■ Crashspot Research – Week 5 Report

Focus: Model Preparation + Baseline Models

Objectives

• Engineer numeric + categorical features
• Define binary target variable
• Perform train/test split
• Train baseline ML models (Logistic Regression, Random Forest)

Workflow

→ Load Monroe dataset (fallback Louisiana if missing)
→ Engineer features: hour, month, weekday, weekend/night flags, counts
→ Define target variable: target_multiveh
→ Split into 80/20 train-test sets
→ Train Logistic Regression and Random Forest models
→ Save outputs (figures + CSV + notebook)

Outputs

■ Confusion matrices → week5_model_eval.png
■ Random Forest feature importances → week5_rf_feature_importance.png
■■ Features CSV → week5_features.csv
■ Notebook → Crashspot_Week5_Starter.ipynb

Insights

✔ Both Logistic Regression and Random Forest achieved 100% accuracy and F1-score on this dataset.
✔ Top predictive features from Random Forest: ve_total, persons, month, hour.
✔ Dataset is very small (60 records), so results risk overfitting.
✔ Next steps: add more contextual features (weather, road type) and apply cross-validation.