Diya Kadakia
Generate: Data Branch Tech Challenge

Understanding the Data

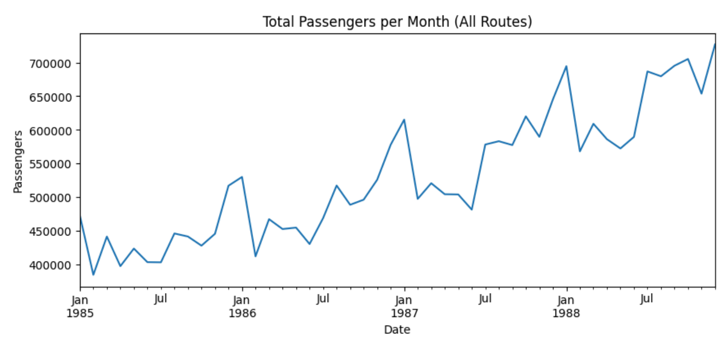
The most and least trafficked routes

```
The Most Trafficked Routes Are:
Route
Sydney to Auckland        2961212
Sydney to Singapore       1440018
Sydney to Tokyo           1292116
Sydney to Hong Kong       1151900
Perth to Singapore         952926
Brisbane to Auckland       893246
Sydney to Christchurch     882357
Melbourne to Singapore     865251
Sydney to Los Angeles      862964
Sydney to Honolulu         861814
Name: Passengers_Total, dtype: int64
```

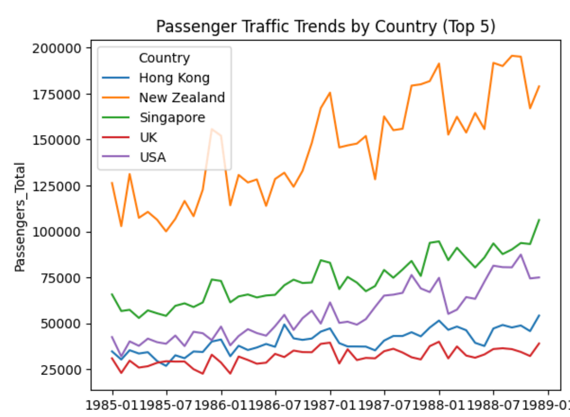
```
The Least Trafficked Routes Are:
Route
Melbourne to Denver            0
Brisbane to Colombo            0
Brisbane to Chicago            0
Perth to Bandar Seri Begawan   0
Adelaide to Harare             0
Townsville to San Francisco    1
Cairns to Honiara              1
Darwin to Zagreb               1
Hobart to Tokyo                1
Hobart to Los Angeles          2
```

Analyze trends and/or geographical patterns and Visualizations

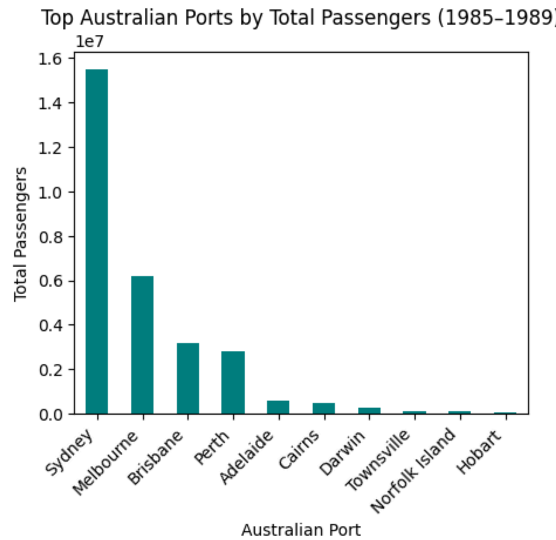
The seasonal pattern of passenger traffic from 1985-1989 shows clear peaks and troughs that align with holiday and travel cycles. Passenger volumes are the highest in December and January, reflecting strong demand across Christmas and the New Year. Travel falls sharply in February and remains relatively low until July, indicating an off peak season. From July onward, volumes rise again, with a second peak in August-September, which is the summer travel season. This shows how air travel is heavily shaped by school and global holiday periods.



From 1985 to 1989, New Zealand dominated passenger traffic, peaking near 200,000 travelers, while Singapore nearly doubled its volumes, emerging as a major Asian hub. The UK showed steady long-haul demand, and Hong Kong and the USA grew modestly but remained secondary markets. These trends suggest AeroConnect should prioritize capacity for New Zealand and Singapore while maintaining targeted services to Europe and other long-haul destinations.

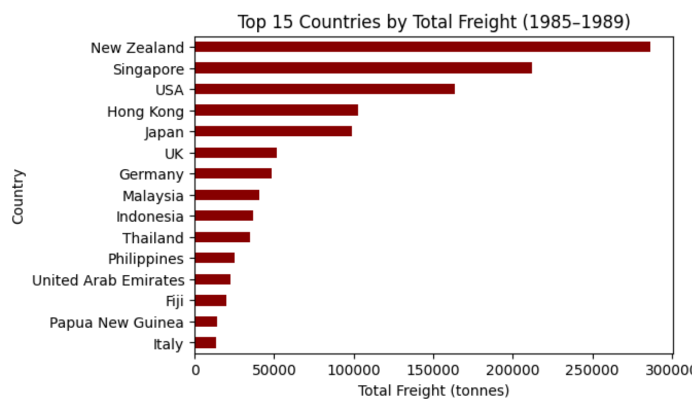


The chart shows that international passenger traffic in Australia from 1985–1989 was highly concentrated in a few major ports. Sydney overwhelmingly dominated, serving more than 15 million passengers and acting as the country's primary international gateway. Melbourne followed as a strong secondary hub, while Brisbane and Perth formed a middle tier of traffic. By contrast, other ports such as Adelaide, Cairns, and Darwin contributed relatively small shares, and locations like Townsville, Norfolk Island, and Hobart saw very limited international volumes.



This pattern highlights that AeroConnect should focus resources—such as larger aircraft and additional crew—on Sydney and Melbourne, while viewing smaller ports as niche or regional connections rather than growth drivers.

The freight analysis highlights that international cargo activity is also concentrated among a few major trade partners. Countries such as New Zealand, Singapore, and the United Kingdom not only account for large passenger volumes but also handle significant freight traffic, reinforcing their importance as strategic hubs. Other countries like the United States and Hong Kong emerge as notable freight destinations, even if their passenger demand is more modest. This suggests that AeroConnect should view these markets as dual-purpose routes, where both cargo and passenger demand can justify sustained investment.



Evaluate your model

For this project, I built a regression model that combines a third-degree polynomial time trend, monthly dummy variables, and ridge regression with an α value of 0.1. I chose the third-degree polynomial because the long-run growth in international air travel is not strictly linear, and a cubic trend provides enough flexibility to capture accelerations and slowdowns in passenger traffic without making the model unnecessarily complex. To address recurring seasonal patterns, I included monthly dummy variables, a technique I first learned in my financial econometrics class, where seasonal dummies are often used to account for predictable month-to-month variations. I also applied ridge regression with an α value of 0.1 to stabilize the model. Ridge regression helps shrink coefficient estimates, which is especially important when working with multiple polynomial terms and seasonal dummies, since this can lead to overfitting if left unchecked. I experimented with different values of α, but 0.1 provided a good balance between flexibility and regularization. These modeling choices were informed by concepts I learned in both my machine learning class, which emphasized polynomial features and regularization techniques, and my econometrics coursework, which emphasized seasonal adjustments.

In terms of performance, the model achieved an MAE of 919.49, an RMSE of 1151.84, and a MAPE of 3.57%. Considering that monthly passenger totals typically range from 20,000 to 30,000, this level of error indicates that the forecasts deviate by less than 4% on average, which is quite strong. Overall, the combination of a cubic trend, monthly dummies, and ridge regression with α = 0.1 allowed me to produce reliable 12-month forecasts for the Sydney–Singapore route while balancing complexity, interpretability, and accuracy.

Provide Recommendations

Based on the exploratory data analysis, AeroConnect should prioritize investment in routes and hubs that consistently demonstrate strong demand. Between 1985 and 1989, New Zealand routes, particularly Adelaide to Auckland, dominated passenger traffic, peaking near 200,000 travelers. Singapore nearly doubled its volumes during the same period, cementing its role as a major Asian hub, while the UK showed steady long-haul demand. Sydney clearly stood out as Australia's primary international gateway, serving more than 15 million passengers, followed by Melbourne as a strong secondary hub. Together, these insights suggest that AeroConnect should direct more resources toward high-traffic connections such as Sydney to Singapore, Adelaide to Auckland, and Sydney to London, as well as maintain robust capacity at Sydney and Melbourne airports where passenger flows are concentrated. By contrast, routes such as Hobart to Los Angeles, Townsville to Manila, and Norfolk Island services had very limited traffic, signaling opportunities to scale back, consolidate, or reallocate capacity to higher-performing routes.

The model further supports these decisions by capturing both seasonal and long-term demand trends. Passenger volumes peak in December–January around Christmas and New Year, fall sharply in February, and then climb again in August–September, reflecting holiday and school travel cycles. By incorporating both a third-degree polynomial time trend and monthly seasonal dummies, the ridge regression model (with α = 0.1) effectively accounts for these predictable fluctuations while smoothing noise in the data. Forecasting demand 12 months ahead allows AeroConnect to proactively adjust fleet allocation, crew scheduling, and pricing strategies to match expected demand, minimizing the risk of overcapacity in low-demand months and maximizing returns during peak seasons.