Instructions:

1. Obtain the following open dataset using the provided REST API endpoint:

   https://data.gov.au/data/dataset/international-airlines-airline-by-country-of-port-data/resource/809c77d8-fd68-4a2c-806f-c63d64e69842

2. Answer the following questions:

   1. How many passengers came in and out of Australia for each Airline in each month of 2019?

      - The output data are in the file Answer_1.csv
      - To get this answer I grouped the data using the air line name and the month number, and applied the aggregate sum to the passenvgers_in and passengers_out

   2. In the last 6 months which port country was responsible for the most passengers coming into the country?
      - The output data are in the file Answer_2.csv
      - Getting this data required to calculate the date six months ago
      - Used the month and the year as condition to eliminate all the rows out of this time period.
      - Used the argmax function to find the record with highest passenger coming from a specific port

   3. What is the 3-months rolling average of freight coming into Australia for each airline in 2018?
      - The output data are in the file Answer_3.csv
      - First, I filter the data to the year of interest (2018)
      - Sort the values according to name, year and month
      - Apply grouping and use the aggregate of Rolling window of 3 with mean

4. Which 3 airlines have the highest mail inbound/outbound ratio across the whole dataset?
   - The output data are in the file Answer_4.csv
   - First, I grouped the data according to the airline name and used the aggregate function for summing all the mail_in and the mail_out
   - I filtered out the records of mail_out equals zero
   - Calculate the ratio by dividing the total mail_in by the total mail_out
   - Finally, I sorted the records descendingly according to the calculated ratios and picked the top 3 records

5. **BONUS:** Using the dataset, identify an airline that may have changed name and explain your approach?
   - The output data are in the file Answer_5.csv
   - I grouped the data using the airline name using the aggregate function max year
   - I grouped the data using the airline name using the aggregate function min year
   - I applied inner join on the records with max year and the records with min year, according to the assumption that the company that changed its name was using the old name one year and started to use the other name in the same year.
   - This is a simple approach, it does not specify exactly which airline changed its name, but it will reduce the number of airlines significantly.
   - This approach can be extended to compare the passengers count and freight movement as I expect the same company will have very close performance the next year to the previous one.

Constraints:

- Must be written in Python or JavaScript.
- Allowed to use any open-source packages for either language. (e.g. NumPy)

Deliverables:

- Executable script or interactive notebook for Python. (e.g. ipynb)
- Results and any explanations should be output to a separate CSV or TXT file. Alternatively, they can be displayed in the interactive notebook.
- All source code must be provided, including the code used to interact with the REST API.
- All results, code and documentation should be packaged in a ZIP file. (Attachment limit - 10 MB)

Please send your submission to Alix Booth alix.l.booth@boeing.com by the agreed submission date.