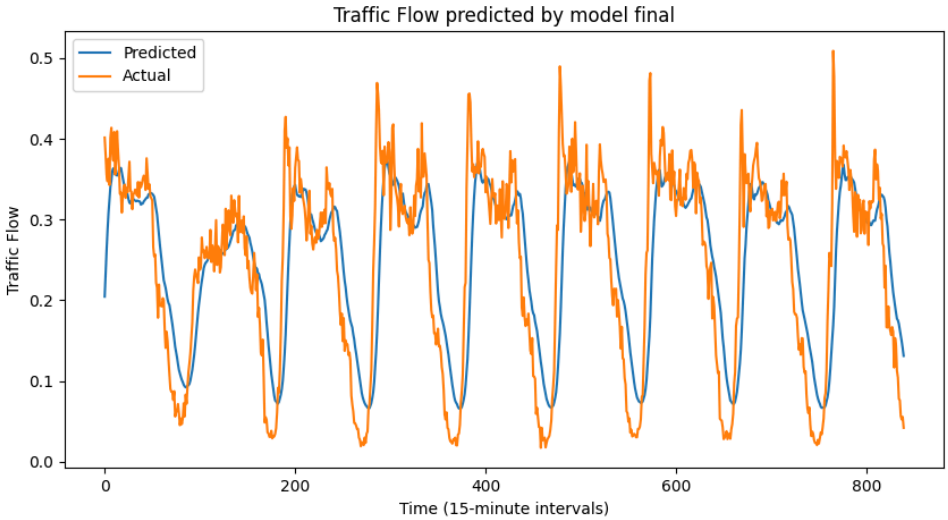
Basic Traffic Flow Prediction

1. **Feature Engineering:**
    a. **Create meaningful features based on your EDA, such as time-based features (e.g., hour of the day, day of the week) and rolling averages.**
    We created the following features
        i. Rolling average
        ii. Weighted average
        iii. Reverse weighted average
        iv. Time of Day on a linear scale
        v. Day of the week on a linear scale
    b. **Explain why each feature was selected and how it is expected to improve model performance.**
        i. Rolling average - We know that traffic is correlated with its previous values. But we want to reduce the number of features to avoid overfitting
        ii. Weighted average - Weighted average might be a better metric as data is highly correlated at higher lags.
        iii. Reverse weighted average - For comparison
        iv. Time of day normalized as one feature - Reduce number of features from 24 to 1 to avoid overfitting
        v. Day of week normalization - Reduce number of features from 7 to 1 to avoid overfitting
2. **Model Implementation and Training:**
    a. **Implement and train the chosen model (e.g., Linear Regression, ARIMA):**
        i. Final model contains a combination of the above features. Time and Average. We used **linear regression** model. Non linear models did not offer a very high prediction accuracy during testing indicating overfitting.
        ii. Training accuracy of final model
            1. Training MAE: 0.02237996446949679
            2. Training RMSE: 0.03221400362827798
3. **Model Performance Evaluation:**
    a. **Evaluate the model using the following metrics:**
        i. Mean Absolute Error (MAE): 0.05453453750280212
        ii. Root Mean Squared Error (RMSE): 0.07580269402221909

b. Include a plot comparing actual vs. predicted values and discuss the results.



- The prediction accuracy is to the expected level.
- Model is able to predict accurately
- Limitations: Initial values are not predicted with as high accuracy as the later values. In real life traffic prediction will be continuous so this is acceptable