1.

a. The 10 most popular passenger routes of total volume across all years are as follows (see dataTech2)
   i. Sydney <-> Auckland (New Zealand) 2961212
   ii. Sydney <-> Singapore (Singapore) 1440018
   iii. Sydney <-> Tokyo (Japan) 1292116
   iv. Sydney <-> Hong Kong (Hong Kong) 1151900
   v. Perth <-> Singapore (Singapore) 952926
   vi. Brisbane <-> Auckland (New Zealand) 893246
   vii. Sydney <-> Christchurch (New Zealand) 882357
   viii. Melbourne <-> Singapore (Singapore) 865251
   ix. Sydney <-> Los Angeles (USA) 862964
   x. Sydney <-> Honolulu (USA) 861814

a. The 10 least popular passenger routes of total volume across all years are as follows (see dataTech3)
   i. Melbourne <-> Denver (USA) 0
   ii. Brisbane <-> Colombo (Sri Lanka) 0
   iii. Brisbane <-> Chicago (USA) 0
   iv. Perth <-> Bandar Seri Begawan (Brunei) 0
   v. Adelaide <-> Harare (Zimbabwe) 0
   vi. Townsville <-> San Francisco (USA) 1
   vii. Cairns <-> Honiara (Solomon Islands) 1
   viii. Darwin <-> Zagreb (Yugoslavia) 1
   ix. Hobart <-> Tokyo (Japan) 1
   x. Hobart <-> Los Angeles (USA) 2

b.
   i. In an effort to investigate a general trend of possible growth as the years went on, I graphed the total passengers for each year and found an interesting aspect where the total passengers travelling dropped by over 50% in 1989. I did some research and found out that there was a workers strike that year, and that influenced some of my later decisions in this challenge, which included disregarding the 1989 data for the prediction model due to this extreme circumstance. See dataTech1 and dataTech4
   ii. I also wanted to investigate from which ports passengers travelled Australia from and found that the ports on the Eastern side of the country were very popular for travelers. See dataTech7

iii. Finally, one more relationship I wanted to get clarity on was how the distance of travel correlated with typical attendance of passengers, and my findings were somewhat consistent with what I had predicted going into it. The farther a journey is, the less likely that many passengers will be apart of it. However, I expected a stronger correlation in this relationship but there was only a loose relationship in the end. Also in doing this, I foundthe top 5 most popular countries that travelers visit from Australia. See dataTech5 and dataTech6 (Routes after filtering: 67 Routes filtered out: 170)

2. Found as model.ipynb
3.
    a. When analyzing the passenger traffic data and thinking about a prediction model, and after researching other general models, I chose SARIMA (Seasonal AutoRegressive Integrated Moving Average) over other time series representations for several reasons. The data had clear seasonal repetition over each year, with relevant peaks and pits at certain months. For the specific SARIMA(1,1,1)(1,1,1,12) parameters, I chose these values based on the patterns in the data and the need to balance accuracy with simplicity. The parameters chosen were largely a result of trial and error, and also looking at examples online. Alternative models like ARIMA would neglect the seasonal component. Additionally, the passenger data showed a clear long-term growth trend over the selected training period, with consistent year-over-year passenger volume increases requiring a more complex model. Simple regression would fail to capture both trend and seasonal dynamics simultaneously. My implementation also uses a streamlined approach with the getPortsData() function, which allows users to input countless city pairs by simply specifying origin airport, destination city, and destination country. This scalability advantage provides a reusable framework where the same model structure applies to any route, with standardized preprocessing for consistent date formatting and route string creation, enabling easy prediction for analyzing routes without code modification.
    b. For the most popular route, Sydney <-> Auckland (New Zealand), testing the model on 1988 on a trained dataset from 1985 to 1987 results in the attached monthly accuracies(See dataTech8). In addition, testing the model up until 1990-12 based on training it with the entire dataset results in dataTech9, with specific numbers attached here.
        i. Monthly standardized accuracy for 1988-01 to 1988-12:

1. Month 01: Accuracy ≈ 0.9935
2. Month 02: Accuracy ≈ 0.9012
3. Month 03: Accuracy ≈ 0.9752
4. Month 04: Accuracy ≈ 0.9200
5. Month 05: Accuracy ≈ 0.9625
6. Month 06: Accuracy ≈ 0.9923
7. Month 07: Accuracy ≈ 0.9641
8. Month 08: Accuracy ≈ 0.8988
9. Month 09: Accuracy ≈ 0.9507
10. Month 10: Accuracy ≈ 0.8767
11. Month 11: Accuracy ≈ 0.8053
12. Month 12: Accuracy ≈ 0.8905

ii. Forecasted passenger traffic for 1989-07 to 1990-12:

1. 1989-07: 54,705 passengers
2. 1989-08: 54,023 passengers
3. 1989-09: 54,021 passengers
4. 1989-10: 55,268 passengers
5. 1989-11: 48,776 passengers
6. 1989-12: 52,595 passengers
7. 1990-01: 56,945 passengers
8. 1990-02: 40,492 passengers
9. 1990-03: 46,126 passengers
10. 1990-04: 36,187 passengers
11. 1990-05: 33,174 passengers
12. 1990-06: 29,579 passengers
13. 1990-07: 43,795 passengers
14. 1990-08: 43,100 passengers
15. 1990-09: 43,024 passengers
16. 1990-10: 44,632 passengers
17. 1990-11: 38,405 passengers
18. 1990-12: 42,639 passengers

4.

a. I believe AeroConnect should definitely scale back on trips out of the Northern ports, as those four ports (Port Hedland, Darwin, Cairns, Townsville) support merely 3.0% of all passenger traffic for the country. This represents a clear misallocation of resources that could be better utilized elsewhere in the system. The low passenger volumes from these Northern

routes suggest limited market demand that don't justify the continued investment of aircraft, staff, and infrastructure. Scaling back these operations would allow AeroConnect to free up valuable resources that could be put to more profitable routes where demand actually exists. I would say the three ports that look to gain the most would be the ones on the Southeastern side of the country, comprising Brisbane, Sydney, and Melbourne, as these three ports make up 84.8% of all passenger traffic. This massive concentration of demand clearly shows where AeroConnect's customers actually want to fly from, and the company should hone in on these proven markets rather than spreading resources thin across suboptimal routes. Furthermore, I believe if AeroConnect were looking to truly maximize their potential, they would provide more support and attention to routes going from these three Southeastern ports to the most popular destinations. The top three countries (New Zealand, Singapore, USA) make up 51.5% of all passenger traffic from the Australian ports, which means focusing on these origins and destinations would capture the highest demand.

b. I believe AeroConnect can use this SARIMA model as a powerful tool for making smarter business decisions and optimizing their operations more effectively, but they need to understand that these predictions work best under normal operating conditions without major disruptions like strikes or other unexpected events. The model shows strong accuracy for monthly forecasting. In the case of this data, it would not be accurate for predicting extraordinary situations like the 1989 pilot strikes that completely threw off actual passenger numbers, so it's best used as a baseline tool that assumes typical market conditions. The seasonal patterns the model captures are particularly valuable as it clearly demonstrates the peaks in December/January and the slower periods mid-year, which means AeroConnect can use this information for smarter pricing strategies and revenue management. I think the most practical way for AeroConnect to implement this would be to use the model for routine monthly planning while keeping flexibility to adjust when unexpected events happen. They should treat these forecasts as their starting point for planning but always have backup plans ready when external factors emerge. The model would be especially useful for identifying their most predictable and profitable routes, allowing them to focus investment and resources on areas where demand patterns are stable and forecastable. This approach would help AeroConnect build a more resilient network that can better handle

unexpected disruptions while maximizing profitability during normal operating periods.