Effects of Weather Events on Flight Delays and Cancellations

BARRETT BRISTER

“Your flight has been cancelled.”

- These may be the five most hated words for an airline passenger.
- They disrupt travel plans, cost everyone time and money, and can hurt an airline's reputation.
- However, sometimes flight cancellations and delays are unavoidable.
- What if we could create an algorithm that lets an airline forecast whether to cancel or delay a flight well ahead of time, minimizing cost to the airline and disruption to customers?

Data

A. Sources
   1. Flight data: Bureau of Transportation Statistics website. All domestic commercial flights, 2016-2019, were collected (over 25 million flights).
   2. Weather data: Kaggle data, which used the work from Moosavi, et al. (2019).

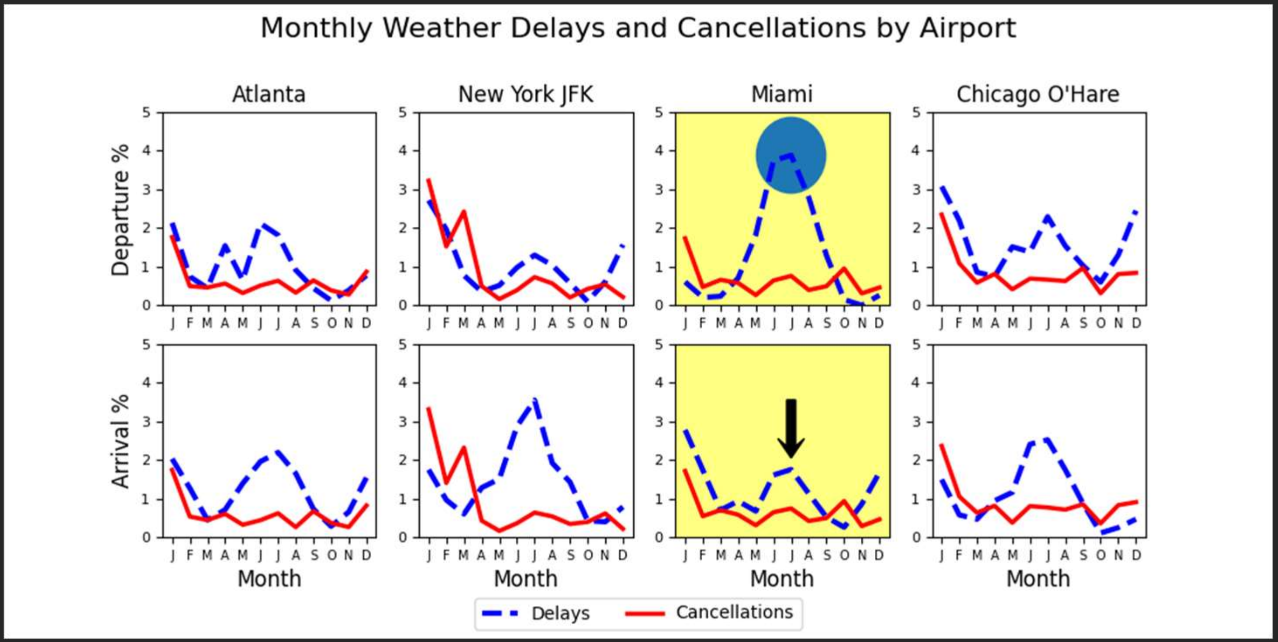
B. Data wrangling
   1. Four major airports were selected as a proof of concept due to flight volume and variety of weather: Atlanta, Chicago-O'Hare, Miami, and New York-JFK.
   2. Weather data were listed as codes, not periodic observations, with inconsistent start/stop times; filtering this took considerable work.

Exploratory Data Analysis

- After merging the weather and flight data, all flights between our four airports were analyzed for weather events and intensity.
- Analysis was conducted for both flight arrivals and flight departures.
- For the purposes of analysis, departure and arrival weather events were considered for their effects on cancellations and delays separately.

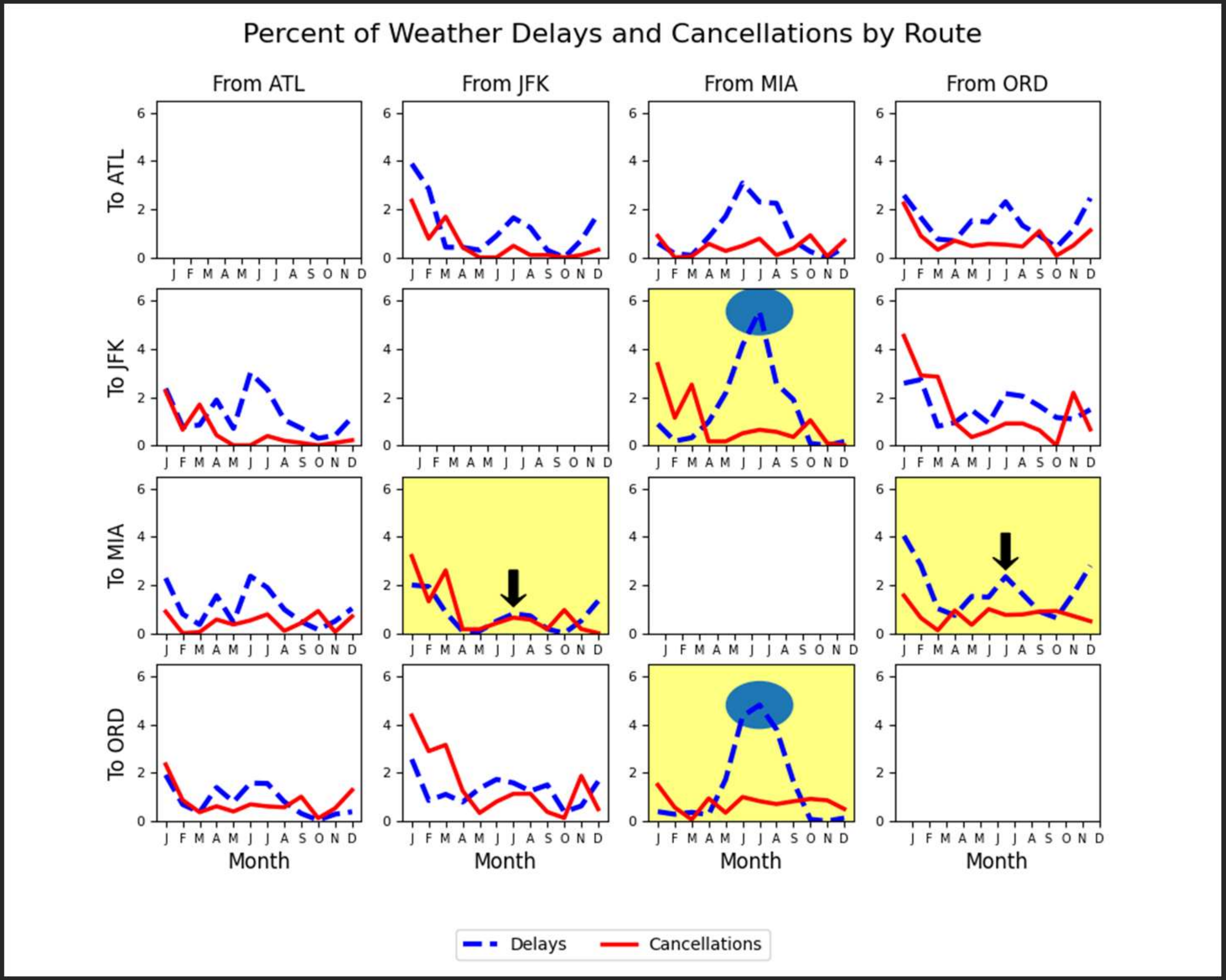
Monthly Disruptions per Airport

- Percentage of flights delayed and cancelled between our four airports by month.
- Note the increase of summertime delays from Miami. These are likely due to Florida's frequent summertime thunderstorms.
- Flights to and from Miami, Sept. 9 – 13, 2017, are removed due to the disruption by Hurricane Irma.



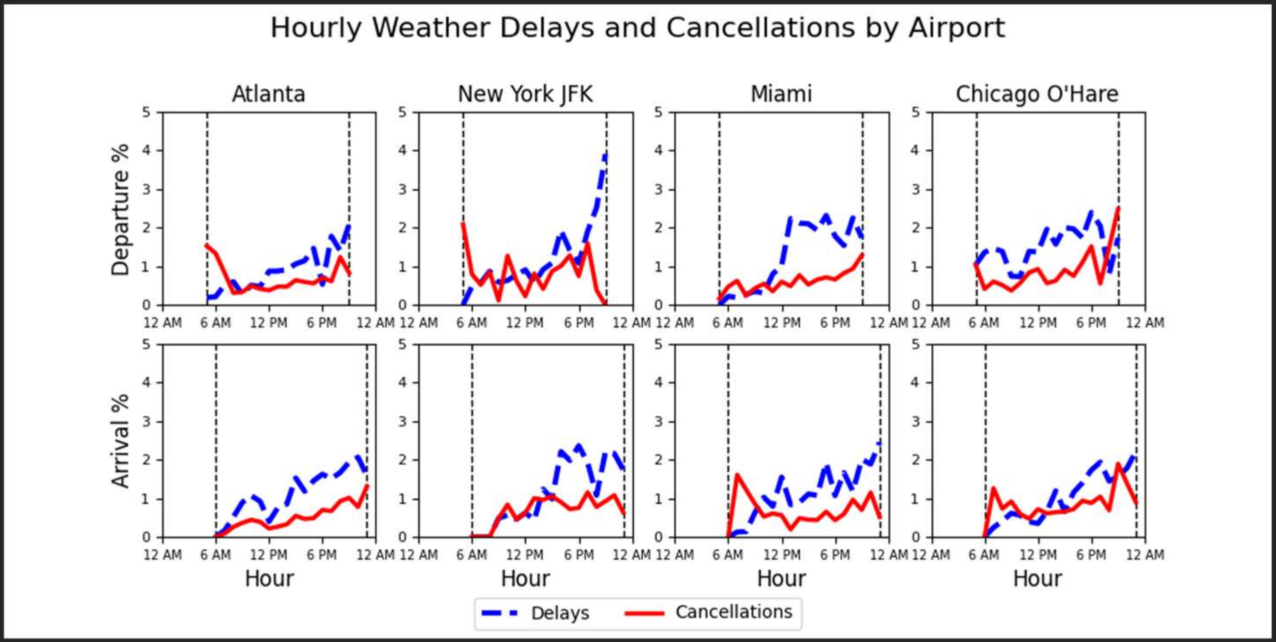
Monthly Disruptions per Route

- The trend of Miami's summertime departure delays continues when the flights are broken down by route.
- The effect is less pronounced for flights to and from Atlanta.



Hourly Disruptions per Airport

- Cancellations are spread throughout the day, while delays are more common after noon.
- JFK Airport has very few departures to ATL, MIA, and ORD after 8 PM, so its evening departure disruptions should be taken with a grain of salt.

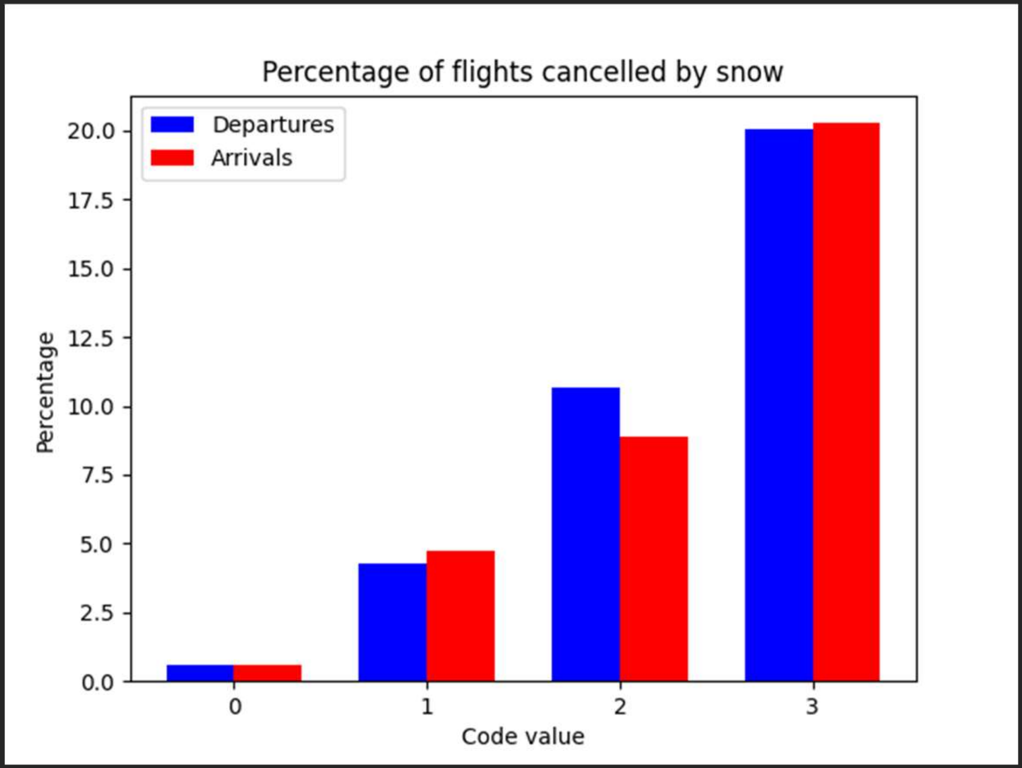


Exploratory Data Analysis, Continued

- With over 3000 snow events and 40,000 rain events affecting departing flights, rain and snow are the two most likely weather candidates for causing cancellations and delays.
- The following graphs shows the following probabilities based on the level of the weather code:
  1. Cancellations due to snow
  2. Delays due to snow
  3. Delays due to rain
- Cancellations due to rain are not significant.
- Rain and snow codes were listed in the Kaggle data as from level 0 (none) to 3 (intense).

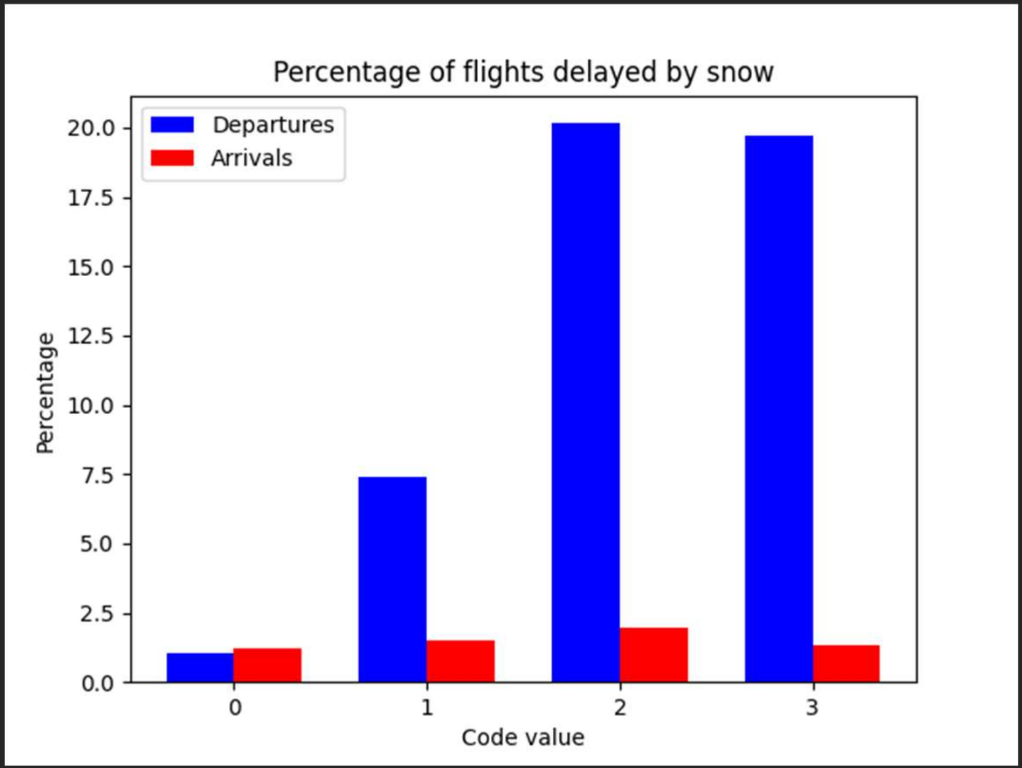
Flights Cancelled by Snow

- The more intense the snow is, the more likely a flight is to be cancelled.
- This effect is seen at both departure and arrival airports.



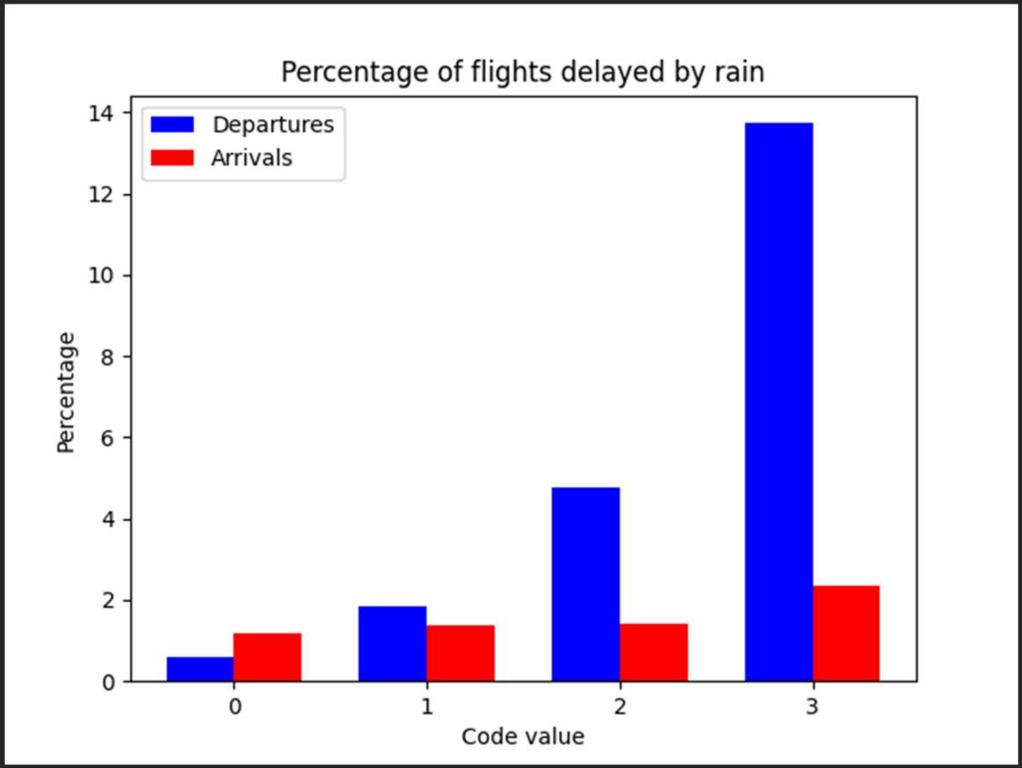
Flights Delayed by Snow

- Far fewer arrivals are delayed by snow than departures are.
- This is likely due to the additional preparation that airplanes need, such as de-icing.
- Many level-3 snow events cause cancellations, hence fewer delays.
- Quadratic variables may be needed to model that level-3 drop.



Flights Delayed by Rain

- As with delays from snow, rain delays far more departures than arrivals.



Machine Learning

A. Variables
   1. *X* variables: Values of the significant weather codes as described earlier. Some codes such as snow departure delays are also encoded as their values squared.
   2. *y* variable: Whether the flight was cancelled/delayed. These are two separate questions, so the training will treat these as two separate problems.

B. Learning methods attempted
   1. Logistic regression: Well-suited for systems with binary outputs such as ours.
   2. Random forests: Good as an out-of-bag, first attempt at training.
   3. Support vector machines: Can be effective, but slow.

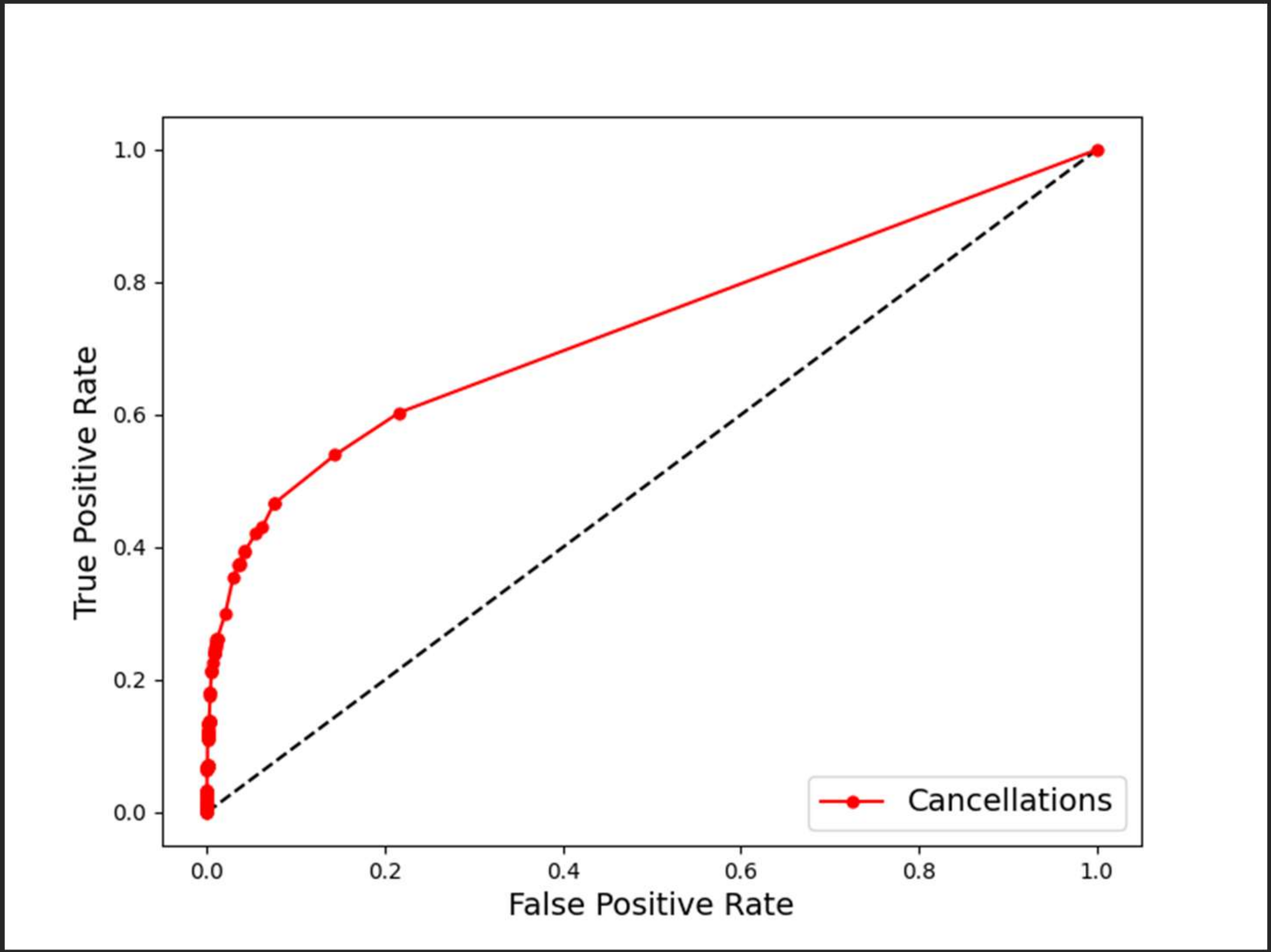
C. All models were trained with 5-fold cross-validation.

Efficacy of the Models

- Of the models tested, logistic regression (C=1) fared the best. Mean test scores: 0.729 for cancellations and 0.646 for delays.
- Logistic regressions work using thresholds: If the predicted probability of y=1 (positive) is greater than the threshold, the model will classify the prediction as positive.
  1. Default threshold: 0.5
  2. We can adjust this between 0 and 1 to see its effect on the model.
- Measures evaluated
  1. True positive rate (TPR): P(positive prediction | actual positive)
  2. False positive rate (FPR): P(positive prediction | actual negative)

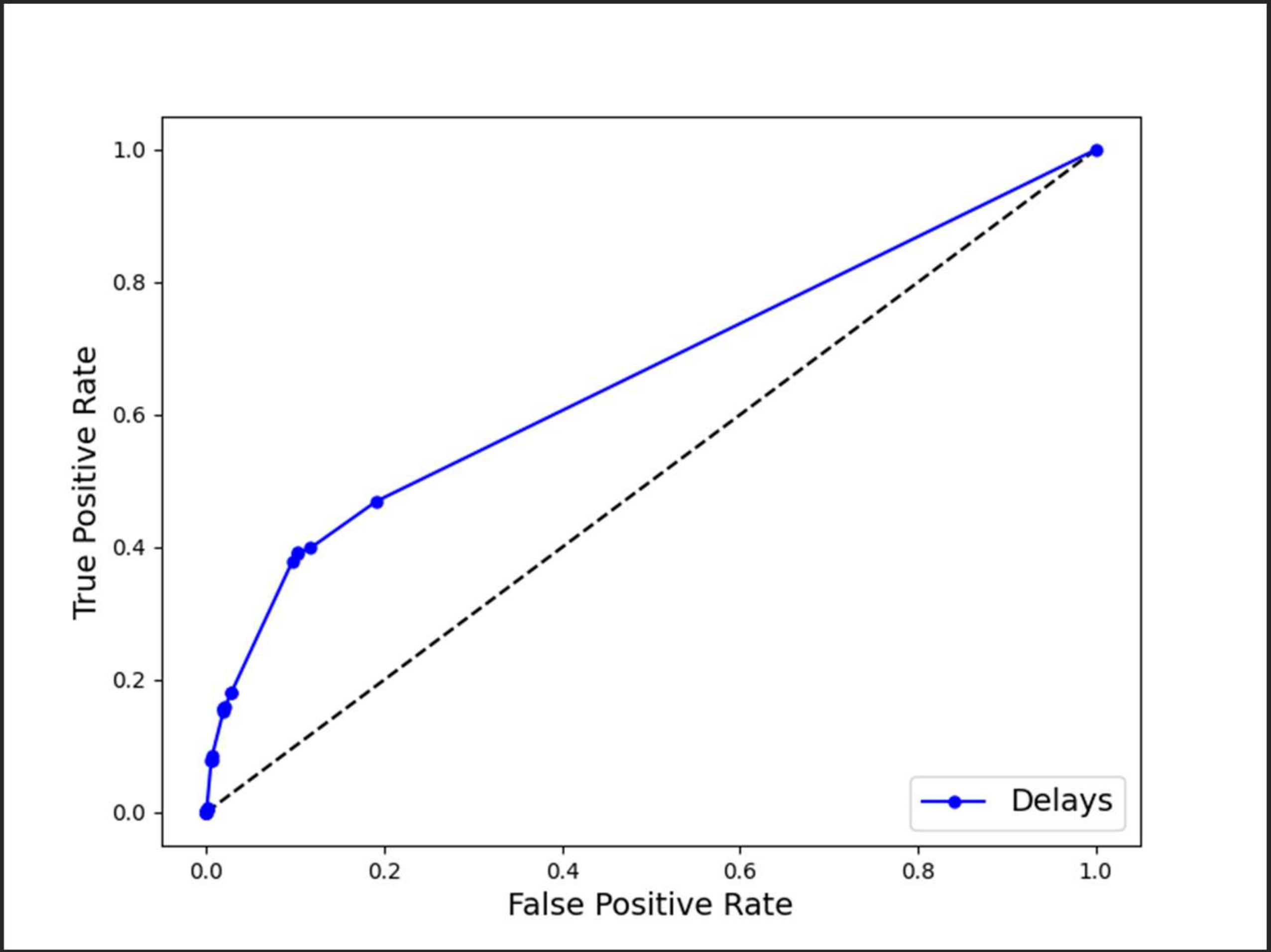
ROC Curve for Cancellations

- The percentage of flights that are cancelled is small, leading to most of the sample data being negative (this was weighted out during training).
- Most of the true positive rates are not very high, but the false positives can be easily controlled.



ROC Curve for Delays

- The percentage of flights that are delayed is small but larger than that of delays.
- The area under the curve (AUC) of this plot is clearly less than that of cancellations, implying that this model provides a worse fit than the cancellation model.

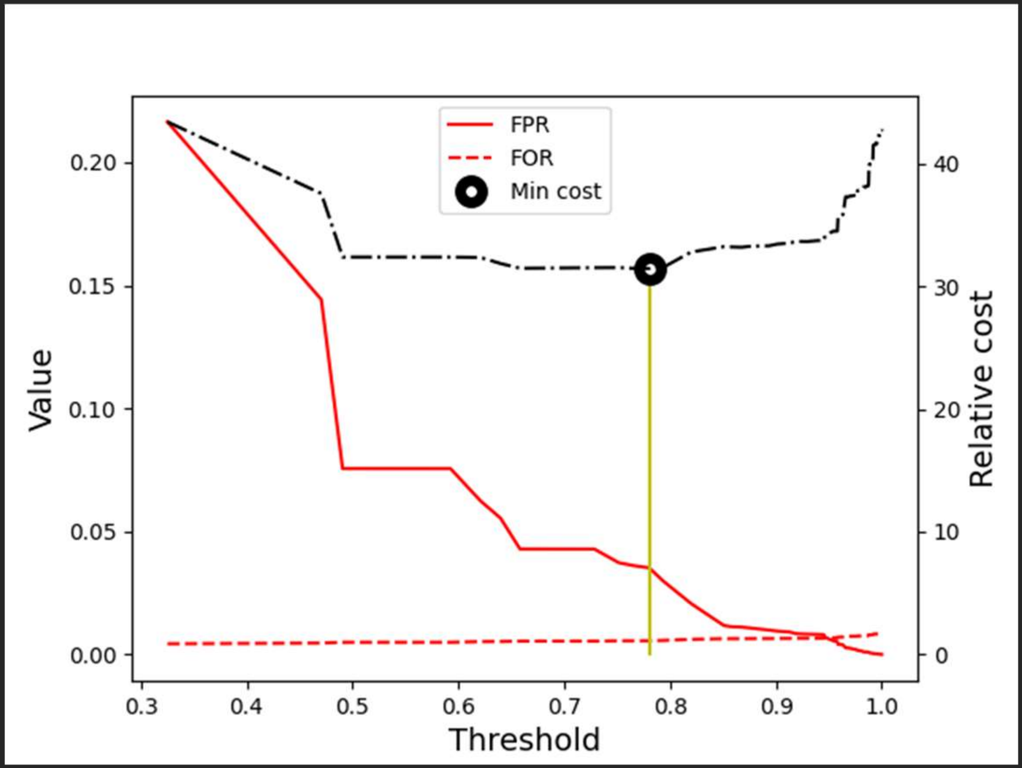


Considerations of the ROC Curve

- Those ROC curves suggest that false positives and false negatives are equally bad, suggesting that we just need to minimize the incorrect predictions.
- In practice, a false negative is a much more serious situation, especially for cancellations. A false positive means cancelling a flight well in advance that should have flown, and a false negative means not cancelling the flight until the last minute. The latter can seriously disrupt customers' travel plans and entice them to never fly with that airline again.
- We make two changes to our analysis.
  1. Replace the TPR with the False Omission Rate (FOR): P(actual positive | negative prediction)
  2. Assume that C(FOR) = 50*C(FPR) for cancellations and C(FOR) = 10*C(FPR) for delays. These are sample ratios to be used as a proof of concept and not final analyses.

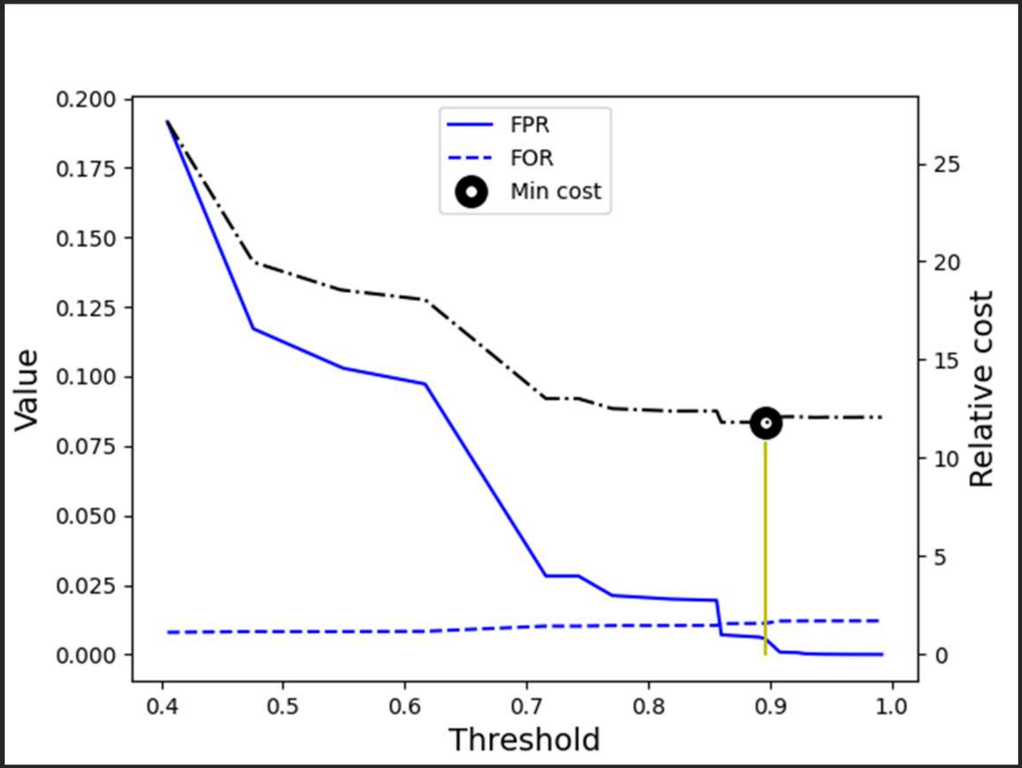
Cost Curve for Cancellations

- Given our assumptions, setting the threshold to 0.78 minimizes the cost from FPRs and FORs.
- The baseline is a threshold of 1. Under our assumptions, we can achieve a 27% expected cost savings over the baseline.
- Precision: 0.0056
- Sensitivity: 0.37



Cost Curve for Delays

- Here, the cost savings are minimal. This probably comes from our assumption of a lower C(FOR)-to-C(FPR) ratio.
- The baseline is a threshold of 1. Under our assumptions, we can only achieve a 2.5% expected cost savings over the baseline.
- Precision: 0.011
- Sensitivity: 0.078



Conclusions

- The 0.373 sensitivity of the cancellations should be taken as a glass-half-full situation. This means that over a third of the times that we need to cancel a flight, this model will correctly flag that flight as needing to be cancelled.
- Most flights are neither cancelled nor delayed by weather. This means that although the false positive rate is high, the percentage of flights that would be false positives would be low.
- Recall, the impact of a false negative far exceeds that of a false positive, so the airline should accept some false positives in order to reduce the risk of false negatives.

Recommendations and Future Work

A. Recommendations
   1. Use a logistic regression with a threshold of 0.78 to determine whether a flight should be cancelled in advance.
   2. Using the model to make the same predictions for delays is not recommended at this time.

B. Future work
   1. Improve the model by finding and using actual, periodic weather observation data
   2. Obtain the actual FOR-to-FPR cost ratio
   3. Consider applying the model to longer events of disruptions, whether days or months, such as natural disasters and pandemics