

Waze User Churn Data Analysis Project Data Summary

by Michelle Aguiar

Overview

This analysis supports Waze's broader initiative to reduce user churn and improve retention. The dataset contains detailed activity records for 14,999 users, including metrics on app usage, trip behavior, and device type. The goal of this phase was to evaluate data quality, surface behavioral patterns, and prepare for predictive modeling.

Project

This project focuses on early-stage exploratory analysis of user churn. Key tasks included:

- Assessing completeness and consistency of the dataset
- Comparing behavioral patterns between churned and retained users
- Engineering additional features to enhance understanding of engagement
- Evaluating the distribution of churn across different user segments

All tasks were conducted in Python using pandas, within a Jupyter Notebook environment.

Key Insights

- Data completeness: Only 700 users (4.7%) were missing churn labels. These missing values appear random and are not associated with device type or user activity, minimizing the risk of bias in future modeling.
- Churn distribution: 82% of labeled users were retained, while 18% had churned. This class imbalance will need to be addressed in later modeling stages.
- User behavior: Churned users exhibited higher-intensity usage over fewer days. They completed more drives, traveled greater distances, and spent more time in the app, but were less consistently active than retained users.
- Engineered metrics such as kilometers per drive, kilometers per driving day, and drives per driving day revealed that churned users often completed more drives and traveled farther per session, indicating a possible segment of high-usage drivers (e.g., commercial or long-haul drivers).
- Device type showed no relationship with churn. iPhone and Android users were proportionally distributed across all segments, including those with missing labels.

Next Steps

- Finalize data preprocessing by addressing missing label values appropriately
- Proceed with feature selection and development of a churn prediction model
- Explore clustering or segmentation techniques to identify user groups with distinct retention risks
- Recommend product improvements or personalized engagement strategies tailored to high-intensity user profiles