AIRPLANE Delay Prediction

For Improved Airlines Operations

Team Members

Name Deekshith Sagar Rangapuram
UBIT Name: drangapu
Mail Id. drangapu@buffalo.edu

Name Indushree Byrareddy
UBIT Name: indushre
Mail Id. indushre@buffalo.edu

Name Akhil Chaitanya Ghanta
UBIT Name: aghanta
Mail Id. aghanta@buffalo.edu

TABLE OF CONTENTS

01

PROBLEM STATEMENT / OBJECTIVES

02

EXPLORATORY DATA ANALYSIS

03

GUI

04

Feedback & Future Work

01

PROBLEM STATEMENT / OBJECTIVES

PROBLEM STATEMENT

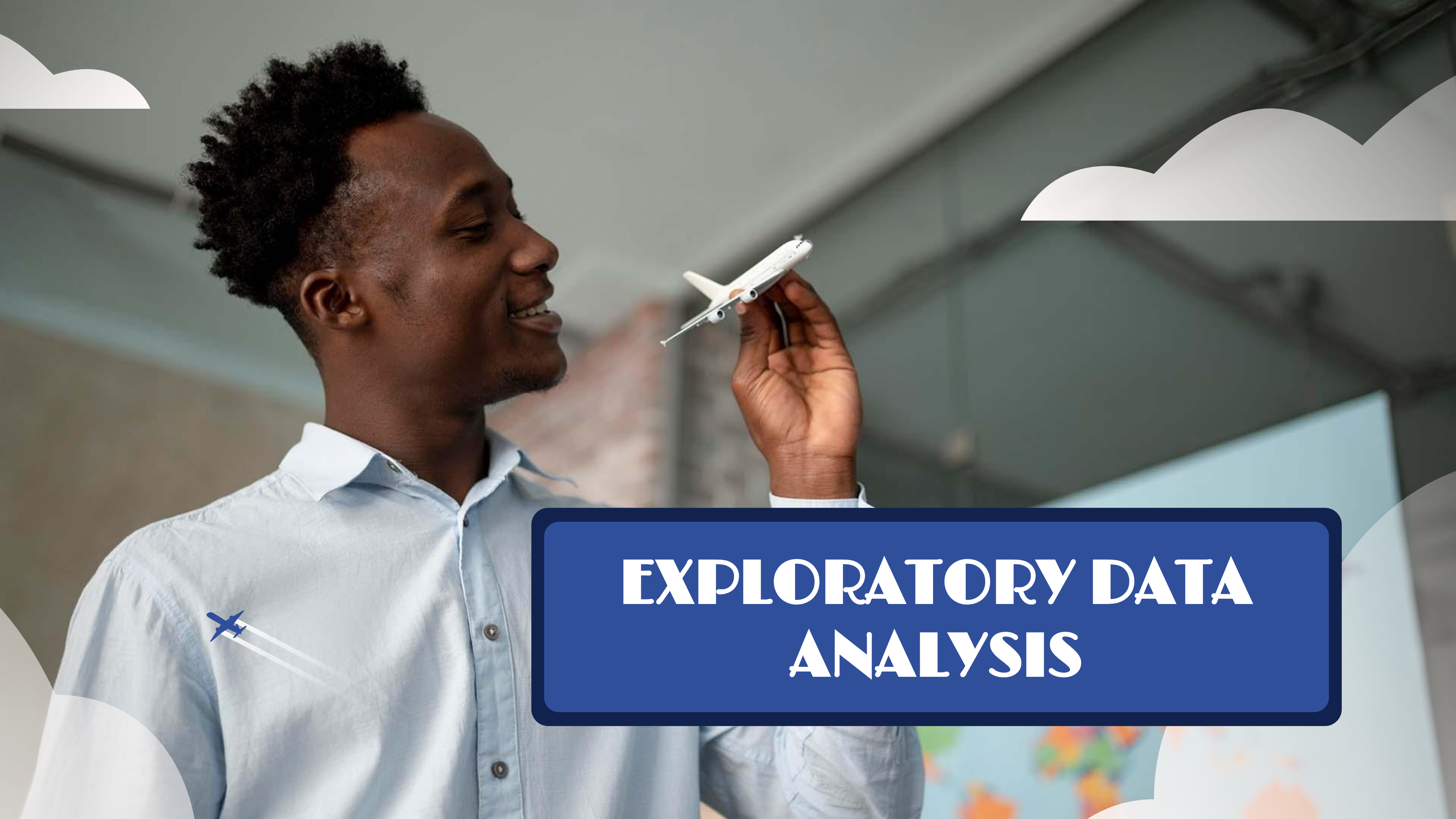
Data-Driven Flight Delay Forecasting to Improve Airlines Operations, addresses the persistent challenge of flight delays that impact airlines worldwide, affecting customer satisfaction and operational costs.

OBJECTIVES

❖ To develop a machine learning model that can accurately predict whether the flight is delayed or not & by how much time.
❖ To create a user interface application that allows users to input journey details and receive a prediction about flight status.
❖ To evaluate the performance of the machine learning model using a variety of metrics, including accuracy and score.
❖ To assess the usability and user experience of the user interface application, including its ease of use and the clarity of the prediction results.
❖ To optimize the machine learning model for robustness and generalizability, using techniques such as Regularization, Feature Engineering and hyperparameter tuning.

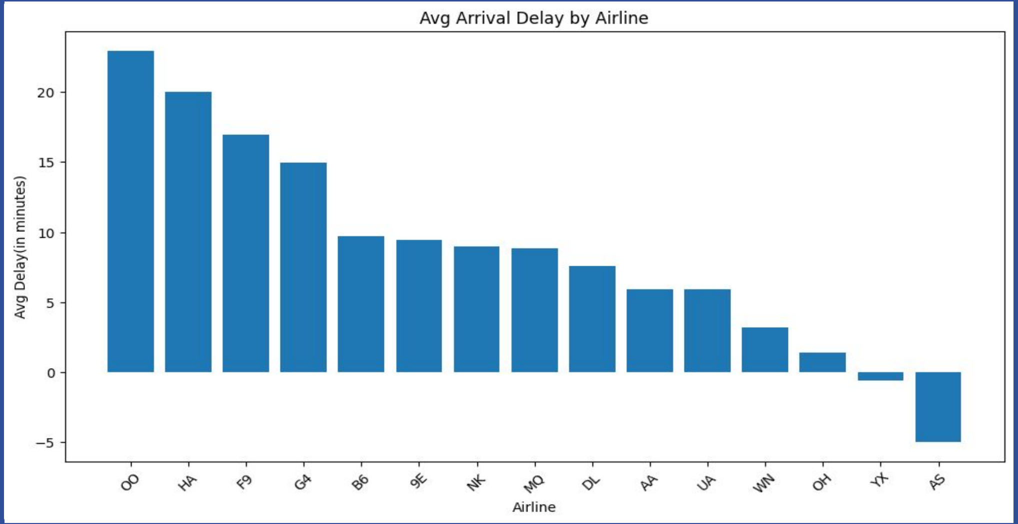
"The only advantage of a flight delay is the possibility of exploring an airport you've never been to before."

—SOMEONE FAMOUS



EXPLORATORY DATA ANALYSIS

Avg Delay Per Airlines

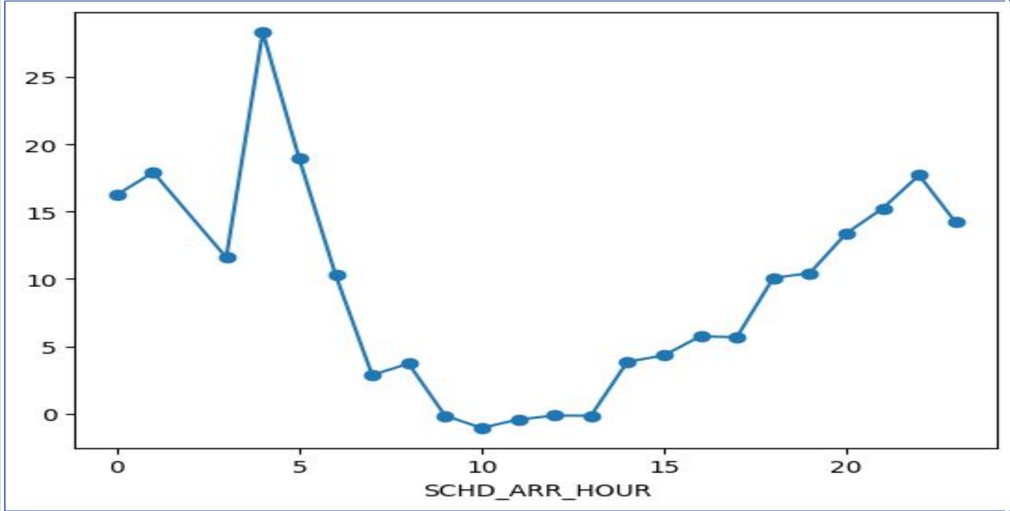


We investigated the typical delay times for several airlines, to analyze how different airlines operate in terms of arrival delays and how they affect the traveling experience.

A bar plot shows how average arrival delays vary across different airlines, making it easier to see which airlines, on average, have longer or shorter arrival delays. The plot is also sorted in descending order , Airlines 'OO' have the highest average delay time & airline 'AS' have the lowest average arrival delay time

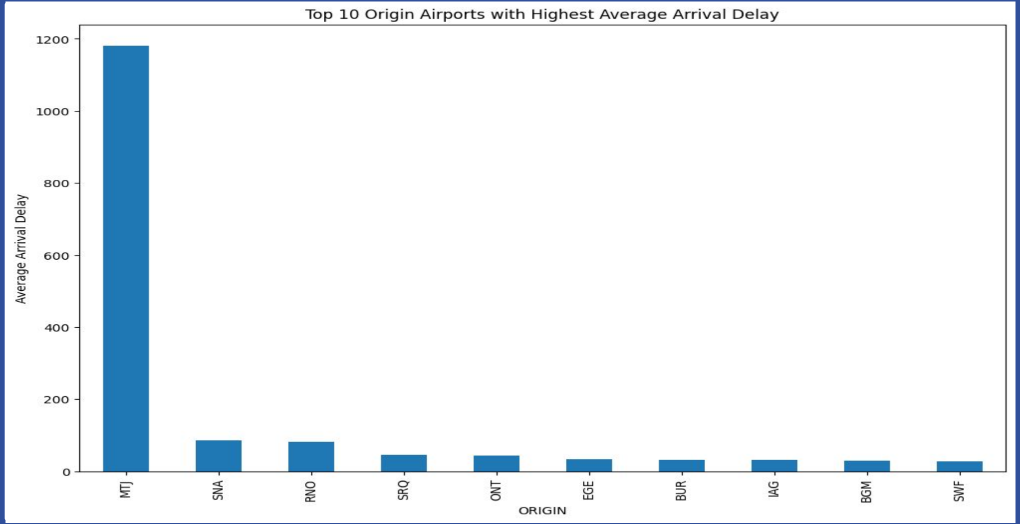
Avg Delay For Each Hour

The figure shows probable peak delays during rush hours and identifies hourly patterns in airline arrival delays.The correlation between scheduled arrival times and typical arrival delays is examined in this section. We have taken the scheduled arrival time for each flight from the dataset, which contains flight data. From the visualization we can say that most delay happens during 0000 to 0500 hrs

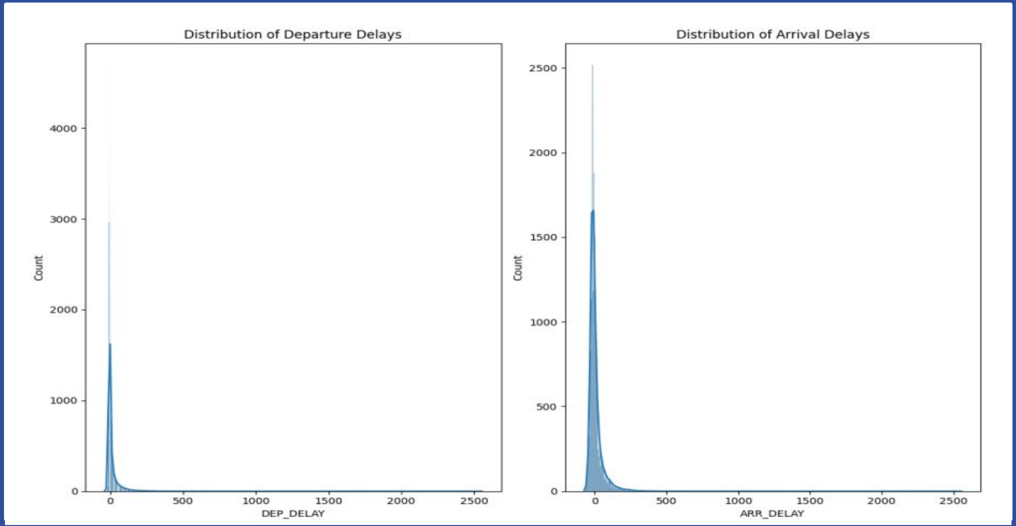


Top 10 Airports With High Avg Delays

The histogram, just to check how is the average delay with respect to different airports from where flights land. The delay in arrival is around the same for all the airports, except for MTJ Airport, which has a very high average Arrival delay. This is not resulting in any trend thus the dataset has some extreme values that might need to be addressed later.

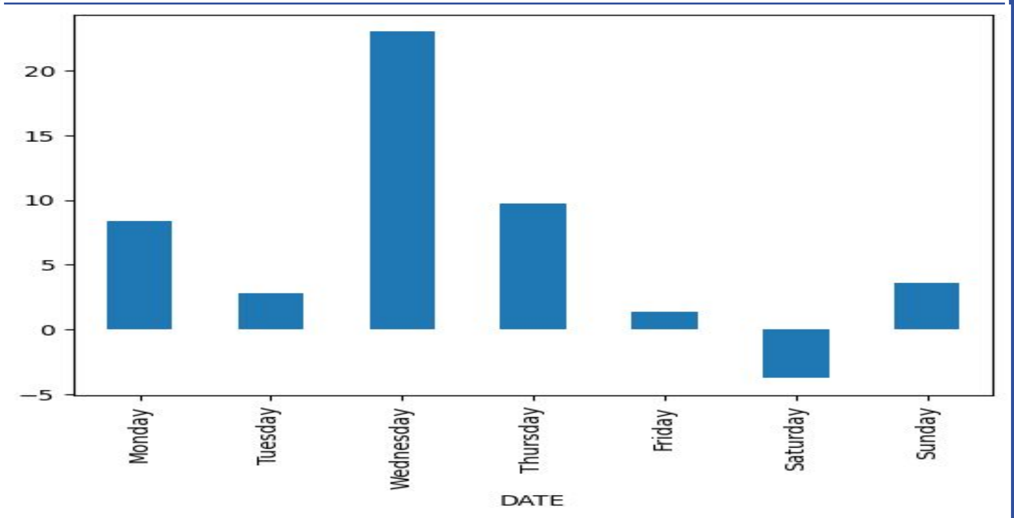


Distribution Of Delays



Histogram plots of delays give the distribution of delays. Which helps in analyzing the central tendency, skewness, Extreme values and how spread are the values of delay. Both Departure and Arrival delays follow the same distribution pattern, both are very narrowly distributed and skewed towards right. There are no such outliers that are very extreme, hence there is no need to find any exceptional circumstances that are impacting delays only at certain times like weather impacts.

Average Delays For Each Day Of The Week



Now to check how the arrival delay is with respect to the day of the week. It is expected that day of the week might have an impact on flight delays and is evident from the below graph. We can say from the graph that more delays are seen on wednesday. This varying probability of average delay based on the day of the week seems like this could be the potential column in further analyzing or arriving at any predictions. Variation along the days of the week is not consistent since there are slight extremes in average delay.

MODEL EVALUATED - PHASE 2

Sl.No	ML Algorithm	MSE	RMSE	R2	MAE
1	Random Forest Regression	155.327	12.463	0.96	8.330
2	Linear Regression	0.037569	0.1938	0.999	0.0486
3	Support vector Regression	64.41996	8.0262	0.9854	2.8262
4	Decision tree Regressor	79.4868	8.9155	0.9820	4.369
5	Gradient Boosting Regressor	68.14	8.25	0.9845	5.39
6	XG Boost	217.1498	14.7360	0.9508	8.9006

LINEAR REGRESSION

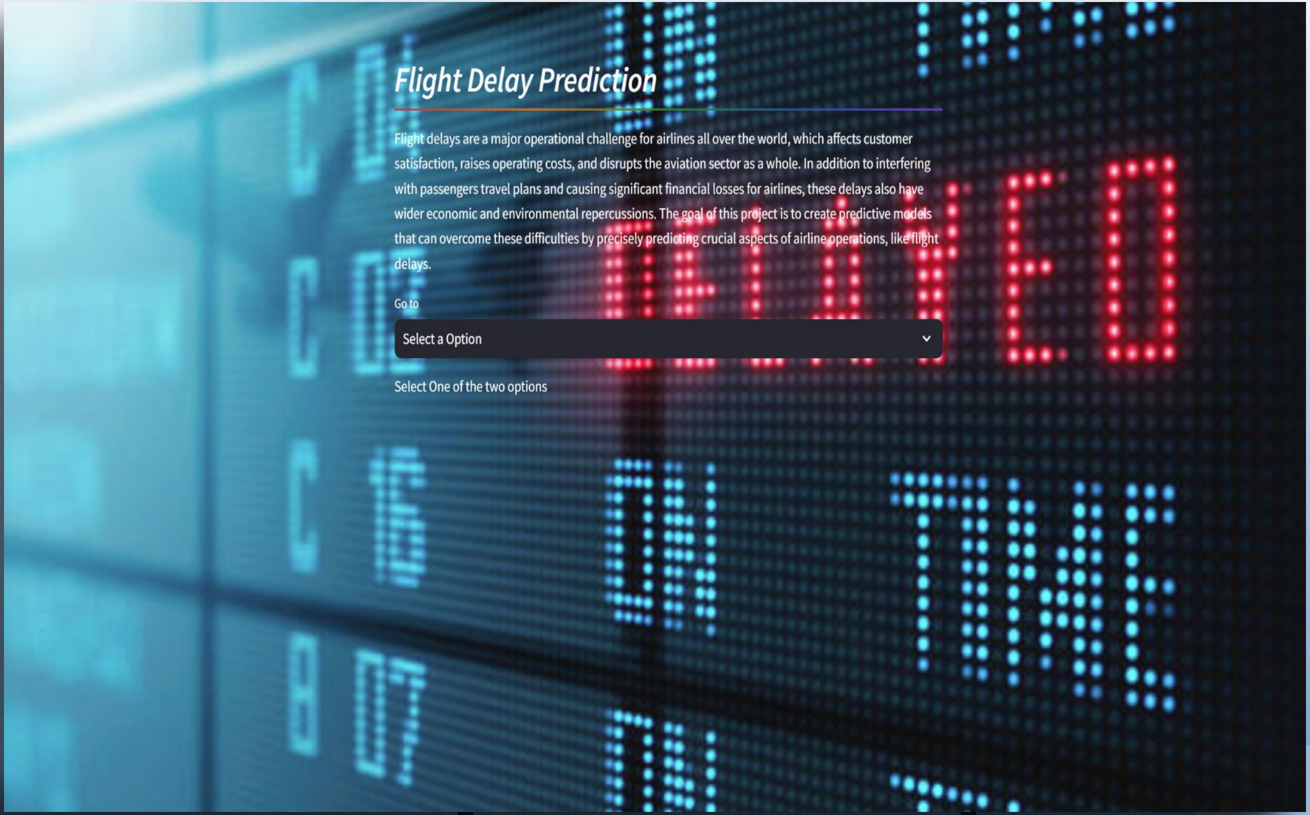
Based on the provided evaluation metrics, Linear Regression is the optimal choice for our flight delay prediction model. Its high R2 score of 0.999 showcases a good level of variance explanation, signifying an efficient fit to the data compared to other models. Moreover, low Mean Absolute Error (MAE) of 0.0486 signifies its precision in predicting the actual arrival delays, outperforming all other models significantly. Both R2 and MAE metrics, demonstrates its capability to predict flight delays with good accuracy and minimal error. Despite the complexity of other models like Random Forest, Gradient Boosting, or XGBoost, Linear Regression performs better by providing comparatively good accuracy with good R2 and exceptionally low MAE. The negligible MSE and RMSE values reinforce the model's precision in minimizing errors when predicting flight delay values. Given the massive datasets in airline operations, Linear Regression's computational efficiency allows us to handle this volume well, enabling faster training and prediction.

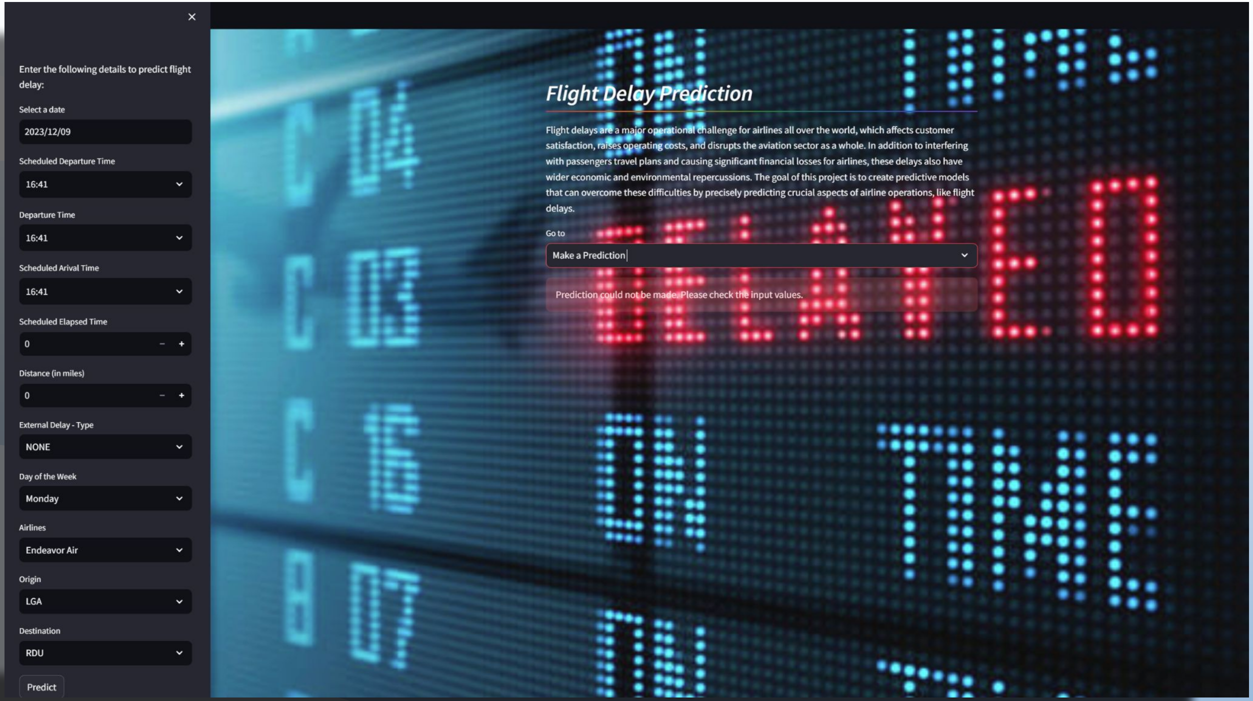
03

Graphical User Interfaces

On the First page user is presented with 2 options to select

- Make a prediction
- Or see the Visualization





If the user selects the 'Make a Prediction' option, they will need to fill out certain fields visible on the left-hand sidebar in order to make a prediction about the flight delay.

If the user selects the 'Visualization' option, they can view trends such as which airlines are most frequently delayed and at which hour most aircraft are delayed, among other insights.

Feedback & Future Work

Feedback & Future Work

Training can be extended over multiple years to capture long-term trends and seasonal shifts, similar to how we understand changing patterns over time. Integrating weather data based on user location could improve predictions, similar to how we consider the impact of weather on our plans. Regular updates or real-time data integration would ensure that predictions remained current, much like keeping up with the latest news. A redesigned user interface with interactive elements and live updates could provide users with a more engaging experience, such as map exploration or real-time flight status updates.


THANK YOU !!!