

User Churn Project | Regression Modeling Results

Prepared for: Waze Leadership Team


Prepared by: Nawaz Quraishi

Date: January 05, 2025

OVERVIEW


The Waze data team is working to reduce monthly user churn - defined as users uninstalling or stopping app usage to drive growth and retention. Using binomial logistic regression, the team identified key factors influencing churn and predicted at-risk users. This report highlights the model's performance, actionable insights, and recommendations, providing a clear roadmap for refining retention strategies and supporting impactful business decisions.

PROJECT STATUS

 **Objective:** Use available data to build and analyze a binomial logistic regression model.

 **Methodology:**

- Explored and prepared the data for analysis
- Checked for assumptions specific to logistic regression
- Built the binomial regression model
- Evaluated model performance

 **Value:** The logistic regression model provides a data-driven foundation to predict churn and understand user behavior. These results enable us to target at-risk users, enhance retention efforts, and align strategies with business objectives.

NEXT STEPS

→ **Add features** like engagement metrics to improve churn prediction.

→ **Test strategies** like daily streaks or notifications to boost retention.

→ **Address class imbalance** with oversampling, undersampling, or class weights.

→ **Explore alternative models** (e.g., Random Forest, Gradient Boosting) for better performance.

KEY INSIGHTS

- **Professional drivers show lower churn rates**, likely due to their frequent usage and reliance on the app for their work. This group represents a valuable, loyal user base, and their behaviors can provide insights to retain casual users.
- **Drivers with more activity days are far less likely to churn**, making engagement a critical factor in user retention. Increasing the number of active days per user can significantly lower churn rates and enhance app loyalty.
- **While the model achieves 83% accuracy, it struggles with identifying churners (6% recall).** This low recall indicates that the model fails to correctly identify most at-risk users, making it unreliable for predicting churn without further improvements.
- **Activity days is the most important predictor of churn**, showing a strong negative correlation. Users who are active on more days are far less likely to churn, underlining the importance of strategies that encourage consistent daily engagement.
- **Features like km_per_driving_day had minimal impact**, suggesting that distance-driven metrics are less relevant to predicting churn. Future models should prioritize features with stronger influence on user behavior, such as activity and engagement metrics.