User Churn Project | Regression Modeling Results

Prepared For: Waze Leadership Team

OVERVIEW

The Waze team is currently developing a data analytics project aimed at increasing overall growth by identifying and preventing monthly user churn. Churn quantifies the number of users who have uninstalled or stopped using the app.
For the purposes of milestone #5, a binomial logistic regression model is being developed to identify which variables have the most predictive power. This information will be used to make data-driven business decisions that will impact the future of the project.

PROJECT STATUS

Target Goal: Utilize user data to build and analyze a binomial logistic regression model.

Methods:
- Engineered features of interest to the business scenario.
- Analyzed features for presence of multi-collinearity.
- Built and fit the logistic regression model.
- Evaluated model performance.

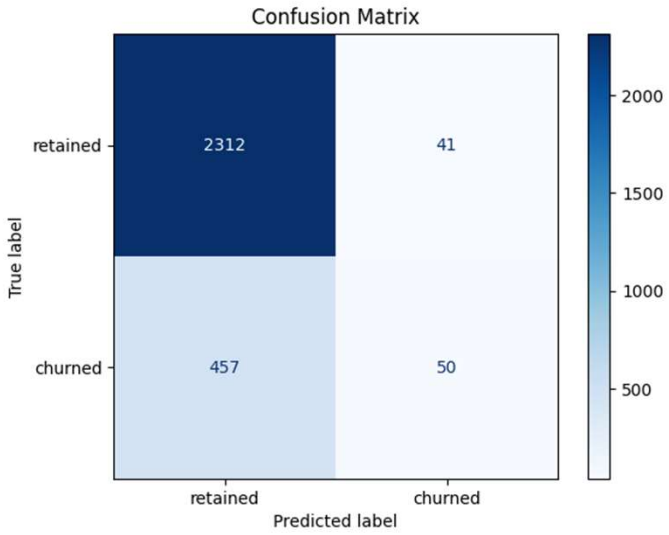
Impact: With enough quality data, binomial logistic regression models can be an effective tool to identify feature relationships and reliably predict binary outcomes, providing businesses with the information to make data driven decisions.

NEXT STEPS

Due to the underperforming models results, the team recommends leveraging the key insights derived from the model to assist with further exploration.

The model cannot be used to make any significant business decisions. That being said, the valuable insights gained from building and analyzing the models results indicate further data related to user churn is necessary. Additionally, more insight and information into the particular user groups that are potentially being targeted with this project would be helpful, as through EDA so far, it's clear there is bias towards a certain group of drivers.

KEY INSIGHTS



- The model has satisfactory precision, scoring 53%. However, the recall score is very low, at only 10%. The recall score reflects the model's capability of identifying users who will churn, and this model produces many false negatives and fails to capture users who will churn.
- 'activity_days' was the most significant predictor, with a negative correlation with churn; as 'activity_days' increases, the likelihood of a user churning decreases.
- Recall is an essential score for this model, as it indicates the number of churned users identified. Binomial logistic regression models rely heavily on precision, accuracy, and recall to determine their effectiveness.
- The engineered variable 'professional_driver' was one of the top 3 variables with the most predictive power.