Feature Representation and Fairness Summary

Modeling Scenarios

We created two modeling setups to test how different feature representations affect model performance, interpretability, and fairness when predicting property crime types (Theft vs Break-in) using LAPD data.

- Baseline Representation:
  Used the cleaned dataset from Phase 1 with raw variables such as Vict Age, Vict Sex, Vict Descent, AREA, LAT, LON, Month, Hour, DayOfWeek, and crime/premise codes. These values were taken as-is, meaning the model saw time as just a number, and age as continuous rather than grouped.

- Improved Representation:
  Applied transformations inspired by our Feature Representation Audit (FRA):

  - Grouped TIME OCC into four categories (Morning, Afternoon, Evening, Night)

  - Binned Vict Age into ranges (<18, 18–30, 31–50, 51+)

  - One-hot encoded Vict Sex to handle missing and unknown cases fairly

  - One-hot encoded AREA to better capture neighborhood-level variation

These changes were made to help the models learn more interpretable patterns and reduce potential bias from raw numeric encodings.

Results

Across Logistic Regression, Decision Tree, and Random Forest models:

- Accuracy remained close to 70–72% for both setups, showing that transformations didn't drastically change overall performance.

- Recall for Break-ins (minority class) improved slightly after feature transformations, indicating the models learned to detect underrepresented crimes better.

- Interpretability increased because features like "time of day" and "age group" make model behavior more understandable.

- Fairness improved slightly since one-hot encoding of sex ensured all groups—including “Unknown”—were represented rather than dropped.

The Random Forest model continued to perform best overall, balancing recall and accuracy while producing clear feature importance results that revealed which victim and neighborhood characteristics mattered most.

Analysis

Feature representation directly influenced what the models learned and how we could interpret their outputs. Grouping and encoding simplified patterns that were previously hidden in raw numbers—for example, it became clearer that break-ins happen more often at night, or that certain age groups were more affected by thefts. These improvements make the results not only more accurate for underrepresented cases but also more ethically interpretable, since they avoid discarding or oversimplifying demographic data. We learned that representation drives meaning: how we structure and define variables shapes both the model’s learning process and the social implications of its predictions.

Next Steps

If given more time or data, we would:

- Apply cross-validation for more stable performance estimates.
- Test resampling or class weighting methods to reduce class imbalance.
- Explore additional contextual variables, like weekend vs weekday, lighting conditions, or population density, to better explain neighborhood-level risk.
- Include fairness metrics (e.g., demographic parity difference) to measure whether performance varies across gender or area.