

User Churn Project | Regression Modeling Results



Prepared for Waze Leadership Team

ISSUE / PROBLEM

The Waze data team is focused on a data analytics project to reduce monthly user churn on the Waze app. Churn is defined as users who have uninstalled or stopped using the app. We're using binomial logistic regression models to guide our decision-making, and this report presents key insights from Milestone 5 to shape the project's future development.

RESPONSE

Regression Modeling:

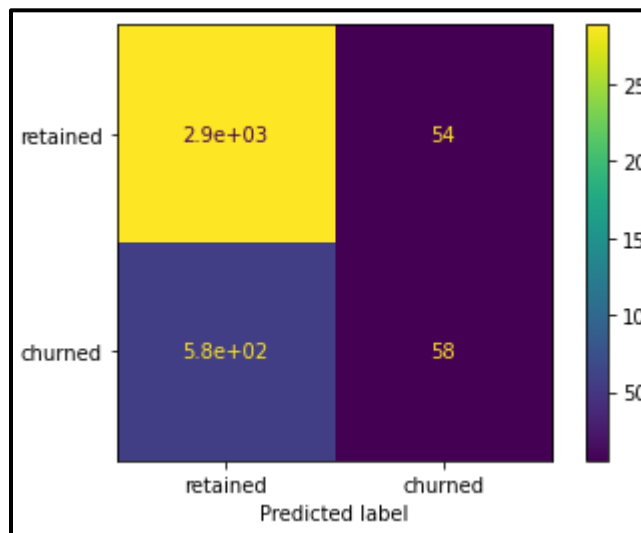
Goal: Analyze the user data by employing a binomial logistic regression model.

Methods:

- Cleaned and wrangled data
- Engineered new feature based on a business scenario
- Assessed features for multicollinearity
- Built the regression model
- Evaluated model performance

IMPACT

Utilizing binomial logistic regression with sufficient data can uncover essential variable relationships and predict binary outcomes, assisting in drawing informed business decisions.



Note: churned = 1, and retained = 0

NEXT STEPS

- After evaluating the model results, our team recommends using key insights from this milestone to perform further exploration.
- This model did not perform well, so leaders should not use it to make significant business decisions. However, we recommend utilizing the insights to find a proper model, which will be more efficient.

KEY INSIGHTS

- To evaluate the effectiveness of a binomial logistic regression model one should check accuracy, precision, and recall scores. Our constructed model has a very low recall rate, 9%, demonstrating many false negative. That is, the model fails to capture users who will churn.
- Our analysis shows that the Activity_days variable was the most important feature in this mode. Activity_days had a negative correlation with the user churn.
- In our previous analysis, we found that when the km_pe_driving_day variable increased, user churn rate increased as well. However, in our current model, distance driven per day was the second least important variable.