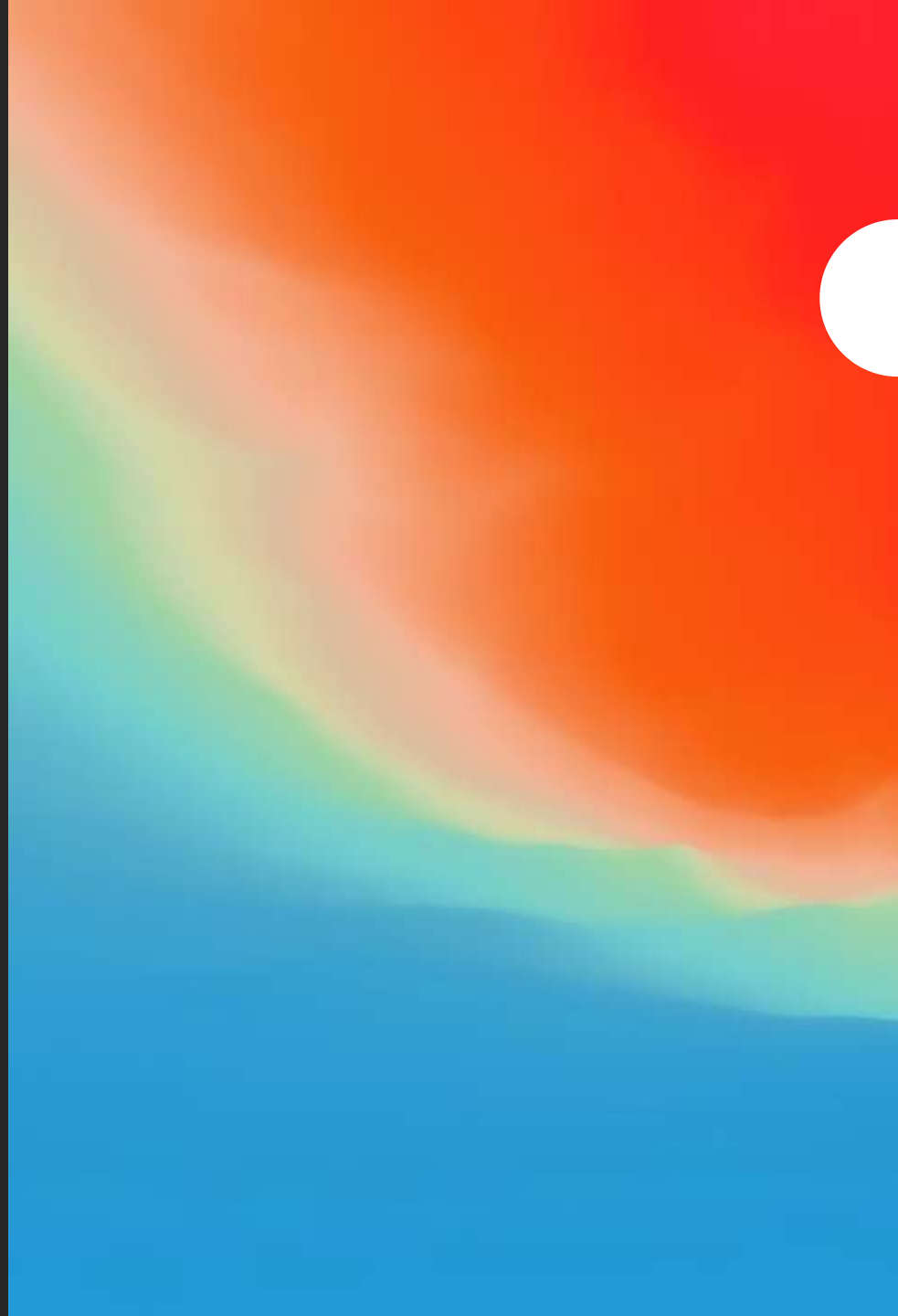
Optimizing Air Travel: A Data-Driven Approach to Flight Delay Analysis and Prediction



Project Overview

Goal: Predict if a flight will be delayed and for how long.

Why? To reduce delay problems and help airlines run more smoothly.

How? Use past flight data to find delay patterns and build prediction models.

Data & Method

•Data: About 1.8 lakh flights from U.S. airports with delay reasons.

•Cleaned and prepared the data (removed wrong values).

•Used important columns like late_aircraft_ct , carrier_ct, arr_flights etc.

•Made two models:

Classifier: Will the flight be delayed? (Yes/No)

Regressor: How many minutes delayed?

•Made a special score called **OAI** to focus on delays that airlines can control.

EDA (Data Analysis) Highlights

1

Most flights have short delays. Few have very big delays.

2

Main delay reasons: Late aircraft and carrier (airline) issues.

3

More delays happen during busy months and hours.

4

Some airports cause more total delays than others.

Model Results

- **Delay Yes/No (Classifier)**:
  - Accuracy: 99.8%
  - Almost all flights predicted correctly!
- **Delay Time (Regressor)**:
  - Error is high (delay time is hard to guess exactly).
  - MAE ≈ 674 min, RMSE ≈ 2185 min.

SHAP Insights

- SHAP tells us **which features affect delay the most**.
- Top 3 reasons:
    - *Delay from late aircraft*
    - *Delay from airline (carrier)*
    - *Total incoming flights*
- SHAP helps us **understand the model's decisions**.

Recommendations

Improve late aircraft handling — avoid one delay causing more delays.

Add extra buffer time during rush hours.

Schedule important flights earlier in the day.

Focus on flights with high OAI — where airline can fix delay causes.

- Thank you