# Initial Investigation and Plan - Lewis

## Overview

The dataset we will be investigating are flight records of 300,261 flight bookings collected from February 11th to March 31st, 2022. The dataset, available on Kaggle.com, is being used under the license of CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, and had 300 thousand unique records with 11 features.

We will need to hypothesis questions which we shall investigate and extract relevant data which will allow us to understand correlations and potentially predict similar results. Questions will include.

* When is the optimum time (days) before booking before price increases drastically?
* Are there considerable differences in price when setting off during times of the day?
* Are certain airlines marketed higher, than competing airlines, if so, what features do they offer, more business class seats, afternoon departures, popular city destinations?
* What airlines have the monopoly out of the 6 cities recorded?

And overall,

* How does the 10 different features affect price?

We will need to analysis each of the features and evaluate if they can help with understanding and answering the questions asked.

## Objectives and expected outputs

We plan on setting achievable objectives which will allow us to monitor our data exploration and put to practice the artificial intelligence techniques we have learnt and apply them to our dataset and target label.

We will split thee dataset into training, validating and testing, as this will allow our artificial intelligence models to be applied appropriately, and due to the scale of our dataset, we will have more opportunity to change the dataset split, and introduce a holdout test set, which we could apply during the end of the project.