USER CHURN PROJECT | Logistic Regression Baseline Model

Prepared for the Waze Leadership Team

waze

Overview

The Waze data team is building predictive models to boost growth by identifying users at risk of churn. This part of the analysis focused on developing and evaluating a baseline **logistic regression model** as a simple, interpretable benchmark for comparison with more advanced models in later phases.

This report summarizes Milestone 5 and its implications for future project development.

Project Status

Milestone 5 - Regression Modeling

Target Goal: Fit a logistic regression model to predict churn and evaluate baseline performance.

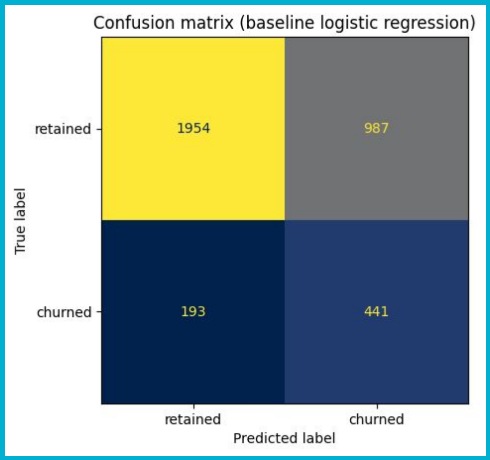
Methods:

- **Cleaned and engineered features**, including ratio metrics and the professional_driver flag.
- **Addressed multicollinearity** by removing highly correlated predictors.
- **Encoded categorical variables** (label_2, device_binary).
- **Standardized numeric features** and built a logistic regression pipeline.
- **Applied class weighting** to account for imbalance (82% retained, 18% churned).

Impact: Produced a baseline model capable of identifying churners at a much higher rate, directly **supporting Waze's business goal of churn prevention.**

Key Insights

- **Churn detection:** The baseline logistic regression model correctly identified **70% of churners**, giving the team strong coverage of at-risk users.
- **Trade-off in accuracy:** Overall accuracy was **67%**, reflecting the model's willingness to predict churn more often in order to avoid missing true churners.
- **Precision vs. recall:** When the model predicts churn, ~31% of those users actually churn. While this introduces false positives, it ensures the model captures the majority of true churners.
- **Balanced performance:** The F1 score (0.43) highlights a reasonable balance between precision and recall, providing a useful baseline for intervention strategies.
- **Model strength:** A **ROC AUC of 0.74** confirms the model is meaningfully distinguishing churners from retained users beyond random guessing.



Next Steps

➔ **Threshold tuning:** Adjust churn probability cutoff to explore different precision/recall trade-offs based on business needs.

➔ **Feature interpretation:** Examine model coefficients to highlight which behaviors are most predictive of churn.

➔ **Model comparisons:** Benchmark against more complex models to test for performance improvements.

➔ **Stakeholder alignment:** Collaborate with product and marketing to decide whether catching more churners (high recall) or reducing false alarms (higher precision) better supports Waze's retention strategy.