30 Days Weather Forecast: Amberpet Mandal, Telangana State - India

- Hilda Nderitu

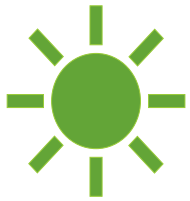
- Advanced Machine Learning
- End of Module Project
- 18th October 2025

Introduction: Telangana State

- Indian state, in South-Central part of Indian subcontinent on the high Deccan Plateau
- Capital city: **Hyderabad**
- Consists of **33 districts**, further divided into **584 mandals**
- **Amberpet mandal** – 1 out of the 16 mandals in Hyderabad District

Introduction: Telangana State Climate

- Semi-arid area, predominantly hot & dry climate
- Annual precipitation
- 3 seasons



Summers

- **March – June**
- **Peaks in min-April**
- **Warm to hot & dry**
- **Peak temp – near or >38 °C**

Monsoon

- **July – September**
- **Tropical rains**
- **Rainy**
  southwest monsoon winds
- **Avg rain: 900mm/year**

Winter

- **October – February**
- **Mild & dry (little humidity)**
- **Avg temp: 22 - 23 °C**
- **Avg min temp: ~ 15 °C,**
  around Jan & Feb

Problem Statement

- Weather forecasting is important as:
  - It allows people to prepare for weather events & make informed decisions
  - Helps save lives & protect property by predicting severe weather
  - Supports farmers in planting & harvesting - agriculture
  - Guides travel planning
  - Enables businesses to manage energy use
- Weather conditions for the next 30 days will be predicted for Amberpet mandal, in Hyderabad district of Telangana State

Objectives

- To develop a **multivariate time series forecast model** for predicting daily weather conditions
- To forecast daily weather conditions for the next 30 days in Amberpet Mandal using the developed model
  - Weather conditions to forecast;
    - Rain (mm)
    - Min Temp (°C)
    - Max Temp (°C)
    - Min Humidity (%)
    - Max Humidity (%)
    - Min Wind Speed (Kmph)
    - Max Wind Speed (Kmph)

Project Overview

Data cleaning

Exploratory Data Analysis
- Line plots
- Box plots & Histograms
- Correlation analysis
- ETS Decomposition
- ACF/PACF Plots

Data preprocessing & Feature Engineering
- Log Transformation
- Robust scaling
- Stationarity check & differencing

Model Fitting
- Splitting dataset
- Training
- Lag order selection

Forecasting & Evaluation

Dataset

- Source: Hugging Face website - https://huggingface.co/datasets/ron-the-code/Telangana_time_series_2023-2025
- Data extracted from Open Data Telangana
- Daily weather conditions for Telangana State from 1st Feb 2023 to 31st Jan 2025
  - 445,212 rows & 10 columns for date, district, mandal, rain, temp, humidity & wind speeds
- Data available for 33 districts & 598 mandals
- Prediction done for Amberpet mandal in Hyderabad District using Amberpet mandal time series data
  - 731 rows, 7 columns/variables

```
1  weather_data.info() #get dataset information
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 445212 entries, 0 to 445211
Data columns (total 10 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   District              445212 non-null  object
 1   Mandal                445212 non-null  object
 2   Date                  445212 non-null  object
 3   Rain (mm)             445212 non-null  float64
 4   Min Temp (°C)         445212 non-null  float64
 5   Max Temp (°C)         445212 non-null  float64
 6   Min Humidity (%)      445212 non-null  float64
 7   Max Humidity (%)      445212 non-null  float64
 8   Min Wind Speed (Kmph) 445212 non-null  float64
 9   Max Wind Speed (Kmph) 445212 non-null  float64
dtypes: float64(7), object(3)
memory usage: 34.0+ MB
```

- The dataset has no missing data for each column.
- There are 445212 rows of data.
- There are 10 columns.
- 3 columns contain categorical data, namely: District, Mandal & Date.
- 7 columns contain data on rain, temperature, humidity and windspee

Data Cleaning

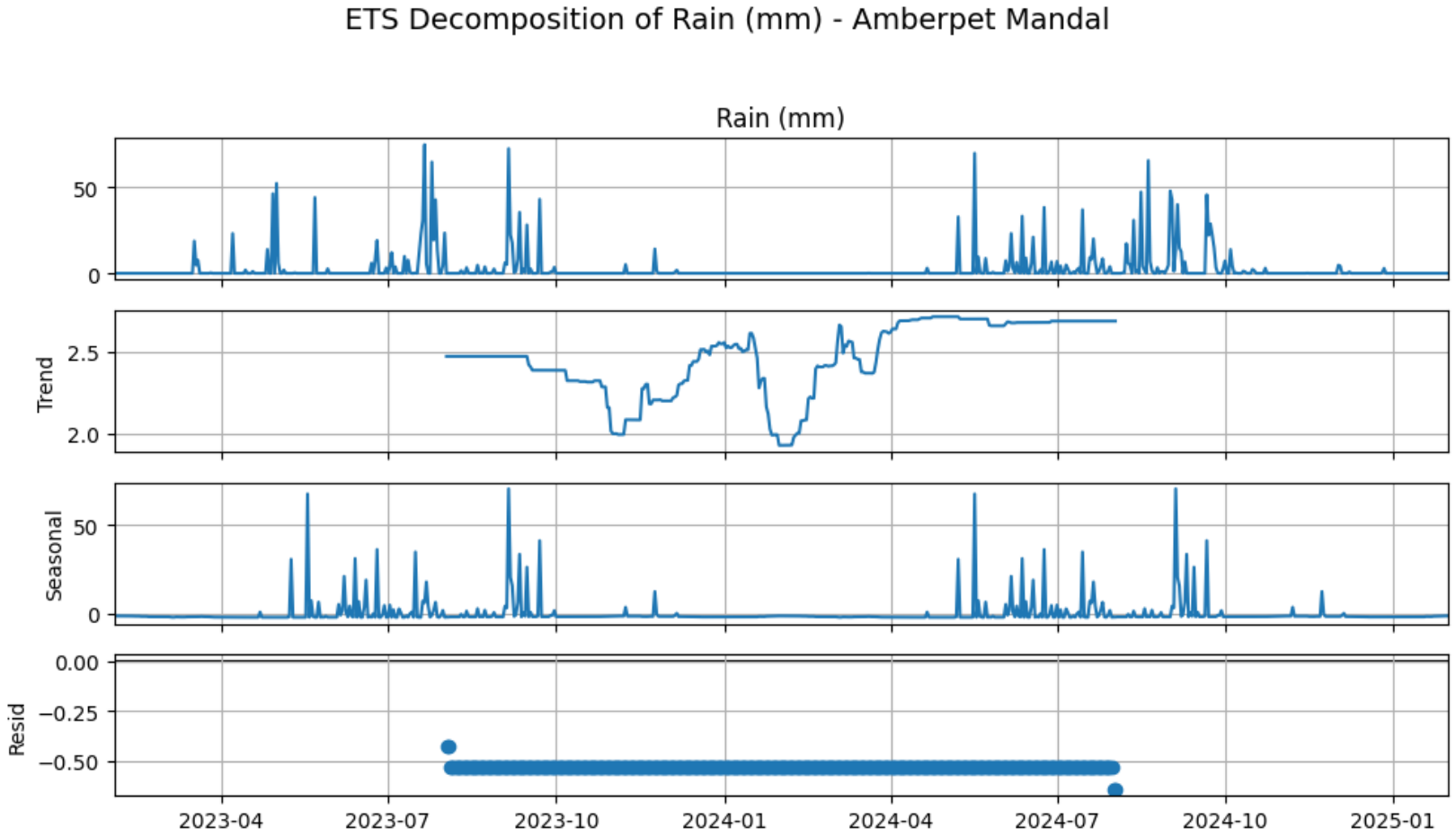
- Select & create data frame for Hyderabad district from larger dataset by grouping data by district
- Select & create data frame for Amberpet Mandal data from Hyderabad district data frame by grouping data by mandals
- District & mandal columns dropped – not necessary for forecasting
- Check for duplicate & missing data
- Convert to time series data

```
 'pandas.core.frame.DataFrame'>
meIndex: 731 entries, 2023-02-01 to 2
olumns (total 7 columns):
olumn                  Non-Null Count
-----                  --------------
ain (mm)               731 non-null
in Temp (°C)           731 non-null
ax Temp (°C)           731 non-null
in Humidity (%)        731 non-null
ax Humidity (%)        731 non-null
in Wind Speed (Kmph)   731 non-null
ax Wind Speed (Kmph)   731 non-null
: float64(7)
 usage: 45.7 KB
```

EDA Insights: distribution, spread & outliers

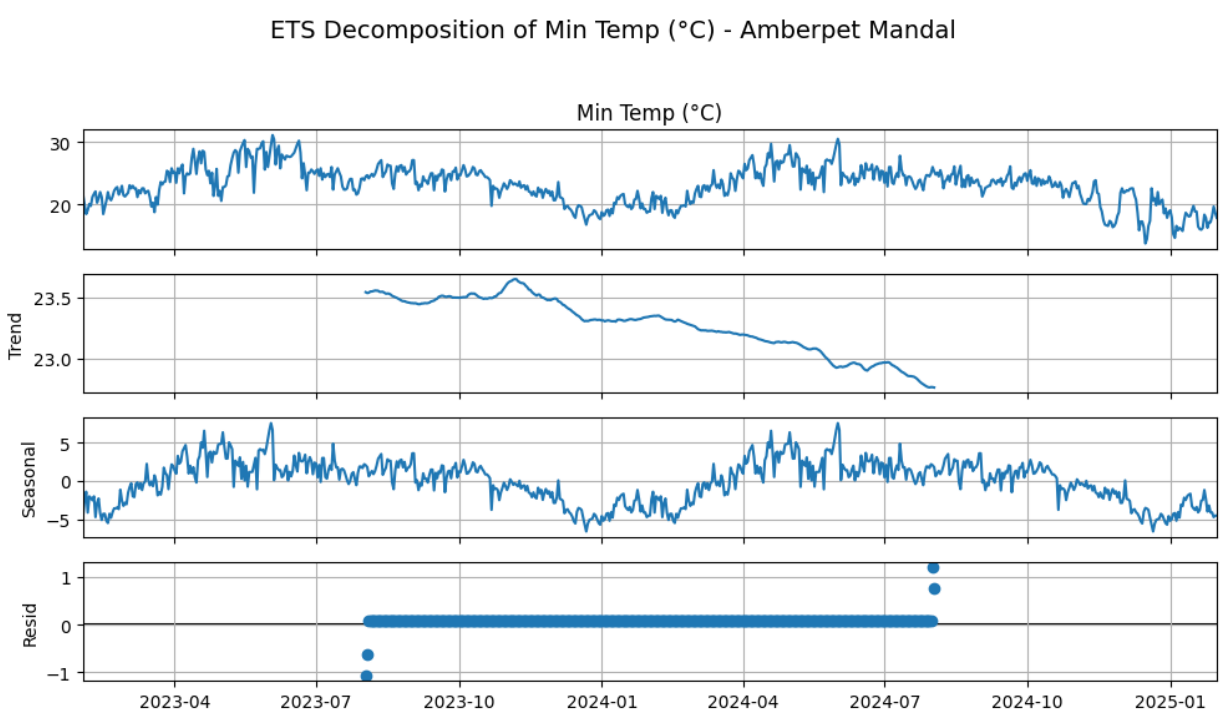
Rain (mm)	• Extreme right skew, most days are dry, rare heavy rainfall events, a lot of zero values, a lot of outliers • Log transformation before model development
Min Temp (º C)	• Slight left skew, but reasonable
Max Temp (º C)	• Nearly perfect normal distribution, outliers present
Min Humidity (%)	• Roughly normal distribution with slight right skew
Max Humidity (%)	• Left-skewed (clustered near 100%). Many days hit 100% max humidity (morning dew/fog)
Min Wind Speed (Kmph)	• A lot of values are 0 kmph, a lot of outliers. • Log transformation before model development
Max Wind Speed (Kmph)	• Most days have low wind (5-15 kmph). Rare extreme wind events (outliers up to 70 kmph), outliers present • Log transformation before model development

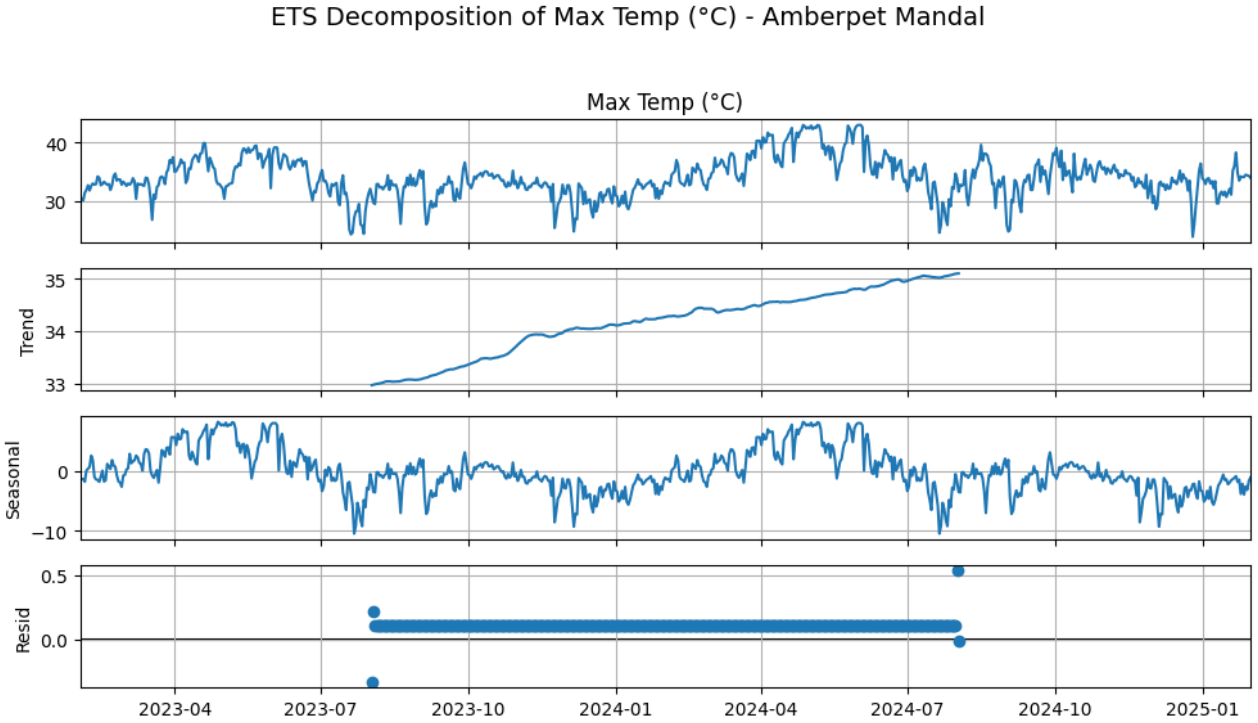
ETS decomposition: Rain



- **Trend -** slight dip in rainfall levels during late 2023 & early 2024, followed by stabilization - changes in overall rainfall patterns over the 2-year period that could be a temporary shift
- **Seasonal -** strong seasonal component present – highlights predictable nature of rainfall cycles, account for monsoon seasonality
- **Residual -** small & scattered, indicating most of variability in rainfall is explained by the trend & seasonal components

ETS Decomposition - Temp





Trend

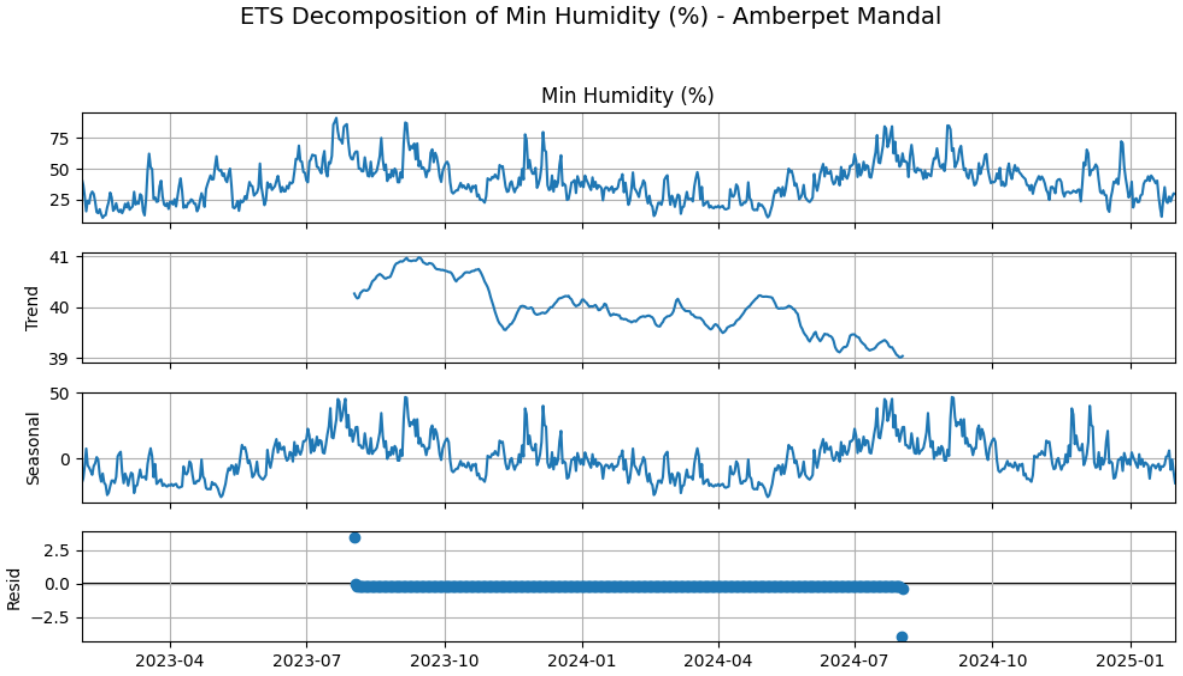
- gradual ↓ in min temps from mid-2023 to early 2025 - due to changes in weather patterns or other environmental factors. Declining trend in min temps can help predict long-term changes in the region's climate.
- gradual ↑ in max temps from mid-2023 to early 2025, could indicate warming trends or seasonal shifts in region's climate

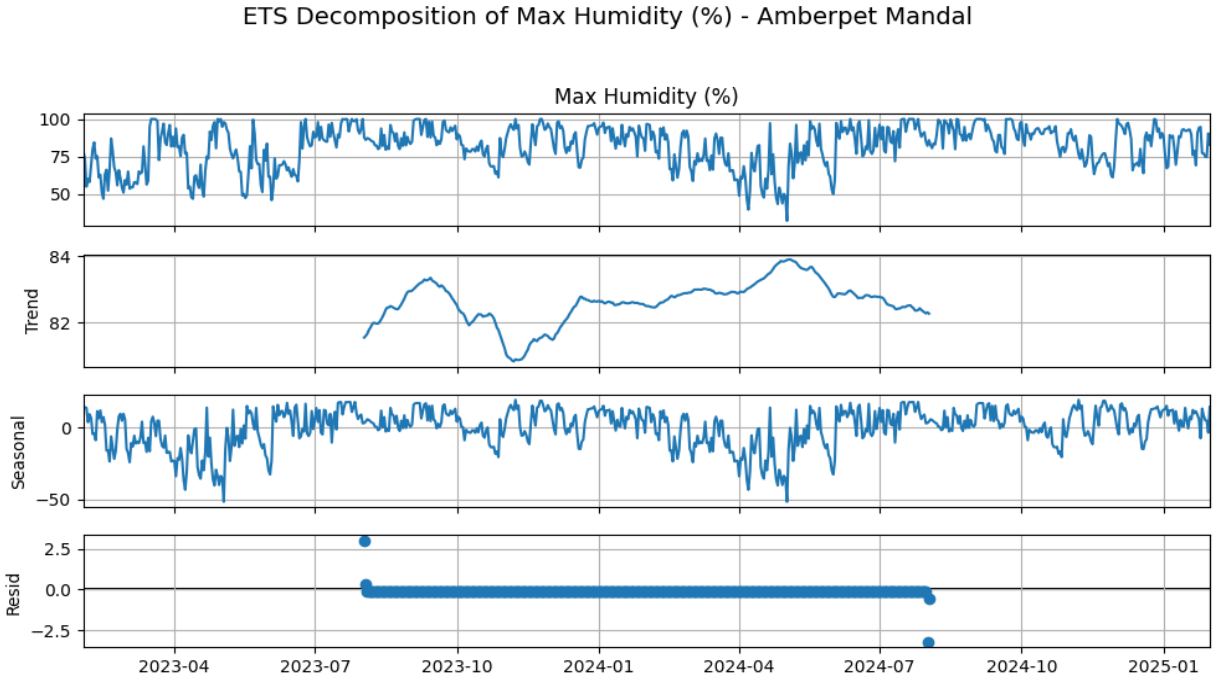
Seasonal

- Repeating seasonal patterns. Peaks & troughs visible, reflecting higher min & max temps during summer months & lower min & max temps during winter months. Predictable nature of temp cycles. Strong seasonal component - account in model

Residual - relatively small, indicating most of variability in min & max temp is explained by the trend & seasonal components

ETS Decomposition - Humidity





Trend

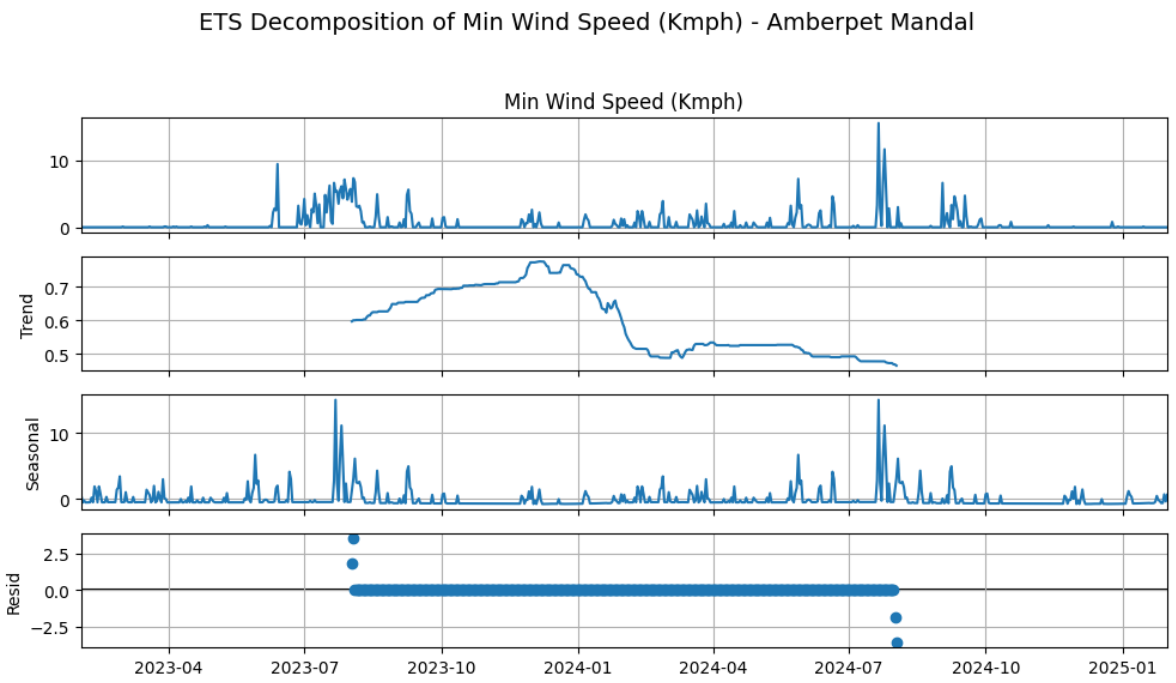
- Initial ↑ in min humidity in mid-2023, followed by gradual decline from late 2023 to early 2025
- Initial ↓ in max humidity in mid-2023, followed by a gradual incline from late 2023 to early 2025
- Could indicate changes in weather patterns or environmental factors affecting humidity levels - help predict long-term changes in the region's climate.

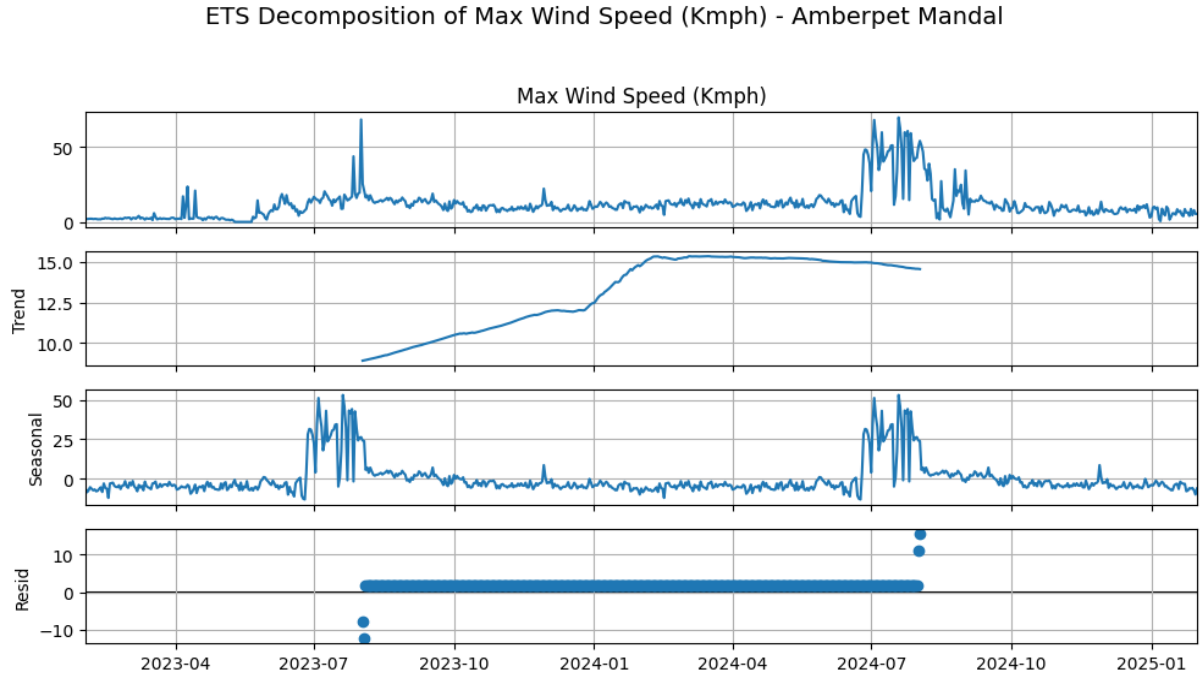
Seasonal

- Repeating seasonal patterns in min & max humidity – highlights predictable nature of humidity cycles, account for seasonality in model
- Peaks & troughs are visible – reflect higher min humidity & max humidity during monsoon months (June–September), lower min & max humidity during dry months (October–May)

Residual - relatively small, indicating most of variability is explained by the trend & seasonal components

ETS Decomposition – Wind Speeds





Trend

- Initial ↑ in min wind speed from around July 2023 to the end of 2023, followed by gradual decline from late 2023 to around March 2024. Then the min weed speed doesn't change much.
- Max wind speed rises steadily from mid 2023, peaks until around March 2024, and then slightly declines. This is indicative of the monsoon season.

Seasonal

- Repeating seasonal patterns in min & max wind speed. Peaks visible, reflecting higher min wind speeds during certain months (likely monsoon or pre-monsoon periods) & lower speeds during calmer months (summer & winter). For max wind speed, pronounced peaks during the same periods as observed spikes, confirms strong seasonality—likely corresponding to monsoon season, hence predictable nature of wind speed cycles, account in models

Residual - relatively small, indicating most of variability is explained by trend & seasonal components. Max wind speed - most residuals are close to zero, but there are occasional outliers, indicating some extreme wind events not explained by regular patterns.

Data Pre-processing & Feature Engineering

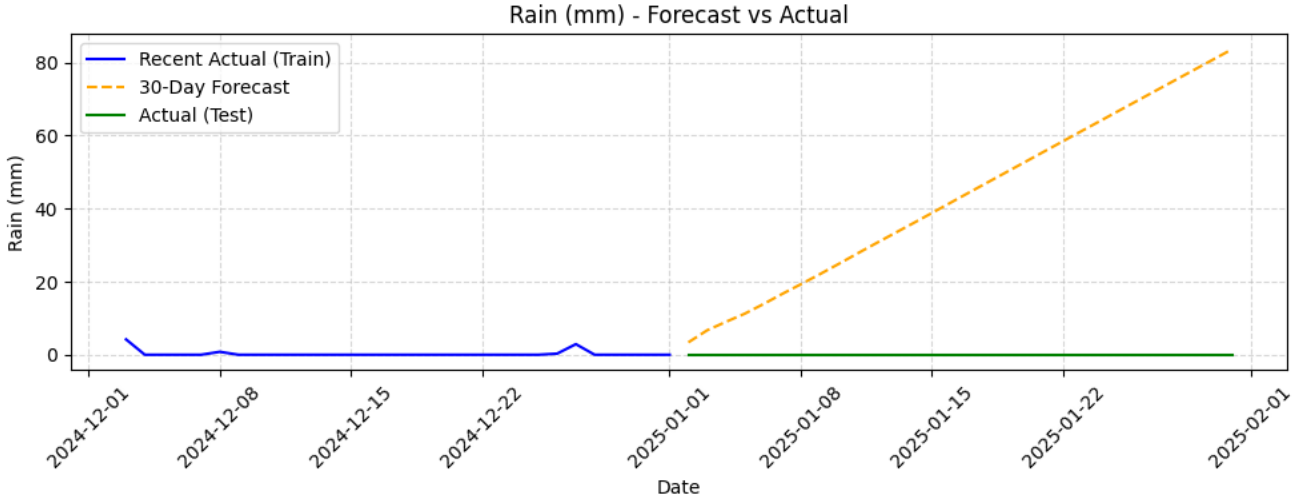
- ACF & PACF plots visualised
- Log transformation of highly skewed variables (Rain, max wind speed & min wind speed)
- Robust scaling of variables with many outliers (Rain, min wind speed, max wind speed & max temp)
- ADF test done – stationarity checked
- 3 variables differenced as they were non-stationary
  - Min Temp (°C): ADF=-1.360097, p=0.601249
  - Min Humidity (%): ADF=-2.734, p=0.06828
  - Max Wind Speed (Kmph) ADF= -2.665582 p = 0.080186

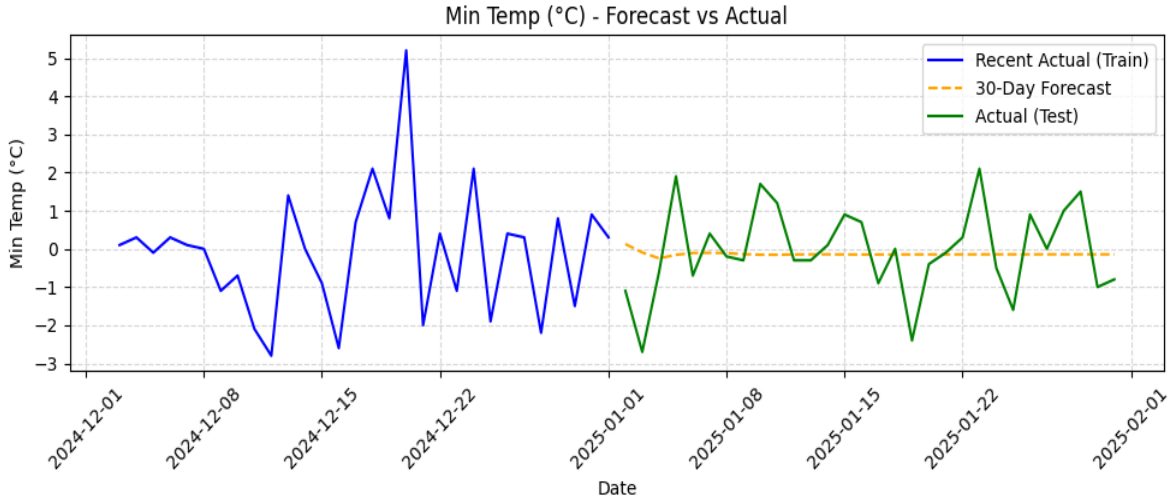
Model

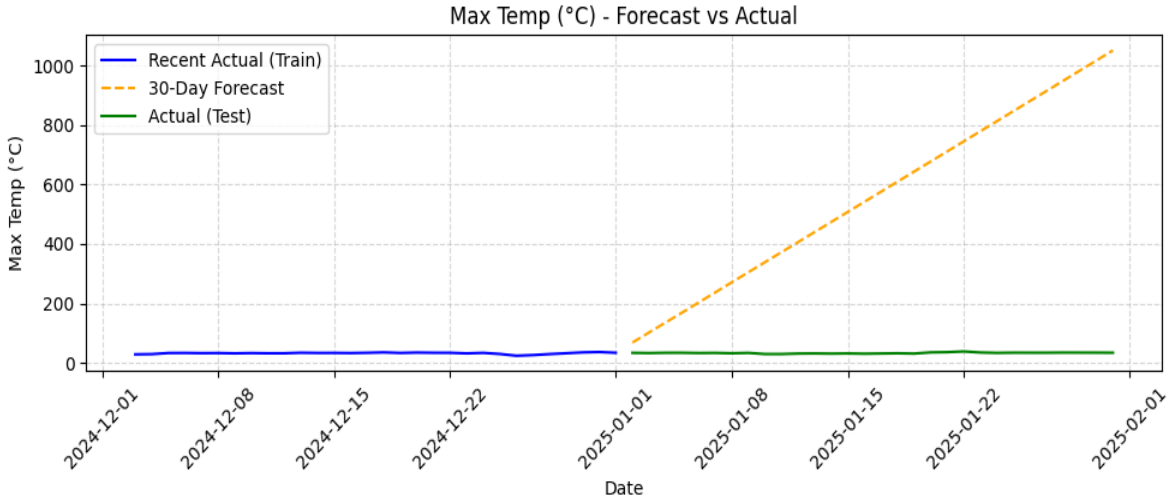
- Vector Autoregression (VAR) (4) model
  - Used to train dataset & predict 30 days weather forecast
  - Optimal Lag order: 4 - based on lowest score of AIC

- All variables that were log transformed, robust scaled & differenced were inverted after prediction

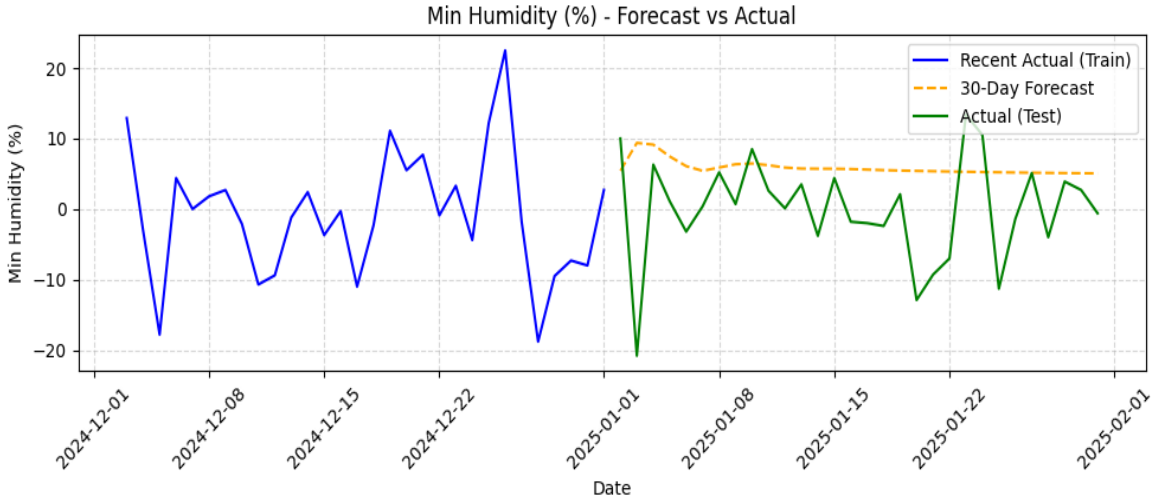
Visualization of Forecast vs. Actual

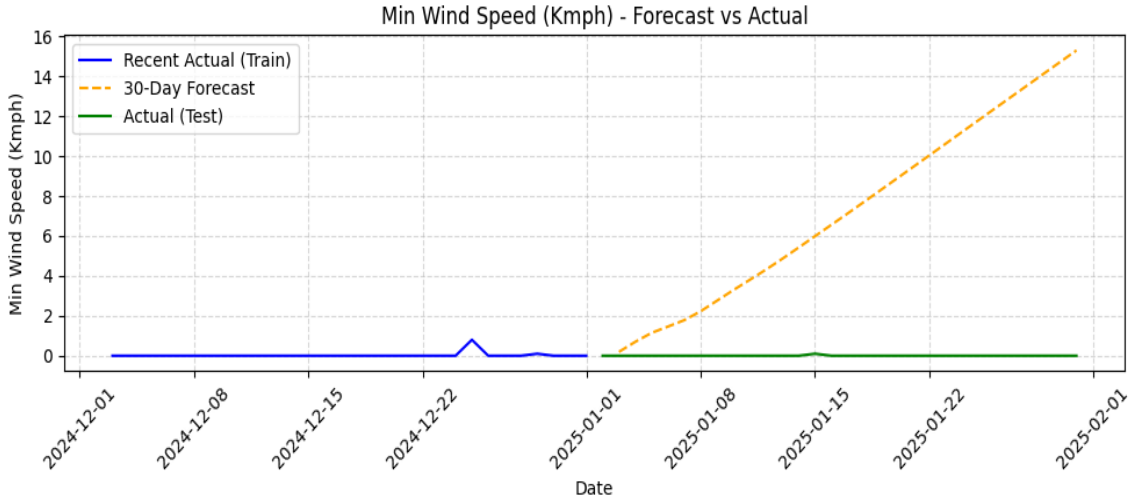


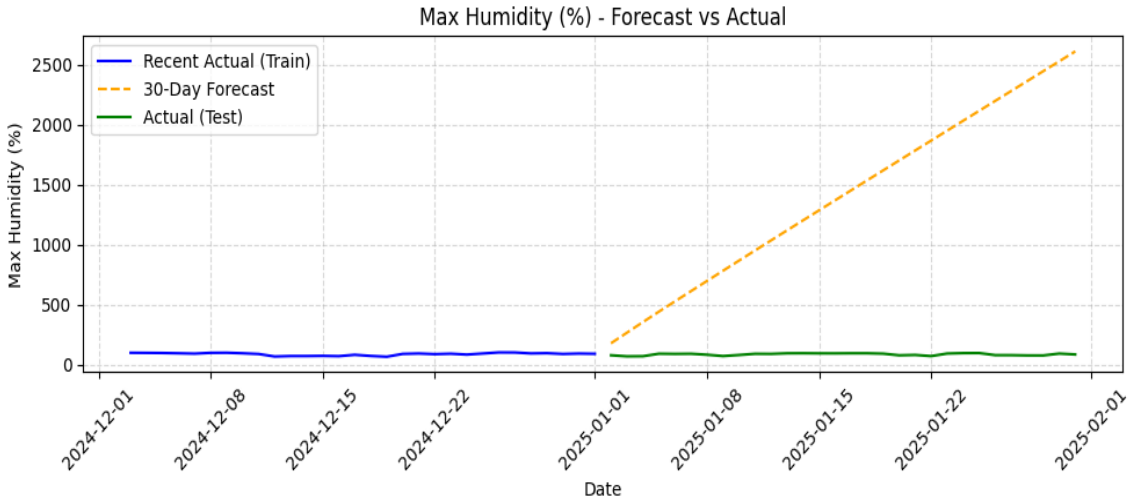


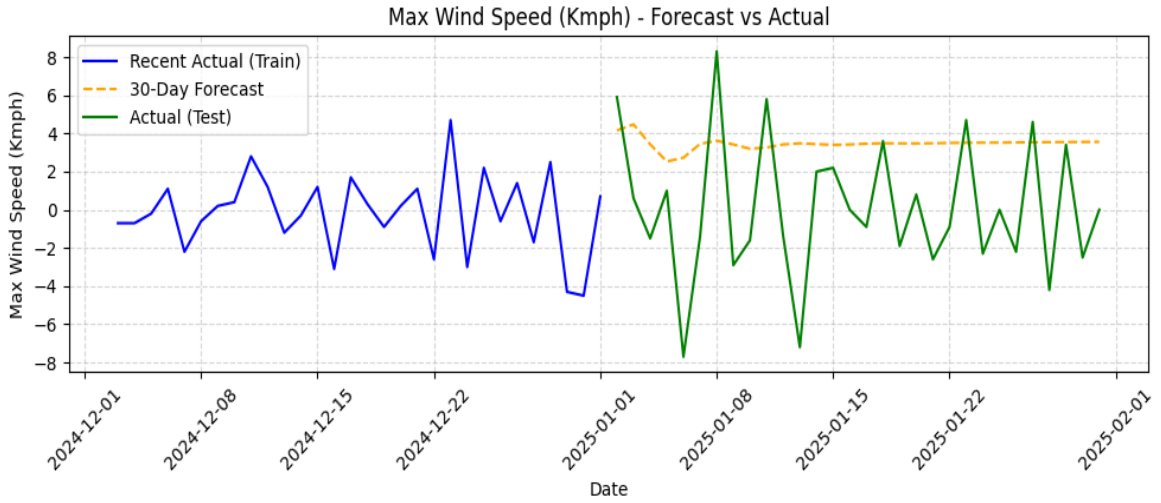


Visualization of Forecast vs. Actual









Model Evaluation

Metrics per Model	Comments
Min Temp (°C) MSE = 1.3324, RMSE = 1.1543, MAE = 0.89	• Excellent performance. • Errors are low
Max Wind Speed (Kmph) MSE = 24.1158, RMSE = 4.9108, MAE = 4.17	• Very good performance. • Errors are low
Min Humidity (%) MSE = 89.7792, RMSE = 9.4752, MAE = 7.19	• Reasonable error range. Could be improved with better lag structure or interaction terms.
Min Wind Speed (Kmph) MSE = 75.0202, RMSE = 8.6614, MAE = 7.33	• Acceptable error range
Rain (mm) MSE = 2442.6890, RMSE = 49.4236, MAE = 43.19	• Relatively high errors, suggests model struggles with rain variability—likely due to its sporadic, non-Gaussian nature
Max Temp (°C) MSE = 362283.1258, RMSE = 601.8996, MAE = 525.93	• Extremely large errors
Max Humidity (%) MSE = 2266311.0979, RMSE = 1505.4272, MAE = 1320.61	• Extremely large errors

Next Steps …

- Determine how best to handle variables with outliers, skewed variables & many zero values
- Create & incorporate a causality-informed feature mask into the VAR model
- Validate model assumptions using residual diagnostics
- Understand how shocks propagate by Impulse Response Analysis
- Extend to VARMAX with exogenous regressors (e.g., seasonal indicators, Fourier terms)
- Do univariate time series forecasting for each variable
- Compare with deep learning models (e.g., XGBoost, LSTM) using same features

References/ Attributes

- https://www.britannica.com/place/Telengana
- Hugging Face website - https://huggingface.co/datasets/ron-the-code/Telangana_time_series_2023-2025