# Come Fly With Me

A visualisation of aviation trends within Australia

Robert A Fraser

# Introduction

Aviation is a growing necessity in keeping our modern world connected. At any moment, hundreds of planes are in the air, and millions of passengers are transiting every month.

Australia has a strong reliance on the aviation industry due to our isolated geography – with no land borders, air travel is key for international passengers to travel in and out of the country. This report will analyse and visualise aviation passenger data within Australia for the past few years, covering both international and domestic flights.

By analysing these trends, I hope to see how the industry has changed and where it is headed in the future; especially after these uncertain times. I also hope to examine how major international hubs connected to Australia have changed.

# Data Source

## Origins

The data sources used in this project are publicly provided by the Bureau of Infrastructure, Transport, and Regional Economics (BITRE); a division of the Department of Infrastructure, Transport, Regional Development and Communications.

BITRE publishes various transportation statistics, including a wide variety of Aviation statistics. This report focuses on Domestic aviation activity and International airline activity. Time series data is provided, with monthly information for routes provided as far back as 1984.

One caveat with this dataset is that reporting of data by Qantas Airways changed in 2003 – for example, a flight reported as Adelaide to London in January 2002 (no direct services between these two cities), would be reported in January 2003 as either Adelaide to Singapore or Melbourne/Sydney to London. This makes it difficult to directly compare data before 2003 to data after 2003. To deal with this in my visualizations, most of my visualizations will only use data from 2003 to 2019 – if required, data before 2003 will not be compared to data after 2003.

The Domestic and International datasets are in very different formats. The International dataset is easier to work with, and most of the visualizations in this report will be based off this. Problems with the domestic dataset include: invalid column header structure, airport codes used instead of city names, and dates being split into two separate columns. The invalid column header structure was fixed manually in Excel before processing.

Additionally, the Domestic dataset only has data in both directions for each row, and each city pair is sorted alphabetically. For example, ABX -> SYD has Sydney as the destination, but SYD -> TSV has Sydney as the origin. This makes the data difficult to work with when it comes to finding city pairs.

## Processing

There was no element of data collection or web scraping in this project.

BITRE provided the data in multiple Excel spreadsheets – for example, the international data was split into 1985-1998, 1989-1993, 1994-1998, 1999-2003, 2004-2008, and 2009-2020. Only data from 2004 onwards was processed.

To combine these spreadsheets together, the Python package **pandas** was used. Multiple spreadsheets are loaded at once using the **read_excel** function, and then the data frames are appended together to create one large data frame. For the Domestic dataset, there was some additional data cleansing which involved standardising the date format.

Since the data is inherently multi-dimensional (Origin, Destination, Time); I also used the package **xarray** to easily work with the dimensions. Xarray provides a powerful n-dimensional data structure with dimensional labelling. These structures are much easier to work with for analysis and visualization purposes.

For the visualisations themselves, continuing with the Python usage, I'm using the library **matplotlib.** I am reasonably unfamiliar with this library, but it should work well with the data.

# Results

## International Traffic

Since this is a large dataset with a wide variety of origins and destinations, it's important to get a look at the bigger picture before jumping in too deep.

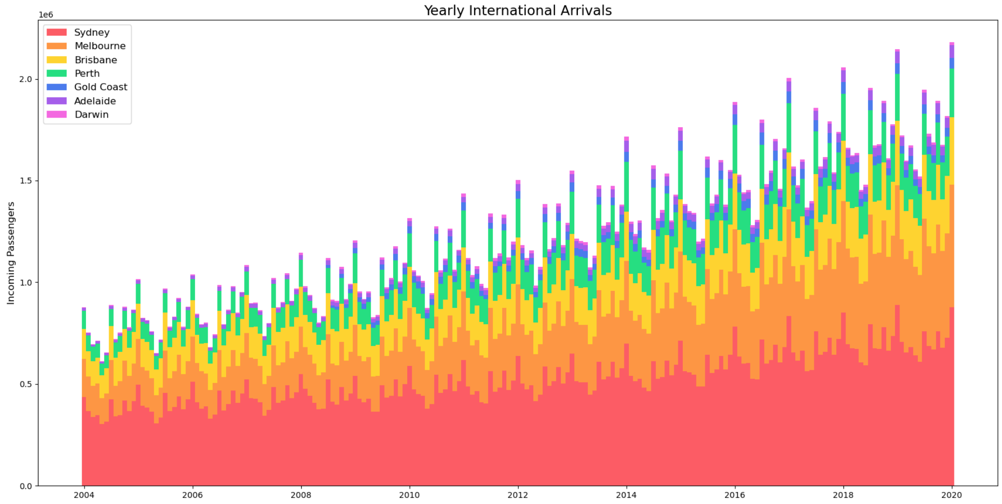


*Figure 1: Stacked bar chart, showing monthly international arrivals in stacked barchart form.*

This stacked bar chart shows two key properties of this dataset. Firstly, activity trends upwards over time. Secondly, aviation activity follows a season pattern, reaching a low in May before reaching a peak around December/January each year.

This visualization is a little ugly, but the bright contrast of hues should help identify each section clearly, even if colour-blind. The order of each bar is also consistent.

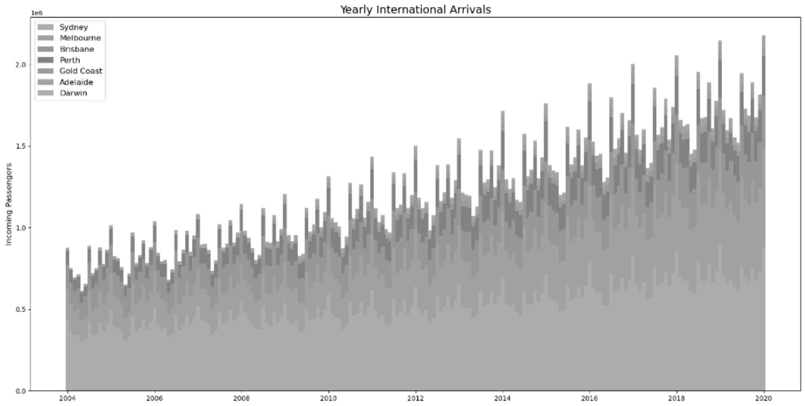


*Figure 2: Grayscale version of the above chart, showing that the colour choices are still legible even in the case of vision impairment*

An alternative to a stacked bar chart is a line chart, which is a better way of comparing individual airports while still seeing overall trends within Australia.

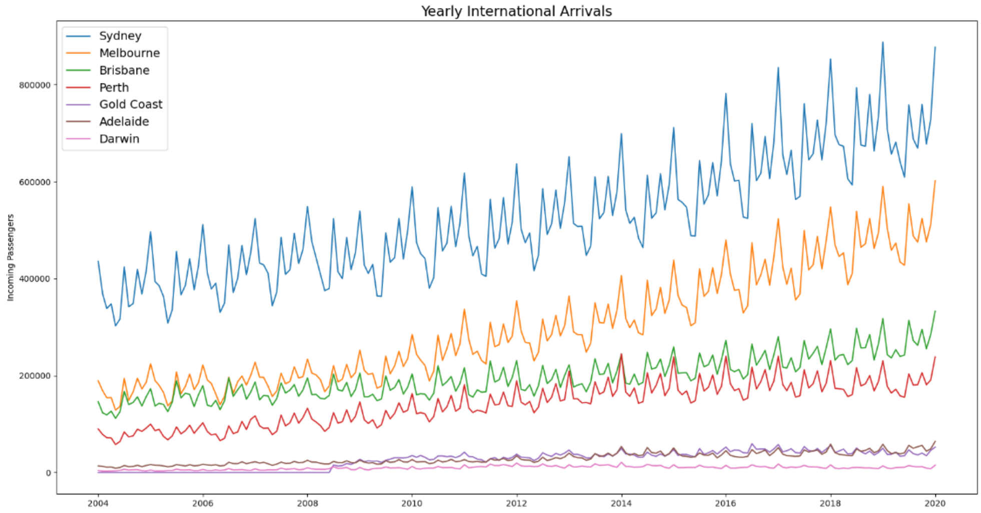


*Figure 3: A basic line chart for international arrivals. This chart allows for comparison between major airports.*

From this chart, we can still see the seasonal variations and the overall increases, but we can also see differences between airports. Notably, we can see that Melbourne airport has done exceptionally well, tending to increase much faster than the other airports. We can also see that for the smaller airports, such as Adelaide and Darwin, seasonal fluctuations have a lot less impact. An exception to this is the Gold Coast airport, which has a stronger reliance on summer tourism.

The colours in this chart are chosen to ensure stronger emphasis, as there is no longer guaranteed order.

For a more detailed look at various connections, we can create a grid of trendlines for various pairs of airports.

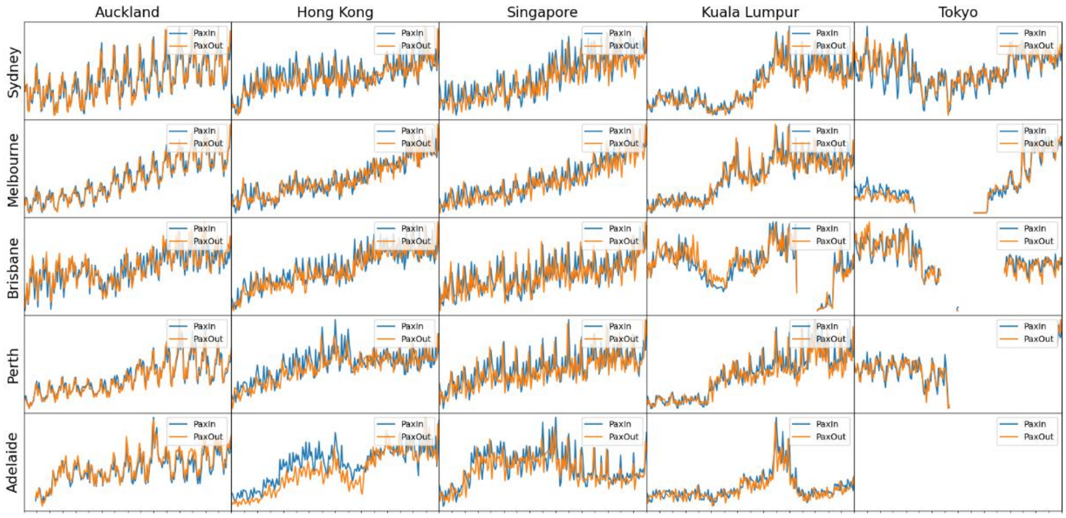


*Figure 4: Trend lines for various pairs of Australian and International airports. Blue lines show incoming passengers, while orange lines show outgoing passengers. These lines are not to scale and are for comparing relative trends only.*

Each trend line is not to scale – this graph is for comparing relative trends only. From this, we can see very interesting data between various airports.

Overall, the trends observed earlier remain true: most routes have the consistent seasonal trends, and overall air traffic is generally increasing. However, there are some exceptions to this. Notable changes can be seen in the Kuala Lumpur routes. Most connections have a sharp descent in the end of 2014 – this is likely due to two major incidents with a major Malaysian airline. Brisbane is hit especially hard, with the route being entirely cancelled between 2016 and 2018. For other anomalies in the data, there's usually a substantial real-world event that causes it.

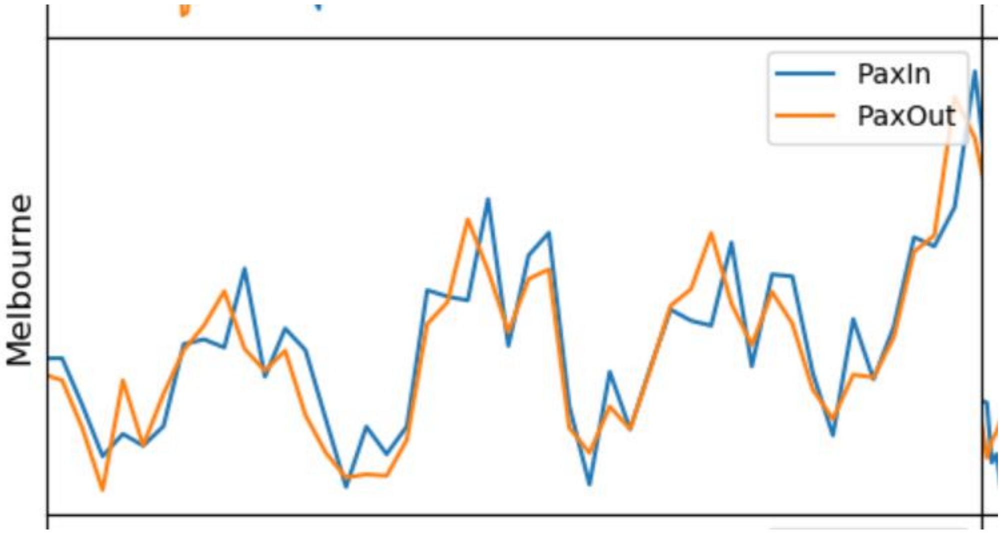


*Figure 5: Zoomed in version of one of the above charts. Seasonal peaks are more clearly visible.*

By zooming in on the trend lines, the seasonal peaks become more visible. Interesting, outgoing passengers tend to peak one month before incoming passengers. Is there a better way to view the seasonal fluctuations?

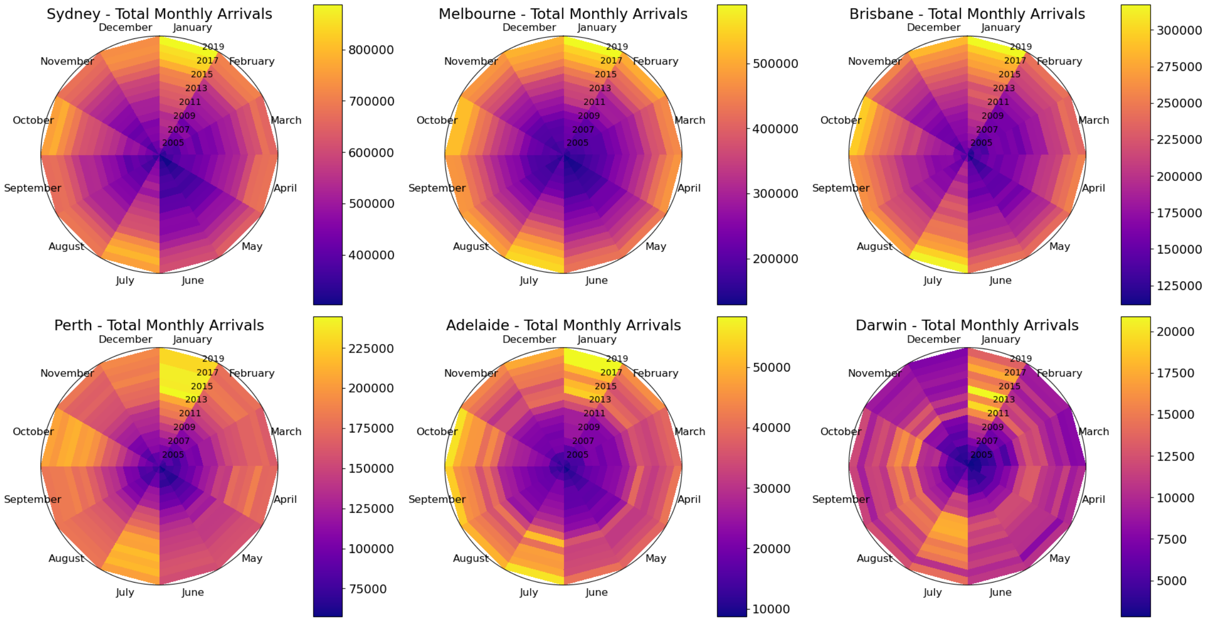


*Figure 6: Time wheels for monthly arrivals for six different Australian airports*

These time wheel plots provide a clear way to view at the monthly variations for various Australian airports. There is a clear major peak in January, with medium peaks in July and October along with a smaller peak in April. This suggests that Australian air travel roughly follows a seasonal cycle, with a dip during the winter months.

Smaller airports are impacted more severely by these cycles – notably, Darwin gets very little traffic outside of the seasonal peak months.

Are these results different for departures?

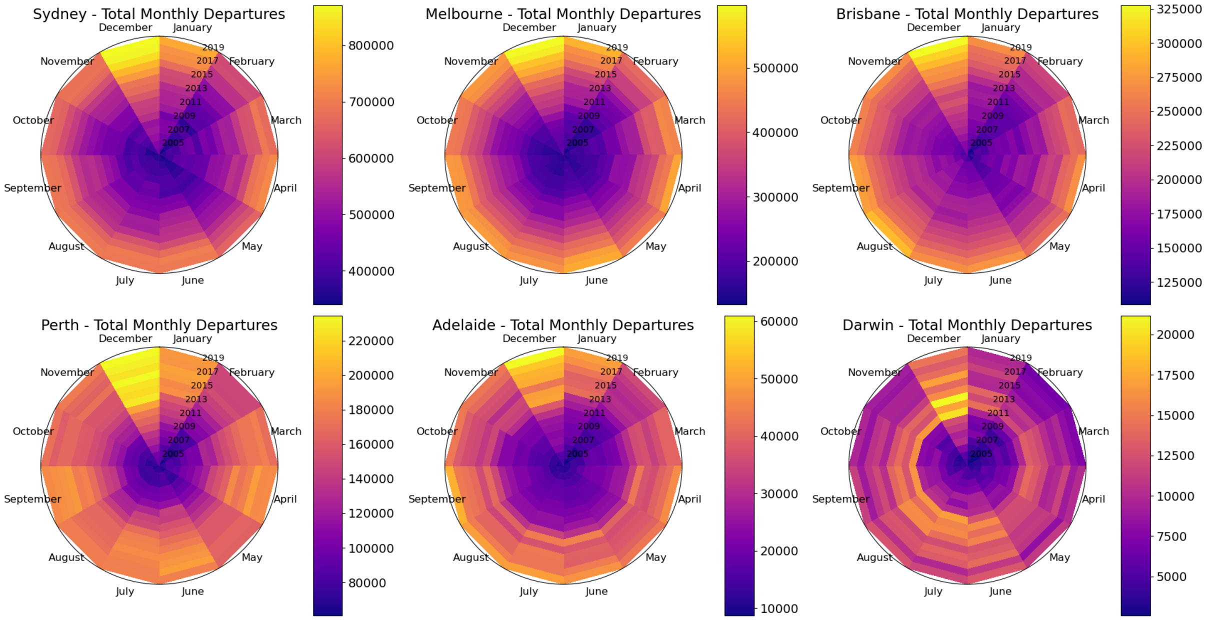


*Figure 7: Time wheels for monthly departures for six different Australian airports*

Departures have a similar monthly cycle, however occurring one month earlier (peaks in December, September, and June). Interestingly, the seasonal variations for departures do not seem as strong – notably Perth experiences consistent traffic year round.

For these graphs, I chose to use the plasma colormap. This is a perceptually uniform colourmap, which makes it very easy to tell at a glance how strong or weak a given value is. For a colourmap like jet, a lack of uniform perceptuality could lead to confusing data at the extrema.

## Domestic Traffic

Since the domestic traffic is in the form of city pairs, it can be displayed in the form of a symmetrical heat map.

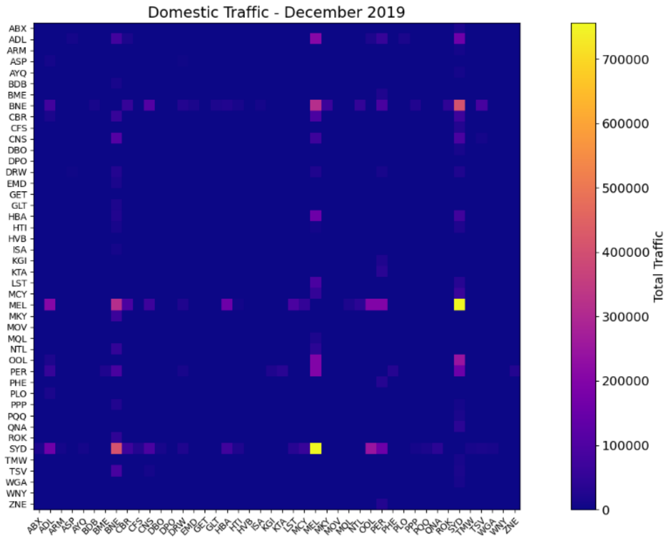


*Figure 8: Heatmap of domestic aviation for the month of December 2019*

This graph does reveal a potential problem with the domestic dataset: only major routes appear to be shown. This gives a good picture of how regional airports are connected to metropolitan airports, however some data for regional connections is lost. For example, there is no data available for Gold Coast – Newcastle flights, even though this route is run daily.

An animated version of this chart can be viewed at: https://www.youtube.com/watch?v=EBht026vlAg

Animating the chart over 15 years does not provide much more additional insights due to the lack of regional routes, however it is possible to see a ‘twinkling’ effect, caused by the seasonal variations discussed earlier. Additionally, with close inspection, it is possible to see the growth of airports such as Perth and Adelaide.

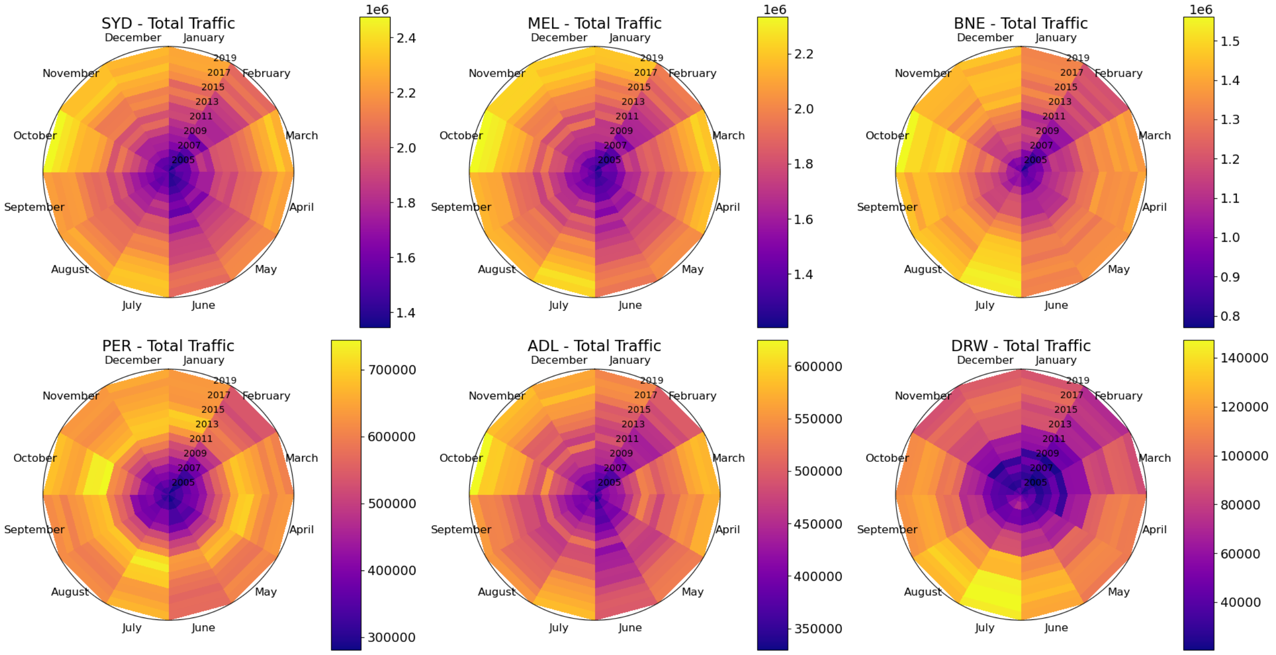


*Figure 9: Timewheel chart for total domestic traffic at various airports*

From this chart, we can see that there is a lot less seasonal variation for domestic travel, although some peaks are still visible in October and July. February appears to be a consistent low point, however, this is likely because February is the shortest month.

Overall, from all the visualisations presented in this report, we can make the following observations:

- Overall, Australian air travel is increasing year by year
- All air travel is affected by seasonal trends, reaching a peak approximately every three months
- International travel is affected more strongly by these seasonal trends
- Regional airports in Australia are generally connected to only one or two international “hub” airports
- International airports in Australia are experiencing stronger growth than smaller airports with less international connections

# Reflection

Processing the data was a difficult step in the process, but I'm incredibly happy with my decision to use xarray for this task. The multi-dimensional structure provided by xarray provided powerful utilities to aggregate the dataset, making it easier to create the engaging visualisations.

The main challenge with the processing was handling the domestic data – the vastly different format required a lot of code to be rewritten, and techniques used to visualise international data would not work with the domestic data. Eventually, I decided to cut my losses and focus primarily on visualising the international data, with some minor visualisations for the domestic data. If returning to this project in future, further analysis of the domestic data would be nice to have.

I'm reasonably happy with the quality of the visualisations that were created. There was a degree of difficulty in determining the suitable formats due to the inherently multi-dimensional nature of the data. Having the two categorical dimensions and one continuous dimension often meant that the data did not fit easily into some common visualisation formats. Additionally, I was unfamiliar with matplotlib, which lead to some delays and imperfections in the charts.

One main issue with the quality of the charts is the font sizing, especially for things like axes and tick labels. I initially created and saved all the charts on a 4K resolution monitor, however, this meant that the size of the text was incredibly small compared to the rest of the chart. With some code adjustment, and saving the figures on 1080p monitor instead, the readability of the text increased dramatically.

Overall, if I repeated this project, I would spend less time focusing on domestic data, and more time broadening the scope of the data. Many other variables are provided such as seat utilisation factors and on time performance, along with other forms of aviation such as freight and mail. I believe that analysing these would provide stronger insights than what I've covered so far.

Aviation data is only released on a biannual basis, which unfortunately means that data for February to April 2020 is unavailable. The trends that occurred in these months would be incredibly interesting to look at, and I'd love the opportunity to come back in a year or two and examine the drastic changes that occurred in 2020.

# Appendix

All code, processed data, and resulting figures can be found in my GitHub repository, https://github.com/rafraser/COSC3000/tree/master/Visualization

This repository was kept private until after the due date of the project.

## Library Documentation

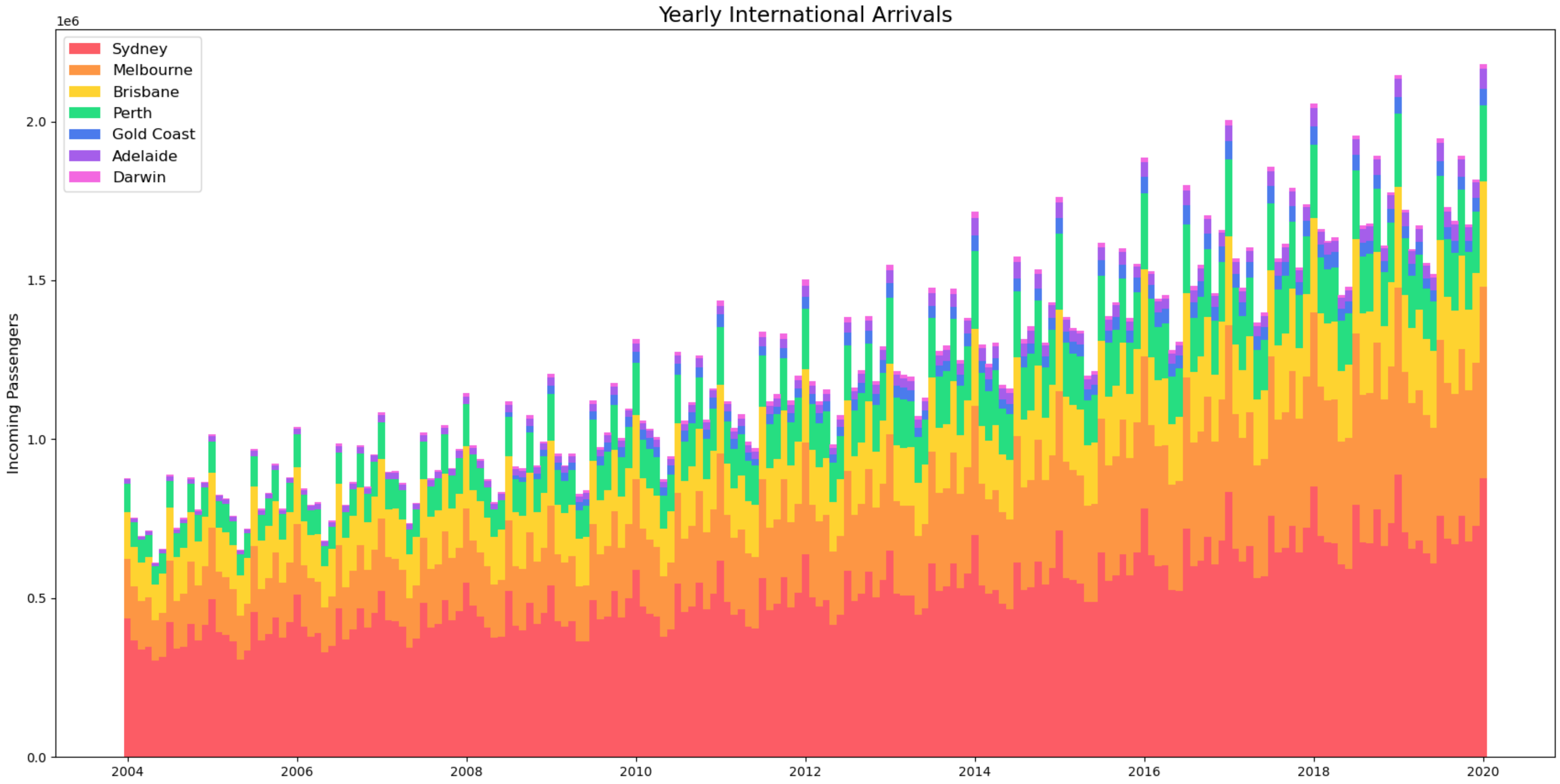
Documentation for Matplotlib can be found at https://matplotlib.org/

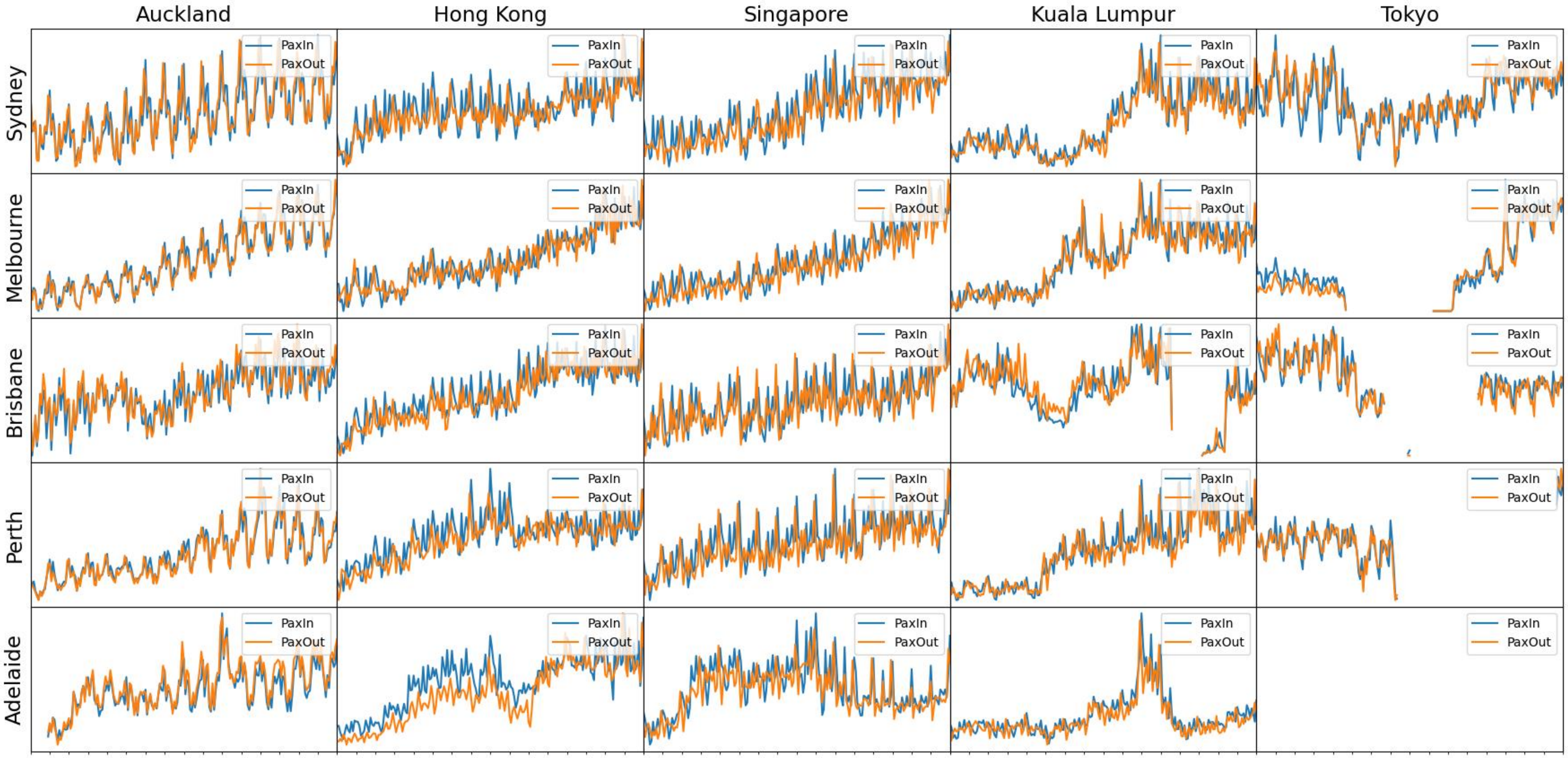
Documentation for Pandas can be found at https://pandas.pydata.org/

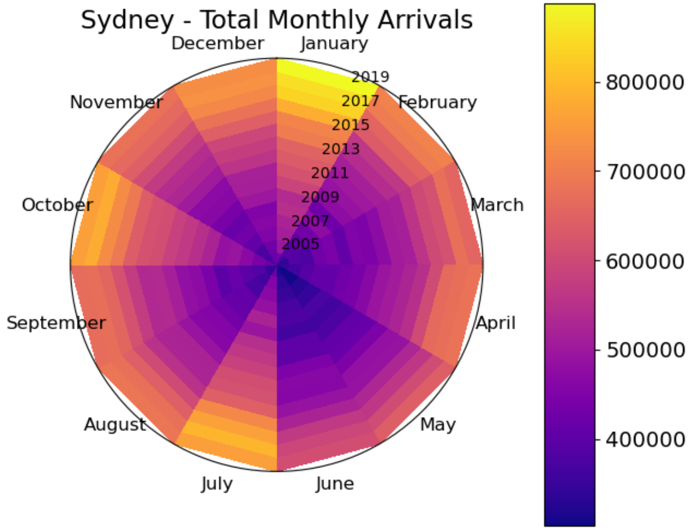
Documentation for Xarray can be found at http://xarray.pydata.org/

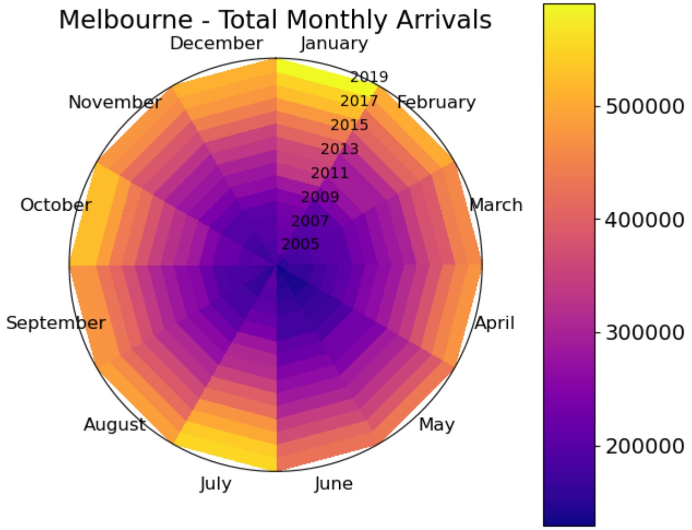
## Data Source

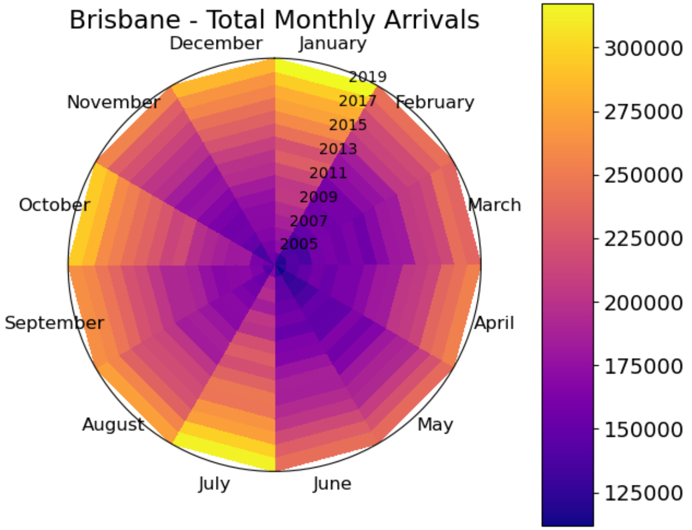
Data for this report was obtained from the Bureau of Infrastructure, Transport, and Regional Economics (BITRE). Aviation statistics can be accessed at https://www.bitre.gov.au/statistics/aviation

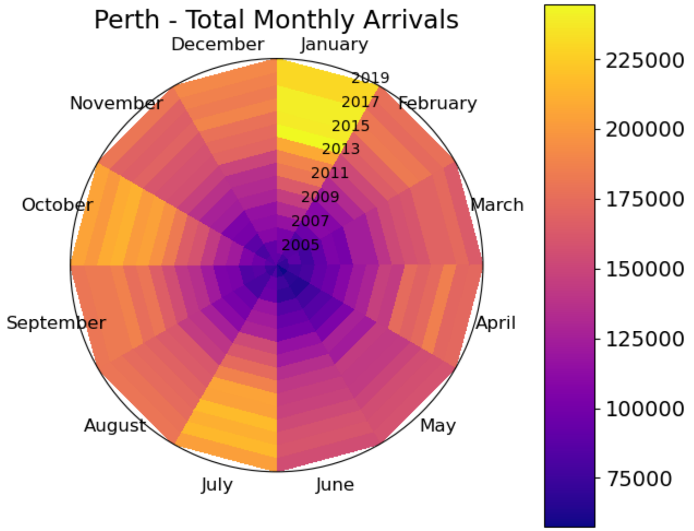

Yearly International Arrivals
1e6
Sydney
Melbourne
Brisbane
Perth
Gold Coast
Adelaide
Darwin
Incoming Passengers
0.0
0.5
1.0
1.5
2.0
2004
2006
2008
2010
2012
2014
2016
2018
2020

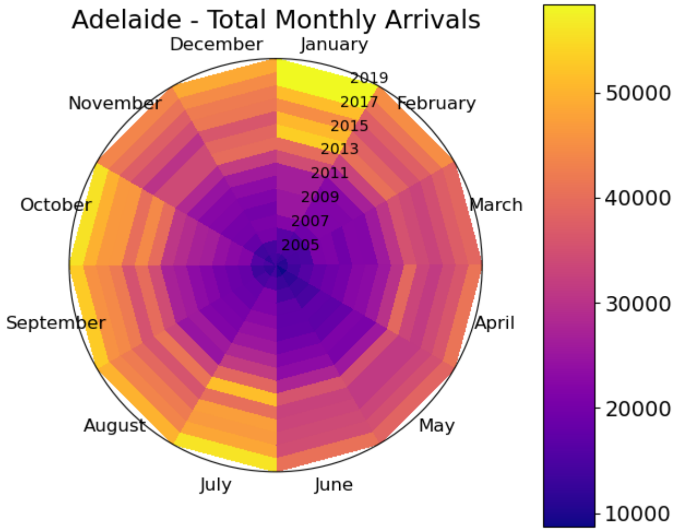

Auckland
Hong Kong
Singapore
Kuala Lumpur
Tokyo
Sydney
Melbourne
Brisbane
Perth
Adelaide
PaxIn
PaxOut

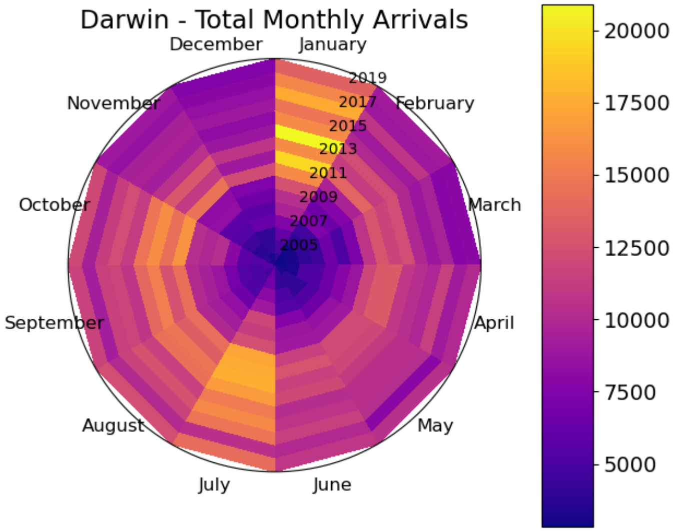

Sydney - Total Monthly Arrivals

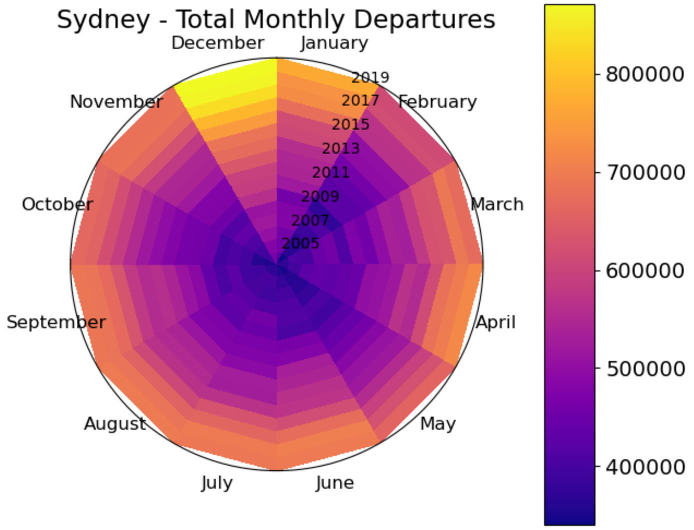

Melbourne - Total Monthly Arrivals


Brisbane - Total Monthly Arrivals

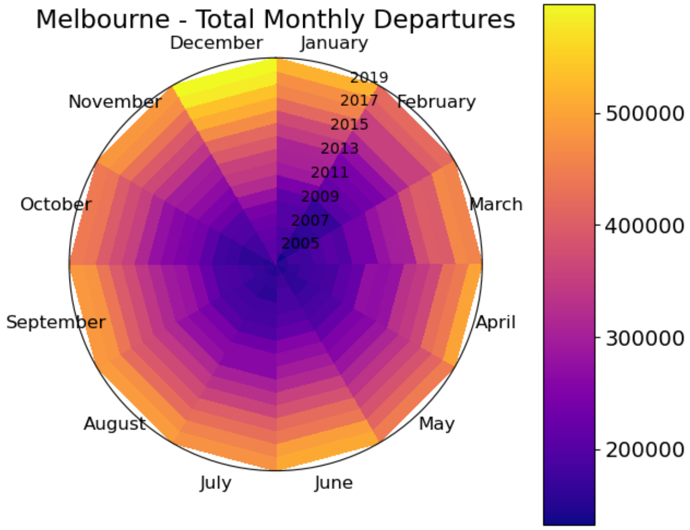

Perth - Total Monthly Arrivals


Adelaide - Total Monthly Arrivals

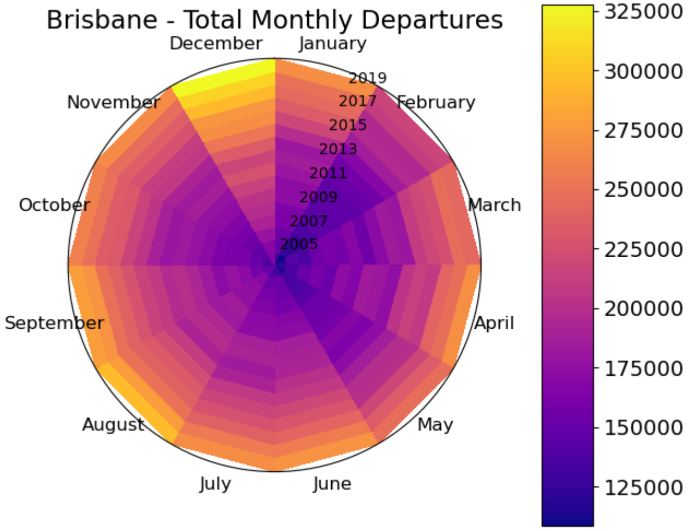

Darwin - Total Monthly Arrivals



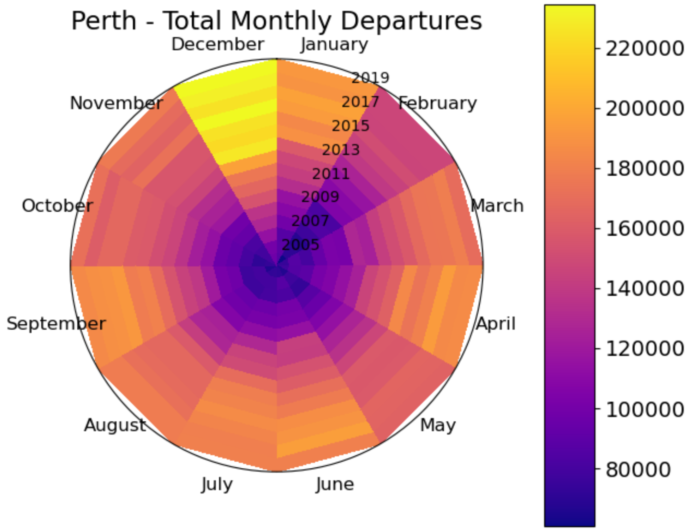
Sydney - Total Monthly Departures



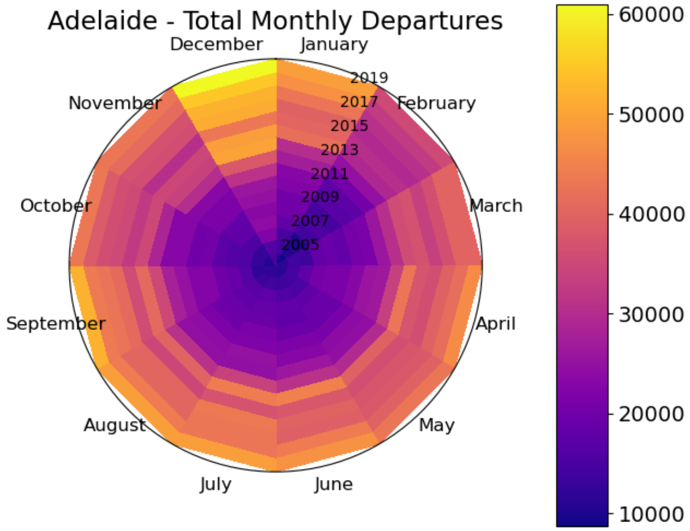
Melbourne - Total Monthly Departures



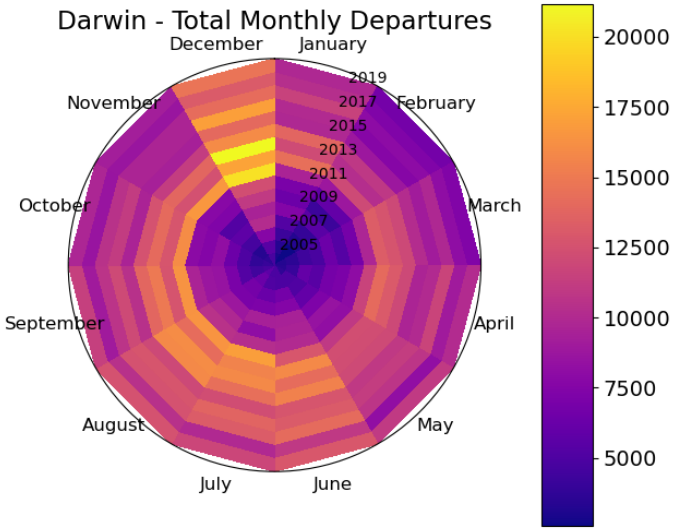
Brisbane - Total Monthly Departures



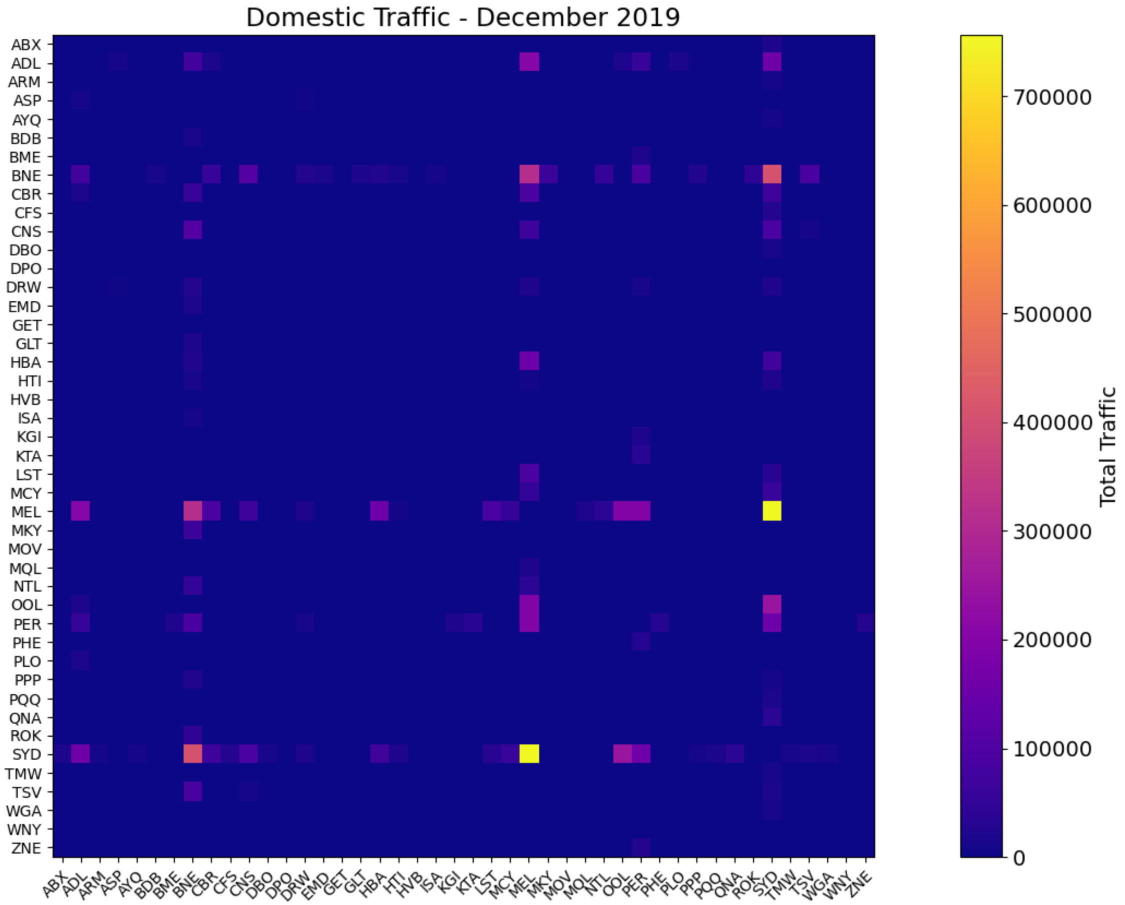
Perth - Total Monthly Departures

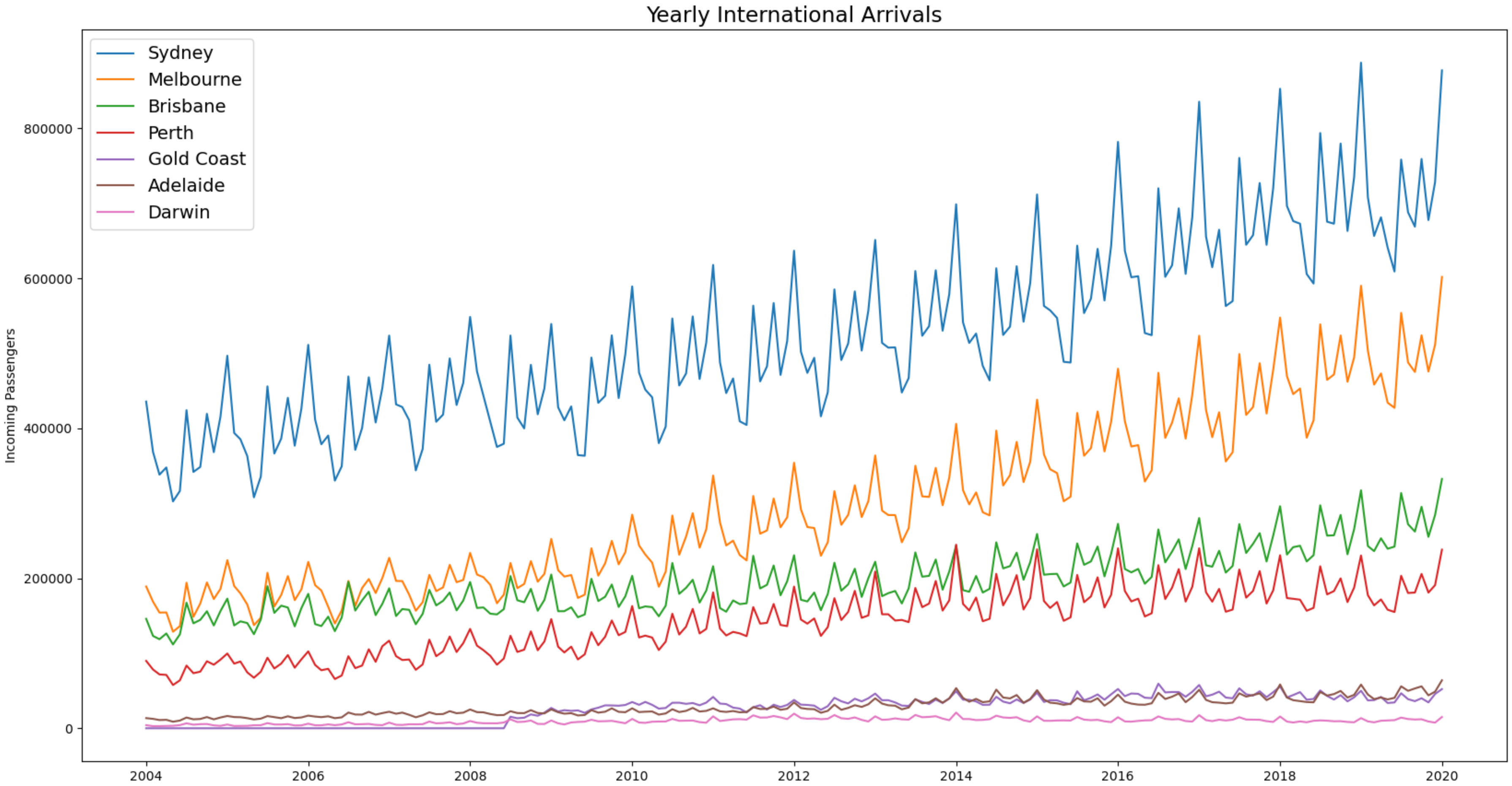


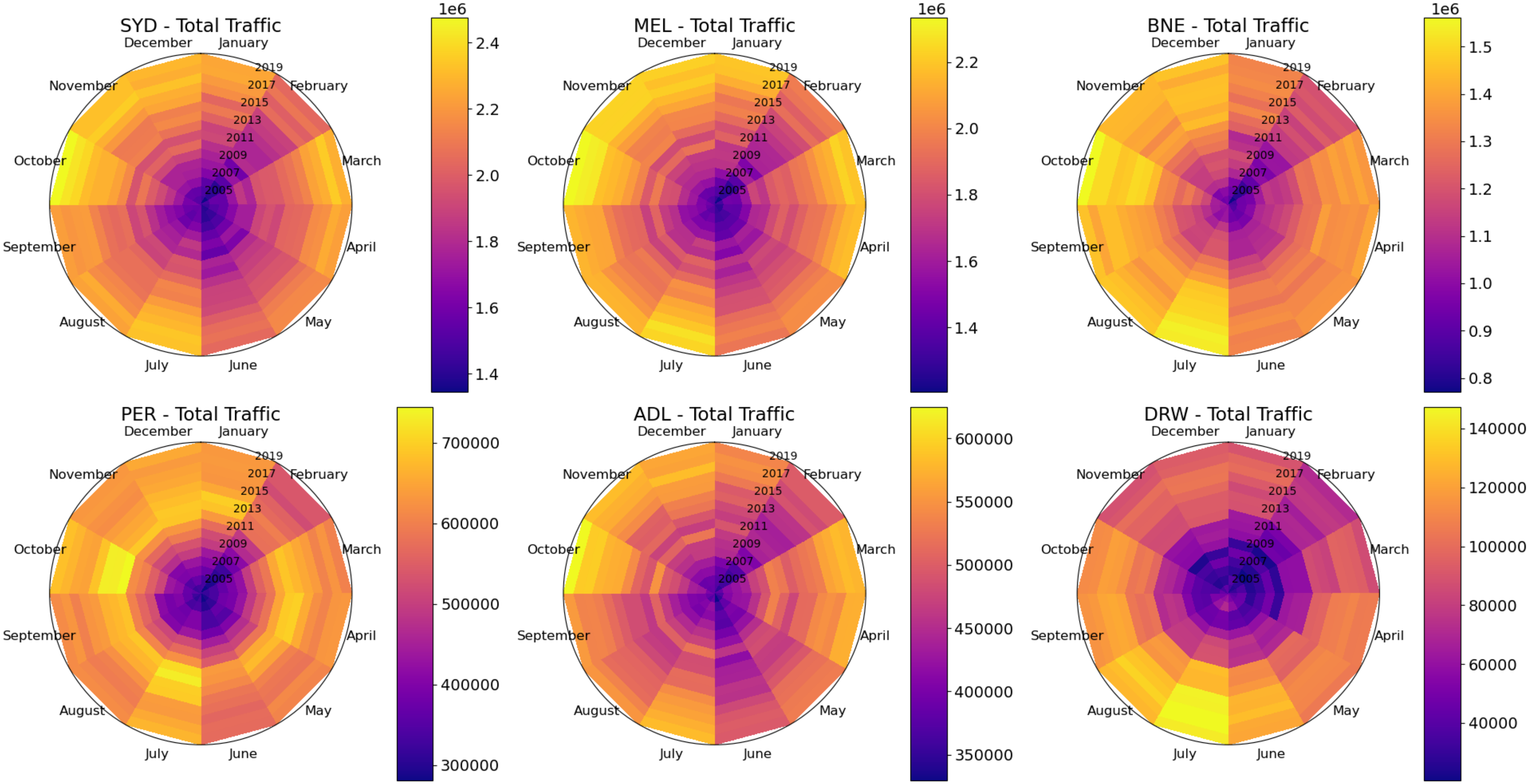
Adelaide - Total Monthly Departures



Darwin - Total Monthly Departures


Domestic Traffic - December 2019
ABX
ADL
ARM
ASP
AYQ
BDB
BME
BNE
CBR
CFS
CNS
DBO
DPO
DRW
EMD
GET
GLT
HBA
HTI
HVB
ISA
KGI
KTA
LST
MCY
MEL
MKY
MOV
MQL
NTL
OOL
PER
PHE
PLO
PPP
PQQ
QNA
ROK
SYD
TMW
TSV
WGA
WNY
ZNE
Total Traffic
700000
600000
500000
400000
300000
200000
100000
0

Yearly International Arrivals
Sydney
Melbourne
Brisbane
Perth
Gold Coast
Adelaide
Darwin
Incoming Passengers
800000
600000
400000
200000
0
2004
2006
2008
2010
2012
2014
2016
2018
2020


SYD - Total Traffic
MEL - Total Traffic
BNE - Total Traffic
PER - Total Traffic
ADL - Total Traffic
DRW - Total Traffic
January
February
March
April
May
June
July
August
September
October
November
December
2005
2007
2009
2011
2013
2015
2017
2019
1e6
2.4
2.2
2.0
1.8
1.6
1.4
1e6
2.2
2.0
1.8
1.6
1.4
1e6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
700000
600000
500000
400000
300000
600000
550000
500000
450000
400000
350000
140000
120000
100000
80000
60000
40000