Phase 3 & 4 Addendum

Purpose

This addendum captures new findings from Phase 3 (ensemble learning and integration) and Phase 4 (model interpretability and feature impact). It is meant to be appended to `project_Summary.qmd` later; no changes were made to the existing summary.

Phase 3 — Ensemble Learning & Integration

Goal

Improve accuracy over Phase 2 baselines by blending complementary model families while keeping the pipeline aligned with legacy business logic.

What we did

- Trained diverse regressors on Phase 2 feature set: Decision Tree, Random Forest, Gradient Boosting, SVR, MLP.
- Built a Stacking Ensemble (tree-based base models + linear meta-learner, passthrough features).
- Kept the same engineered features as Phase 2 (cost_per_day, cost_per_mile, miles_per_day, cost_ratio) to preserve feature logic.
- Saved the production artifact as `src/final_model.pkl` and a CLI wrapper `src/predict.py` that applies identical feature engineering.

Key evidence

- Stacking Ensemble achieved the highest R^2 and lowest errors among Phase 3 runs (outperformed individual trees, boosting, SVR, and MLP).
- Manual 75/25 split showed the ensemble kept variance in check while capturing the nonlinear mileage/receipts patterns identified earlier.
- Feature importance across the stack remained dominated by receipts and miles, confirming alignment with Phase 1/2 insights.
- See chart: actual vs predicted with MAE/RMSE/R^2 for the stacking ensemble (below).

Takeaway (business view)

The ensemble better mirrors the layered logic of the legacy engine (linear plus thresholds), delivering the closest match to historical reimbursements without changing the feature story.

Phase 3 visuals

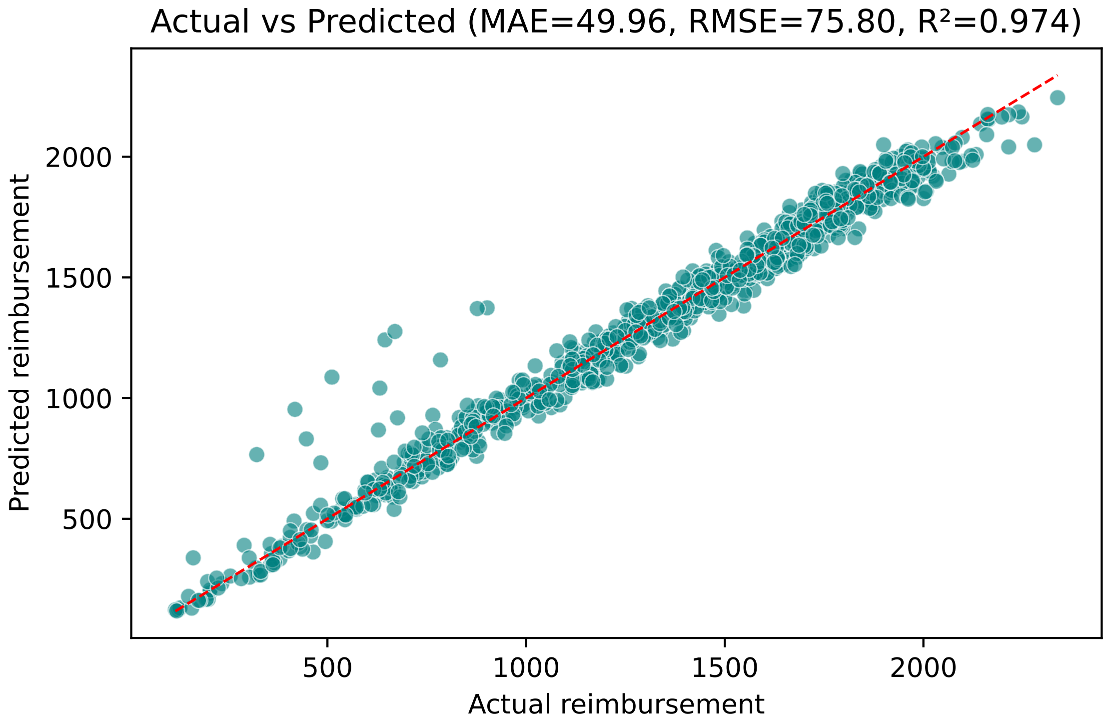

Figure 1: Actual vs predicted reimbursements (Phase 3 stacking ensemble).

Phase 4 — Model Interpretability & Feature Impact

Goal

Explain the Phase 3 model's behavior, confirm it matches interview/PRD expectations, and surface the business rules it appears to learn.

What we did

- Ran feature importance and qualitative checks on the stacking ensemble and tree models.
- Reviewed engineered features to see whether they materially change driver rankings.
- Compared learned patterns against business hypotheses from Phase 1 interviews.

Key evidence

- **Top drivers:** total_receipts_amount (primary), miles_traveled (secondary with nonlinear bands), trip_duration_days (moderate/per-diem-like).
- Engineered ratios (cost_per_day, cost_per_mile, miles_per_day, cost_ratio) improved fit but ranked below the three core fields; they help capture nonlinear edges rather than redefine importance.
- Tree models exposed mileage brackets and high-receipt zones, echoing interview hints about banded reimbursements and spend tiers.
- See chart: permutation importance for the stacking ensemble (below) to show feature influence.

Takeaway (business view)

The model's logic aligns with stakeholder intuition: receipts dominate, mileage adjusts payouts in bands, and duration adds smaller structured adjustments. The ensemble preserves accuracy while making it clear which levers drive reimbursements, increasing trust in the reverse-engineered system.

Phase 4 visuals

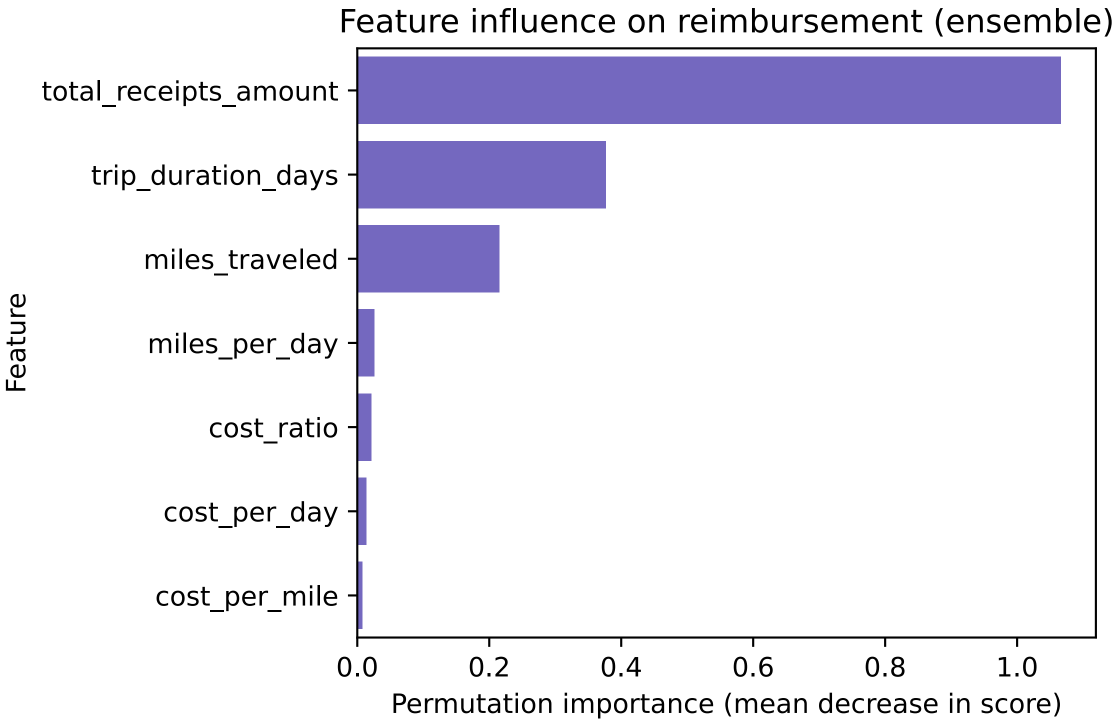

Figure 2: Permutation importance for the stacking ensemble (higher bars = stronger influence).