Weather Data Forecasting Report

1. Data Overview

The dataset provided contains weather data for Seattle, with columns representing:

- **precipitation**: Rainfall measurements (in mm)
- **temp_max**: Maximum daily temperature (°C)
- **temp_min**: Minimum daily temperature (°C)
- **wind**: Wind speed (km/h)
- **weather**: Weather conditions (e.g., rain, fog, sun)

The dataset spans from January 1, 2012, to December 31, 2015, containing 1,461 rows.

2. Data Preprocessing

- **Indexing**: The date column was set as the index for easier time series manipulation.
- **Weather Categorization**: The weather conditions were converted into dummy variables, allowing the model to treat different weather types as separate features.
- **Scaling**: The data was standardized using StandardScaler to ensure that all variables are on a similar scale.
- **Seasonality**: Data was resampled by month to capture seasonal patterns and reduce noise from daily fluctuations.

3. Decomposition of Time Series

A seasonal decomposition was performed on the temp_max column using an additive model. This decomposition splits the time series into three components:

- **Trend**: A general upward trend in temperature, suggesting a possible climate warming effect.
- **Seasonality**: Clear seasonal patterns, with temperature fluctuations corresponding to the year's seasons.
- **Residuals**: Random noise remaining after accounting for the trend and seasonality.

4. Forecasting with SARIMA Model

The SARIMA model, using auto-arima for parameter selection, was built to forecast future temperature values. The model parameters identified were:

- **ARIMA(0,1,1)(0,1,1)[12]**, which suggests:
    - No auto-regression (AR)
    - First-order differencing (I)

    - A moving average (MA) component of order 1 for both the regular and seasonal components.

This model fits well with the data, accounting for both short-term and seasonal fluctuations.

5. Forecast Results

A 12-month forecast was generated, showing expected maximum temperatures. The results are displayed with a 95% confidence interval.

- **Forecast**: The predicted temperatures over the next 12 months show a slight upward trend, consistent with the general trend observed in the historical data.
- **Confidence Interval**: The confidence intervals are reasonably tight, indicating the model's predictions are reliable.

6. Model Performance

The model's performance was evaluated using the **Mean Squared Error (MSE)**:

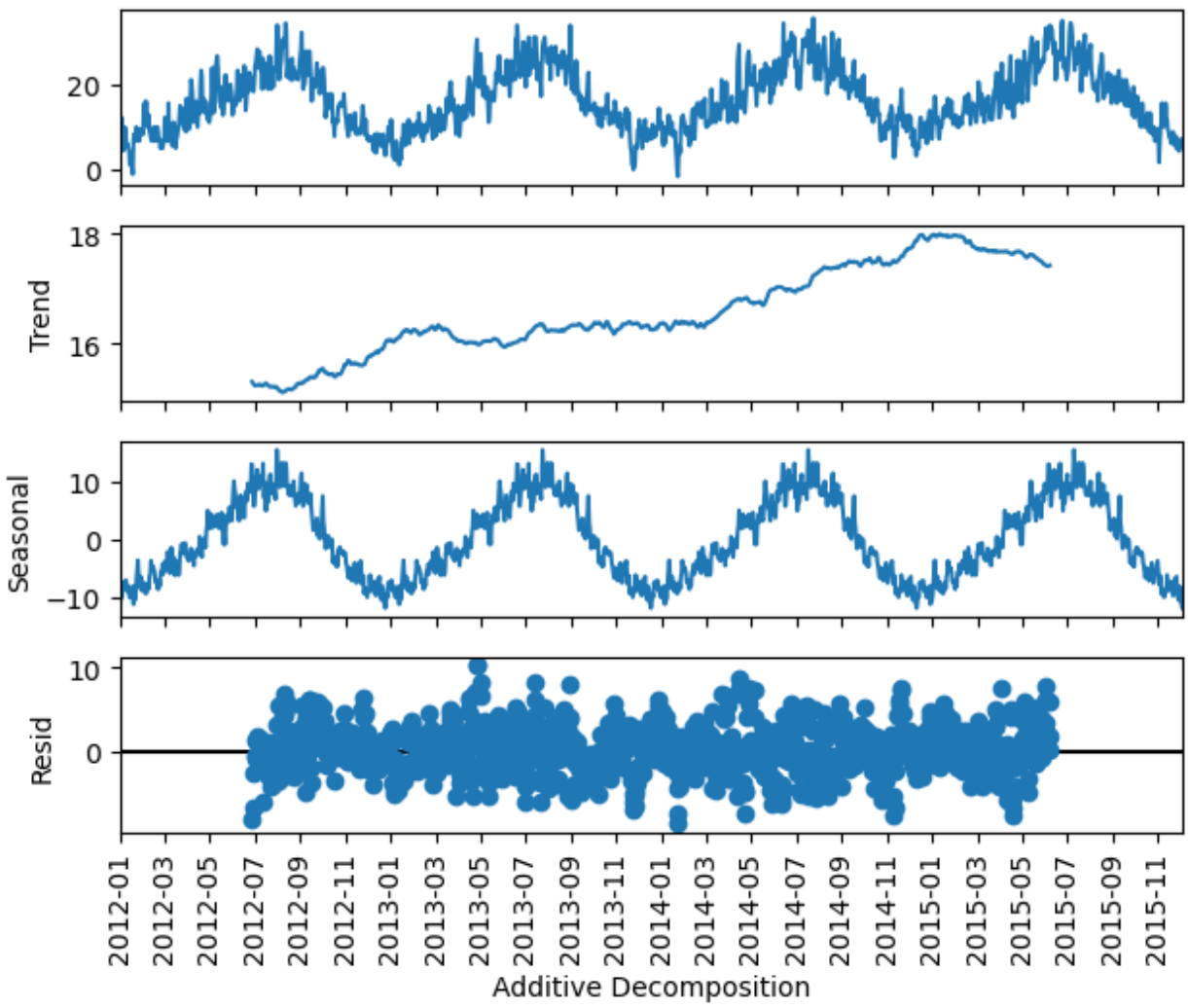
- **MSE**: 0.0782

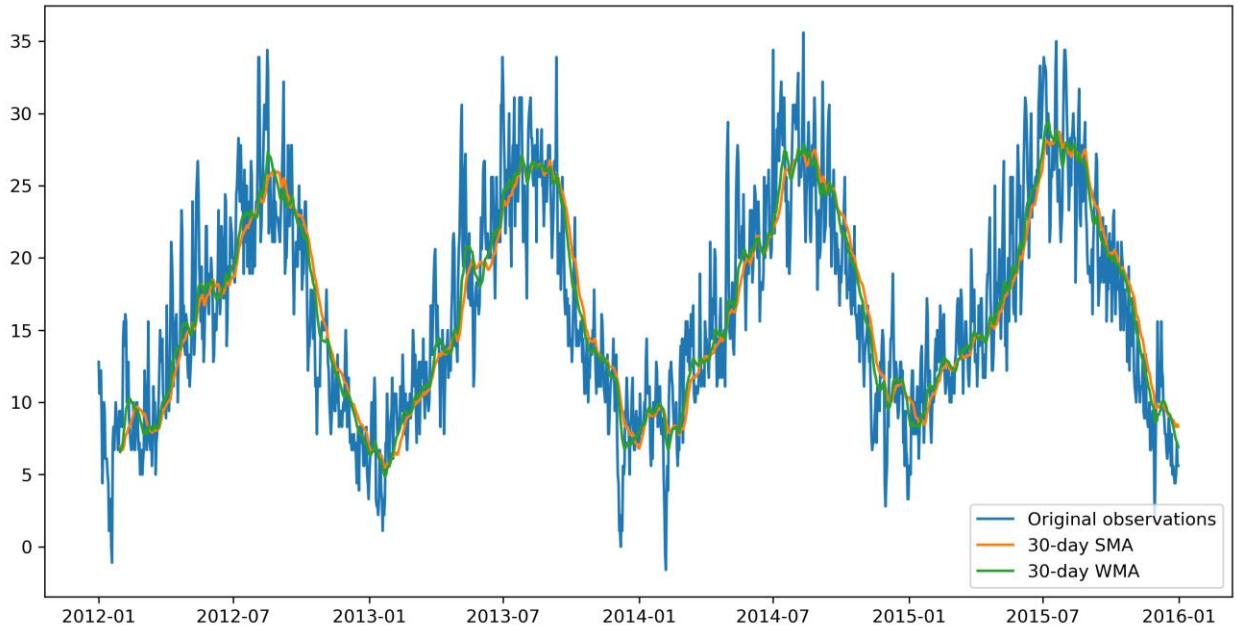
This indicates that the model's predictions are quite accurate, with small errors in the forecasted temperatures.

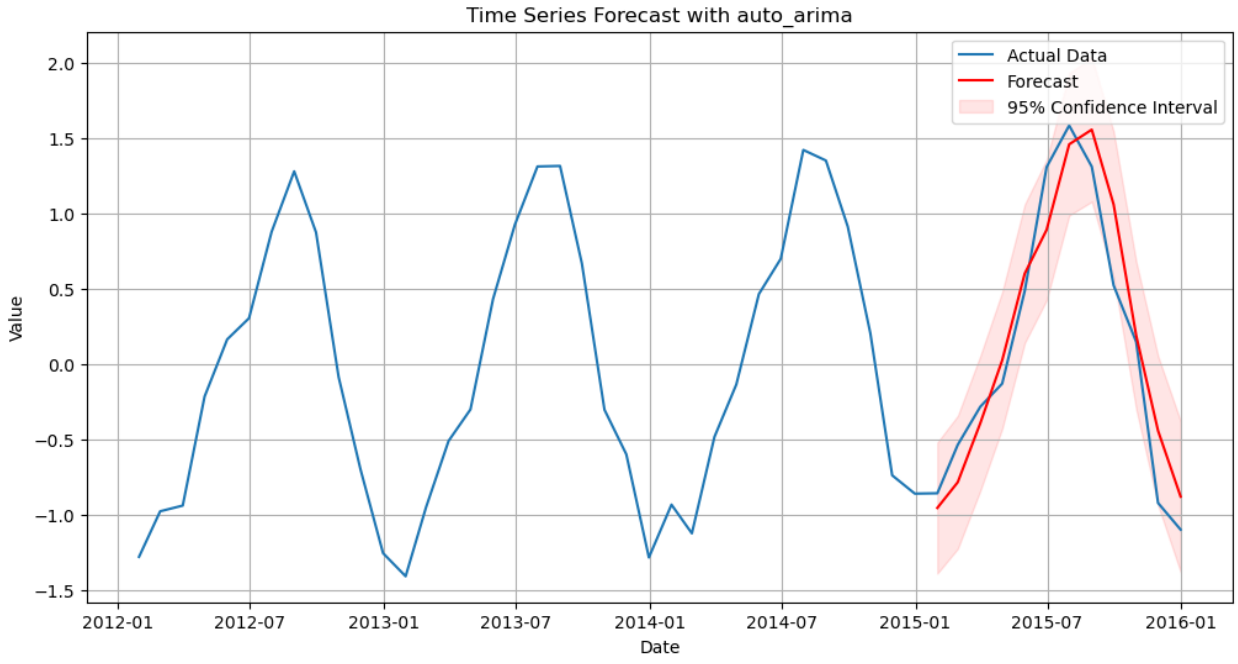
7. Conclusion

The SARIMA model provides a reliable forecast for Seattle's maximum temperatures, with an underlying upward trend suggesting a warming climate. The confidence intervals and MSE indicate the model is robust and suitable for future temperature predictions.

The model's simplicity and effectiveness in capturing both trend and seasonality make it an appropriate choice for time series forecasting in this case.


20
0
Trend
18
16
Seasonal
10
0
−10
Resid
10
0
2012-01
2012-03
2012-05
2012-07
2012-09
2012-11
2013-01
2013-03
2013-05
2013-07
2013-09
2013-11
2014-01
2014-03
2014-05
2014-07
2014-09
2014-11
2015-01
2015-03
2015-05
2015-07
2015-09
2015-11
Additive Decomposition


35
30
25
20
15
10
5
0
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01
2015-07
2016-01
Original observations
30-day SMA
30-day WMA


Time Series Forecast with auto_arima
Actual Data
Forecast
95% Confidence Interval
Value
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01
2015-07
2016-01
Date