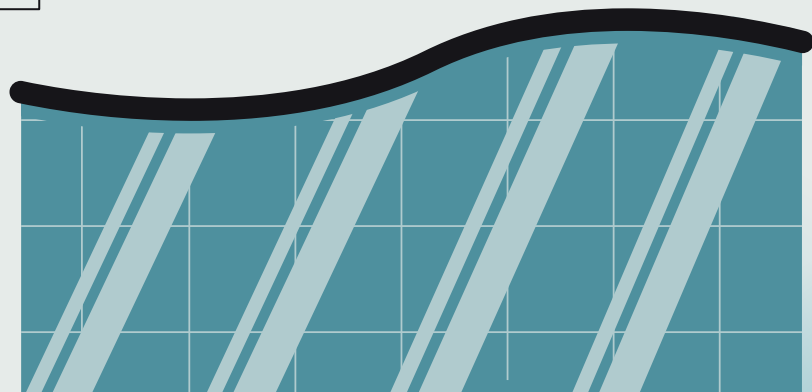# Detecting Severe Air Travel Delays with Machine Learning

By Robert Harrow

01
Business Problem
What we set out to solve

02
Data Methods
How we solved it

03
Product Demo
Presenting our MVP

04
Next Steps
What happens next

01

Business Problem

Getting Ahead of Delays

Delays Are a Major Headache

"In the first five months of 2022, **more than 1 in 5 domestic flights were delayed**" - ValuePenguin

"Almost 79% of 2022 travelers experienced problems during a trip[...] **During the July 4 holiday weekend, airlines reported thousands of canceled and delayed flights**" - CBS News

"**61% of Summer Travelers Had Flight Delayed** or Canceled, 83% Of Those Lost Money" - Forbes

Flight Chicken wants to alleviate the worst problems

Flight Chicken wants to build a **service that warns travelers if an upcoming flight runs the risk of being severely delayed** by 1 hour or more.

Their hope is that giving travelers this heads up allows them to adjust plans and prepare.

As such, our job is to:

- **Build a model** that detects delays of 1 hour or more
- **Productionize** the model so a non-technical stakeholder can use it
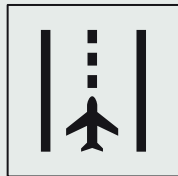- **Set goals** for future product iterations

02

DATA METHODS

Data & Methods

7M Domestic Flights

We analyzed and cleaned over **7M flight records** between **June 2021 and June 2022**
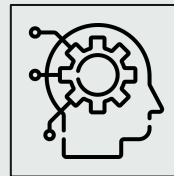
Source: Bureau of Transportation Statistics

Weather API

We leveraged a weather API to pull in detailed daily **weather data on over 124 locations** over the last year

Source: weatherAPI.com

Machine Learning

We then trained **machine learning algorithms** on our data to **detect severe flight delays**.

Success Metrics

F1 Score

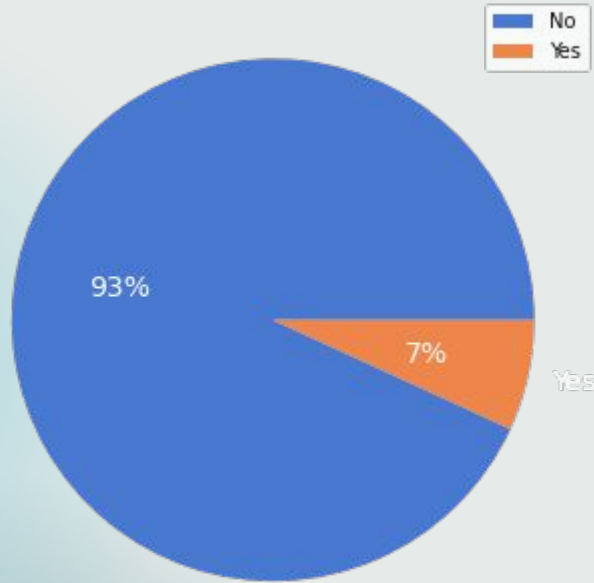Balance between precision (no false positives) and recall (tolerating some false positives)

Accuracy

While F1 is the primary metric, we can use Accuracy as a tiebreaker

Though bad, severe delays are infrequent

Is there a severe delay?

■	No
■	Yes

93%

7%

Yes

- Flight Chicken is going after the most severe delays first though they happen just **7% of the time**
- Future versions of the product can explore other types of delays

The Relationship Between Delays and Different Factors is Complicated

Percent of Flights with Severe Delays Throughout the Week - Atlanta

# What that same plot looks like for Los Angeles International Airport

Percent of Flights with Severe Delays Throughout the Week - LAX
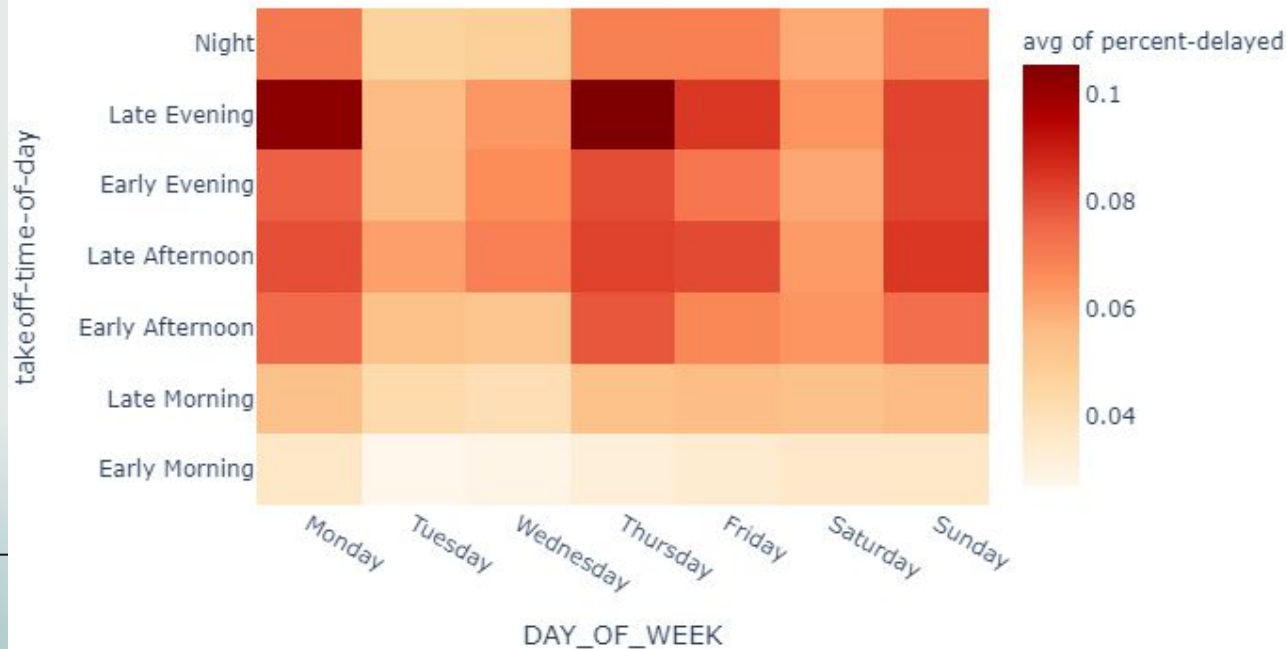
XGBoost was our best model

Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.645	0.121	0.655	0.204
Random Forest	0.625	0.111	0.631	0.189
🏆 XGBoost	0.72	0.137	0.622	0.24
LightGBM	0.697	0.136	0.638	0.224

Final Model Evaluation on Unseen Data
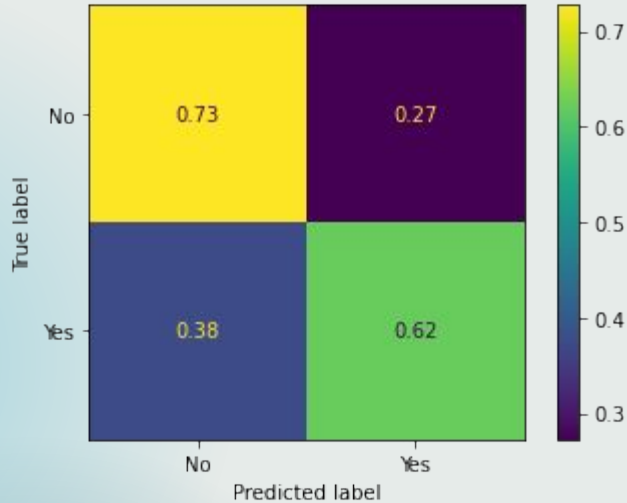
Our final model is able to **predict a severe delay in 62% of cases**

- Accuracy: 0.72
- Precision: 0.14
- Recall: 0.62
- F1 score: 0.24

03

PRODUCT

And Demonstration

Product *overview*
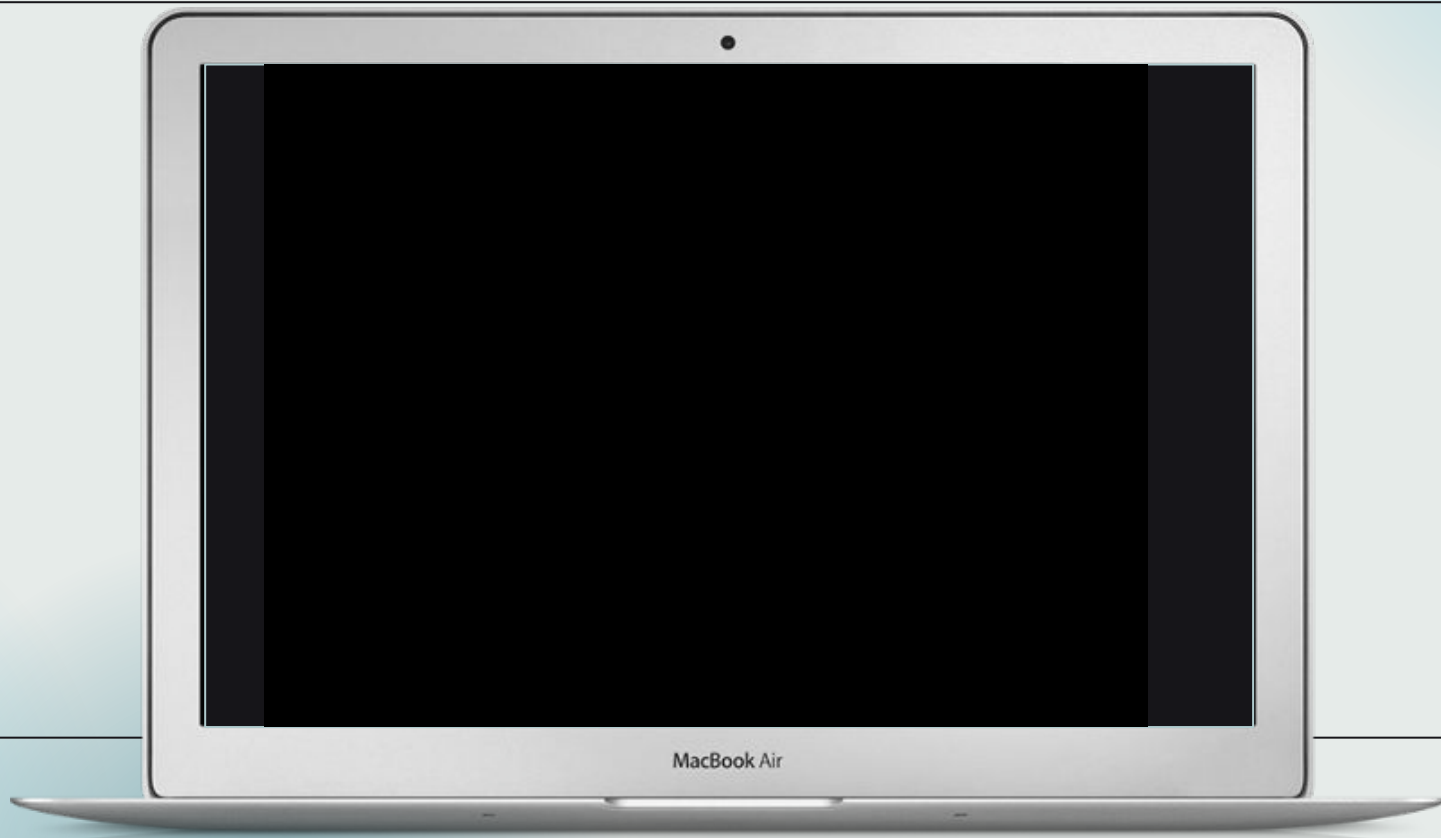
Flight Predictions

Get prediction about whether your flight will experience a severe delay

Airport Dashboard

See when delays are most common at your home airport

DEMO

04

NEXT STEPS

& Conclusions

CONCLUSIONS & NEXT STEPS

- We built a model that **detects major delays 64%** of the time
- The model is currently **fully usable by non-technical users**
- Future updates to product may include:
  - Training model on more data
  - Exploring other types of delays (sub 1-hour)
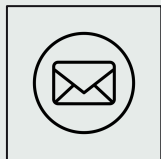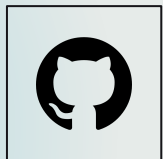 - Expand coverage to more airlines & airports

Thanks

Do you have any questions?

rharrow928@gmail.com
718 483 1580
robertharrow.com