# **CAPSTONE Project – Berkley Haas College of Engineering March 2024 Cohort**

# **Title: Flight Delay Predictions**

## Objective:

To use machine learning algorithms to predict flight delays based on various factors such as:

* Flight status
* Weather
* Air traffic
* Historical data

## Business Use Case:

### Airline Route Optimization

**Proactive Rerouting:** By predicting delays, airlines can proactively reroute flights to avoid congested airspace or airports experiencing delays. This can help minimize the impact on overall flight schedules and reduce the total delay time.

**Weather Avoidance:** Predicting weather-related delays allows for route adjustments to avoid adverse weather conditions, ensuring safer and more efficient flights.

### Airline Operations Management

**Scheduling Adjustments:** Delay predictions help airlines adjust their schedules dynamically, reallocating resources like gates, ground crew, and equipment to manage disruptions effectively.

**Resource Allocation:** Knowing potential delays in advance allows airlines to allocate resources more efficiently, such as repositioning aircraft and crew to minimize the impact on subsequent flights.

**Passenger Communication:** With accurate delay predictions, airlines can inform passengers in advance, and manage expectations, and provide better customer service, such as rebooking options or compensation.

## Example:

By predicting flight delays, an AI system can:

1. Suggest alternate flight paths that are less likely to experience delays
2. Provide timely updates and rebooking options to Passengers
3. Adjust flight schedules dynamically to manage disruptions effectively.
4. Allocate resources efficiently to minimize the impact on subsequent flights.

## Key Performance Metrics:

This project is expected to address two key performance metrics incrementally:

1. **Identify Bottlenecks:**
   * Identify bottlenecks in flight schedules.
   * Flag flights that are highly likely to be delayed even before their take-off.
2. **Real-time Dashboard:**
   * Provide a real-time dashboard
   * Identify the percentage of flights in the air that are expected to be delayed by more than 15 minutes.Top of Form

Bottom of Form

## Data Needed:

1. Flight Status Records – Historical data and possible real-time feeds of flight routes, departure and arrival times.
2. Weather Data – Historical data that includes precipitation, temperature, wind speed that impact flight operations.
3. Air Traffic Data – Information on air traffic patterns, airport and hub congestion that lead to flight delays.

### Kaggle 2019 Data Set with needed data:

* <https://www.kaggle.com/datasets/threnjen/2019-airline-delays-and-cancellations?select=raw_data_documentation.txt>
* To get access to more recent data-sets, the cleanup process with their data-sources is in <https://www.kaggle.com/code/threnjen/dataset-cleanup-how-the-train-test-sets-were-made>

### Kaggle’s Data Set Source for any data-sets post- 2019:

1. 5 informational datasets from the Bureau of Transportation Statistics:

* T3\_AIR\_CARRIER\_SUMMARY\_AIRPORT\_ACTIVITY.csv
* B43\_AIRCRAFT\_INVENTORY.csv
* AIRPORT\_COORDINATES.csv
* CARRIER\_DECODE.csv
* P10\_EMPLOYEES.csv

1. Weather datasets from the National Centers for Environmental Information

* Airport\_Weather.csv
* Airport\_list.csv

1. Links for above data sources

* Flight On-Time Status <https://transtats.bts.gov/ONTIME/>
* Air Carrier Ground Crew support with P10 Employees <https://www.transtats.bts.gov/Tables.asp?QO_VQ=EGI>
* Airport Weather data <https://www.ncei.noaa.gov/access>

**Creating New Features from Existing Data to consider these as possibilities:**

**Delay Categories:** Classify delays into categories like No Delay, Moderate Delay, and Severe Delay for both departure and arrival times.

**Time-Based Features:** Extract day of the week and part of the day from departure and arrival times.

**Weather Impact:** Create features for weather conditions at airports, such as visibility, temperature, and overall severity.

**Route-Based Features:** Include normalized or categorized flight distance and duration, and flag busy routes.

**Using Clustering Results:**

* **Cluster Membership:** Add cluster labels as new categorical features to indicate patterns in delays or weather.
* **Anomaly Scores:** Create a feature representing the distance of each flight from its cluster center to identify potential outliers.

**Interaction Features:**

* **Departure-Arrival Interaction:** Combine departure and arrival delays to capture their relationship.
* **Weather and Delay Interaction:** Create interaction terms between weather severity and departure delays.

**Aggregation Features:**

* **Historical Averages:** Calculate average delays for routes, days, or time blocks based on historical data.
* **Rolling Averages:** Implement moving averages for delay times or weather conditions to capture trends.

**Dimensionality Reduction:**

* **PCA:** Use Principal Component Analysis to reduce dimensionality if this results in many features to retain most of the variance.

**Handling Categorical Features:**

* **One-Hot Encoding:** Convert categorical features into one-hot encoded vectors.