Worker No-Show Prediction - Key Insights

Dataset and features:

- Input fields include pay, shift time, worker rating, previous no-shows, commute distance and a facilities flag.

- Derived features: shift category (morning/afternoon/night), commute_long (>5 km), high_prev_no_show (>=2), and pay buckets.

Main drivers of no-shows:

- Long commutes (commute_km > 5 km) increase the chance of no-show.

- Low-pay gigs and lack of basic facilities (food/water) correlate with higher no-show rates.

- Workers with prior no-shows are much more likely to miss future gigs.

Modeling summary:

- Tested Logistic Regression, Decision Tree, Random Forest and Gradient Boosting.

- Best model chosen by ROC AUC (recommended for imbalanced target).

- The production step saves the best model for deployment.

Recommended interventions:

- Flag bookings with high predicted no-show risk and apply targeted interventions:
  - Reminder SMS / call 1–2 hours before shift.
  - Small travel/top-up incentive for long-commute workers.
  - Automatic replacement suggestions when risk is high.

Next steps:

- Collect more labeled data and use time-based cross-validation.

- Add features from worker app behavior (last-active timestamp, accept/decline times).

- Monitor model drift once deployed.