Mandy Langlois

Air Carolina

Introduction

This report presents an analysis of passenger demand projections for Air Carolina's proposed new routes. Using a dataset that includes flight route details, passenger volumes, travel distances, and socio-economic indicators from the Census Bureau for both origin and destination cities, I was able to develop a regression model to predict passenger demand. The dataset contains 9,500 observations, each representing a unique flight route with variables such as origin and destination airport codes, number of passengers per route, travel distance in miles, and CBSA (Core-Based Statistical Area) characteristics. Socio-economic data includes median income and population size for each CBSA, which helps capture regional economic profiles potentially influencing air travel demand. The model uses these variables to predict passenger demand on routes from Raleigh-Durham International Airport (RDU) to four new destinations: Portland, OR (PDX); El Paso, TX (ELP); Tallahassee, FL (TLH); and Sacramento, CA (SMF). This analysis aims to inform Air Carolina investors, about anticipated route popularity, expected passenger volumes, and possible returns on investment.
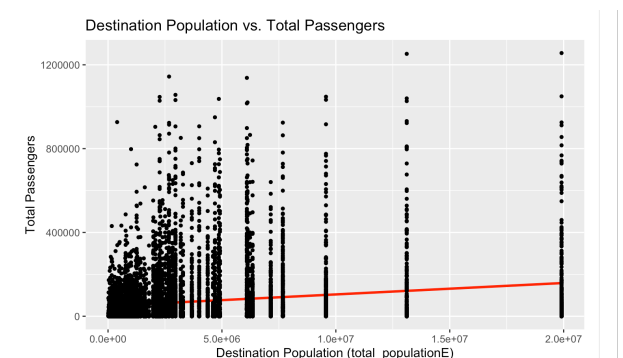
Scatterplot interpretation

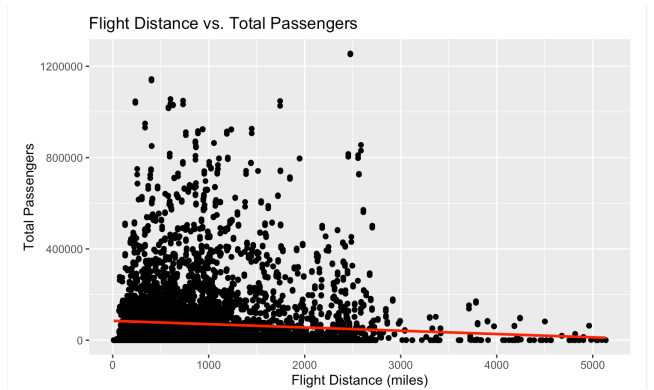**Figure 1. Destination V. Total Passengers**
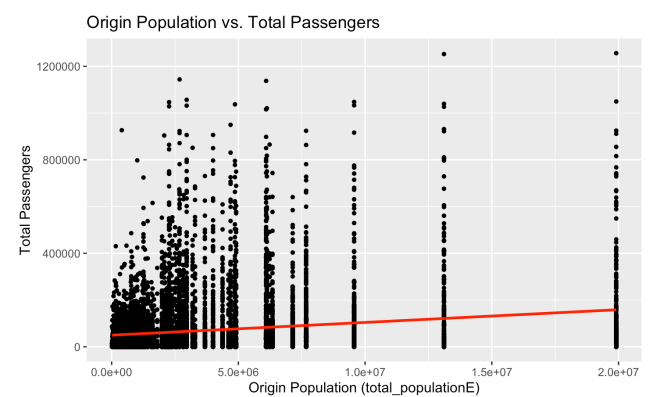
Figure 2. Distance V. Total Passengers.

Figure 3. Origin Population V. Total population.

The scatterplots reveal interesting patterns in passenger demand. In the Origin Population vs. Total Passengers plot **(figure 3),** I observe a positive trend, indicating that routes originating from cities with larger populations tend to have higher passenger volumes. This suggests that larger populations at the origin contribute to higher demand for flights. Similarly, the Destination Population vs. Total Passengers plot **(figure 1)** shows a positive correlation, implying that destinations with larger populations also attract more passengers, likely due to increased tourism or business travel opportunities in populous areas. In contrast, the Flight Distance vs. Total Passengers plot **(figure 2)** displays a negative trend, suggesting that longer flights generally see fewer passengers, maybe due to higher costs and time commitments associated with extended travel distances. These trends indicate that both origin and destination populations positively influence demand, while longer distances may reduce passenger volume

Model summary

```
Call:
lm(formula = passengers ~ median_incomeE + distancemiles, data = airport_pairs)

Residuals:
    Min      1Q  Median      3Q     Max
-157946  -64607  -45710    4733 1200553

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    6337.08330 6753.79195   0.938    0.348
median_incomeE    1.02817    0.08111  12.676   <2e-16 ***
distancemiles   -18.66439    1.92008  -9.721   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 132000 on 9395 degrees of freedom
  (104 observations deleted due to missingness)
Multiple R-squared:  0.02337,   Adjusted R-squared:  0.02316
F-statistic: 112.4 on 2 and 9395 DF,  p-value: < 2.2e-16
```

The linear regression model to predict passenger demand for Air Carolina's proposed routes uses three main factors: median income at the destination city, the distance between origin and destination, and the total population of the destination CBSA. These variables capture the effects of economic capacity, convenience, and market size on air travel demand. The median income at the destination is a significant predictor, with a positive coefficient suggesting that cities with higher incomes attract more passengers. Specifically, a $1,000 increase in median income is linked to an additional 1,028 passengers annually, reflecting greater travel affordability in wealthier areas.

The distance between origin and destination has negatively affected passenger volume, reducing it by approximately 18.7 passengers per mile. This aligns with the idea that longer flights see less demand due to higher costs and time commitments. Although this effect is moderately significant, it suggests that while distance matters, it may be less consistent as a predictor compared to income or population size. The total population of the destination CBSA also positively influences demand, as larger populations typically generate more passengers. The model estimates that a 1,000-person increase in population is associated with 2.9 additional passengers annually. Though this effect is small, it is statistically significan and reinforces the demand advantage in more populated areas.

Overall, these predictors match with expected air travel patterns, where higher income and larger population increase demand, while longer distances decrease it. However, the model has a low R-squared of 0.023, explaining only 2.3% of the variation in passenger volume. This suggests that other unmodeled factors, such as tourism appeal, business needs, or seasonality, likely influence actual passenger demand, which may lead to deviations in predicted outcomes.

Forecasted Demands

Description: df [6 × 5]

	origin <chr>	dest <chr>	distancemiles <dbl>	median_income <dbl>	predicted_pass... <dbl>
1	RDU	PDX	2363	90451	55232.13
2	RDU	ELP	1606	59757	37802.43
3	RDU	THL	496	55344	53982.59
4	RDU	SMF	2345	50000	13977.59
5	PDX	RDU	2363	90451	55232.13
6	ELP	RDU	1606	59757	37802.43

I calculated the projected passenger demand for each proposed route, both from RDU to the destination and in the reverse direction. These predictions are summarized above. Among these routes, RDU to Portland (PDX) and RDU to Tallahassee (TLH) have the highest projected demand, with estimated passenger counts of 55,232 and 53,983. The route to Sacramento (SMF) has the lowest forecasted demand at 13,978 passengers, most likely due to the combination of longer distance and lower median income. The popularity of the routes to Portland and Tallahassee may have something to do with their moderate distance coupled with favorable socio-economic conditions in the destination areas.

Recommendations

Based on my projections, Air Carolina will have to prioritize launching the routes with the highest estimated demand (Portland and Tallahassee) to maximize initial returns. However, it is important to approach these estimates with caution. The low R-squared value indicates that many additional factors influence passenger demand beyond income, distance, and population. For example, airport size, number of terminals, local tourism patterns, business travel demands, and seasonal variations could significantly impact the actual demand observed once the routes are implemented. Therefore, I would recommend conducting further research on these factors and considering a trial period for new routes to gather real-world data that may enhance future predictions.