What's My Flight Status?

Using Flight Data to Predict Flight Delays

By: Andre Layton

Overview:

Objectives:

- ➔ Using machine learning and deep learning algorithms to build a model that will predict flight delays.
- ➔ A flight delay = departing/arriving 15 minutes past the scheduled time. I will focus on **departure delays**.

Process:

- ➔ Employ sampling techniques and advanced classification algorithms to enhance precision.

Results:

- ➔ The machine learning model serves as my best algorithm for predicting flight delays, yielding 27.2% precision.

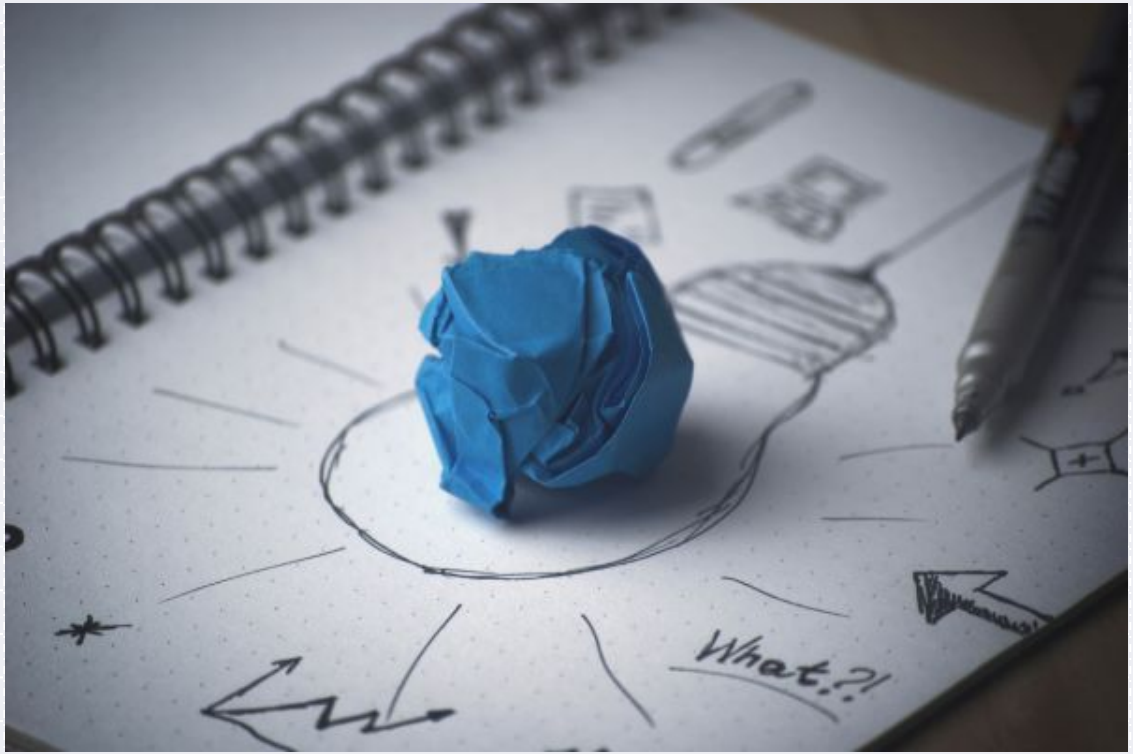
Outline:

01 Business Problem

02 Data & Methods

03 Modeling

04 Conclusions

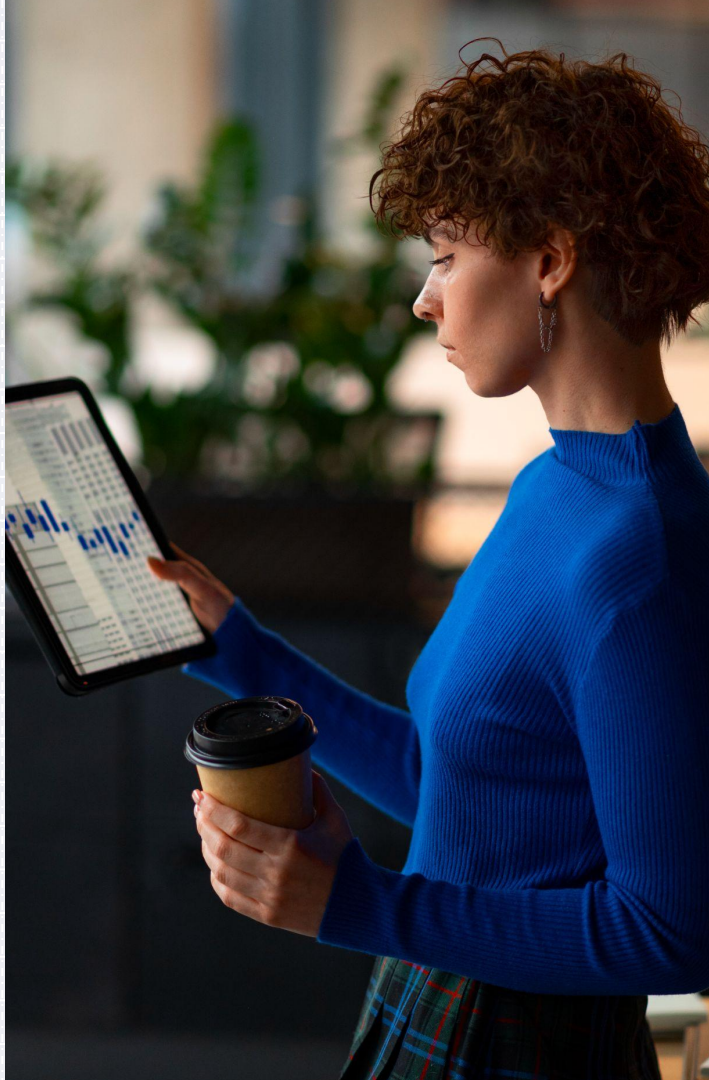


Business Problem:

United Airlines is looking to regain consumer confidence by addressing their flight delays.

- ➔ United Airlines loses approximately $102/minute when facing delays. (Schonland, 2023)

- ➔ Use both machine learning and deep learning to predict delays with highest precision.

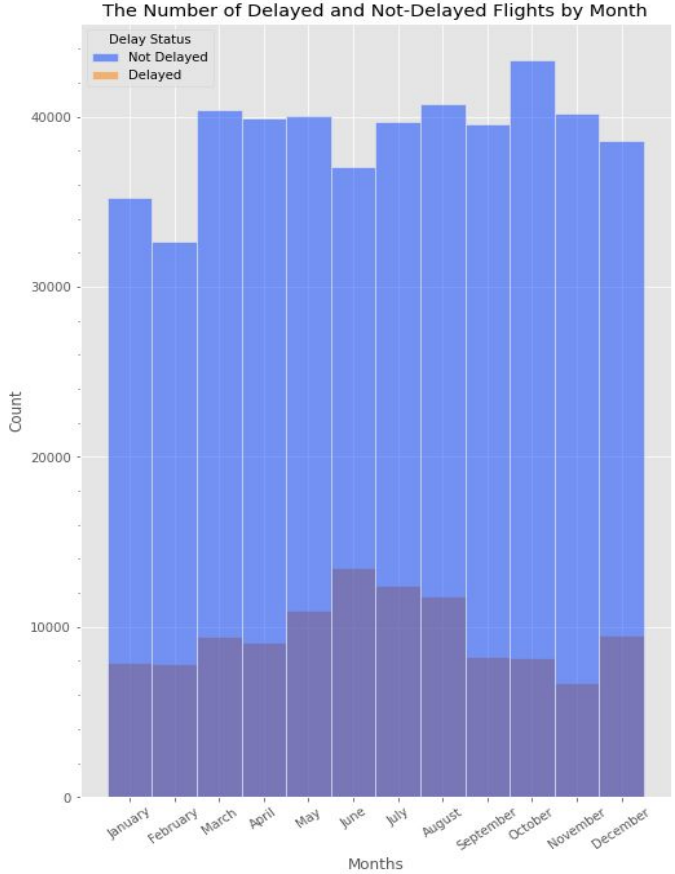


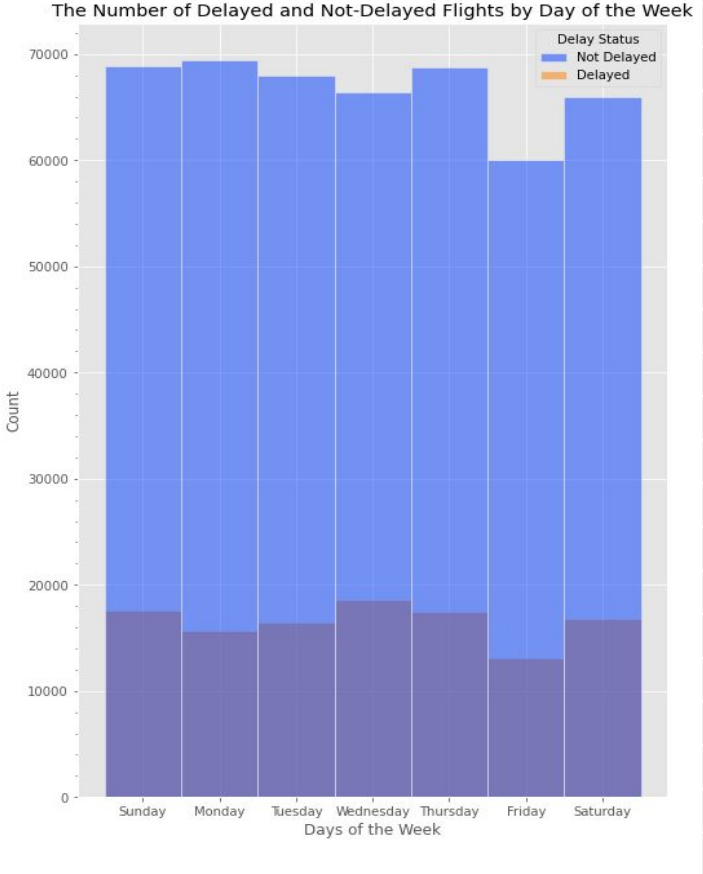
Data & Methods:

The Data:

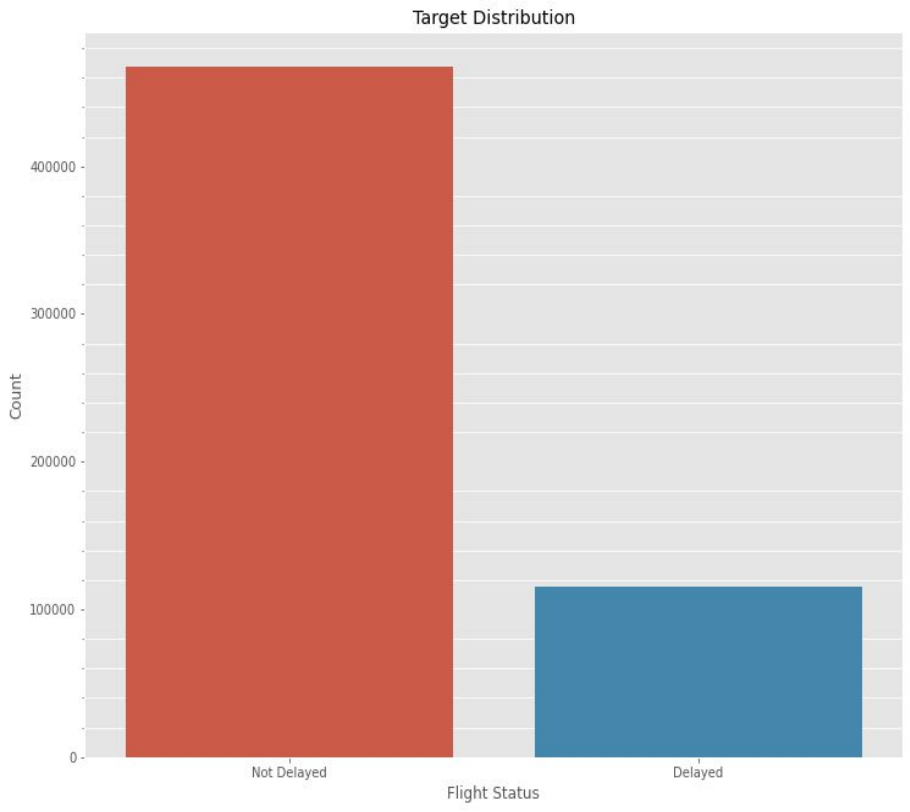
- Contains airline, weather, and airport information from 2019

- Here are two features and their distributions, classified by flight status.





Data & Methods (cont.):



Class Distribution:

- **80. 18%** of United flights were on-time
- **19.82%** of United flights were delayed

Methods:

- Before modeling, I balance the data's target distribution by oversampling.

- Final dataset: **701,026** United flights analyzed (gained 263,822 observations from sampling).

Modeling:

Evaluation:

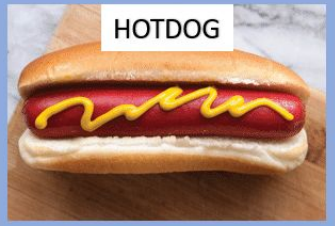
- ➔ Models were evaluated on precision and the number of false positive predictions, in addition to the number of correct delay predictions.

True negative

False positive

False negative

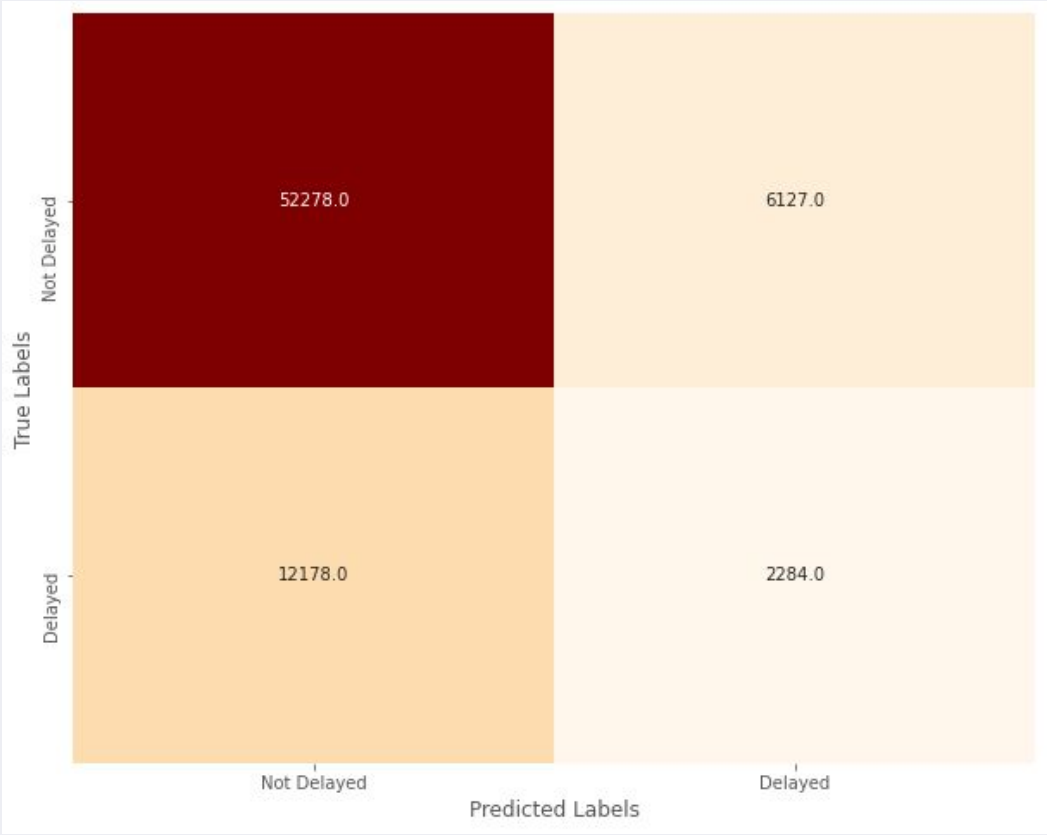
True positive



Modeling Results:

- ➔ The ML algorithm (my final model) is 27.2% precise.
  - ➔ Baseline model (Random Forests) was 58.0% precise.

- ➔ Out of 72,867 predictions, the model had 6,127 false positive cases (the baseline had 1,505).



Conclusions:

1. The machine learning algorithm performed best.

2. The model is 27.2% precise when testing and classifying flights as delayed.

Limitations & Further Work:

- ➔ Computational constraints
- ➔ Imbalanced classes in data
- ➔ Apply deeper ML techniques given improved hardware
- ➔ Gain external feedback and perspectives

Works Cited:

Schonland, A. (2023, February 13). *The value of time for an airline | AirInsight*.

https://airinsight.com/the-value-of-time-for-an-airline/

Thank You!

Do you have any questions? Contact me!

Email: alaygt6@gmail.com
GitHub: @therookiescientist-andre
LinkedIn: linkedin.com/in/ak-layton/