Dengue Prediction Model Training Summary

2. Cleaned missing values using forward fill (ffill).

3. Encoded the 'city' column to numerical values: 'sj' 0, 'iq' 1.

4. Created new features:

- Lag features (lag_1_total_cases, lag_2_total_cases)
- Rolling statistics (rolling_mean_3, rolling_std_3)
- Temperature difference (temp_diff = max_temp - min_temp)

5. Performed Exploratory Data Analysis (EDA):

- Missing value visualization
- Boxplots for city-wise and weekly dengue cases
- Correlation heatmap for feature selection

6. Feature selection using Random Forest: Top 10 most important features selected.

7. Trained and evaluated different regression models (separately for each city):

a) Random Forest Regressor

- MAE (San Juan): 7.99
- MAE (Iquitos): 4.44

b) Gradient Boosting Regressor

- MAE (San Juan): 8.88
- MAE (Iquitos): ~5+ (less accurate)

c) LightGBM Regressor

- MAE (San Juan): ~7.3 (approx)
- MAE (Iquitos): ~4.0+

d) XGBoost Regressor (Best Model)

- MAE (San Juan): 5.29
- MAE (Iquitos): 1.68

8. Performed final prediction using trained XGBoost models on test data.

9. Predictions were clipped to remove negative case values.

10. Final predictions exported to 'dengue_predictions.csv'.