Traffic Forecasting — EDA & Modeling Summary

Author: Karthik Date: Date: October 05, 2025

Prepared for: Exploratory Data Analysis & Data Preparation Component

Data & Feature Engineering

- Source: Traffic sensor counts aggregated to hourly intervals per junction
- Key features: hour, day_of_week, is_weekend, hour_sin/hour_cos, lag features, rolling means, is_event
- Optional: weather (temp, precipitation, wind), event metadata merged by hour

Modeling Approach

- Baseline: Persistence (previous hour)
- Tree-based: RandomForest, LightGBM (feature importance + robustness)
- Time-series: ARIMA/SARIMA and sequence models (LSTM)
- Use time-based splits and TimeSeriesSplit for validation; evaluate MAE, RMSE, R^2

Evaluation & Diagnostics

- Metrics: MAE (primary), RMSE, R^2; per-junction breakdown for diagnosis
- Diagnostics: residual plots, error distribution, worst-performing junctions analysis
- Hyperparameter tuning: RandomizedSearch with TimeSeriesSplit; prefer LightGBM for speed

Deliverables & Next Steps

- Deliverables: Research PDF, Integrated dataset CSV, Modeling notebook (.ipynb), Models (.joblib/.txt/.h5), Summary presentation
- Next: integrate weather & event data, run time-series CV, produce per-junction error report, train LightGBM baseline
- Optional: deployment pipeline and dashboard