The senior pilots have significant flexibility in selecting flights according to their vacations and trainings. These absences will not span an entire month, so part of the pilot’s schedule stays intact.

This flexibility creates difficulty in forecasting pilot availability and estimating the required number of pilots for covering the unselected flights.

Objectives

Automate the process of forecasting unselected flying time by implementing an R algorithm

Accurate predictions using R to prevent under or over forecasting of unselected flying time

Impact

Elimination of superfluous costs resulting from over/under estimation of required pilots

Assists Alaska Airlines in hiring required number of pilots

Improved reliability of predicted variable as compared to traditional methodology

Capitalizing a pilot’s flying time will lead to increased high pilot productivity

Data Modeling

Analyzed and determined relevant input variables affecting the output variable.

Modified datasheets to include all input variables (trainings, vacations, etc.) for all the years.

Research

Experimented with different modeling techniques such as Binomial model, Generalized Linear Model, Linear Regression Model, etc.

Examined different methods to establish confidence in the predictions.

Evaluated various front end systems and GUI’s such as RExcel to run R code and present the data.

Implementation

Created two data sets for training and testing the model.

Trained the linear regression model with historical data set to predict values for output variable.

Tested for accuracy in the predictions with a 95% confidence interval.

Predicted Output Value

The output variable ‘Open hours’ is predicted using the trained Linear Regression Model.

The output consists of three values, Fit value, which is closest to the actual value and a Range (Lower and Upper Limit) which provides intervals around the predicted value.

Front end: RExcel

RExcel is an add-in for MS Excel. It makes R functionality accessible through menus and dialog boxes, and with just few clicks output variable can be predicted.

Data set can be easily modified and updated.