# AML - Challenge 1, Airplane Delays

# Context

Every day, in US, there are thousands of flights departures and arrivals (prior to COVID): unfortunately, as you may have noticed yourself, flight delays are not a rare event!!

In this challenge, we play the role of a data scientist working in the travel industry, specifically on air transportation of passengers. We want to **explore** the **data** collected by the *Department of Transportation* (*DoT*) to **understand passengers' behavior**, as well as the **properties** of all **flights**, across several airline companies.

Now, given historical data about flights in the country, including the delay information that was computed a-posteriori (so the ground truth is available), we want **to build a model that can be used to predict how many minutes of delay a flight might experience in the future**. This model should provide useful information for the airport to manage better its resources, to minimize the delays and their impact on the journey of its passengers. Alternatively, astute passengers could even use the model to choose the best time for flying, such as to avoid delays.

## Goal

The goal of this challenge is to construct a model for predicting arrival delays.

## Metric

A leaderboard for this challenge will be ranked using the Root Mean Squared Error (RMSE) between the predicted values and the observed arrival delays.

## Submission File Format

Your submission is a CSV file containing your final model's predictions on the given test data. This file should contain a header and have the following format:

```
ID,ARRIVAL_DELAY
1561528,30.497905409756292
845015,32.75533601062071
717459,34.758212570279404
451496,32.73011833829676
etc.
```