Predictive Waze users churn Project Proposal

Objective	The data team's objective is to develop a machine learning model to predict user churn that quantifies the number of users who have uninstalled the Waze app monthly.

Planning and Analyzing stages

Milestone	Tasks	Outcome/Deliverables	Estimated Time
Milestone 1	• Outline project workflow • Gather data from various sources • Identify software/hardware needs	• Stakeholders updated	2–3 weeks
Milestone 2	• Clean, convert, and format data	• Database ready for modeling • Stakeholders updated	2–3 weeks

Constructing and Executing stages

Milestone	Tasks	Outcome/Deliverables	Estimated Time
Milestone 3	• Finalize modeling strategies • Build machine learning models • Test models for accuracy	• Machine learning model • Stakeholders updated	4 weeks
Milestone 4	• Finalize results • Share findings with stakeholders • Incorporate feedback	• Visualizations • Executive summary • Results	3 weeks

waze

User Churn Project | Preliminary Data Summary

Prepared for: Waze Leadership Team

OVERVIEW

The Waze data team is currently developing a 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.

This report offers a preliminary data summary, information on the project status and key insights of Milestone 2, which impact the future development of the overall project.

PROJECT STATUS

Milestone 2 - Compile Summary Information

🎯 **Target Goal:** Inspect user data to learn important relationships between variables.

🎯 **Methods:**
- Built a dataframe
  - Each row represents a single observation, and each column represents a single variable
- Collected preliminary statistics
- Analyzed user behavior

🎯 **Impact:** Our team determined important relationships between variables that will guide further analysis of user data.

NEXT STEPS

➔ **Our team recommends gathering more data on the super-drivers**. It's possible that the reason they're driving so much is also the reason why the Waze app does not meet their specific set of needs, which may differ from the typical driver.

➔ **The immediate next step is to conduct thorough EDA and develop data visualizations** to illustrate the narrative behind the data and guide future project decisions.

KEY INSIGHTS

- This dataset contains **82% retained users** and **18% churned users**.

- The dataset contains 12 unique variables with types including objects, floats, and integers; the label column is missing 700 values with no indication that the omissions are non-random.

- Churned users averaged ~3 more drives in the last month than retained users.

- Retained users used the app on over twice as many days as churned users in the last month.

- The median churned user drove ~200 more kilometers and 2.5 more hours during the last month than the median retained user.

- Churned users had more drives in fewer days, and their trips were farther and longer in duration. Perhaps this is suggestive of a user profile; our team will have to continue exploring!

- The median user who churned drove 698 kilometers each day they drove last month, which is about 240% the per-drive-day distance of retained users.

- Regardless of user churn, the users represented in this data drive a lot! It is probably safe to assume that this data does not represent typical drivers at large.

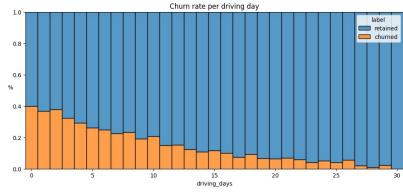# User Churn Project | Exploratory Data Analysis

Prepared for: Waze Leadership Team

Project Overview

The Waze data team is currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn on the Waze app. Thorough exploratory data analysis (EDA) enables Waze to make better decisions about how to proactively target users likely to churn, thereby improving retention and overall customer satisfaction. **This report offers details and key insights from Milestone 3, which impact the future development of the overall project.**

Details

The churn rate is highest for people who didn't use Waze much during the last month.

The proportion of churned users to retained users is consistent between device types.
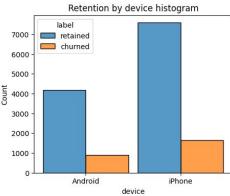
Key Insights

- **The more times users used the app, the less likely they were to churn.** While 40% of the users who didn't use the app at all last month churned, nobody who used the app 30 days churned.

- **Distance driven per driving day had a positive correlation with user churn.** The farther a user drove on each driving day, the more likely they were to churn.

- **Number of driving days had a negative correlation with churn.** Users who drove more days of the last month were less likely to churn.

- **Users of all tenures from brand new to ~10 years were relatively evenly represented in the data.**

- **Nearly all the variables were either very right-skewed or uniformly distributed.**
    - For the right-skewed distributions, this means that most users had values in the lower end of the range for that variable.
    - For the uniform distributions, this means that users were generally equally likely to have values anywhere within the range for that variable.

- **Several variables had highly improbable or perhaps even impossible outlying values**, such as: driven_km_drives, activity_days and driving_days.

Next Steps

➔ **Investigate the erroneous or problematic discrepancies between number of sessions, driving_days, and activity_days.**

➔ **Continue to explore user profiles with the greater Waze team; this may glean insights on the reason for the long distance drivers' churn rate.**

➔ **Plan to run deeper statistical analyses on the variables in the data to determine their impact on user churn.**

User Churn Project | Two-Sample Hypothesis Test Results
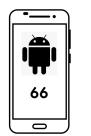
Prepared for: Waze Leadership Team

Overview

The Waze data team is currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn on the Waze app. As part of the effort to improve retention, Waze wants to learn more about users' behavior. **This report offers information on the project status and results of Milestone 4, which impact the future development of the overall project.**

Objective

🎯 **Target Goal:** Develop a two-sample hypothesis test to analyze and determine whether there is a statistically significant difference between mean number of rides and device type – Android vs. iPhone.

🎯 **Impact:** Statistical tests, such as the one conducted for Milestone 4, enable the Waze data team to make inferences about the populations from which the data was drawn and help them learn more about their user base.

Results

Average Number of Drives

66

68

Note: The mean number of drives shown here – 66 for Android and 68 for iPhone – have been rounded up.

- Based on the calculations, drivers who use an iPhone to interact with the application have a higher number of drives on average.

- **The t-test results concluded there is not a statistically significant difference in mean number of rides between iPhone users and Android users.**

Next Steps

➜ **Due to the results rendered from this specific hypothesis test, the Waze data team recommends running additional t-tests on other variables to learn more about user behavior.**

➜ **Additionally, since the user experience is the same, temporary changes in marketing or user interface may be impactful rendering more data to investigate user churn behavior.**

Additional information from source

Data dictionary

14,999 rows – each row represents one unique user

13 columns

Column name	Type	Description
ID	int	A sequential numbered index
label	obj	Binary target variable ("retained" vs "churned") for if a user has churned anytime during the course of the month
sessions	int	The number of occurrence of a user opening the app during the month
drives	int	An occurrence of driving at least 1 km during the month
device	obj	The type of device a user starts a session with
total_sessions	float	A model estimate of the total number of sessions since a user has onboarded
n_days_after_onboarding	int	The number of days since a user signed up for the app
total_navigations_fav1	int	Total navigations since onboarding to the user's favorite place 1
total_navigations_fav2	int	Total navigations since onboarding to the user's favorite place 2
driven_km_drives	float	Total kilometers driven during the month
duration_minutes_drives	float	Total duration driven in minutes during the month
activity_days	int	Number of days the user opens the app during the month
driving_days	int	Number of days the user drives (at least 1 km) during the month