**Business Objective: -** To build a model to predict the price fluctuations over time.

**Datasets: -** The dataset that is received from client have the following details:

Variables are –

1.) Invoice Date: - This provide time stamp information at which a particular product is sold

The data sets contain data from April, 2018 to June, 2019

2.) NetFare :- This provide information about the price of the product is sold out

3. ProductType: - This provide information about the type of product being sold.

The following are the Type of products:

1.) Air

2.) Air Cancellation

3.) Air Debit Note

4.) Air Loss

5.) Charge

6.) Hotel

7.) Hotel Cancellation

8.) Hotel Debit Note

9.) Hotel Loss

10.) Other Product

11.) Other Product Cancellation

12.) Other Product Debit Note

13.) Payment

14.) Refund

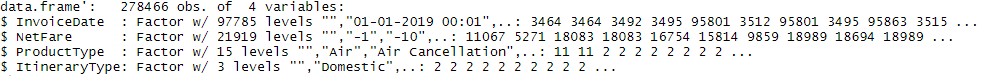
4. Itinernary Type:- Domestic or International

**Dependent Variable (Y ):- NetFare**

**Independent Variable :- Airfare, InvoiceDate**

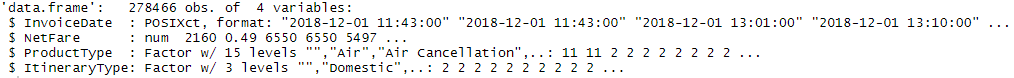
**Exploratory Data Analysis:-**

Structure of dataset

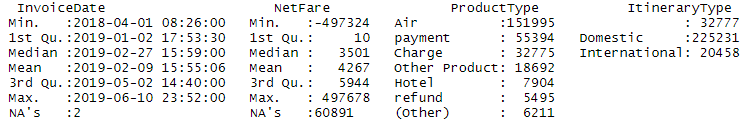


The InvoiceDate and NetFare should be in Date and Numeric format respectively.

After conversion into proper format



Summary



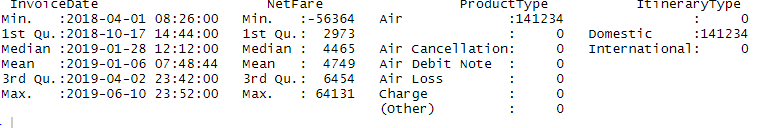
After arranging the data in ascending order by InvoiceDate,

**Analysis of DOMESTIC AIRLINES**

Filtering out domestic data from airlines and selecting

**Statistical Analysis of the variables**

Data Summary:-

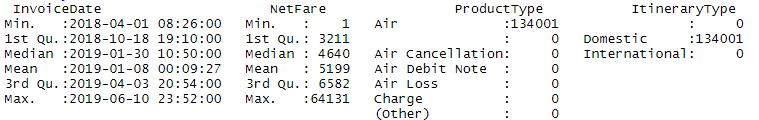


There is no NA values

From the above summary, we get to know that there are some –ve and 0 values for net fare for Air and based on domain knowledge we can say that there can’t be any negative value or zero for NetFare

So arranging the data in ascending order and excluding the data which contain negative values.

Summary

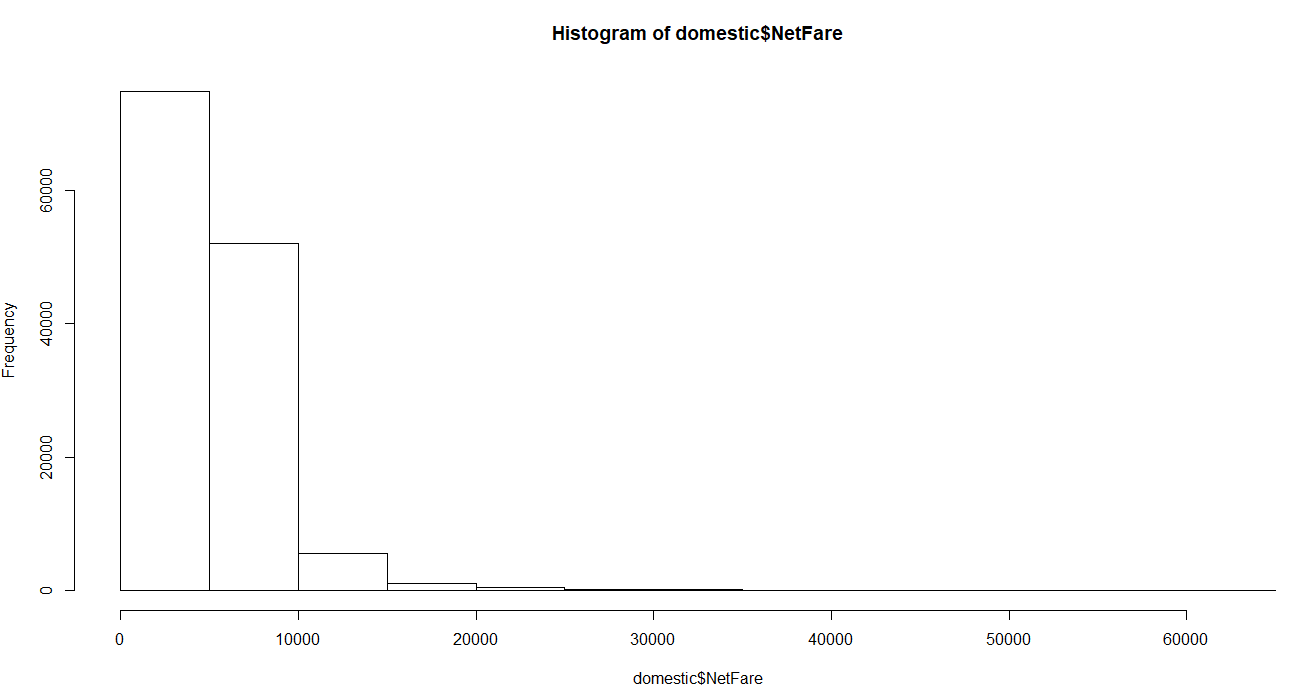


**i.) NetFare**

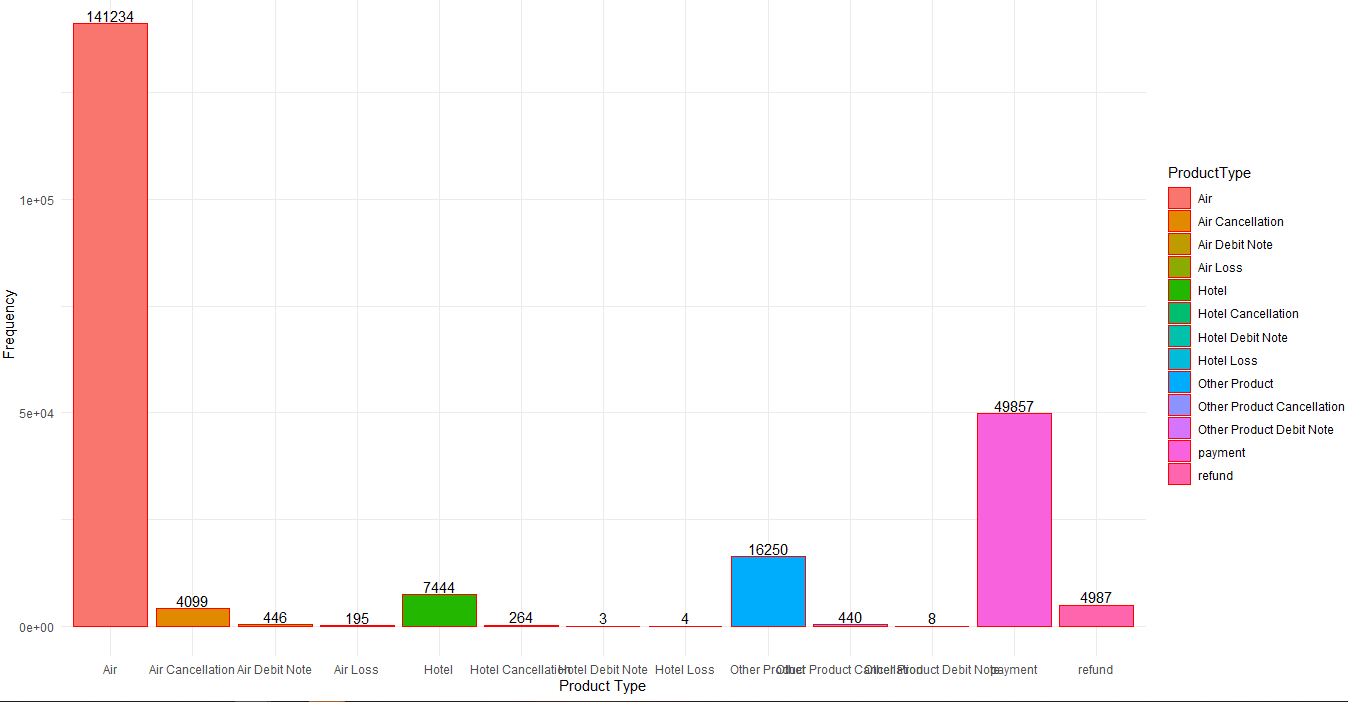
Statistical Summary

|  |  |
| --- | --- |
| Mean | 5199 |
| Median | 4640 |
| Variance | 8996601 |
| StdDev | 2999.433 |

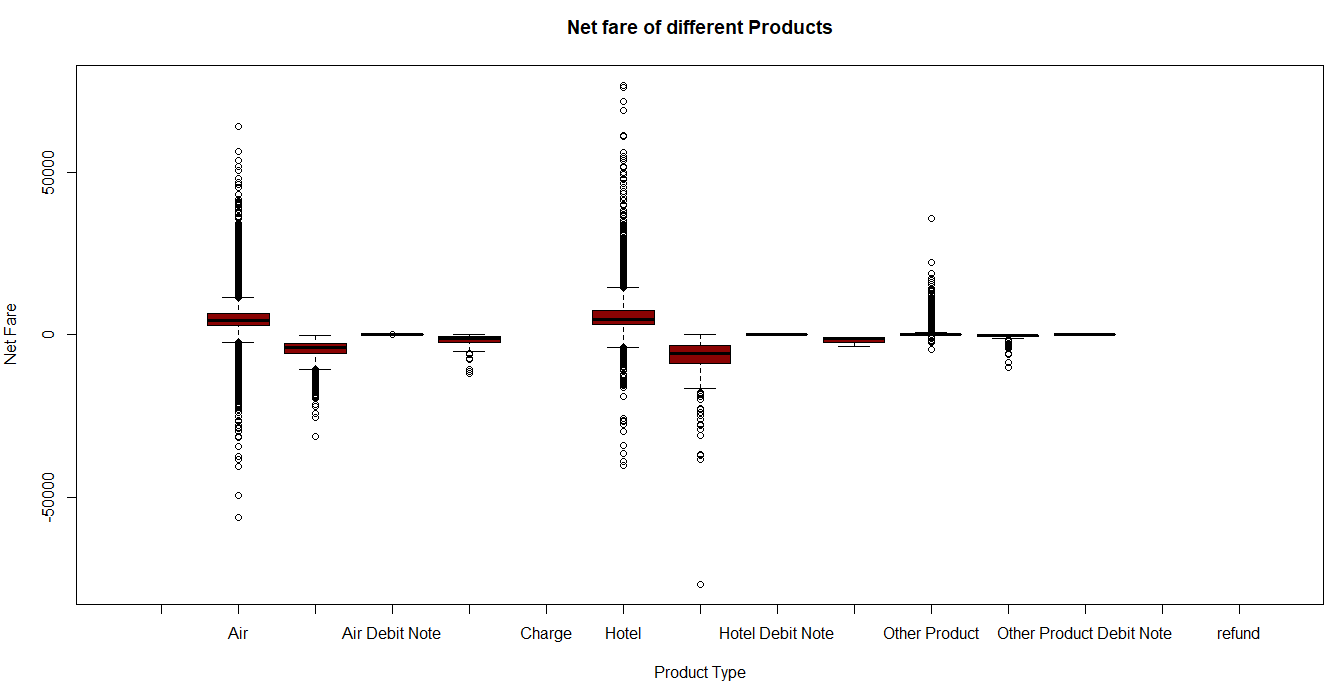
Graphical representation



**ii.) Product Type**

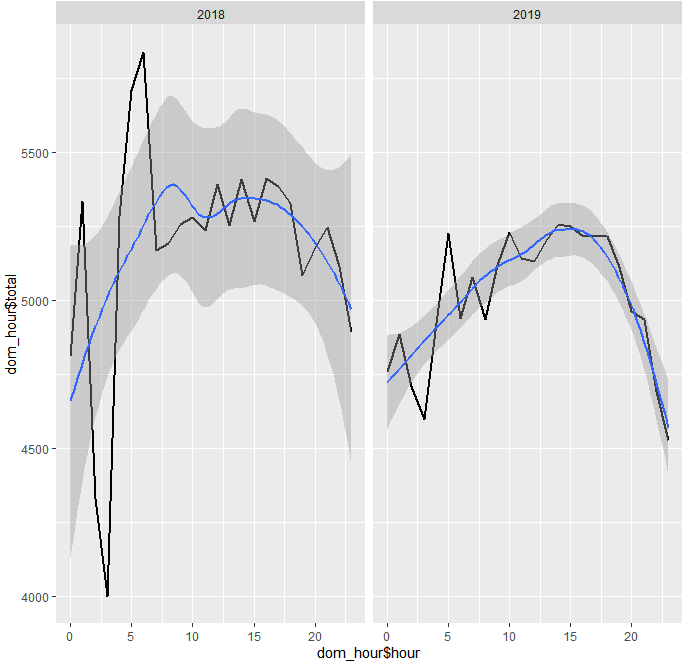


Seperate the InvoiceDate into different columns namely, year, month, day, wday, hours, minute.



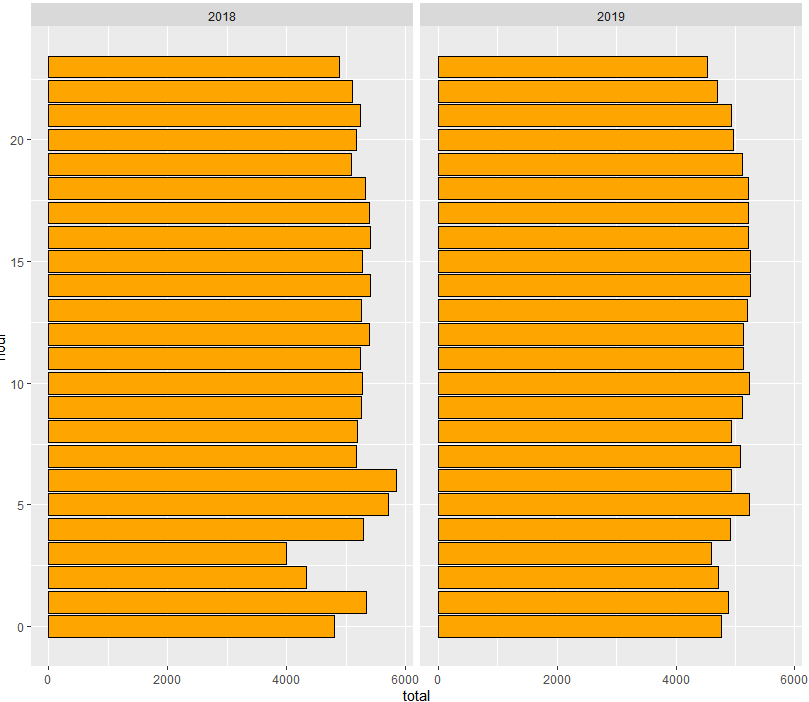
Trend Analysis For Net fare of AIR

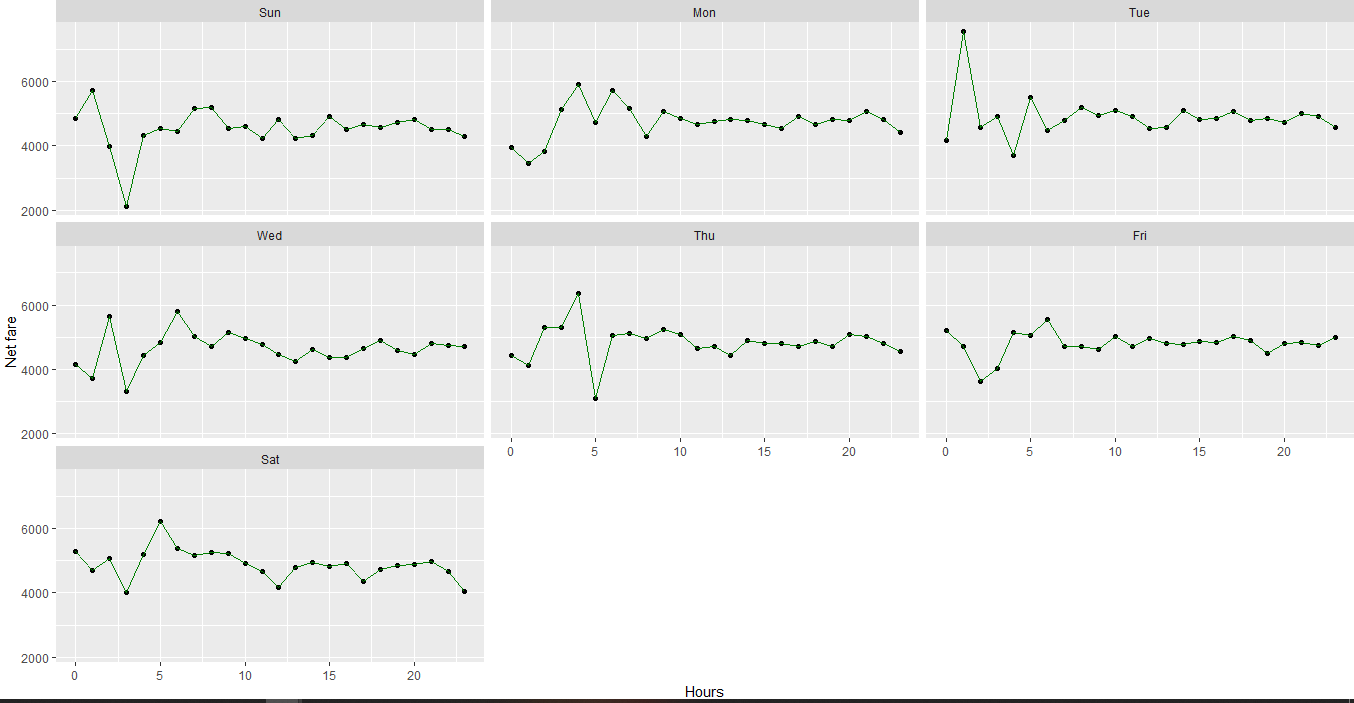
1. Average NetFare for different hours of day



The above graph depicts about the average Netfare hours of day during different for the year 2018 and 2019 .The black line indicates the fluctuations in Netfare during different hours of day and blue line indicates the smoothing average line over the period of 24 hrs. It can be said that, Netfare value picks during morning hours and falls to minimum after midnight.

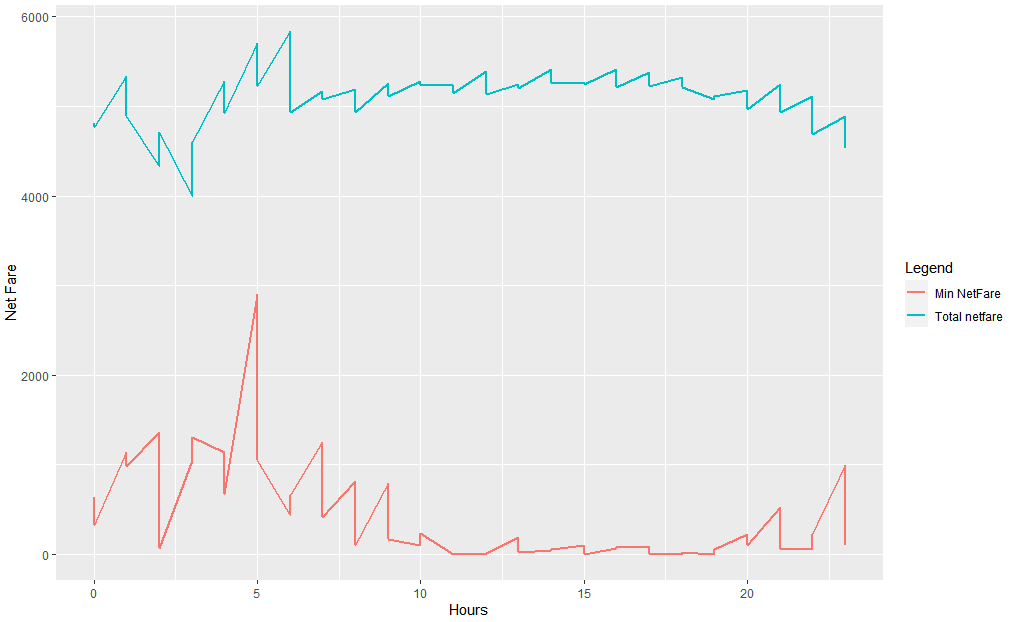
The same analysis can be interpreted from the barplot below.



Hour Wise trend for each day of the week

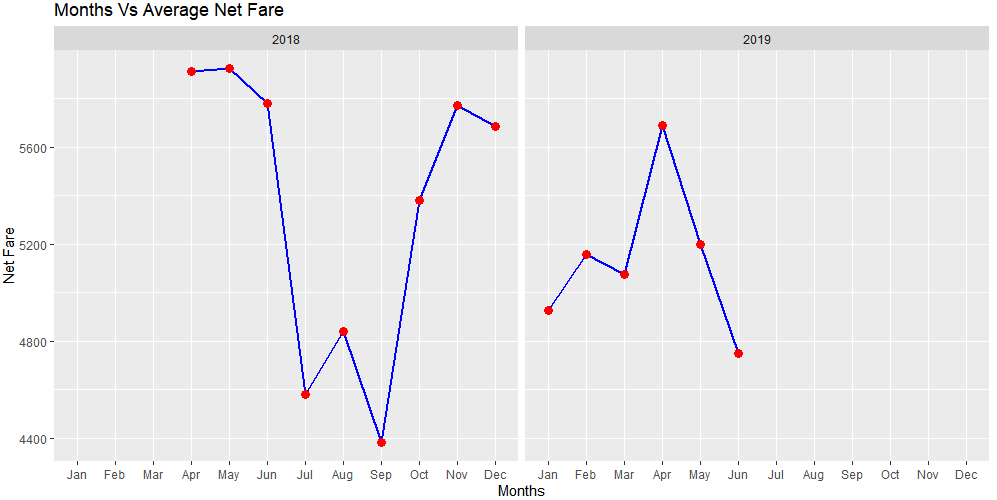
From the above plot, it is able to observe that NetFare reduces at midnight and starts increasing from early morning and remains almost constant with little fluctuations through the day hours.

2. Hours of the day vs Total Average Netfare total and Min netFare



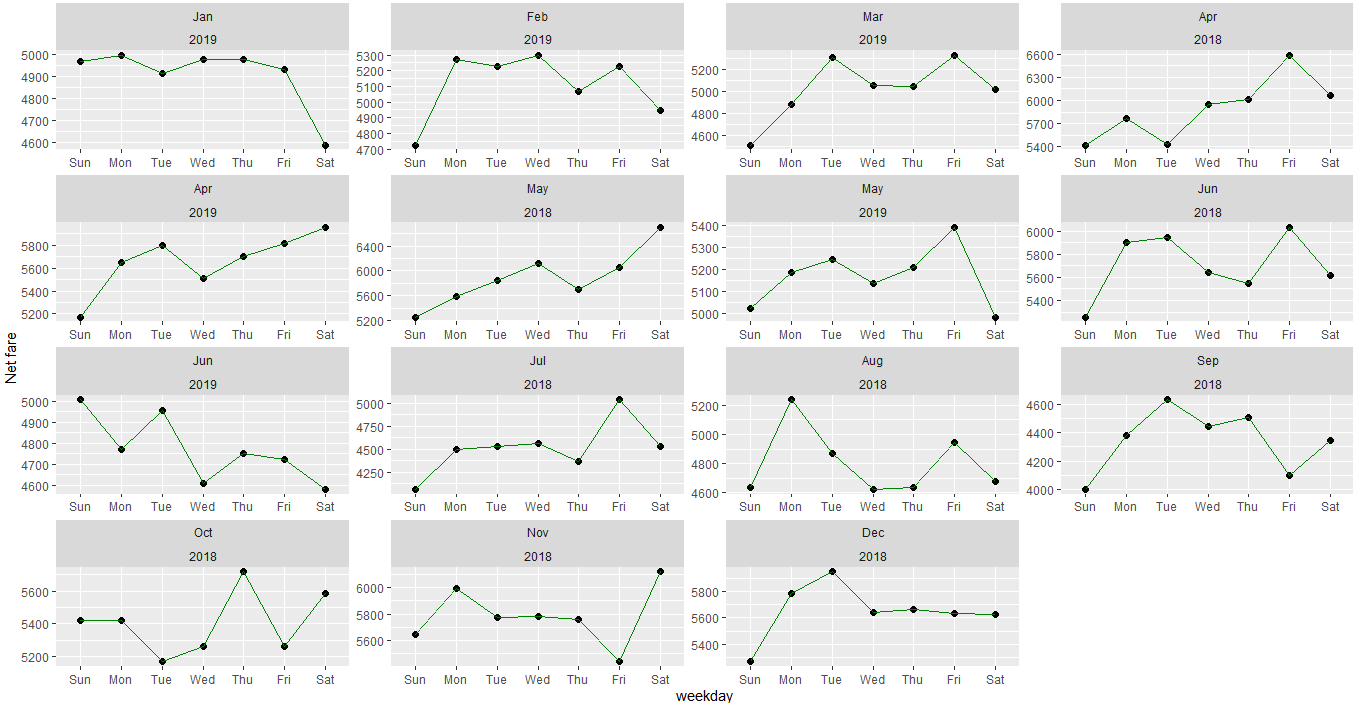
The red line shows about the min NetFare and blue line indicates TotalAverage Netfare for different hours of the day.

2.) Average Monthly NetFare for Air for 2018 and 2019

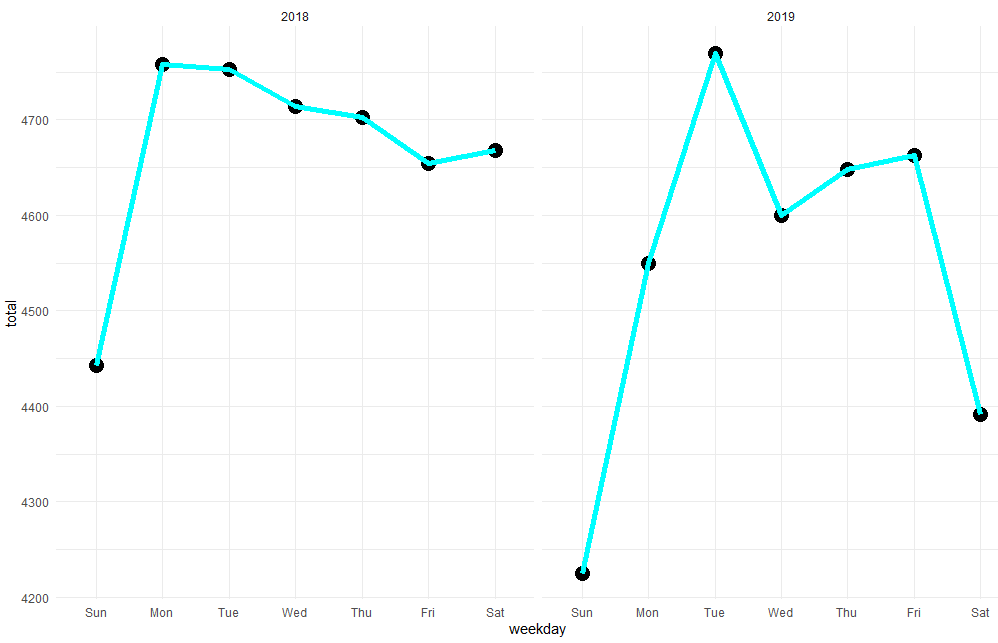


From the above graph, For 2018, the average Air fare is lowest for July and September and highest in the month of April and October. For 2019, the highest Air fare is in the month of April. It can be assumed that for the month of April, as summer vacation starts may be there is increase in NetFare and gradually decreases upto July and from September it again increases as because festival seasons starts during that time.

3.) Weekday trend for every month for the year 2018 and 2019



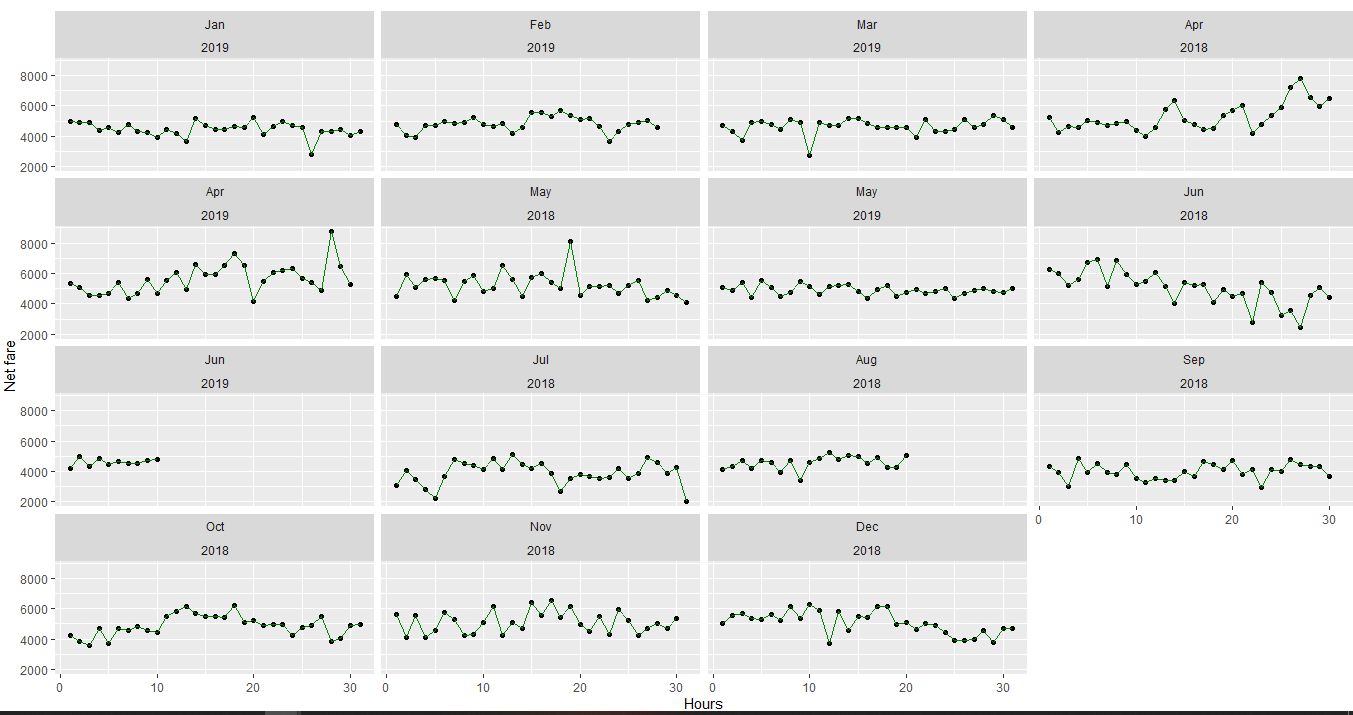
The above plot, it can be said that there is no fixed weekday pattern for all the months of both the year



The above plot represents the average NetFare in Weekday basis of both the year.

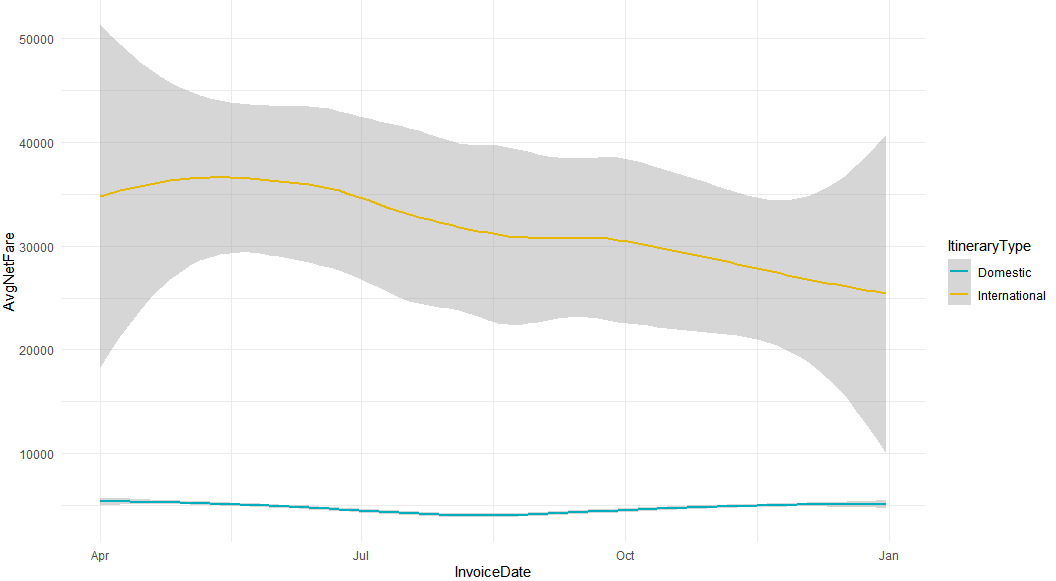
For the year 2018, AirFare is higher in weekdays and same situation arises for the year 2019 as well. AirFare is lower at Sundays for both the year.

v.) Month wise trend for each day for the year 2018 and 2019



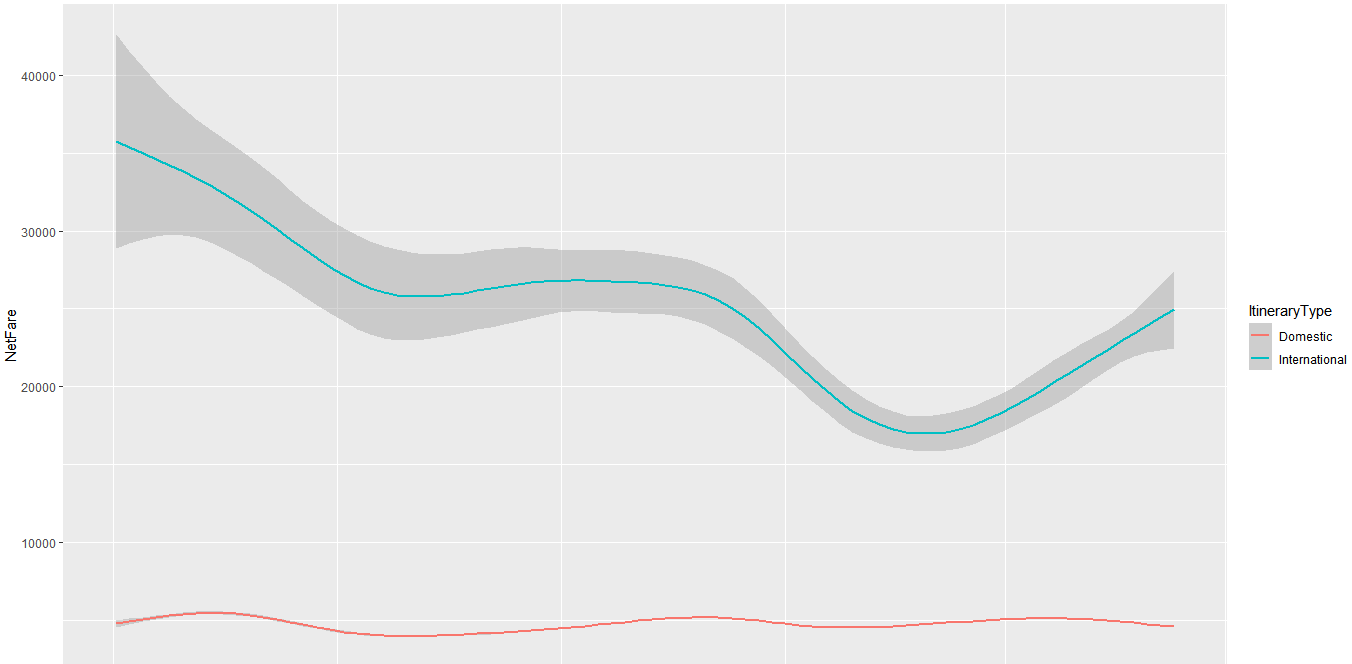
**Analysis for Domestic and International NetFare**

Comparison of average values for domestic and International

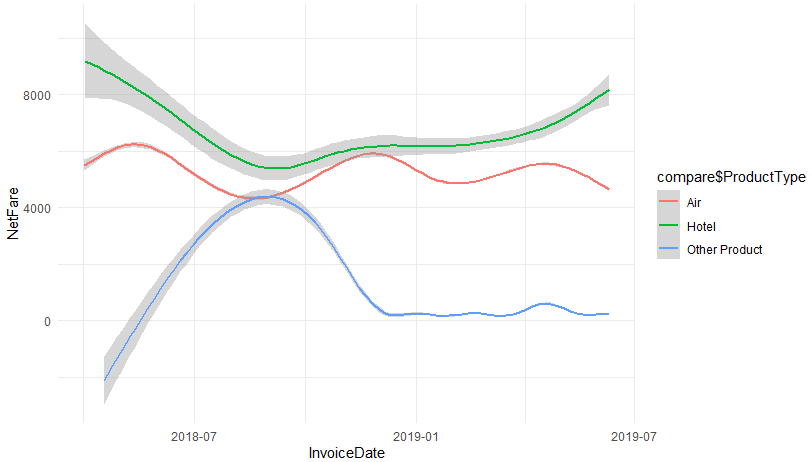


The above is comparison of a segment of data which shows the trend of both domestic and international trend of NetFare. Domestic fare ranges is less than the international NetFare.

**Over all trend of NetFare for the period of both Domestic and International**

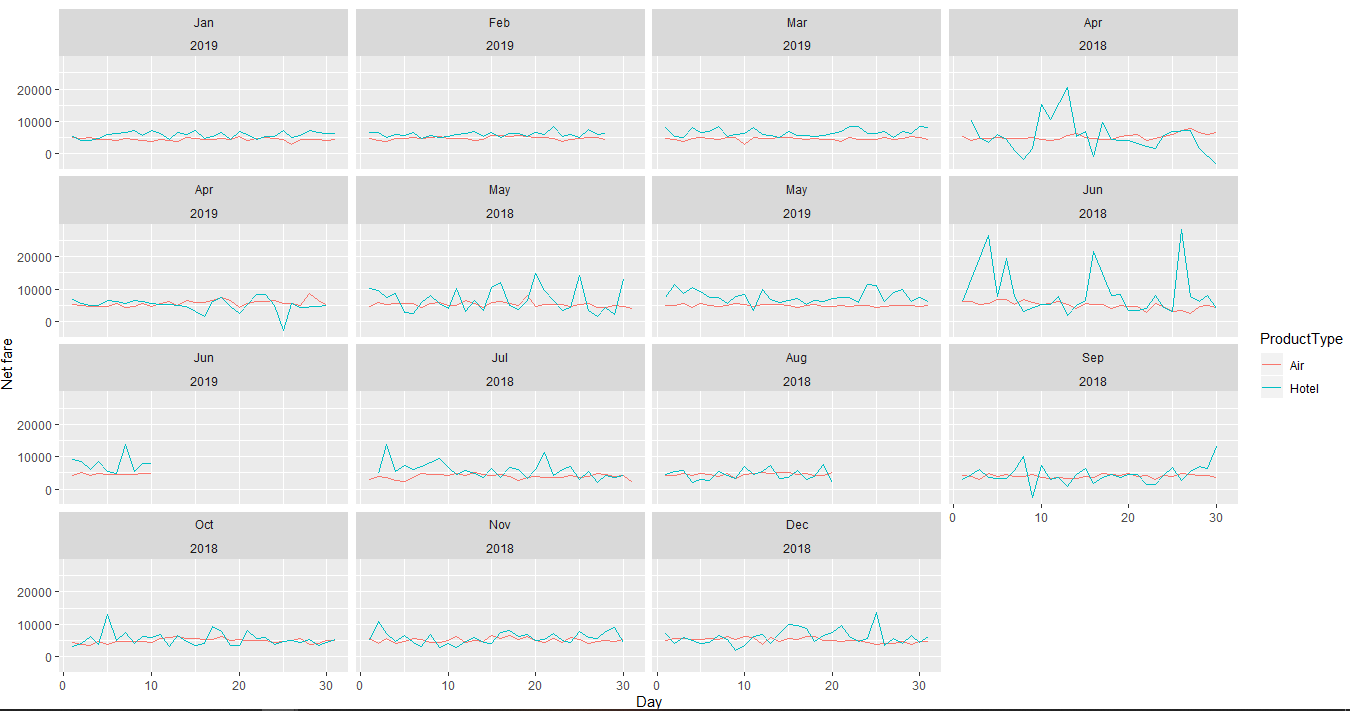


**Trend Comparison among different product type only for domestic data**



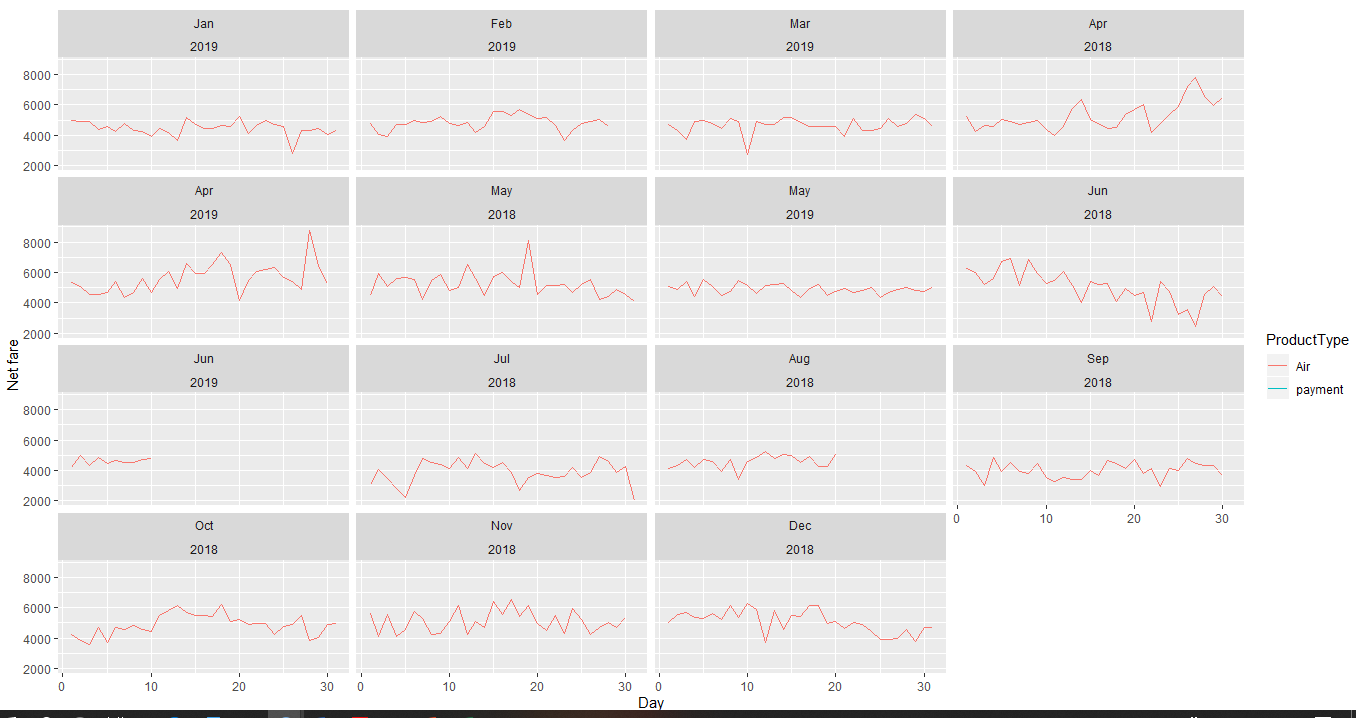
The above product represents the graph of three products viz Air, Hotel, Other Product together to explain how the trend varies with each other.

1. AIR and HOTEL



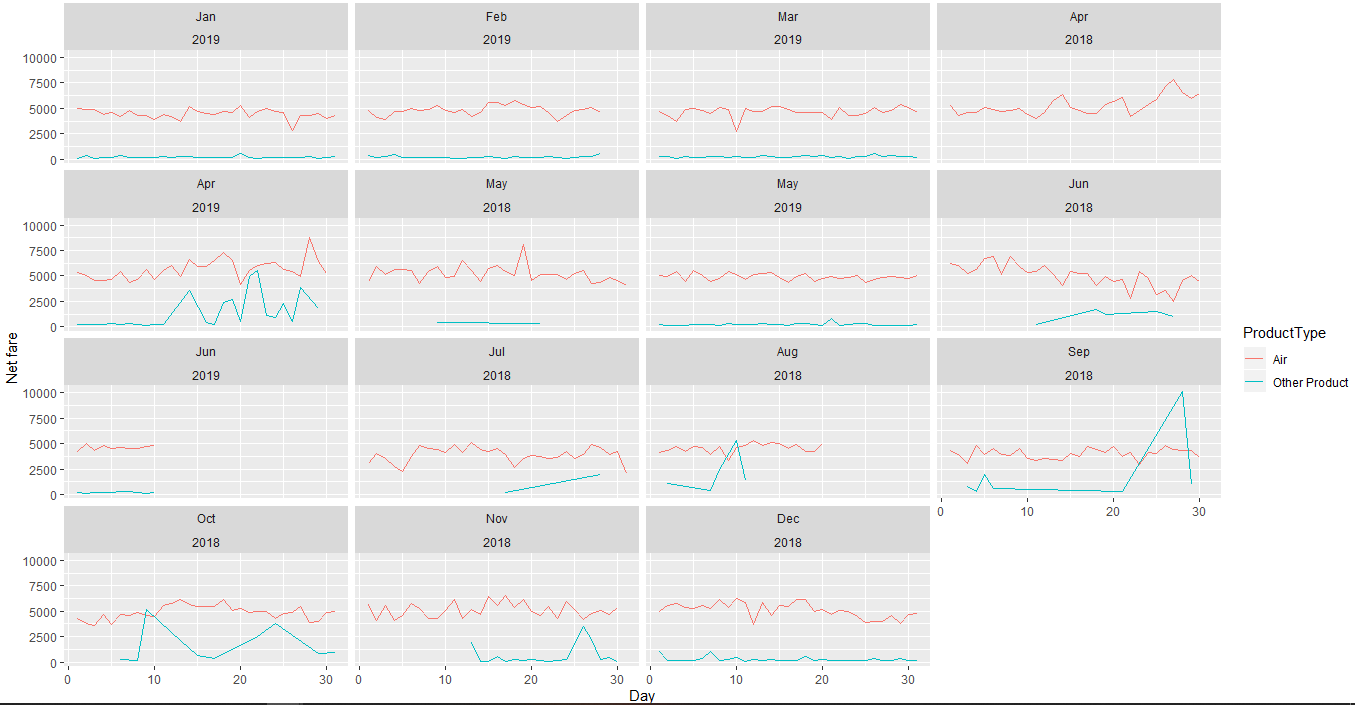
From the above plot it can be said that the airfare and hotel fare do not have any influence that can be compared with each other.

2. AIR and Payment



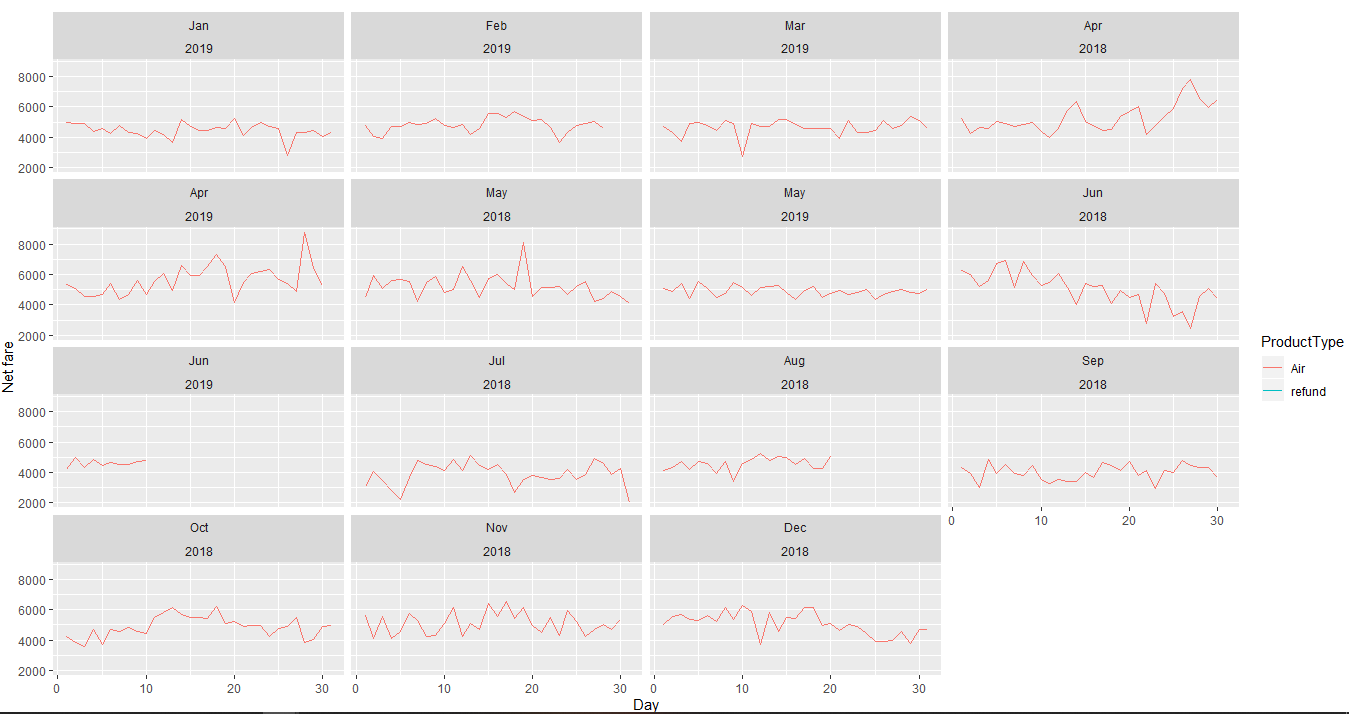
Payment doesnot not contain any netfare value so couldnot able to compare.

3. AIR and OTHER PRODUCT



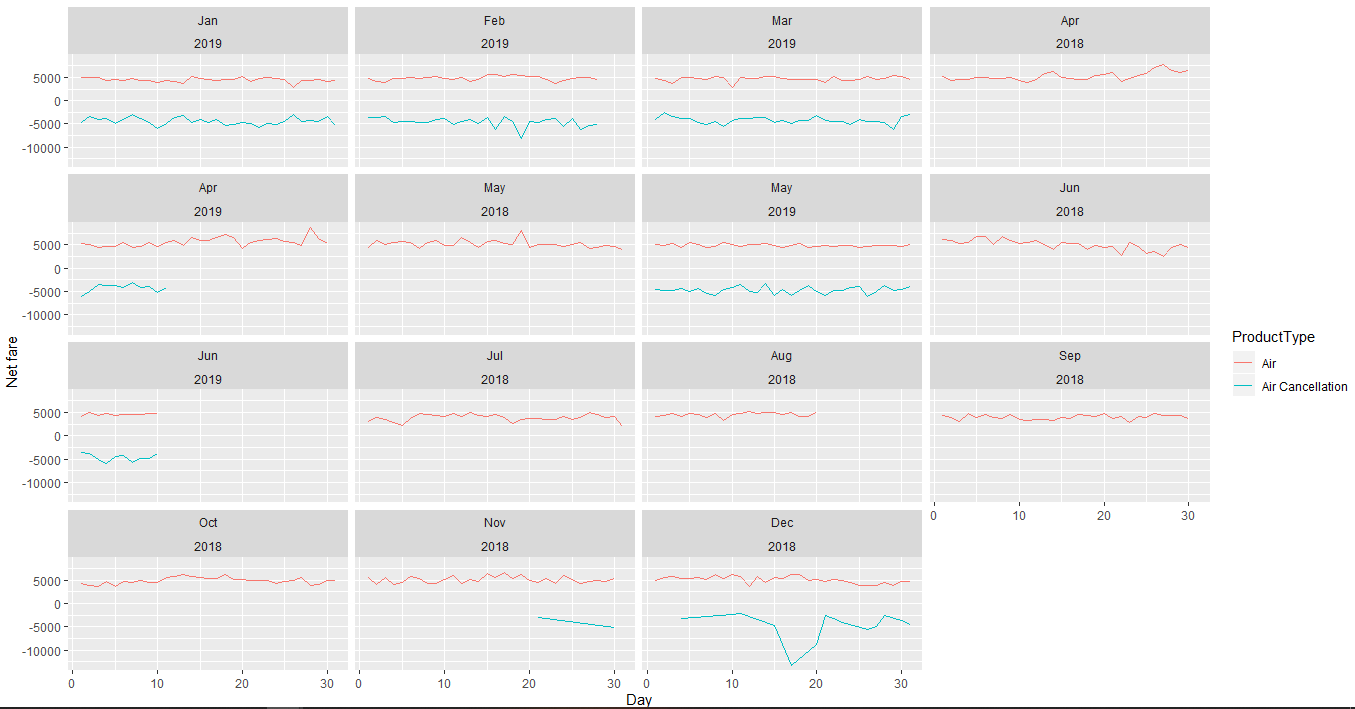
The rise of fare of other product is not having any influence in the raise and fall of airfare which can be visible from above month wise trends for the year 2018 and 2019.

4.) AIR And REFUND



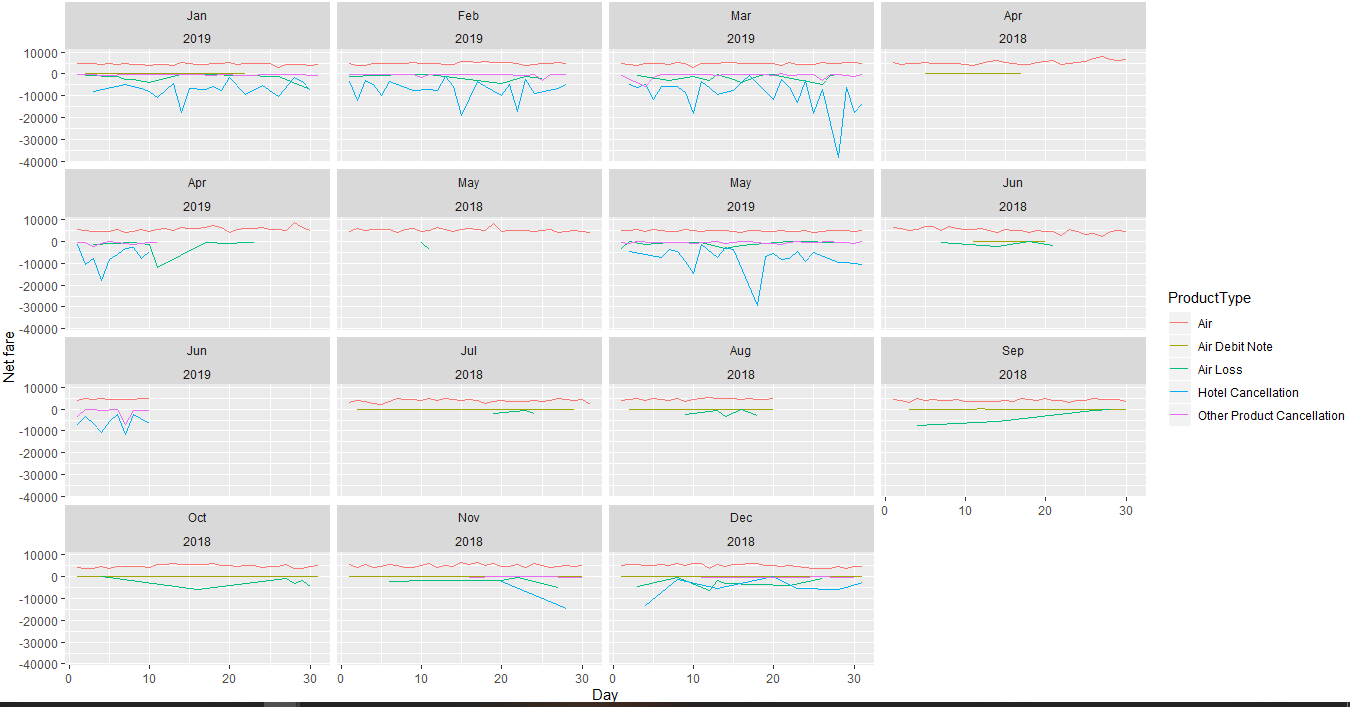
Even there is no value for refund

5.) AIR and AIR FARE CANCELLATION



There is no influence of airfare cancellation on air

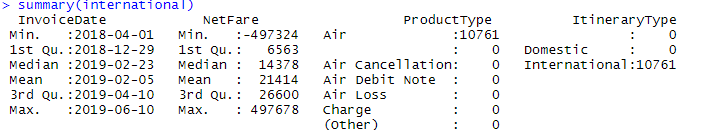
6.) AIR and OTHER PRODUCT CANCELLATION, HOTEL CANCELLATION, AIR DEBIT NOTE, AIR LOSS



Even the charges of other products like Other Product Cancellation, hotel cancellation, air debit note and Air loss also do not have any influence on Air Fare

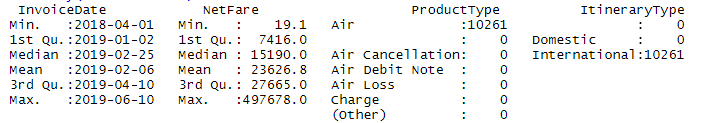
**Analysis for International Airlines**

Filtering out the international data from airlines



From the above summary, we get to know that there are some –ve and 0 values for net fare for Air and based on domain knowledge we can say that there can’t be any negative value or zero for NetFare

So arranging the data in ascending order and excluding the data which contain negative and zero values.

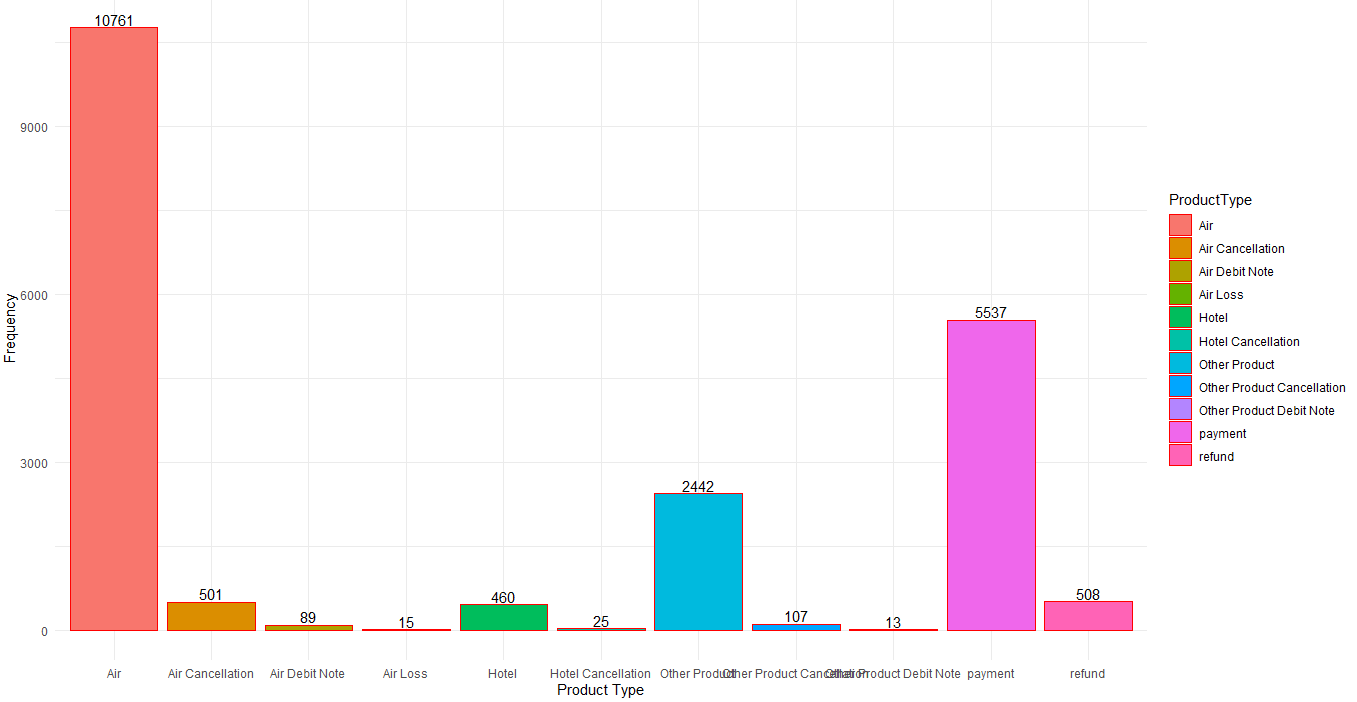
Summary 

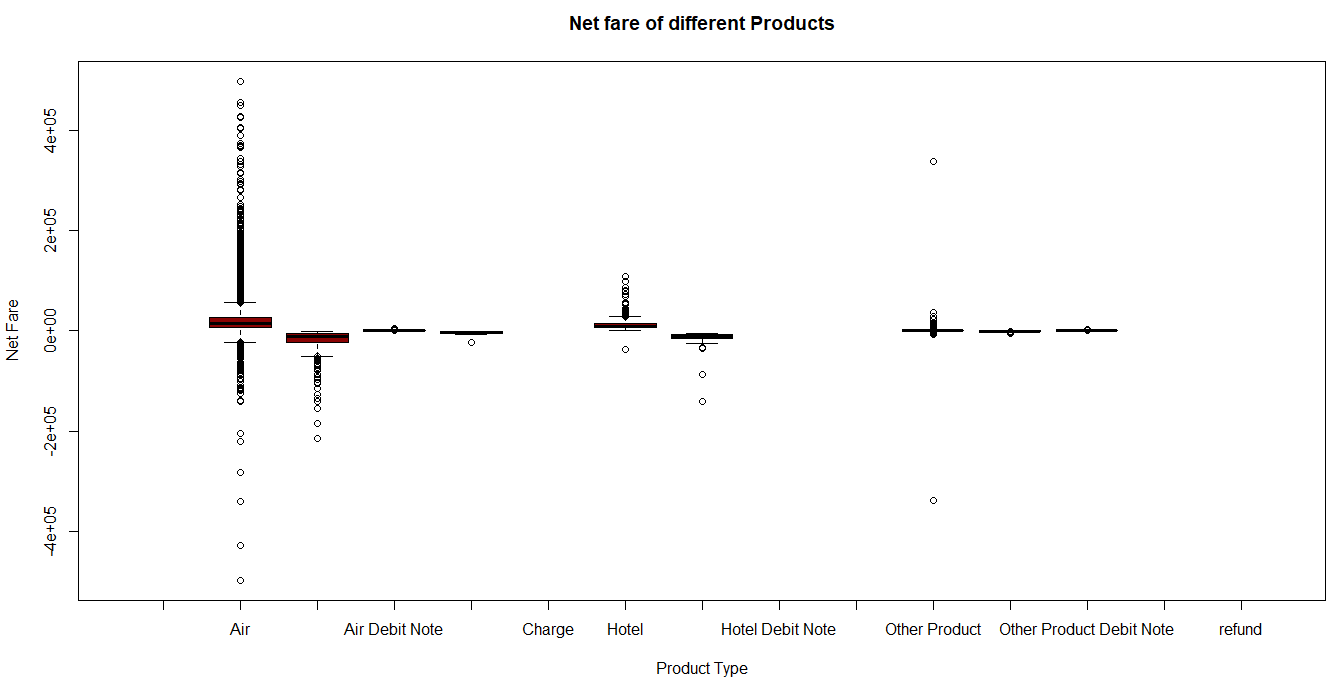
**i.) NetFare**

Statistical Summary

|  |  |
| --- | --- |
| Mean | 23627 |
| Median | 15190 |
| Variance | 1064600277 |
| StdDev | 32628 |

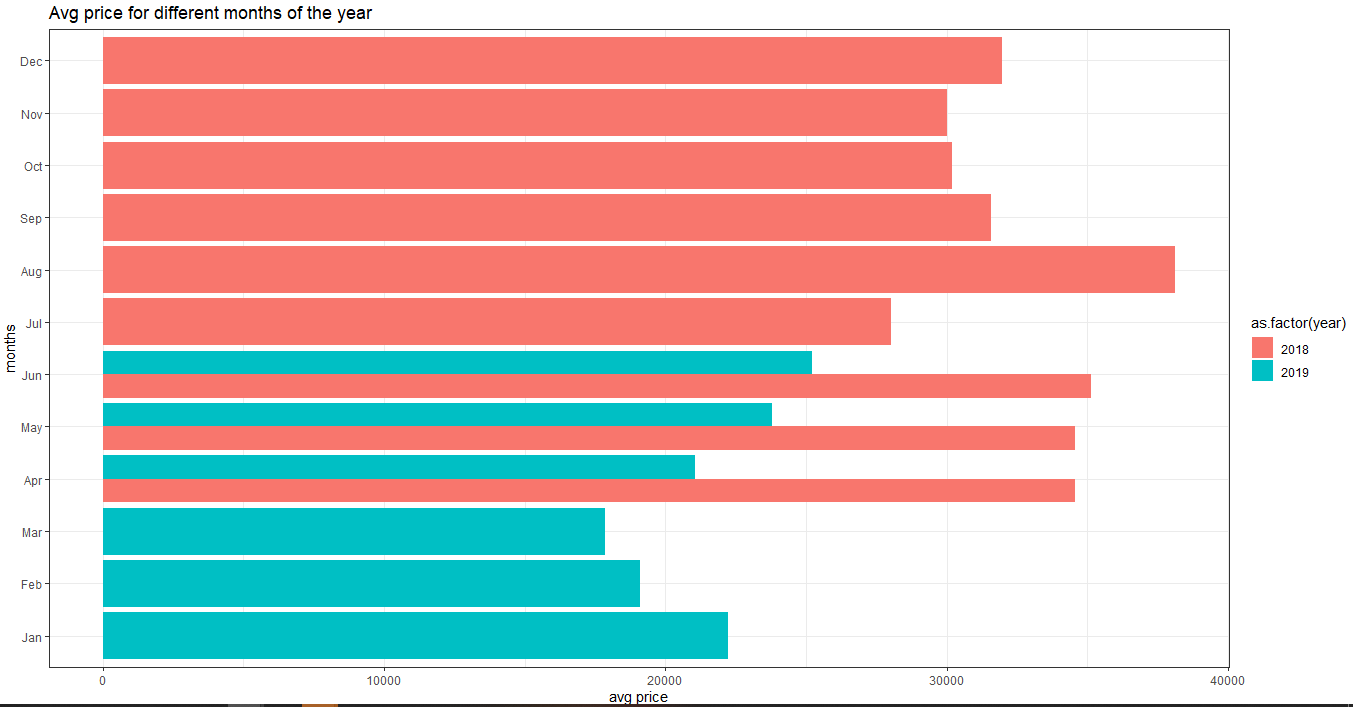
**ii.) Product Type**





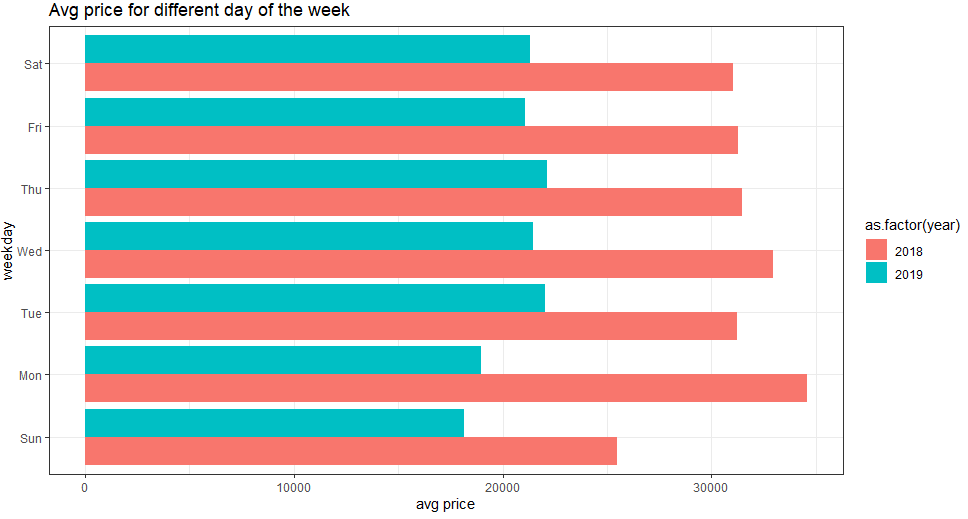
**Trend Analysis of International data**

1. Monthly analysis



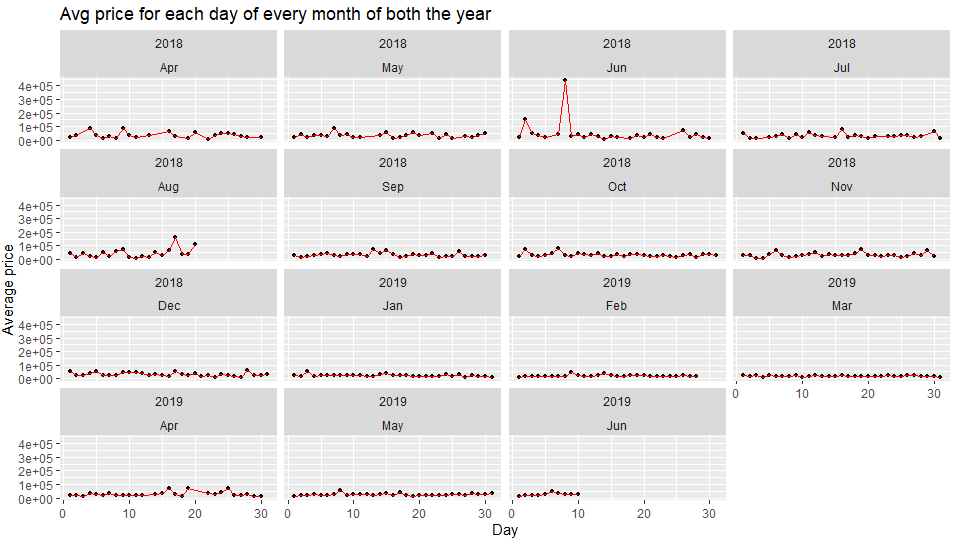
The above plot represents the average price for each month for 2018 and 2019. In August 2018 has the highest peak in fare. And In 2019, June has the highest peak

2. Weekly Analysis

an

For 2018, Monday has highest fare and Sunday has the lowest fare. For 2019, Thursday and Tuesday has the highest fare.

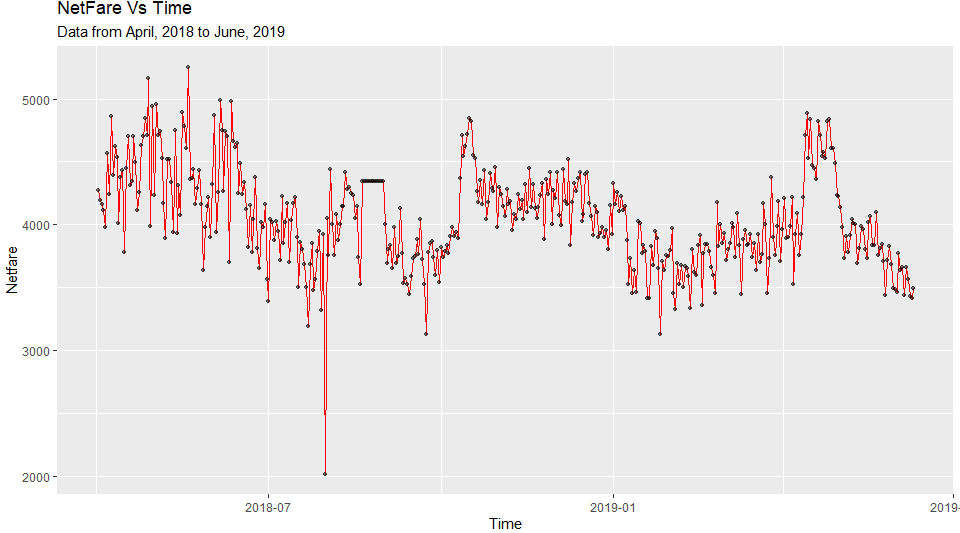
**3. Everyday analysis**



There is no specific pattern can be identified from the above plots.

**Model Building:-**

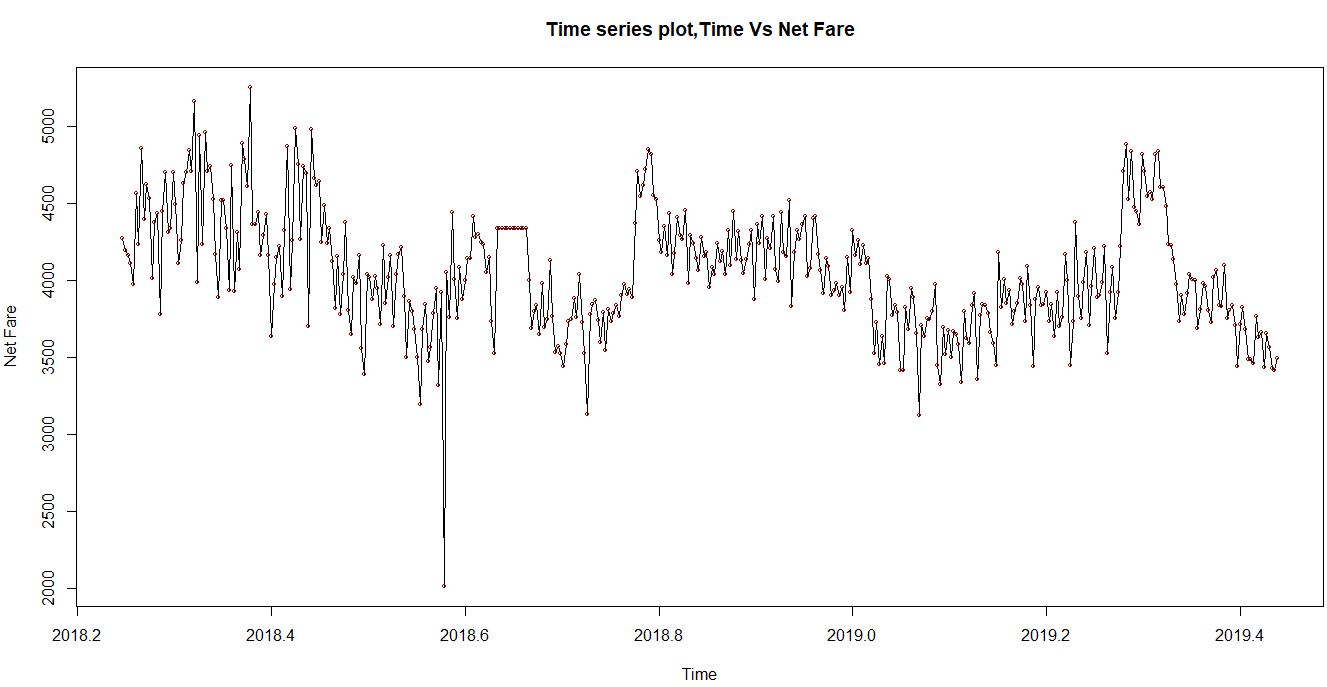
**Domestic Data**



Data is converted to time series data.

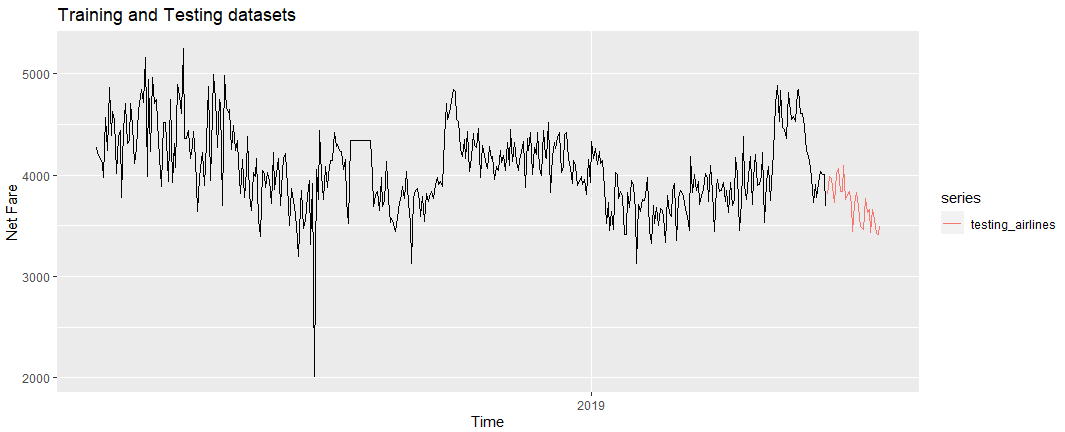
The following is the summary of the time series data.





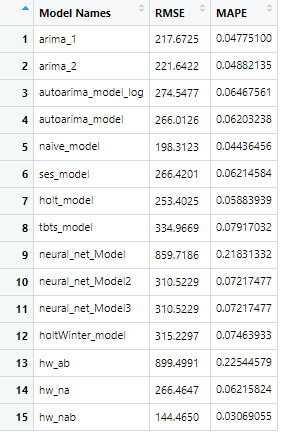
Splitting the data into train and test.

406 days as train and 30 days as test data



The above plot represents the train and test data trend of domestic airlines.

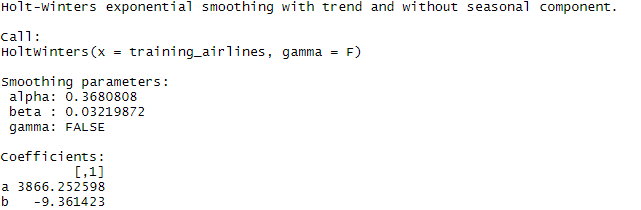
The following table represents the RMSE and MAPE value for different types of model built on different forecasting model.

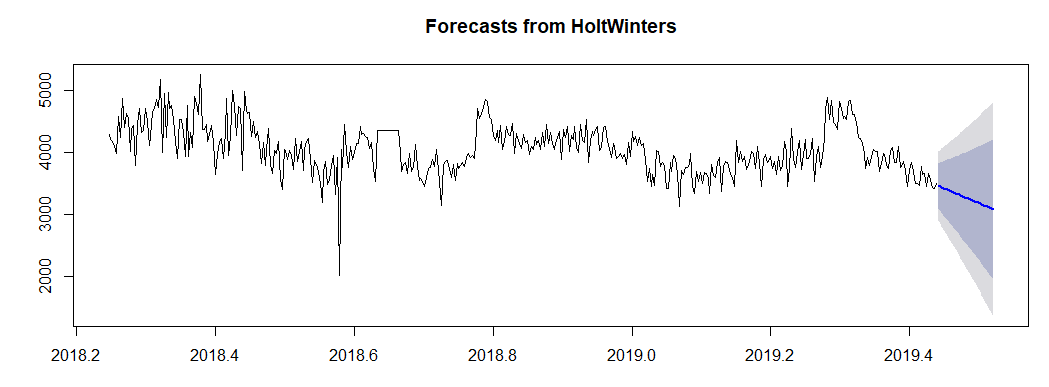


From the above table, the lowest RMSE model is Holt Winters’ model where it takes no Gamma value and takes optimum alpha and Beta value.

The following is the summary of the Holt Winter’s Method

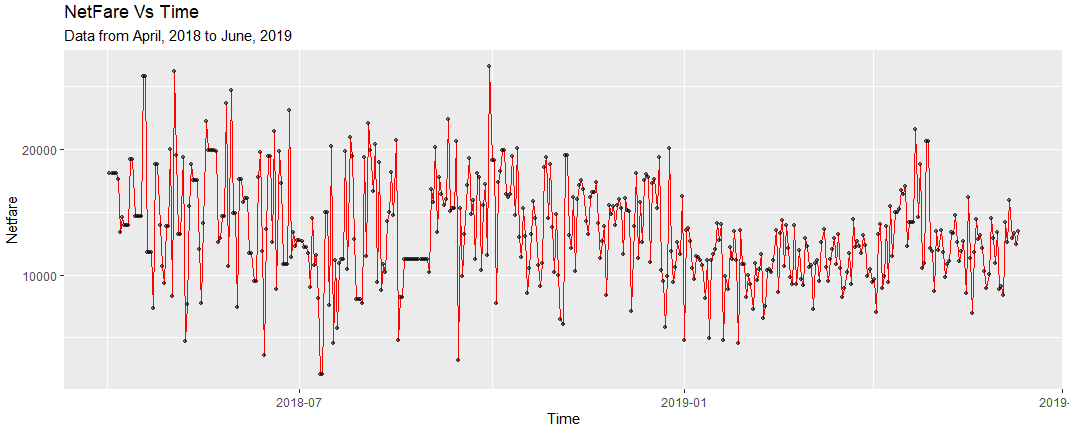
Summary:



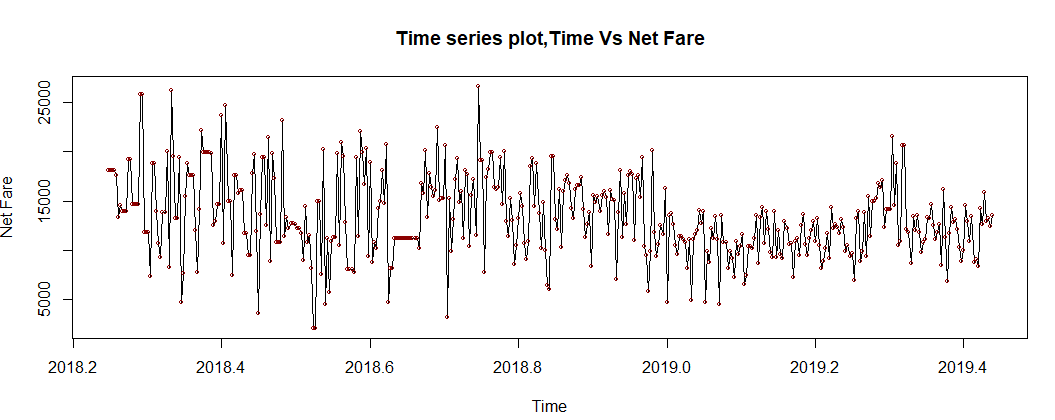


The blue colour line indicates the forecasted 30 days.

**Model Building for International Airlines**



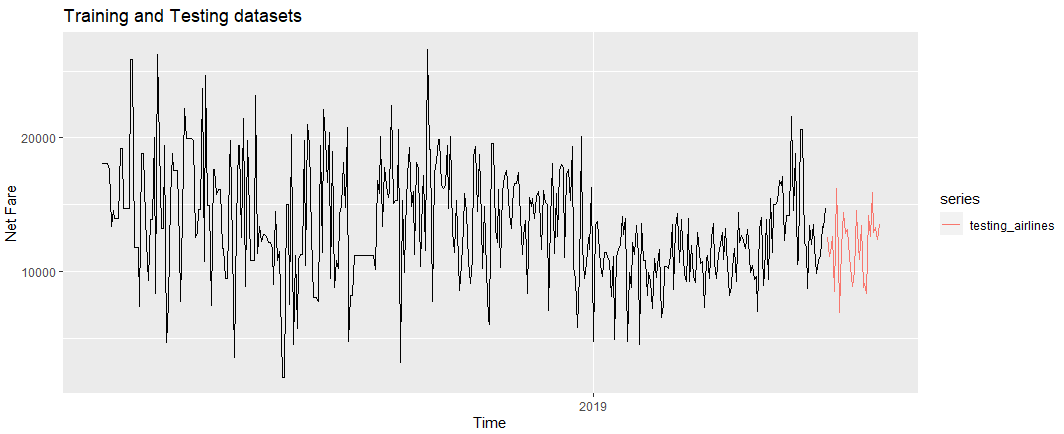
Converting to time series object





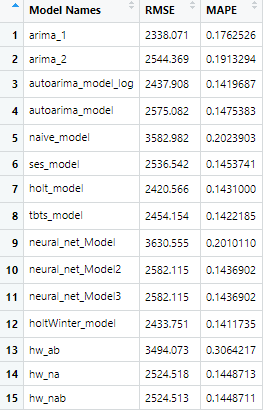
Splitting the data into train and test

406 days data is for train and 30 days data for test

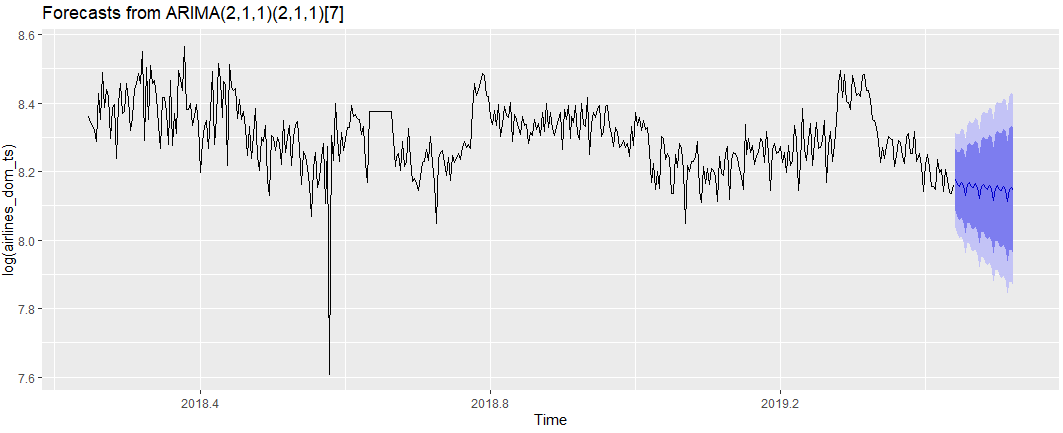


The above plot represents from train and test of international airlines.

The following table represents the RMSE and MAPE value for different types of model built on different forecasting model.



From the above table arima model has lowest RMSE model



Deployment

The following dash board is built on R shiny to forecast airfare for next 30 days form June10, 2019 to July 9 2019. The user can select domestic as well as international and can predict the netfare for respective dates.

