HW 4: Air Carolina Route Demand Forecasting Report

Anna Fetter

2025-03-30

Github Repo with source code & dataset: https://github.com/aefetter/plan372-air-carolina-regression

1. Summary

I used U.S. Census data and flight volume data from October 2021-September 2022 to help Air Carolina decide where to launch its next nonstop route out of Raleigh-Durham (RDU). I built a linear regression model to predict passenger volumes based on metro area population, median income, and distance between cities. I used data from all existing flight routes in the country with above 10,000 passengers, not just flights flying to or from Raleigh. Using that model, I forecasted traffic on four possible new routes from RDU: Portland (PDX), El Paso (ELP), Tallahassee (TLH), and Sacramento (SMF).

RDU–TLH had the highest forecasted passenger volume at around 49,000 passengers annually. Routes to PDX and SMF also performed well, each with forecasted demand around 37,000, making them strong contenders. The route to ELP had the lowest predicted demand at 19,000 passengers. For context, among all moderate-to high-volume routes (over 10,000 passengers per year), the median volume is about 70,400. This reflects the hub-and-spoke structure of the airline industry, which tends to concentrate traffic through major hubs rather than direct flights between smaller metro areas.

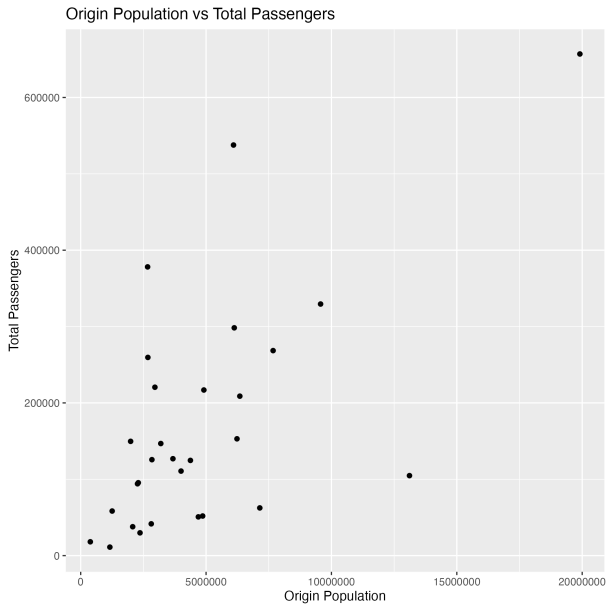
The model doesn't account for seasonality, competition from other airlines, or tourism trends. Still, metro-level population, income, and distance provide useful starting points for identifying promising new routes.

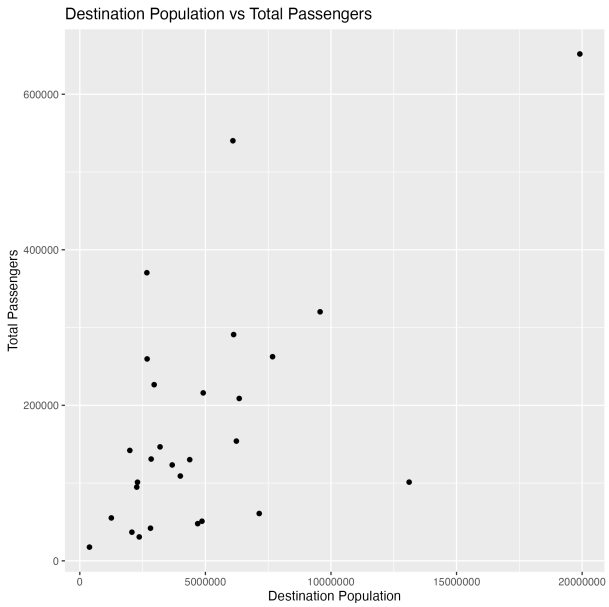
2. Market Saturation at RDU

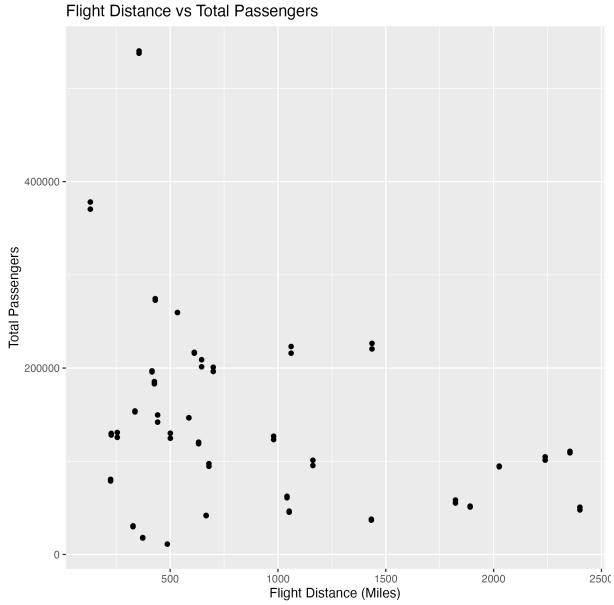
most popular routes, include the scatter plots RDU currently has 34 departing direct flights. The five most popular outbound routes from RDU between October 2021 and September 2022 were Atlanta (ATL, 540,190 passengers), Charlotte (CLT, 370,480), New York–LaGuardia (LGA, 272,810), Orlando (MCO, 259,640), and Denver (DEN, 226,520). The list of direct flights is dominated by major airline hubs and major US cities. Only one direct route from RDU serves a metro area with fewer than 1 million residents: Trenton–Princeton, NJ. However, that route likely functions as a proxy for larger nearby markets, as Trenton is located less than 60 miles from both Manhattan and Philadelphia.

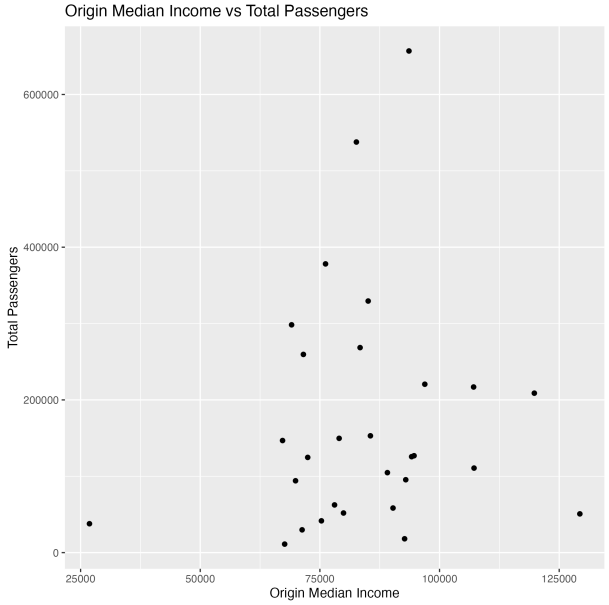
3. Regression: What drives air travel demand?

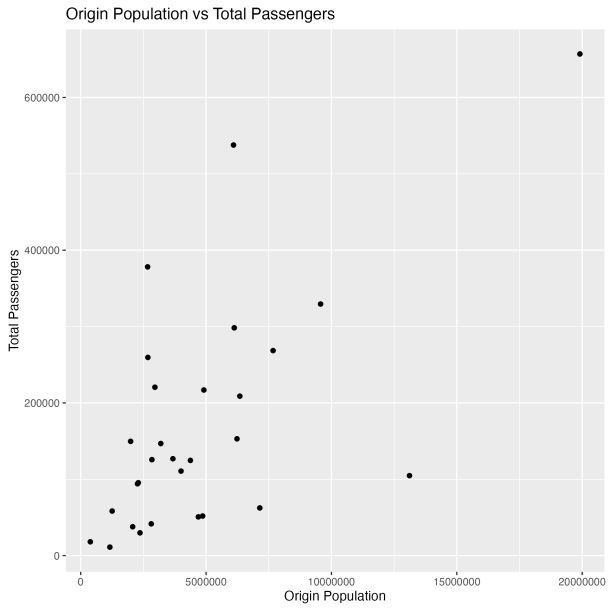
As preliminary exploration before building my regression, I contructed a series of plots to see if there were any obvious trends.

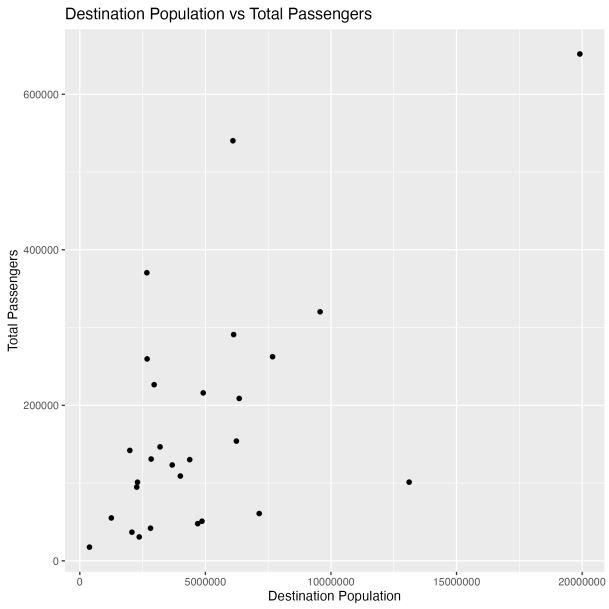

Origin Population vs Total Passengers
Total Passengers
600000
400000
200000
0
0
5000000
10000000
15000000
20000000
Origin Population

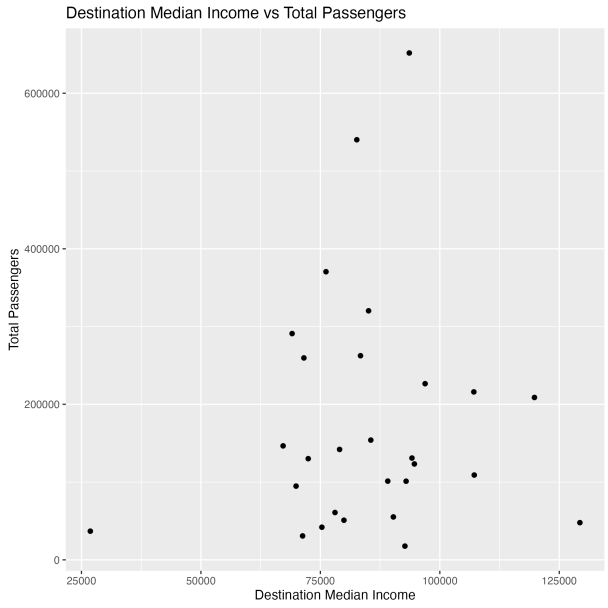

Destination Population vs Total Passengers
Total Passengers
600000
400000
200000
0
0
5000000
10000000
15000000
20000000
Destination Population


Flight Distance vs Total Passengers
Total Passengers
400000
200000
0
500
1000
1500
2000
Flight Distance (Miles)


Origin Median Income vs Total Passengers
Total Passengers
600000
400000
200000
0
25000
50000
75000
100000
125000
Origin Median Income







Population of origin and destination cities appeared to be the most impactful on the number of passengers, at least visually.

To estimate air travel demand between metro areas, I built a linear regression model using passenger volume data and U.S. Census variables. I filtered the dataset to include only routes with more than 10,000 annual passengers, using data from October 2021 to September 2022. The model predicts total passenger volume based on origin and destination metro area population and median income levels as well as the distance between the two cities and miles. All demographic variables came from the American Community Survey (ACS) 2023 via the `tidycensus` package. Distance was calculated between the origin and destination airports. The model was trained on all qualifying routes in the U.S.. This model could be used to forecast routes from other airports, not just RDU. The table below shows the model output:

Regression Model: Predicting Passenger Volumes on All Flights

Variable	Estimate	Std. Error	t-Value	p-Value	Significance
(Intercept)	-74107.555	9241.287	-8.02	0	***
origin_pop	0.006	0.000	18.82	0	***
destination_pop	0.006	0.000	18.86	0	***
distancemiles	-26.100	1.880	-13.89	0	***
origin_income	0.781	0.083	9.44	0	***
destination_income	0.790	0.083	9.52	0	***

All variables proved statistically significant at the $p < 0.001$ level, denoted by the triple stars.

Higher origin and destination populations correlated with more passengers. For every 1,000 additional people in the origin or destination metro, the model predicts about 6 more passengers on the route annually.

Median income at origin and destination also correlated to more passengers on the routes. A $1,000 increase in median income is associated with roughly 780 to 790 more passengers annually.

Distance had a negative coefficient. Each additional mile between cities is associated with an average of 26 fewer passengers.

4. Prediction: What new routes should Air Carolina fly?

Using the regression model explained above, I forecasted annual passenger volumes for four proposed nonstop routes from Raleigh-Durham (RDU): Portland (PDX), El Paso (ELP), Tallahassee (TLH), and Sacramento (SMF).

Here are the results:

Prediction Model: Predicting Passenger Volumes on All Flights

Route	Origin CBSA	Destination CBSA	Origin Population	Destination Population	Origin Income	Destination Income	Distance (mi)	Forecasted Passengers
RDU → TLH	39580	45220	1449594	388298	96066	63078	496	48728
RDU → PDX	39580	38900	1449594	2510529	96066	94573	2363	37628
RDU → SMF	39580	40900	1449594	2406563	96066	93986	2345	37009
RDU → ELP	39580	21340	1449594	869606	96066	58800	1606	19273

Among the four options, RDU to Tallahassee (TLH) had the highest predicted demand, with approximately *48,700 annual passengers. This was somewhat surprising given TLH's relatively small metro area, but its short distance from Raleigh (496 miles) and moderate income levels helped boost the forecast.

Routes to Portland (PDX) and Sacramento (SMF) followed closely, with around 37,000 passengers each. These cities are much farther from RDU, but are larger and wealthier metros areas.

El Paso (ELP) had the lowest projected demand, with just 19,300 passengers annually. This is likely due to its smaller population, lower median income, and mid-range distance.

Based on the model, TLH appears to be the strongest short-haul addition. SMF and PDX may be worthwhile long-haul bets, while ELP does not appear promising given current demand drivers including population, distance, and income levels.