Waze Churn Project | Preliminary Data summary

Prepared for: Waze leadership team

OVERVIEW

Milestone 5 - Regression Modeling

The Waze data team is now working on a data analytics initiative targeted at boosting overall growth by reducing monthly user attrition on the Waze app. For the purposes of this project, churn refers to the number of users who uninstalled or ceased using the Waze app. Binomial logistic regression models are often flexible and predictive, making them useful for informing broader business choices. Our team attempted to create one using the data provided for this project. This report provides Milestone 5 data and significant insights that will have an impact on the project's future progress.

PROJECT STATUS

Goal: Use user data to create and analyze a binomial logistic regression model. Methods: Developed features relevant to stakeholders and the business environment.

Method:

- Features were assessed for multicollinearity.
- Built the regression model.
- Evaluation of model performance .

Impact: With enough data, binomial logistic regression model findings can highlight crucial variable correlations and forecast binary outcomes, guiding marketing and product development decisions, for example.

NEXT STEPS

➔ **Based on the model results, our team suggests adopting the main insights from this project milestone to guide future exploration.**
➔ **This model should not be used to make significant business decisions; however, it provides valuable insights in that it demonstrates a significant need for additional data (features) that correlates with user churn, as well as a potential need to better define the user profile Waze seeks to target in order to increase overall growth by preventing monthly user churn on the app.**

KEY INSIGHTS
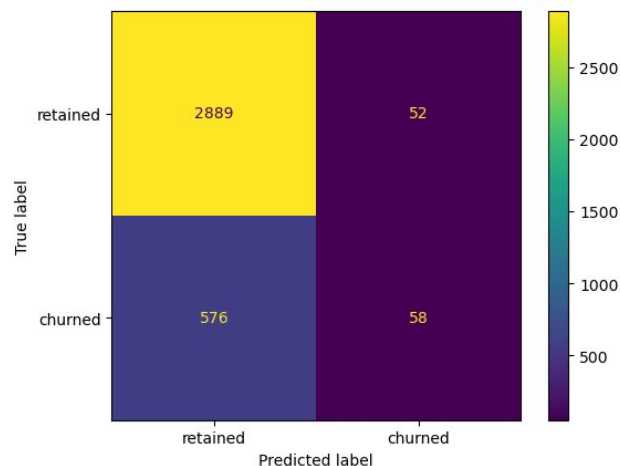
Note: 1 = churned and 0 = retained

- The effectiveness of a binomial logistic regression model is assessed by its accuracy, precision, and recall scores; recall, in particular, is critical to this model because it indicates the amount of churned users.
- The model has mediocre precision (53% of positive predictions are right), but extremely low recall, identifying only 9% of churned users. This means that the algorithm generates a lot of false negative predictions and fails to identify customers who would churn.
- Activity_days was by far the most essential component of the model. It was negatively correlated with user churn.
- In the previous EDA, user churn grew as the km_per_driving_day numbers climbed. Distance driven per day was the model's second-least important variable.