User Churn Project | Regression Modeling Results

Prepared for: Waze Leadership Team

OVERVIEW

The Waze data team is working on a data analytics project focused on boosting overall growth by reducing monthly user churn(defined as users who uninstall or stop using the app). To support this goal, the team developed a binomial logistic regression model, known for its flexibility and predictive power in informing business decisions. This report highlights key findings from **Milestone 5**, which will shape the next phase of the project's development.

PROJECT STATUS

Milestone 5 - Regression Modeling

🎯 **Target Goal:** Build and analyze a logistic regression model using user data to predict churn.

🎯 **Methods:**

- Created stakeholder-relevant features
- Checked for multicollinearity
- Built logistic regression model
- Evaluated model performance

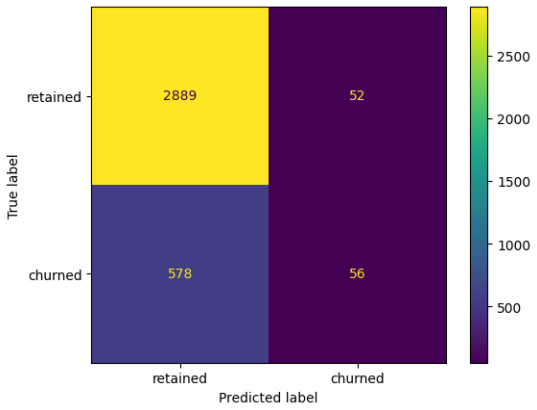
🎯 **Impact:** With sufficient data, a binomial logistic regression model can uncover key variable relationships and predict binary outcomes, supporting decisions in areas like marketing and product development.

NEXT STEPS

➔ **Based on the model results, we recommend using the key insights from this milestone to guide further analysis.**

➔ **This model is not suited for major business decisions, but it offers valuable insights. It highlights the need for more relevant data linked to user churn and suggests refining the target user profile to support Waze's goal of reducing monthly churn and driving growth.**

KEY INSIGHTS



Note: 1 = churned and 0 = retained

- The effectiveness of a binomial logistic regression model is measured by accuracy, precision, and recall. **In this case, recall is especially important, as it reflects how well the model identifies churned users.**
- The model shows moderate precision, with 52% of its positive predictions being correct, but very low recall—only identifying 9% of churned users. This indicates a high number of false negatives and a poor ability to detect users likely to churn.
- **Activity_days** was the top feature, negatively correlated with churn—more active days meant lower churn.
- Earlier EDA showed churn increased with higher **km_per_driving_day**, but in the model, this feature was the second least important.