# UChicago Time Series Final Project (Example)

Student A Student B Student C

2025-03-11

## Table of contents

# 1 Introduction

This project uses the built-in **AirPassengers** dataset, which records monthly totals of international airline passengers from 1949 to 1960. The dataset exhibits a clear upward trend and strong seasonal patterns, making it an ideal candidate for forecasting analysis. In this report, we compare three different forecasting approaches:

- **ARIMA**: Automatically identifies optimal parameters to model trends and autocorrelations.
- **ETS**: Decomposes the series into error, trend, and seasonal components for clear interpretability.
- **TBATS**: Utilizes trigonometric representations and transformations to handle complex seasonality and non-linear trends.

Each model's detailed analysis is provided in its respective child document, enabling a modular workflow that facilitates collaborative development and maintenance.

# 2 Data Overview

The **AirPassengers** dataset contains monthly totals of international airline passengers over a 12-year period. Its combination of an upward trend and seasonal fluctuations presents a robust challenge for forecasting methods, making it a common benchmark in time series analysis.

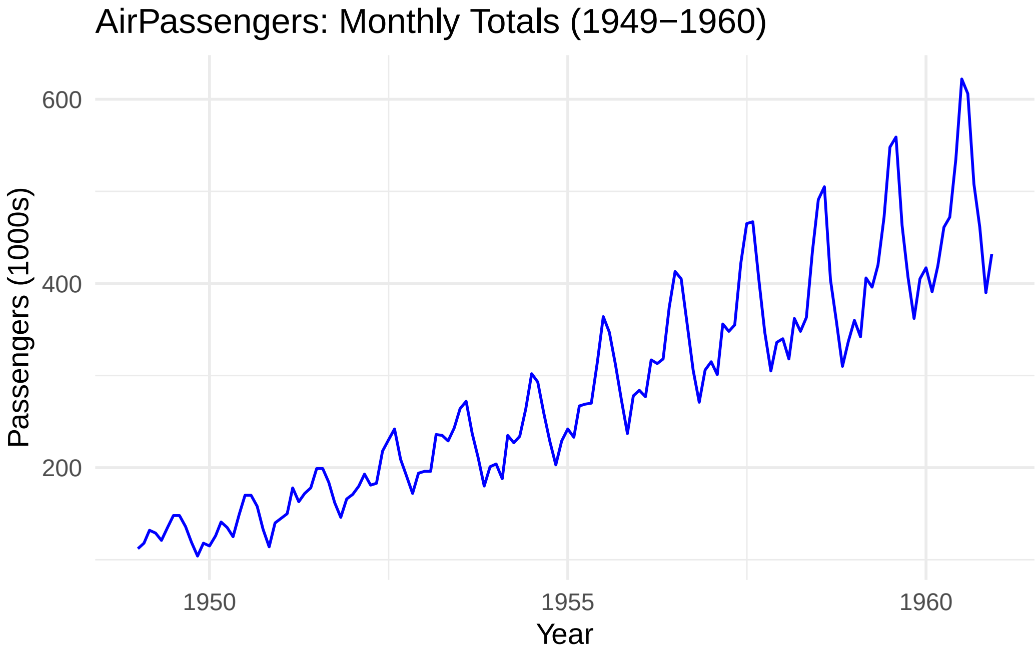
# 3 EDA

# 4 Exploratory Data Analysis (EDA)

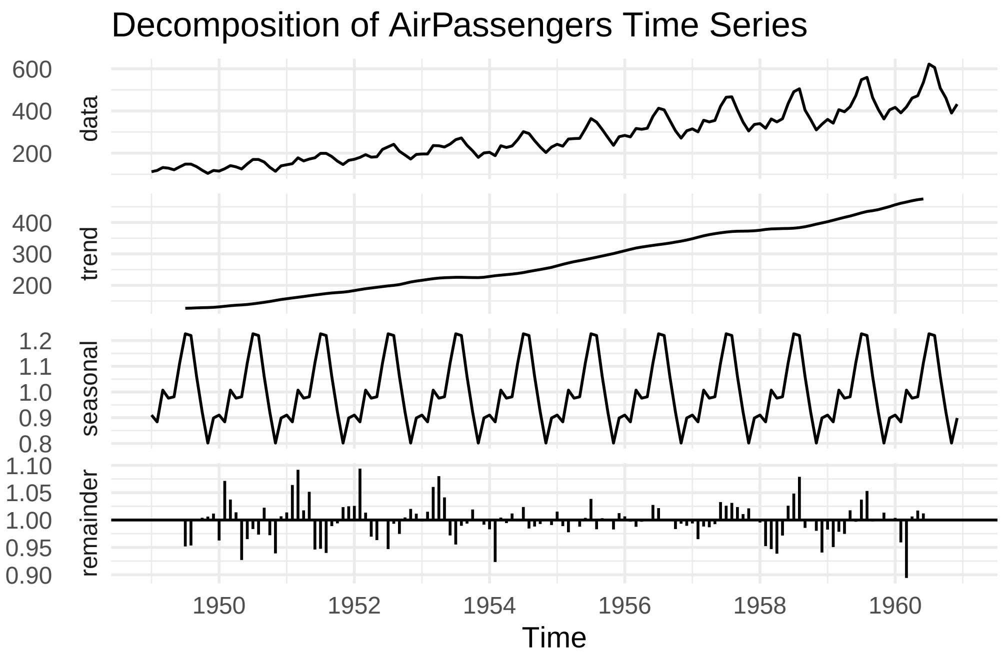
```
This dataset starts in 1949 and ends in 1960 with 144 monthly observations.
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  104.0   180.0   265.5   280.3   360.5   622.0 
```

## 4.1 Raw Time Series Plot



### 4.2 Time Series Decomposition




Observations from EDA: - The number of passengers generally increases over time, indicating a growing trend. - There is a clear seasonal pattern (peaks in mid-year). - The variance grows over time, hinting at non-stationarity. - The multiplicative decomposition visually shows strong seasonality and an upward trend.

## 5 Model 1: ARIMA Analysis

### 5.1 Model Overview

ARIMA (AutoRegressive Integrated Moving Average) is used for time series forecasting by combining autoregression, differencing, and moving average components. It is particularly effective for data with trends and seasonality, and the `auto.arima` function automatically selects the best model parameters based on criteria like AICc.

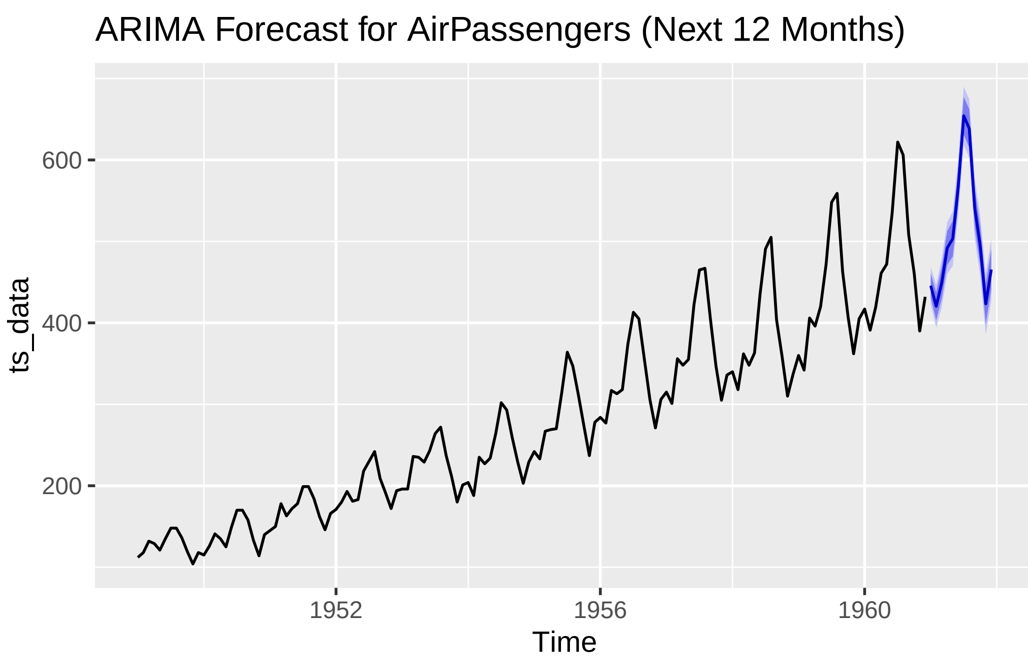
### 5.2 Model Fitting

In this section, we use the built-in `AirPassengers` dataset. We fit an ARIMA model using the `auto.arima` function from the `forecast` package.

```
Series: ts_data
ARIMA(2,1,1)(0,1,0)[12]

Coefficients:
         ar1     ar2      ma1
      0.5960  0.2143  -0.9819
s.e.  0.0888  0.0880   0.0292

sigma^2 = 132.3:  log likelihood = -504.92
AIC=1017.85   AICc=1018.17   BIC=1029.35

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE
Training set 1.342306 10.84619 7.867539 0.4206996 2.800458 0.245628
                     ACF1
Training set -0.001248451
```

### 5.3 Model Evaluation

We evaluate the ARIMA model by forecasting the next 12 months and visualizing the results. This visualization shows how well the model captures the underlying patterns in the `AirPassengers` data.



### 5.3.1 RMSE and MAPE Metrics

We calculate the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) using the accuracy function to evaluate the model's performance on the training set.

Table 1: ARIMA Model Performance Metrics

| Metric | Value |
|---|---|
| RMSE | 10.846187 |
| MAPE | 2.800457 |

Evaluation Summary: - Model selection is based on AICc and diagnostic checks. - Further residual analysis and out-of-sample validation can be used to refine the model.

# 6 Model 2: ETS Analysis

## 6.1 Model Overview

The ETS (Error, Trend, Seasonal) model decomposes a time series into its components directly. It is particularly useful when the data exhibits clear level, trend, and seasonal components. ETS models automatically select the best combination of error, trend, and seasonality structures.

## 6.2 Model Fitting

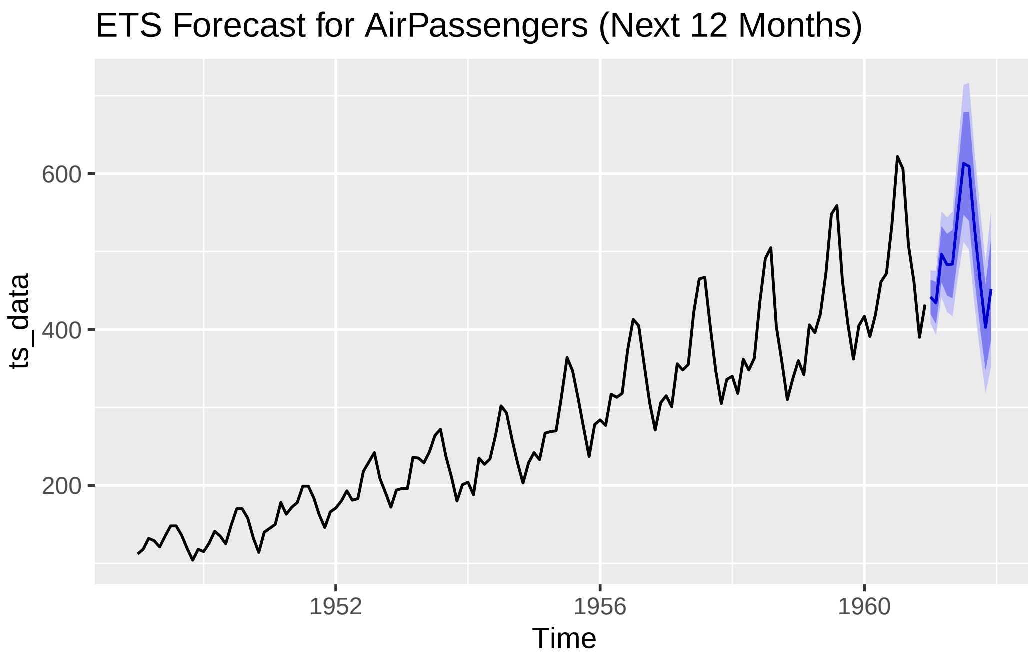
We again use the built-in `AirPassengers` dataset. The ETS model is fitted using the `ets` function from the `forecast` package.

```
ETS(M,Ad,M)

Call:
ets(y = ts_data)

  Smoothing parameters:
    alpha = 0.7096
    beta  = 0.0204
    gamma = 1e-04
    phi   = 0.98
```

```
  Initial states:
    l = 120.9939
    b = 1.7705
    s = 0.8944 0.7993 0.9217 1.0592 1.2203 1.2318
           1.1105 0.9786 0.9804 1.011 0.8869 0.9059

  sigma:  0.0392

     AIC     AICc      BIC
1395.166 1400.638 1448.623

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 1.567359 10.74726 7.791605 0.4357799 2.857917 0.2432573 0.03945056
```

### 6.3 Model Evaluation

We evaluate the ETS model by forecasting the next 12 months and plotting the forecast. This visualization shows how well the model captures the underlying patterns in the `AirPassengers` data.



### 6.3.1 RMSE and MAPE Metrics

We calculate the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) using the accuracy function to evaluate the model's performance on the training set.

Table 2: ETS Model Performance Metrics

| Metric | Value |
|---|---|
| RMSE | 10.747256 |
| MAPE | 2.857917 |

Evaluation Summary: - ETS provides an alternative perspective to ARIMA, particularly when dealing with multiplicative seasonal effects. - Accuracy measures and visual inspection aid in comparing ETS with other models.

# 7 Model 3: TBATS Analysis

## 7.1 Model Overview

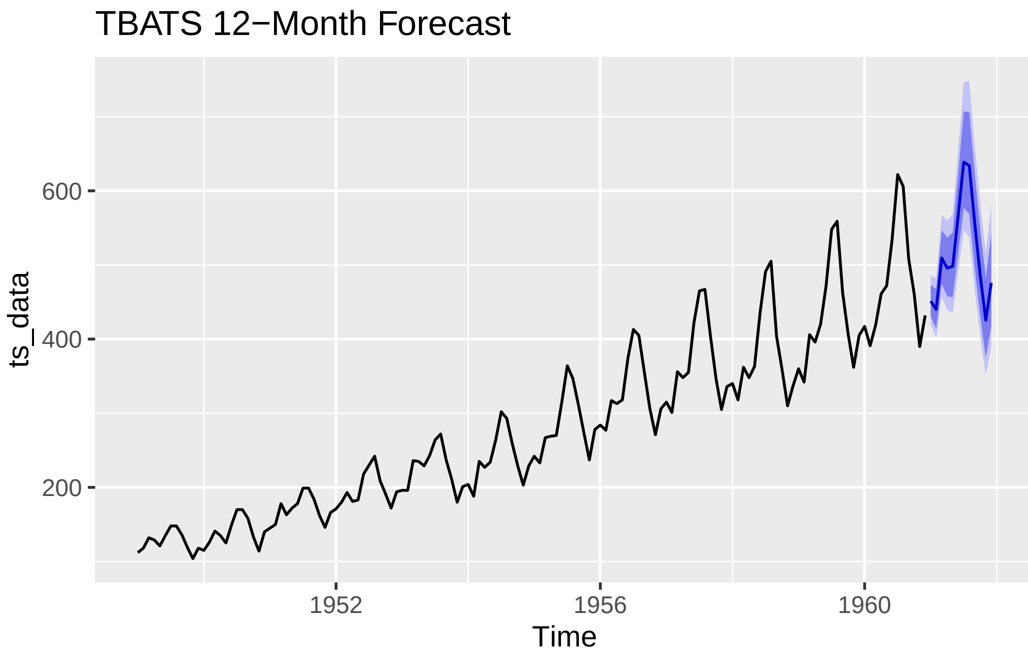
TBATS (Trigonometric, Box-Cox transformation, ARMA errors, Trend, and Seasonal components) is designed to handle complex seasonal patterns and non-linear trends. It is especially useful for data with multiple seasonal periods or subtle non-linear behaviors.

## 7.2 Model Fitting

For this analysis, we use the built-in `co2` dataset, which records atmospheric CO2 concentrations. We fit a TBATS model using the `tbats` function from the `forecast` package.

## 7.3 Model Evaluation

We evaluate the TBATS model by forecasting the next 12 months and plotting the forecast. This visualization shows how well the model captures the underlying patterns in the `AirPassengers` data.




### 7.3.1 RMSE and MAPE Metrics

We calculate the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) using the accuracy function to evaluate the model's performance on the training set.

Table 3: TBATS Model Performance Metrics

| Metric | Value |
|---|---|
| RMSE | 10.661439 |
| MAPE | 2.851916 |

Evaluation Summary: - TBATS summary provides insights into the transformation and seasonal components used. - Forecast visualization aids in assessing the model's predictive performance.

# 8 Model Comparison and Discussion

- ARIMA automatically identifies parameters that capture both trend and autocorrelation.

- ETS explicitly decomposes the time series, making it easier to understand the underlying level, trend, and seasonal components.
- TBATS offers flexibility when dealing with complex seasonal patterns or subtle non-linear behaviors.

The forecasts generated by each model provide complementary insights into future passenger numbers. The choice of model may ultimately depend on factors such as forecasting horizon, interpretability, and the specific nuances present in the data.

# 9 Conclusion

Our analysis of the **AirPassengers** data using ARIMA, ETS, and TBATS models demonstrates that each approach has its own strengths. ARIMA effectively handles autocorrelation, ETS offers a clear component-based decomposition, and TBATS is well-suited for complex seasonal patterns. Future work might include cross-validation for forecast accuracy or even combining model forecasts to further improve performance.

# References

Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. OTexts. https://otexts.com/fpp3/.

Taylor, S. J., and B. Letham. 2003. "Forecasting: Methods and Applications." *Journal of Business Research* 56 (2): 97–110. https://doi.org/10.1016/S0148-2963(02)00146-6.