# User Churn Project | ML Model Results

Prepared for: Waze Leadership Team

## ISSUE / PROBLEM

The Waze data team is currently working on a data analytics project to boost overall growth by preventing users from leaving the Waze app each month. Churn refers to the number of users who uninstall or stop using the app. The project's ultimate goal is to create a machine learning (ML) model that predicts user churn. **This report provides details and key insights from Milestone 6, which could influence the project's future development if further work is done.**

## IMPACT

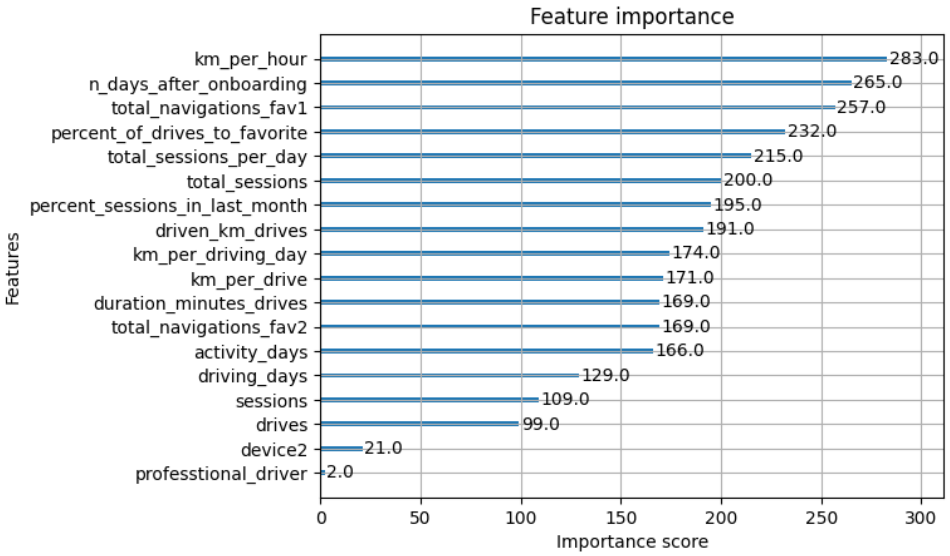
- ➔ **The ML models from Milestone 6 highlight a clear need for more data to improve churn prediction accuracy.**
- ➔ **This modeling effort confirms that the current data is not enough to reliably predict churn. Additional drive-level details—such as drive times and locations—would help. More granular data on user interactions, like reporting or confirming road hazards, and monthly counts of unique start and end locations, could also improve the model.**
- ➔ **Since engineered features can greatly improve ML model performance, the Waze team recommends a second iteration of the User Churn Project.**

## RESPONSE

- **To maximize predictive power, the Waze data team built and compared two models: Random Forest and XGBoost.**

- The data was split into training, validation, and test sets. **While this reduces the amount of data available for training, it allows model selection using the validation set and gives a more reliable estimate of future performance by testing the final model separately on the test set.**

## KEY INSIGHTS




- **Engineered features made up six of the top 10 features:** km_per_hour, percent_of_drives_to_favorite , total_sessions_per_day, percent_sessions_in_last_month, km_per_driving_day, km_per_drive.

- **The XGBoost model fit the data better than the random forest model**. Additionally, it's important to note that the recall score (~14%) shows a meaningful improvement over the previous logistic regression model(~ 9%) developed in Milestone 5.

- **The ensemble tree-based models used in this project milestone are more effective than the singular logistic regression model, as they achieve higher scores across all evaluation metrics and require less data preprocessing. However, their predictive processes are more complex and less interpretable.**