BANA 630 Final Project - Southwest Airlines Co. Flight Assignments

Group: Joshua Cabal, Krish Viswanadhan Nair, Denise Becerra, Namrata Patil
Date: Dec 12, 2024

Contents:

Project Proposal - 3 Pages

Southwest's business model is distinct from that of other US airlines because it uses a point-to-point network. **Point-to-point transit** is a transportation system in which a plane, bus, or train travels directly to a destination, rather than going through a central hub. The concept of this project is to replicate a simplified and small portion of a point-to-point network using Southwest Airline flight data.

We will choose 2 distinct sets of nodes, one set of origin nodes and one set of destination nodes. We will look at all the data for a given month. Both the supply and the capacity of the nodes will be predicted using a model such as regression. The objective is to develop a Linear Programming model that optimizes the assignment of flights from these two distinct sets of airports within Southwest Airlines' network. The goal is to minimize total operational delays.

Phase II: Predictive Analysis - 12 Pages

Using publicly available airline data provided by the Bureau of Transportation Statistics (BTS), we aim to predict the following values for Southwest Airlines Co. for June 2025: **Origin Node Supply Amounts, Destination Node Demand Amounts, Delays (Node Edges).** Our origin airports are BWI, MDW, DAL, DEN, LAS, and the destination airports are LAX, OKC, SAN, SEA, LGA, CHS, DCA, HNL, MIA. Once these values are predicted, we will have the constraints for our point-to-point network as well as the values of node edges (interpretable as the delay cost for sending a flight from origin to destination).

Phase III: Prescriptive Analytics - 16 Pages

With the airport supply, demands, and route delays predicted for June 2025, we aim to use these values to formulate and solve a linear integer programming assignment problem. In this problem, the cost is the predicted delay on each route. Given that our total demand exceeds our supply, we cannot exceed the demand for each destination node, but we must assign all flights from each origin node, and finally we must meet a minimum number of flights for each route.

Professor, do not accept any PDFs without this digital signature: Joshua Cabal

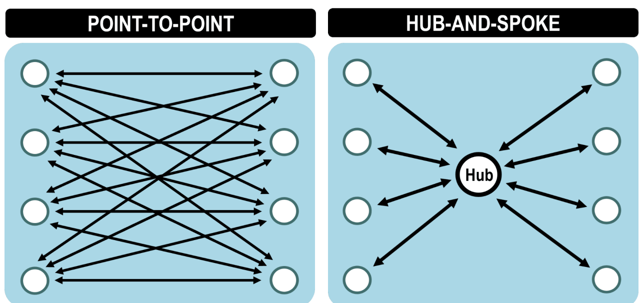
Digitally signed by Joshua Cabal
Date: 2024.12.08 16:14:35 -08'00'

BANA 630 Project Proposal

Group: Joshua Cabal, Krish Viswanadhan Nair, Denise Becerra, Namrata Patil
Date: Nov 18, 2024

Topic Selection - Optimizing Flight Assignments for Southwest Airlines' Using Linear Programming

Southwest's business model is distinct from that of other US airlines because it uses a point-to-point network. **Point-to-point transit** is a transportation system in which a plane, bus, or train travels directly to a destination, rather than going through a central hub.



The concept of the project is to replicate a simplified and small portion of a point-to-point network using Southwest Airline flight data.

We will choose 2 distinct sets of nodes, one set of origin nodes and one set of destination nodes. We will look at all the data for a given month. Both the supply and the capacity of the nodes will be predicted using a model such as regression.

The objective is to develop a Linear Programming model that optimizes the assignment of flights from these two distinct sets of airports within Southwest Airlines' network. The goal is to minimize total operational delays.

Practical Implications

Operational Efficiency: Optimizing flight assignments can lead to significant reductions in delays, enhancing the reliability of Southwest Airlines' service.
Cost Reduction: Minimizing delays and optimizing resource allocation directly translates to cost savings in fuel, crew overtime, and maintenance.
Passenger Satisfaction: Improved on-time performance fosters greater passenger satisfaction and loyalty, crucial for Southwest's brand reputation.
Competitive Advantage: Efficient scheduling allows Southwest to better compete with other airlines by offering more reliable and cost-effective services.

Role of Predictive and Prescriptive Analysis

Incorporating Predictive Analysis:

- **Delay Prediction:** Utilize historical flight data to forecast potential delays on specific routes.
- **Supply / Demand Forecasting:** Predict future flight demand between chosen origin and destination airports to inform supply constraints in the LP model.

Incorporating Prescriptive Analysis:

- **Optimal Flight Assignments:** Apply the LP model to prescribe the most efficient flight assignments that minimize delays.

Integration:

By combining predictive insights with prescriptive optimization, the project ensures that flight assignments are not only efficient under current conditions but also adaptable to future variations in demand and operational challenges.

Known and Unknown Variables

Known Variables:

- **Origin Airports:** Two distinct sets of major Southwest Airlines' source airports
- **Destination Airports:** Two distinct sets of major destination airports
- **Scheduled Departures:** The planned departure times for each flight.
- **Actual Departures:** The recorded actual departure times for each flight.
- **Actual Delays of Each Flight:** Historical data on reported delays for each origin-destination route.

Origin Airports: BWI, MDW, DAL, DEN, LAS *(Chosen per the base airports as defined here)*

Destination Airports: LAX, OKC, SAN, SEA, LGA, CHS, DCA, HNL, MIA *(Chosen to have an even coverage around the United States)*

Unknown Variables:

- **Predicted: Origin Node Supply Amounts**
- **Predicted: Destination Node Demand Amounts**
- **Predicted: Delays (Node Edges)**
- **Prescriptive: Optimal Flight Assignments:** The number of flights assigned from each origin airport to each destination airport to minimize total delays.

Objective Function and Limitations

Objective Function

Assuming that every flight from a source to a destination has some delay, we aim to minimize the total delay within the network model.

Constraints

- Solution must adhere to the predicted supply of the source nodes and the predicted demand of the destination nodes.
- Due to range constraints some source airports do not service some destinations.

Dataset Source

https://www.transtats.bts.gov/ontime/ - Departure and arrival statistics (scheduled departure time, actual departure time, scheduled elapsed time, departure delay, wheels-off time and taxi-out time) by airport and airline; airborne time, cancellation and diversion by airport and airline.

We are exporting the data per the following configuration:

Click “**Departures**”

Airline: Southwest Airlines Co.
Statistics: All
Years: 2000-2024
Days: All

Origin Airports: BWI, MDW, DAL, DEN, LAS

Project Phase II: Predictive Analytics

Group: Joshua Cabal, Krish Viswanadhan Nair, Denise Becerra, Namrata Patil
Date: Dec 3, 2024

Introduction

Using publicly available airline data provided by the Bureau of Transportation Statistics (BTS), we aim to predict the following values for Southwest Airlines Co. for June 2025: **Origin Node Supply Amounts, Destination Node Demand Amounts, Delays (Node Edges).** Our origin airports are BWI, MDW, DAL, DEN, LAS, and the destination airports are LAX, OKC, SAN, SEA, LGA, CHS, DCA, HNL, MIA. Once these values are predicted, we will have the constraints for our point-to-point network as well as the values of node edges (interpretable as the delay cost for sending a flight from origin to destination).

The code, notebooks, and logs related to this project are contained in the project GitHub Repository.

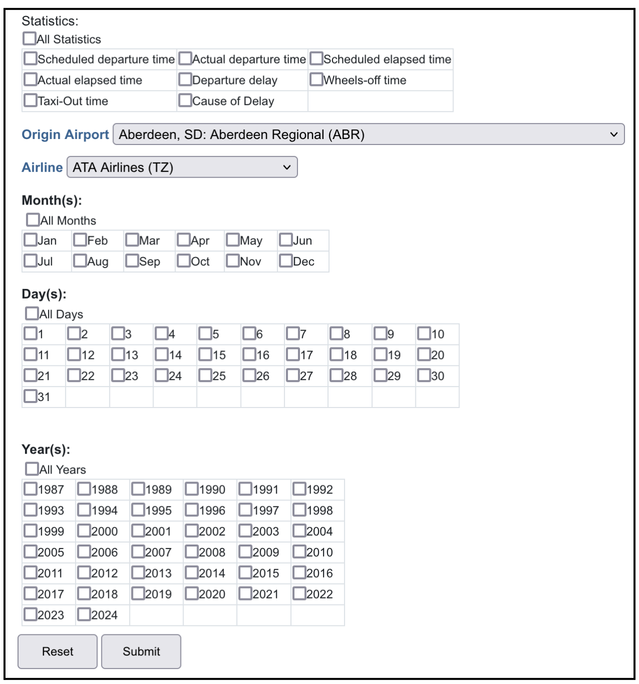
Dataset Preparation

The Bureau of Transportation Statistics (BTS), part of the United States Department of Transportation, is a government office that compiles, analyzes, and publishes information on the nation's transportation systems across various modes; and strives to improve the DOT's statistical programs through research and the development of guidelines for data collection and analysis. BTS is a principal agency of the U.S. Federal Statistical System.

The dataset is publicly accessible and is from the website https://www.transtats.bts.gov/ontime/. The data includes the following detailed departure and arrival statistics: scheduled departure time, actual departure time, scheduled elapsed time, departure delay, wheels-off time and taxi-out time by airport and airline; airborne time, cancellation and diversion by airport and airline.

Export Settings

The data was exported per the following configuration: Navigate to “**Departures**”.



Airline: Southwest Airlines Co.
Statistics: All
Years: 2000-2024
Days: All

Origin Airports: BWI, MDW, DAL, DEN, LAS

Because we aim to predict the data for 5 origin nodes, we have 5 different CSV files that were exported.

Data Dictionary

Below is the data dictionary provided by the BTS. The highlighted fields are the pertinent fields that will be used. Origin airport is also pertinent and is implied by the file name, which was named after data export.

Column Name	Definition
Carrier Code	A unique identifier assigned to the airline operating the flight.
Date (MM/DD/YYYY)	**The date on which the flight is scheduled to depart, formatted as month/day/year.**
Flight Number	The unique number assigned to the flight, used to identify it among other flights.
Tail Number	The unique alphanumeric identifier assigned to the aircraft.
Destination Airport	**The airport where the flight is scheduled to land.**
Scheduled Departure Time	The planned time at which the flight is scheduled to depart from the origin airport.
Actual Departure Time	The actual time at which the flight departed from the origin airport.
Scheduled Elapsed Time (Minutes)	The planned duration of the flight from departure to arrival, measured in minutes.
Actual Elapsed Time (Minutes)	The actual duration of the flight from departure to arrival, measured in minutes.
Departure Delay (Minutes)	**The difference between the actual departure time and the scheduled departure time, measured in minutes.**
Wheels-off Time	The exact time when the aircraft's wheels leave the ground during departure.
Taxi-Out Time (Minutes)	The time taken by the aircraft to taxi from the gate to the runway for takeoff, measured in minutes.
Delay Carrier (Minutes)	The amount of departure delay attributed to the carrier (airline), measured in minutes.
Delay Weather (Minutes)	The amount of departure delay attributed to weather conditions, measured in minutes.
Delay National Aviation System (Minutes)	The amount of departure delay attributed to the National Aviation System (e.g., air traffic control), measured in minutes.
Delay Security (Minutes)	The amount of departure delay attributed to security-related issues, measured in minutes.
Delay Late Aircraft Arrival (Minutes)	The amount of departure delay attributed to the late arrival of the aircraft from a previous flight, measured in minutes.

Data Dictionary for Dataset: Airline On-Time Statistics

Key Features

This section is a brief section on the dataset characteristics. We have the recorded counts for origin and destination nodes, the yearly amounts, and the distribution of departure delays by origin.

Destination Airport	CHS	DCA	HNL	LAX	LGA	MIA	OKC	SAN	SEA	
Origin Airport										
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	1395	0	0	1176	377	410	384	1331	396	**5469**
Chicago, IL Chicago Midway International (MDW)	832	1779	0	3724	2584	272	517	2618	2263	**14589**
Dallas, TX Dallas Love Field (DAL)	275	1178	0	1166	1083	173	2775	974	197	**7821**
Denver, CO Denver International (DEN)	190	0	0	3174	913	107	1490	2831	2180	**10885**
Las Vegas, NV Harry Reid International (LAS)	0	0	226	8516	0	0	946	8454	2430	**20572**
	2692	**2957**	**226**	**17756**	**4957**	**962**	**6112**	**16208**	**7466**	

Table 1: Cross tabulation of the 59366 recorded flights. Values of 0 mean there was no recorded flight data and will therefore be excluded from the model predictions and the network.

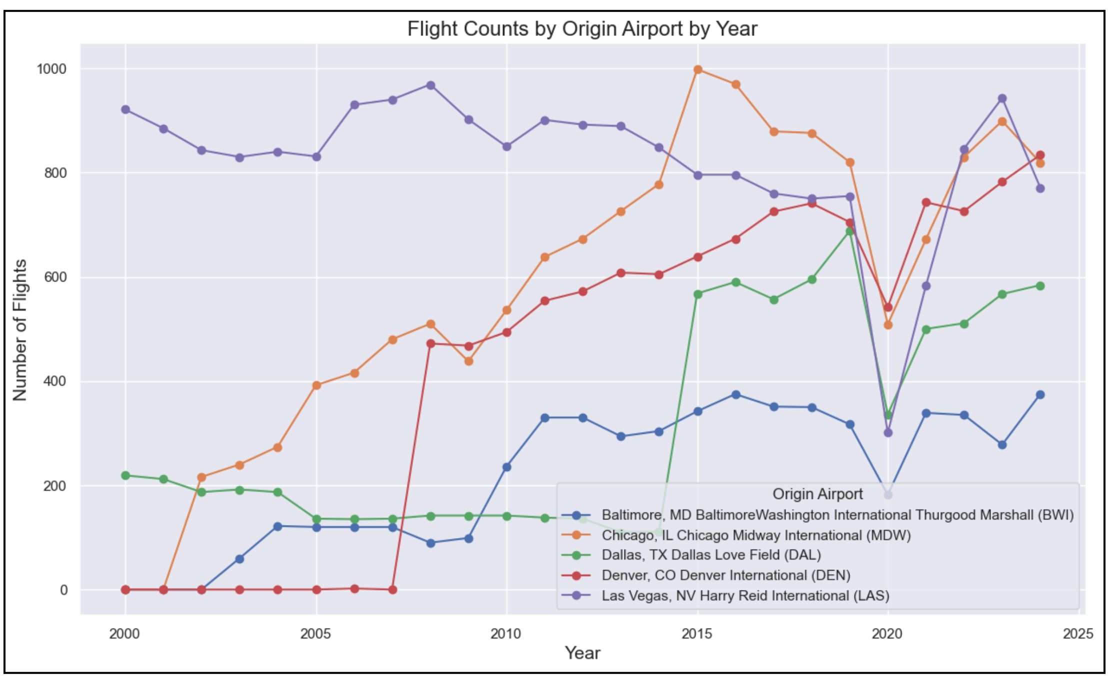


Figure: Flights Counts by Origin Airport by Year

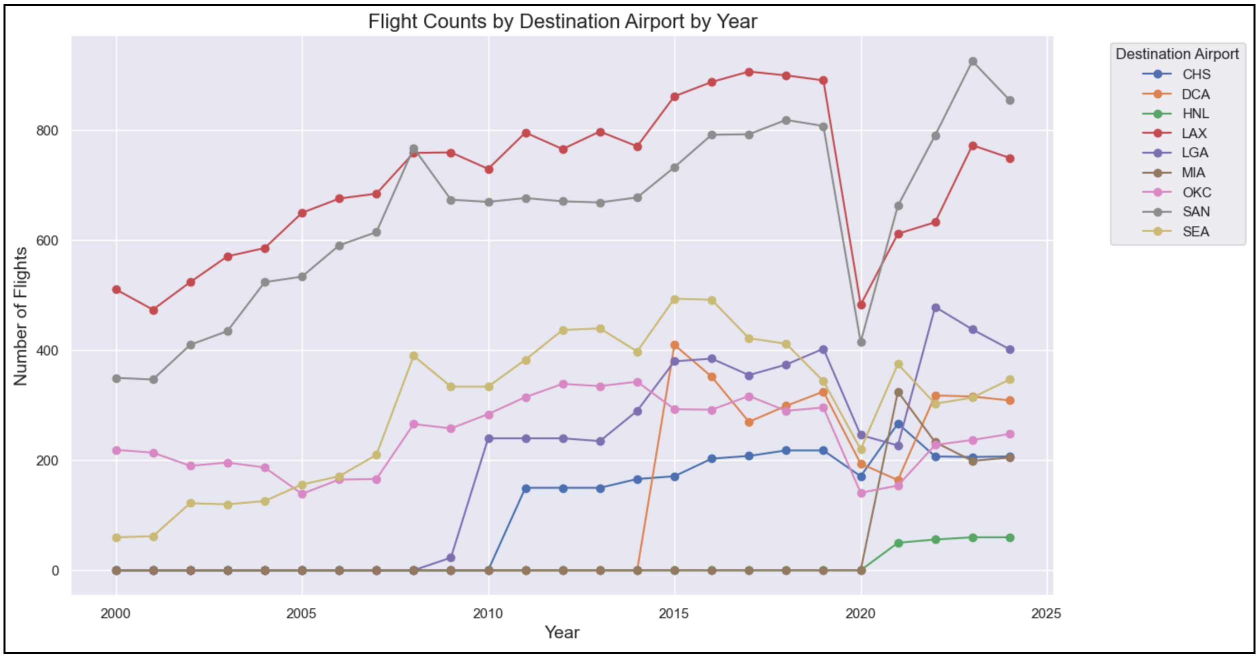


Figure: Flights Counts by Destination Airport by Year

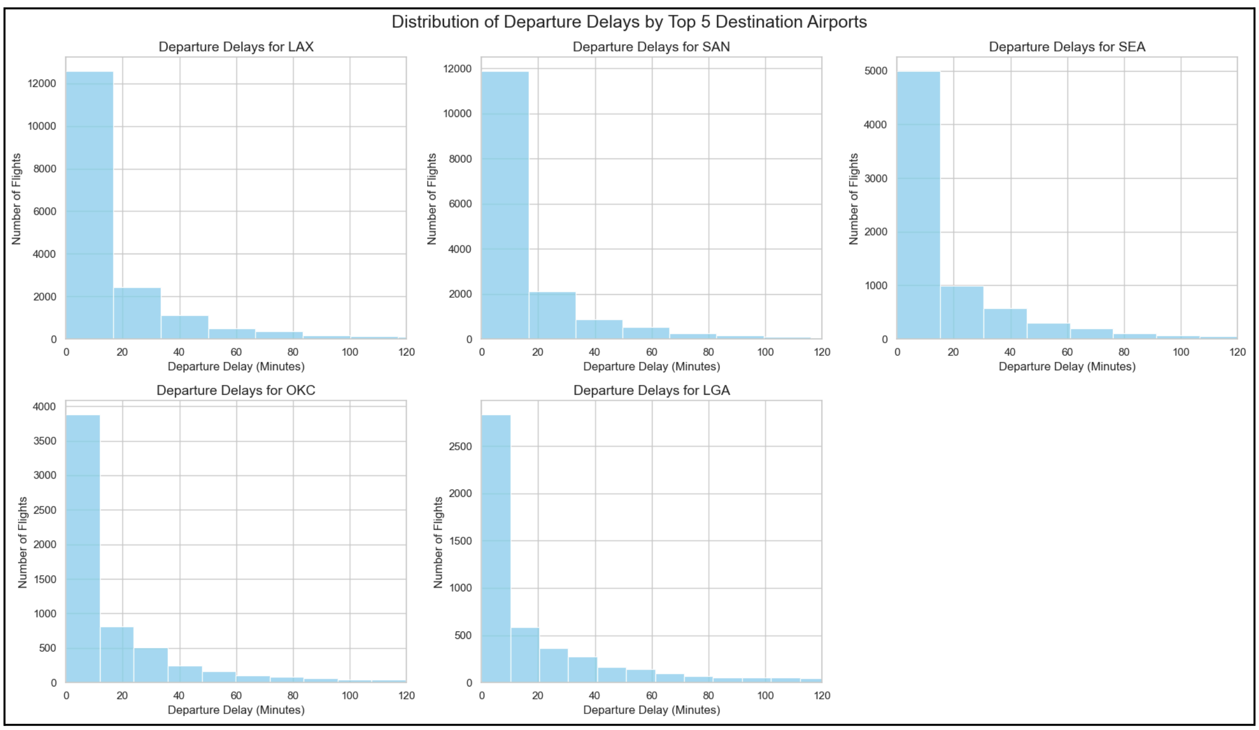


Figure: Distribution of Departure Delay. X-Axis is limited to 120 minutes.

Preprocessing Operations

Because we are predicting the data for 5 origin nodes, we have 5 different CSV files that were exported. In order to keep operations consistent, a function library file was created flight_forecasting.py.

Data Cleaning After Export

All operations were contained in the function `clean_and_prepare_data(df)`. This function is defined in the library file flight_forecasting.py.

The following operations were completed immediately against the exported datasets:

1. Convert Date field to DateTime data type
2. Drop any records in which `Date` was NULL
3. Drop any records in which `Destination Airport` was NULL
4. Drop any records in which the `Destination Airport` was not one of the following: `[LAX, OKC, SAN, SEA, LGA, CHS, DCA, HNL, MIA]`
5. Fill missing departure delays with the median delay of the route
6. Replace negative delays (early departures) or no delay recorded with zero
7. Reset dataframe index after all operations

Aggregating All Data into a Single DataFrame

In order to easily pass the data into the predictive models, all data was merged onto a single DataFrame. This was handled by the defined function `clean_and_prepare_data(df)` (flight_forecasting.py) and the first for loop of the `main()` function. A description of this block is given below:

1. Initialize empty master DataFrame
2. For the CSV file (each origin airport data), we extract the origin airport name from the file name, and append this value onto a column for the entire table
3. Execute the function `clean_and_prepare_data(df)` and append the final cleaned DataFrame to the master DataFrame previously initialized
4. Repeat for each CSV file

Predictive Model Choice

Predictive Objective is the following- To build predictive models to forecast the following values for June 2025:

1. **Origin Node Supply Amounts**
    a. We have 5 origin nodes
2. **Destination Node Demand Amounts**
    a. We have 9 destination nodes
3. **Delay (Cost) Amount from each Origin Node to Each Destination Node**
    a. We have 45 combinations

Thus, we have a total of 59 different values to predict. Our approach was not to pick a *single* model to forecast for each value, but instead, to build several prediction models for each value with *different* methods and then choose the best performing model by RSME. The models that were implemented and tested were the following:

1. **Naïve Forecast:** Predicts the next value as the last observed value in the training data.
2. **Linear Regression:** Uses a linear model to predict future values based on the trend in the historical data. (Y value is delay and X value is year)
3. **ARIMA (AutoRegressive Integrated Moving Average):** A time series model that captures autocorrelations in the data. Configured with parameters (1,1,1) indicating one autoregressive term, one differencing operation, and one moving average term.
4. **Exponential Smoothing:** A time series forecasting method that applies weighted averages with exponentially decreasing weights over time.
5. **Prophet:** A forecasting tool developed by Facebook, suitable for time series with strong seasonal effects and multiple seasons.

For example, we will predict the origin node amounts for airport BWI with the 5 different models above. We will predict the values for the next available year, and after walking through each year and tracking the performance via RSME, the best performing model will be selected to predict the value for June 2025.

We chose these models because all of the above models are time series and are based only on the historical data. We chose to predict based only on historical data (trends of the previous years) because building predictive models with the features of the exported BTS data is unsuitable. The columns of the exported dataset are either not relevant or not suitable to predict total delay. Using the delay component fields would lead to high multicollinearity and all other fields are irrelevant. Therefore, the forecasting methods mentioned above were deemed to be the most appropriate algorithms to use given the nature of the dataset and the problem.

Model Training and Evaluation

The process for model fit and testing is as follows:

1. **Models are fit using the previous year(s), then the next available year is predicted.**
    a. For example, use the data from Years 2000 and 2001 to predict the value for year 2002.
2. **Record the residual**
    a. Record the error between the predicted value for 2002 and the actual value for 2002.
3. **Repeat these two steps until all available year have been predicted**
    a. Walk through all available years until 2025 is reached.

The functions which handle the above operations are the five different forecast functions defined in the library file.

IATA Code	PREDICTED SUPPLY	Model Used
BWI	375	Naïve Forecast
MDW	800	Naïve Forecast
DAL	584	ARIMA(1,1,1)
DEN	834	Naïve Forecast
LAS	771	Naïve Forecast
	3364	

Origin Airport Predictions

IATA	PREDICTED DEMAND	Model Used
LAX	750	Naïve Forecast
OKC	248	Naïve Forecast
SAN	855	Naïve Forecast
SEA	347	Naïve Forecast
LGA	402	Naïve Forecast
CHS	207	Naïve Forecast
DCA	309	Naïve Forecast
HNL	60	Naïve Forecast
MIA	205	Naïve Forecast
	3383	

Destination Airport Predictions

Origin Airport	Destination Airport	2025 Predicted Avg Departure Delay (Minutes)	Model Used
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	CHS	16.01	Linear Regression
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	LAX	32.58	Naïve Forecast
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	LGA	26.64	Naïve Forecast
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	MIA	25.15	Linear Regression
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	OKC	20.17	Naïve Forecast
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	SAN	12.9	Linear Regression
Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)	SEA	19.75	Naïve Forecast
Chicago, IL Chicago Midway International (MDW)	CHS	17.71	Linear Regression
Chicago, IL Chicago Midway International (MDW)	DCA	16.69	Linear Regression
Chicago, IL Chicago Midway International (MDW)	LAX	23.56	Linear Regression
Chicago, IL Chicago Midway International (MDW)	LGA	15.33	Linear Regression
Chicago, IL Chicago Midway International (MDW)	MIA	19.41	Naïve Forecast
Chicago, IL Chicago Midway International (MDW)	OKC	27.19	Naïve Forecast
Chicago, IL Chicago Midway International (MDW)	SAN	21.68	Linear Regression
Chicago, IL Chicago Midway International (MDW)	SEA	22.74	Linear Regression
Dallas, TX Dallas Love Field (DAL)	CHS	15.92	Linear Regression
Dallas, TX Dallas Love Field (DAL)	DCA	17.59	Linear Regression
Dallas, TX Dallas Love Field (DAL)	LAX	22.98	Linear Regression

Dallas, TX Dallas Love Field (DAL)	LGA	20.68	Linear Regression
Dallas, TX Dallas Love Field (DAL)	MIA	27.04	Linear Regression
Dallas, TX Dallas Love Field (DAL)	OKC	24.75	Prophet
Dallas, TX Dallas Love Field (DAL)	SAN	18.84	Linear Regression
Dallas, TX Dallas Love Field (DAL)	SEA	19.76	Linear Regression
Denver, CO Denver International (DEN)	CHS	14.9	Naïve Forecast
Denver, CO Denver International (DEN)	LAX	23.05	Linear Regression
Denver, CO Denver International (DEN)	LGA	32.35	Naïve Forecast
Denver, CO Denver International (DEN)	MIA	25.13	Naïve Forecast
Denver, CO Denver International (DEN)	OKC	24.25	Linear Regression
Denver, CO Denver International (DEN)	SAN	21.75	Naïve Forecast
Denver, CO Denver International (DEN)	SEA	22.48	Linear Regression
Las Vegas, NV Harry Reid International (LAS)	HNL	15.47	Naïve Forecast
Las Vegas, NV Harry Reid International (LAS)	LAX	22.2	Naïve Forecast
Las Vegas, NV Harry Reid International (LAS)	OKC	17.39	ARIMA(1,1,1)
Las Vegas, NV Harry Reid International (LAS)	SAN	15.96	Prophet
Las Vegas, NV Harry Reid International (LAS)	SEA	20.54	ARIMA(1,1,1)

Delay Predictions

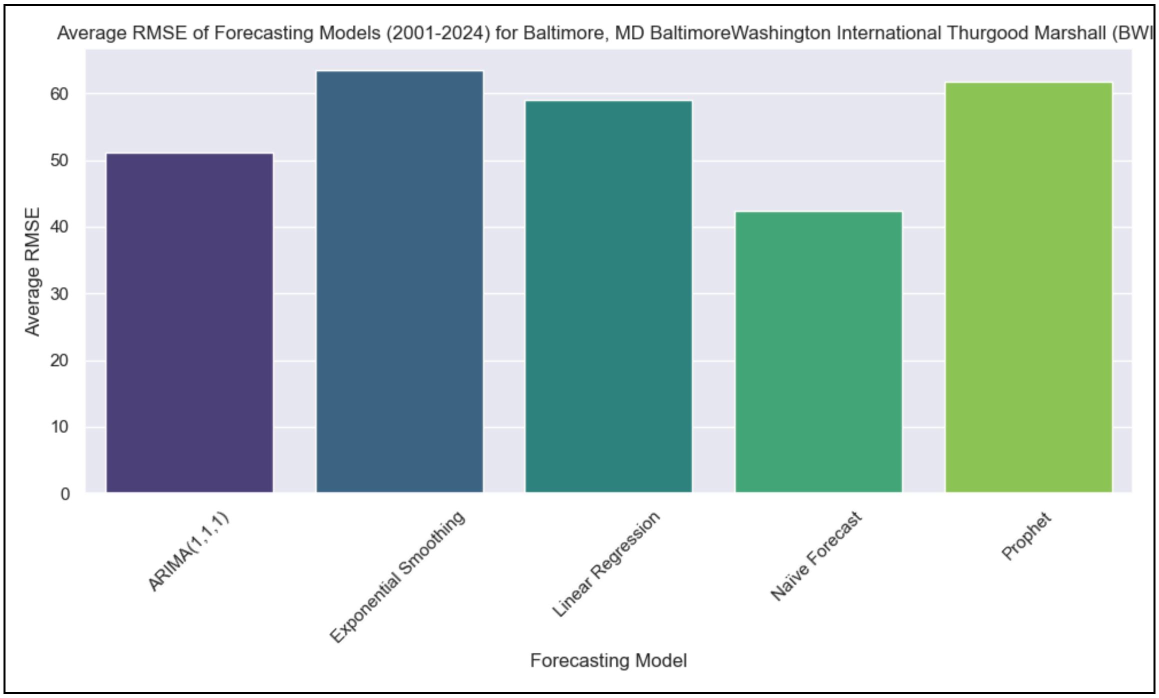
For model validation, we used the metrics RMSE and R2. Because the models walk through each available year and predict the next unseen year, all models are being tested on unseen data. The RMSE and R2 for all models used are logged and recorded in the directory 630-SOUTHWEST-SCHEDULING/Forecasts/. Within this directory are folders for each airport. Contained in each folder are the plots and logged results related to the performance of models.

Below are the plots and logged results for predicting the supply amount for origin node **Baltimore, MD BaltimoreWashington International Thurgood Marshall (BWI)**

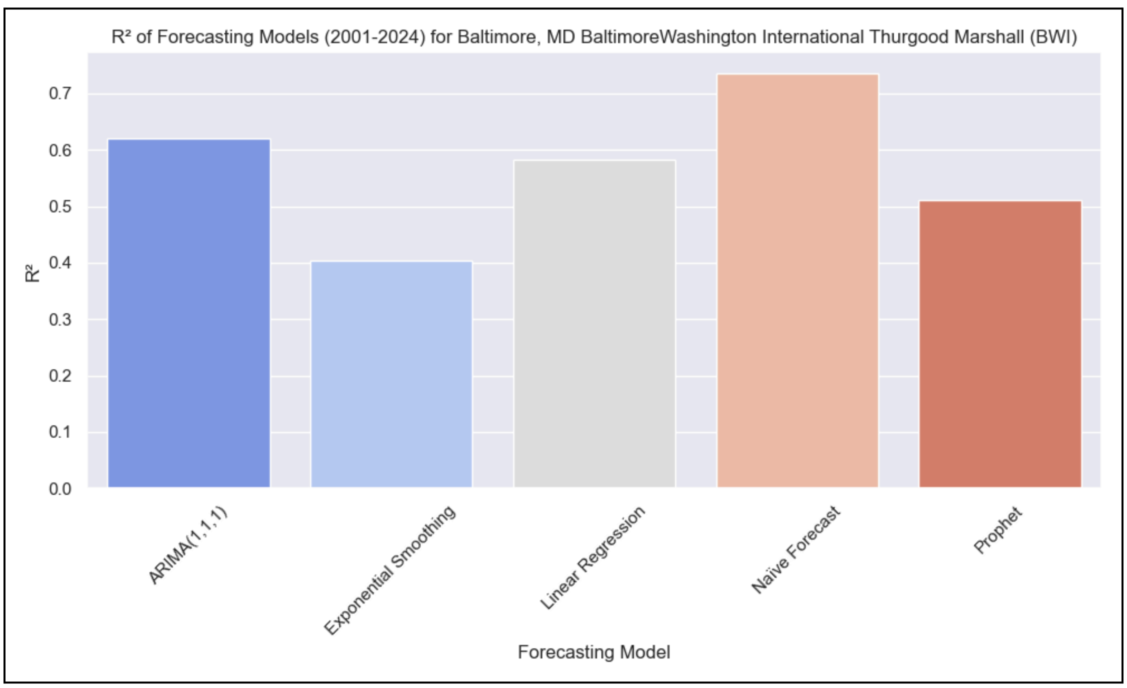
Forecast_vs_Actual

Model	RMSE	R2	Adjusted_R2
ARIMA(1,1,1)	51.06428033	0.6201997775	
Exponential Smoothing	63.4707051	0.4040174512	
Linear Regression	59.08562733	0.5835513676	0.5646218843
Naïve Forecast	**42.45833333**	**0.7356753722**	
Prophet	61.6894564	0.5109999701	

Performance_Summary



Average_RMSE



R2

Many files are generated in this process because we repeat this process for all predicted values. As in, each origin node, each destination node, and each route has had their related data predicted and results logged. To view all results, navigate to 630-SOUTHWEST-SCHEDULING/Forecasts/.

Project Phase III: Prescriptive Analytics

Group: Joshua Cabal, Krish Viswanadhan Nair, Denise Becerra, Namrata Patil
Date: Dec 12, 2024

Introduction

With the airport supply, demands, and route delays predicted for June 2025, we aim to use these values to formulate and solve a linear integer programming assignment problem. In this problem, the cost is the predicted delay on each route. Given that our total demand exceeds our total supply, we cannot exceed the demand for each destination node, but we must assign all flights from each origin node, and finally we must meet a minimum number of flights for each route.

Problem Formulation

Nodes

1. O: Set of origin airports (e.g., $\mathrm{BWI}, \mathrm{MDW}, \mathrm{DAL}, \mathrm{DEN}, \mathrm{LAS}$)
2. D: Set of destination airports (e.g., $\mathrm{LAX}, \mathrm{OKC}, \mathrm{SAN}, \mathrm{SEA}, \mathrm{LGA}, \mathrm{CHS}, \mathrm{DCA}, \mathrm{HNL}, \mathrm{MIA}$)

Parameters

- Supply_o: Predicted supply at origin airport o
- Demand_d: Predicted demand at destination airport d
- $\mathrm{Delay}_{o,d}$: Predicted average departure delay (minutes) from origin o to destination d
- $\mathrm{MinMix}_{o,d} = 0.05 \times \mathrm{Supply}_o$: Minimum flight assignment mix for route (o, d)

Decision Variable

- $X_{o,d}$: Number of flights (or assigned units) from origin o to destination d

Objective Function

Minimize the total delay across all routes:

$$\text{Minimize} \quad Z = \sum_{o \in O} \sum_{d \in D} \text{Delay}_{o,d} \cdot X_{o,d}$$

Constraints

The total supply constraint is configured with an equality to ensure that all flights are sent. The node demands constraints as configured as maximums because demand exceeds supply in our models. The flight assignment constraint is a constraint that could be changed to reflect minimum flight requirements between certain airports and contracts. Trivially, the decision variable value must be an integer.

Constraint 1: Supply

For each origin airport o, the total assigned amount cannot exceed the predicted supply:

$$\sum_{d \in D} X_{o,d} = \text{Supply}_o \quad \forall o \in O$$

Constraint 2: Demand

For each destination airport d, the total assigned amount must not exceed the predicted demand. Demand cannot be set as a minimum in this case because total demand exceeds the total supply

$$\sum_{o \in O} X_{o,d} \leq \text{Demand}_d \quad \forall d \in D$$

Constraint 3: Minimum Flight Assignments per Route

For each route (o, d), the number of flights assigned must meet the minimum mix constraint, defined by the 5% of the supply node:

$$X_{o,d} \geq 0.05 \times \text{Supply}_o \quad \forall (o, d)$$

Constraint 4: Integer Assignments

The decision variables must be integers:

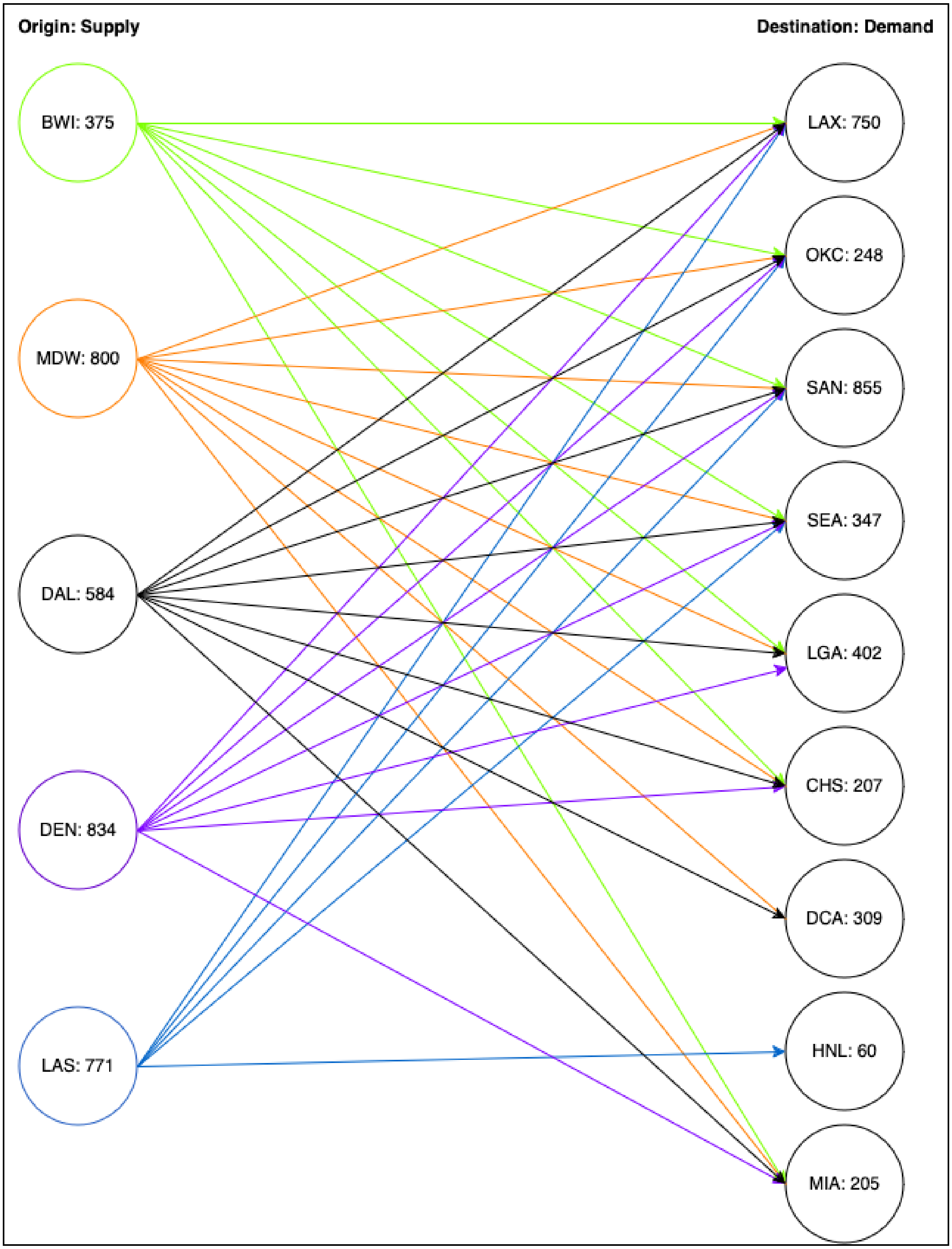
$$X_{o,d} \in \mathbb{Z} \quad \forall (o, d)$$

Mathematical Formulation

$$\text{Minimize:} \quad Z = \sum_{o \in O} \sum_{d \in D} \text{Delay}_{o,d} \cdot X_{o,d}$$

Subject to:

1. $$\sum_{d \in D} X_{o,d} = \text{Supply}_o \quad \forall o \in O$$
2. $$\sum_{o \in O} X_{o,d} \leq \text{Demand}_d \quad \forall d \in D$$
3. $$X_{o,d} \geq 0.05 \times \text{Supply}_o \quad \forall (o,d)$$
4. $$X_{o,d} \in \mathbb{Z} \quad \forall (o,d)$$



Visual representation of the network including the supply amounts, possible routes, and demand amounts.

IATA Code	Predicted Supply
BWI	375
MDW	800
DAL	584
DEN	834
LAS	771
	3364

Predicted Supply Table

IATA Code	Predicted Demand
LAX	750
OKC	248
SAN	855
SEA	347
LGA	402
CHS	207
DCA	309
HNL	60
MIA	205
	3383

Predicted Demand Table

Origin Airport	Destination Airport	2025 Predicted Avg Route Delay in Minutes - (Node Edges)
BWI	CHS	16.01
BWI	LAX	32.58
BWI	LGA	26.64
BWI	MIA	25.15
BWI	OKC	20.17
BWI	SAN	12.9
BWI	SEA	19.75

MDW	CHS	17.71
MDW	DCA	16.69
MDW	LAX	23.56
MDW	LGA	15.33
MDW	MIA	19.41
MDW	OKC	27.19
MDW	SAN	21.68
MDW	SEA	22.74
DAL	CHS	15.92
DAL	DCA	17.59
DAL	LAX	22.98
DAL	LGA	20.68
DAL	MIA	27.04
DAL	OKC	24.75
DAL	SAN	18.84
DAL	SEA	19.76
DEN	CHS	14.9
DEN	LAX	23.05
DEN	LGA	32.35
DEN	MIA	25.13
DEN	OKC	24.25
DEN	SAN	21.75
DEN	SEA	22.48
LAS	HNL	15.47
LAS	LAX	22.2
LAS	OKC	17.39
LAS	SAN	15.96
LAS	SEA	20.54

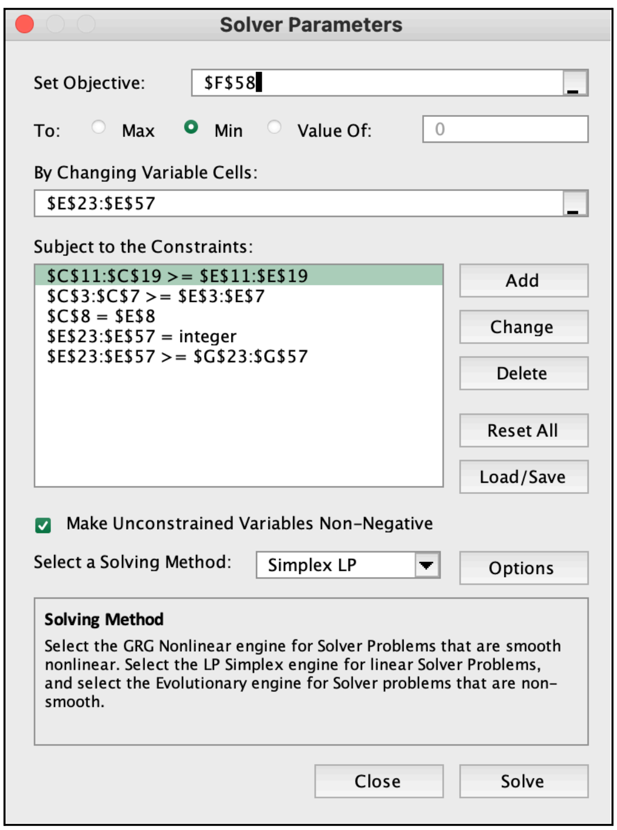
Predicted Delay Table. Routes which are not listed did not have recorded data in the dataset.

Optimal Solution

Color key: **Constraint**, **Decision Variable**, **Objective**

IATA Code	Predicted Supply	Sign	Assigned Amount
BWI	375	>=	375
MDW	800	>=	800
DAL	584	>=	584
DEN	834	>=	834
LAS	771	>=	771
	3364	=	**3364**
IATA Code	**Predicted Demand**	**Sign**	**Assigned Amount**
LAX	750	>=	731
OKC	248	>=	248
SAN	855	>=	855
SEA	347	>=	347
LGA	402	>=	402
CHS	207	>=	207
DCA	309	>=	309
HNL	60	>=	60
MIA	205	>=	205
	3383		**3364**

Origin and Destination Constraints



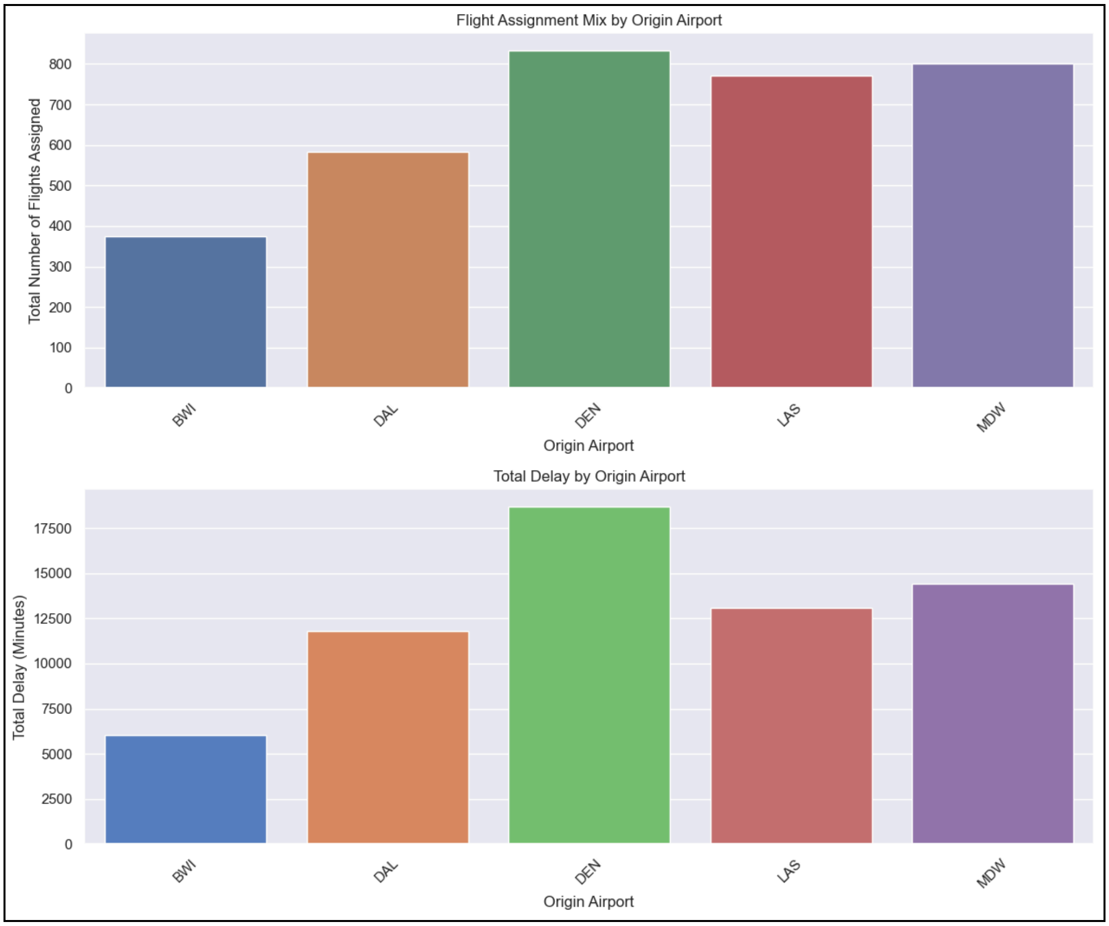
Solver Configuration

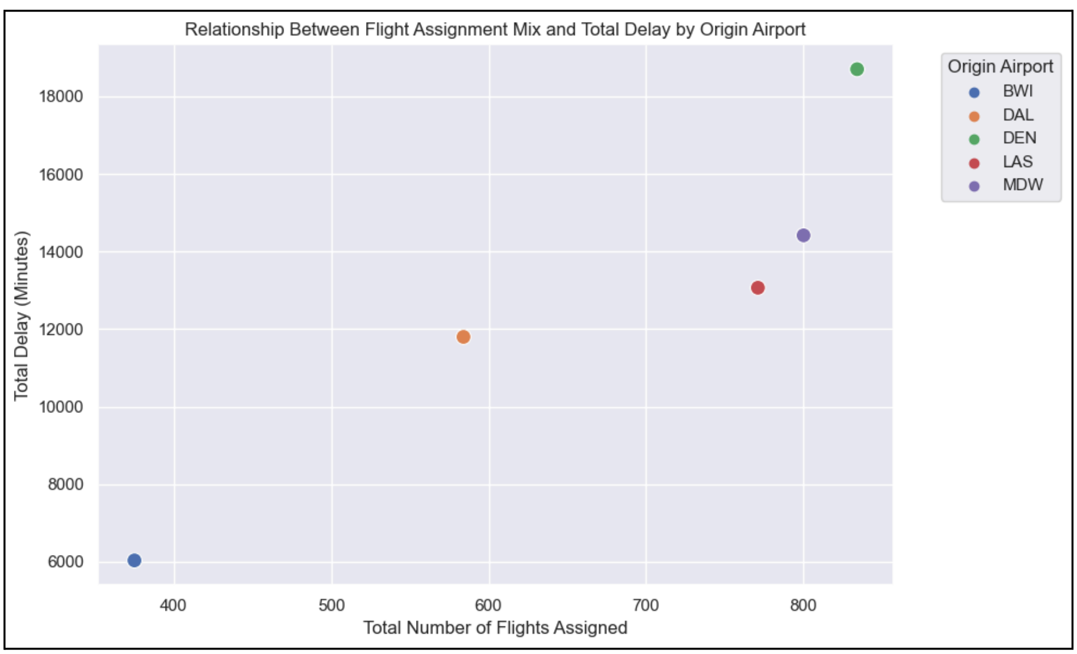
Origin Airport	Destination Airport	2025 Predicted Avg Departure Delay (Minutes)	Decision Variable	Total Delay from Route	Flight Assignment Mix Minimum
BWI	CHS	16.01	19	304.19	19
BWI	LAX	32.58	19	619.02	19
BWI	LGA	26.64	19	506.16	19
BWI	MIA	25.15	19	477.85	19
BWI	OKC	20.17	19	383.23	19
BWI	SAN	12.9	261	3366.9	19
BWI	SEA	19.75	19	375.25	19
MDW	CHS	17.71	40	708.4	40
MDW	DCA	16.69	173	2887.37	40
MDW	LAX	23.56	40	942.4	40
MDW	LGA	15.33	312	4782.96	40
MDW	MIA	19.41	115	2232.15	40
MDW	OKC	27.19	40	1087.6	40
MDW	SAN	21.68	40	867.2	40
MDW	SEA	22.74	40	909.6	40
DAL	CHS	15.92	29	461.68	29
DAL	DCA	17.59	136	2392.24	29
DAL	LAX	22.98	96	2206.08	29
DAL	LGA	20.68	29	599.72	29
DAL	MIA	27.04	29	784.16	29
DAL	OKC	24.75	29	717.75	29
DAL	SAN	18.84	29	546.36	29
DAL	SEA	19.76	207	4090.32	29
DEN	CHS	14.9	119	1773.1	42
DEN	LAX	23.05	505	11640.25	42
DEN	LGA	32.35	42	1358.7	42
DEN	MIA	25.13	42	1055.46	42
DEN	OKC	24.25	42	1018.5	42
DEN	SAN	21.75	42	913.5	42
DEN	SEA	22.48	42	944.16	42
LAS	HNL	15.47	60	928.2	39
LAS	LAX	22.2	71	1576.2	39
LAS	OKC	17.39	118	2052.02	39
LAS	SAN	15.96	483	7708.68	39
LAS	SEA	20.54	39	801.06	39
			Grand Total Delay (obj is to minimize)	**64018.42**	

Optimal Solution

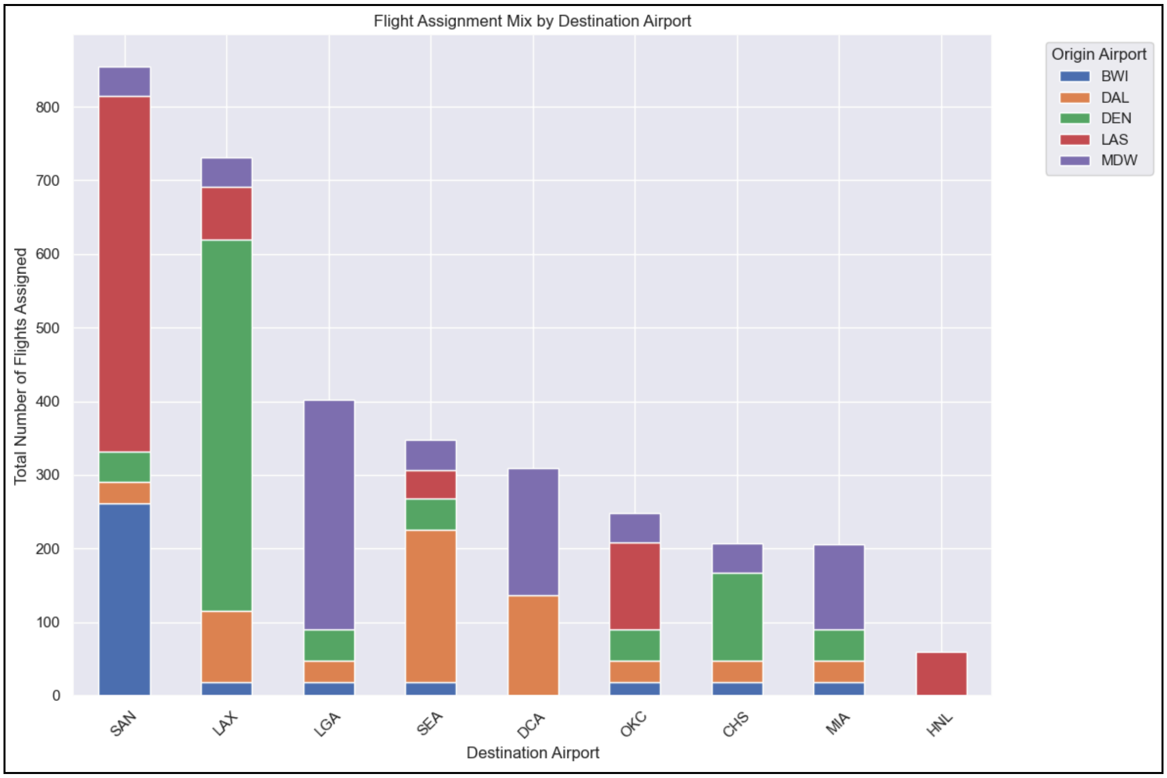
Solution Interpretation

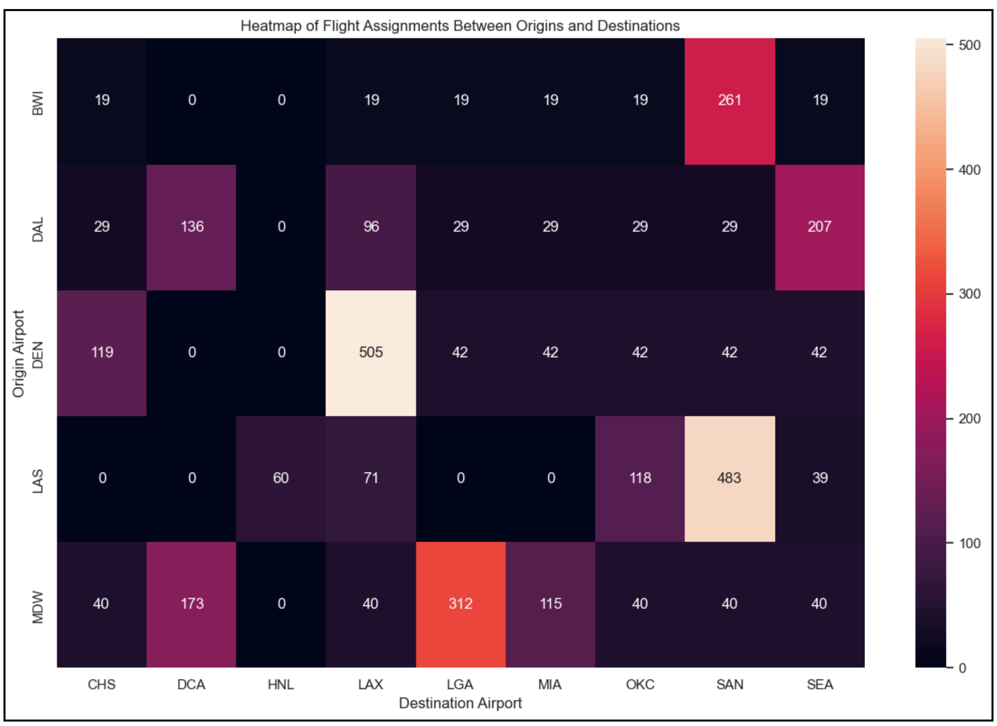
Without loss of generality, we assume that the route delays predicted here will be the observed averages. In practice, delays are subject to large fluctuation and our models only predict the averages which are expected to be observed. The LP model minimized the total delay across all routes to a total of 64018.42 minutes. As expected, the routes with lower delays tended to have more flights assigned and the router with higher delays had less flights assigned, usually with the lower bound of flights being assigned.





Flight assignment mix and total delay by origin airport. We have a strong linear relationship between these two values.





Stacked bar chart and heatmap to show the distribution of flights from origin to destinations. The route values which are 0 are due to airliner range constraints or dataset availability, as those nodes were not connected in the model. Destinations tended to be mostly fulfilled by a single origin, then the remaining assignments were even.

Sensitivity Analysis

Variable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
$E$23	BWI to CHS Decision Variable	19	5.02	16.01	1E+30	5.02
$E$24	BWI to LAX Decision Variable	19	13.44	32.58	1E+30	13.44
$E$25	BWI to LGA Decision Variable	19	14.25	26.64	1E+30	14.25
$E$26	BWI to MIA Decision Variable	19	8.68	25.15	1E+30	8.68
$E$27	BWI to OKC Decision Variable	19	5.84	20.17	1E+30	5.84
$E$28	BWI to SAN Decision Variable	261	0	12.9	3.83	1E+30
$E$29	BWI to SEA Decision Variable	19	3.83	19.75	1E+30	3.83
$E$30	MDW to CHS Decision Variable	40	3.78	17.71	1E+30	3.78
$E$31	MDW to DCA Decision Variable	173	0	16.69	0.97	2.67
$E$32	MDW to LAX Decision Variable	40	1.48	23.56	1E+30	1.48
$E$33	MDW to LGA Decision Variable	312	0	15.33	4.45	1E+30
$E$34	MDW to MIA Decision Variable	115	0	19.41	2.67	1E+30
$E$35	MDW to OKC Decision Variable	40	9.92	27.19	1E+30	9.92
$E$36	MDW to SAN Decision Variable	40	5.84	21.68	1E+30	5.84
$E$37	MDW to SEA Decision Variable	40	3.88	22.74	1E+30	3.88
$E$38	DAL to CHS Decision Variable	29	1.09	15.92	1E+30	1.09
$E$39	DAL to DCA Decision Variable	136	0	17.59	2.67	0.97
$E$40	DAL to LAX Decision Variable	96	0	22.98	0.07	1.56
$E$41	DAL to LGA Decision Variable	29	4.45	20.68	1E+30	4.45
$E$42	DAL to MIA Decision Variable	29	6.73	27.04	1E+30	6.73
$E$43	DAL to OKC Decision Variable	29	6.58	24.75	1E+30	6.58
$E$44	DAL to SAN Decision Variable	29	2.1	18.84	1E+30	2.1
$E$45	DAL to SEA Decision Variable	207	0	19.76	1.56	1E+30
$E$46	DEN to CHS Decision Variable	119	0	14.9	1.09	1E+30
$E$47	DEN to LAX Decision Variable	505	0	23.05	2.65	0.07
$E$48	DEN to LGA Decision Variable	42	16.05	32.35	1E+30	16.05
$E$49	DEN to MIA Decision Variable	42	4.75	25.13	1E+30	4.75
$E$50	DEN to OKC Decision Variable	42	6.01	24.25	1E+30	6.01
$E$51	DEN to SAN Decision Variable	42	4.94	21.75	1E+30	4.94
$E$52	DEN to SEA Decision Variable	42	2.65	22.48	1E+30	2.65
$E$53	LAS to HNL Decision Variable	60	0	15.47	6.73	1E+30
$E$54	LAS to LAX Decision Variable	71	0	22.2	0.85	2.1
$E$55	LAS to OKC Decision Variable	118	0	17.39	4.81	1E+30
$E$56	LAS to SAN Decision Variable	483	0	15.96	2.1	3.83
$E$57	LAS to SEA Decision Variable	39	1.56	20.54	1E+30	1.56

Sensitivity Table for Variable Cells. Color scale on reduced cost: lighter red = lower value, darker red = higher value. Coefficient values for variable cells which are in red must be reduced *by the reduced cost in order for the model to determine that it is beneficial to assign more flights to.*

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
$C$11	LAX Predicted Demand	750	0	0	19	1E+30
$C$12	OKC Predicted Demand	248	4.81	0	19	32
$C$13	SAN Predicted Demand	855	6.24	0	19	32
$C$14	SEA Predicted Demand	347	3.22	0	19	67
$C$15	LGA Predicted Demand	402	6.75	0	19	67
$C$16	CHS Predicted Demand	207	8.15	0	19	463
$C$17	DCA Predicted Demand	309	5.39	0	19	67
$C$18	HNL Predicted Demand	60	6.73	0	19	32
$C$19	MIA Predicted Demand	205	2.67	0	19	67
$C$3	BWI Predicted Supply	375	3.91	0	0	444
$C$4	MDW Predicted Supply	800	0.97	0	0	107
$C$5	DAL Predicted Supply	584	0.07	0	0	463
$C$6	DEN Predicted Supply	834	0	0	0	1E+30
$C$7	LAS Predicted Supply	771	0.85	0	0	463
$C$8	Predicted Supply	3364	-23.05	0	463	0

Sensitivity Table for Constraints. Shadow price is the observed change in objective function value when we change the RHS value of the constraint by 1 unit. This is assuming we are varying by values within the allowable range of the constraint. Non-zero shadow price implies the constraint is binding. Remark: Excel Solver seems to have a limit on the amount of constraints which are reported in this table. The minimum route constraints were not included in this report.

Cell	Name	Cell Value	Formula	Status	Slack
Constraints					
$C$11	LAX Predicted Demand	750	$C$11>=$E$11	Not Binding	19
$C$12	OKC Predicted Demand	248	$C$12>=$E$12	Binding	0
$C$13	SAN Predicted Demand	855	$C$13>=$E$13	Binding	0
$C$14	SEA Predicted Demand	347	$C$14>=$E$14	Binding	0
$C$15	LGA Predicted Demand	402	$C$15>=$E$15	Binding	0
$C$16	CHS Predicted Demand	207	$C$16>=$E$16	Binding	0
$C$17	DCA Predicted Demand	309	$C$17>=$E$17	Binding	0
$C$18	HNL Predicted Demand	60	$C$18>=$E$18	Binding	0
$C$19	MIA Predicted Demand	205	$C$19>=$E$19	Binding	0
$C$3	BWI Predicted Supply	375	$C$3>=$E$3	Binding	0
$C$4	MDW Predicted Supply	800	$C$4>=$E$4	Binding	0
$C$5	DAL Predicted Supply	584	$C$5>=$E$5	Binding	0
$C$6	DEN Predicted Supply	834	$C$6>=$E$6	Binding	0
$C$7	LAS Predicted Supply	771	$C$7>=$E$7	Binding	0
$C$8	Total Predicted Supply	3364	$C$8=$E$8	Binding	0

Answer Report for demand and supply constraints. Almost all of the constraints are binding.

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$23	BWI to CHS Decision Variable	19	$E$23>=$G$23	Binding	0
$E$24	BWI to LAX Decision Variable	19	$E$24>=$G$24	Binding	0
$E$25	BWI to LGA Decision Variable	19	$E$25>=$G$25	Binding	0
$E$26	BWI to MIA Decision Variable	19	$E$26>=$G$26	Binding	0
$E$27	BWI to OKC Decision Variable	19	$E$27>=$G$27	Binding	0
$E$28	BWI to SAN Decision Variable	261	$E$28>=$G$28	Not Binding	242
$E$29	BWI to SEA Decision Variable	19	$E$29>=$G$29	Binding	0
$E$30	MDW to CHS Decision Variable	40	$E$30>=$G$30	Binding	0
$E$31	MDW to DCA Decision Variable	173	$E$31>=$G$31	Not Binding	133
$E$32	MDW to LAX Decision Variable	40	$E$32>=$G$32	Binding	0
$E$33	MDW to LGA Decision Variable	312	$E$33>=$G$33	Not Binding	272
$E$34	MDW to MIA Decision Variable	115	$E$34>=$G$34	Not Binding	75
$E$35	MDW to OKC Decision Variable	40	$E$35>=$G$35	Binding	0
$E$36	MDW to SAN Decision Variable	40	$E$36>=$G$36	Binding	0
$E$37	MDW to SEA Decision Variable	40	$E$37>=$G$37	Binding	0
$E$38	DAL to CHS Decision Variable	29	$E$38>=$G$38	Binding	0
$E$39	DAL to DCA Decision Variable	136	$E$39>=$G$39	Not Binding	107
$E$40	DAL to LAX Decision Variable	96	$E$40>=$G$40	Not Binding	67
$E$41	DAL to LGA Decision Variable	29	$E$41>=$G$41	Binding	0
$E$42	DAL to MIA Decision Variable	29	$E$42>=$G$42	Binding	0
$E$43	DAL to OKC Decision Variable	29	$E$43>=$G$43	Binding	0
$E$44	DAL to SAN Decision Variable	29	$E$44>=$G$44	Binding	0
$E$45	DAL to SEA Decision Variable	207	$E$45>=$G$45	Not Binding	178
$E$46	DEN to CHS Decision Variable	119	$E$46>=$G$46	Not Binding	77
$E$47	DEN to LAX Decision Variable	505	$E$47>=$G$47	Not Binding	463
$E$48	DEN to LGA Decision Variable	42	$E$48>=$G$48	Binding	0
$E$49	DEN to MIA Decision Variable	42	$E$49>=$G$49	Binding	0
$E$50	DEN to OKC Decision Variable	42	$E$50>=$G$50	Binding	0
$E$51	DEN to SAN Decision Variable	42	$E$51>=$G$51	Binding	0
$E$52	DEN to SEA Decision Variable	42	$E$52>=$G$52	Binding	0
$E$53	LAS to HNL Decision Variable	60	$E$53>=$G$53	Not Binding	21
$E$54	LAS to LAX Decision Variable	71	$E$54>=$G$54	Not Binding	32
$E$55	LAS to OKC Decision Variable	118	$E$55>=$G$55	Not Binding	79
$E$56	LAS to SAN Decision Variable	483	$E$56>=$G$56	Not Binding	444
$E$57	LAS to SEA Decision Variable	39	$E$57>=$G$57	Binding	0

Answer Report for route minimum assignment constraints. Of these constraints 22 are binding and 13 are non-binding.

Sensitivity Interpretation

The model tends to have the origin nodes prefer sending the majority of flights to a single destination. The routes in which the minimum number of flights was *not* assigned were non-binding. Routes with Zero Reduced Cost are the routes which are optimal and active in the solution. No improvement in their delay coefficient is required for these routes to remain part of the solution. The routes which are identified as binding are bounded by the minimum flight assignment mix constraint. In reality, the agreed upon number of flights between airports are determined by factors such as route profitability and air service agreements. We should prioritize the binding constraints and increase supply where the shadow price is positive. Any predicted demand nodes can have their capacity increased by 19, and then the optimal solution would change.

Trade-Off Analysis

When minimizing delays through this linear programming model, the majority of flights from each origin airport are allocated to one or two key destinations, while the remaining flights satisfy only the minimum route requirements. This minimizes overall delays in the network. However, this approach results in an unbalanced distribution of flights across the network, where certain routes become heavily saturated and others remain underutilized.

This imbalance can lead to customer dissatisfaction due to limited flight options on less prioritized routes, forcing passengers to seek alternatives from competitors that offer a broader range of flight schedules and destinations. Consequently, travelers may opt for airlines that provide more flexible and extensive flight offerings. Balancing delay minimization with a more equitable flight distribution is essential to maintain customer satisfaction and competitive advantage.

Limitations

While this model provides valuable insights into flight assignments and delay minimization for Southwest Airlines Co., several limitations must be acknowledged to contextualize the findings and guide future improvements.

- The dataset likely does not contain all flights the airliner has sent.
    - **Data Availability:** The model assumes that the dataset encompasses all relevant flights operated by Southwest Airlines Co. However, it is likely that not all flights are captured within the provided dataset. Missing data can lead to underestimation or overestimation of flight volumes and associated delays on certain routes.
    - **Missing Routes:** For instance, the dataset lacks information on flights from Las Vegas (LAS) to LaGuardia Airport (LGA), despite this being an active route for Southwest Airlines Co. The absence of such data points can skew the model's accuracy and reliability for those specific routes.

- **Static Delay Estimates:** The model relies solely on historical data to predict average departure delays. However, delays are influenced by a multitude of dynamic factors such as weather conditions, Notices to Airmen (NOTAMs), airport construction activities, passenger-related issues, and operational disruptions. These factors introduce variability and unpredictability that historical averages may not fully capture.
- **Uniform Minimum Mix Requirement:** The model assumes a uniform minimum flight assignment mix (e.g., 5% of supply) across all routes. In reality, demand variability, route profitability, and strategic priorities necessitate differentiated minimum flight assignments for different routes to optimize operational efficiency and customer satisfaction.

Future Directions

By leveraging the findings through the sensitivity analysis and considering the trade-off evaluations, airlines can identify key points where strategic adjustments yield substantial improvements in delay reduction. The results give a tangible list of routes and nodes to target and increase volume for. By prioritizing specifically the binding constraints, Southwest Airlines Co. can strategically schedule crew, plan maintenance, and allocate resources to drive efficient operations.

Building on the findings of this analysis, many improvements can be taken on the model to further enhance flight assignment strategies and minimize delays. Firstly, integrating real-time data such as weather conditions, air traffic, and operational disruptions can significantly improve the accuracy of delay predictions and enable dynamic flight reassignments. Additionally, expanding the dataset to include all operational routes and historical flight volumes will provide a more comprehensive foundation for optimization models, ensuring that no critical routes are overlooked. Implementing advanced optimization techniques, such as simulation or more advanced machine learning-based forecasting, can also address the inherent uncertainties in flight operations, leading to more resilient and adaptive scheduling solutions. Furthermore, establishing a balanced flight distribution by incorporating constraints that prevent over-concentration on specific routes will help mitigate customer dissatisfaction and enhance overall service reliability.

With a limited number of flights any airliner can operate in a given year and time period, prescriptive analytics and operations research play a pivotal role in enhancing decision-making processes. For example, by adopting a goal programming approach, Southwest Airlines Co. can establish an entirely new framework for strategic decision-making, providing another comprehensive reference point for optimizing flight assignments and minimizing delays.