Weihao Yuan (Melvin), Zeyang Wu (Duncan)
IS 5201
Professor Lee
March 14, 2018

Final Project: Airline Demand Forecasting

The Airline Demand Forecasting project is intended to help the management of XYZ Airline to predict the future sales and increase revenues. Based on the size of the dataset, advancing booking model is adapted to generate predictions of sales. The goal of this project is to minimize the forecast errors and generate the most precise predictions. Out of the two advanced forecasting methods, the additive method simply because it returns fewer errors. This method uses the cumulative booking tickets at a particular day prior to the departure date and the historical average remaining demand for days prior to departure to estimate the final sales.

The model uses the training dataset, which contains information of departure date, booking date, and cumulative booking, to find the historical average remaining demands for days prior to departure and for each day of the week. After importing the training data and validation data, the function use departure date and booking date to calculate the days prior to departure, (the difference between departure date and booking date). It also calculates the weekday for each departure date. These two steps are applied to both datasets.

Next up, the function converts the training data into a pivot table, by using weekday and departure date as indexes, days prior to departure as columns, and cumulative bookings as values. Within the pivot table, the function uses a for loop to calculate the remaining demands for each given departure date and day prior to departure. After that, the model uses group function to calculate the historical average remaining demands based on each weekday and merges this information to the validation dataset on weekdays and days prior to departure.

The model adds the historical average remaining demands on the cumulative booking in the validation dataset to calculate the final forecasting demand. To find out the precision of the forecasting result, the function calculates the mean absolute errors (MAD) for both forecasting result and the naive forecast. Errors mean the difference between the final demands and the forecasting result (or the naive forecast).

Finally, we use the mean absolute scaled error ratio (MASE) to examine the accuracy of the forecasting model. We use the MAD of our forecast model and divide it by the MAD of the naive forecast to get the MASE.

The model returns a MASE ratio of **0.78**, which means the model successfully reduces 22% of forecast errors.