User Churn Project | Regression Modeling Results

OVERVIEW

The data analytics project aimed at increasing overall growth by preventing monthly user churn on the Waze app. For the purposes of this project, churn quantifies the number of users who have uninstalled the Waze app or stopped using the app. Binomial logistic regression models typically offer flexibility and predictive power, which can be used to inform larger business decisions. A logistic regression model have been built with the dataset provide. **This report offers details and key insights about this regression model analysis.**

PROJECT STATUS

Regression Modeling

🎯 **Target Goal:** Apply user data to build and analyze a binomial logistic regression model.

🎯 **Methods:**

- Created features of interest for data-driven decision making
- Assessed features for multicollinearity
- Built the regression model
- Evaluated model performance

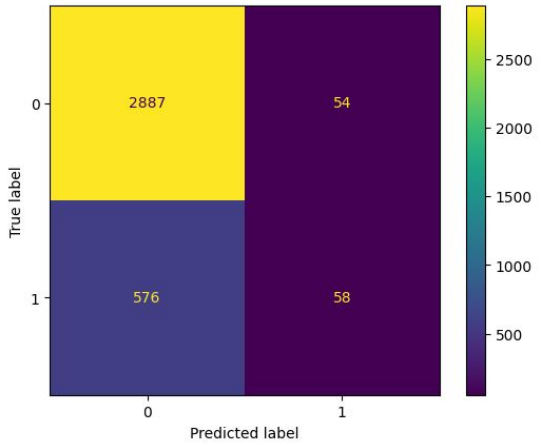
🎯 **Impact:** With enough data, binomial logistic regression model results can reveal important variable relationships and predict binary outcomes, which can inform decisions for marketing and product development, for example.

NEXT STEPS

➔ **Due to the model results, key insights will be used to guide further exploration.**

➔ **This model should not be used to make significant business decisions; however, it has valuable insights insofar as it demonstrated a great need for additional features that correlates with user churn, and specifically target some specific user groups, as Waze seeks to increase overall growth by preventing monthly user churn on the app.**

KEY INSIGHTS



Note: 1 = churned and 0 = retained

- The efficacy of a binomial logistic regression model is determined by accuracy, precision, and recall scores; in particular, **recall is essential to this model as it shows the number of churned users.**
- **The model has mediocre precision (53% of its positive predictions are correct) but very low recall, with only 9% of churned users identified.** This means the model makes a lot of false negative predictions and fails to capture users who will churn.
- **Activity_days was by far the most important feature in the model.** It had a negative correlation with user churn.
- In previous EDA, user churn rate increased as the values in km_per_driving_day increased. **In the model, distance driven per day was the second-least-important variable.**