Waze Predictive Modelling Project

Initial Data Preparation and Coding Report

Project Overview

The project aims to develop a machine learning model to predict user churn to optimise Waze's retention strategy and enhance user experience. This report will provide an overview of the insights obtained from preliminary data preparation.

Details

The median churned user drove approximately 240% more distance per day than the retained users. A similar ratio was seen for the number of drivers per day.

The 12 variables are of integer, float and string data types, and most of the columns do not show any signs of outliers being present.

64.5% of the dataset consists of iPhone users while the other 35.5% are Android users. These ratios are consistent with the values for churned (64.9%-35.1%) and retained users (64.4%-35.6%)

Key Insights

Out of the 14999 observations included, 82.3% of the users were retained while 17.7% churned.

The dataset consists of 12 variables, containing missing values only in the 'label' column (700 rows) and this is shown to have no discernible effect on the overall dataset.

A large portion of the churned users are possibly long-haul truckers who possibly have needs that differ from a typical driver.

The device type is shown to have no noticeable effect on the churn rate.

Considering the median values for churned and retained users, the former are shown to have higher number of drives and drive durations, while the latter have a great number of active days and driving days.

Next Steps

It would be helpful to collect more information about the churned users, particularly the super-drivers to identify how their needs and reasons for churning. This could be used to decide on a strategy to improve retention by addressing any user experience issues.

For the predictive model, the next step would be carrying out Exploratory Data Analysis before proceeding to model development.