DECODING DELTA

A Comprehensive Exploration of Flight Delays and Cancellations

Iris Brook, Lucy Liu, Krishanu Datta, Devin Wasilefsky

AGENDA

01 **INTRODUCTION**
Importance of flight disruptions and problem statement

02 **METHODOLOGY**
Dataset overview and models tested

03 **RECCOMENDATIONS**
What actions should Delta Air Lines take in the future

04 **CONCLUSION**
Recap of project and next steps

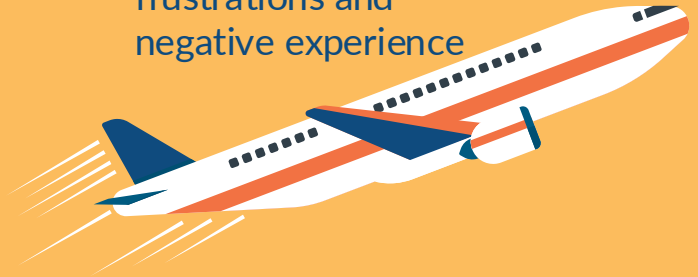
SIGNIFICANCE OF FLIGHT DELAYS & CANCELLATIONS

DELAYS & CANCELLATIONS

Common and disruptive events in air travel

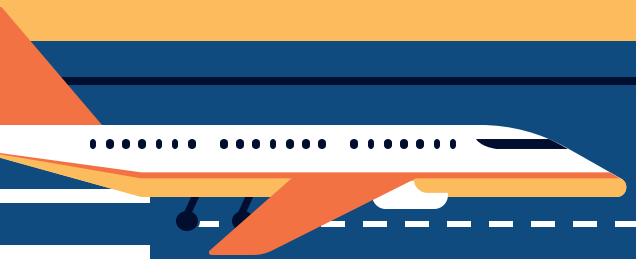
IMPACT ON PASSENGERS

Lead to traveler frustrations and negative experience



BROADER IMPLICATIONS

Can damage an airline's reputation and customer loyalty



AIRLINE OPERATIONAL CHALLENGES

INCREASED OPERATIONAL COSTS

Delays and cancellations often result in financial losses due to additional staffing, fuel, and accommodation costs

RESOURCE MANAGEMENT

Disruptions lead to inefficient use of resources like aircraft and crew, affecting overall operational flow

RIPPLE EFFECT ON SCHEDULES

A single delay can have a cascading effect, disrupting the schedule of multiple flights and causing broader network inefficiencies

Create a feature-focused model to predict flight delays and cancellations to provide recommendations for future airline operations

PROBLEM STATEMENT

METHODOLOGY

Dataset Overview and Models

DATASET

OVERVIEW

- **Millions of flights**
  - Multiple airlines
  - 2018-2022
- **61 original features**
  - Origin, destination, departure time, delayed status, etc.

FEATURE GENERATION

- Holiday closeness
- Delta Hubs (Origin & Destination)
- Flight congestion

MODELS

XGBoost

- Performs Well in Practice
- Fitting on Residuals

Random Forest

- Ensemble Learning
- Reduces Overfitting

Naive Bayes

- Simple Model
- Handles High Dimensional Data

SVM

- Handles Non-Linearity
- Robust Generalization

Logistic Regression

- Works Well for Linear Relationships
- Interpretable

RESULTS AND ANALYSIS

Model Performance & Prediction Accuracy

MODEL PERFORMANCE

XGBoost

Highest Accuracy with 0.724

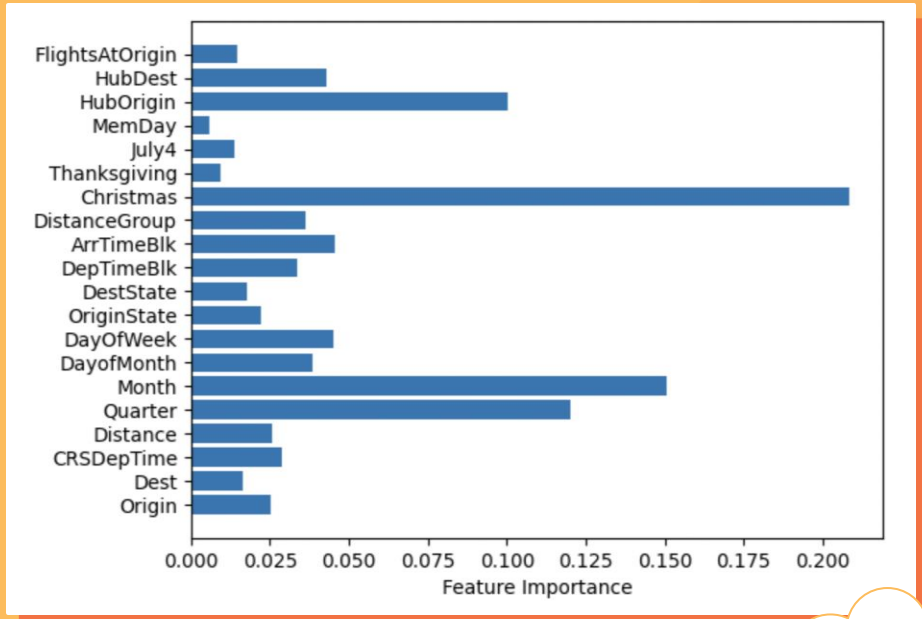
Naive Bayes

Lowest Accuracy with 0.591

XGBoost has the capability to handle complex relationships within an extensive feature set

	Accuracy
Logistic Regression	0.708386
Random Forest	0.713459
Naive Bayes	0.591443
Support Vector Machine (SVM)	0.704056
XGBoost	0.723925

FEATURE IMPORTANCE



CHRISTMAS

Most Important Feature

MONTH & QUARTER

Second highest

MEM DAY

Least Important Feature

FEATURE PREDICTION ACCURACY

Airports:

Top 5 accuracies:
XNA, MSN, DSM, TLH, JAN

Day of Week:

Tuesday has the highest accuracy

Day in Month:

1st day has the highest accuracy

Arrival Time Block:

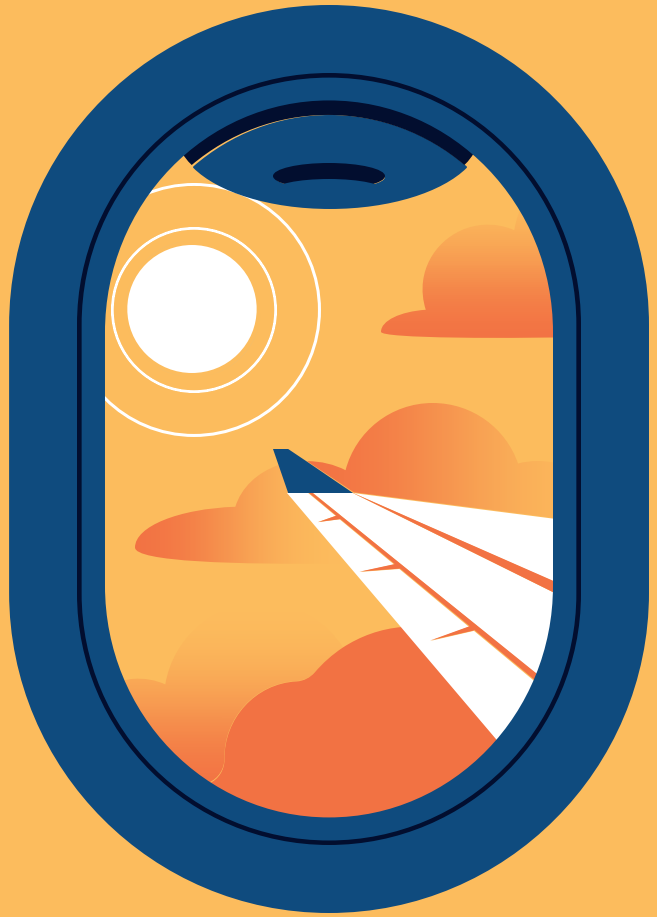
Block 4 (4:00 – 5:00 AM) has the highest accuracy

Hub or Not:

Hub Origin: Low accuracy
Hub Destination: High accuracy

RECCOMENDATIONS TO DELTA

What has been working and what needs to change



LEAST DISRUPTIVE AIRPORTS

AIRPORTS:

XNA, MSN, DSM, TLH, JAN

IMPLICATION:

Apply successful strategies from these airports to more challenging ones

MOST DISRUPTIVE AIRPORTS

AIRPORTS:

EYW, GNV, MTJ, PWM, MDT

CONTRIBUTING FACTORS:

Geographic location, airport infrastructure, operational constraints

RECCOMENDED APPROACH:

Targeted operations handling, enhanced resource allocation, apply successful strategies from previous airports

DAY OF WEEK DISRUPTIONS

HIGHER DISRUPTION DAYS

Fridays and Sundays (overall weekends)

REASONING

Increased passenger traffic & tighter schedules lead to cascading delays

STRATEGY

Implement more robust strategies on peak days

TIME OF DAY DISRUPTIONS

MORE DISRUPTIONS IN NIGHT FLIGHTS

04:00 am	**TIME BLOCK 4**	Least disruptions
12:00 am	**TIME BLOCK 0**	Most disruptions
RESULTS IN CUMULATIVE DELAYS		Focus on morning punctuality to mitigate night delays

OPTIMIZING FLIGHT SCHEDULING

Predictive analytics is essential in improving strategies against disruptions.
Key to Reducing Disruptions:

- Avoid scheduling during peak congestion
- Distribute flights evenly throughout the day
- Focus on worse performing airports
- Apply successful strategies to other airports

With more data can enhance accuracy of predictive model in the future.

PROJECT RECAP

REVOLUTIONARY PREDICTIVE MODEL

XGBoost Capable of Forecasting Flight Status with 72.4% Accuracy

ACTIONABLE INSIGHTS FOR DELTA

Identified successful strategies to apply to flights that are likely to face challenges

FEATURE-ORIENTED ANALYSIS

Uncovered factors influencing disruptions: flight congestion and proximity to major holidays

POTENTIAL FUTURE DIRECTIONS

Investigate external factors: weather and global events and implement models with real-time data

THANK YOU

Iris Brook
MBAn 2024
irisb211@mit.edu

Lucy Liu
MBAn 2024
feifanl@mit.edu

Krishanu Datta
MBAn 2024
krishanu@mit.edu

Devin Wasilefsky
MBAn 2024
dwaz23@mit.edu

QUESTIONS?