User Churn Project | Regression Modeling Results

For Waze Leadership Team

OVERVIEW

Our data analytics project aims to boost the overall growth of Waze by reducing the monthly user churn rate. User churn refers to the number of users who have either deleted the Waze app or stopped using it. We decided to use a binomial logistic regression model for its flexibility and predictive power, which can help us make better business decisions. **In this report, we share the details and key insights from Milestone 5, which affect the future direction of our project.**

PROJECT STATUS

Milestone 5 - Regression Modeling

Target Goal: Using the data collected from the users, construct and examine a binomial logistic regression model.

Methods:

- Developed relevant attributes for the business case and stakeholders
- Checked attributes for interdependence
- Constructed the regression model
- Measured model performance

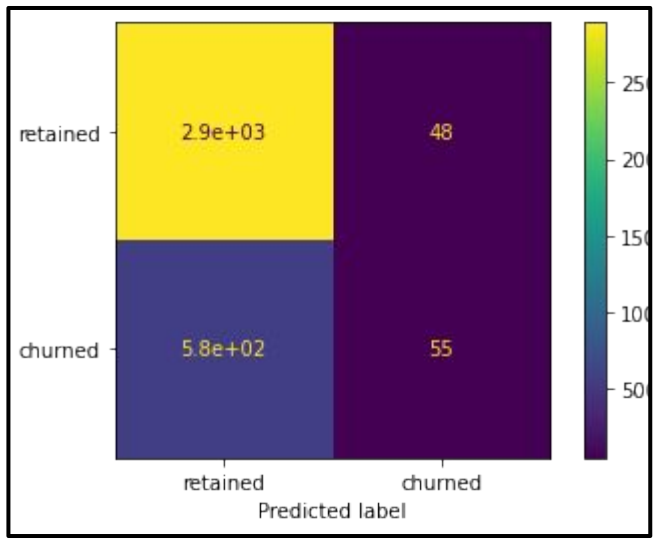
Impact: A binomial logistic regression model can help us understand how different variables affect binary outcomes, such as customer behavior or product performance. This can be useful for making decisions in areas like marketing and product development, if we have enough data to support the model.

NEXT STEPS

➔ **Based on the model outcomes, our team suggests applying the main findings from this project stage to inform future investigation.**

➔ **This model is not suitable for making important business decisions, but it reveals some useful information. It shows that more data (features) that relate to user churn are needed, and that the target user profile for Waze may need to be redefined to achieve higher growth by reducing monthly user churn on the app.**

KEY INSIGHTS



- A binomial logistic regression model can be evaluated by three metrics: accuracy, precision, and recall. Among these, **recall is especially important for this model because it indicates how many users who left the service were correctly identified.**
- **The model's precision is not very high (only 53% of its positive predictions are accurate) and its recall is extremely low, with just 9% of churned users detected.** This implies that the model produces many false negative predictions and misses users who will churn.
- **The model showed that the number of days a user was active had the strongest influence on user retention.** This feature had an inverse relationship with user churn.
- As we saw in the exploratory data analysis, the more kilometers a user drove per day, the more likely they were to churn. However, **the model ranked km_per_driving_day as the second least important predictor of churn.**