Jetblue Experienced the Greatest Odds of Delays from its Departing Flights at Sacramento International Airport

Jessica Cerda

2023-05-06

Abstract After running tests and interpretations on Flight Delay Data from the Bureau of Transportation, it was discovered that the global pandemic during these years caused an increase in flight delays. It was also found JetBlue and Southwest airlines had the greatest odds of being delayed in the years 2019,2021, and 2022. Those who flew Skywest and Delta airlines during this time span faced the lowest odds of having a flight delay.

Model Description

This model analyzes the data set *SMFflights* and explores the probability a flight is delayed 15 minutes or longer when departing from Sacramento International.

Each row in the column set represents a different departing flight from Sacramento International Airport. The data set includes 9862 observations of 7 different variables

For this data analysis focuses on the following variables within the data set

- `DEP_DEL15` Whether or not a flight was departing over 15 minutes flight
- `YEAR` The year each flight departed from Sacramento International Airport
- `Airlines` A categorical variable with containing each airline that had departing flights from Sacramento International Airport

Below is a table containing the first few rows within the dataset

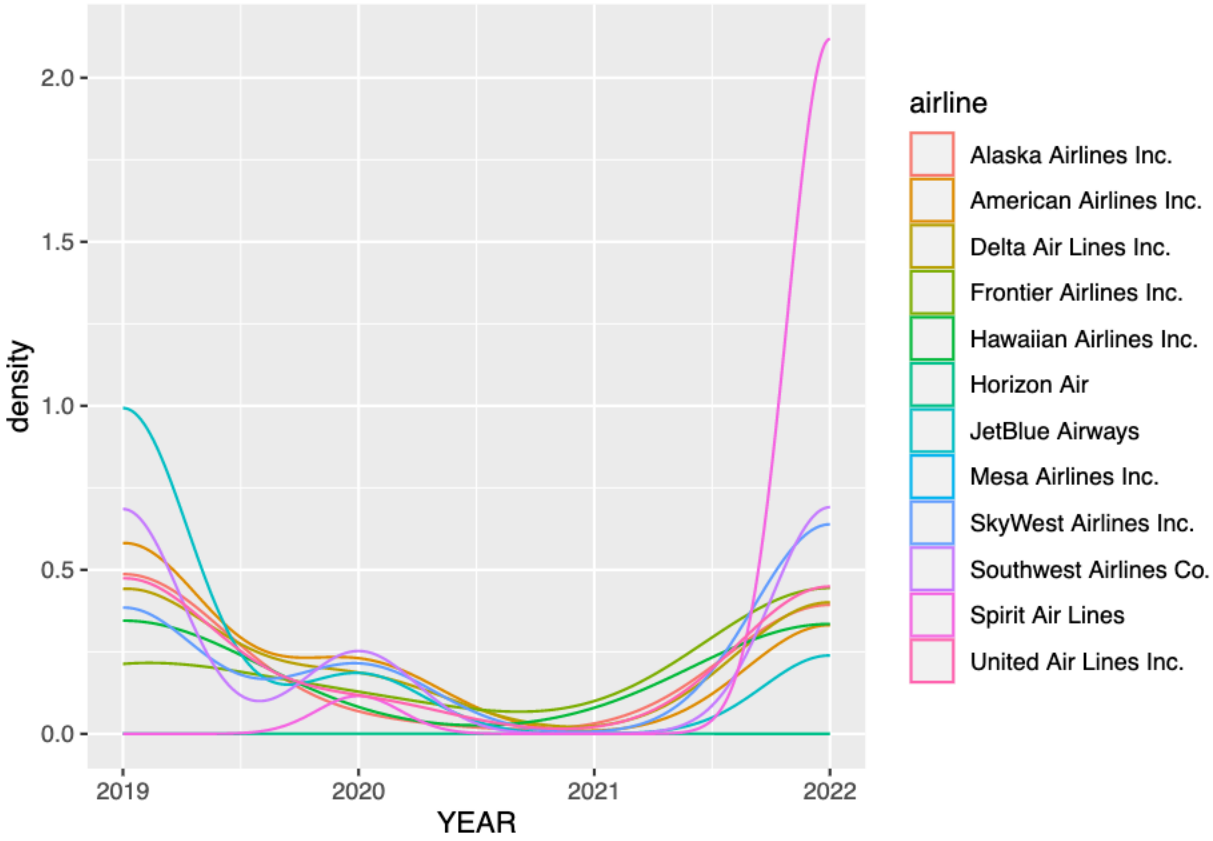
```
##       OP_UNIQUE_CARRIER YEAR               FL_DATE ORIGIN DEST DEP_DELAY_NEW
## 6243                 WN 2020 4/26/2020 12:00:00 AM    SMF  PHX             0
## 9349                 WN 2022 4/12/2022 12:00:00 AM    SMF  BUR            33
## 5903                 WN 2020 4/23/2020 12:00:00 AM    SMF  ONT             0
## 5331                 WN 2022  4/8/2022 12:00:00 AM    SMF  LAS             0
## 183                  AA 2022  4/4/2022 12:00:00 AM    SMF  PHX            29
## 10376                WN 2019 4/23/2019 12:00:00 AM    SMF  SEA             0
##       DEP_DEL15                airline
## 6243          0 Southwest Airlines Co.
## 9349          1 Southwest Airlines Co.
## 5903          0 Southwest Airlines Co.
## 5331          0 Southwest Airlines Co.
## 183           1 American Airlines Inc.
## 10376         0 Southwest Airlines Co.
```

Included below is a visual representation of the logistic regression model used for the exploratory data analysis

$$y = \beta_0 + \beta_{1year}x_{i1} + \beta_{AmericanAirlines}x_{i2} + \beta_{3Southwest}x_{i3} + ... + \beta_{9UnitedAirlines}x_{i9}$$

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



The above graph gives a visual comparison of the amount of departing flights occurring from Sacramento International Airport during April of 2019-2022.

- Not Pictured is Year 2021

Coefficient Interpretation

The coefficients within the model *delaymod* made the following conclusions. The following variables increased the odds a flight was delayed 15 minutes or longer before the flight departs:

- `YEAR` increased odds by .28697 for each year added on
- `American Airlines` increases odds by .42433
- `Frontier Airlines` increases odds by .16199
- `Hawaiian Airlines` increases odds by .44993
- `Horizon Air` increases odds by .04421
- `JetBlue Airways` increases odds by .93049
- `Southwest Airlines` increases odds by .59748
- `Spirit Airlines` increases odds by .38376
- `United Airlines` increases odds by .25076

The following coefficients decreases odds that flights are delayed 15 minutes or longer before the flight departs:

- `Delta Air Lines` decreases odds by .15863
- `Mesa Airlines` decreases odds by 9.86362
- `SkyWest Airlines` decreases odds by .45159

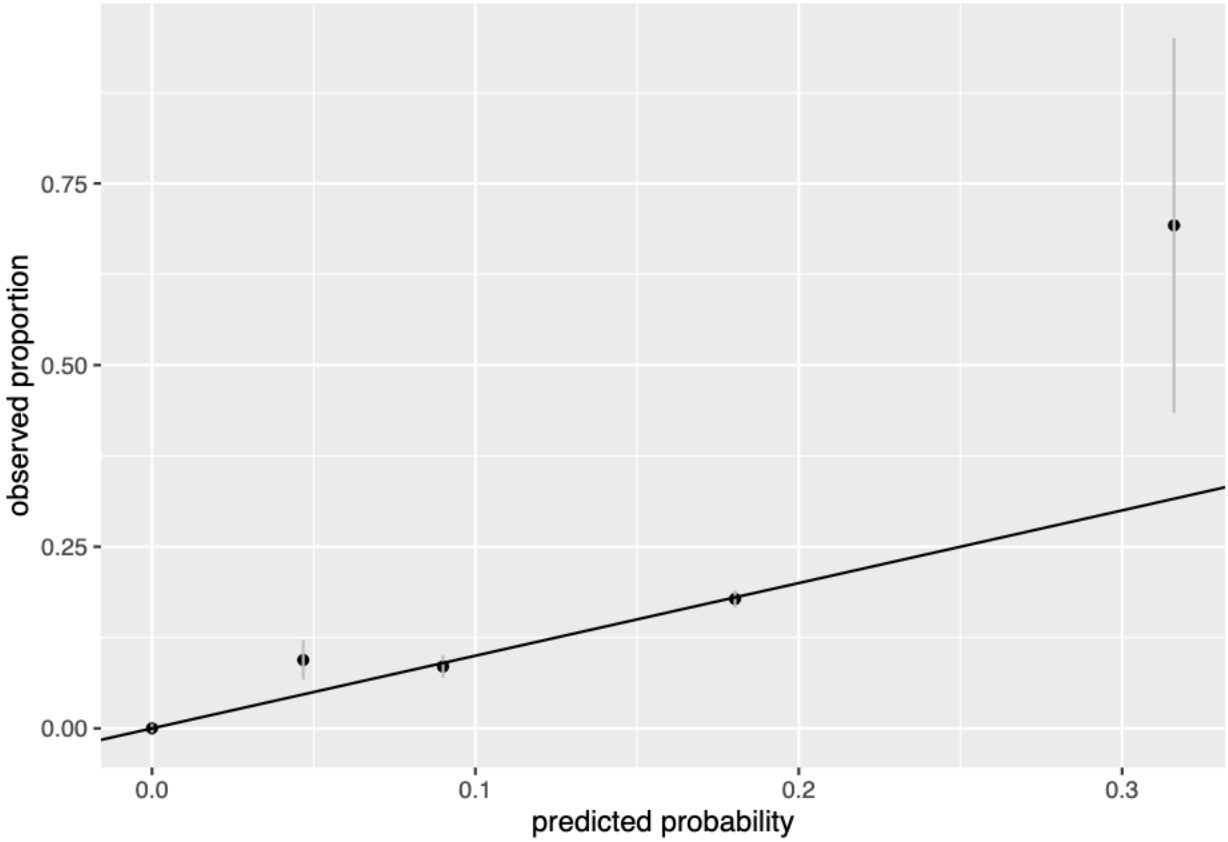
Odds Confidence Interval

```
## Waiting for profiling to be done...
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##                                       2.5 %        97.5 %
## (Intercept)                    3.210388e-300 4.089554e-207
## YEAR                            5.582726e-01  5.842490e-01
## airlineAmerican Airlines Inc.   4.851111e-01  7.170431e-01
## airlineDelta Air Lines Inc.     3.268614e-01  6.002671e-01
## airlineFrontier Airlines Inc.   3.205745e-01  7.270854e-01
## airlineHawaiian Airlines Inc.   4.126814e-01  7.679474e-01
## airlineHorizon Air              2.948636e-01  7.040456e-01
## airlineJetBlue Airways          5.528230e-01  8.348041e-01
## airlineMesa Airlines Inc.                 NA  1.000000e+00
## airlineSkyWest Airlines Inc.    2.845993e-01  5.105033e-01
## airlineSouthwest Airlines Co.   5.506736e-01  7.370719e-01
## airlineSpirit Air Lines         4.013852e-01  7.543598e-01
## airlineUnited Air Lines Inc.    4.405638e-01  6.814599e-01
```

A 95% confidence interval was constructed to determine the probability of flight delay for each airline carrier when departing from Sacramento International Airport. Southwest and Jetblue Airlines had confidence intervals constructed beyond .50 probability, meaning with 95% confidence these flights had at least a 50% chance of being delayed. Majority of the confidence intervals ranged between (0.3,0.7). Mesa Airlines only had one flight within the data set, explaining the large confidence interval.

Residuals

```
## Warning in geom_abline(intercept = 0, slope. = 1): Ignoring unknown parameters:
## `slope.`
```



After analyzing the binned residual plot, there is little indication of significant outliers indicating that the model is not ill fitting for predicting the probability of departing flights being delayed at least 15 minutes at Sacramento International Airport.

Hypothesis Testing

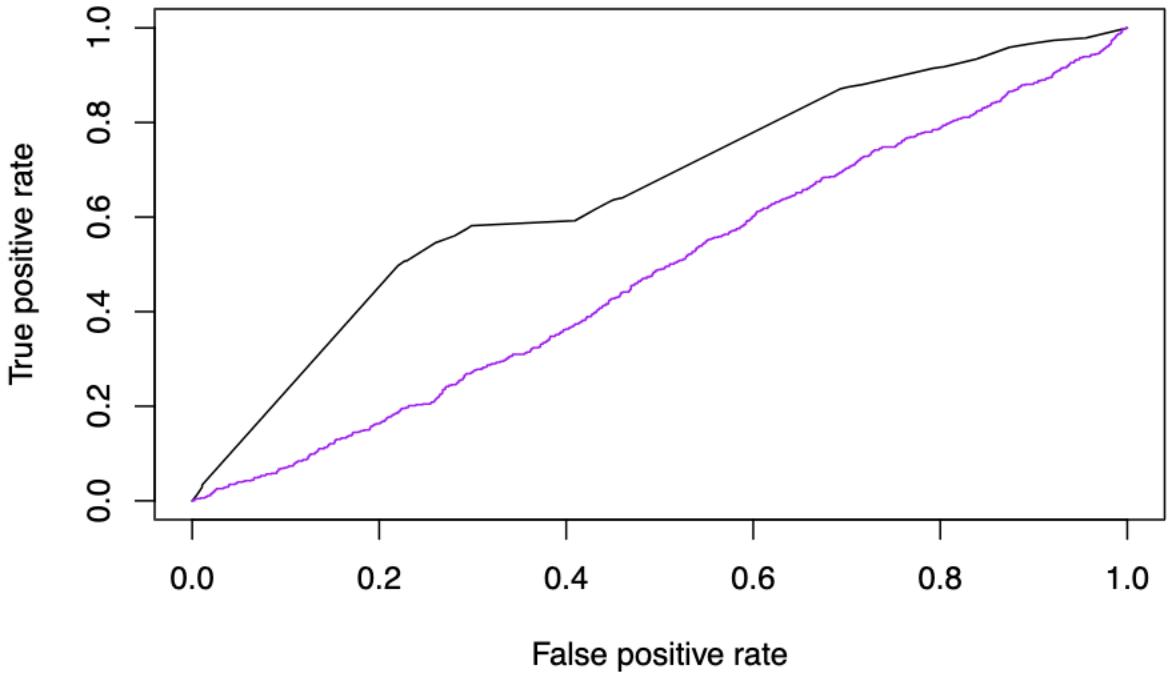
```
##    2.5 %   97.5 %
## 1.597291 1.974828
```

```
##
## Call:
## glm(formula = DEP_DEL15 ~ YEAR + airline, family = binomial,
##     data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8717  -0.5905  -0.5120  -0.3853   2.4750
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -581.95731   54.67567 -10.644  < 2e-16 ***
## YEAR                            0.28697    0.02706  10.607  < 2e-16 ***
## airlineAmerican Airlines Inc.   0.42433    0.25150   1.687  0.09157 .
## airlineDelta Air Lines Inc.    -0.15863    0.28669  -0.553  0.58006
## airlineFrontier Airlines Inc.   0.16199    0.43636   0.371  0.71047
## airlineHawaiian Airlines Inc.   0.44993    0.39201   1.148  0.25107
## airlineHorizon Air              0.04421    0.43826   0.101  0.91966
## airlineJetBlue Airways          0.93049    0.35698   2.607  0.00915 **
```

```
## airlineMesa Airlines Inc.       -9.86362  196.96779  -0.050  0.96006
## airlineSkyWest Airlines Inc.    -0.45159    0.24495  -1.844  0.06524 .
## airlineSouthwest Airlines Co.    0.59748    0.21027   2.841  0.00449 **
## airlineSpirit Air Lines          0.38376    0.38525   0.996  0.31918
## airlineUnited Air Lines Inc.     0.25076    0.25401   0.987  0.32356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5096.6  on 5916  degrees of freedom
## Residual deviance: 4903.2  on 5904  degrees of freedom
## AIC: 4929.2
##
## Number of Fisher Scoring iterations: 10
```

- Based on the confidence interval above, with 95% confidence, flying in 2 years increases the odds of a delay occurring from nearly 60% to 97%
- Upon conducting a Step wise model selection, it was found that all variables were kept signifying that the model is ideal for predicting delays in departing flights from Sacramento International

ROC Curve



The Receiver Operating Characteristics (ROC) views performance of a model for all possible cut off values. Based on the ROC curve, there is no specific cut off value that would make the model work significantly better for predictions. This is normal given the multiple factors in flight delays leaving it arduous to accurately predict flight delays. Outside of the ROC curve, the model still provides insights about which airlines get affected the most.

Conclusion After analyzing flight data from the past few years, there is statistically significant indications that the probability a flight experiencing a delay 15 minutes or longer has increased over time. This can be attributed to the global pandemic when many flights were delayed or cancelled due to flight restrictions and fewer staff members. Apart from this, passengers flying Jet blue have the greatest probability of experiencing

a delay of 15 minutes or longer, while those flying Delta had the lower probability. With the global pandemic ending, flight schedules are going back to Pre-Covid regulations, however with staffing issues and predictive weather, delays cannot be completely eliminated but some solutions can provide relief to delays. These solutions can include hiring additional staff and watching weather patterns and changing flight schedules accordingly.