## **Predictive modelling**

### **Brief**

Qantas provided me a dataset of airfare observations, with one price per flight per day over a three-year period from 2017 to 2019. The prices are normalised to make them comparable across different flights. These data are provided as a CSV file (**airfares.csv**) with the following fields:

* **Date**: the date the observation was made;
* **ID**: an ID of the flight in question; and
* **Price**: the price on that date.

While individual observations are synthetic, they are representative of actual airfares, with the underlying distributions based on data published by the Bureau of Infrastructure, Transport and Regional Economics (BITRE).

They also provided school holiday data as a CSV file (**holidays.csv**) with the following columns:

1. **Month**: the month the row pertains to.
2. **Term**: an indicator showing the month when school terms finish, with five possible values: 0 (no end of term) and 1-4, indicating the end of term 1-4 respectively. When different states finish in different months, the earliest is taken.

**My work is to:**

1. Write a Python function or module to calculate the average price (over all flights) for each month in the dataset and combine this average price with the school holidays data. This function should output 36 records, one for each month over the 3-year period. Each record should have three fields: the year and month, the average price and the end of school term indicator. Qantas asked me to design this function such that it can scale to handle a significantly larger dataset (e.g. 1000x more records) than the dataset have been given.
2. Build a model of the average monthly price and use this model to forecast of the average monthly price for the period Jan-2022 and Dec-2022 inclusive.