

Background

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





Predictive Analytics for Pilot Productivity

Dre	-	int.	-			.+	1/-1	
FI	-u	ICU	20		սւթւ	aι	vai	u
Sectors 1	arring S.	all horn is	en els	ing ang been	ing two here	14	utuar local	-
and a	108	January 1	1.0	44.45	4 1000.04	1008.00	LOSS ADDRESS	1.00
2767	208	34016.40	100	88.00	10 NOWYO	MIT-INE	DOWN MUTCH	-
100.2	47	1048.07		41.0	EVANCE	74.90	708.0023708	175
440	- 44	1048.00		43.5	10400	701.100	487 005 706 3	475
794.5	- 40	10008-080	130	41.44	4 PORCH	799.2390	442,7090,963	405
705	47	1004.00	150	41.44	4 40410	10.470	417.423.7543	
108.5	15	14000.407	1.00	41.08	3-ANCCA	843.0545	Test (channel)	100
50.5	40	1403.47	1.00	81.08	IR ANOTO	-	104.006017	-
440.5	45.2	pression of	104	83.49	5-30a0a	1406-118	1108.21004	-
10.00	275	Design of the local division of the local di	104	81.49	5.50470	1445-047	1000.127562	104
10.0.0	84	5.000	1.00	62.75	d'unité.	0174305	THE LOTION	-
101.0	10	8.000	1.00	62.75	LA LANTO	107.1403	PLANOWIDES	-
1018.2		1075.40	1.07	63.47	8 PONCE	600.000	248.7524767	-
444	- 40	10.75.40	1.09	41.47	8.00MPO	-	101.001.001	THE
104		100412-200	-	41.4	LE ANCOR	470,0400	796.0004291	-
146	10	1040.76	100	41.4	48 AACPO	40,100	THE OWNER.	-
401.5	403	20070.00	7.8	41.44	# 30404	1765-217	INCOME.	-
MORTS.	144	200.75.00	114	41.44	5.30410	1075.048	In the Association	1.01
160	41	5400.0	340	45.01	\$-uescie	104-04	site associate	100
800	41	5400.0	140	45.01	14 LAND	88.10	THE LABOR	100
100.0	10	1000.7.000	100	84.00	18 POINTA	40.140	754-4577046	1.05
10.7		10017-000	1.00	88.70	4 40410	PRIATO	and Principle	
1000.0	128	1410	6.82	45.35	1.45004	101.076	100.0124075	
1070.4	62	1414	6.82	45.36	LP ANORE	101.007	400.47002	100
4000	100	2404.10	100	41.04	10-10HOA	101.00	NUMBER OF TAXABLE	100
880.0	188	3404.10	140	81.04	LA NUMPO	100.00	ICOL AND THE	142
104.5	45	1074.70	100	45.04	LE-LANCA	80,500	ALC: NOTE: NO.	-
488.3	88	1074.75	100	45.04	10400	100.000	No. on Address	817
1000	14	100.00	100	10.00	10 STATE	100.705	and includes	-





Data Modeling



<u>Research</u>



Experimented with different modelling 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.

Sponsors: Chad Koehnke Karen Biesiadka Toni Crowell

Team: Divya Yadav, MSIM-Full time Laxmi Patil, MSIM-Full time Pratima Malviya, MSIM-Full time



Analyzed and determined relevant input variables affecting the output variable.

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

Error Analysis



Investigated for 'Outliers' and 'Influential Points' in the Training Data Set for inaccuracy, if any, in predicted output values.

Calculated 'Leverage' to measure the amount by which the predicted value would change if the observation was shifted one unit in the 'y-direction'.

Calculated 'Cook's Distance'to check for any influential points.

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.





Information School UNIVERSITY of WASHINGTON