1. Understanding the Data

   a) Identify the most and least trafficked routes

   Using pivot tables and conditional formatting in Excel, I was able to create a heat map of passenger traffic. Based on the “Passengers_Total” field, the top 3 most trafficked routes are Sydney to Auckland, Sydney to Singapore, and Sydney to Tokyo. The least 3 trafficked routes (with at least 1 flight flown on the route) are Hobart to Tokyo, Cairns to Honiara, and Darwin to Zagreb.

   Using a similar approach, I was able to create a heat map of the countries most traveled to. The top 3 include New Zealand, Singapore, and USA, providing supplementary information to help us understand the data. The bottom 3 countries include Malta, Denmark, and Cyprus.

   b) Analyze trends and/or geographical patterns

   The goal of this project is to recommend profitable routes for AeroConnect. Beyond the passenger traffic (volume) we analyzed in part A, another crucial factor for investment is a route’s consistency and growth patterns over time.

   To analyze consistency, I created a pivot table and line chart to visualize the average monthly passenger totals for each route in the years given. While the Sydney to Auckland route has very high traffic, it also shows month-to-month variability. In contrast the Sydney to Singapore route is more consistent, but has lower total traffic. The Sydney to Tokyo route is the least trafficked of the top three and also varies the most.

   Next, I analyzed growth and decline trends over time by creating two pivot tables. The first pivot table visualizes traffic per route per year, showing which routes are growing and which are declining. The second is a heat map showing the change in traffic to each country over the years. The analysis of these pivot tables reveals that most of the top-performing countries show positive growth trends over the years, whereas countries with lower traffic have less consistent and even declining trends.

   Finally, for international flights, some research in the industry reveal that profitability is often a combination of both passenger and cargo (in our case freight) traffic. The raw data includes passenger, freight, and mail traffic, which can be analyzed further to identify routes that are profitable due to a combination of the three.

   c) Create visualizations to demonstrate trends & patterns determined in part b

2. Build a Model

   a) Your model should predict passenger traffic for the next 6–12 months on at least 1 city pair

   NOTE: Make sure to use proper coding practices (i.e. commenting, camelcase, etc.)!

   I built a time-series forecasting model using the Prophet Python library to predict passenger traffic for the Sydney to Auckland pair for the year 1989. The model was trained on the data from 1985-1988. To do so, I did have to strip the data to only two columns: the date and the passenger total. For the date, I used the first of the month. I also had to reformat the columns to ‘ds’ (data) and ‘y’ (passenger total), as required by the library.

3. Evaluate your model

a) Explain your model choices — why did you choose the elements you did

I landed on the Prophet model after first trying a linear regression model and doing some research based on the use case. My linear regression model performed decently, with a 6097 mean absolute error (around a 10% error rate). This was a good start, however, it was clear the linear regression model was not able to capture the full extent of the trends present in the data. I had heard of time series modeling in the past, so I decided to look into this. I figured out that this modeling approach would work and stumbled upon the ARIMA approach. ARIMA stands for auto regressive, integrated, and moving average. Auto regressive means the model uses the relationship between current and past values. Integrated makes the time series stationary by differencing (rather than forecasting the number of passengers, we forecast the change in number of passengers). Moving average models the relationship between the current and past values. Taking this one step further is SARIMA modeling, which is Seasonal ARIMA. This adds on to the ARIMA approach with another layer strictly for time series that show seasonality. The SARIMA approach seemed great, however, it seemed to require a lot of data preprocessing to get started, so I looked for another approach. This brought me to Prophet, Facebook’s tool for forecasting time series data. This tool is similar to the other approach, however, is very streamlined and easier to get started with. It required just one step of data preprocessing. There are tons of parameters to mess around with in Prophet. For my model, I added the following parameters: seasonality prior scale, custom seasonality mappings (monthly and quarterly), and holiday effects. Seasonality prior scale controls the strength of automatic seasonalities (yearly, weekly, etc.). The custom seasonality mappings help capture patterns at the monthly and quarterly level that the default Prophet version (yearly) might miss. Since our predictions were on a monthly basis, it was important to look for monthly patterns - taking it a step further to quarterly improved the results as well. Finally, our data analysis proved that holiday travel had a significant effect on trends, so I used Prophet’s in-built holiday effects to automatically include Australian holidays. The result of adding these parameters was a multi-layered seasonal model which was pretty accurate in predicting the 1989 data. The model was also able to predict future data.

b) Evaluate the model's performance & report the accuracy of the model

My model performed pretty well. To evaluate my model I decided to use mean absolute error. I decided to use this scoring method because it's simple, easily interpretable, and is scale consistent with the original data. So, with our time series data, we can say “On average, my forecast was off by 1000 passengers.”. In my case, I was able to get my

MAE (mean absolute error) down to 3069, meaning, on average my predictions were off by 3069 passengers. In the first six months of the year 1989, there were around 51,208 passengers on average. So, we can calculate our error rate (3069/51,208) as an average of ~6%. This is pretty good for our model and a good basis for making business decisions.

4. Provide Recommendations

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

   AeroConnect should continue to invest in the top 3 routes: Sydney to Auckland, Sydney to Singapore, and Sydney to Tokyo. These routes have the highest passenger traffic and generally show positive growth trends. Based on predictions, the Sydney to Auckland route is expected to maintain strong performance with monthly traffic varying between 45,334 and 68,042 (an all time high) over the next 12 months. Further information on predictions can be found in the future_predictions.csv file. A further, deep dive into the profitability of freight and mail traffic for these routes could help optimize operations and provide further allocation. Other than the top 3 routes, routes that have a clear, increasing trend in passenger traffic and routes that have minimal month to month variability are primed for investment. Perth to Singapore and Brisbane to Auckland (the 5th and 6th busiest routes) have high passenger traffic and are worth investing in. Other than the top routes, some interesting routes to continue to watch and potentially invest in include Cairns to Port Moresby, Darwin to Singapore, and Melbourne to Singapore. These routes aren't the busiest, but have strong passenger traffic and more importantly, very clear upward trends.

   b) How can AeroConnect use this model going forward?

   At the end of my recommendations, I mentioned three routes that have potential investment opportunities. This is where the predictive model comes into play. The Cairns to Port Moresby, Darwin to Singapore, and Melbourne to Singapore routes all have upward trends, so running the predictive model on these routes can help determine if these routes will continue to have investment opportunities. Moving forward, the model should be used to predict route traffic and determine investment and resource allocation, helping AeroConnect run their operations more efficiently. As we continue to collect data, we can add to our model. Firstly, we can predict more into the future as we have better historical data and a grasp of the underlying patterns and trends. Secondly, we can add other layers to our model. So far, our model only predicts using passenger traffic, but freight traffic is also very profitable. In the future, adding another layer to our model to predict freight traffic could take profitability and investment opportunities further.