

User Churn Prediction Project | Exploratory Data Analysis

Prepared for: Waze Leadership Team

OVERVIEW

The Waze data team is currently developing a machine learning model to predict user churn. An accurate model will help prevent churn, improve user retention, and grow Waze’s business.

The data is ready for exploratory data analysis (EDA) and further data visualization.

PROJECT STATUS

Goal: Conduct exploratory data analysis and create data visualizations.

Tasks:

- Clean data
- Handle outliers
- Perform EDA
- Visualize data

KEY INSIGHTS

- Analysis revealed that the overall churn rate is 17.7%, and that this rate is consistent between iPhone users and Android users. 82.3% of users were retained.
- The ‘sessions’ variable is a right-skewed distribution with half of the observations having 56 or fewer sessions. However, some users have more than 700.
- The ‘drives’ information follows a distribution similar to the sessions variable. It is right-skewed, approximately log-normal, with a median of 48. However, some drivers had over 400 drives in the last month.
- The ‘total_sessions’ is a right-skewed distribution. The median total number of sessions is 159.6.
- The ‘n_days_after_onboarding’ is a uniform distribution with values ranging from near-zero to about 3,500 (about 9.5 years).
- The ‘driven_km_drives’ is a right-skewed distribution with half the users driving under 3,495 kilometers.

NEXT STEPS

- Determine any unusual data points that could pose a problem for future analysis in predicting user churn.
- Determine the variables that have the largest impact on user churn.
- Find out why so many users used the app so much in just the last month.
- Find out why users who drive very long distances on their driving days are more likely to churn, but users who drive more often are less likely to churn.