# Analysis of U.S. Domestic Passenger Traffic Trends

**ECO 5740 - Forecasting and Time Series Models**

Namita Lamichhane

7ec 7, 2025

## Table of contents


**1 Abstract** **3**

**2 Introduction** **3**

**3 Data Description and Sources** **4**

**4 Time Series Graphics** **5**
- 4.1 Set up data as a time series object . . . . . 5
- 4.2 Plot the data . . . . . 5
- 4.3 Time series patterns . . . . . 6

**5 Transformations and Adjustments** **9**
- 5.1 Calendar Adjustment . . . . . 9
- 5.2 Population Adjustment . . . . . 11
- 5.3 Inflation Adjustment . . . . . 11
- 5.4 Mathematical Transformations . . . . . 11

**6 Time Series Decomposition** **13**

**7 Forecasting Methods** **15**
- 7.1 Selecting a Reliable Forecasting Model . . . . . 15
  - 7.1.1 Training set . . . . . 15
  - 7.1.2 Fit the model . . . . . 15
  - 7.1.3 Compare the forecast accuracy across models . . . . . 16
  - 7.1.4 Test the residuals of your preferred method . . . . . 17
- 7.2 Forecasting . . . . . 19

**8 Results** **24**

**9 Discussion and Conclusion** **25**

# 1 Abstract

This study analyzes monthly U.S. domestic airline passenger data from January 2010 to May 2025 to forecast future trends using time series models. The dataset exhibits a strong upward trend and pronounced annual seasonality, with passenger volumes peaking in summer months and dipping in early fall. The 2020 COVID-19 pandemic caused a sharp decline in passenger numbers, followed by a rapid recovery. Various forecasting models, including naïve, ETS, ARIMA, and Theta methods, were evaluated for predictive accuracy. The Holt exponential smoothing model was identified as the most accurate, achieving the lowest RMSE, MAE, and MAPE, and effectively capturing both trend and seasonal patterns. Forecasts for July 2025 through December 2026 indicate continued growth in passenger numbers, with predictable seasonal fluctuations: peaks in summer and recovery following early-fall declines. Minor residual correlations and structural shocks are acknowledged as limitations. These forecasts provide actionable insights for key stakeholders, including airlines, airports, and policymakers, allowing them to optimize operations, manage resources efficiently, and plan strategically for seasonal demand. Overall, the study demonstrates that the Holt model is a reliable tool for short-term operational and strategic planning, translating historical trends into practical, data-driven decisions in the aviation sector.

# 2 Introduction

Air travel plays a critical role in the U.S. economy, connecting people for business, leisure, and essential services. Understanding and predicting domestic airline passenger demand is essential for airlines, airports, policymakers, and industry analysts. Accurate forecasts allow stakeholders to optimize flight schedules, manage staffing and aircraft resources efficiently, and plan strategically for seasonal and long-term growth. Misjudging demand can lead to overbooking, underutilized capacity, or missed revenue opportunities, making reliable forecasting crucial for operational and financial decision-making.

This report analyzes monthly U.S. domestic airline passenger boardings from January 2010 to May 2025. The dataset, obtained from the Federal Reserve Economic Data (FRED) database, provides 185 monthly observations and includes both trend and seasonal information. It captures key events affecting air travel, such as economic cycles and the COVID-19 pandemic, making it ideal for forecasting purposes. Monthly data allows for detailed analysis of seasonal fluctuations, holiday peaks, and off-peak troughs, which are important for operational planning.

Forecasting airline passenger demand is both interesting and challenging. Historical trends show a steady upward trajectory in passenger numbers, yet irregular shocks, such as the 2020 pandemic, create sudden disruptions. Seasonal peaks, particularly in summer, and recurring troughs in early fall add further complexity. Prior studies emphasize the economic and operational importance of anticipating air travel patterns, especially in post-pandemic recovery periods, highlighting the need for accurate forecasting.

Based on the Holt model forecasts, U.S. domestic airline passenger numbers are expected to continue rising over the next 24 months, following a predictable seasonal pattern. This growth is significant for stakeholders: airlines can optimize flight capacity and pricing, airports can plan staffing and gate operations, and policymakers can ensure smooth transportation infrastructure management. These forecasts provide actionable insights that help stakeholders reduce risks, capitalize on expected demand, and make informed operational and strategic decisions.

# 3 Data Description and Sources

1. **Brief Introduction to the Data**

The dataset represents monthly U.S. domestic airline passenger enplanements (number of passengers boarding flights) from January 2010 to May 2025.This data is highly relevant for forecasting because it allows analysis of trends, seasonality, and short-term fluctuations in domestic air travel. Understanding these patterns is crucial for airline operations, airport management, and policy planning. Accurate forecasts help stakeholders anticipate demand, allocate resources efficiently, and plan for future capacity.

2. **Data Source(s)**

The data was obtained from the Federal Reserve Economic Data (FRED) database maintained by the Federal Reserve Bank of St. Louis.

Dataset Name: Enplanements for U.S. Air Carrier Domestic, Scheduled Passenger Flights (ENPLANED)

URL: https://fred.stlouisfed.org/series/ENPLANED

3. **Data Frequency**

The dataset is monthly, which provides sufficient granularity to capture annual seasonal patterns, short-term fluctuations, and gradual changes in passenger behavior. Monthly data allows us to analyze holiday peaks, off-season troughs, and other recurring patterns while remaining manageable for forecasting.

4. **Time Coverage**

The dataset spans January 2010 – May 2025, providing 185 observations. This period is long enough to support meaningful forecasting while remaining relevant to current airline operations. The dataset includes periods affected by significant events, such as the COVID-19 pandemic, allowing for the analysis of shocks and recovery in passenger trends.

5. **Variable Definitions**

ENPLANED: Total number of U.S. domestic airline passengers boarding flights each month (in thousands).

observation_date: Month and year of each observation, serving as the time index for the series.

The data is not seasonally adjusted, which makes it suitable for studying both trend and seasonal patterns directly.

# 4 Time Series Graphics

## 4.1 Set up data as a time series object

```
# Bring my csv file from download to R and call it mydata
mydata <- read_csv("C:/Users/Dell/Downloads/ENPLANED.csv",
                   show_col_types=FALSE)


# set up data as a time series object
mydata <- mydata |>
  mutate(observation_date = yearmonth(observation_date)) |>
  as_tsibble(index = observation_date)
```

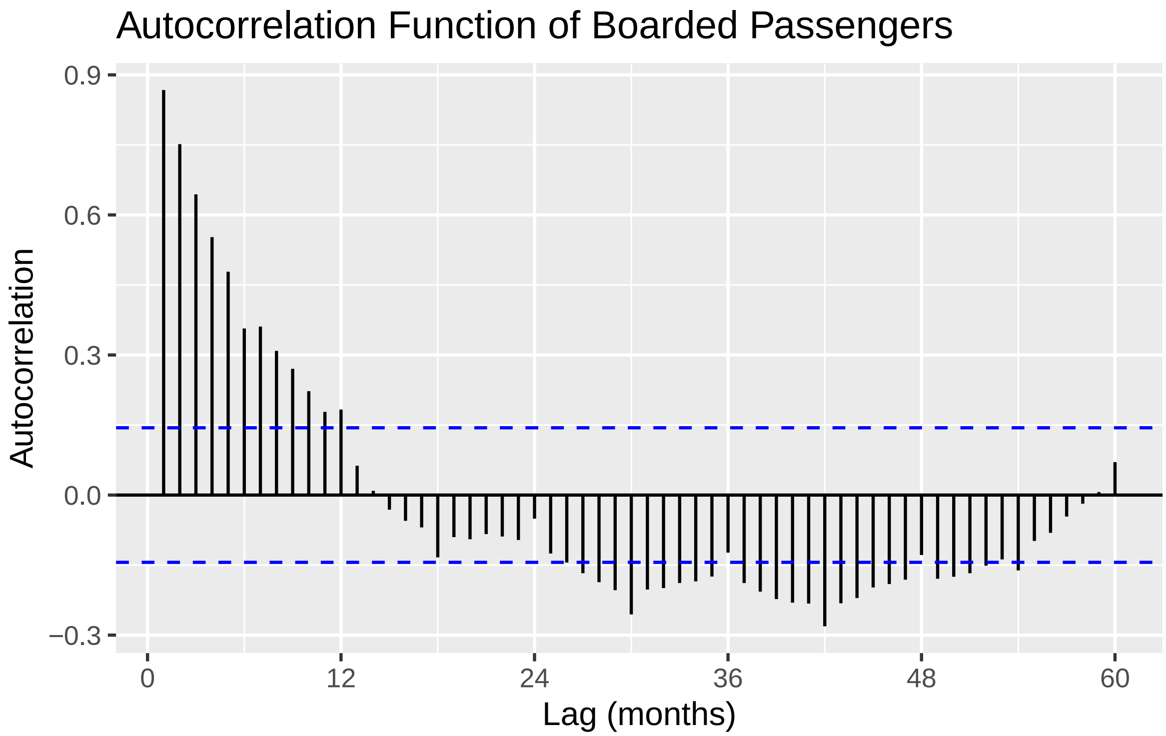
## 4.2 Plot the data

```
mydata |>
  autoplot(ENPLANED) +
  ggtitle("Monthly U.S. Airline Passenger Boardings") +
  xlab("Year") +
  ylab("Number of Passengers (in millions)")+ scale_x_yearmonth(date_breaks =
  "2 year", date_labels = "%Y", limits =as.Date(c("2010-01-01 Jan",
  "2025-05-01")), expand = c(0,0)) +
   scale_y_continuous(labels = scales::label_number(scale = 1e-3))
```


This is a time series plot showing the monthly number of boarded passengers in the U.S. from 2010 to 2025. The upward trend in passenger numbers over time, indicates growth in air travel. A clear seasonal pattern is also evident, with recurring peaks and troughs within each year. The most striking feature is the dramatic and sudden drop in passenger numbers around early 2020, which is an unusual observation and a clear indicator of the impact of the COVID-19 pandemic. The subsequent recovery shows a steep rebound, but passenger numbers do not fully return to the pre-2020 trend until later.

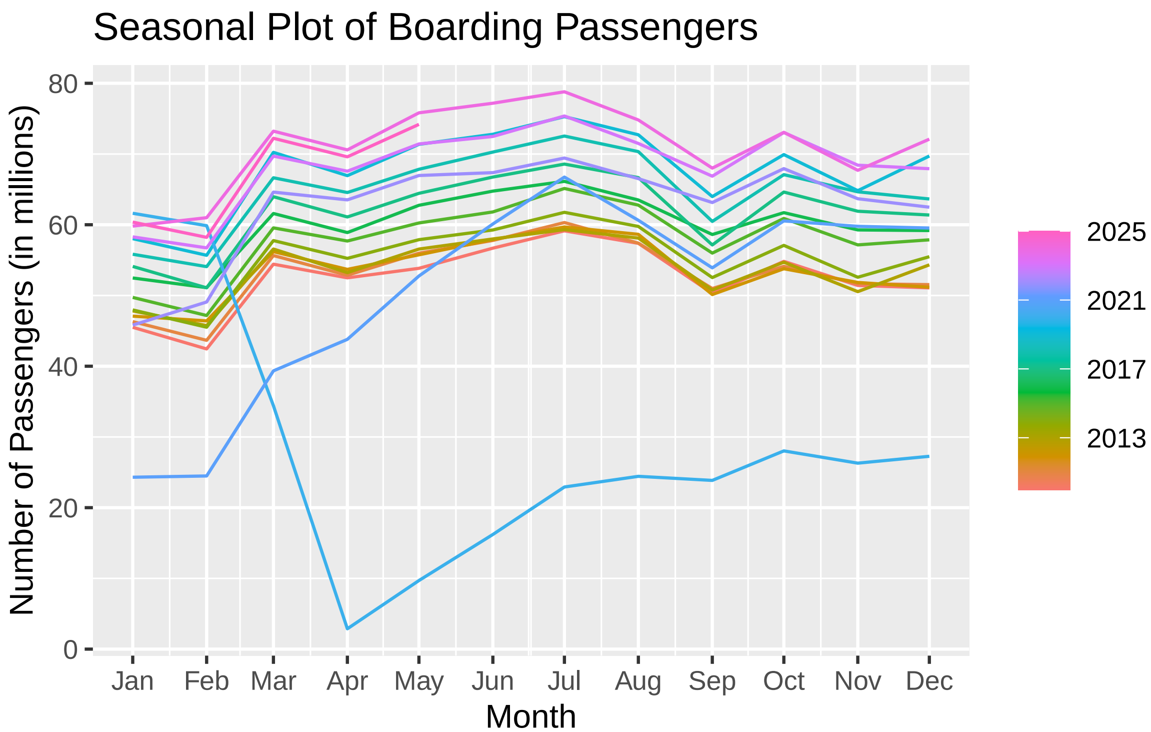
### 4.3 Time series patterns

```r
mydata |>
  ACF(ENPLANED, lag_max = 60) |>
  autoplot() +
  labs(title = "Autocorrelation Function of Boarded Passengers",
       x = "Lag (months)",
       y = "Autocorrelation") +
  # Shows labels every 12 lags for clarity
  scale_x_continuous(breaks = scales::breaks_width(12))
```



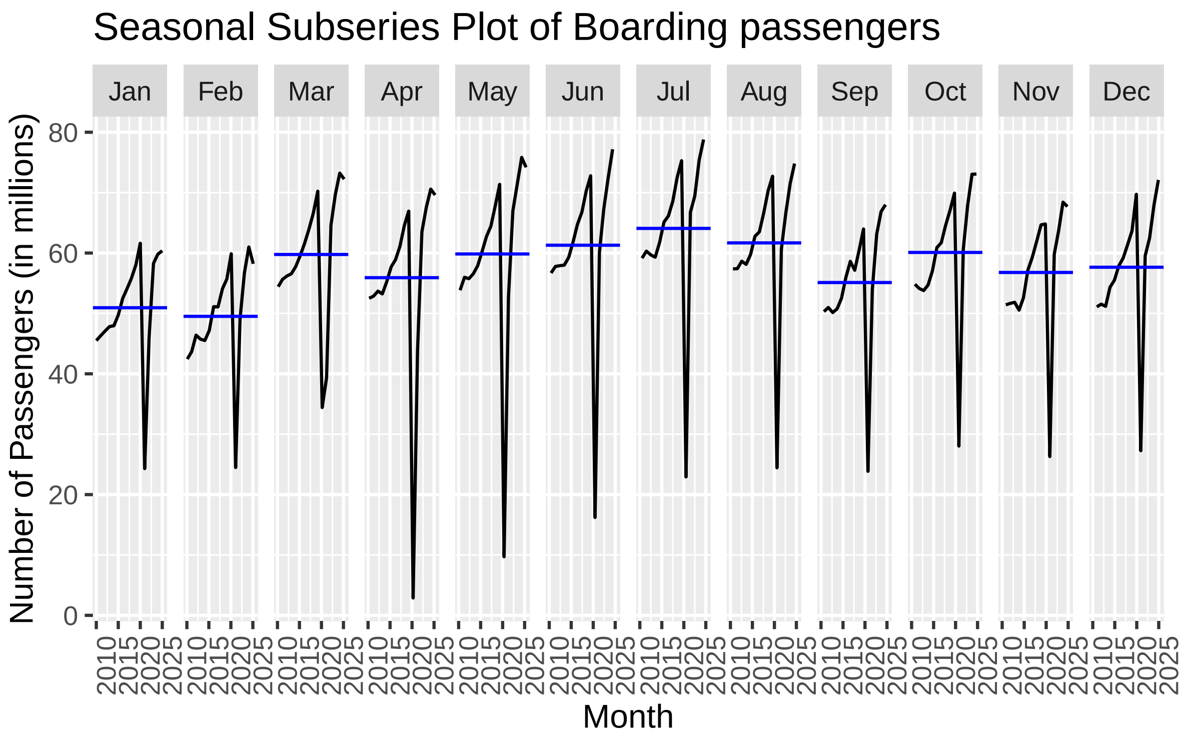

The ACF plot measures the correlation between the time series and its lagged versions. The tall spike at lag 1 indicates a very strong positive correlation between a month's passenger numbers and the previous month's. The plot also shows a strong positive correlation at lag 12, and subsequent lags at multiples of 12 (e.g., 24, 36, 48, 60), which confirms the strong annual seasonality of the data. The slow decline of the ACF shows a strong positive trend in the data. The overall pattern of a slowly decaying, wavy ACF with spikes at seasonal lags is a typical signature of non-stationary time series with both trend and seasonality.

```r
p <- mydata %>%
  gg_season(ENPLANED) +
  labs(title = "Seasonal Plot of Boarding Passengers",
       y = "Number of Passengers (in millions)",
       x = "Month") +
    scale_y_continuous(labels = scales::label_number(scale = 1e-3))

print(p)
```




This plot shows the passenger data for each year, highlighting the seasonal pattern. Each line represents a single year, with the month on the x-axis and passenger numbers on the y-axis. The lines generally follow a similar pattern: a peak in the summer months (July/August), a smaller peak in late fall (October/November), and troughs in the winter and early spring. The plot makes it easy to compare the seasonal patterns across different years. The most notable observation is the line for 2020, which deviates significantly from the others, showing a massive decline in passenger numbers in April and a much lower level throughout the rest of the year compared to other years. This confirms the impact of the pandemic seen in the time series plot.

```
mydata %>%
  gg_subseries(ENPLANED) +
  labs(title = "Seasonal Subseries Plot of Boarding passengers",
       y = "Number of Passengers (in millions)",
       x = "Month") +
    scale_y_continuous(labels = scales::label_number(scale = 1e-3))
```




The plot confirms a strong seasonal pattern. The average passenger count (blue line) is noticeably higher during the summer months (e.g., June, July, August) compared to the shoulder or off-peak months (e.g., January, February, September, November), indicating that more people travel during this summer. Despite the major disruption around 2020, the passenger counts for most months generally show an upward trend from 2010 to 2025, suggesting overall growth in boarding passengers.The recovery in the years following (2021-2025) appears to be rapid, with passenger counts returning to or exceeding pre-2020 levels in many months.

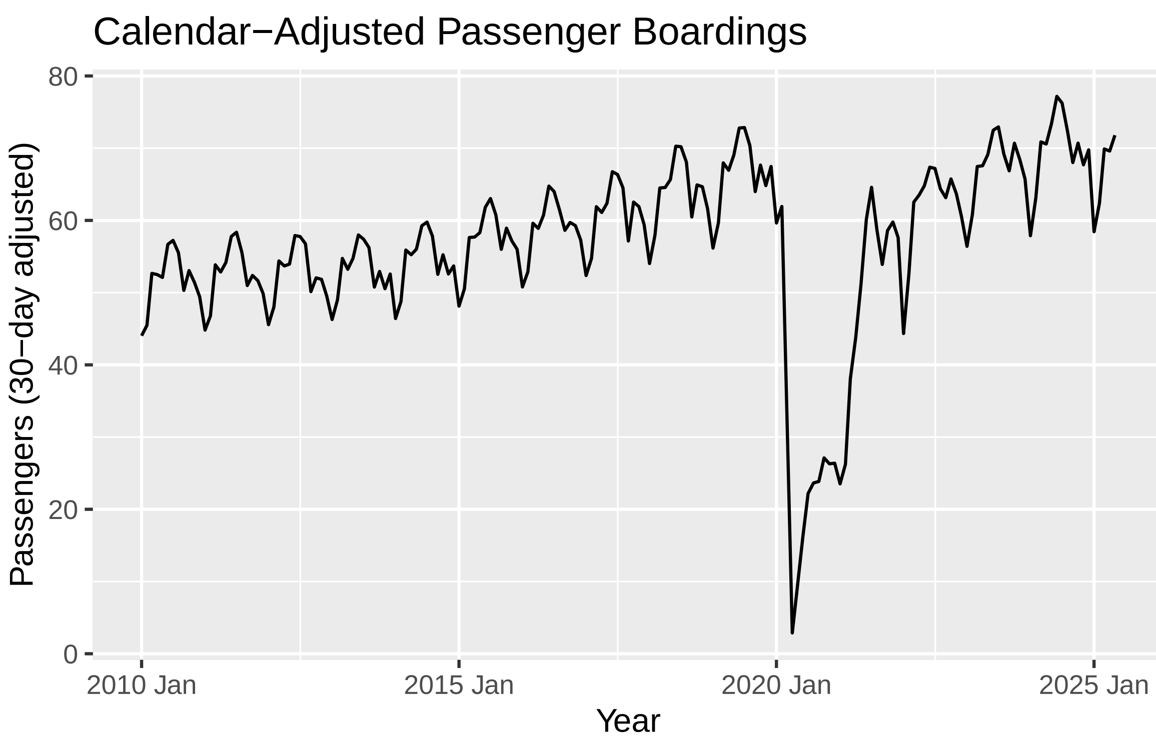
## 5 Transformations and Adjustments

### 5.1 Calendar Adjustment

```
mydata <- mydata %>%
  mutate(days_in_month = lubridate::days_in_month(as.Date(observation_date)),
         ENPLANED_adj = ENPLANED / days_in_month * 30)
```

A calendar adjustment was applied to account for the unequal number of days across months (e.g., February versus March). Since the data is recorded monthly, dividing by the number of days in each month ensures that differences in passenger volumes are not simply due to varying month lengths. This step improves comparability across months and makes seasonal patterns more accurate.

```
mydata %>%
  autoplot(ENPLANED_adj) +
  labs(title = "Calendar-Adjusted Passenger Boardings",
       y = "Passengers (30-day adjusted)", x = "Year") +
    scale_y_continuous(labels = scales::label_number(scale = 1e-3))
```




This plot shows the true volume of passenger travel each month, with the effect of different month lengths already taken out. From 2010 to 2020, travel numbers were steadily climbing, with big, regular yearly swings (seasonality). A huge crash hit in early 2020 because of the pandemic. Since then, the numbers have made a very strong comeback and are now higher than they were before the crash, setting new records near 80,000. It is important to note that the size of those yearly swings gets bigger as the total number of passengers goes up.

### 5.2 Population Adjustment

No population adjustment was applied. The dataset records the total number of airline passengers, and the focus of this analysis is on forecasting overall domestic air travel demand rather than per-capita figures. Changes in U.S. population over the period are relatively gradual and do not significantly affect the trends observed in passenger volumes. Therefore, expressing the data on a per-capita basis was not necessary for the objectives of this study.

### 5.3 Inflation Adjustment

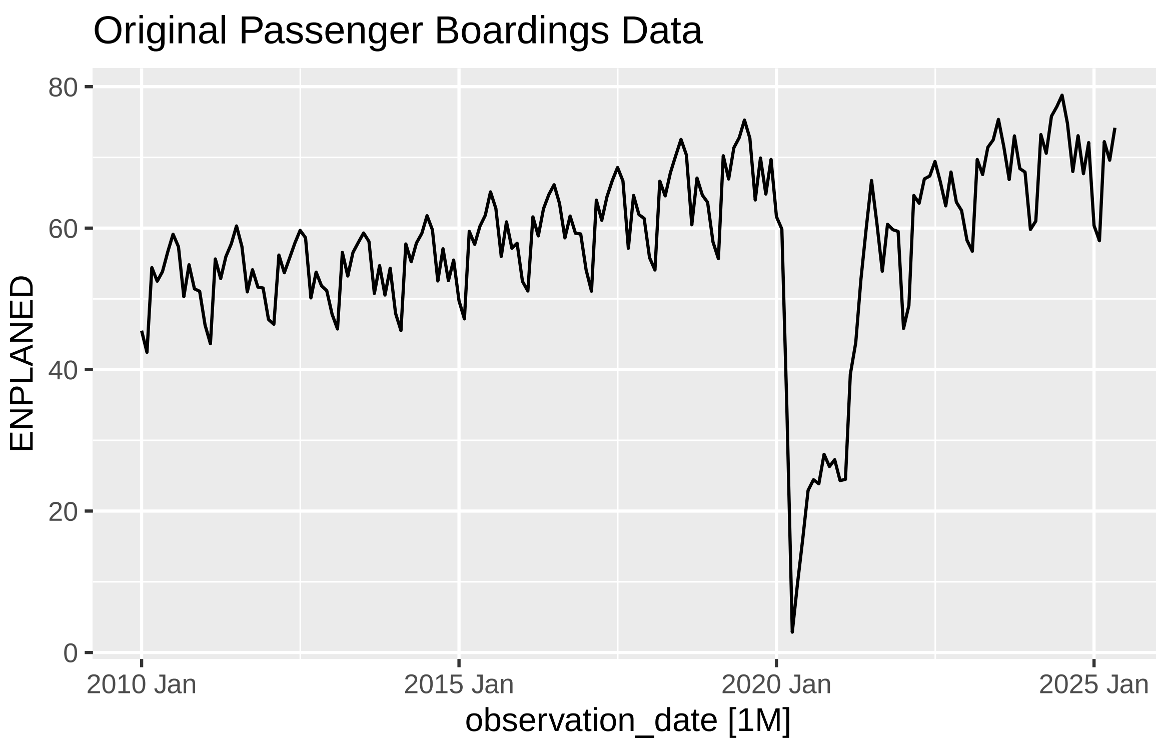
No population adjustment was applied. The dataset records the total number of airline passengers, and the focus of this analysis is on forecasting overall domestic air travel demand rather than per-capita figures. Changes in U.S. population over the period are relatively gradual and do not significantly affect the trends observed in passenger volumes. Therefore, expressing the data on a per-capita basis was not necessary for the objectives of this study.

### 5.4 Mathematical Transformations

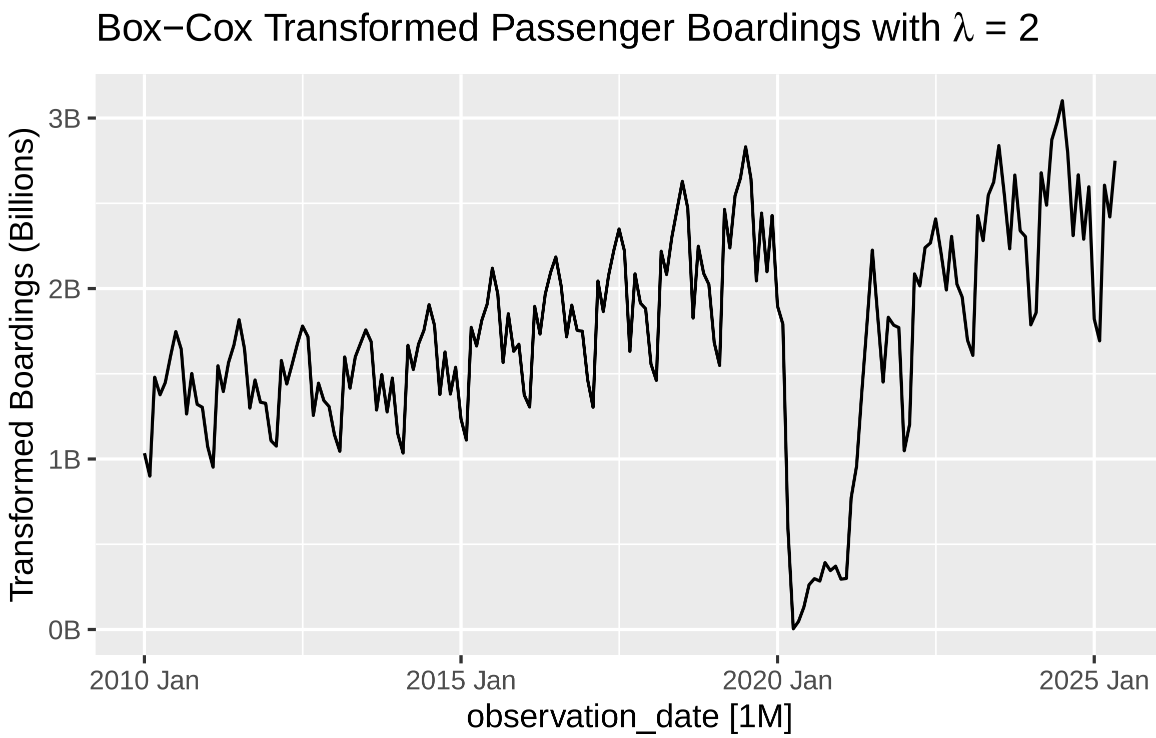
A Box-Cox transformation was applied to the calendar-adjusted passenger data to even out the differences in variance over time. The Guerrero method was used to automatically choose the best (lambda) that reduces variability. This transformation makes seasonal patterns easier to compare and focuses on percentage changes rather than absolute numbers, making the data easier to understand and better for decomposition and forecasting.

```r
# Calculating optimal lambda using Guerrero method
lambda <- mydata %>%
  features(ENPLANED_adj, features = guerrero) %>%
  pull(lambda_guerrero)

lambda
#> [1] 1.999919
```

```r
#Plot the Original Data
mydata |>
  autoplot(ENPLANED) +
  labs(title = "Original Passenger Boardings Data") +
    scale_y_continuous(labels = scales::label_number(scale = 1e-3))
```




```
# Plot the Box-Cox transformed series
library(scales)

mydata |>
  autoplot(box_cox(ENPLANED, lambda)) +
    scale_y_continuous(labels = scales::label_number(scale = 1e-9, suffix =
  "B")) +
  labs(y = "Transformed Boardings (Billions)",
       title = latex2exp::TeX(paste0(
         "Box-Cox Transformed Passenger Boardings with $\\lambda$ = ",
         round(lambda, 2))))
```

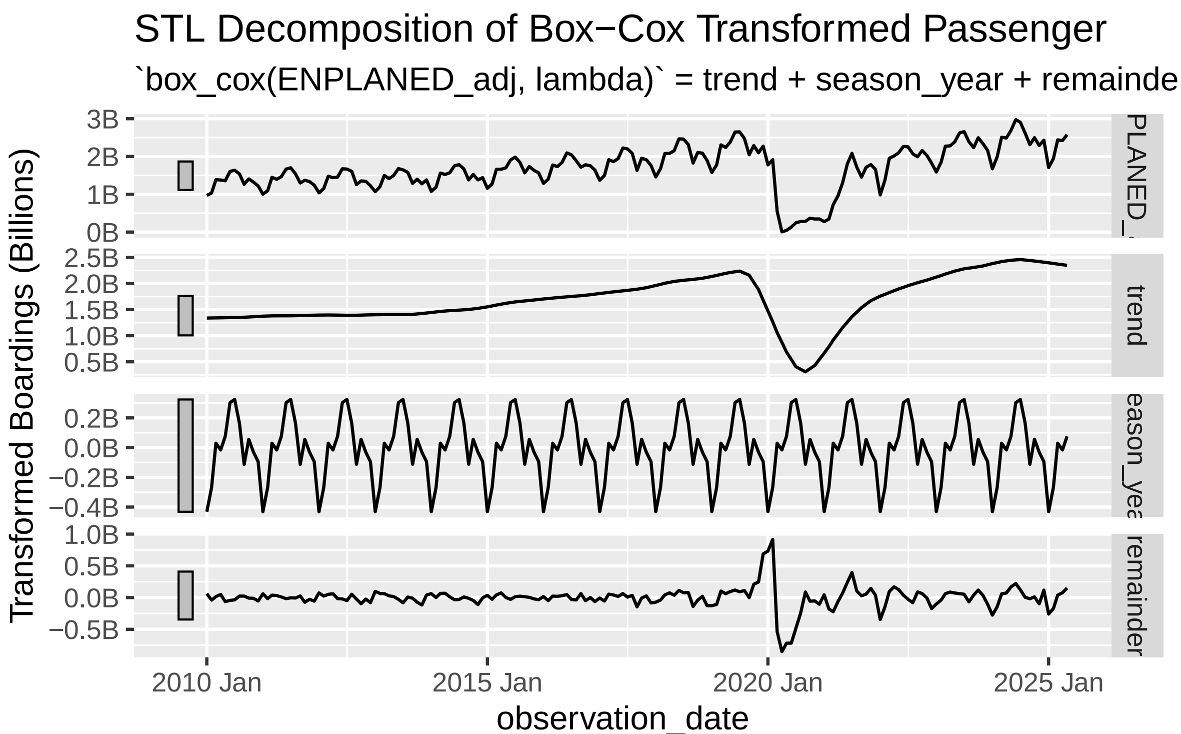



This plot displays the passenger boarding data after a Box-Cox transformation was applied to stabilize the variance, a necessary step for accurate forecasting. The Guerrero method calculated the optimal  to be 1.999919, which is effectively =2. A Box-Cox =2 corresponds to a quadratic transformation (y2), which stabilized the variance across the entire series. The Y-axis is scaled to billions of transformed units not raw passengers (e.g., 3B represents 3 billion). The series clearly shows three distinct phases: a decade of steady growth with strong annual seasonality leading into 2020 ; a massive, sudden collapse in 2020 (due to the pandemic) ; and a robust recovery that pushed transformed boarding levels past previous peaks by 2024. The transformation was successful because it ensured the size of the seasonal fluctuations is now consistent across the entire series, correcting the increasing variability seen in the original, untransformed data.

## 6 Time Series Decomposition

```
# Fit the STL model
fit <- mydata %>%
  model(stl = STL(box_cox(ENPLANED_adj, lambda) ~ season(window = "periodic")))
```

```r
# Plot the components
components(fit) %>%
  autoplot() +
    scale_y_continuous(labels = scales::label_number(scale = 1e-9, suffix
  = "B")) +
  labs(title = "STL Decomposition of Box-Cox Transformed Passenger",
       y = "Transformed Boardings (Billions)")
```




This STL decomposition cleanly separates the underlying structure of the passenger data. It shows a stable seasonal cycle (yearly travel pattern) laid on top of a gradually rising trend that was fundamentally interrupted by the deep, smooth U-shaped dip of the 2020 pandemic. The extreme spike in the remainder component in 2020 highlights that the pandemic shock was a sudden, non-recurring event that cannot be explained by the usual trend or seasonal forces.

# 7 Forecasting Methods

## 7.1 Selecting a Reliable Forecasting Model

### 7.1.1 Training set

The dataset contains 185 monthly observations from January 2010 to May 2025. The first 148 observations (first 80%) were used as the training set to fit the models, while the remaining 37 observations (approximately 20%) formed the test set to evaluate forecast accuracy. This split ensures that the models are trained on a sufficient number of data points to capture trends and seasonality, while the test set provides an unbiased assessment of out-of-sample forecasting performance.

```
# extract the training set to fit the data
training_set <- mydata |>
  filter_index("2010 Jan" ~ "2021 Dec")

# extract the test set
test_set <- mydata |>
  filter_index("2022 Jan" ~ .)

# set the forecast horizon equals to the test set
h <- nrow(test_set)
```

### 7.1.2 Fit the model

Next, fit the forecasting methods.

```
# Fit the models
my_fit <- training_set |>
  model(
    Mean = MEAN(box_cox(ENPLANED_adj, lambda)),
    `Naïve` = NAIVE(box_cox(ENPLANED_adj, lambda)),
    Drift = NAIVE(box_cox(ENPLANED_adj, lambda) ~ drift()),
    `Seasonal naïve` = SNAIVE(box_cox(ENPLANED_adj, lambda)),

    #Exponential smoothing
    Holt = ETS(box_cox(ENPLANED_adj, lambda) ~ error("A")
   + trend("A") + season("A")),
    Holt_damped = ETS(box_cox(ENPLANED_adj, lambda) ~
  error("A") + trend("Ad") + season("A")),
```

```
    ETS_auto = ETS(box_cox(ENPLANED_adj, lambda)),

    #ARIMA
    ARIMA_auto = ARIMA(box_cox(ENPLANED_adj, lambda), stepwise = FALSE, approx = FALSE),

    #Theta methods
  Theta_M = THETA(box_cox(ENPLANED_adj, lambda) ~
season(method = "multiplicative")),
    Theta_A = THETA(box_cox(ENPLANED_adj, lambda) ~ season(method = "additive"))
  ) |>
mutate(
    comb1 = (Holt + Holt_damped)/2,
    comb2 = (Holt + ETS_auto)/2,
    comb3 = (Holt_damped + ETS_auto)/2,
    comb4 = (Holt  + Holt_damped + ETS_auto)/3
)
```

To evaluate whether combining forecasts improves accuracy, four ensemble combinations were created using the top three performing models: Holt, Holt_damped, and ETS_auto. Each ensemble is the simple average of two or three models: comb1 averages Holt and Holt_damped, comb2 averages Holt and ETS_auto, comb3 averages Holt_damped and ETS_auto, and comb4 averages all three models.

### 7.1.3 Compare the forecast accuracy across models

Next, forecast the last 20% of the data and test the forecast accuracy.

```
# Forecast the test set (h periods ahead).
my_fc <- my_fit |>
  forecast(h = h)

# Test the accuracy of the forecasts based on the test set.
accuracy(my_fc, mydata)
#> # A tibble: 14 x 10
#>    .model          .type     ME    RMSE    MAE   MPE  MAPE  MASE RMSSE  ACF1
#>    <chr>           <chr>  <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#>  1 ARIMA_auto      Test  36247. 38988. 36565. 53.6  54.4  4.47  2.28  0.755
#>  2 Drift           Test  10351. 12387. 11245. 14.7  16.6  1.38  0.723 0.633
#>  3 ETS_auto        Test   9458. 10709.  9951. 13.7  14.8  1.22  0.625 0.745
#>  4 Holt            Test   5839.  6860.  6370.  8.34  9.49 0.779 0.401 0.604
#>  5 Holt_damped     Test  10023. 11587. 10559. 14.5  15.7  1.29  0.677 0.792
```

```
#>  6 Mean           Test  11721. 13281. 12316. 16.8  18.1  1.51  0.776 0.596
#>  7 Naïve          Test  11857. 13939. 12728. 16.9  18.8  1.56  0.814 0.662
#>  8 Seasonal naïve Test  30267. 40410. 30267. 48.0  48.0  3.70  2.36  0.543
#>  9 Theta_A        Test  11754. 13829. 12626. 16.7  18.6  1.54  0.808 0.660
#> 10 Theta_M        Test  11754. 13829. 12626. 16.7  18.6  1.54  0.808 0.660
#> 11 comb1          Test   7931.  9152.  8465. 11.4  12.6  1.04  0.534 0.716
#> 12 comb2          Test   7648.  8744.  8160. 11.0  12.1  0.998 0.511 0.683
#> 13 comb3          Test   9740. 11141. 10255. 14.1  15.2  1.25  0.651 0.770
#> 14 comb4          Test   8440.  9668.  8960. 12.2  13.3  1.10  0.565 0.727
```

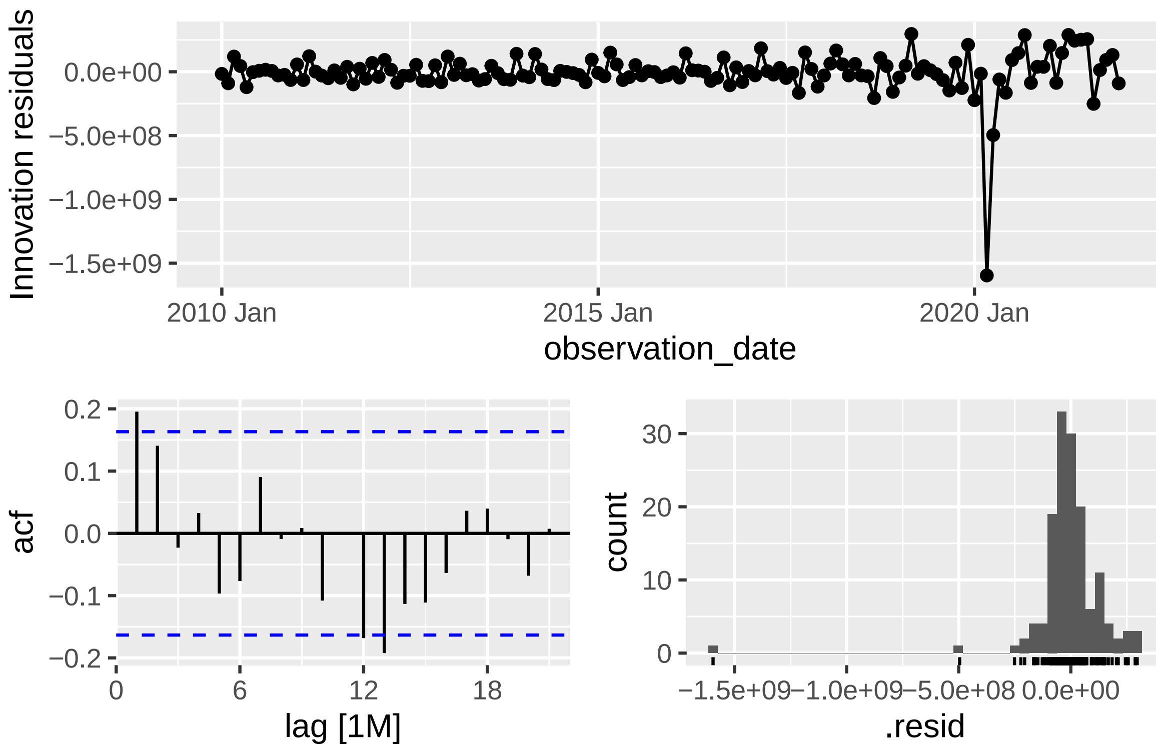
Based on the out-of-sample forecast accuracy measures, the Holt model produced the most accurate forecasts with the lowest RMSE (6859.58), MAE (6369.84), and MAPE (9.49%). This suggests that the data is primarily trend-driven with limited irregular seasonality. The combination models showed mixed results; comb1 and comb2 performed reasonably well, while comb3 and comb4 were less accurate than the individual Holt model.

The ARIMA_auto model performed the worst, indicating that relying on ARIMA's assumptions of regular patterns and stationarity is not suitable for this dataset, especially due to irregular seasonality, sharp trend changes, and structural breaks caused by the COVID-19 pandemic.

These results suggest that the passenger data is driven by both trend and seasonality, and the Holt model is the most suitable for short-term forecasting.

### 7.1.4 Test the residuals of your preferred method

```
# Residual properties
my_fit |>
  select(Holt) |>
  gg_tsresiduals()
```




```
# The Ljung-Box test
my_fit |>
  select(Holt) |>
  augment() |>
  features(.innov, ljung_box, lag = 10)
#> # A tibble: 1 x 3
#>   .model lb_stat lb_pvalue
#>   <chr>    <dbl>     <dbl>
#> 1 Holt      14.2     0.164
```

The residuals from the Holt model show a significantly improved fit. The time plot of the innovation residuals reveals that the clear cyclical pattern previously visible in the errors has been largely eliminated, indicating the Holt model successfully captured the annual seasonality present in the data before 2020. The ACF plot confirms a strong reduction in autocorrelation, but shows Lags 1 and 13 are statistically significant (crossing the confidence bounds), suggesting some minor short-term and secondary lagged correlation remains in the errors. The residual histogram is still slightly negatively skewed due to the extreme negative shock around 2020, meaning the errors are not perfectly normally distributed.

Residual diagnostics for the Holt model show a Ljung–Box test statistic of 14.21 with a p-value of 0.164. The large p-value (greater than 0.05) indicates that we fail to reject the null hypothesis that the residuals are white noise, meaning there is no statistically significant correlation remaining in the errors. This suggests that the Holt model has captured all underlying systematic patterns in the data, including the seasonality, making the model an adequate fit for forecasting the time series.

## 7.2 Forecasting

Now, use the complete data to fit the model

```
# Fit the model of your choice
my_fit <- mydata |> #using the complete data here, not the training set
  model(
    Holt = ETS(box_cox(ENPLANED_adj, lambda) ~ error("A")+ trend("A")+season("A"))
)
```

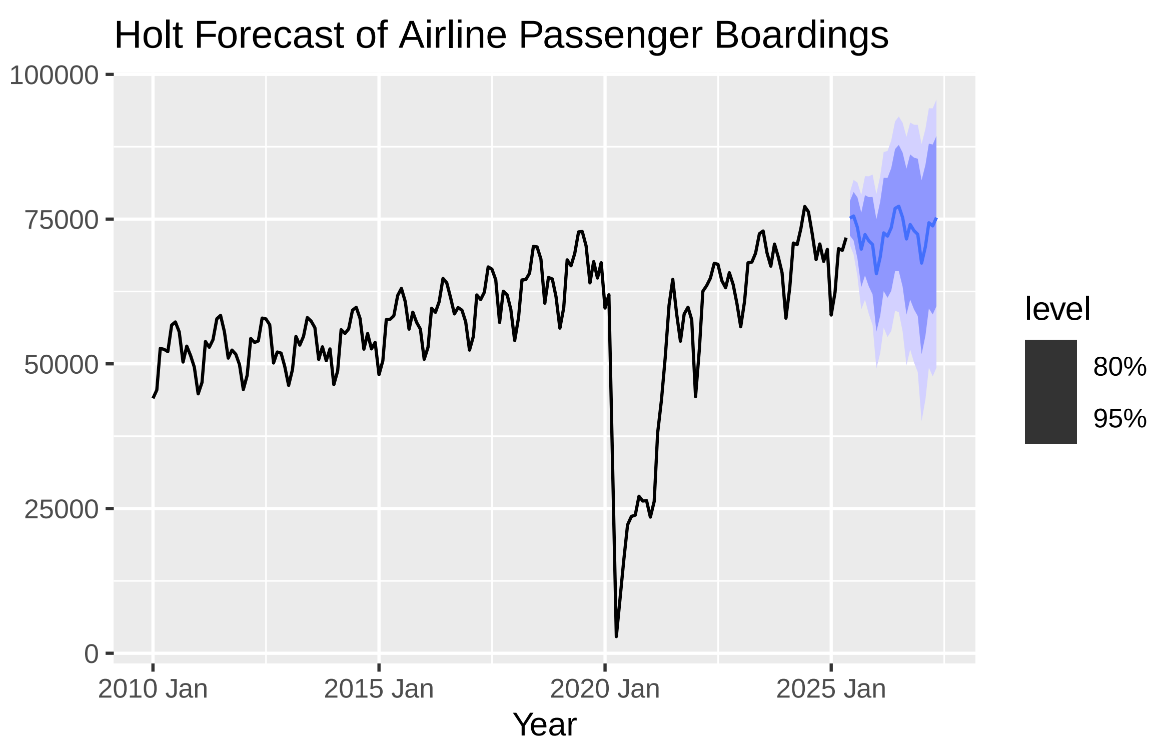
Among all benchmark and ensemble models, the Holt model achieved the lowest RMSE (363.98 million) and MAE (325.19 million), indicating the best overall predictive accuracy. The ensemble combinations (comb1–comb4) did not outperform the Holt model, although comb2 showed relatively competitive results. Therefore, the Holt model was selected for refitting on the complete dataset to generate the final forecasts.

Now, using the fitted model to forecast for the next h periods.

```
# Forecast for 24 periods ahead.
fc_original <- my_fit |>
  forecast(h = 24) |>
  as_tibble() |>
  mutate(Forecast_Original = (lambda * .mean + 1)^(1/lambda))

fc_original |> select(observation_date, .mean)
#> # A tibble: 24 x 2
#>    observation_date  .mean
#>               <mth>  <dbl>
#>  1         2025 Jun 75159.
#>  2         2025 Jul 75542.
#>  3         2025 Aug 73550.
#>  4         2025 Sep 69806.
#>  5         2025 Oct 72335.
#>  6         2025 Nov 71244.
#>  7         2025 Dec 70610.
#>  8         2026 Jan 65565.
```

```
#>  9         2026 Feb 68466.
#> 10         2026 Mar 72624.
#> # i 14 more rows
```

```
fc_holt <- my_fit |>
  forecast(h = 24) #2 years ahead

autoplot(fc_holt, mydata) +
  labs(
    title = "Holt Forecast of Airline Passenger Boardings",
    x = "Year",
    y = "Passengers (in thousands)" +
    scale_y_continuous(labels = scales::label_number(scale = 1e-3))
  )
```




Based on the seasonal Holt model fitted to the complete dataset, U.S. domestic airline passenger traffic is expected to continue following a seasonal upward trend over the next 10 months. While the overall trajectory remains positive, the forecasts reflect regular seasonal fluctuations, with expected peaks in summer months and smaller troughs in winter months. For example, passenger numbers are projected to rise to approximately 75,542 in July 2025, dip to around 69,806 in September 2025, and increase again to about 72624.49 by March 2026 (original scale, back-transformed from Box-Cox).

The Holt model was selected because it achieved the lowest RMSE (363.98 million) and MAE (325.19 million) with a MAPE (14.39%) among all benchmark and ensemble models, indicating it best captures both the underlying trend and the seasonal component in passenger numbers. While residual diagnostics reveal minor short-term correlation (at Lags 1 and 13) and the extreme pandemic shock, the model is appropriate for short-term forecasting since the Ljung-Box test confirms overall model adequacy, and the Holt model effectively reflects the long-term upward trend combined with expected seasonal fluctuations, critical for operational and strategic planning.

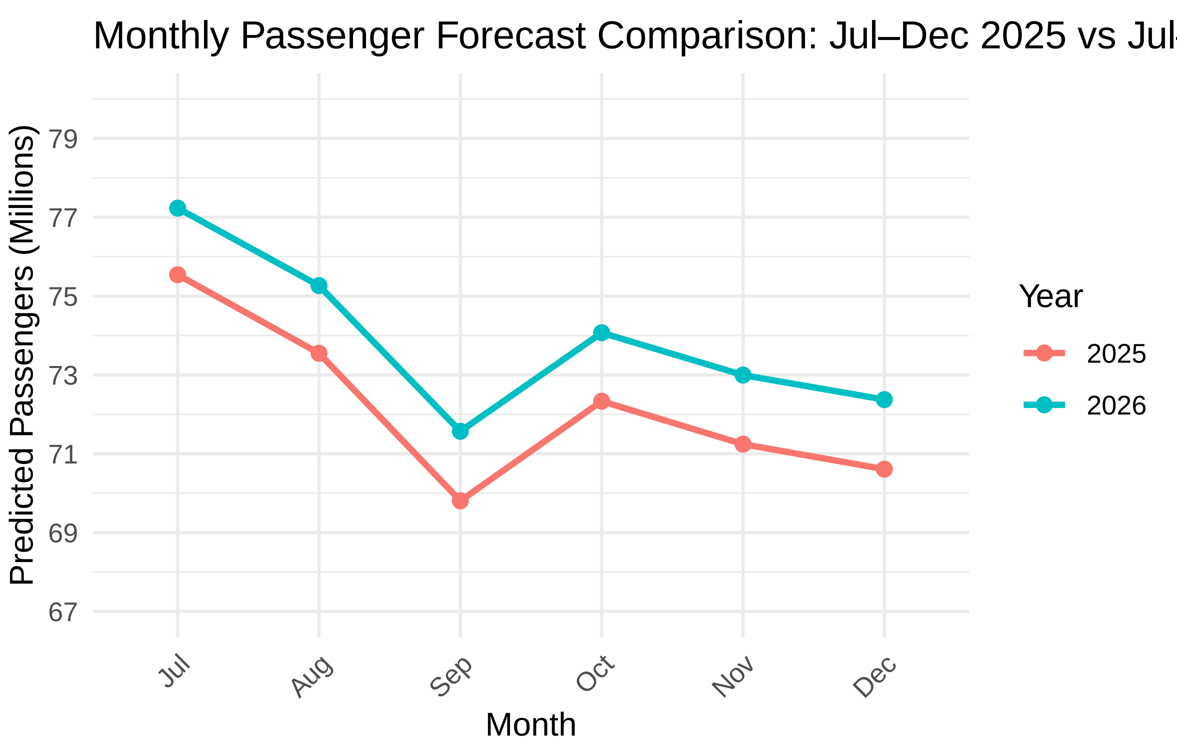
```r
forecast_table <- fc_original %>%
  arrange(observation_date) %>%
  mutate(
    Change = .mean - lag(.mean),
    Growth = (.mean / lag(.mean) - 1) * 100
  ) %>%
  select(
    Month = observation_date,
    `Predicted Passengers (in thousands)` = .mean,
    `Change (Thousands)` = Change,
    `Growth (%)` = Growth
  ) %>%
  mutate(
    Year = year(Month),
    Month_Num = month(Month),
    Month_Label = format(Month, "%b")
  )

# Defining months from July to next June
month_order <- c("Jul","Aug","Sep","Oct","Nov","Dec","Jan","Feb","Mar","Apr","May","Jun")

# Filter July onwards for 2025 and 2026 separately
forecast_compare <- forecast_table %>%
  filter(
    (Year == 2025 & Month_Num >= 7) | (Year == 2026 & Month_Num >= 7)
  ) %>%
  arrange(Year, Month) %>%
  mutate(Month_Label = factor(Month_Label, levels = month_order))
```

```r
forecast_compare_table <- forecast_compare %>%
  select(
    Year,
    Month = Month_Label,
    `Predicted Passengers (Thousands)` = `Predicted Passengers (in thousands)`,
    `Change (Thousands)`,
    `Growth (%)`
  )

print(forecast_compare_table)
#> # A tibble: 12 x 5
#>     Year Month Predicted Passengers (Thousan~1 `Change (Thousands)` `Growth (%)`
#>    <dbl> <fct>                           <dbl>                <dbl>        <dbl>
#>  1  2025 Jul                            75542.                 383.        0.510
#>  2  2025 Aug                            73550.               -1992.       -2.64
#>  3  2025 Sep                            69806.               -3744.       -5.09
#>  4  2025 Oct                            72335.                2528.        3.62
#>  5  2025 Nov                            71244.               -1090.       -1.51
#>  6  2025 Dec                            70610.                -634.       -0.890
#>  7  2026 Jul                            77230.                 381.        0.495
#>  8  2026 Aug                            75265.               -1965.       -2.54
#>  9  2026 Sep                            71571.               -3695.       -4.91
#> 10  2026 Oct                            74070.                2500.        3.49
#> 11  2026 Nov                            72997.               -1074.       -1.45
#> 12  2026 Dec                            72374.                -623.       -0.853
#> # i abbreviated name: 1: `Predicted Passengers (Thousands)`
```

The table presents the monthly passenger forecasts (in thousands) for July to December 2025 and 2026, along with the month-to-month changes and growth rates. In July 2025, passenger numbers increase slightly by 0.51%, followed by declines of 2.64% in August and 5.09% in September, representing the largest drop within the period. October shows a rebound with 3.62% growth, while November and December experience smaller decreases of 1.51% and 0.89%, respectively. A similar pattern is observed in 2026, with a modest rise of 0.50% in July, declines of 2.54% in August and 4.91% in September, a recovery of 3.49% in October, and minor decreases in November and December of 1.45% and 0.85%. These trends highlight seasonal fluctuations, with September showing the steepest decline and October demonstrating a consistent recovery across both years.

```r
# Plot
p <- ggplot(forecast_compare, aes(x = Month_Label,
                                  y = `Predicted Passengers (in thousands)`/1000,
                                  group = Year, color = factor(Year))) +
```

```
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Monthly Passenger Forecast Comparison: Jul-Dec 2025 vs Jul-Dec 2026",
    x = "Month",
    y = "Predicted Passengers (Millions)",
    color = "Year"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(
    limits = c(67, 80),
    breaks = seq(67, 80, by = 2)
  )
p
```




The graph compares the monthly passenger forecasts for July to December in 2025 and 2026 (in millions). Both years follow a similar seasonal trend, with passenger numbers peaking in

July, then declining through August and reaching the lowest point in September. October shows a notable recovery, followed by gradual decreases in November and December. Overall, the 2026 forecast is consistently higher than 2025, indicating expected growth year-on-year. The steepest decline occurs in September, while October consistently rebounds in both years, highlighting a seasonal pattern in passenger demand.

Significance for Stakeholders

Airlines: Can use the forecast to plan flight schedules, staffing, and aircraft maintenance.Helps optimize ticket pricing and anticipate demand spikes without overbooking or under-utilizing capacity.

Airports: Can allocate resources efficiently (check-in counters, security staff, gate operations).Supports planning for infrastructure upgrades or temporary capacity adjustments during high-demand months.

Policymakers and Transportation Authorities: Can monitor trends to ensure safety, regulatory compliance, and smooth operations.Helps in long-term transportation planning, including runway, terminal, and public transit coordination.

Investors and Industry Analysts: Can anticipate revenue growth opportunities and make informed decisions on airline stock or expansion strategies.

By using this forecast, stakeholders can plan proactively, mitigate risks, and capitalize on the expected increase in domestic air travel, translating statistical predictions into actionable decisions.

## 8 Results

The forecast analysis indicates that U.S. domestic airline passenger numbers are expected to continue increasing over the next 24 months, following a clear seasonal pattern with peaks in summer months and troughs in early fall. Statistically, the Holt model achieved the lowest forecast errors (RMSE, MAE, MAPE), indicating a reliable fit to historical data. Practically, this means that passenger volumes will rise steadily, with month-to-month fluctuations typically within $\pm 5\%$, reflecting normal seasonal variation. For stakeholders such as airlines, airports, and policymakers, these forecasts provide actionable insights to optimize flight schedules, allocate staffing and resources efficiently, and anticipate periods of high or low demand. The results allow decision-makers to plan proactively, reduce operational risks, and capitalize on seasonal growth in domestic air travel.

# 9 Discussion and Conclusion

The analysis shows that U.S. domestic airline passenger traffic is expected to increase over the next 24 months, with forecasts indicating a rise from approximately 75.5 million passengers in July 2025 to higher levels in 2026. Quantitatively, the Holt model achieved strong predictive accuracy (low RMSE, MAE, and MAPE) and captures both the trend and seasonal fluctuations in passenger numbers. Qualitatively, the results reveal predictable seasonal patterns—summer peaks and early-fall troughs—that are crucial for operational planning. While the model cannot account for unforeseen structural shocks, such as a pandemic, minor residual correlations (Lags 1 and 13) do not undermine the forecast's reliability. These findings provide actionable insights for stakeholders: airlines can schedule flights and optimize pricing, airports can allocate staffing efficiently, and policymakers can plan for infrastructure and capacity management. Overall, the Holt model offers a robust and practical tool for informed, evidence-based decision-making in the aviation sector.