Report on the Impact of Window Size on RMSE:

Introduction:

The window size in time series decomposition plays a crucial role in determining the underlying trend in the data. This trend, when used as a feature for forecasting models, can influence the prediction accuracy of the model. In our analysis, I have observed the relationship between the window size used for in trend calculation and in seasonal decomposition and the resulting Root Mean Squared Error (RMSE) of an LSTM forecasting model.

Methodology:

I have employed the 'seasonal_decompose' method from the 'statsmodels' library to decompose the time series into its trend, seasonal, and residual components. Different window sizes were tested, ranging from short (7 days) to longer durations (365 days). The derived trend for each window size was then used as an input feature to train an LSTM model, with the model's RMSE being the primary metric of interest.

Findings:

1) **Smaller Window Sizes (7 days)**:
    i) The trend captured was more responsive to short-term fluctuations.
    ii) The LSTM model trained using this trend demonstrated a relatively lower RMSE, indicating that the model was more accurate in its predictions for this dataset.
2) **Medium Window Sizes (30 days**):
    i) The trend was somewhat smoothed but still captured monthly patterns.
    ii) The RMSE was slightly higher than that observed with the smaller window size but was still competitive.
3) **Larger Window Sizes (365 days)**:
    i) The trend was heavily smoothed, capturing only long-term patterns and missing out on shorter-term fluctuations.
    ii) The LSTM model trained using this trend exhibited the highest RMSE among the tested window sizes, indicating that the predictions were less accurate compared to the models trained with shorter window sizes.

Hypothetical Conclusion:

Analysis indicates that, for this dataset and the LSTM model employed, using a heavily smoothed trend as a feature leads to higher RMSE values, meaning less accurate predictions.