# User Churn Project | Logistic Classification

## OVERVIEW

A first model for prediction of churning was made base on the logistic regression algorithm

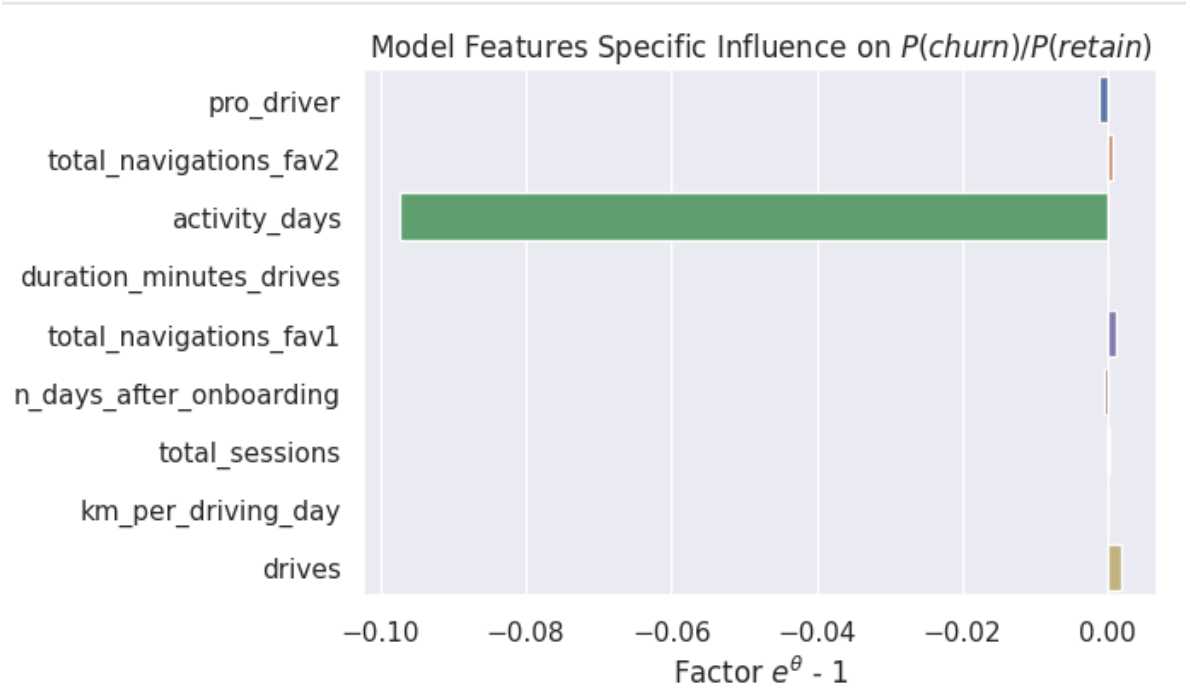
## Methodology

The model was constructed with selected variables according to colinearity criteria. The model was trained witn 70/30 train/test-sets. The test group included 3988 users out of which 3306 were retained and 682 lost as customers.

| | score_counts |
|---|---|
| **TP** | 55 |
| **TN** | 3266 |
| **FP** | 40 |
| **FN** | 627 |
| **Total** | 3988 |

## Model observations

- **It brings poor churn-prediction power: Only 55 matches out of 682**
- It is good at predicting retention: 3266 out of 3306
- Reducing the number of features just increases the bias (poorer results)
- Most influential feature is *activity_days*. About -10% change rate in odds per activity day. Followed by *drives* and *pro_driver*
- User device was already discarded as relevant feature.




## Next Steps

- Try other models than logistic regression
- Get more data:
  - Long term data
  - Measure other factors