Global Weather Forecasting Analysis & Prediction

TEMPERATURE PREDICTION MODEL USING TIME SERIES ANALYSIS

WILSON ZHENG

PM Accelerator's Mission

By making industry-leading tools and education available to individuals from all backgrounds, **we level the playing field for future PM leaders.** This is the PM Accelerator motto, as we grant aspiring and experienced PMs what they need most – Access. We introduce you to industry leaders, **surround you with the right PM ecosystem**, and discover the new world of AI product management skills.

Predicting temperature patterns using time series analysis

Dataset: Global Weather Repository (~58,270 entries from May 2024 to March 2025)

Focus: Temperature forecasting for the United States

Approach: SARIMA time series modeling with exogenous variables

Methodology

Data Cleaning & Preprocessing

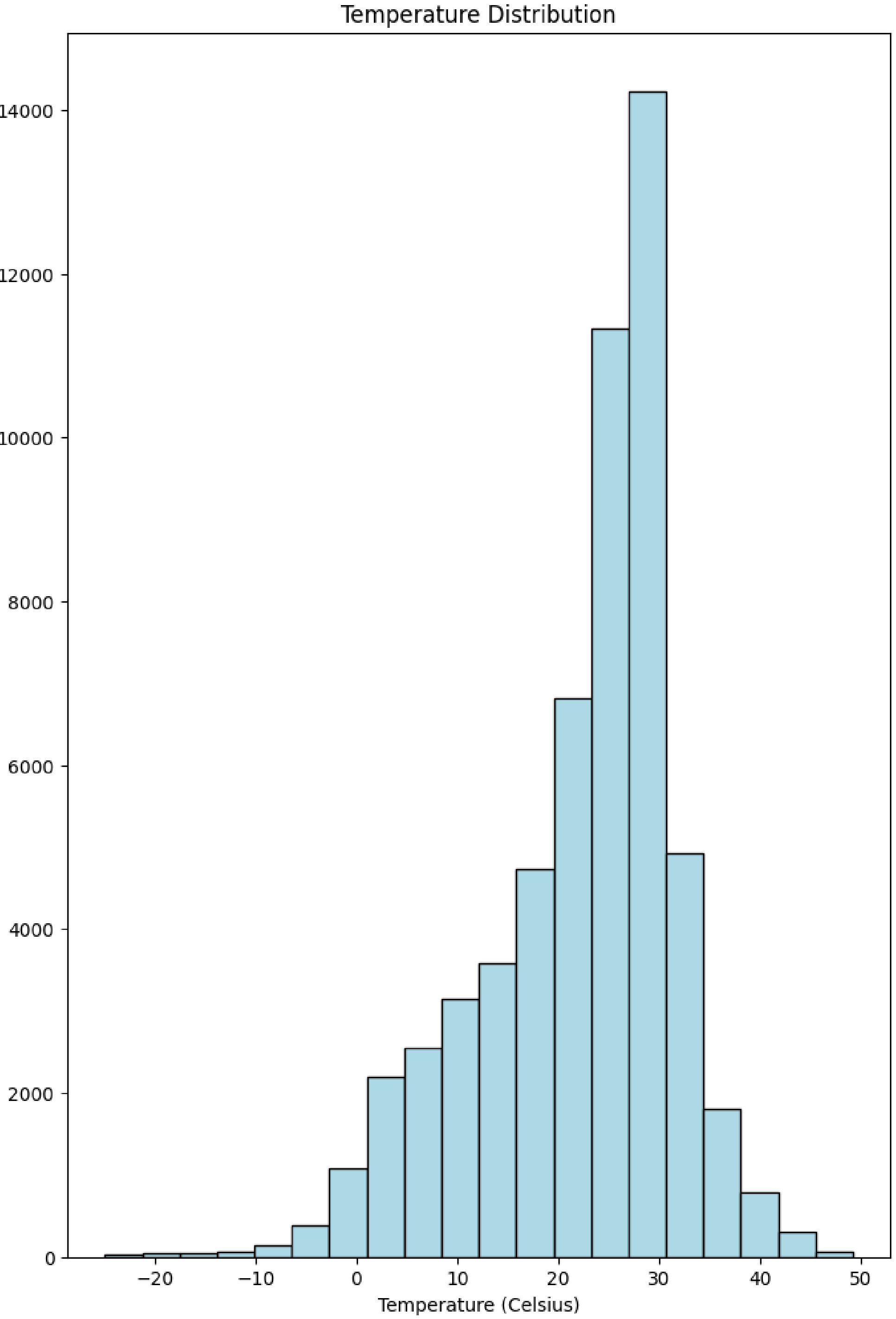
- Libraries: Pandas, NumPy
- Handling missing values (found none in the dataset)
- Outlier management using IQR method
- Feature selection and redundancy removal

Exploratory Data Analysis

- Libraries: seaborn, matplotlib
- Uncovering distributions, trends, and correlations in weather data
- Visualizing temperature and precipitation patterns

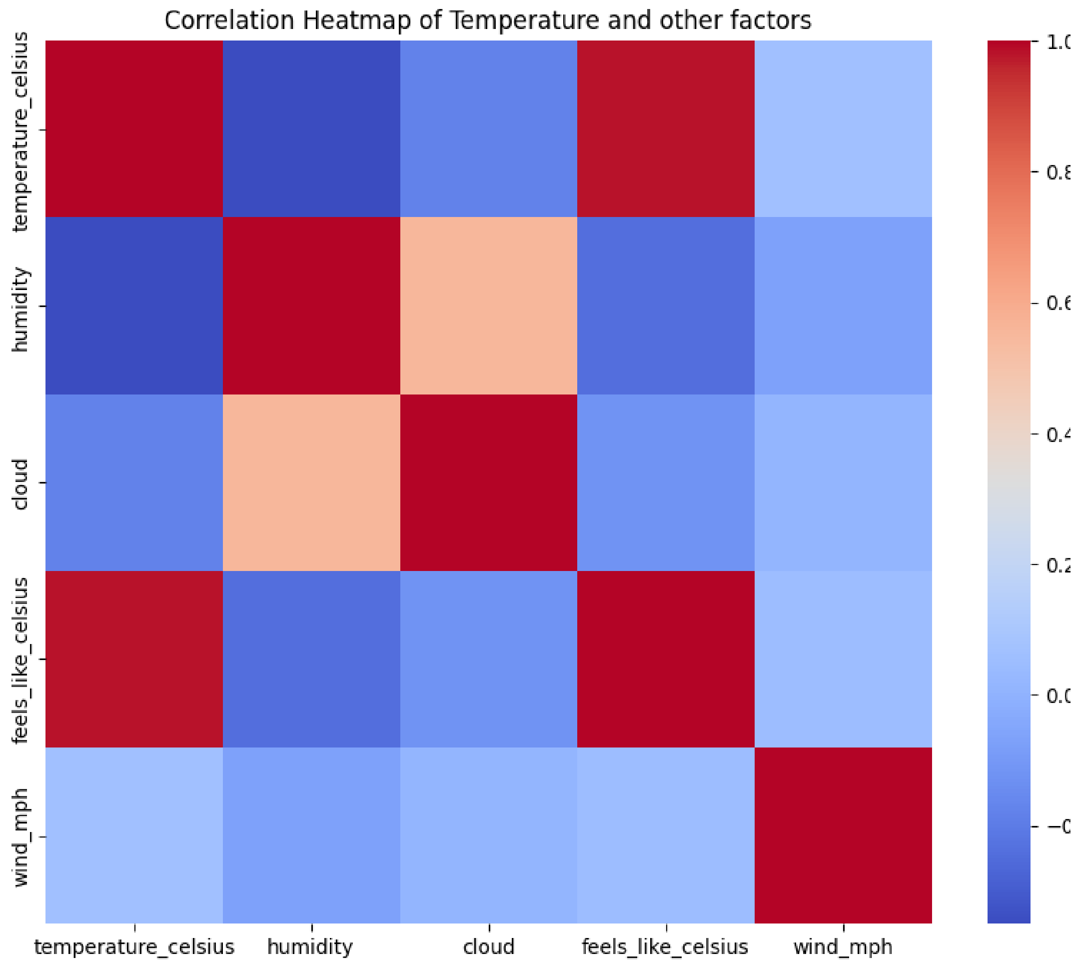
Model Building

- Libraries: pmdarima, statsmodels
- Time series forecasting with SARIMA
- Auto_arima for automated parameter selection
- Performance evaluation with RMSE



Exploratory Data Analysis

TEMPERATURE DISTRIBUTION HISTOGRAM WITH KEY FINDING: MOST FREQUENT TEMPERATURES BETWEEN 25-30ºC



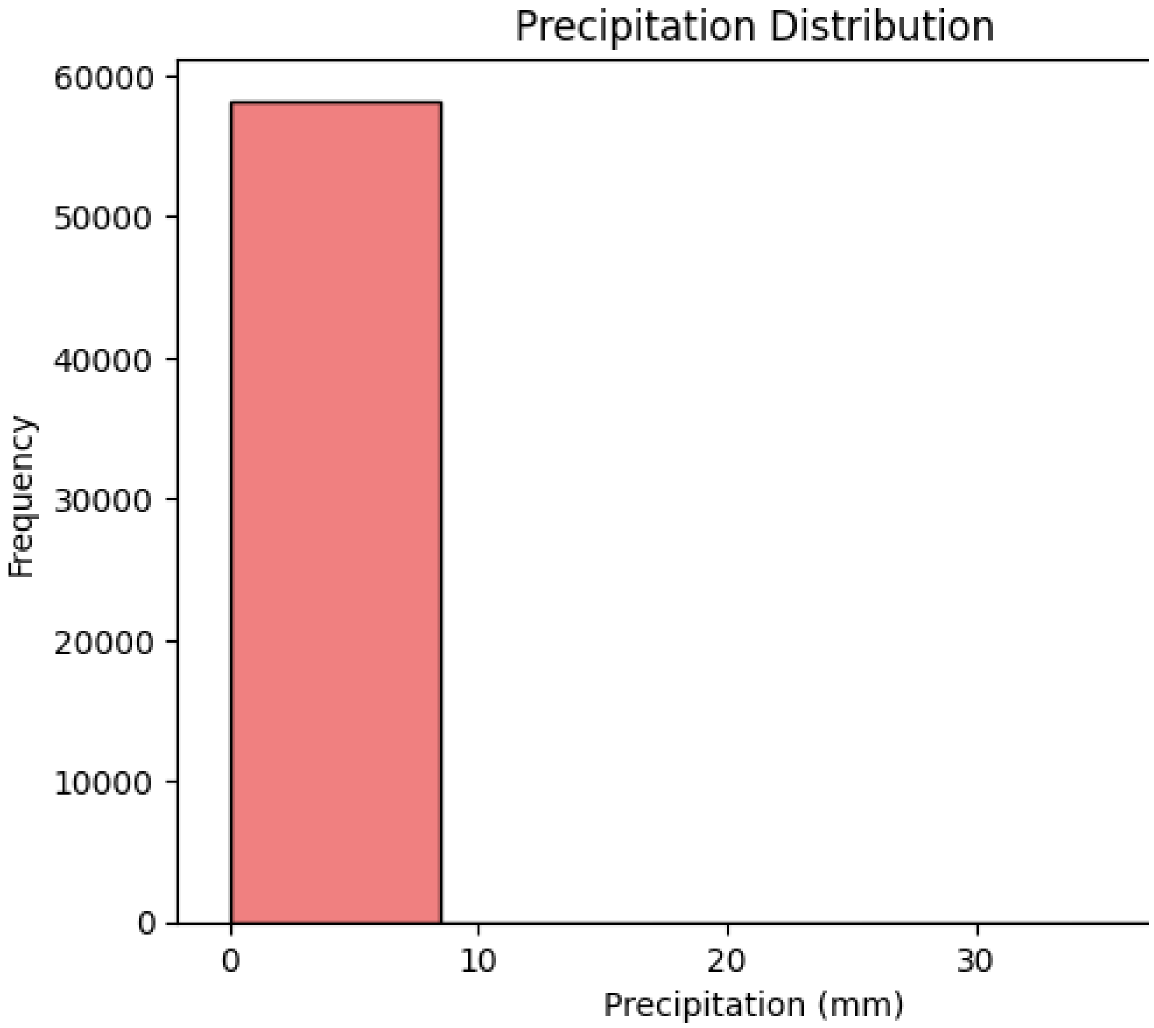
Correlation heatmap highlighting key relationships

Correlation heatmap highlighting key relationships:

Strong correlation between actual and feels-like temperature

Weak negative correlation between humidity and temperature

Minimal correlation between other variables



Precipitation insight

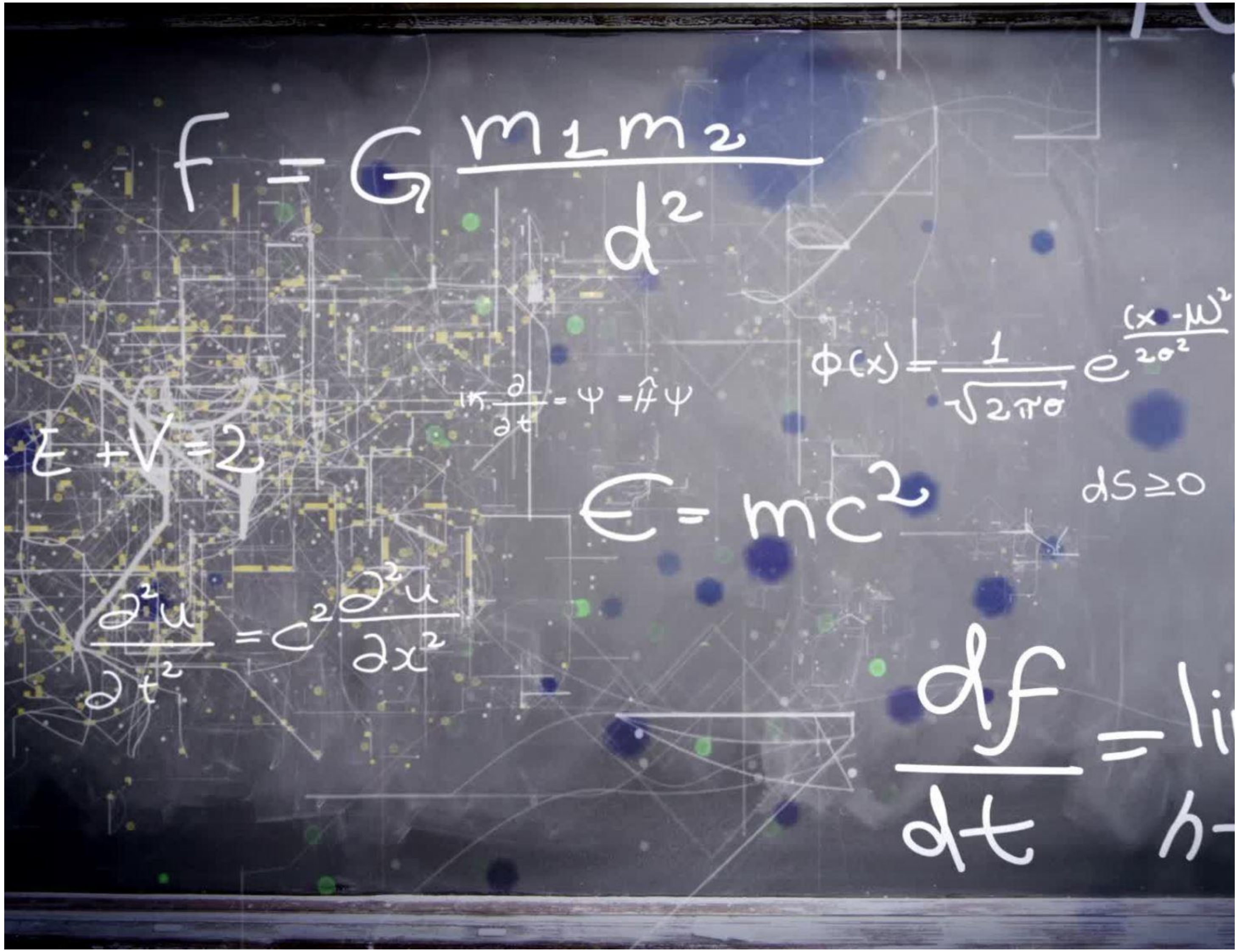
VERY LOW PRECIPITATION FREQUENCY (0.05% OF MEASUREMENTS)

Data Preprocessing

Country selection: United States of America

Preprocessing pipeline:

- DateTime conversion and indexing using 'last_updated' column
- Feature reduction from 41 to 6 key features
- Outlier removal using IQR method
- Daily resampling and forward filling
- Final dataset: 301 daily observations ready for modeling



Model Training

Approach: SARIMA

Model selection process:

- Auto_arima library for optimal parameter identification
- First-order differencing for stationarity
- 80/20 train-test split

Key variables:

- Target: temperature_celsius
- Features: wind_kph, wind_degree, pressure_mb, humidity, gust_kph

Results & Evaluation

Forecast vs. actual temperature visualization from the test set

Performance metric: RMSE = 4.04°C

Residual analysis:

- Random distribution indicating good model fit
- Near-normal distribution with slight left skew
- Interpretation: Model captures general trends but struggles with some fluctuations

Insights & Next Steps

Key insights:

- Weather patterns show complex dependencies
- Model achieves moderate prediction accuracy

Limitations:

- Limited dataset size after filtering (~300 entries)
- 10-month timeframe insufficient for full seasonal analysis

Future work:

- Expand data collection timeframe
- Incorporate additional variables
- Explore region-specific models