Understanding Waze User Churn | Regression Modeling Results

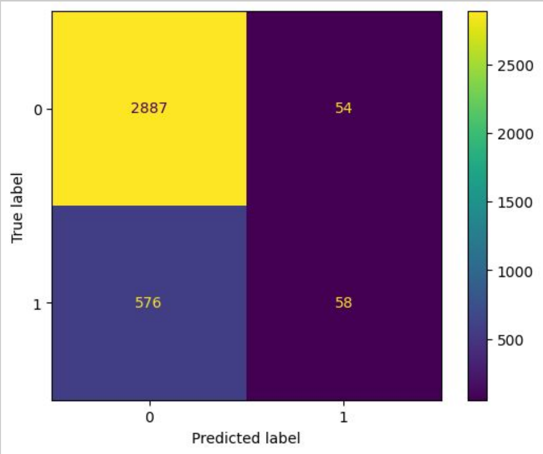
Prepared for: Waze Leadership Team

Project Overview

The Waze data team is currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn (users who have uninstalled/stopped using) on the Waze app. Binomial logistic regression models typically offer flexibility and predictive power, which can be used to inform larger business decisions. Our team sought to build one from the data provided for this project. **This report offers details and key insights from Milestone 5, which impact the future development of the overall project.**

Key Insights

- The model had **mediocre precision** and **very low recall** meaning it makes a lot of false negative predictions and fails to capture users who will churn.
- **Activity_days** was the most important feature in the model and was **negatively correlated** with user churn (-0.106032), implying that more activity days reduce the likelihood of churn. Specifically, **for each additional day of activity, there is a 10.6% reduction in the likelihood of churn**, assuming all other factors are constant.
- In the previous EDA, the churn rate amongst users increased as km_per_driving_day increased, but in this model, that variable was the **second least important variable.**



Note: 1 = churned and 0 = retained

Details

Milestone 5 - Regression Modeling

- **Target Goal:** Apply user data to build and analyze a binomial logistic regression model.
- **Methods:**
  - Created features of interest to the stakeholders and business scenario
  - Assessed features for multicollinearity
  - Built the regression model
  - Evaluated model performance
- **Impact:** With enough data, binomial logistic regression model results can reveal important variable relationships and predict binary outcomes, which can inform decisions for marketing and product development, for example.
- **Considerations:** The current MLR model uses all available data, and improving its accuracy would require additional, relevant data. Simply adjusting the model won't enhance performance without incorporating more comprehensive features.

Next Steps

- This model **should not** be used to make significant business decisions but **reinforces** the need for **additional data** that correlates with user churn.
- If no other data is available, the most predictive factor for churn is **activity days**, so prioritizing strategies to **increase user activity** via the app should be the main focus. As for the other variables, their small coefficients suggest **minimal impact** on churn, so it's prudent to **avoid making business decisions** based on them at this stage.