User Churn Dataset Inspection

Milestone 2

Overview

Waze has initiated a new project to develop a machine learning model to predict user churn. The team has just received access to user data and is in the early stages of preparing the dataset for analysis. This summary outlines the initial inspection and analysis of the dataset to ensure it is ready for exploratory data analysis (EDA) and model development.

Key Insights

The dataset contains **14,999 rows and 13 columns**, including 11 numeric and 2 categorical variables.

There are **700 missing values** in the `label` column, representing users without a known churn status. These missing values appear **randomly distributed** across device types (64.5% iPhone, 35.5% Android), consistent with their overall proportions in the dataset.

About **82% of users are retained**, and **18% are churned**.

Churned users drive more frequently, farther, and longer in fewer days compared to retained users.

- Median kilometers per drive: Retained – 75.0 km, Churned – 74.1 km
- Median kilometers per driving day: Retained – 289.5 km, Churned – 697.5 km
- Median drives per driving day: Retained – 4.06, Churned – 10.0

These patterns suggest a segment of **"super-drivers"**, possibly long-haul drivers, whose behaviors differ significantly from typical users.

Device usage (iPhone vs. Android) is balanced across churned and retained groups, indicating **no meaningful churn difference by device type**.

Next Steps

1. **Gather additional data on super-drivers** to better understand their behaviors, needs, and potential reasons for disengagement from the app.
2. **Perform comprehensive exploratory data analysis (EDA)** to identify usage trends, behavioral patterns, and anomalies across user segments.
3. **Create visualizations** to support key findings,