Unveiling Flight Disruptions Insights

Team BA7

Jinke Han, Sneha Sunil Ekka, Tanvi Sheth, Mingze Wu, YuChun Su

BOARDING PASS

FLIGHT
BA843

GATE
07

SEAT
1A

Table of Contents

01 **Introduction**
- Motivation
- Business Description
- Data Overview

02 **Descriptive Analysis**
- Location/Airports
- Airlines
- Time

03 **Machine Learning**
- Model comparison
- Feature Importance
- Hyperparameter tuning

04 **Implications & Challenges**
- Recommendations
- Limitations
- Future Scope

BOARDING PASS

FLIGHT

BA843

GATE

07

SEAT

1A

01

INTRODUCTION

Motivation

PROBLEM

Flight delays and cancellations are inconvenient, disruptive, stressful, and costly.

GOAL

Identify unique patterns among airlines and airports to enhance operational efficiency, improve travel experiences, and mitigate economic losses.

Business Description

01

Analyze delays to discern seasonal trends and variations

Offer valuable insights into the factors affecting flight schedules

02

Develop predictive models to estimate delay durations

Empower stakeholders with proactive strategies to mitigate risks

03

Provide passengers with travel time recommendations

Facilitate informed decisions to minimize disruptions' impact on plans

Dataset

Time Range

5 years of flight data from 2018 to 2022 for US commercial airlines

Columns

Total 61 columns including flight dates, airline information, aircraft information, departure and arrival times

Target Column

Detailed information about the flight status with respect to the cancellations and delays

BOARDING PASS

FLIGHT

BA843

GATE

07

SEAT

2A

02

Exploratory Data Analysis

BOARDING PASS

FLIGHT
BA843

GATE
07

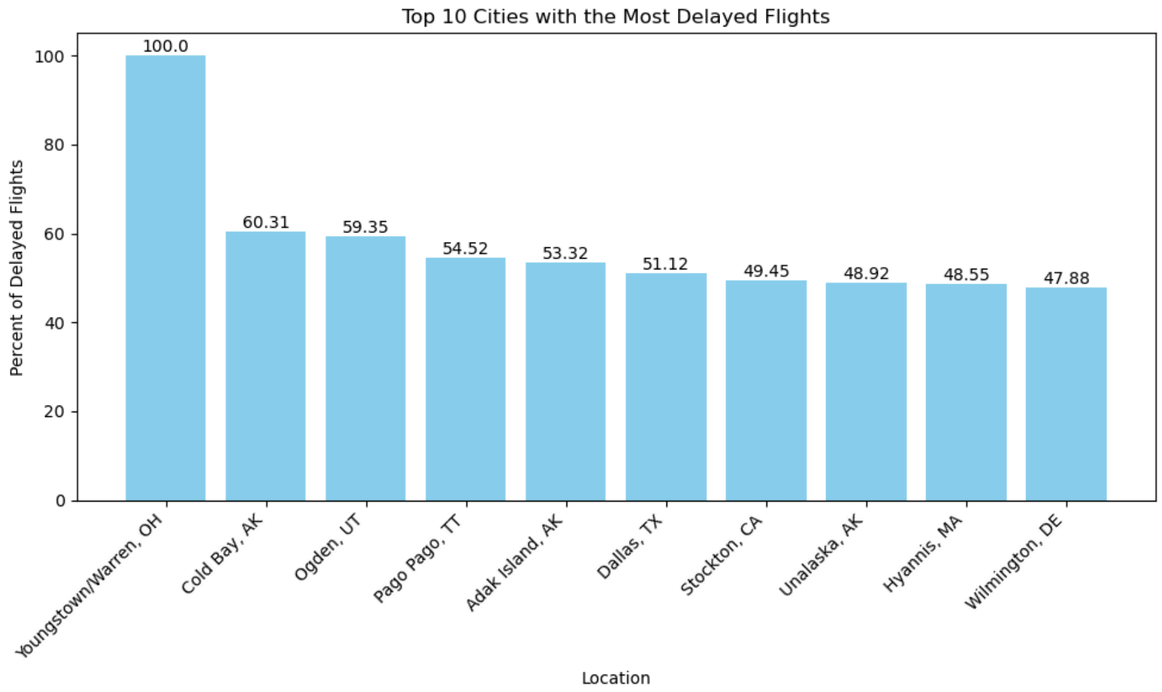
SEAT
2A

02

EDA

Analysis by “Location”

Which cities experienced the highest percentage of delayed flights?



Top 10 Cities with the Most Delayed Flights

OriginCityName	Count	Delayed Count	Delay Proportion
Youngstown/Warren, OH	2	2	100.00%
Cold Bay, AK	262	158	60.31%
Ogden, UT	460	273	59.35%
Pago Pago, TT	299	163	54.52%
Adak Island, AK	452	241	53.32%
Dallas, TX	292,538	149,560	51.12%
Stockton, CA	3,521	1,741	49.45%
Unalaska, AK	1,392	681	48.92%
Hyannis, MA	379	184	48.55%
Wilmington, DE	165	79	47.88%

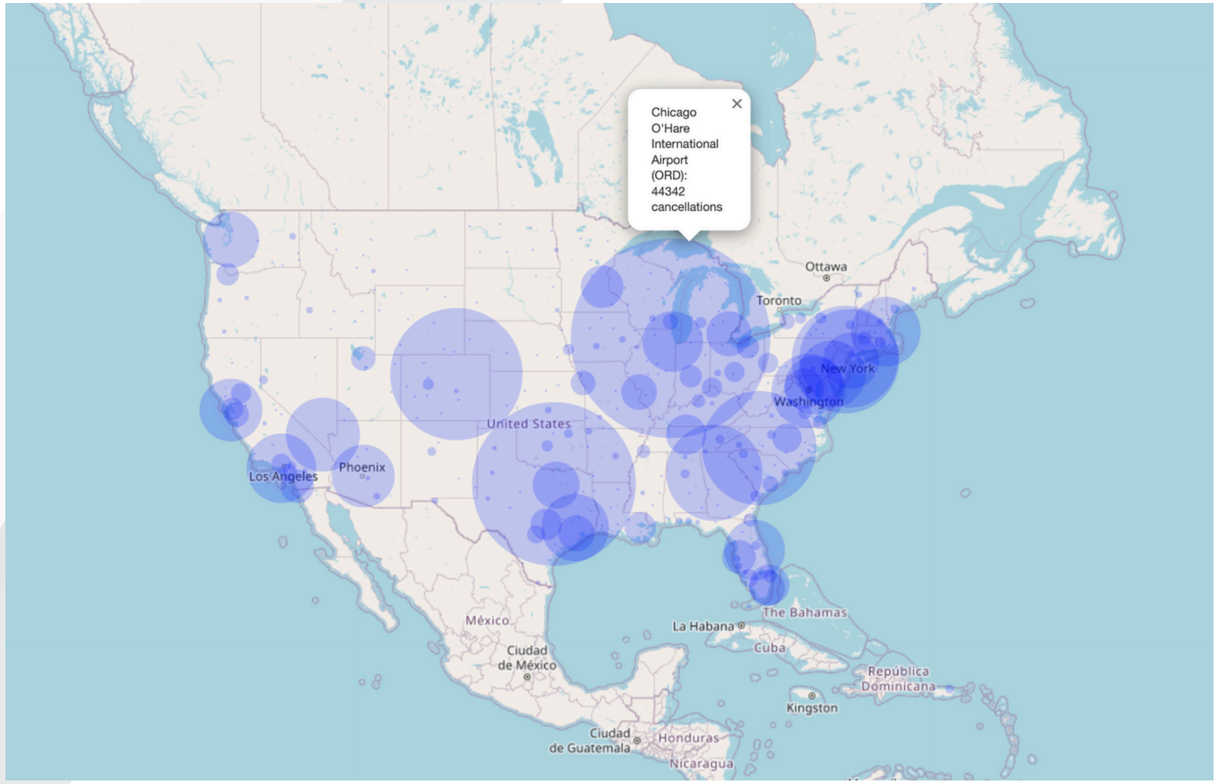
What happened in Dallas?

Origin	Airline	Total_delay	Flight_count	Average_delay
DAL	SkyWest Airlines Inc.	100,061.0	2,286	43.77
DAL	Delta Air Lines Inc.	72,566.0	1,830	39.65
DAL	Alaska Airlines Inc.	32,668.0	850	38.43
DAL	Endeavor Air Inc.	787.0	21	37.48
DAL	Virgin America	5,877.0	196	29.98
DFW	GoJet Airlines, LLC	7,456.0	80	93.20
DFW	Compass Airlines	27,630.0	305	90.59
DFW	Endeavor Air Inc.	72,875.0	1,074	67.85
DFW	Republic Airlines	103,899.0	1,591	65.30
DFW	ExpressJet Airlines LLC	99,216.0	1,556	63.76

Airports most prone to cancellations

Top 3 Airports with the Most Cancelled Flights

Origin	CancelledFlights
ORD	44342
DFW	36698
DEN	29676



- Flight cancellations across the USA are unevenly distributed, with a higher concentration in the eastern regions

BOARDING PASS

FLIGHT
BA843

GATE
07

SEAT
2A

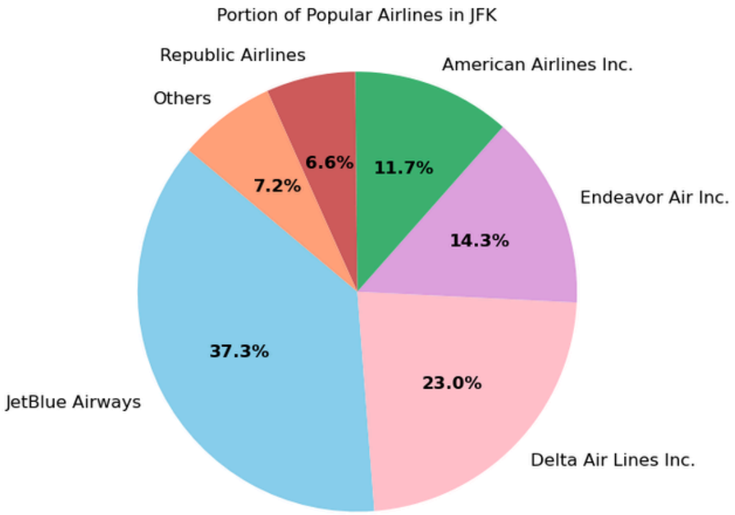
02

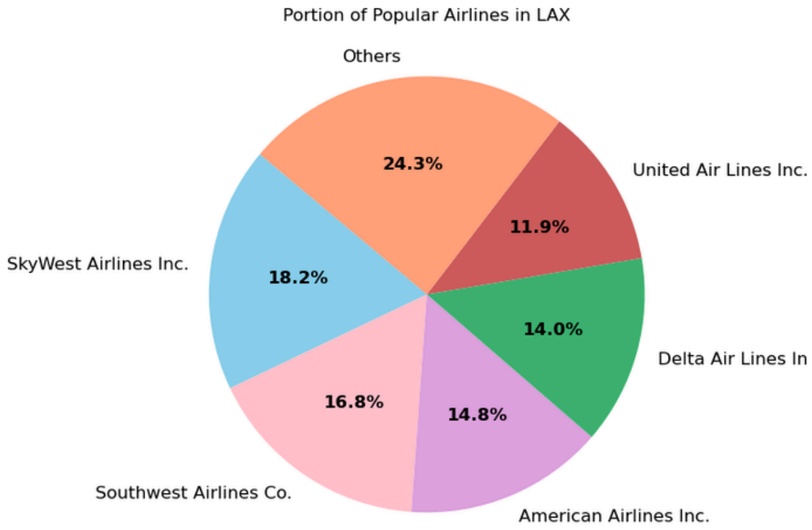
EDA
Analysis by “Airline”

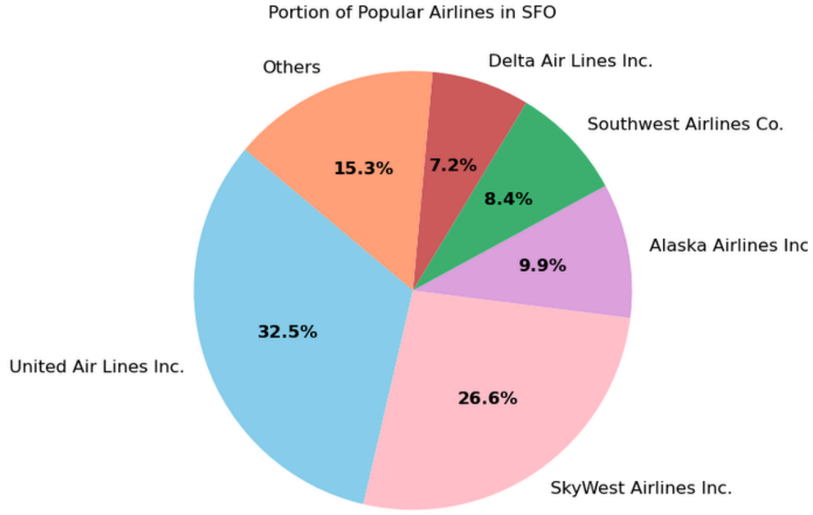
Proportion of Airlines at the busiest airports

- The top airlines that fly the most are mainly international.
- United Airlines is more common in northern cities while American Airlines are more popular in the south.

Proportion of major airlines at 3 Popular International Airports







- For JFK，JetBlue Airways and Delta Airlines handle more than half of all flights, making them the biggest players there.

- For LAX, this airport has a mix of many airlines, with no single airline having more than 20% of the flights.

- For SFO, Similar to JFK, two airlines at San Francisco airport control nearly 60% of the flights.

BOARDING PASS

FLIGHT
BA843

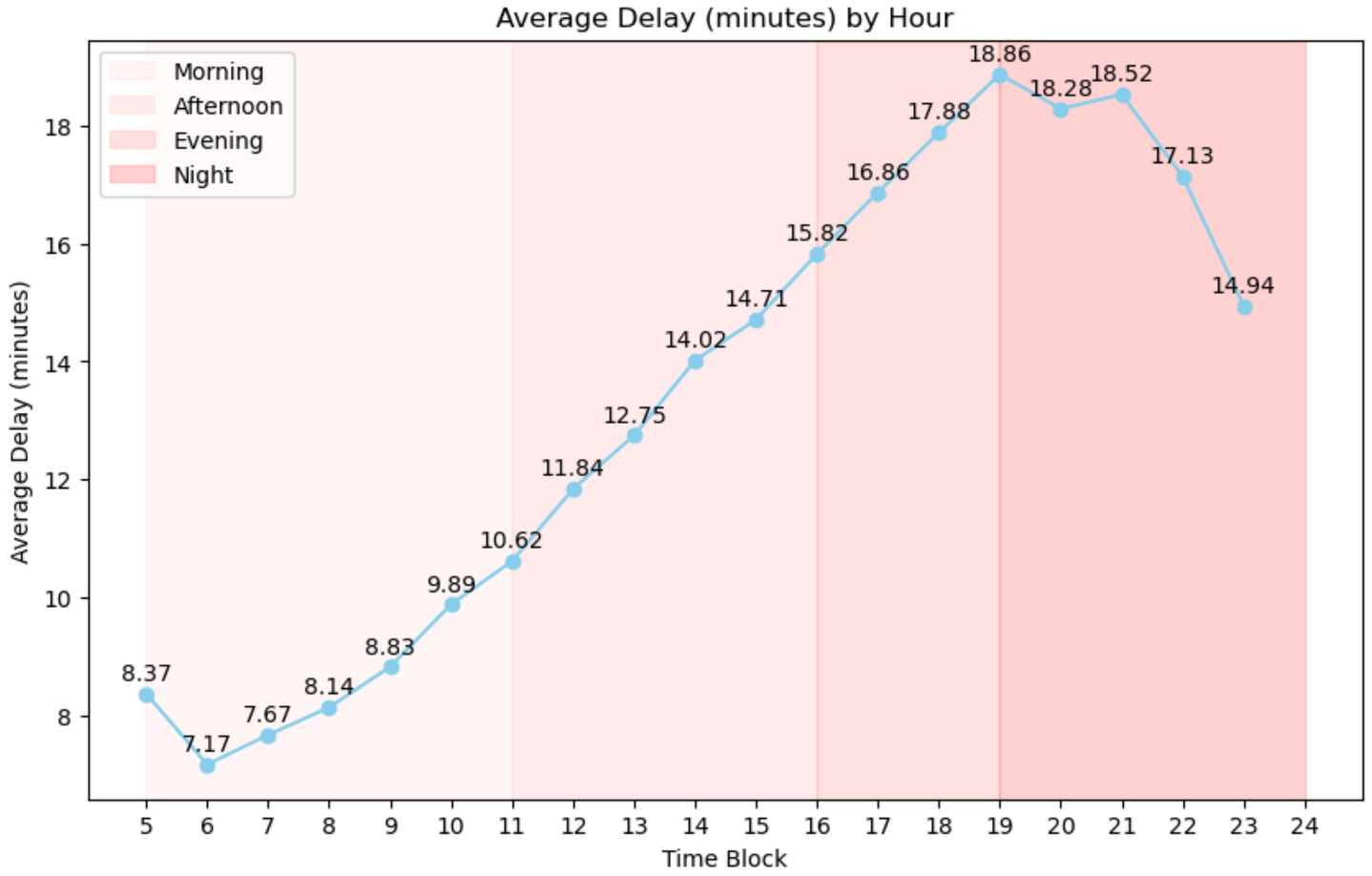
GATE
07

SEAT
2A

02

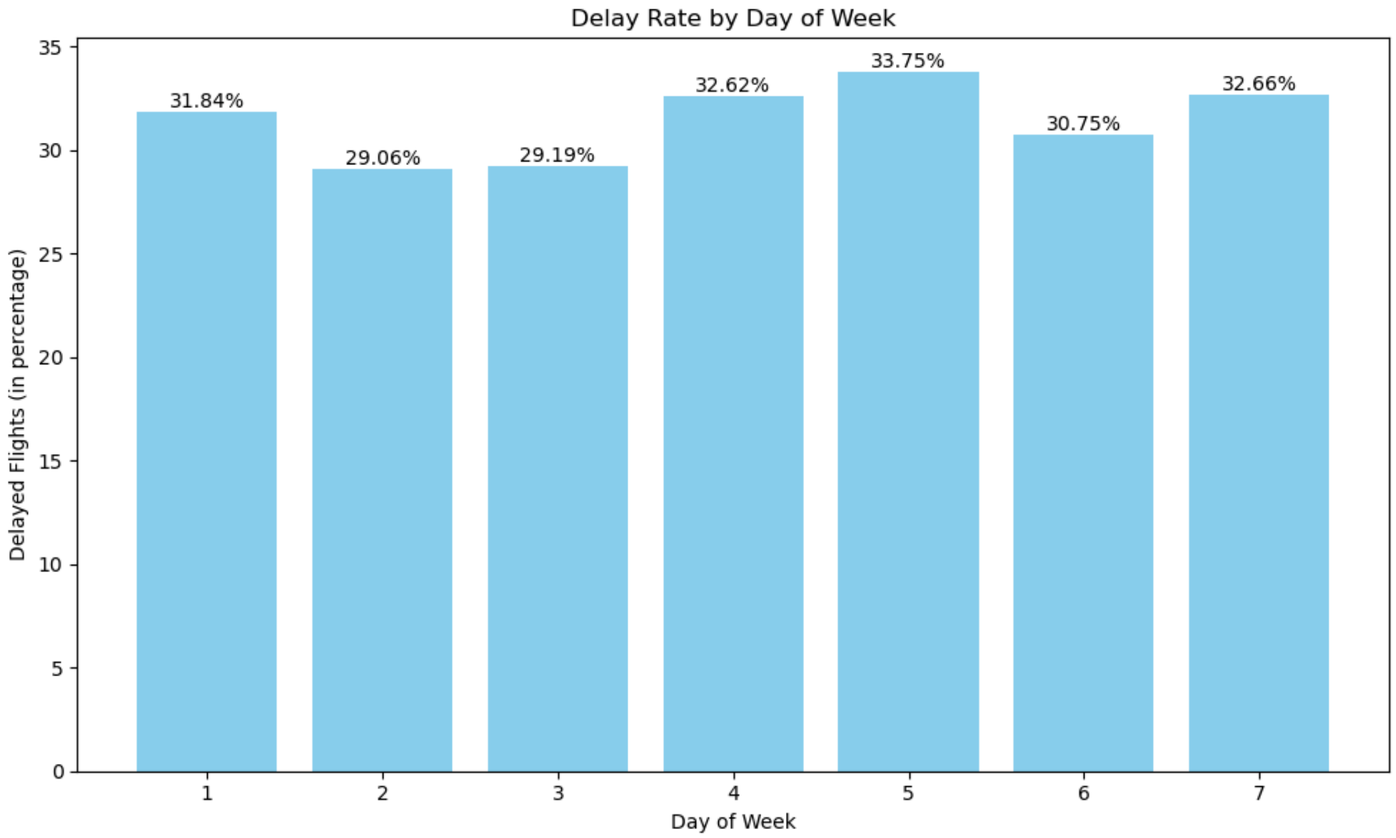
EDA
Analysis by “Time”

Delay Time by Hour of Day



- Rising trend throughout a day
- Evening and night periods experience the longest delays
- Peak at about 19 minutes on average at around 7 PM

Delay Rate by Day of Week

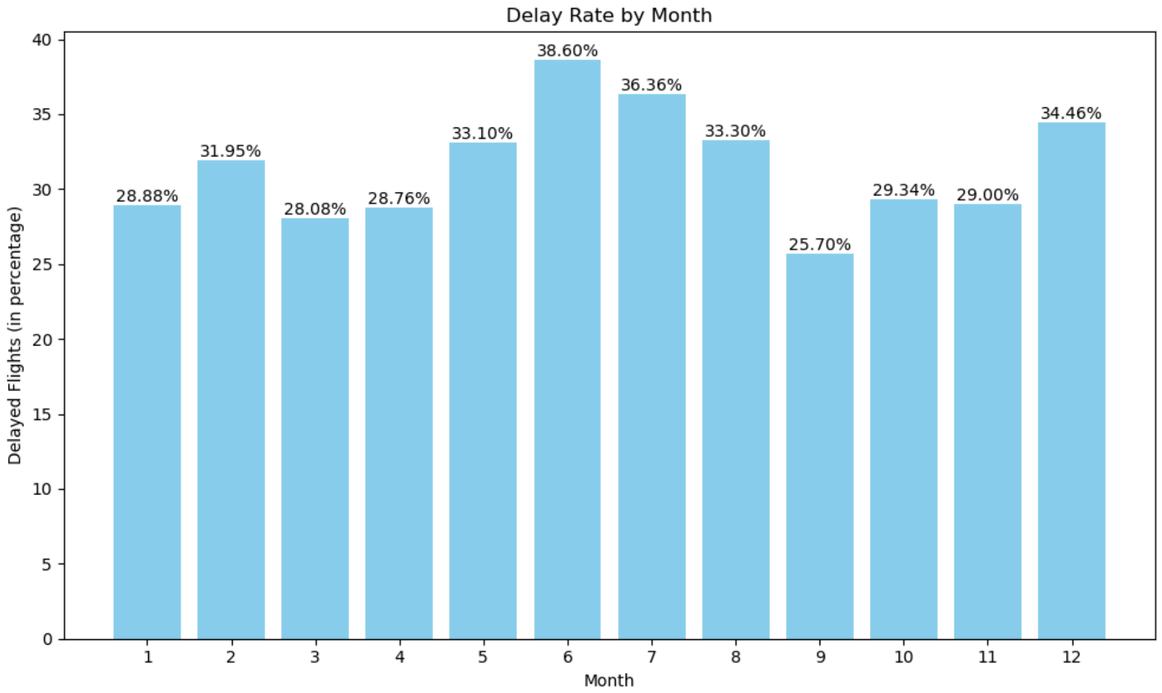


DayType	delay_proportion
Weekday	31.20
Weekend	31.63

- No obvious patterns based on weekdays or weekends.
- Peak on Friday
- The rise in travel toward weekends intensify delays

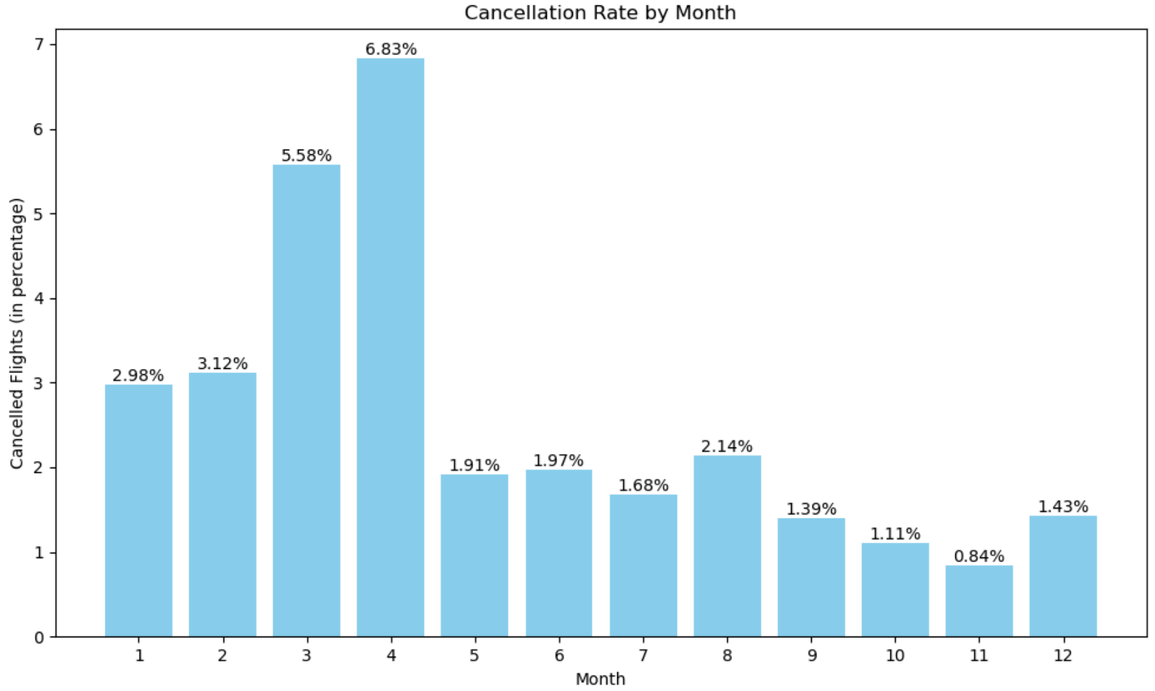
Delay & Cancellation Rate by Month

Delays



- Flight delays have a seasonal pattern
- Higher delay rates during the summer months

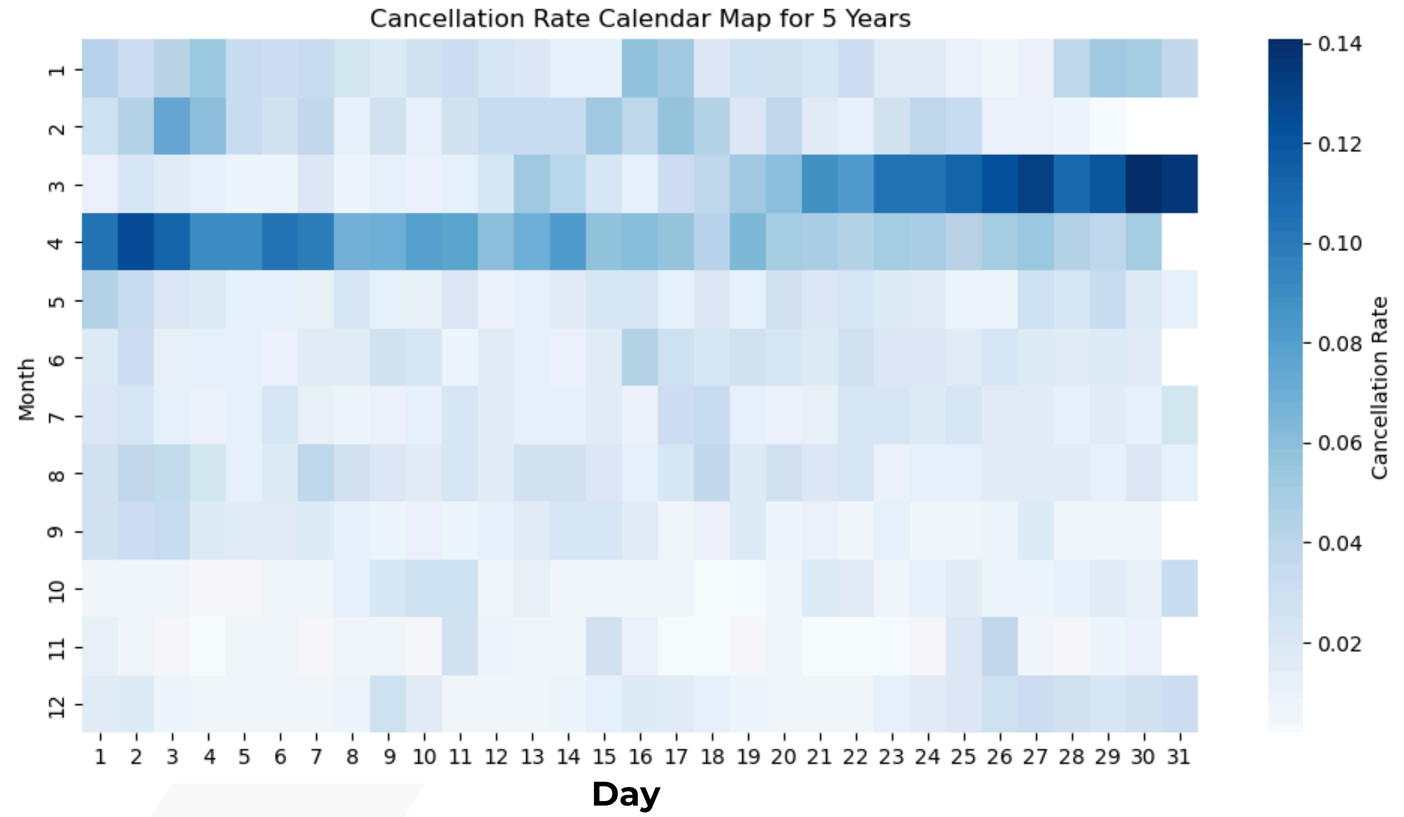
Cancellation



- Higher cancellation rates in the winter and early spring months
- Cancellation rates shoot up in the months of March and April

Average Daily Cancellation Rates over 5 Years

Average Cancellation %



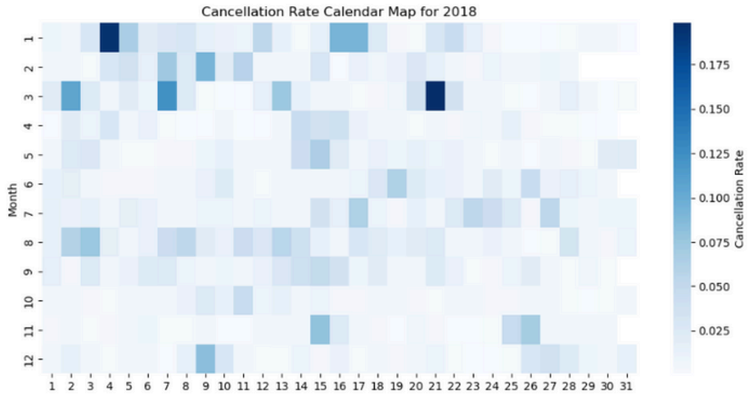
March to April

Month

- The most cancellations in March and April

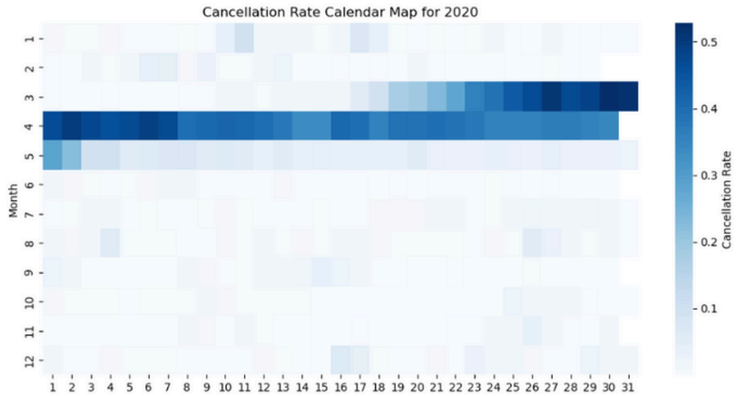
Average Daily Cancellation Rates for each year

18 %



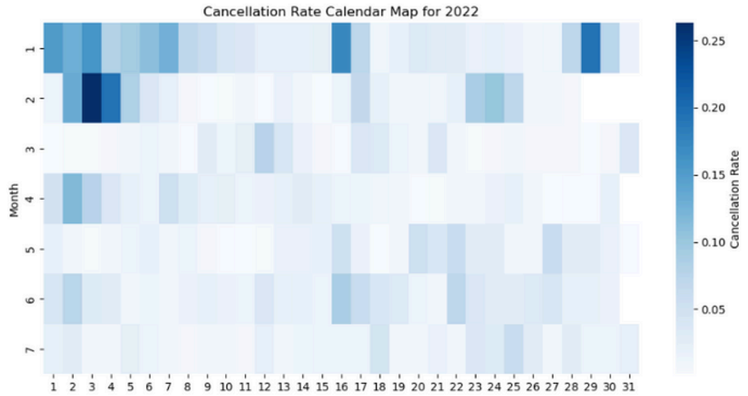
2018

50 %



2020

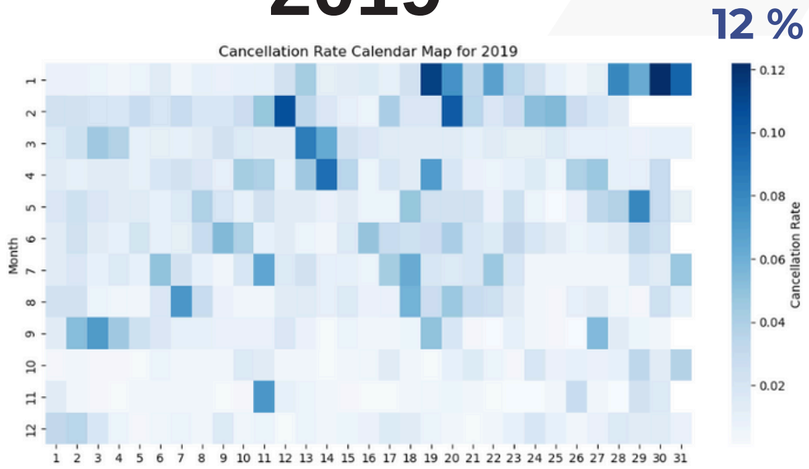
25 %



2022

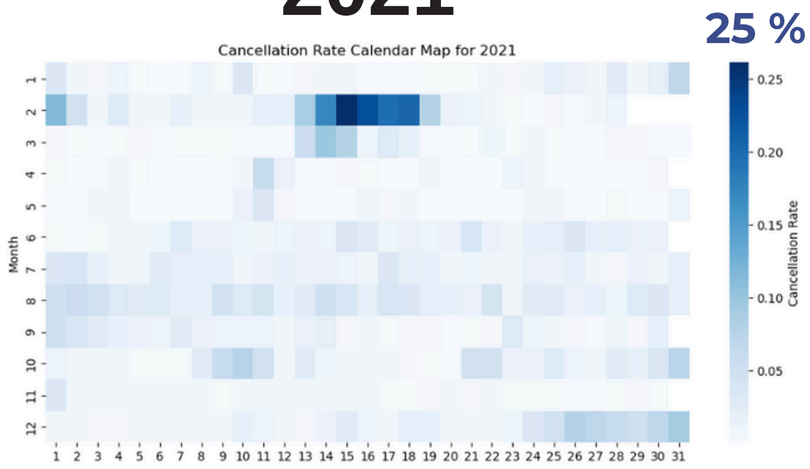
2019

12 %



2021

25 %



03

MACHINE LEARNING

Classification

BOARDING PASS

FLIGHT

BA843

GATE

07

SEAT

3A

Prediction Problem

TARGET: Flight Delay

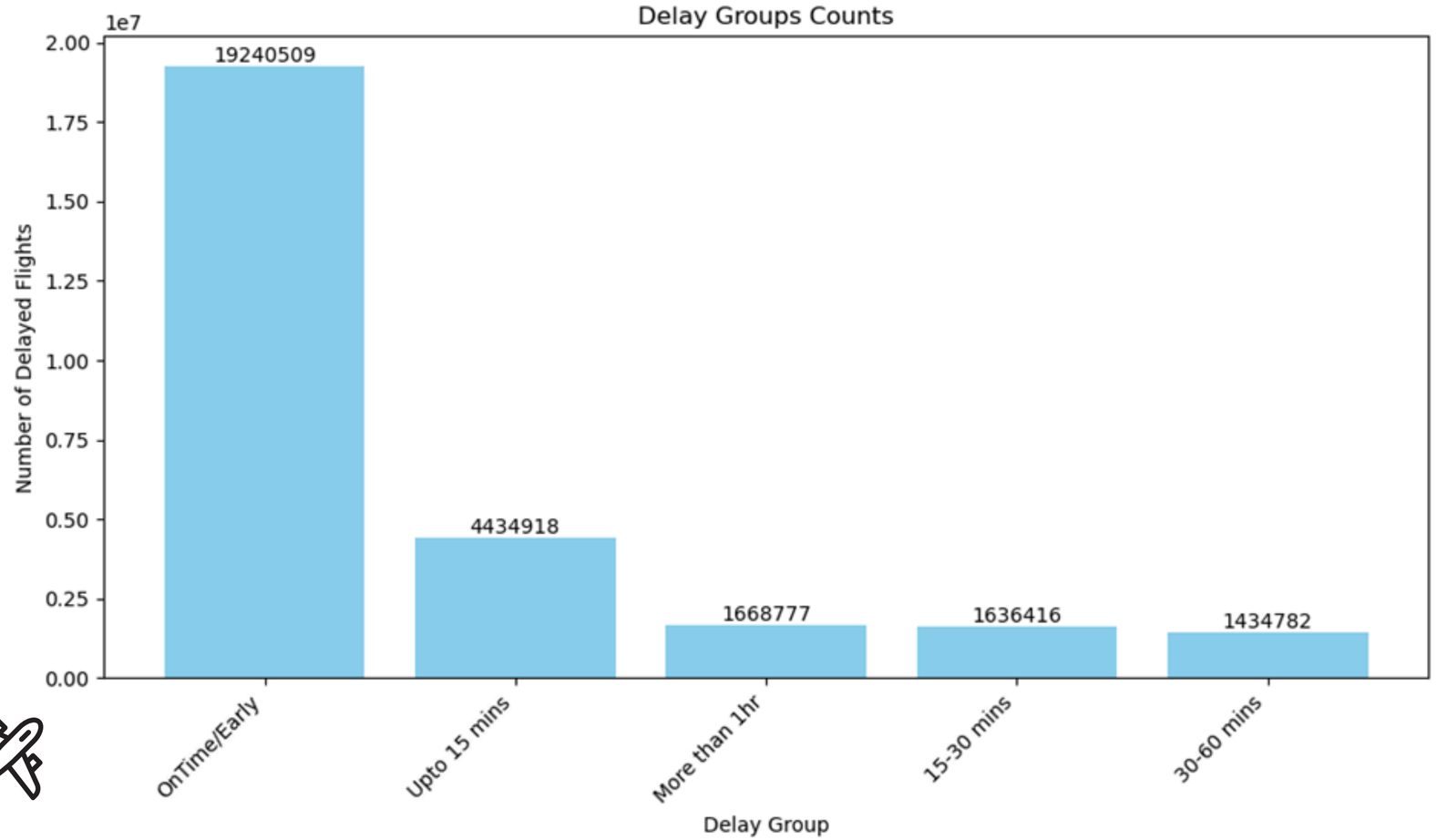
Upto 15 minutes | 15 - 30 minutes | 30 - 60 minutes | More than 1 hour

FEATURES

Origin Airport | Destination Airport | Airtime | Time of Day | Day of Week

Machine Learning

Distribution of Delay Categories



Insights

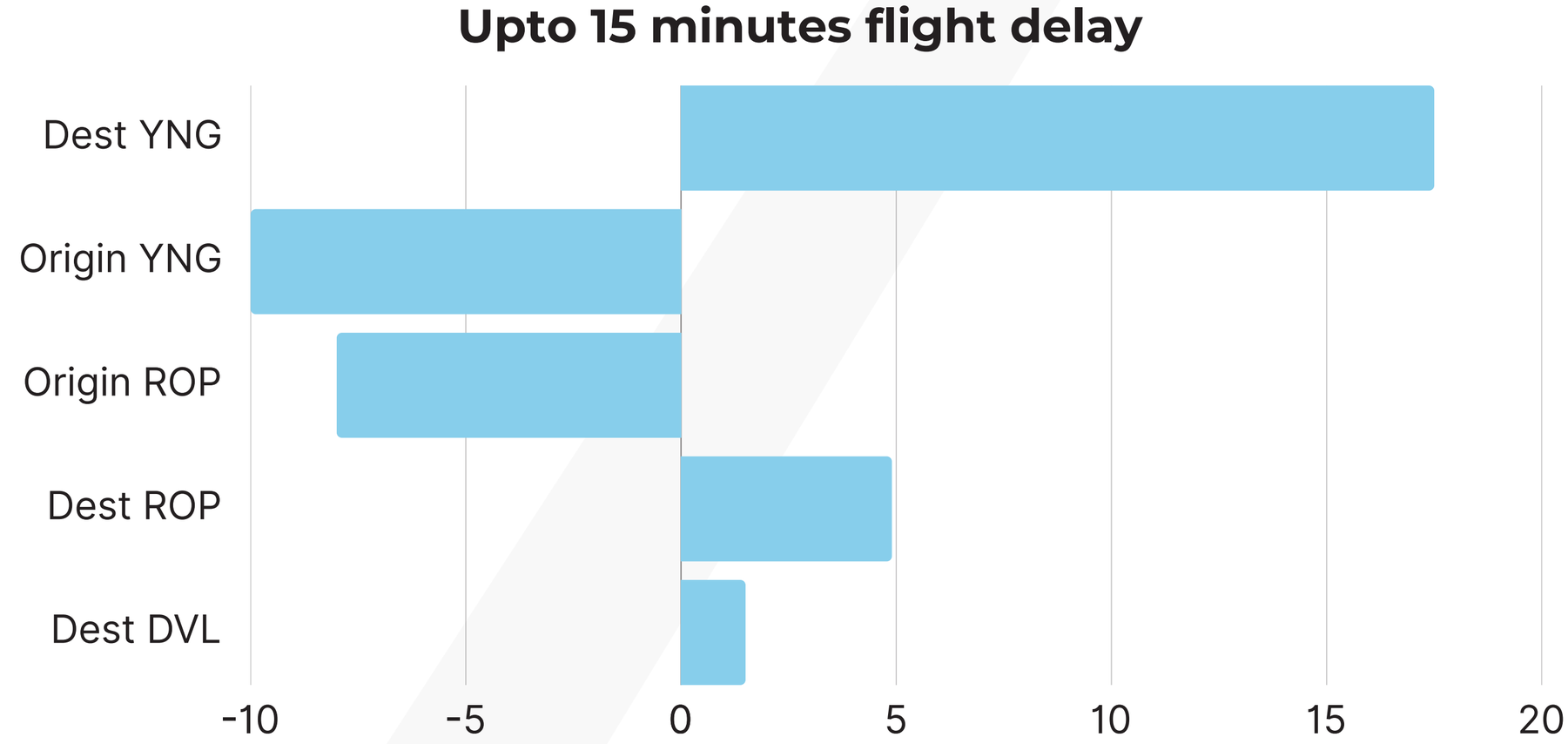
- Imbalanced groups
- Focusing on delays
- Predicting delay groups

Comparing Classification Models

MODELS	ACCURACY	PRECISION	RECALL	F1 SCORE
LOGISTIC REGRESSION	0.50	0.53	0.50	0.38
DECISION TREE	0.49	0.32	0.49	0.36
RANDOM FOREST	048	0.23	0.48	0.31

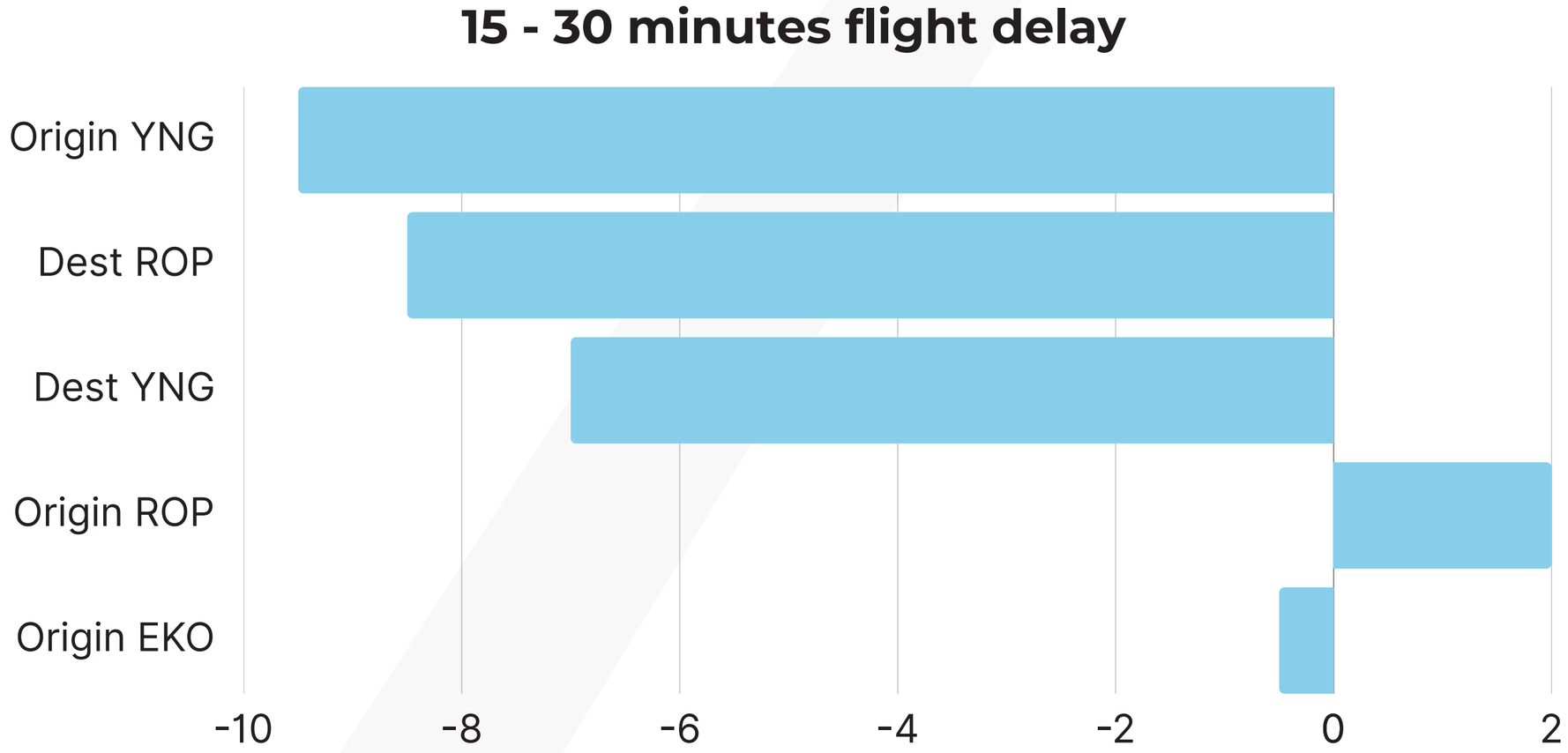
Logistic Regression emerges as the best base model with an accuracy of 50% in predicting each delay category.

Logistic Regression: Feature Importance



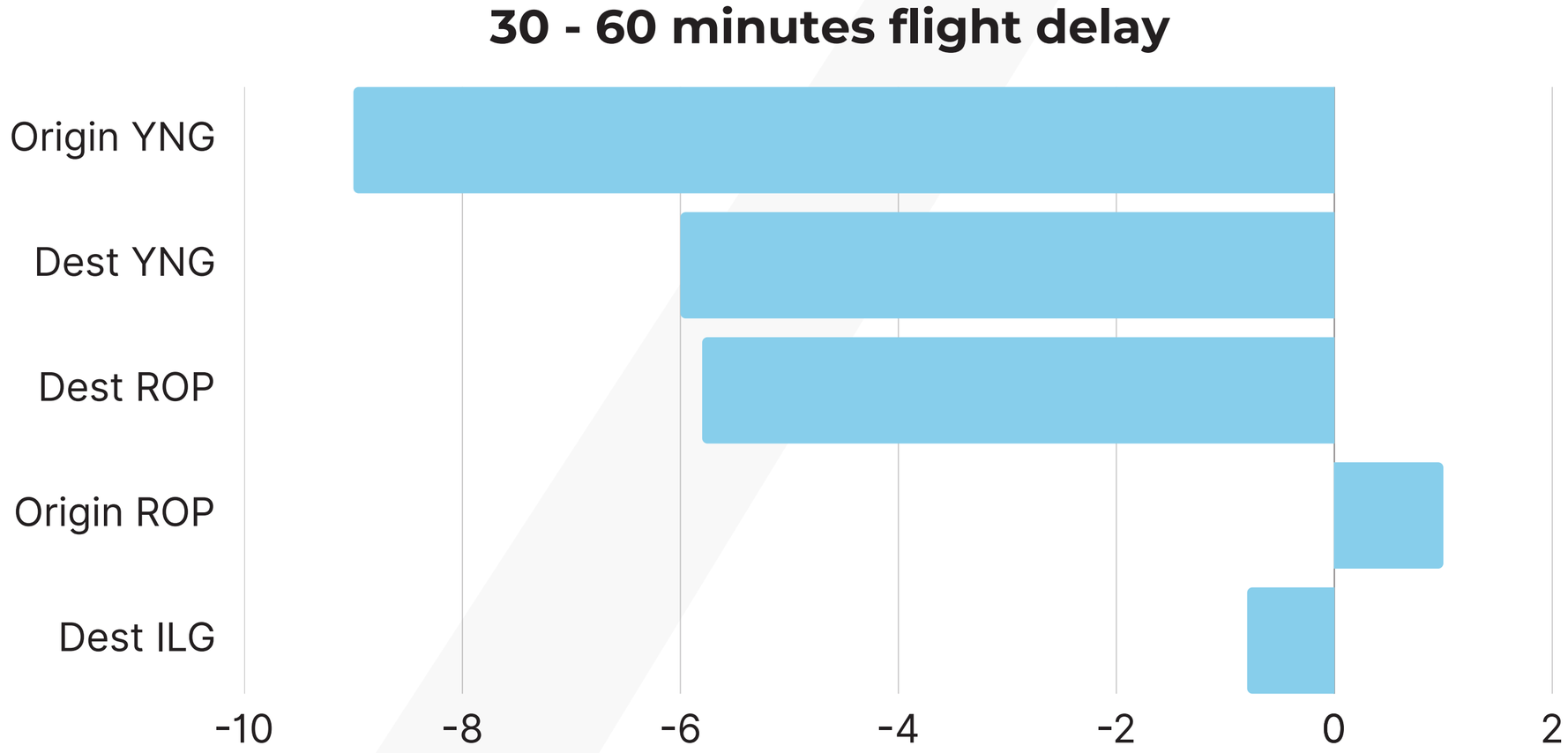
Important features: Youngstown-Warren Regional Airport, Ohio | Rota International Airport, Rota Island | Devils Lake Regional Airport, North Dakota

Logistic Regression: Feature Importance



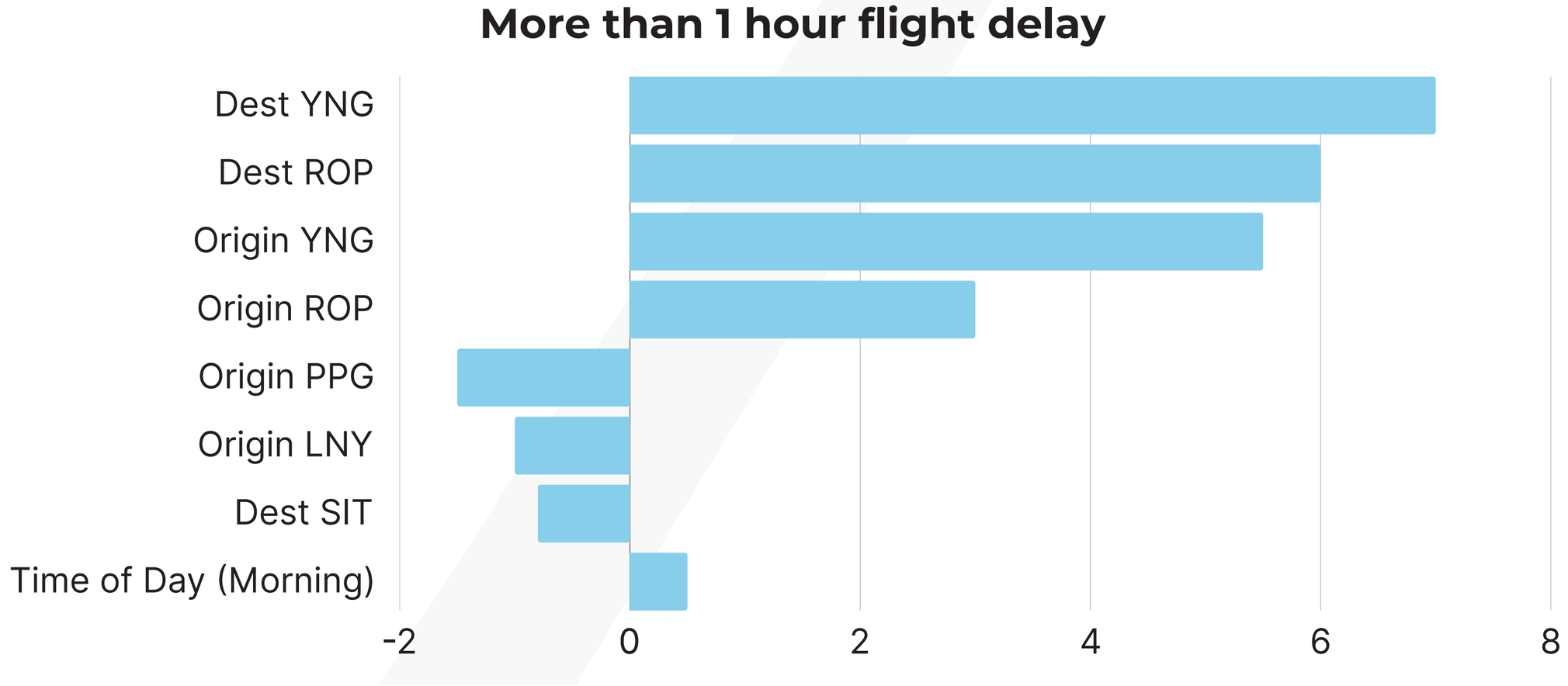
Important features: Youngstown-Warren Regional Airport, Ohio | Rota International Airport, Rota Island | Elko Regional Airport, Nevada

Logistic Regression: Feature Importance



Important features: Youngstown-Warren Regional Airport, Ohio | Rota International Airport, Rota Island | Wilmington Airport (Delaware)

Logistic Regression: Feature Importance



Important features: Youngstown-Warren Regional Airport, Ohio | Rota International Airport, Rota Island | Lanai Airport, Hawaii | Sitka Rocky Gutierrez Airport, Japonski Island | **Time of Day (Morning)**

Logistic Regression: Hyperparameter Tuning

MODELS	ACCURACY	PRECISION	RECALL	F1 SCORE
INITIAL MODEL	0.50	0.53	0.50	0.38
TUNED MODEL	0.49	0.36	0.49	0.37

BOARDING PASS

FLIGHT

BA843

GATE

07

SEAT

4A

04

CONCLUSION

Conclusion

Regional Airlines and Airports

Cancellation Hotspots

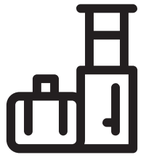
External Factors

Timing & Seasonality

Airlines Performance

Challenges

01 **Computational limitations**

02  **Missing Data**

03 **Skewed Data**

Thank You for flying with us today!

BOARDING PASS

FLIGHT
BA843

GATE
07

SEAT
BYE!!!