MACHINE LEARNING MODEL THAT PREDICTS USER CHURN

Preliminary Data Summary

Project Overview:

This project is developed to help prevent user churn on the Waze app, focusing on monthly user churn. This report offers a preliminary data summary and key insights of Milestone 2, which impact the future development of the overall project.

Key Insights

- Database contains 12 fields and 14,999 observartons
- Since now it is estimated that Label field could act as dependent variable (churned or retained)
- There are 700 missing values in Label field; 447 (~ 64%) for iphone and 253 (~ 36%) for Android. The proportion for whole data set it seem to keep same ratios.
- Regarding Label field, proportion are ~ 82% for retained and ~ 18% for churned.
- Grouping by Label, churned median user drove ~200 more kilometers and around 2.5 more hours during the last month than the median retained user.
- Creating km_per_driving_day calculated field, median churned user drove ~ 2,4 times median retained users. For drives_per _driving _day calculated field was similar proportion.

Details

- For this preliminary data summary was use Median instead of Mean minimize influce of outliers influence and skewed distributions.
- There was no difference in churn rate between iPhone and Android users; iPhone (~ 17.0%) and Android (~ 16.7%)
- Regardless of user churn, the users represented in this data drive a lot! It is probably safe to assume that this data does not represent typical drivers at large.

Next Steps

➔ Carrying out EDA.

➔ To create visualization to find more insights frim data.

➔ To get more data for users who tended to drove more, because it's possible that the reason they're driving so much is also the reason why the Waze app does not meet their needs.