



GROUP 4 FINAL PROJECT PRESENTS:

Home Flight ☰

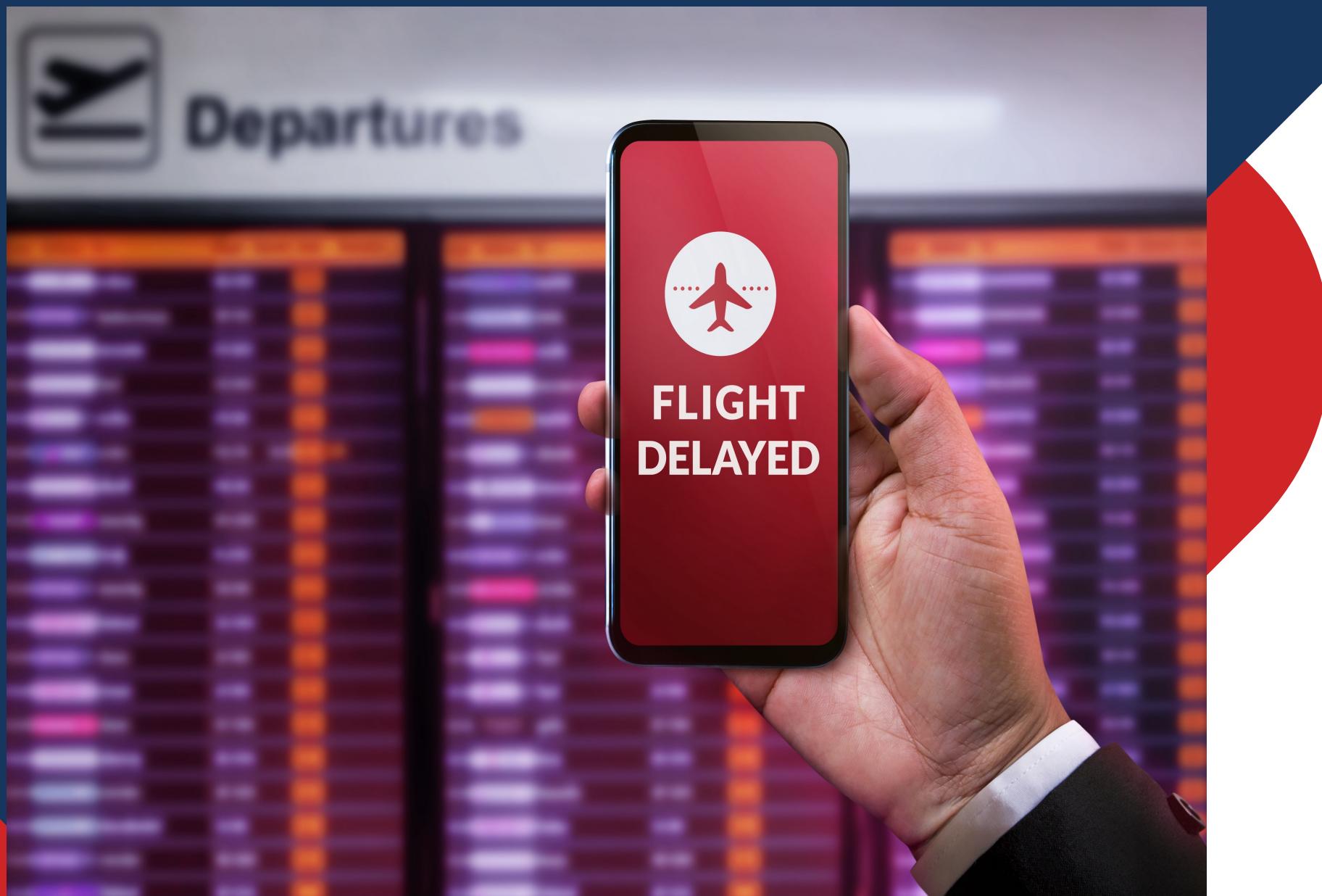
FLIGHT DELAYS:

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



Group Members:
Kenan Foust
John Harris
Marcia Joseph





Project Overview

January 2020 Flight Delay Analysis

1. Predictive Problem
Ability to predict the arrival of flight delays with real world data
2. Determine the Best and Worst Airlines for flight delay arrivals and departures
3. Determine the Best and Worst Airports for flight delay arrivals and departures

Flight Delays

Why we chose this topic?

- We thought it would be interesting!
- Flight delays are prevalent and a hot topic in all news cycles.
- FYI - A flight is considered delayed if it arrived at (or departed) the gate 15 minutes or more after the scheduled arrival (departure) time as reflected in the Computerized Reservation System.
- The purpose of this project is to use exploratory data analysis and to build machine learning models to predict flight delay arrivals for airlines and airports in the United States.
- Share our Top 10 list of Best and Worst Airlines and Airports before you plan your next trip/vacation.



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 of flight delays by Airline and Airport





Technology, Languages, Tools, & Algorithms Used for Project

- Python
 - ???
- Jupyter Notebook
- Tableau
- CSV Editor
- PgAdmin
- Quick DBD
- PowerPoint
- Github
- AWS





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

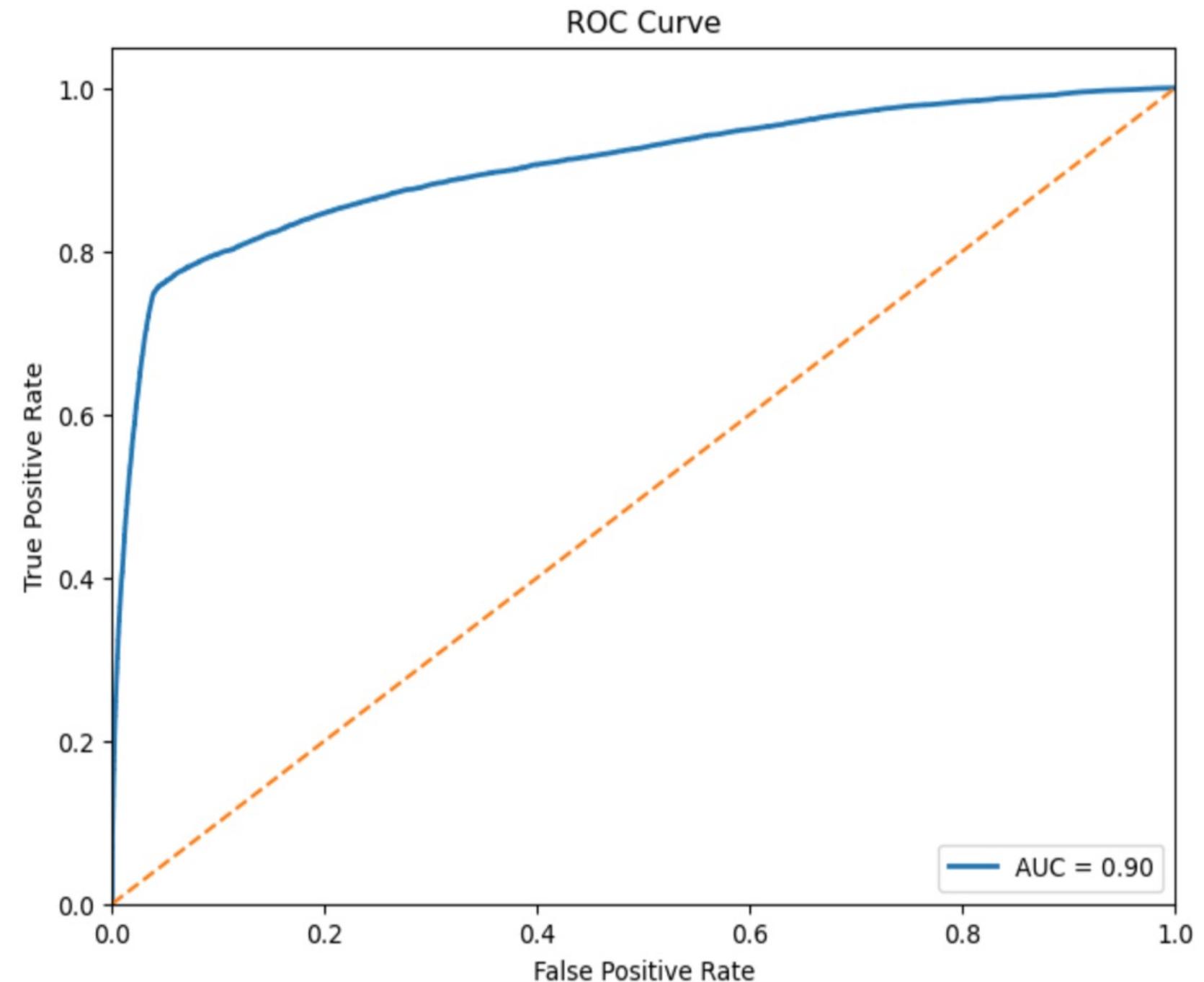
```
Training Accuracy: 0.9320879240072255
Testing Accuracy: 0.932100722545764

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
```

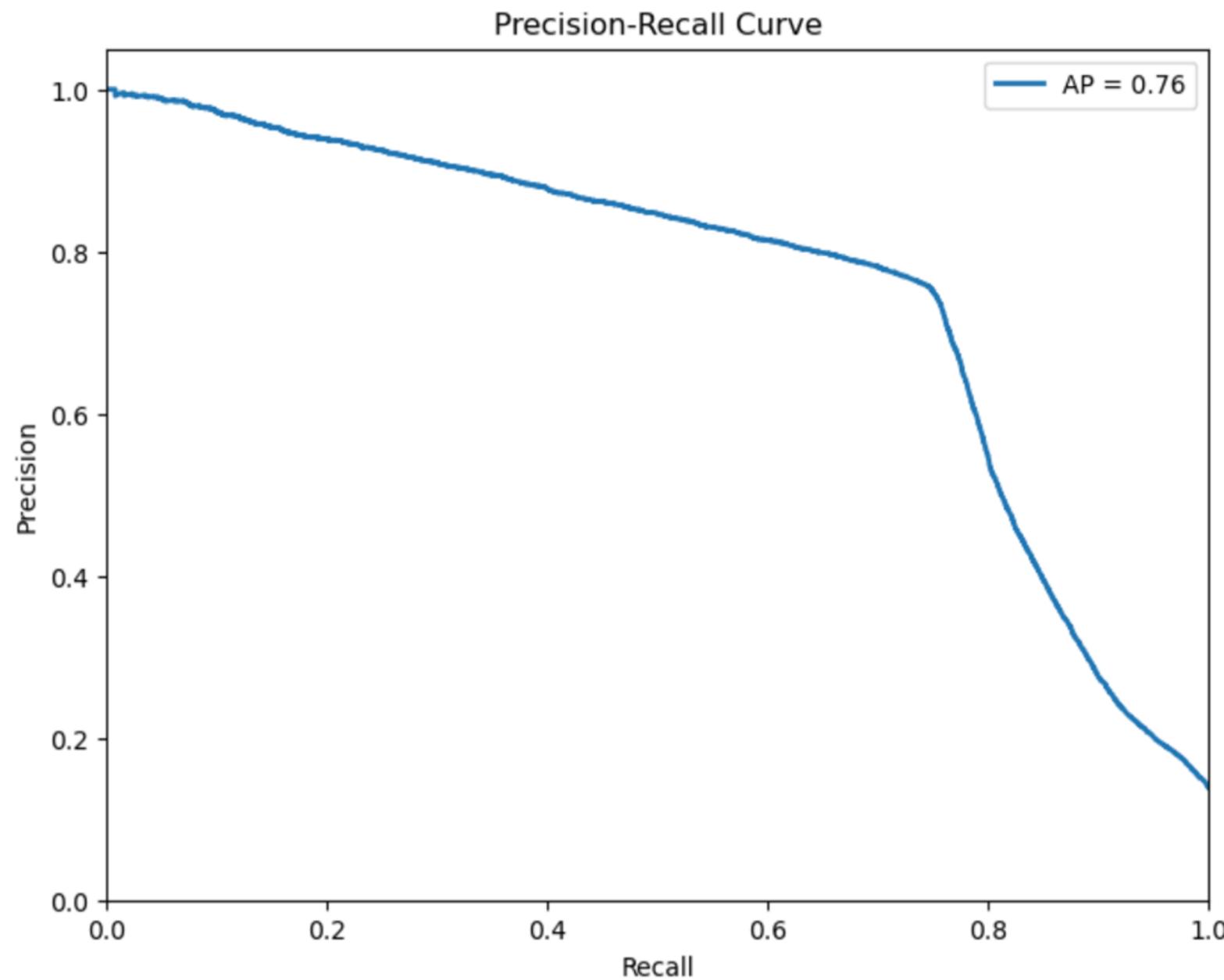
2. ROC Curve



Gradient Boosting Classifier



3. Precision-Recall Curve



4. Confusion Matrix

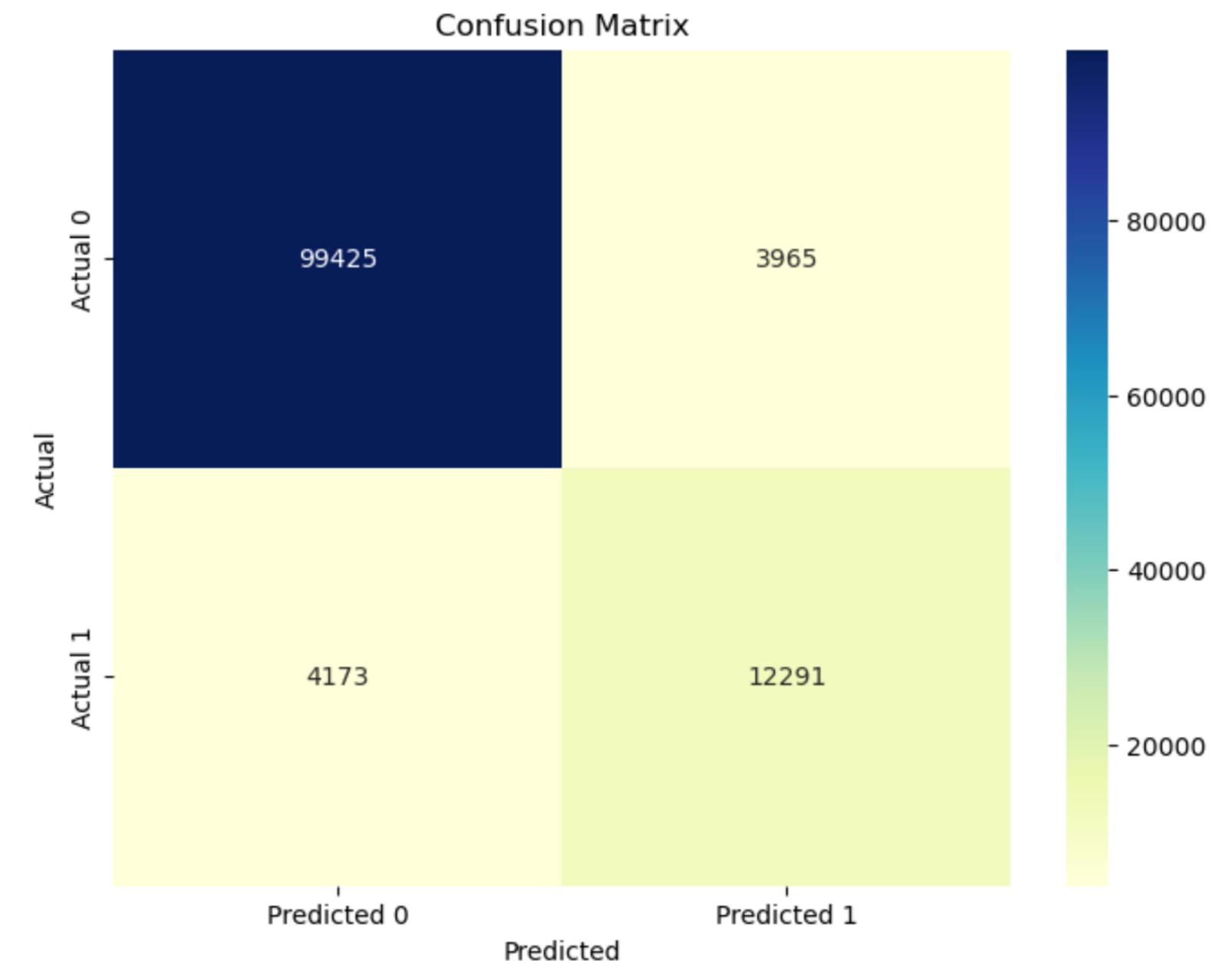


Tableau Visualization Dashboard

Tableau Dashboard

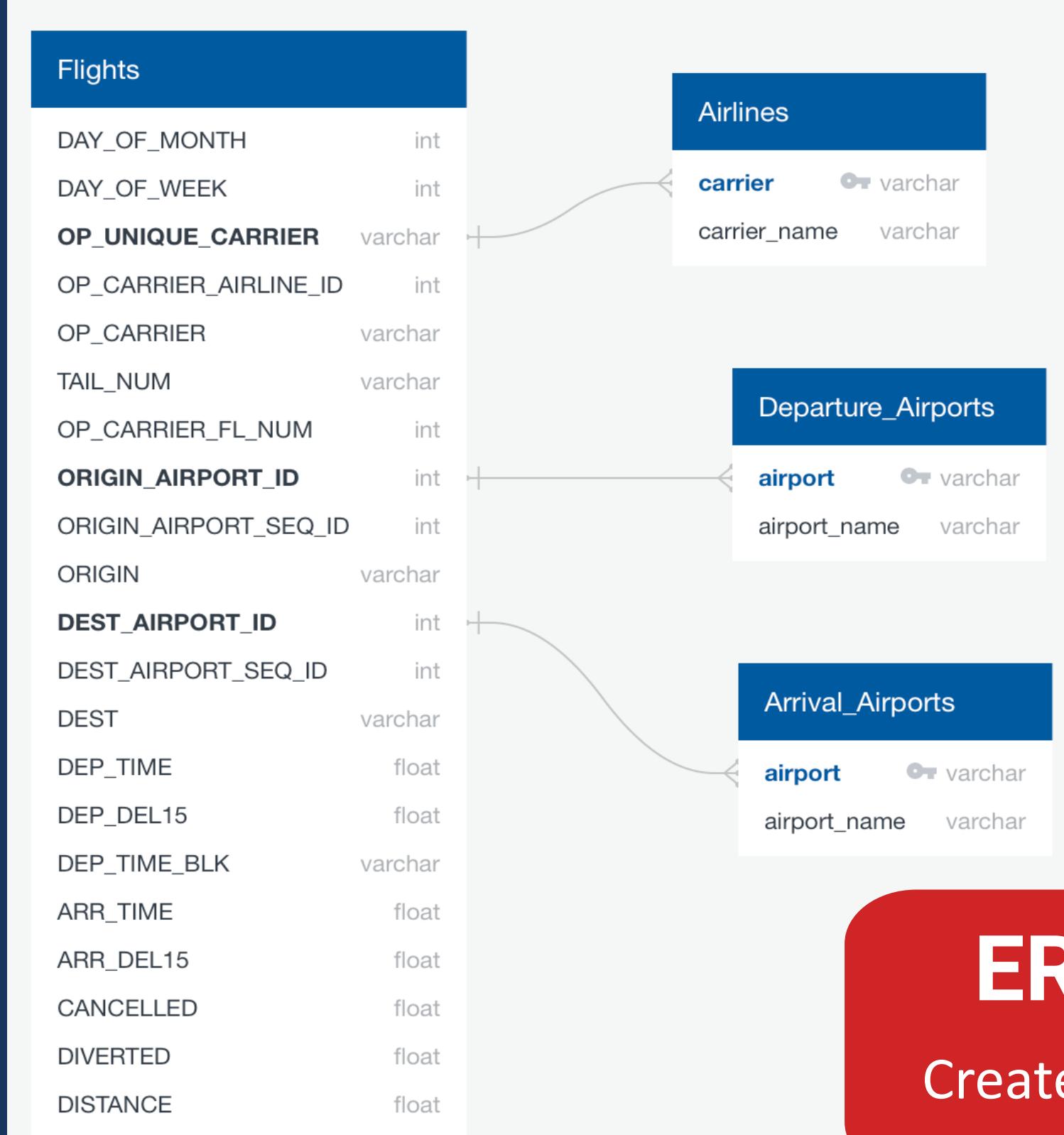




Database

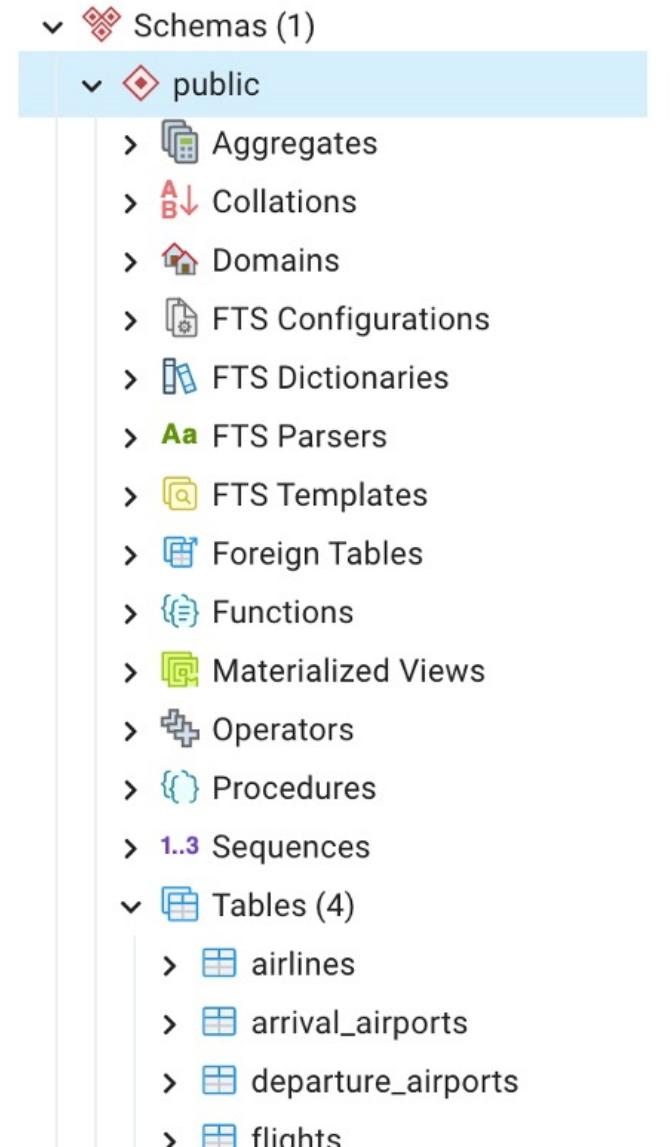
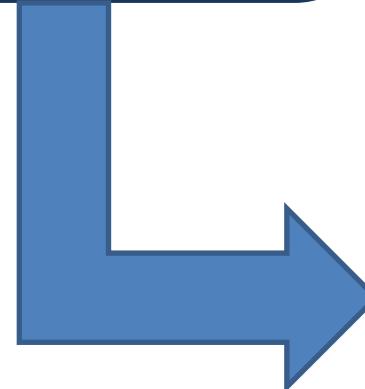
1. A database was setup using AWS
2. ER Diagrams were created using Quick DBD
3. PgAdmin – Postgress SQL Database Management tool
4. Flight Delay Database Query Tables

*Samples of work will be shown on the next slide.



PgAdmin

Postgress SQL Database Management tool



ER Diagrams

Created using Quick DBD

ANALYSIS & RESULTS

- Best Arrival Airports
- Worst Arrival Airports
- Best Departure Airports
- Worst Departure Airports
- Best Airlines
- Worst Airlines



ANALYSIS & RESULTS

Best Airport Arrivals

No.	Airport Name	Total Flights	Delayed Flights	Percentage Delayed
1.	Owensboro, KY: Owensboro Daviess County Regional	9	0	0%
2.	St. Cloud, MN: St. Cloud Regional	13	0	0%
3.	Devils Lake, ND: Devils Lake Regional	47	1	2%
4.	Twin Falls, ID: Joslin Field - Magic Valley Regional	89	4	4%
5.	Lewiston, ID: Lewiston Nez Perce County	64	3	5%
6.	Butte, MT: Bert Mooney	58	3	5%
7.	Watertown, SD: Watertown Regional	56	3	5%
8.	Pocatello, ID: Pocatello Regional	92	5	5%
9.	Cedar City, UT: Cedar City Regional	54	3	6%
10.	Hilo, HI: Hilo International	535	30	6%



ANALYSIS & RESULTS

Worst Airport Arrivals

No.	Airport Name	Total Flights	Delayed Flights	Percentage Delayed
1.	Pago Pago, TT: Pago Pago International	9	4	44%
2.	Kodiak, AK: Kodiak Airport	50	18	36%
3.	Riverton/Lander, WY: Riverton Regional	50	17	34%
4.	Sitka, AK: Sitka Rocky Gutierrez	81	27	33%
5.	North Bend/Coos Bay, OR: Southwest Oregon Regional	21	7	33%
6.	Aspen, CO: Aspen Pitkin County Sardy Field	860	269	31%
7.	Ogdensburg, NY: Ogdensburg International	49	15	31%
8.	Meridian, MS: Key Field	60	18	30%
9.	Dubuque, IA: Dubuque Regional	86	25	29%
10.	Phoenix, AZ: Phoenix - Mesa Gateway	487	141	29%



ANALYSIS & RESULTS

Best Airport Departures

No.	Airport Name	Total Flights	Delayed Flights	Percentage Delayed
1.	Owensboro, KY: Owensboro Daviess County Regional	9	0	0%
2.	Quincy, IL: Quincy Regional-Baldwin Field	10	0	0%
3.	Cody, WY: Yellowstone Regional	1	0	0%
4.	Cedar City, UT: Cedar City Regional	54	3	6%
5.	Casper, WY: Casper/Natrona County International	105	4	4%
6.	Lewiston, ID: Lewiston Nez Perce County	64	3	5%
7.	Butte, MT: Bert Mooney	58	3	5%
8.	Lake Charles, LA: Lake Charles Regional	153	8	5%
9.	Jamestown, ND: Jamestown Regional	74	4	5%
10.	Brunswick, GA: Brunswick Golden Isles	86	5	6%



ANALYSIS & RESULTS

Worst Airport Departures

No.	Airport Name	Total Flights	Delayed Flights	Percentage Delayed
1.	Adak Island, AK: Adak	9	3	33%
2.	Pago Pago, TT: Pago Pago International	9	3	33%
3.	Ogden, UT: Ogden-Hinckley	9	3	33%
4.	North Bend/Coos Bay, OR: Southwest Oregon Regional	21	7	33%
5.	Aspen, CO: Aspen Pitkin County Sardy Field	813	247	30%
6.	Portsmouth, NH: Portsmouth International at Pease	20	7	33%
7.	Sioux City, IA: Sioux Gateway Brig Gen Bud Day Field	108	32	30%
8.	Riverton/Lander, WY: Riverton Regional	48	14	29%
9.	Ogdensburg, NY: Ogdensburg International	48	25	29%
10.	Marquette, MI: Sawyer International	92	26	28%



ANALYSIS & RESULTS

Best Airlines



No.	Carrier Name	Total Flights	Delayed Flights	Percentage Delayed
1.	Southwest Airlines Co.	107,708	10,321	10%
2.	Hawaiian Airlines Inc.	7,015	688	10%
3.	Delta Air Lines Inc.	79,928	8,459	11%
4.	Endeavor Air Inc.	22,955	2,515	11%
5.	Republic Airline	28,826	3,428	12%
6.	United Air Lines Inc.	48,019	6,235	13%
7.	Spirit Air Lines	17,333	2,334	13%
8.	Frontier Airlines Inc.	12,312	1,795	15%
9.	JetBlue Airways	24,621	3,675	15%
10	American Airlines Inc.	75,472	11,348	15%

ANALYSIS & RESULTS

Worst Airlines



No.	Carrier Name	Total Flights	Delayed Flights	Percentage Delayed
1.	PSA Airlines Inc.	23,999	5,166	22%
2.	Alaska Airlines Inc.	20,783	4,393	21%
3.	Allegiant Air	7,663	1,538	20%
4.	Envoy Air	25,284	4,979	20%
5.	Mesa Airlines Inc.	17,473	3,118	18%
6.	ExpressJet Airlines Inc.	10,376	1,601	15%
7.	SkyWest Airlines Inc.	69,501	10,692	15%
8.	American Airlines Inc.	75,472	11,348	15%
9.	JetBlue Airways	24,621	3,675	15%
10	Frontier Airlines Inc.	12,312	1,795	15%

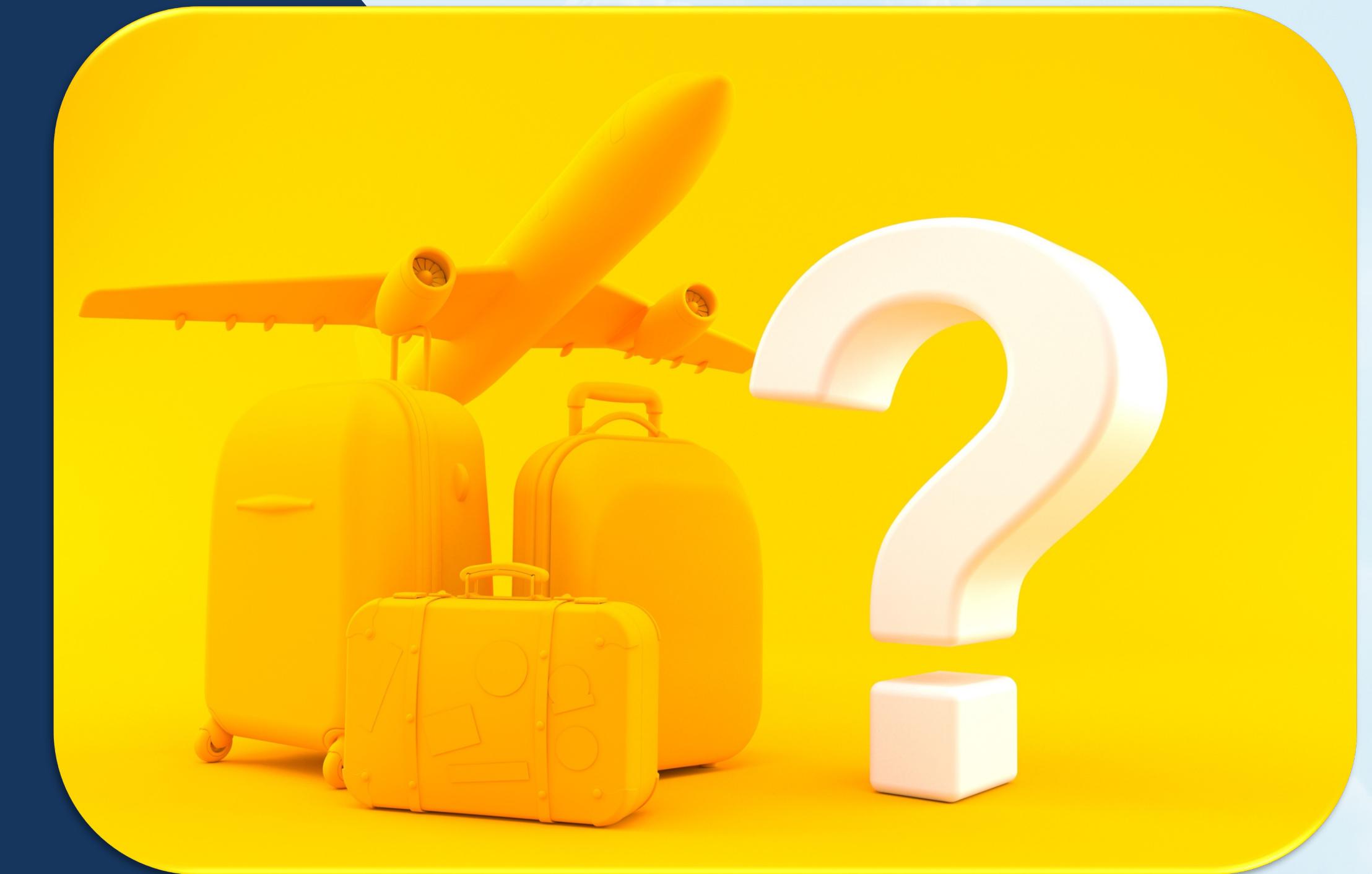
RECOMMENDATIONS

Future Analysis & What we would have done differently?



- ✓ Get datasets on the actual 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, years, time of day, etc...
- ✓ Use a different Machine Learning Model for testing
- ✓ Best and worst times of day for arrivals analysis
- ✓ Have we addressed "Future Analysis"???

Questions And Answers



THE END!

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

