Airline Passenger Traffic Analysis and Forecasting Report

Introduction

This report presents an in-depth exploratory data analysis, imputation of missing values, and forecasting of monthly airline passenger traffic data spanning from 1949 to 1960. The objective is to understand the trends, seasonality, and to apply multiple forecasting techniques to accurately predict future passenger volumes.

Data Overview

- **Dataset Description:** Monthly recorded international airline passengers (in thousands) from January 1949 to December 1960.
- **Format:** Time series data with a single feature – number of passengers per month.
- **Missing Values:** Imputation techniques were employed to handle missing data points to ensure integrity of time series analysis.

Missing Data Handling

Three methods were used to impute missing values:

1. **Mean Imputation:** Filling missing values with the mean of observed data.
2. **Linear Interpolation:** Estimating missing values by linear trends between observed points.
3. **Comparison:** Linear interpolation preserved the continuity better than mean imputation, improving data quality for downstream forecasting.

Exploratory Data Analysis (EDA)

- The time series shows a clear upward long-term trend, indicating increasing airline passenger traffic over the years.
- Seasonality is evident, with peaks typically occurring annually around the same months.

- The data is multiplicative in seasonality, where seasonal fluctuations increase with the trend level.

Visualizations included:

- Line plots of passenger counts over time.
- Seasonal decompositions showing trend, seasonal, and residual components.

Forecasting Methods Applied and Performance

Multiple forecasting techniques were evaluated by training on historical data and comparing forecasts using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error):

Method	**RMSE**	**MAPE (%)**
Naive Method	137.51	23.63
Simple Average Method	219.69	44.28
Simple Moving Average Method	103.33	15.54
Simple Exponential Smoothing	107.65	16.49
Holt's Exponential Smoothing	71.94	11.11
Holt's Winter Additive	35.10	6.53
Holt's Winter Multiplicative	34.46	6.82
Auto Regressive (AR) Method	93.39	13.77
Moving Average (MA) Method	91.21	13.39
Auto Regressive Moving Average (ARMA)	88.88	12.89
ARIMA Method	88.88	12.89
SARIMA Method	37.32	7.32
SARIMAX Method (with exogenous variables)	**26.63**	**4.65**

Table 1: Forecasting methods performance comparison.

Observations

- The **Naive method** has relatively high error but serves as a baseline.
- **Simple average methods** performed the worst, highlighting the importance of accounting for seasonality and trend.
- Moving average and AR-based models improve performance.
- **Holt's Winter methods (additive and multiplicative)** significantly reduce errors by capturing trend and seasonality.
- The **SARIMA** and **SARIMAX** models are the best performers, with SARIMAX delivering the lowest RMSE (26.63) and MAPE (4.65%), indicating excellent forecasting accuracy.

Event Data

The event variable, which could represent external occurrences impacting passenger volume (e.g., strikes, economic shocks, or special events), was integrated in SARIMAX models to improve forecasting accuracy.

Conclusion and Recommendations

- The airline passenger traffic shows a strong seasonal pattern combined with an upward trend.
- Handling missing data using linear interpolation is preferred for maintaining time series consistency.
- Advanced forecasting models that capture seasonality and trend (Holt-Winters, SARIMA, SARIMAX) outperform simple methods.
- **SARIMAX** is recommended for future forecasting tasks especially when external variables/events are available.
- This analytical approach can support airlines and airports in demand planning and capacity management.

If you need any specific visualizations or further insights from the notebook, feel free to ask!