Analysis of Data

---

1. Dataset Overview:

- The dataset contains monthly passenger data from 1949 to 1960.
- Columns:
    - **Month**: The date in "YYYY-MM" format.
    - **Thousands of Passengers**: Number of passengers in thousands.

---

2. Summary Statistics:

- **Mean Passengers**: ~280.3 thousand.
- **Range**: Minimum of 104 thousand (low season) and maximum of 622 thousand (high season).
- **Standard Deviation**: 119.97 thousand, showing moderate variability in passenger numbers.
- **No Missing Values**: The dataset is complete.

---

3. Trends and Seasonal Patterns:

a. Overall Trend:

- A clear upward trend in the number of passengers over time, reflecting growth in air travel demand.

b. Seasonal Patterns:

- **Peaks**: Passenger numbers increase significantly during **summer months** (July and August), likely due to vacations.
- **Troughs**: Numbers drop during **winter months** (November to February), indicating a post-holiday and off-season decline.

c. Cyclical Nature:

- The data exhibits predictable yearly cycles with consistent peaks and troughs, indicating strong seasonal effects.

---

4. Growth Analysis:

- **Year-over-Year (YoY) Growth**:
    - Passenger numbers grow steadily year after year.
    - Growth rates fluctuate, with some years showing double-digit increases.

- Example: YoY growth in 1951 was 21.84%.

---

5. Anomaly Detection:

- Anomalies are months with unusually high or low passenger counts compared to the overall trend.
- These were identified using Z-scores (>2 or <-2).
- Example: **July 1960** had a very high passenger count (622 thousand, Z-score: 2.86).

---

6. Decomposition of Time Series:

- **Observed**: Shows the actual data.
- **Trend**: Indicates a steady increase in passengers over time.
- **Seasonal**: Captures predictable yearly fluctuations.
- **Residuals**: Represents the noise or random variations in the data.

---

7. Forecasting Future Passenger Counts:

- Forecasting was performed using the **Holt-Winters Exponential Smoothing** method.
- Predictions for the last 12 months (1960) were compared to actual values:
  - The model captured trends and seasonality effectively.
  - Some deviations occurred, but the forecast aligns well with the data.

---

8. Visualizations:

- **Line Plot**: Displays overall trends in passenger numbers.
- **Bar Chart**: Shows average passengers by month, highlighting seasonal effects.
- **Growth Plot**: Year-over-year growth rates in passenger numbers.
- **Anomalies Plot**: Highlights months with significant deviations.
- **Forecast Plot**: Compares actual and forecasted passenger numbers.

Key Insights:

1. **Seasonal Dependence**: Summer months are consistently the peak period, while winter months show a slowdown.
2. **Growth Trend**: Air travel grew steadily from 1949 to 1960, likely due to increasing popularity and accessibility.
3. **Anomalies**: Significant deviations from trends provide insights into unusual travel patterns or external factors.
4. **Forecasting**: The model effectively predicts future values, leveraging historical trends and seasonality.