Waze User Churn Dataset Exploratory Data Analysis

Executive Summary

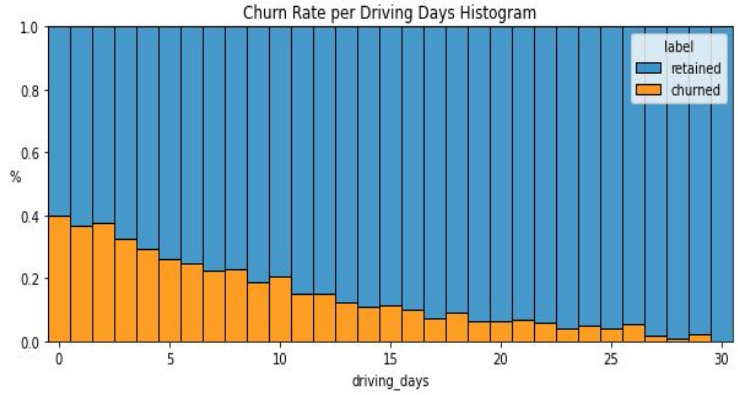
Project Overview

The Waze data team is currently in the process of creating a machine learning model to predict user churn rates but prior to the creation of said model exploratory data analysis (EDA) must be completed on the dataset. The completion of this EDA will allow the waze data team to be proactive in minimizing the monthly churned users.

Key Insights

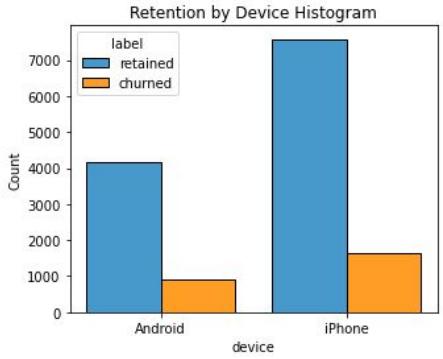
- **Both long-time users and brand new users were well represented in the dataset.**
- **Negative correlation between driving days and churn rate.**
  - The more days a user spent driving with the waze app the less likely they were to churn.
- **Positive correlation between kilometers driven and churn rate.**
  - The more kilometers a user drove the more likely they were to churn.
- **Most variables were evenly distributed or rightly-skewed.**
  - The right-skewed data means that most users fell in the lower range of the variable whereas for the evenly distributed data there was a consistent amount of users throughout all ranges of the variable.
- **Some variables had extreme values that are most likely impossible.**
  - Example: The max value for kilometers driven in a day was 15 420 kms.

Details



The more active a user was the less likely they were to churn. 40% of users who didn't use the app last month churned while those who used the app for 30 days didn't churn at all.

The percentage of churned users per device is consistent between iPhones and Androids.



Next Steps

- Take a deeper look at certain discrepancies in the variables such as the number of days in a 'month' for the driving days and activity days variables.
- Speak with the rest of the waze team to see if any more insights can be gained on why the more kilometers driven the higher chance of churning. Could the potential longer drive distances be due to issues with the app causing users to be frustrated and churn? It is important to try to make sense of these issues.
- Prepare for deeper analysis on individual variables within the dataset to uncover the potential links to user churn rates.