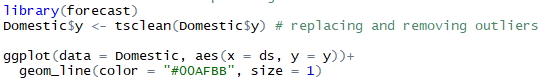
For detecting and replacing anomalies in data we use anomalize library developed by Business Science.

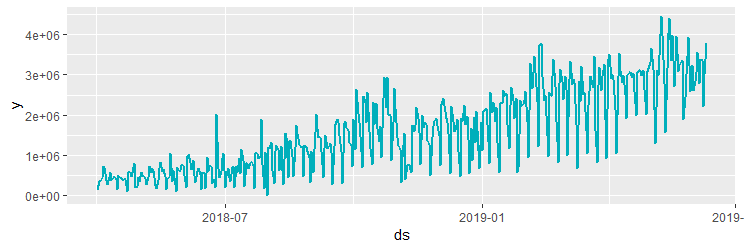






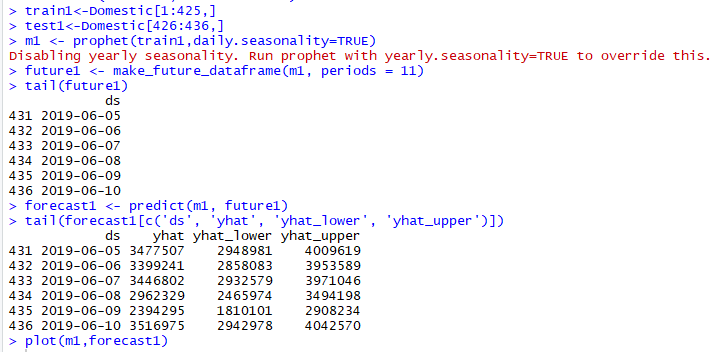


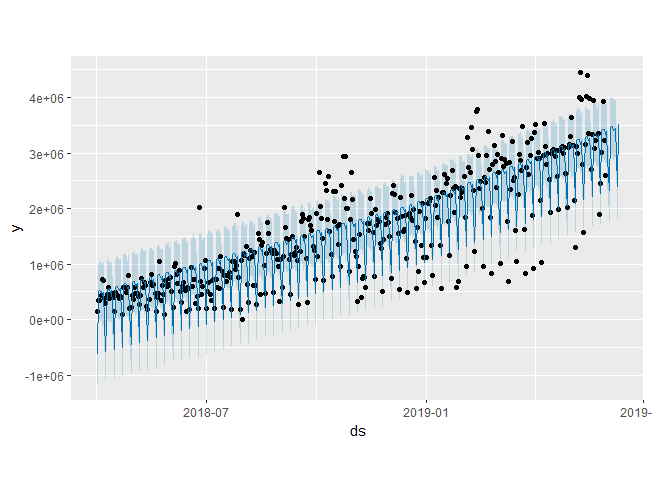


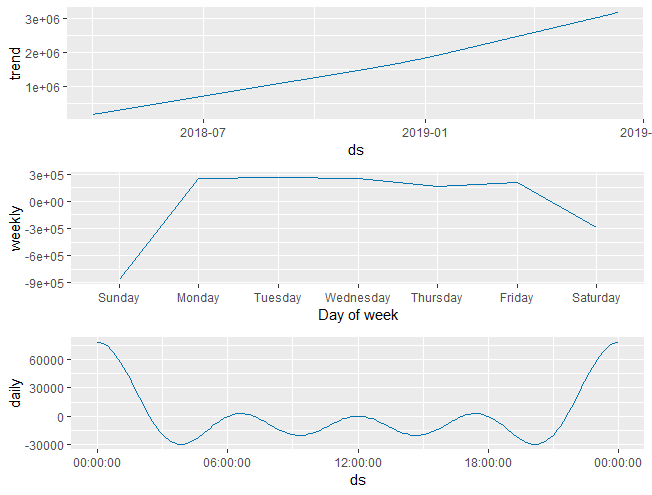


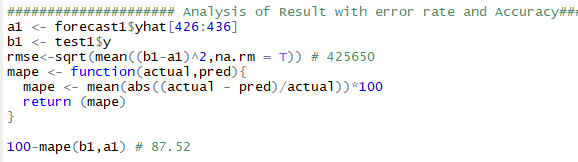
As we can see from the above graph that the fire trail has gone.

1. Building the Model with new data frame









As we see from the above the

RMSE = 425650 which is nearly 4.2 lakh far less than daily revenue generated.

Accuracy = 87.52% which increase by almost 11%

Following the same above procedure for Domestic Hotel , International Air and International Hotel we get the following results

1. Domestic Hotel

RMSE = 153436.729

Accuracy = 90.56%

1. International Air

RMSE = 846267.679

Accuracy = 88.779 %

1. International Hotel

RMSE = 18526.55 (Using auto arima) , 68799(prophet)

Accuracy = 78.43%(Using auto arima), 79%(prophet)