Title Waze's Data Team Logistic Regression Model Project

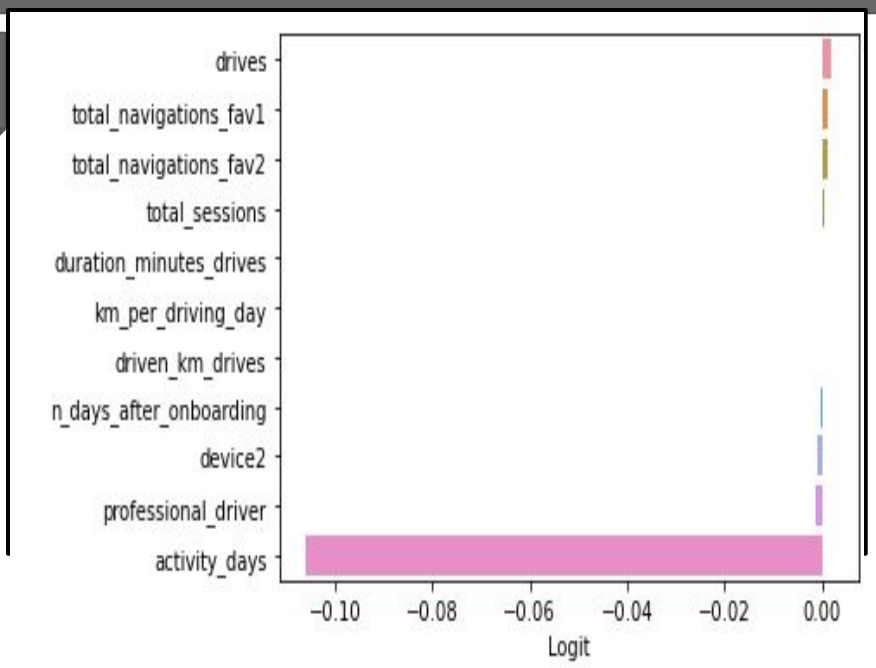
Project Overview

This project aims to build a logistic regression model to predict the probability of churned user rate based on the number of features. By that, we can know which feature(s) might significantly impact the target variable, and determine the decisions about it afterward.

Key Insights

- The model results show that the logistic regression model we built is ineffective.
- The model's predictions majority contain the **retained** users' values over the churned users. It predicts the number of **retained** users is 3480. However, the main problem is that 580 users are **churned**, but the model predicts it as **retained**. Needless to say, if we choose to implement this model, we might expect to lose the users rather than continue using our app.
- The model's score is also very low. Specifically, the **recall score** (which determines the proportion the model accurately predicts the churned users, out of the total of **churned** users that are true) is 0.09%. This low proportion is due to the model incorrectly predicting **retained** users but **churned** users.
- The model also predicts a total of 112 **churned** users, which contains 54 incorrectly. In other words, 54 users are **retained**, but the model predicts as **churned**.
- Even though the model is ineffective, some variables have a significant impact on the user rate probability. The **activity_days** variable is the most influential variable in the model which shows a negative correlation to the target variable. Specifically, if we increase one unit of **activity_days** (increase one day) and hold other variables constant, we might expect the odds of the **churned**-users rate to decrease by 10%.
- Besides the activity_days, others don't show much significant impact on the target variable.

Details



*This plot shows the features importance of the model. Specifically, we can easily detect that **activity_days** is the most important variable to the users churn rate (negative correlation).*

Next Steps

- Since this logistic regression model we built is ineffective, we might recommend not implementing it for business decisions about user churn rate.
- Our Waze data team will further examine the results of the model performance to adjust. For example, we might have to carefully choose the variables that significantly impact the target variables. If necessary, we can add some more features to account for more impactful as well.
- Perhaps we might then choose to remodel to determine if it brings more effective results.