

Waze: Predicting User Churn

Exploratory Data Analysis

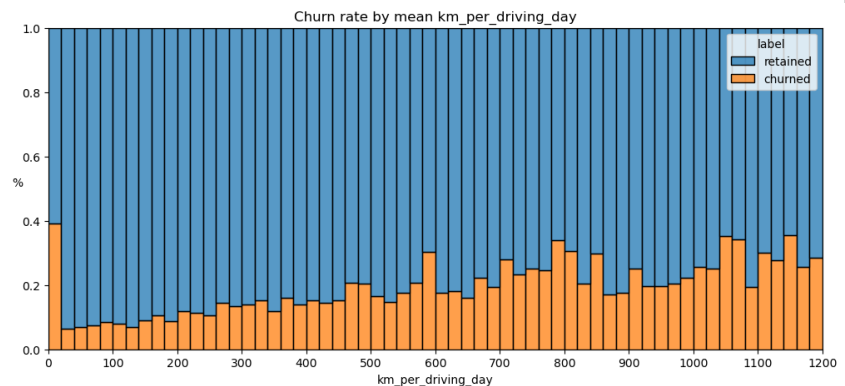
Project Overview

The objective of this project is to conduct Exploratory Data Analysis (EDA) on the Waze dataset to better understand the variable distributions, data quality and the behavioural patterns associated with user churn. The analysis focused on identifying missing values, detecting outliers, validating derived metrics and examining how user engagement differs between churned and retained users.

Key Insights

- Majority of usage-related variables exhibit **right-skewed distributions**, indicating a small group of highly active users.
- User **churn increases with driving intensity**, as the average km_per_driving_day increases the proportion of churned users rises.
- Several extreme outliers were identified and handled using **percentile-based capping** to reduce distortion.
- The churned users tended to show **less consistent engagement**, despite sometimes exhibiting higher intensity on active days.
- The device type (Android vs. iPhone) shows **no meaningful difference** in churn behaviour.
- Newer users are more prevalent in the dataset than long-tenured users.

Details



Exploratory Data Analysis revealed that a majority of the user engagement variables are highly right-skewed, with a small subset of users exhibiting extreme driving behaviour. To ensure interpretability, the physically implausible values were excluded, and outliers were capped using percentile-based thresholds. Analysis of the average km driven per driving day shows a clear relationship with churn, this being as the driving intensity increases, the proportion of churned users rises steadily. The retained users are concentrated in lower to moderate usage ranges while the high-intensity users display substantially higher churn rates. This pattern therefore suggests that consistent engagement across days, rather than high daily driving intensity is more strongly associated with retention.

Next Steps

- Perform feature engineering via selection and transformation based on EDA findings.
- Segment users by engagement patterns, such as consistent vs. high-intensity drivers.
- Address class imbalance when modelling churn.
- Build and evaluate baseline churn prediction models using cleaned and engineered features.