1. **Understanding the Data**
   a) Identify the most and least trafficked routes

I understod traffic as the total number of passengers per route on the entire dataset (considering all years). The busiest international routes are Sydney-based connections, especially:

- Sydney–Auckland, Sydney–Singapore, and Sydney–Tokyo. These routes carried millions of passengers over the dataset, making them critical air corridors. Secondary hubs like Perth–Singapore and Brisbane–Auckland also appear in the top 10, highlighting their role as important regional links. Clearly Sydney is the dominant hub in Australia's international air travel network.

On the other hand, the least trafficked routes are primarily from smaller cities (like Hobart, Cairns, Townsville, Darwin) to distant or niche destinations such as Los Angeles, Tokyo, San Francisco, Zagreb. These connections show extremely low passenger totals (1 or 2).

Hobart-Los Angeles had 2 total passengers

Hobart-Tokyo 1 total passenger

Cairns-Honiara 1 total passenger

Townsville-San Francisco 1 total passenger

Darwin-Zagreb 1 total passenger

Additionally, there were other routes in the dataset that had 0 total passengers so I didn't considered those, these were: Melbourne-Denver, Perth-Bandar Seri Bagwan, Brisbane-Colombo, Adelaide-Harare, and Brisbane-Chicago

b) Analyze trends and/or geographical patterns

Passenger Traffic Over Time:

Between 1985 and 1988 there was an overall growth: Passenger traffic shows steady growth from around 5.2M in 1985 to a peak of around 7.8M in 1988. In fact, 1988 marks the highest recorded passenger traffic in this period. However, in 1989 there was a drastic decline in traffic; Passenger numbers dropped dramatically to ~3.8M in 1989, breaking the upward trend. Despite the 1989 dip, the long-term trend (1985–1988) suggests increasing demand and traffic growth.

Average Monthly Passenger Traffic

Highest passenger traffic occurs in January (around 3,350 avg passengers). In February, traffic is very low, in fact it drops to the lowest point, marking a sharp decline from January to around 2,700 total passengers. March–June remain relatively steady with minor fluctuations, marking the Spring season. After that, there's a gradual rise from July, peaking again in August (aprox 3,050). Finally, traffic picks up strongly in December (aprox 3,250), suggesting holiday travel demand as it is the beginning on summer in Australia at that month. Overall, there are clear seasonal cycles with peaks in January, August, and December.

Top 10 Countries by Total Passenger Volume

New Zealand dominates with the highest passenger volume (~8M), far ahead of others. Singapore and the USA follow as the 2nd and 3rd most significant markets. Hong Kong, Japan, and the UK form the mid-tier group with strong traffic. Indonesia, Fiji, Thailand, and Malaysia contribute smaller but still notable volumes. Overall there's a clear regional concentration: most top routes are within the Asia-Pacific region, suggesting strong regional connectivity.

Passenger Traffic Over Time for the most trafficked Route (Sydney-Auckland)

Passenger traffic shows strong seasonal fluctuations, with peaks and troughs recurring each year. Overall, there is a growth trend from 1985 to 1988, with volumes climbing from aprox 40k to aprox 70k passengers. 1988 marks the peak, with passenger traffic reaching its highest sustained levels. There's a sharp decline beginning in 1989, suggesting possible external disruptions like policy, economic, or operational changes.

Model

2a, 3a, 3b

For this part of the challenge I built 2 models, evaluated them and then decided which one to keep to make the predictions for the most trafficked route (Sydney-Auckland) for the next 12 months.

I first tried using a Random Forest because this ML model captures complex nonlinear relationships:

-passenger demand can depend on more than jusr seasonal cycles, like promotions, economic charges etc

Also because RFs can handle feature engineering like lagged and rolling features:

-For example, it can detect relationships like: “Passenger demand this month is strongly related to the last 3 months’ average” .

This model gave me initially a MAE of around 6523, a RMSE of around 7496 and an R squared value of aprox 0.5. However, I then used GridSearch for hyperparameter tuning and ended up improving these metrics. The tuned RF had a MAE of around 6263, a RMSE of 7117, an R squared value of 0.54 and a MAPE of 11.9% which is almost excellent for time series forecasting (under 10% is considered excellent).

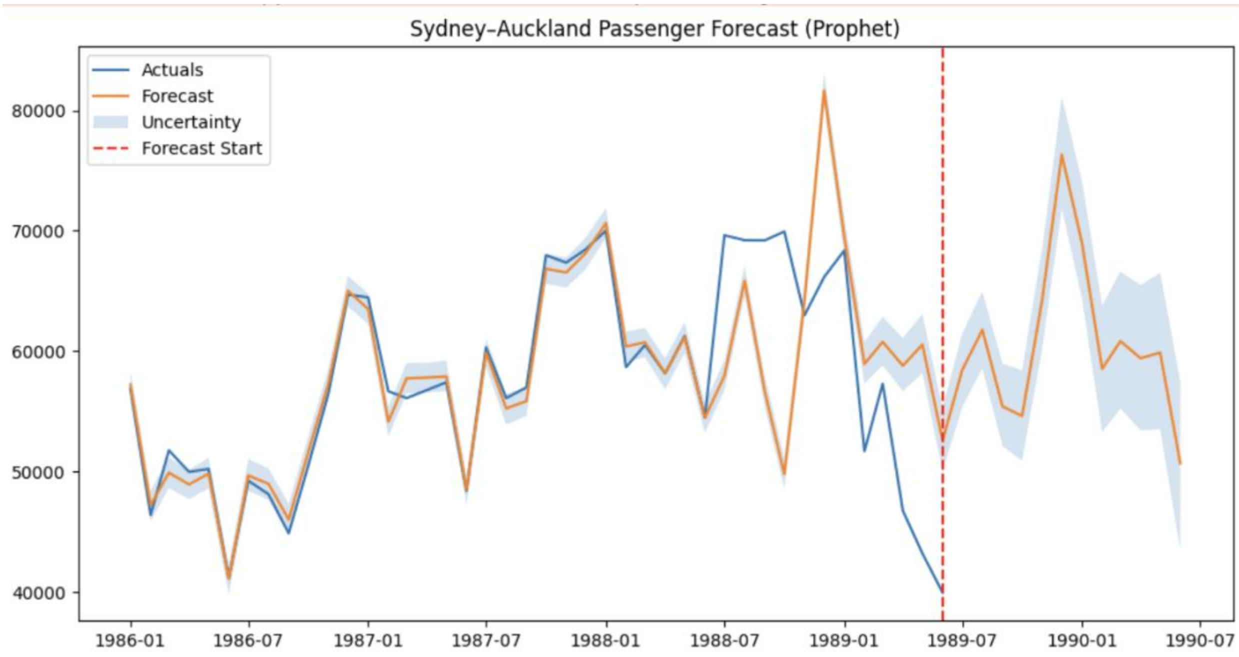
Regardless, I later benchmarked this model with the Prophet model which is a model generally suitable for time series forecasting. The metrics were the following ones: a MAE of 9870, a RMSE of 11636 and a MAPE of 17.8% which is decent. Clearly results weren’t as good as the RF tuned but Prophet was able to capture seasonal fluctuations in traffic which is what we needed for long term (6-12 months) predictions.

Therefore I decided to stick with the prophet model. I then proceeded to make 12 months predictions with Prophet for the Sydney-Auckland route, which were the following ones:

ds	yhat	yhat_lower	yhat_upper
1989-07-01	58418.327109	55353.906652	61462.952623
1989-08-01	61775.124781	58597.959883	64972.434648
1989-09-01	55411.244145	52122.576106	58965.785578
1989-10-01	54618.513544	50891.433637	58410.073941
1989-11-01	64291.274915	60240.894970	68545.787147
1989-12-01	76326.680987	71850.827689	81146.697501
1990-01-01	68986.486959	64370.516737	74108.601378
1990-02-01	58536.472469	53249.968310	63795.040546
1990-03-01	60834.364812	55240.259386	66601.674087
1990-04-01	59412.069473	53418.937251	65489.547447
1990-05-01	59882.221470	53515.485094	66509.190952
1990-06-01	50691.381522	43488.599163	57518.859862

(ds is the date, yhat is the actual prediciton, and yhat_lower and yhat_upper represent the rang of values yhat could be)

After that I plotted Prophet's forecasting passenger traffic history against the actual passenger traffic:



The plot shows the historical passenger traffic for the Sydney–Auckland route (blue line) alongside Prophet's forecast (orange line). The shaded blue region represents Prophet's uncertainty interval, indicating the range within which future values are likely to fall. The red dashed line marks the start of the forecast period (mid-1989). As seen, Prophet tracks historical passenger patterns closely, capturing both the seasonal peaks and troughs. After the forecast start, the model predicts continued fluctuations with some uncertainty, reflecting natural variability in passenger demand. This suggests Prophet is able to model the underlying trend and seasonality of the route, while also providing a measure of forecast confidence.

4. **Provide Recommendations**

a) Which routes should AeroConnect invest more in or scale back from?

Although the forecasting model was applied to only one route, the exploratory data analysis highlights clear opportunities:

High-Priority Investment Routes:

Sydney–Auckland: By far the busiest route, with nearly 3M passengers. This is AeroConnect's strongest market and should remain a priority for more capacity and competitive pricing.

Sydney–Singapore, Sydney–Tokyo, Sydney–Hong Kong: Strong demand, making them suitable candidates for increased investment, larger aircraft allocation, or more frequent flights.

Other Growth Hubs: Routes like Perth–Singapore and Melbourne–Singapore also show solid volumes and connect to major international hubs, making them valuable for strengthening long-haul connectivity.

Routes to Reconsider or Scale Back:
Hobart–Los Angeles, Hobart–Tokyo, Cairns–Honiara, Darwin–Zagreb: These routes show very low passenger volumes and may not be sustainable without subsidies or niche strategies. Scaling back capacity or consolidating these with nearby hub airports could improve profitability.

This means AeroConnect should focus on scaling up in New Zealand and Southeast Asian routes, while reviewing long-haul flights from secondary airports like Hobart or Darwin.

b) How can AeroConnect use this model going forward?

Strategic Planning: Prophet allows AeroConnect to anticipate passenger volumes 12+ months in advance. This can guide fleet management, staffing, and pricing strategies.

Revenue Management: By predicting when demand will be high or low, AeroConnect can adjust ticket pricing dynamically, raising fares during peak demand and offering promotions during slower periods.

Route Optimization: Applying the same forecasting approach to multiple routes will provide a portfolio view of AeroConnect's network, enabling data-driven decisions on where to allocate aircraft and resources.

Scenario Testing: The model can be updated regularly with new data, making it a living tool for ongoing route monitoring. AeroConnect can simulate "what-if" scenarios (e.g., fuel price increases, new competitor entry) and assess potential impacts on passenger volumes.