# Predict_413_Discussion_Week_3

*Singh, Gurjeet*

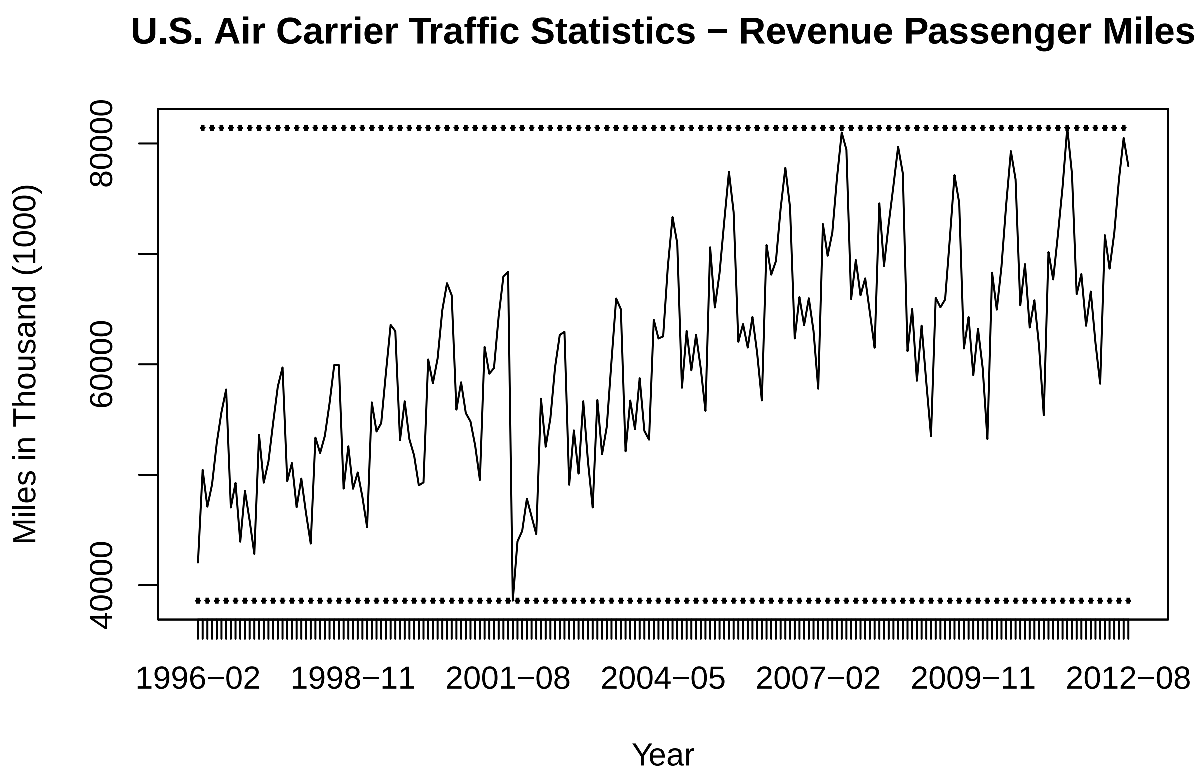
*January 24, 2018*

## Contents

## 1. Data

For this week's discussion, I am using the same dataset (U.S Air Carrier Traffic Statistics – Revenue Passenger Miles) from the last week. The figure below shows that this data has strong seasonality within each year. Along with the seasonality, there is also an upward trend to it.

```
#Import the data
US.air.traffic  <- read.csv("us-air-carrier-traffic-statistic.csv")
colnames(US.air.traffic) <- c("YearMonth", "Revenue_Miles")
plot(zoo(US.air.traffic$Revenue_Miles/1000, US.air.traffic$YearMonth),
        main = paste0("U.S. Air ",
                      "Carrier Traffic Statistics ",
                      "- Revenue Passenger Miles"),
        xlab = "Year",
        ylab = "Miles in Thousand (1000)")
```




### 1.1 Split Data - Training/Test set

We split the data into the training set and test set using 70/30 split approach. We will use the training set to build the model and test data for validation and check the accuracy of the model.

```
#Split the data into training set and test set
train.df <- ts(US.air.traffic[1:139,"Revenue_Miles"], frequency = 12) #training set
test.df <- ts(US.air.traffic[140:199,"Revenue_Miles"], frequency = 12) #test set
```

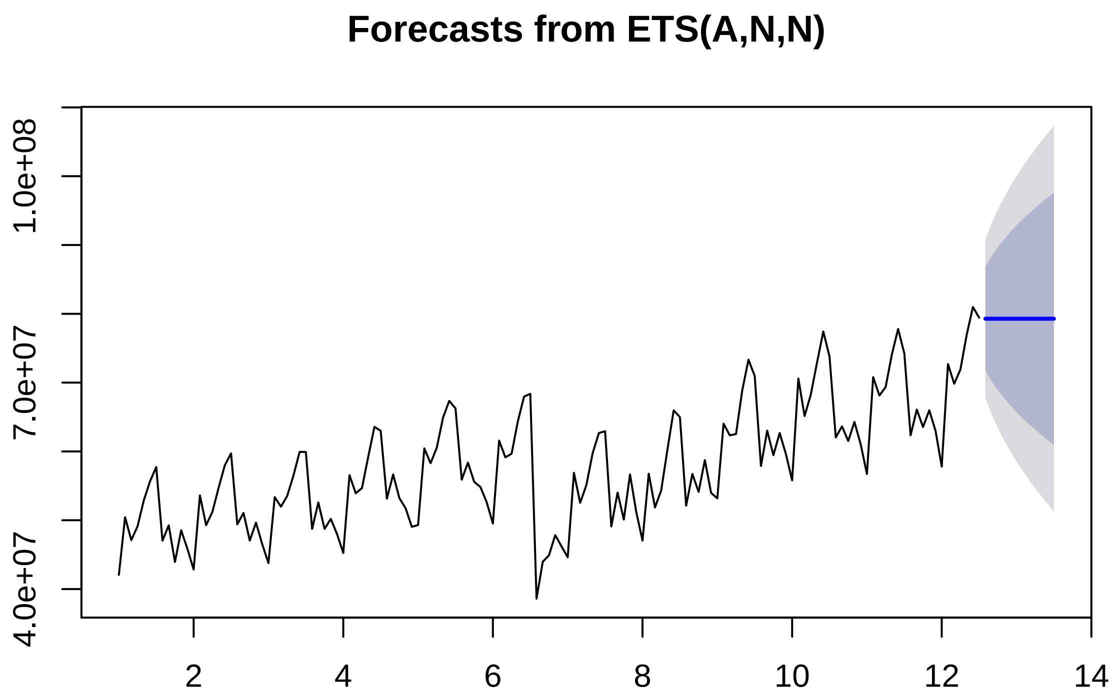
## 2. Model

For this exercise, we create five (5) different EST models.

### 2.1 EST Model ANN

The figure below shows our first EST model ANN created using Additive for season part only. It looks like this model gave us the flat forecast. Not sure if this is the correct model to use for this data. Therefore, we will move to the next model.

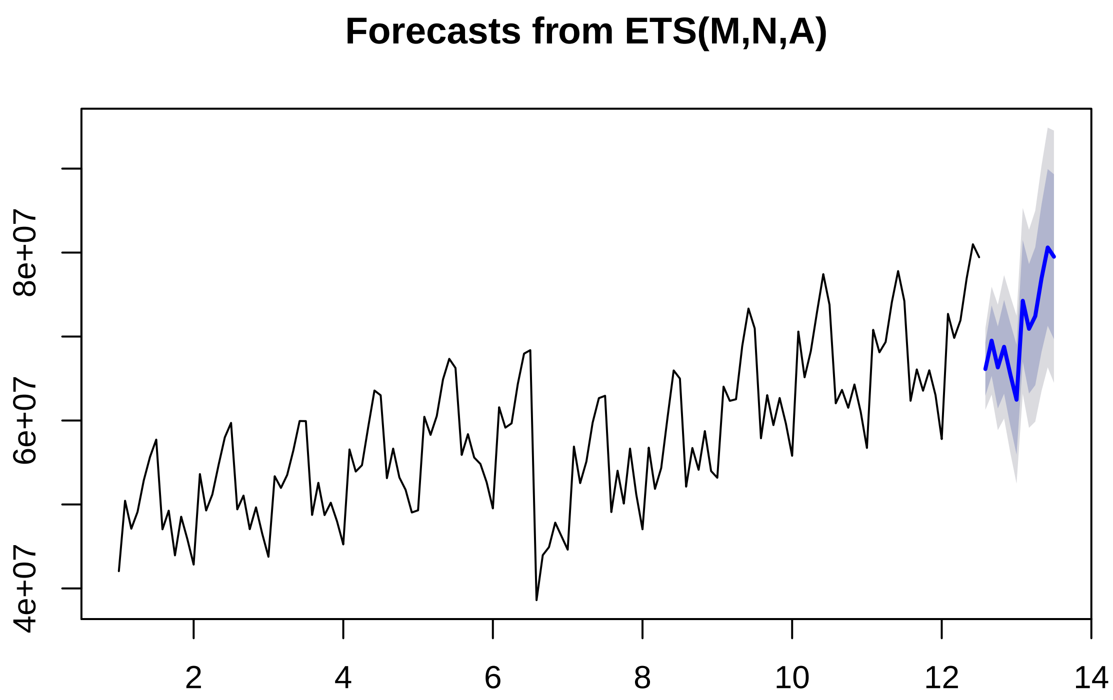
```
#ETS ANN
EST.Model1 <- ets(train.df,model="ANN")
forecast.Model1 <- forecast(EST.Model1,12)
acc.Model1 <-accuracy(forecast.Model1,test.df[1:12])
plot(forecast.Model1)
```



### 2.2 EST Model MNA

Our next EST model MNA was created using Multiplicative for season and Additive for random. From the figure below, this model seems to forecast much better than our first model. This one is clearly the winner among the two. However, we would like to experiment with a few more model to make sure we pick the best model.

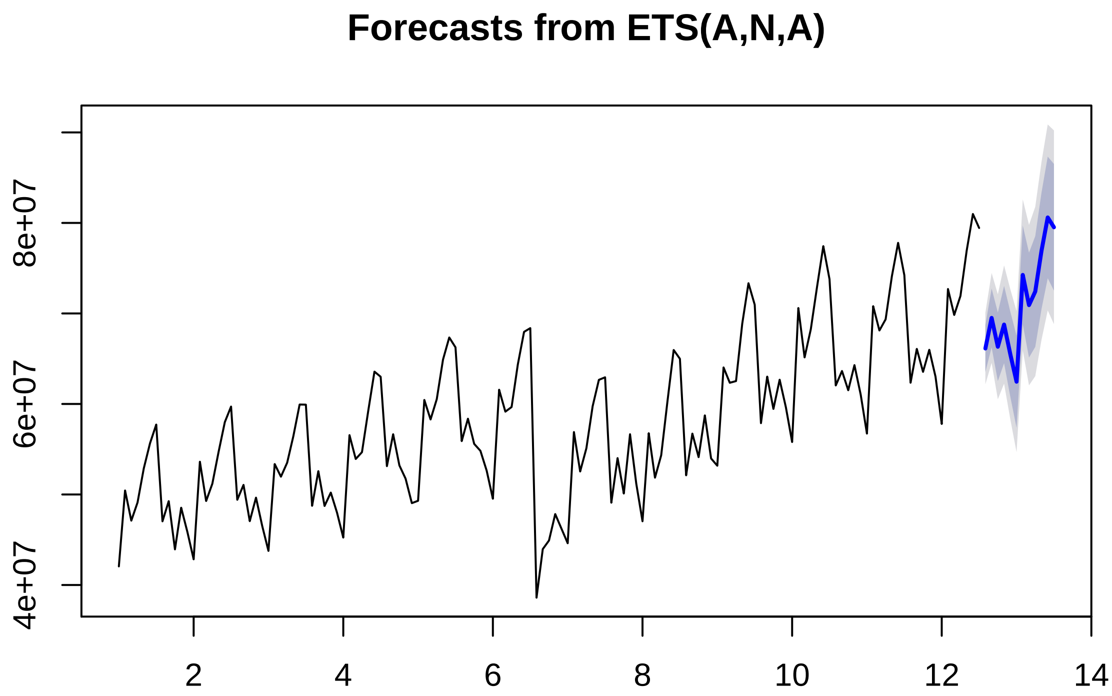
```
#ETS MNA
EST.Model2 <- ets(train.df,model="MNA")
forecast.Model2 <- forecast(EST.Model2,12)
acc.Model2 <-accuracy(forecast.Model2,test.df[1:12])
plot(forecast.Model2)
```




### 2.3 EST Model AUTO

Our next EST model AUTO was created automatically by the est function. It seems like it has selected additive for season and random. This makes a lot of sense because our data has a strong season to it.

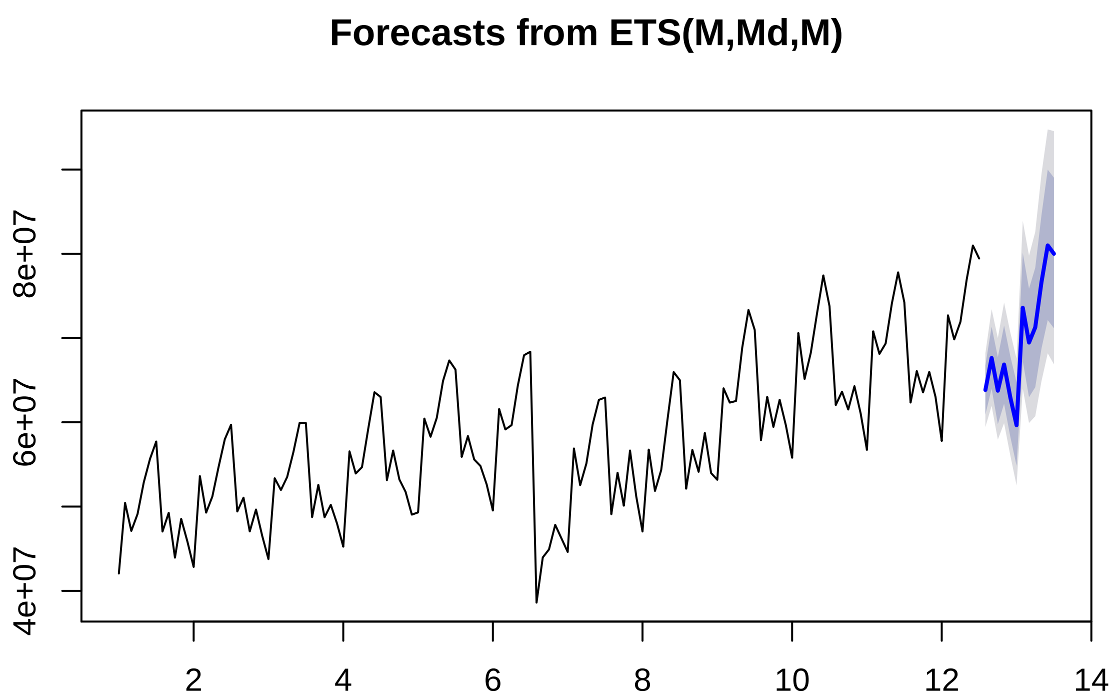
```
#ETS AUTO
EST.Model3 <- ets(train.df)
forecast.Model3 <- forecast(EST.Model3,12)
acc.Model3 <-accuracy(forecast.Model3,test.df[1:12])
plot(forecast.Model3)
```


Forecasts from ETS(A,N,A)
4e+07
6e+07
8e+07
2
4
6
8
10
12
14

### 2.4 EST Model MMM

Our next EST model MMM was created using Multiplicative for the season, trend, and random. It is difficult to determine from the figure below for any changes from the previous model's graph. Looking at the metrics will definitely give us some sense.

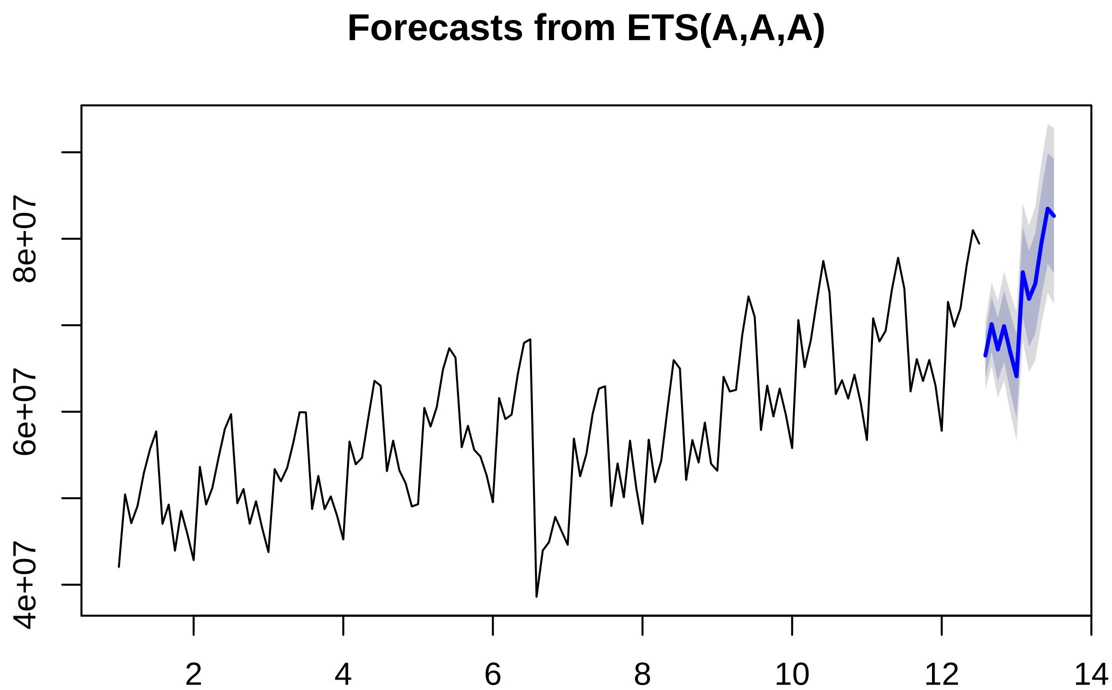
```
#ETS MMM
EST.Model4 <- ets(train.df,model="MMM")
forecast.Model4 <- forecast(EST.Model4,12)
acc.Model4 <-accuracy(forecast.Model4,test.df[1:12])
plot(forecast.Model4)
```




### 2.5 EST Model AAA

Our last EST model AAA was created using Additive for the season, trend, and random. Again, it is difficult to determine from the figure below for any changes from the previous models' graph. Looking at the metrics will definitely give us some sense.

```
#ETS AAA
EST.Model5 <- ets(train.df,model="AAA")
forecast.Model5 <- forecast(EST.Model5,12)
acc.Model5 <-accuracy(forecast.Model5,test.df[1:12])
plot(forecast.Model5)
```


Forecasts from ETS(A,A,A)
4e+07
6e+07
8e+07
2
4
6
8
10
12
14

## 3. Metrics

The tables below list all the metrics for each model. Model ANN, MMM, and AAA performed really poorly in all the metrics. Only two models MNA and our AUTO (ANA) model performed really well. Both the models' metrics results are very close.

Table 1: Metrics for EST Model ANN

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | 321481.9 | 5899462 | 4330349 | -0.2471401 | 7.722379 | 0.9389349 | 0.022054 |
| Test set | -8874102.6 | 10386800 | 8944377 | -13.2635274 | 13.351693 | 1.9393790 | NA |

Table 2: Metrics for EST Model MNA

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | 177762.2 | 2018775 | 1251316.7 | 0.1969194 | 2.328212 | 0.2713187 | -0.0085417 |
| Test set | -699916.0 | 1047376 | 802749.9 | -0.9922800 | 1.131876 | 0.1740575 | NA |

Table 3: Metrics for EST Model AUTO (ANA)

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | 188768.0 | 2017198 | 1255135.2 | 0.2136787 | 2.336910 | 0.2721467 | 0.0285181 |
| Test set | -697919.1 | 1047349 | 801421.7 | -0.9890646 | 1.129569 | 0.1737696 | NA |

Table 4: Metrics for EST Model MMM

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | 37156.81 | 1909798 | 1171873 | -0.0663533 | 2.181977 | 0.2540933 | 0.0270323 |
| Test set | 688347.59 | 1672603 | 1526493 | 1.1069976 | 2.199053 | 0.3309843 | NA |

Table 5: Metrics for EST Model AAA

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -122979.1 | 2008905 | 1225950 | -0.3269398 | 2.305273 | 0.2658186 | 0.0446196 |
| Test set | -2445528.1 | 2823053 | 2445528 | -3.4170953 | 3.417095 | 0.5302556 | NA |

## 4. Conclusion

In conclusion, we would select Model AUTO (ANA) because additive decomposition is a good choice here because there's no fluctuations or the variation in the trend cycle in the dataset.