Waze Project

Milestone 5 - Regression Modeling

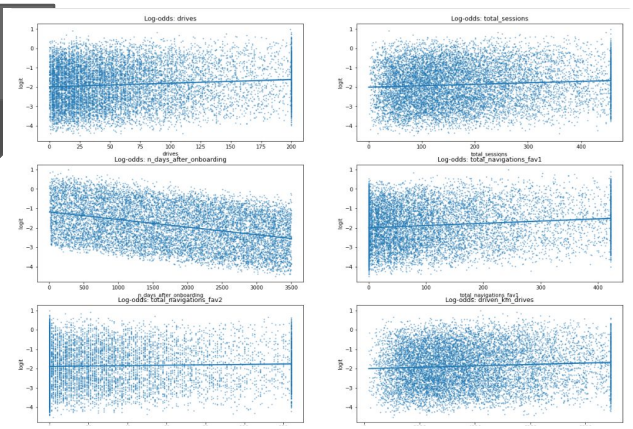
Project Overview

Waze leadership has asked the data team to analyze the relationship between mean amount of rides and device type. The goal is to build a binomial logistic regression model to predict user churn based on a variety of variables and evaluate the model's performance.

Key Insights

- activity_days was by far the most important feature in the model. It had a negative correlation with user churn.
- **Since we did not find any statistically significant difference in the mean amount of drives between iPhone users and Android users, we can conclude that the average rides variable is no obvious relationship between mean amount of rides and device type.**
- To improve this model, we can try to create some new features like professional_drive which is the third important feature in this model, or transform predictors to better identify the linear relationship, or reconstruct this model with other combination of variables.

Details



- Linearity assumption is not met.
- Recall is 9%.
- Precision is 52%.

	precision	recall	f1-score	support
retained	0.83	0.98	0.90	2941
churned	0.52	0.09	0.16	634
accuracy			0.82	3575
macro avg	0.68	0.54	0.53	3575
weighted avg	0.78	0.82	0.77	3575

Next Steps

- It would be helpful to have drive-level information for each user (such as drive times, geographic locations, etc.).
- It would probably also be helpful to have more granular data to know how users interact with the app.