Group Project

*Han Li*

The goal of this project is to predict the number of sold tickets for each flight. With historical data, we can get how many tickets have been sold before n days of departure. Here n is from 0 to 60 in historical data, where n is equal to departure date minus current date in Dataframe. When n is equal to 0, the tickets represents for the final sold quantities for each flight. Then we can get the mean of how many tickets have been sold before departure for each n. Finally, there is a relation between n and the mean of the quantity of sold tickets on that day.

With this relation, there are two methods to do the prediction:

1. Addictive model
2. Multiplicative model

Using addictive model to forecast is to add how many tickets have been sold from validation data and the mean of quantities of historical tickets that would be sold between current date and departure date from this relation. In order to get this quantities, we could use the mean of final sold quantities minus each mean of quantities of historical sold tickets on current date from the above relation. Therefore, we can get all the forecast result only using addictive model.

The same as addictive model, we could run a multiplicative model and get the forecast result on with multiplicative method. The only difference is the method to predict. In this time, we use the quantities of each n days before departure divided by the final sold tickets and get this ratio. Therefore, we get the result of prediction using multiplicative model as the quantities of sold tickets on each day divided by this ratio.

With addictive result and multiplicative result, we can combine them together for the best forecasting result. In this project, we select the condition to be 8 days before departure. Because from comparison of addictive result and multiplicative result, when the day is less than 8, addictive result is more accurate to the real result. On the other hand, when the day is equal to or more than 8 days, multiplicative model does a better prediction. Therefore, we should combined them together.

MASE in this project is calculated as 71.46%, which means comparing with Naive prediction, the forecast error is 71.46% and the total reduced error is 28.54%.