Jamal Haruna's Flight Volume Investigation Report

Within this report, I document and explain my solutions and analysis on the tasks provided to me within the hourly pattern problem sheet.

TASK 1 (CODING)

For this task, I was challenged with creating a simple script that can download data from the OpenSky API for Heathrow over the period of a year.

To complete this task, I first created a virtual environment. I did so as there are many libraries I will be relying upon to complete this task. Once I had completed this, I installed the OpenSky library from their official website and installed it into the virtual environment using pip.

Once this was complete, I began scripting using the OpenSky REST documentation. As I could only download 7 days of data at a time, I iterated 52 times to cover the 52 weeks within the year 2018. I wrote all of these requested data within the JSON file into a CSV file. This will allow me to easily visualise it within Excel or utilising a library such as Pandas.

The implementation for this task can be found within taskOne.py, the file can be run so long as it is run within an environment that has access to the OpenSkyAPI.

TASK 2 (DATA RELATED)

Here, I was tasked with analysing the 2019 data that was included with the challenge. Unfortunately, I was unable to deduce how to solve the first question ("How the departure flight hourly volume changes across the year? Are there any periodicity or seasonality in the data?") and thus attempted the second and third part only.

For the second question, I had to find out the peak hours within the airport throughout the year. To do so, I decided to visualise the data in a bar graph. I created this bar graph using two libraries, Pandas and Matplotlib.

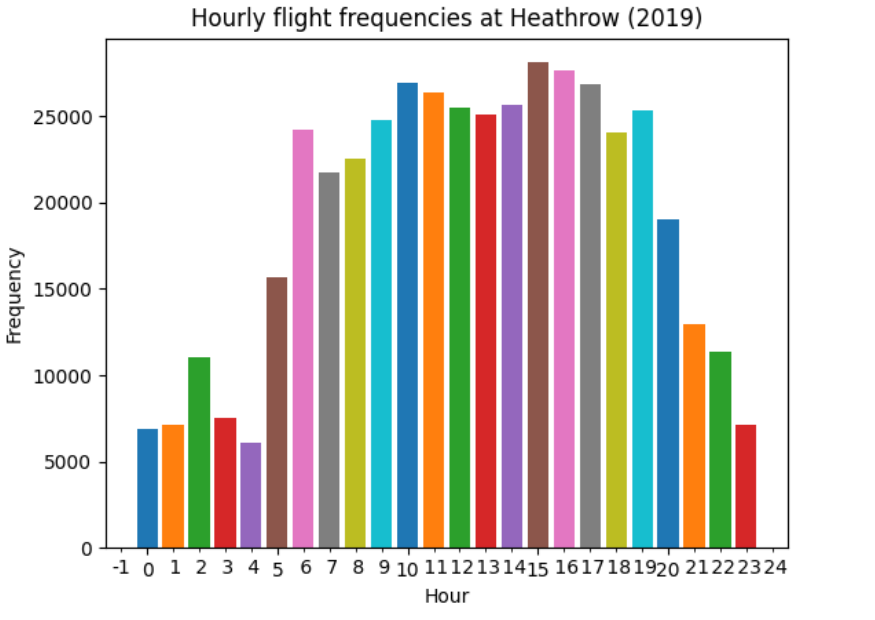
First, I utilised pandas read the included csv file. I then once again used pandas to convert the UNIX timestamps from the firstSeen column into regular GMT datetime. Once I had a regular GMT datetime, I created a pandas series indexed by a time ranging from 00:00 to 23:00. Each hour was initialised with a value of 0:

00:00 0
01:00 0
02:00 0
03:00 0
04:00 0
.
.
.

Once I had created my series, I had to increment each hour every time there was a flight with a matching hour. Thus I iterated through the data, updating the time series each iteration dependent on the hour. Below is the final series output that I received.

```
00:00:00     6908
01:00:00     7105
02:00:00    11035
03:00:00     7560
04:00:00     6076
05:00:00    15678
06:00:00    24195
07:00:00    21788
08:00:00    22544
09:00:00    24794
10:00:00    26966
11:00:00    26389
12:00:00    25497
13:00:00    25073
14:00:00    25684
15:00:00    28129
16:00:00    27697
17:00:00    26906
18:00:00    24099
19:00:00    25325
20:00:00    19068
21:00:00    12986
22:00:00    11376
23:00:00     7134
dtype: int64
```

Once I had completed my hourly frequency table, I wanted to visualise the data using a bar graph so that I could easily spot trends and irregularities. To do so, I utilised matplotlib to generate a bar chart. I chose to visualise the data using a bar chart as it will clearly show differences in hourly data. I considered utilising a line graph, as this would show me the transitions between hourly trends. However, I decided against this as I was purely interested in the hourly accumulated frequencies. Below was the bar chart that was output.



From this bar chart I was able to retrieve several pieces of information:

- Peak time appears to be 3pm – 5pm.These are the hours throughout the year that consistently had the highest number of departures and arrivals. This may be due to the fact that this is often the most active part many people's day.
- Early morning hours (12am – 4am) appear to be the least busy hours, with the lowest overall frequency average throughout the whole chart. This is because any arrivals/departures during this time are considered night flights, of which there are less.

- There is a significant increase in the number of departures and arrivals from 5am. This is due to the fact that flights begin arriving to Heathrow at 4:30am (https://www.heathrow.com/company/local-community/noise/operations/arrival-flight-paths)

Outliers and Anomalies

There was one particular anomaly I noticed with the output. At 2am, there is a large increase in the number of departures/arrivals. The number goes back down for 3am and 4am. There are two reasons that could be causing this:

- A large volume of flights may have been delayed to this time, thus 2am is a common time for these flights to depart.
- This may be due to busy periods within airports across the globe, resulting in a spike in arrivals at Heathrow at 2am.

BONUS

I had attempted the bonus challenge, however, unfortunately the pandas library was unable to read one of the columns from my file. Due to time restraints, I regrettably left this part of the challenge. However, it should theoretically output the hourly frequencies for 2018 by a simple change in the file that is read.