



Airfare Forecast

By: Shweta Agrawal
Date: May, 2, 2021

Problem Statement

Based on a comparison of a online airfare quote vs. models airfare prediction, a user can make a decision if he/she book the ticket right away or later for that certain route.



Motivation

According to FAA around 3 million passengers fly every day in and out USA

There are more than 10 routes with in the USA, around 1 million passengers fly quarterly.





Agenda:

1. Data Collection
2. EDA
3. Model
4. Evaluation
5. Application Deployment
6. Conclusion/ Takeaways

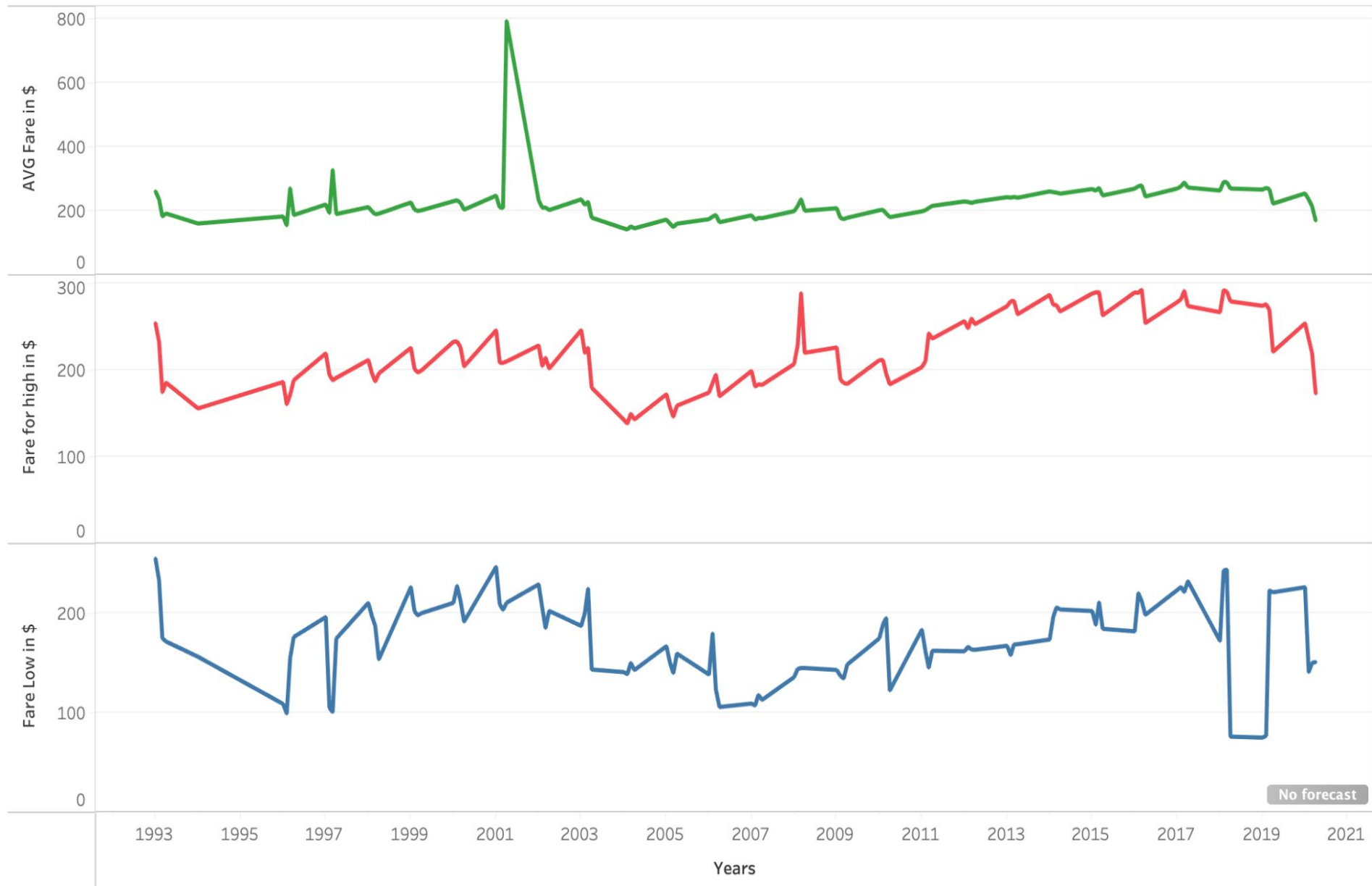
Data Collection

1. Consumer Airfare quarterly report based on airport pair markets for year 1993 to 2020.
2. Jet fuel price monthly report for year 1990 to 2020
3. Airport - city and city population monthly data for year 1993-2020
4. Monthly City weather data for year 1990 to 2020



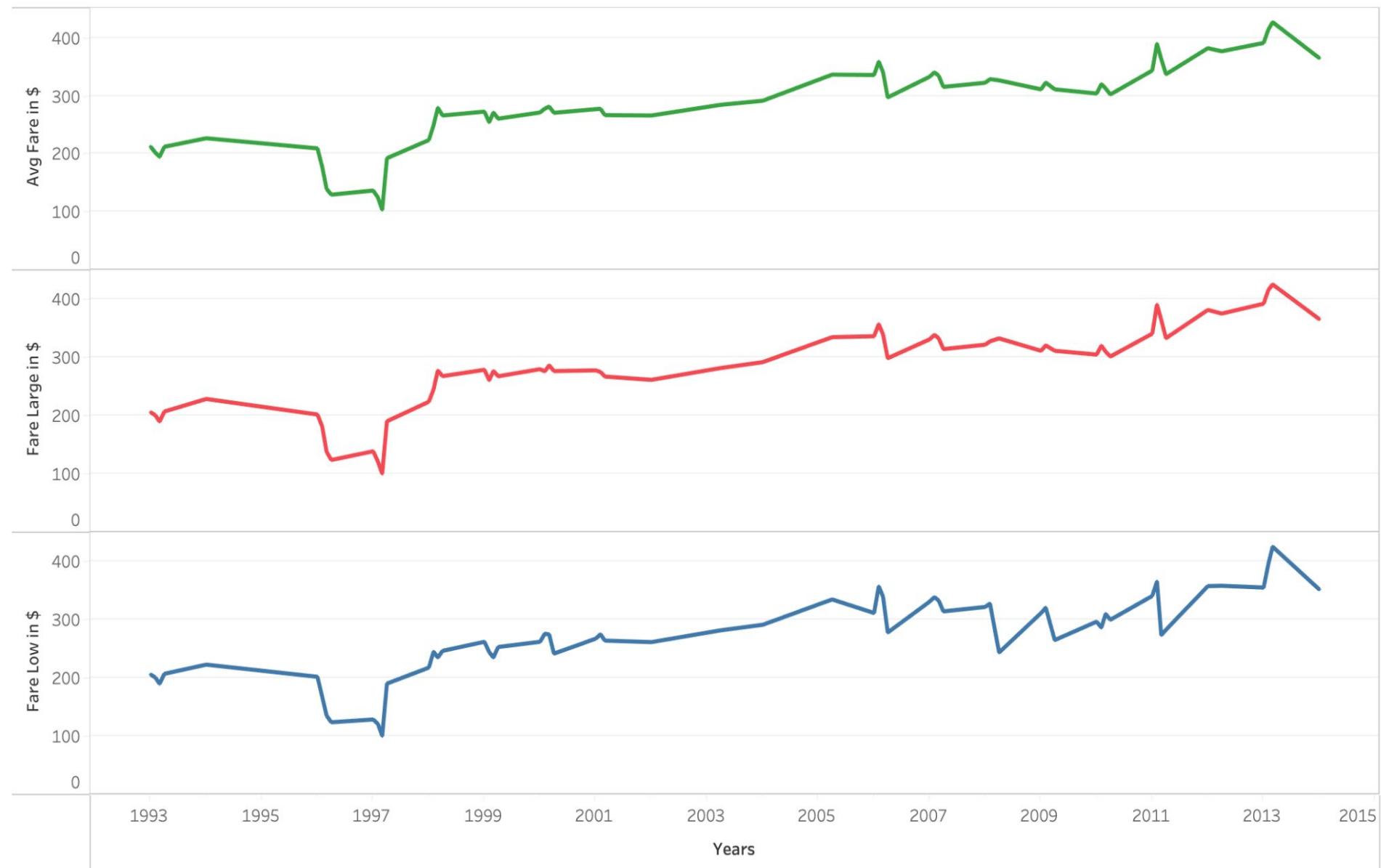
EDA/Preprocessing

NC to FL airfare comparision for low mkt share carrer to high mkt share carrer



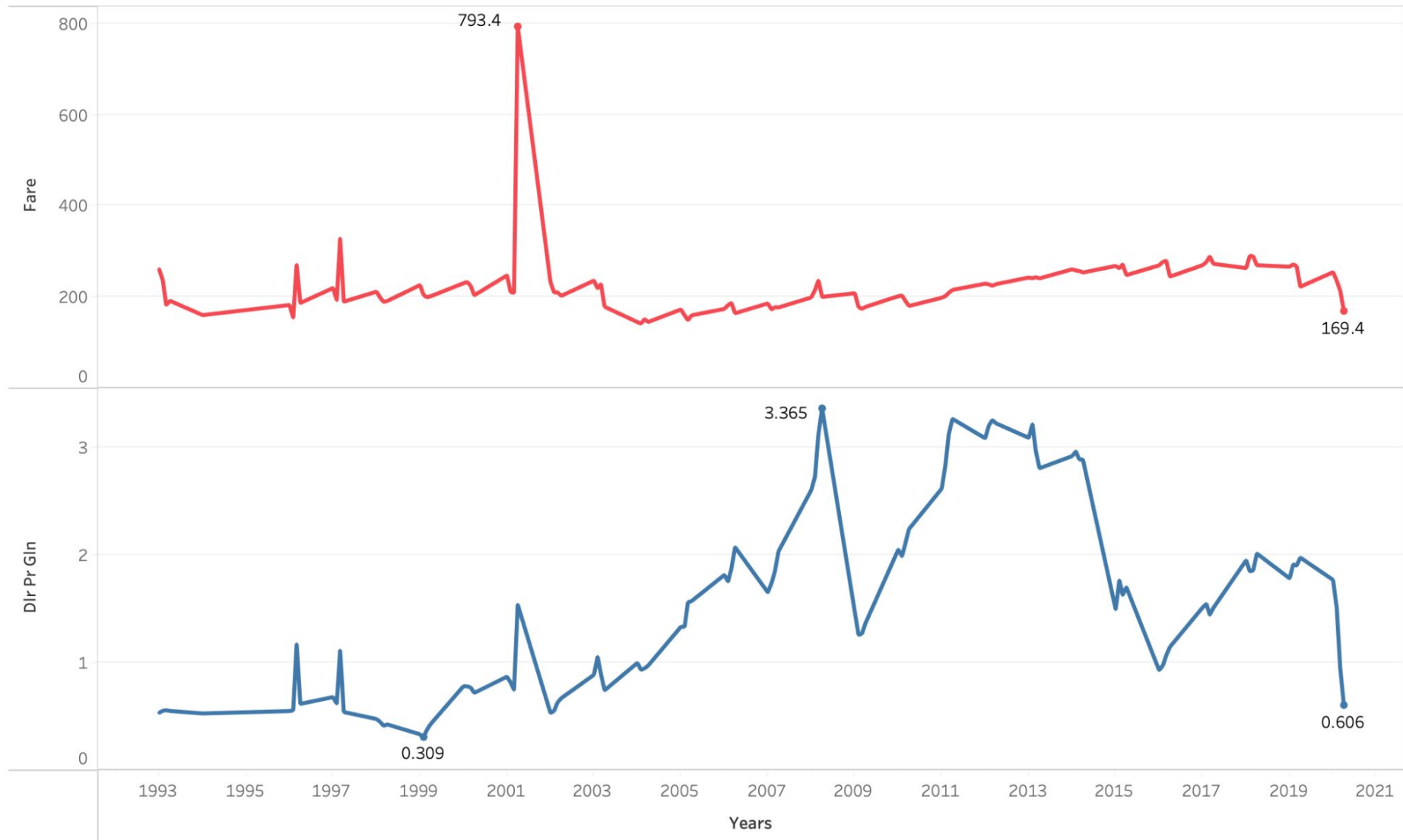
EDA/Preprocessing

TX to MI airfare comparison for low to high mkt share carriers



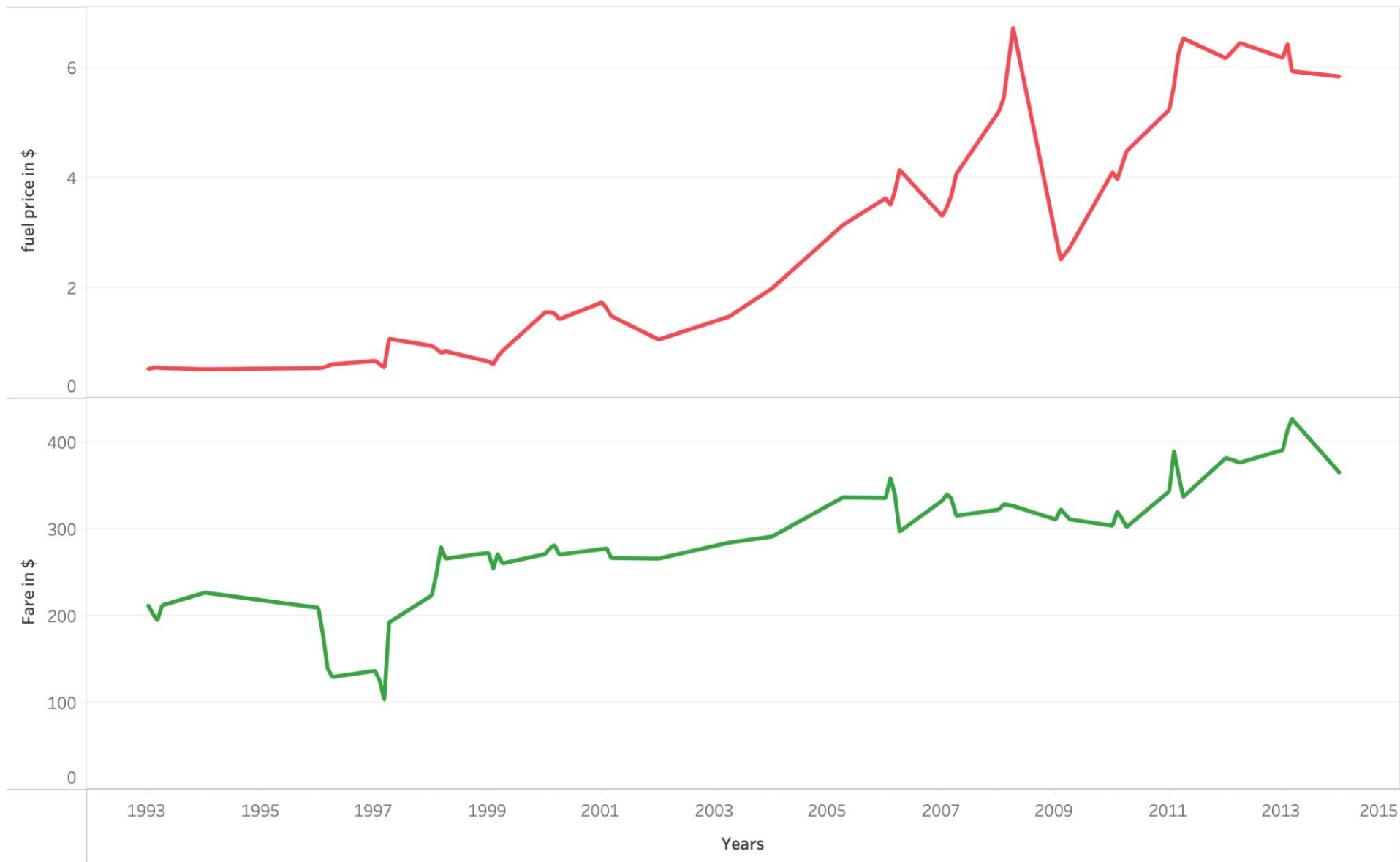
EDA/Preprocessing

NC to FL airfare and fuel price comparision



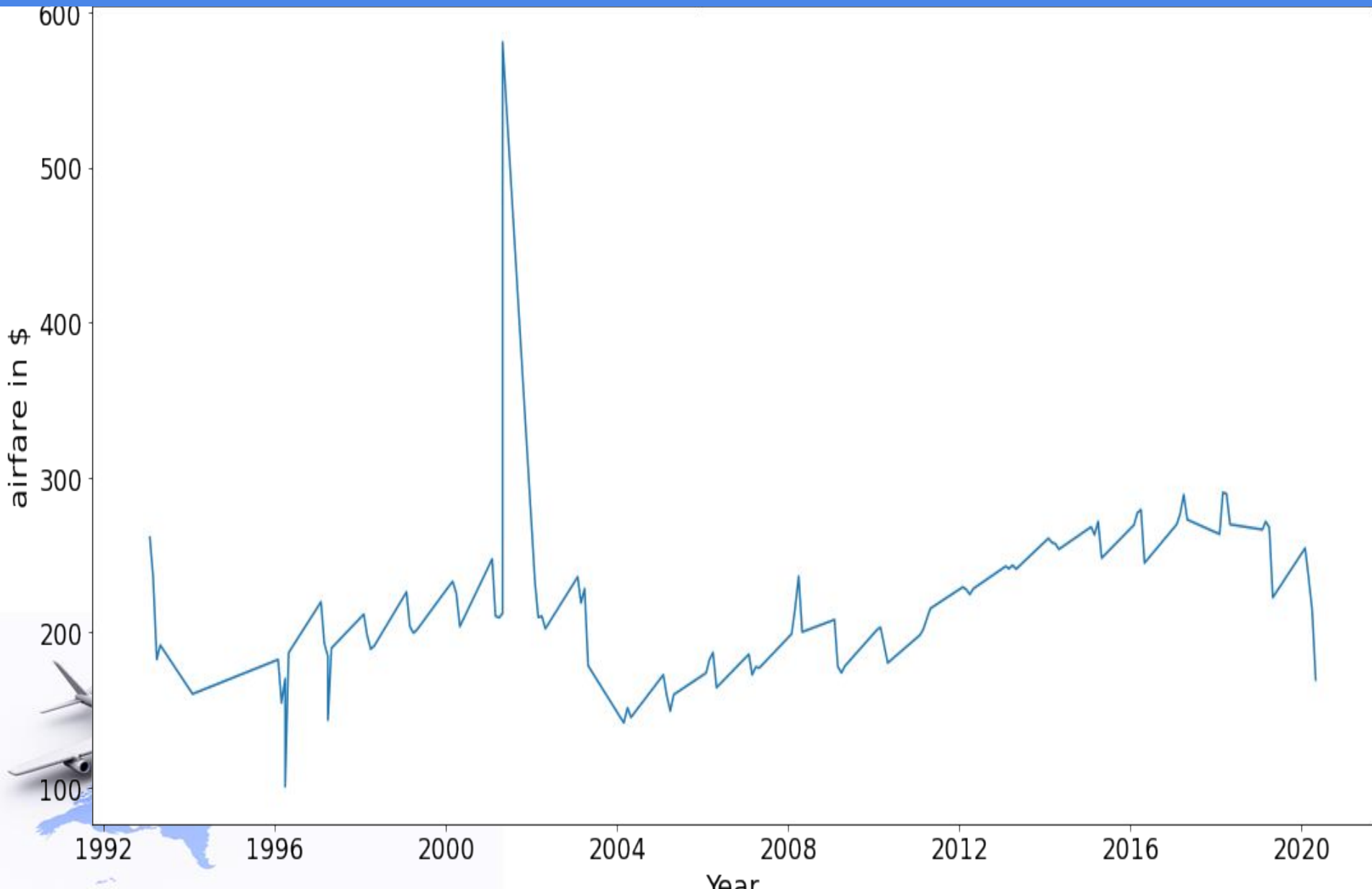
EDA/Preprocessing

TX to MI airfare comparision with fuel prices



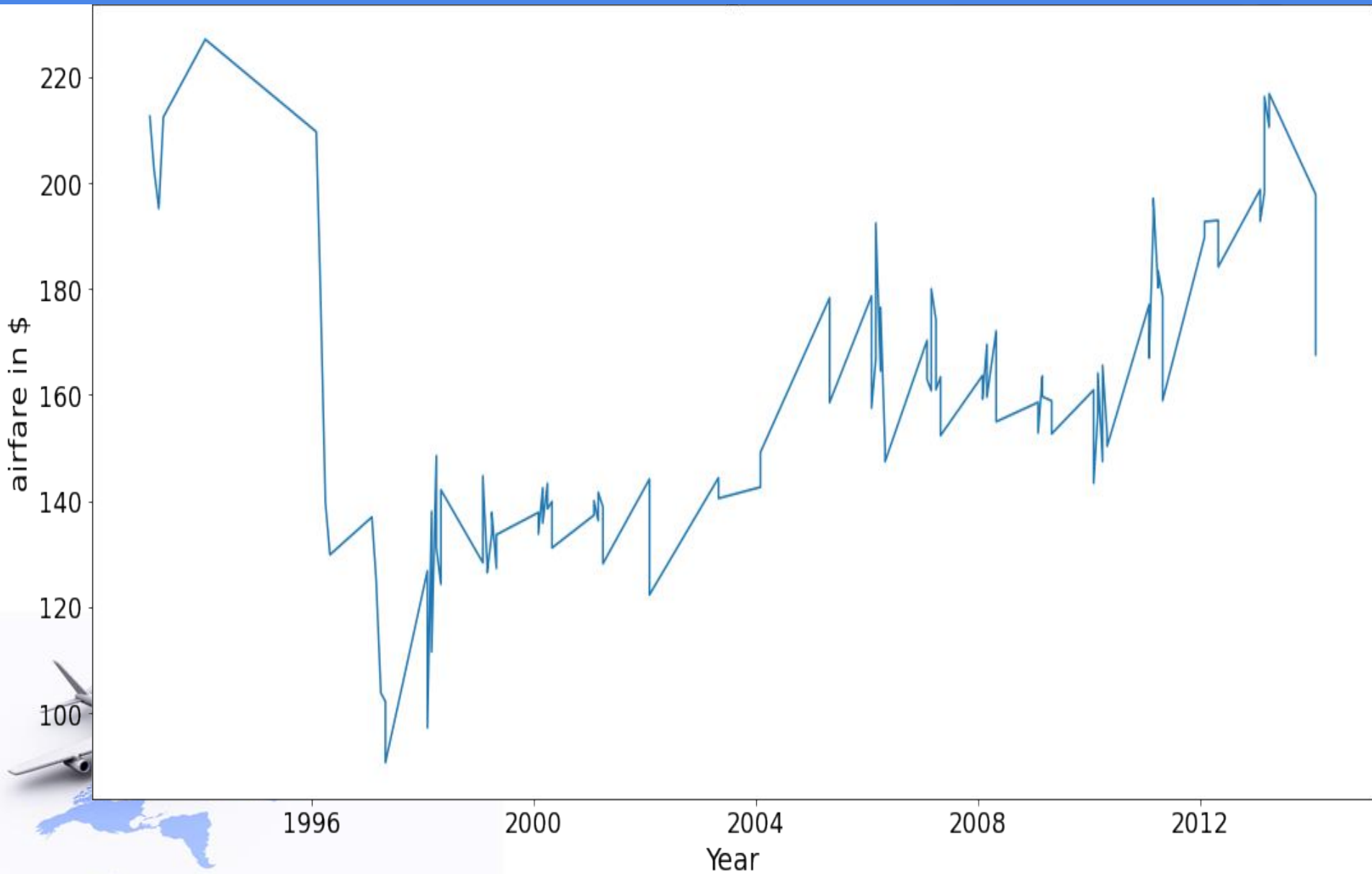
EDA / Preprocessing

Charlotte, NC to Tampa, FL airfare over the years

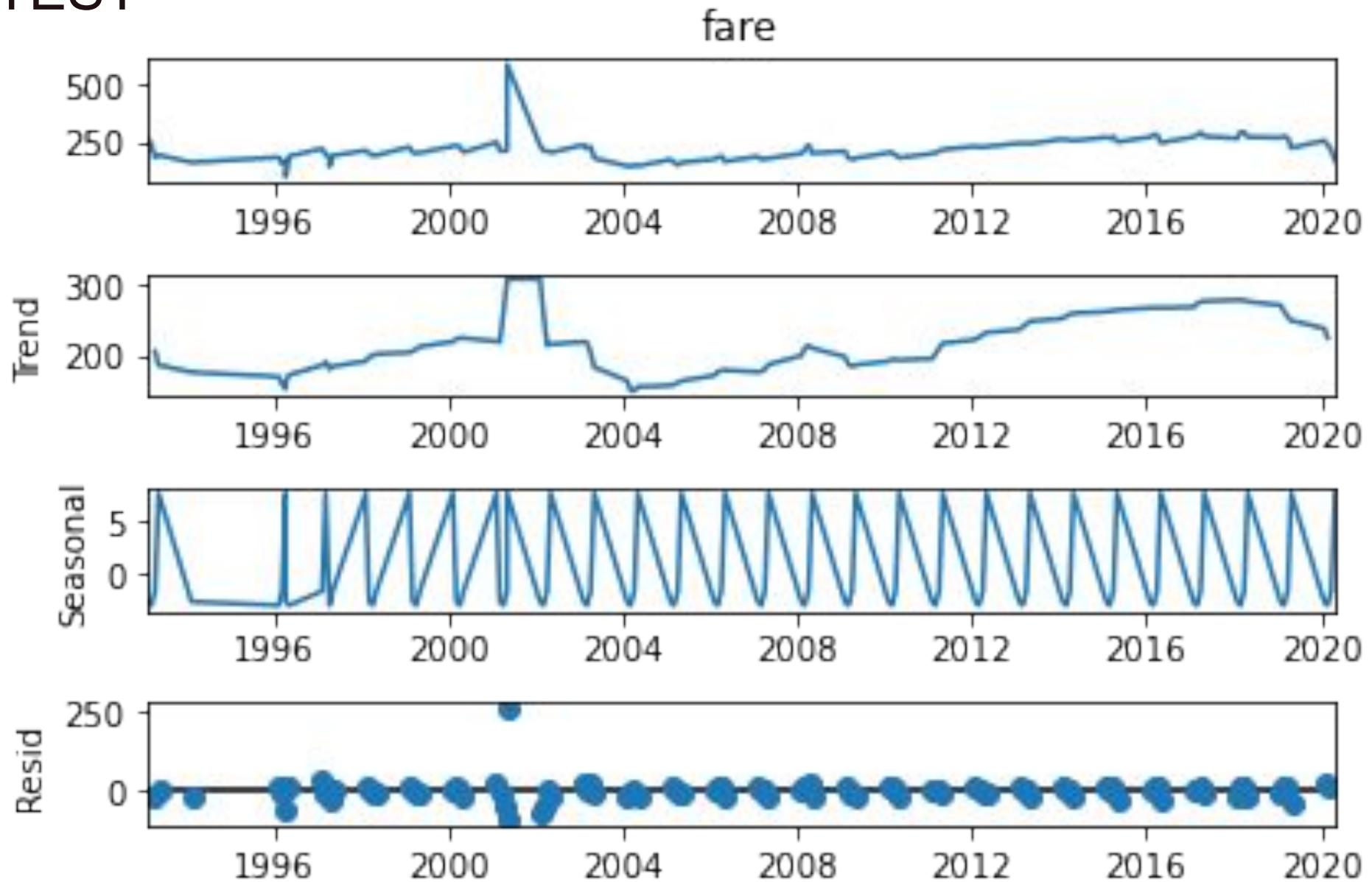


EDA/Preprocessing

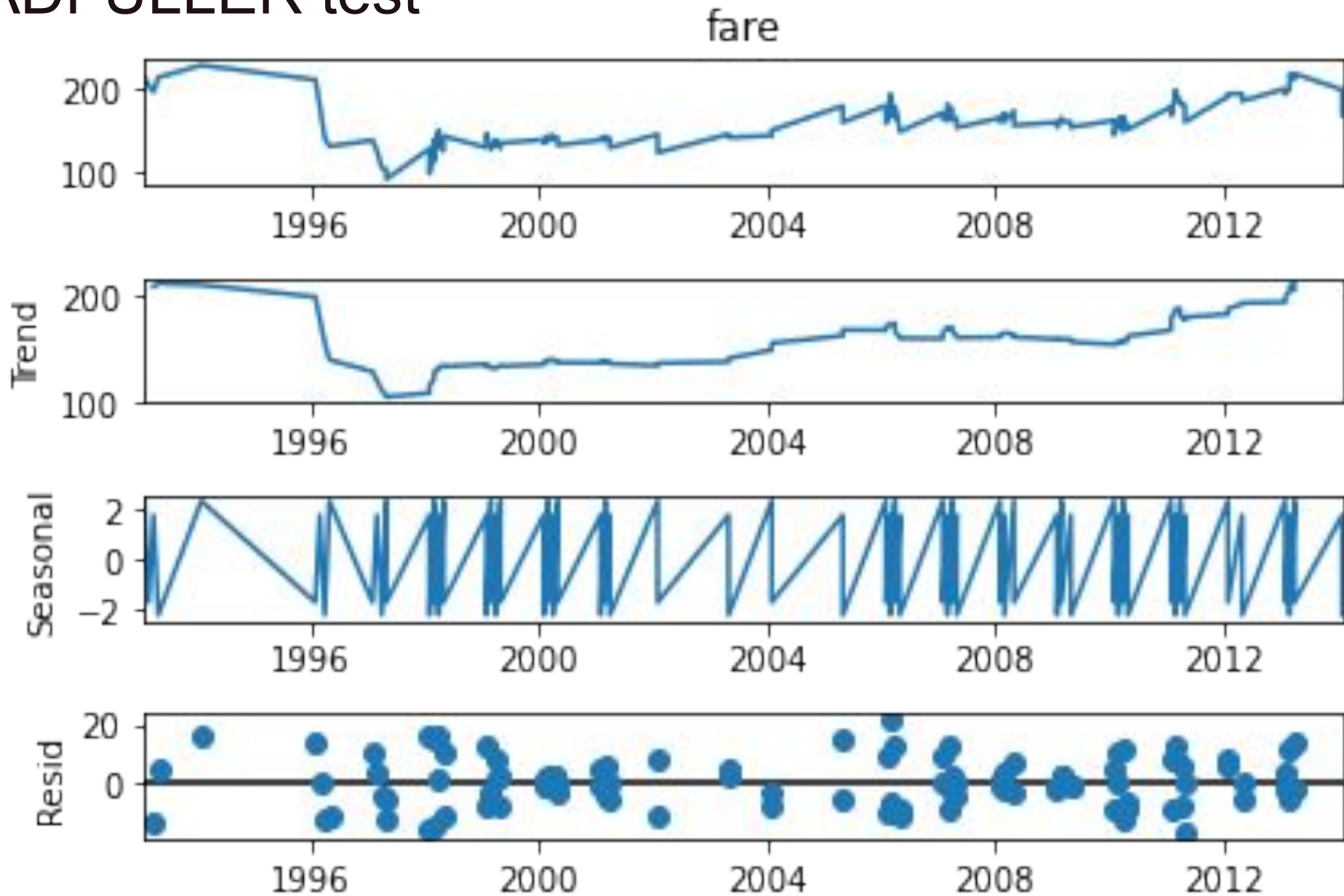
Dallas, TX to Vicksburg, MS airfare over time



NC to FL Data is not stationary conduct ADFULLER TEST



TX to MS data is not stationary conduct
ADFULLER test



Model

- Build multiple OLS (Original Least Squares) models with optimal features on routes (1996-2006)
 - Analyze Statsmodel Summary and ensure features fall below a 5% threshold for p value
 - Iteration: Remove features & retest new model (machine learning)
- Apply our best model on each routes (1996-2003)
- Apply our best model on unseen data for each route (2004-2007)



Weighted moving Average

As we know in the Moving average method, we equally weigh the past 'n' observations.

But we might encounter situations where each of the observation from the past 'n' impacts the forecast in a different way.

Such a technique which weighs the past observations differently is called Weighted Moving Average technique.



SARIMAX MODEL

- For seasonal ARIMA model first thing we do is identify if the data is stationary or non – stationary. if data is non-stationary we will try to make them stationary then we will process further.
- Dickey-Fuller Test : To check stationarity



Model performances on RMSE And R2

Model	RMSE	R2
OLS	36.7%	82%
Moving Average	43.7%	82%
SARIMAX	20%	94%



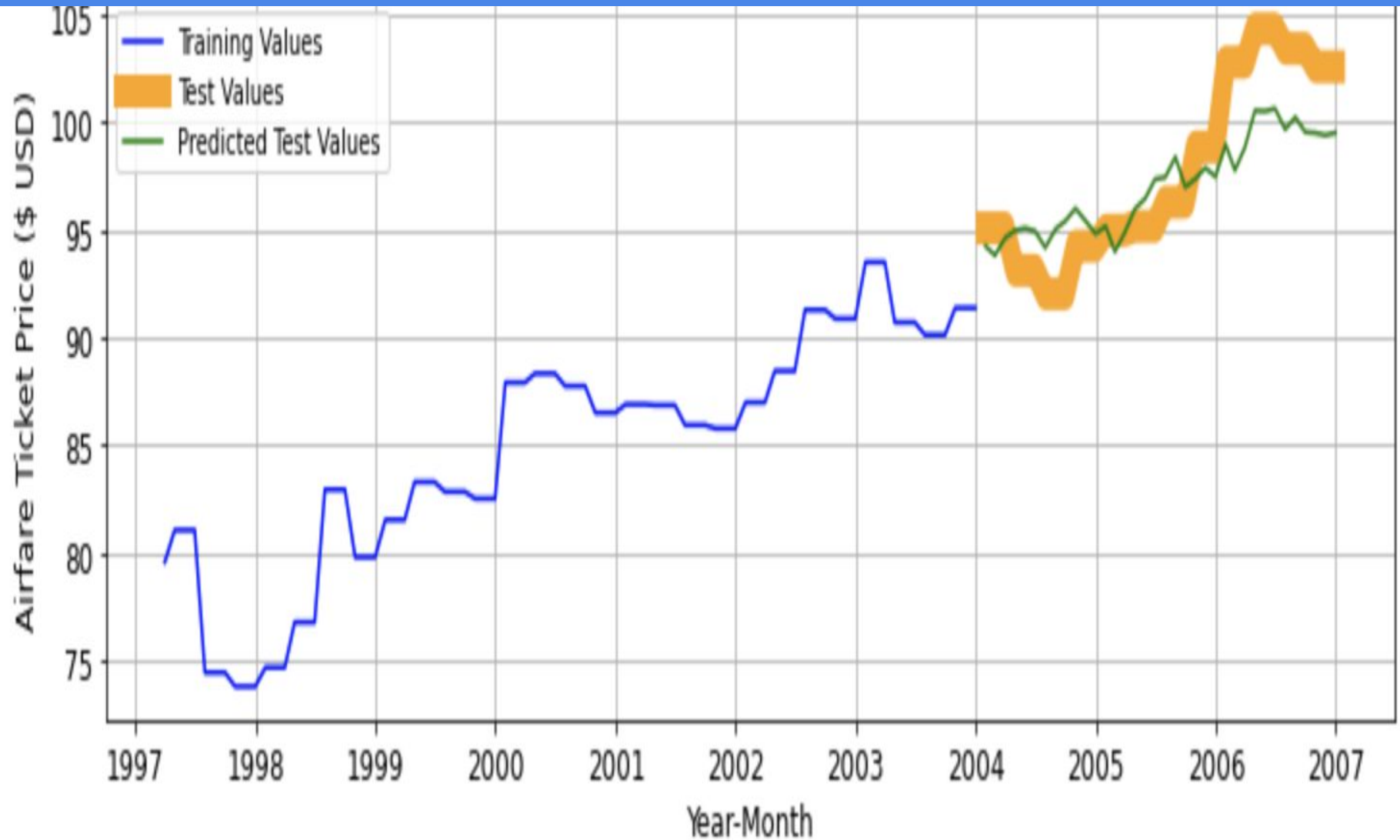
Model predictions

Train Test predictions For NC to FL airfare in \$



Model Predictions

Train Test Predictions for TX to MS airfare in \$



Web application and deployment

To forecast the quarterly airfare for future .

<https://ds-demo-airfare.pages.dev/>



Conclusion:

1. Got the predictions with R2 score 93% for more than 73 routes.
2. Beat the baseline R2 score 52% for NC to FL route with R2 score 93%
3. Beat the baseline R2 score 22% for TX to MS route with R2 score 88%



Takeaways:

1. Explore more routes.
2. Use graph theory to compute shortest route.
3. If I could get airfare data on daily basis then forecasting airfare for next week and so, would be more useful.
4. Build application in a more interactive and robust.



Thank you!!!

