



GROUP 4 FINAL PROJECT PRESENTS:

Home Flight ☰

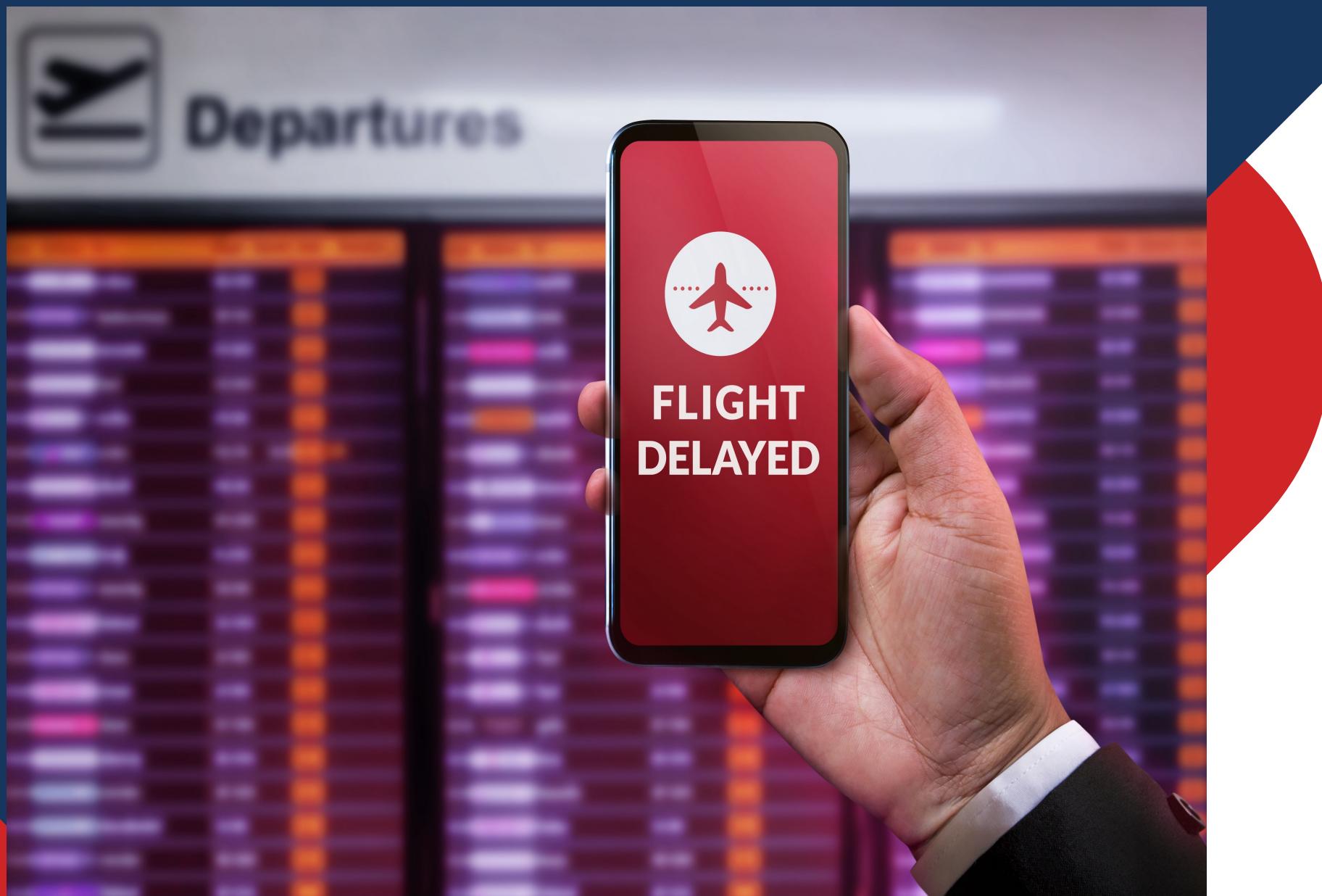
FLIGHT DELAYS:

**Flight Delays,
Best & Worst Airports, and
Best & Worst Airlines**



Group Members:
Kenan Foust
John Harris
Marcia Joseph





Why we chose this topic?

- Airports and airlines have been in news a lot due to flight delays and cancellations
- Curiosity: Wanted to learn about the causes for flight delays
- There was a great deal of information/datasets available about flight delays
- We wanted to share our top 10 list of Best and Worst Airports and Airlines for flight delays
- ???

Project Overview

January 2020 Flight Predictions

1. Predictive Problem
Flight Delays for the total flights in January 2020
2. Determine the Best and Worst Airlines for flight delays
3. Determine the Best and Worst Airports for flight delays

Data Source Description

- Data collected from Kaggle.com, “January Flight Delay Prediction”
- Data is collected from the Bureau of Transportation Statistics, Government of the United States of America
- Data is open-sourced under the U. S. Government Works
- Contains all the flights starting from January 1 – 31, 2020
- There are more than 400,000 flights in the month of January itself throughout the United States
- We used the data to predict the arrival and departure flight delays by Airline and Airport





Technology, Languages, Tools, & Algorithms Used for Project

- Python
- Jupyter Notebook
- Tableau
- PgAdmin
- Quick DBD
- PowerPoint
- Github
- ???





Data Exploration Machine Learning Models

Classification Model Testing:

1. Logistic Regression Model
2. Random Forest Classifier
3. Gradient Boosting Classifier

Machine Learning Model Classification Model Testing



Logistic Regression Model

- **Training Accuracy**



- **Testing Accuracy**



- **Precision (weighted average):**



- **Recall (weighted average)**



- **F1-Score (weighted average)**



Machine Learning Model Classification Model Testing



Random Forest Classifier

- Training Accuracy



- Testing Accuracy



- Precision (weighted average):



- Recall (weighted average)



- F1-Score (weighted average)



Machine Learning Model Classification Model Testing

Gradient Boosting Classifier



- Training Accuracy 93.21%
- Testing Accuracy 93.21%
- Precision (weighted average) 93%
- Recall (weighted average) 93%
- F1-Score (weighted average) 93%

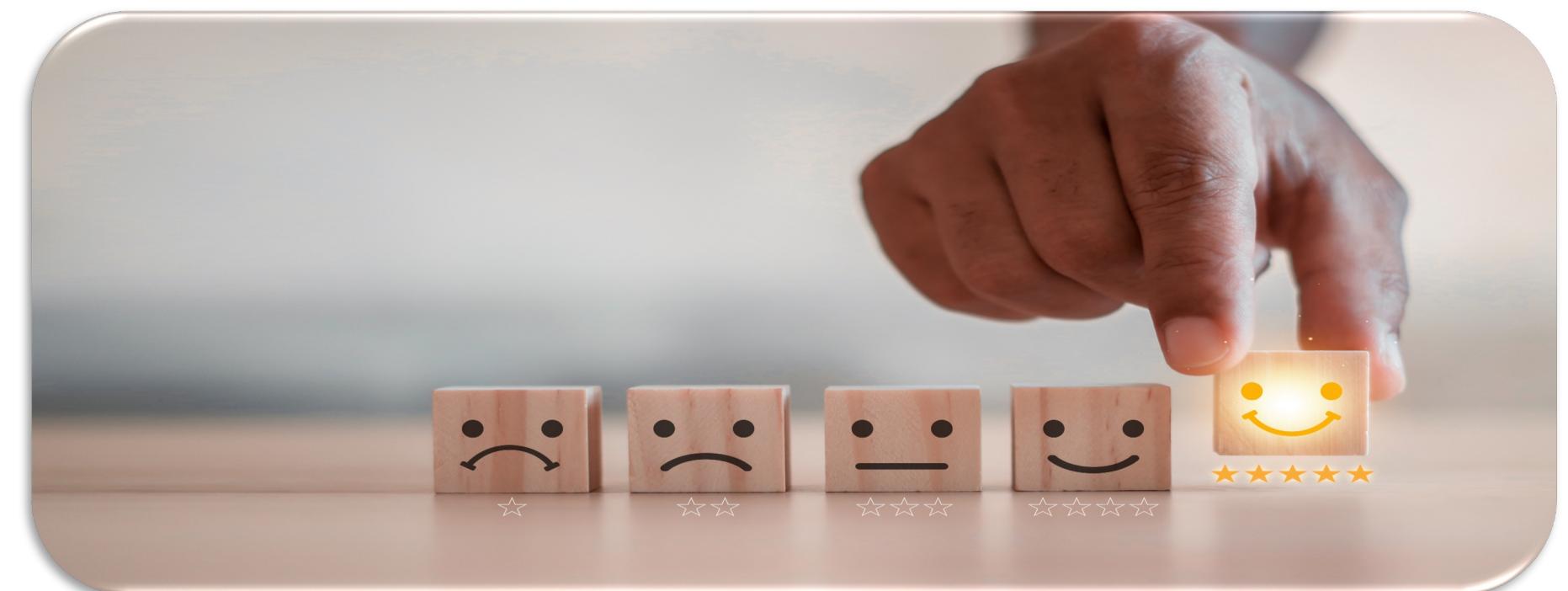
Best Machine Learning Model for Project: Gradient Boosting Classifier

1. Highest Testing Accuracy

- Among the three models, it has the highest testing accuracy (93.21%), which shows it works well with unseen data.
- Has a strong balance between precision, recall, and F1-Score, indicating that it performs well across different metrics.

2. Most Reliable Choice

- The model is not only good at identifying true positives but also at minimizing false positives and false negatives, making it the most reliable choice.
- **Best model for our final project!**



Gradient Boosting Classifier

1. Highest Testing Accuracy

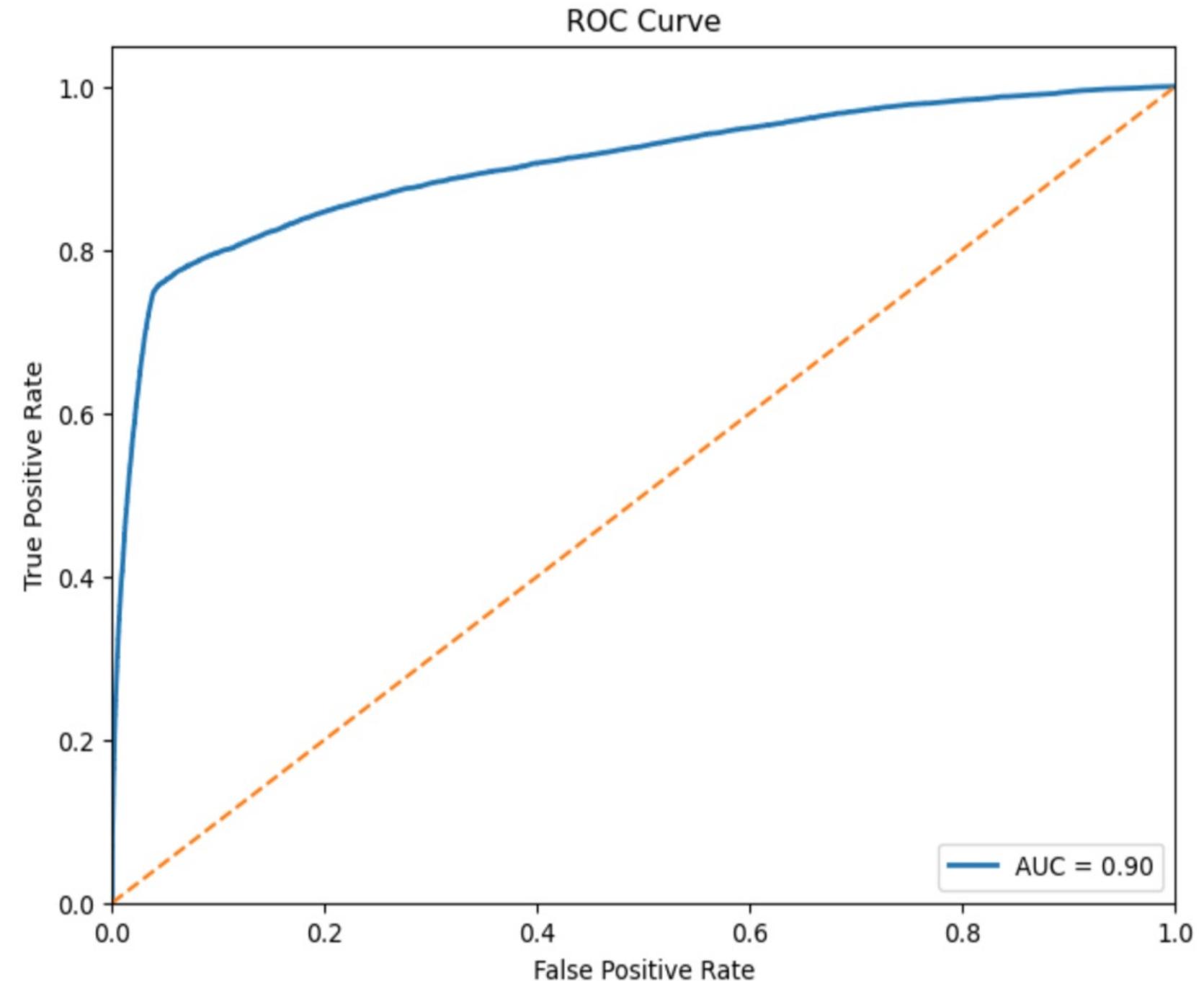
```
Training Accuracy: 0.9320879240072255
Testing Accuracy: 0.932100722545764

Confusion Matrix:
[[99425 3965]
 [ 4173 12291]]

Classification Report:
precision    recall    f1-score   support
0.0          0.96     0.96      0.96      103390
1.0          0.76     0.75      0.75      16464

accuracy         0.93      0.93      0.93      119854
macro avg       0.86     0.85      0.86      119854
weighted avg    0.93     0.93      0.93      119854
```

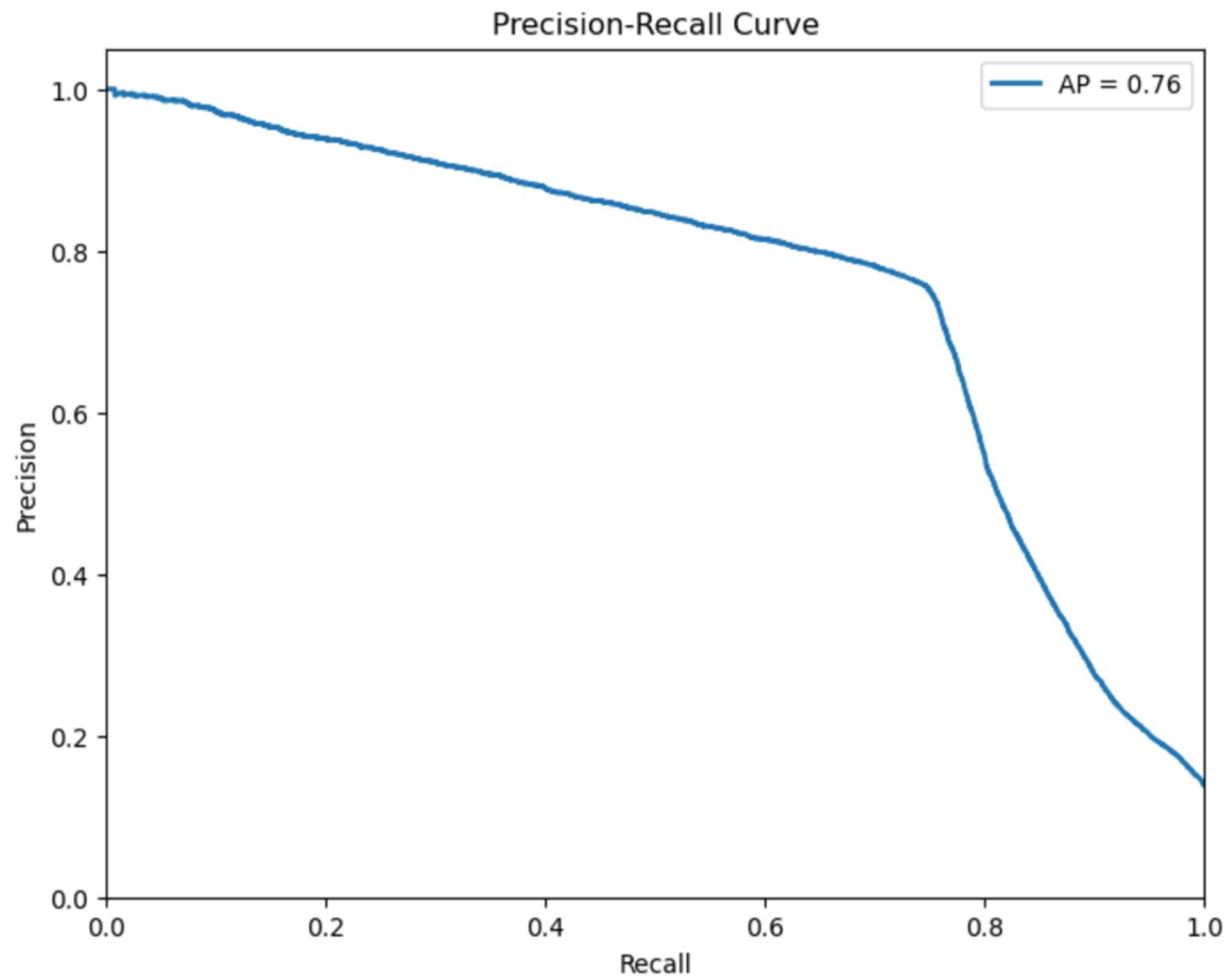
2. ROC Curve



Gradient Boosting Classifier



3. Precision-Recall Curve



4. Confusion Matrix

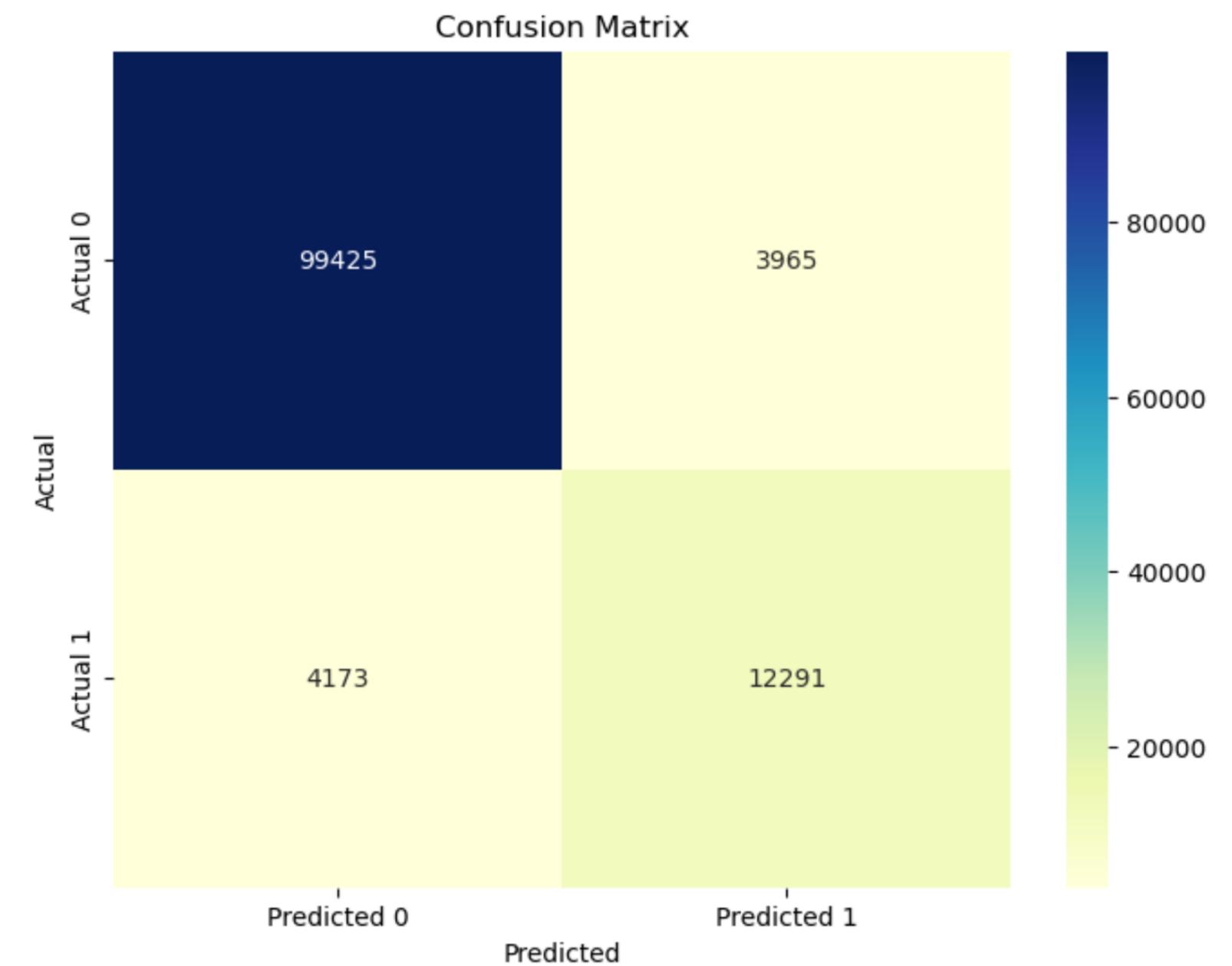


Tableau Visualization Dashboard

Tableau Dashboard

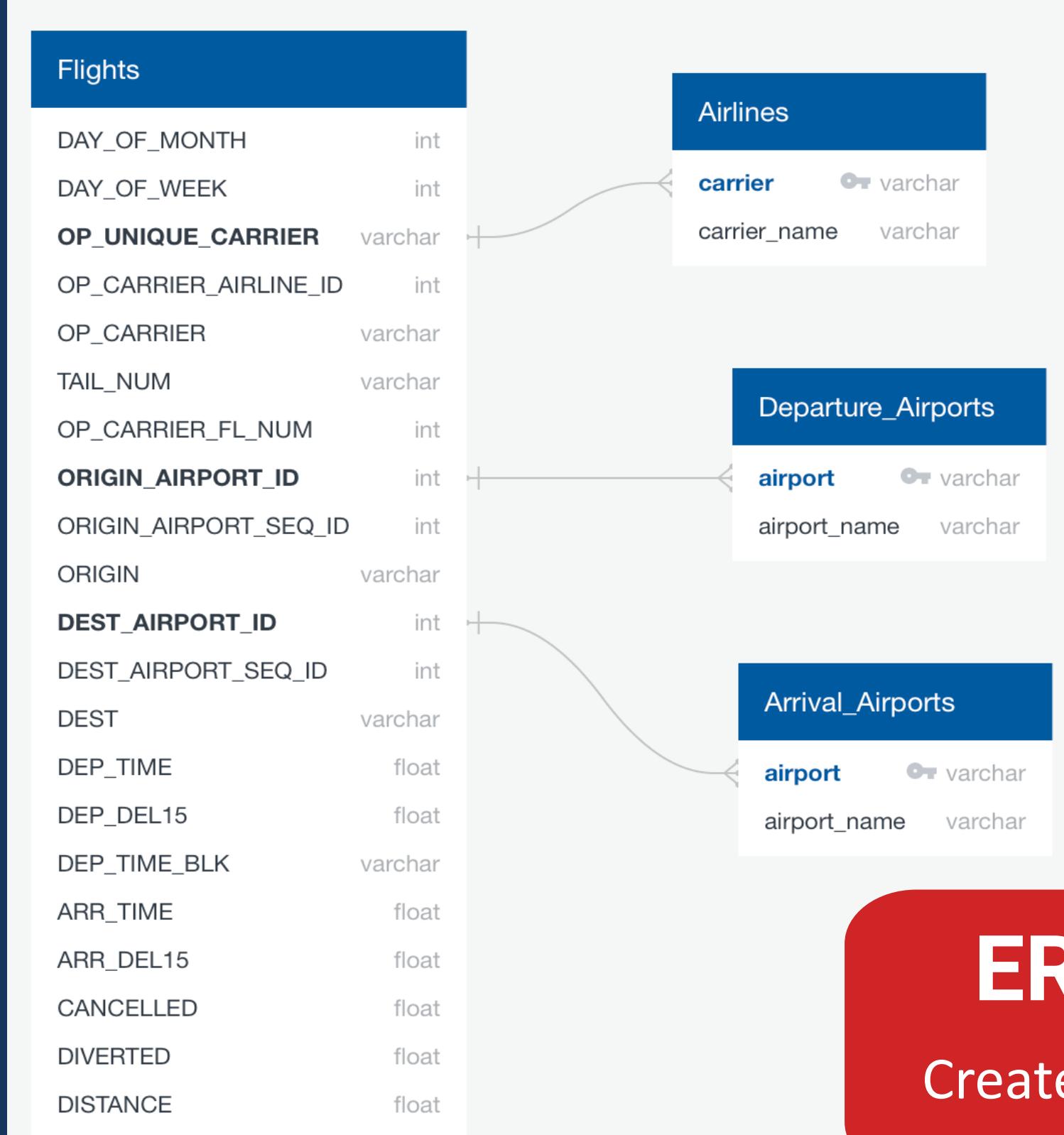




Database

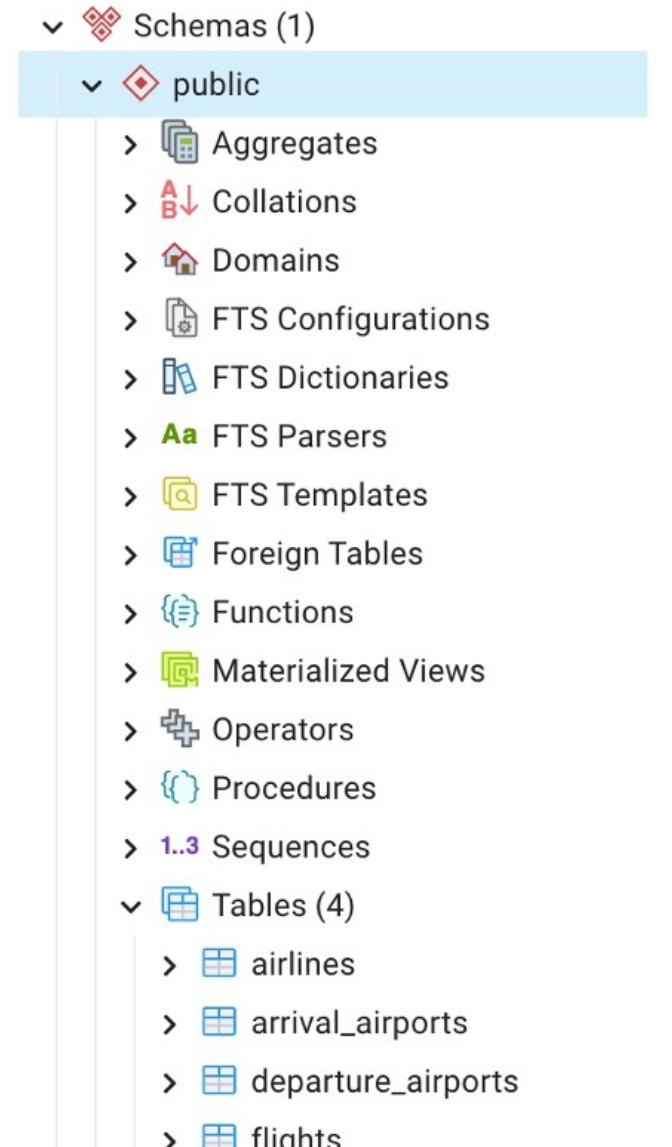
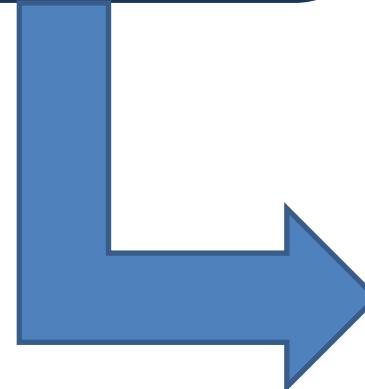
1. A database was setup using AWS.
2. ER Diagrams were created using Quick DBD
3. PgAdmin – Postgress SQL Database Management tool
4. Flight Delay Database Query Tables

*Samples of work will be shown on the next few slides.



PgAdmin

Postgress SQL Database Management tool



ER Diagrams

Created using Quick DBD

ANALYSIS & RESULTS

Best & Worst Airports: Arrival Delays and Departure Delays

Best

1. ???
2. ???
3. ???
4. ???
5. ???
6. ???
7. ???
8. ???
9. ???
10. ???

Worst

1. ???
2. ???
3. ???
4. ???
5. ???
6. ???
7. ???
8. ???
9. ???
10. ???



ANALYSIS & RESULTS

Best & Worst Airlines: Arrival Delays and Departure Delays



Best

1. ???
2. ???
3. ???
4. ???
5. ???
6. ???
7. ???
8. ???
9. ???
10. ???

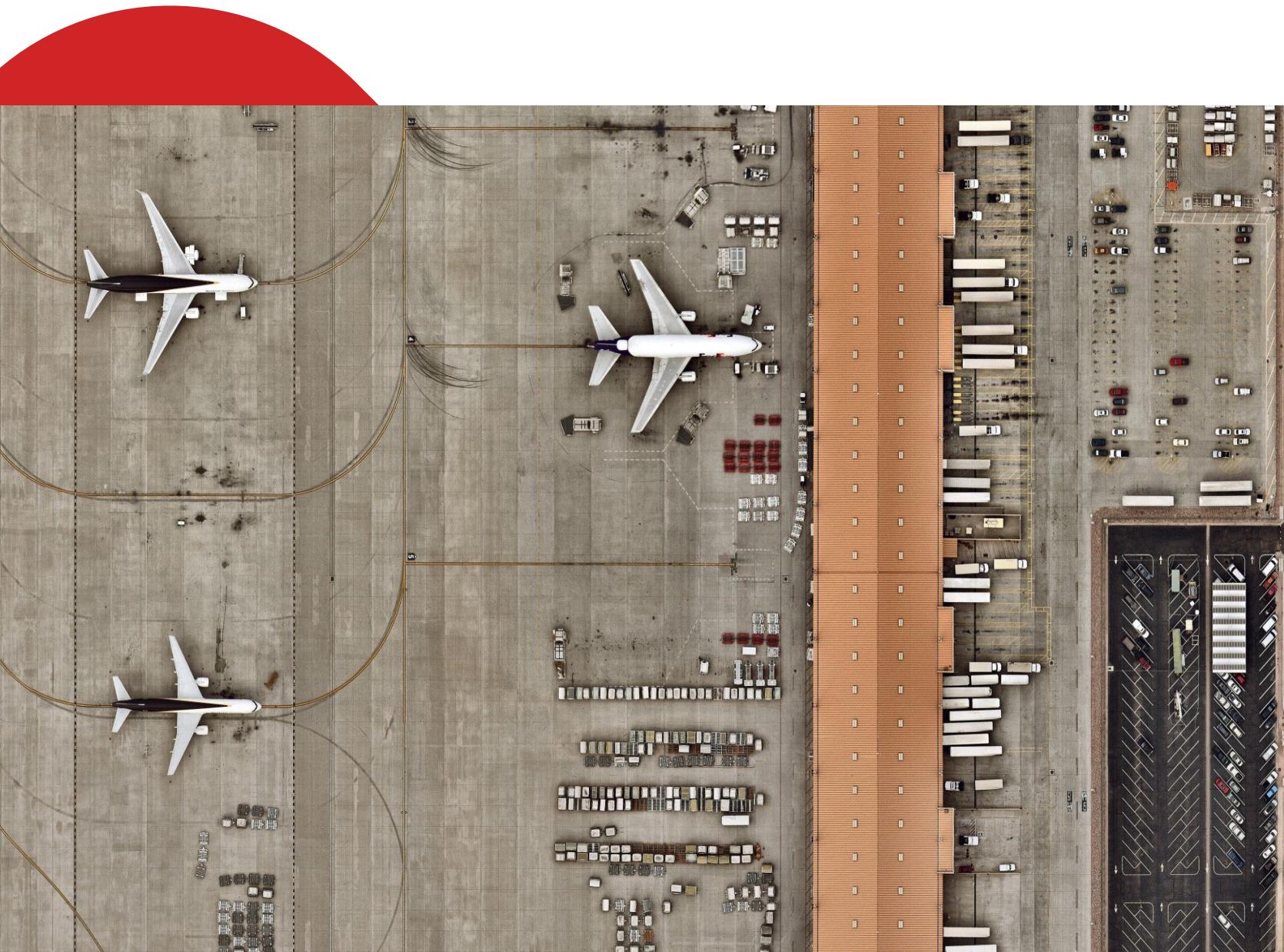
Worst

1. ???
2. ???
3. ???
4. ???
5. ???
6. ???
7. ???
8. ???
9. ???
10. ???



RECOMMENDATIONS

Future Analysis & What we would have done differently?



- ✓ Get datasets on actual the causes of flight delays (i.e. weather, maintenance issues, over booking, etc...)
- ✓ Study the impact of flight delays on airline carriers (i.e. costs, loss of passengers and business, etc...)
- ✓ Collect data to study the impact of flight delay trends by months and/or years
- ✓ Use a different Machine Learning Model for testing
- ✓ ???
- ✓ Have we addressed "Future Analysis"???

THE END!

**Thank you all for
your time and
attention!**

