

Executive summary for all the projects

Use the Layout dropdown menu to select a template or build your own using these layouts as inspiration.

Salifort Motors (Capstone Project)

Employee Retention

ISSUE / PROBLEM

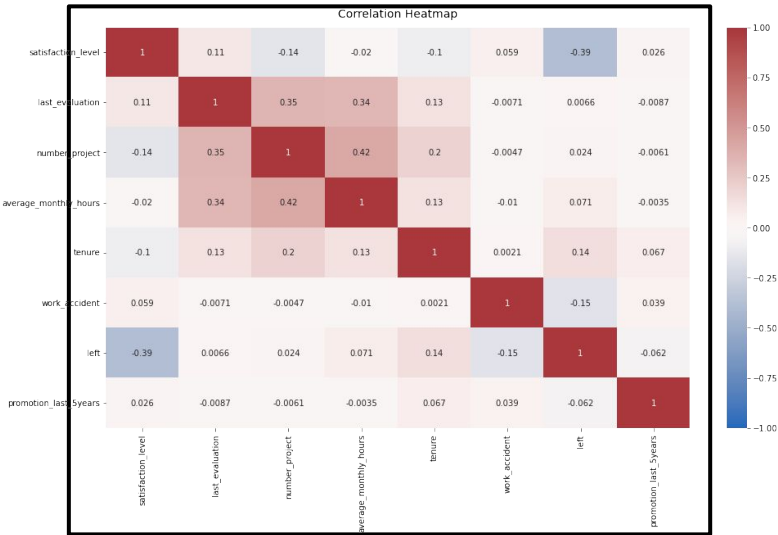
Salifort seeks to improve its employee retention rate.

RESPONSE

Either a binomial logistic regression model or a tree based model would be useful at predicting and understanding the causes for employees leaving since the outcome variable is categorical.

IMPACT

This model will help predict whether an employee will leave or stay. This will allow HR to make changes to company policies and improve retention



Heat map shows positive correlation between number of projects, average number of hours worked and last_evaluation. There is strong negative correlation between employees left and satisfaction_level.

	precision	recall	f1-score	support
0	0.86	0.93	0.90	2321
1	0.44	0.26	0.33	471
accuracy			0.82	2792
macro avg	0.65	0.60	0.61	2792
weighted avg	0.79	0.82	0.80	2792

KEY INSIGHTS

- After conducting the logistic regression model weighted average precision,recall,accuracy and f1 score was 0.79, 0.82, 0.82 and 0.80.
- The employees who are overworked and have over 4 projects seems to be the ones least satisfied and are leaving. The highest working employees are also not getting promoted in the last 4 years.
- Employee work life balance need to improve. Either promote employees who work longer hour or encourage them to work more standard hours. Maybe provide overtime pay if working overtime is necessary.
- Limit the number of projects an employee is working to 3-4. Have company wide surveys to get employee satisfaction anonymously. Also make sure that employee evaluation follows up with support and actionable improvements.
- Have programs such as mental health days, good Paid time Off programs and overall any initiative that would improve work life balance so that employees do not get burnt out. Evaluation scores should not focus on hours worked but more to effective work being done.

User Churn Project/ ML Model Results

For Waze Leadership Team

➤ ISSUE / PROBLEM

The Waze Data team is developing a data analytics project to predict user churn rate on the waze app. Churn is quantified by the no of users who have stopped using the app. The goal is to build and ML model that will successfully predict user churn rate.

➤ IMPACT

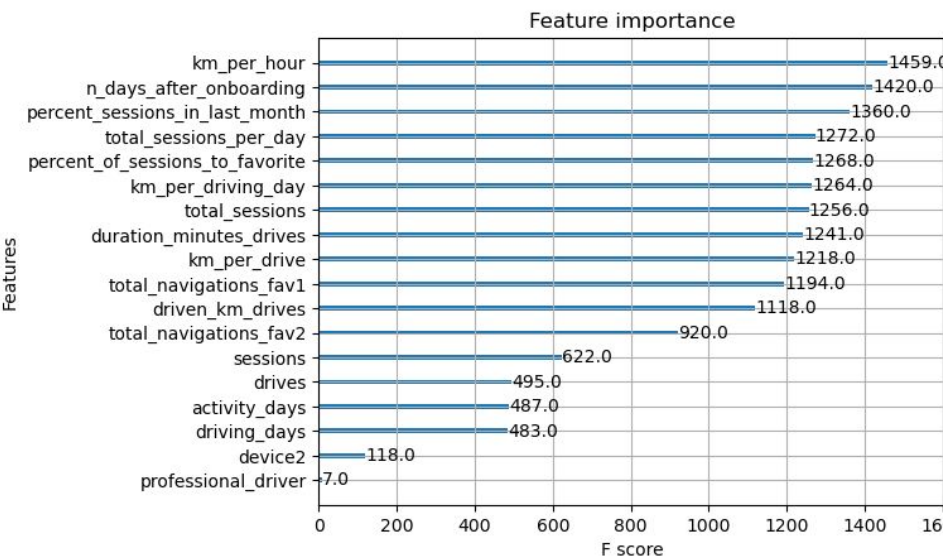
The ML model built lacks sufficient data to predict user churn rate consistently. More granular data is required for the driver and their interaction with the app. How often they post about road hazards.

Another iteration of this project should be initiated with the aforementioned data.

➤ RESPONSE

The Waze team created two different models using the same data: Random Forest and XGBoost for cross comparison. The data was split into Train, Validation and Test data. Both the models were trained on training data and based which ever one performed better on the validation data was chosen to be run on test data. This gave an accurate representation of how the model is going to perform in the future.

➤ KEY INSIGHTS



- Engineered features are 6 of the top 10 features.
- The XGBoost recall score is 17% which is almost twice that gotten from the logistic regression model.
- Overall the model performed better in all the metrics compared to the regression model. It requires less processing but difficult to interpret its predictions.

User Churn project

Prepared for waze leadership team

OVERVIEW

The waze data team is currently developing a ML model that will help reduce user churn rate on the app. User churn rate is the no of users who have uninstalled or stopped using the app.

PROJECT STATUS

Milestone-2

Compiling summary information about the data

- Began exploring the data
- Building a dataframe
- Collected initial insights about churned users

NEXT STEPS

- Need to do more EDA on the churned user data. Maybe our waze app is not suiting the needs of our long haul drivers.
- Want to know how this dataset was gathered
- Need data visualization

KEY INSIGHTS

- Dataset contain 82% retain users and 18% churned user
- There are 700 missing values from the 'label' column.
- The dataset has 12 unique variable containing objects,floats and integers
- Churned users drive 608 kms per driving day vs 247 km for retained users.
- .Churned users also drive around 2.5 times more drives per day.
- Churned users seem to be more long distance drivers and drive further and longer distance.
- Drivers in this dataset drive longer than average drivers

User Churn project-EDA

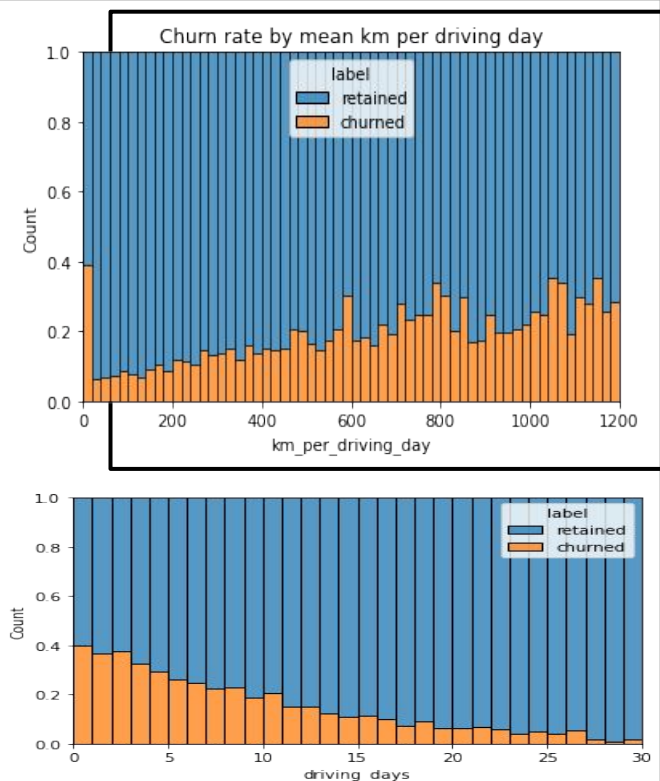
Prepared for waze leadership team

The waze data team is currently developing a ML model that will help reduce user churn rate on the app. User churn rate is the no of users who have uninstalled or stopped using the app. Doing EDA helps uncover insights from the data on what factors might be impacting user churn and how to reduce it.

Key Insights

- Users who drive more km per day have a higher churn rate.
- Users who drive less number of days have a higher churn rate
- There is no correlation between user churn and type of device.
- 40% of users who did not use the app at all churned.
- The number of years the churned users have been using the app has a uniform distribution.
- Most of the distributions of the data were right skewed or had uniform distribution.
 -

Details



Next Steps

- Further information is needed on user demographics and why long distance traveller have a higher churn rate
- Need to understand why so many users started using the app last month
- Need to do further analysis on the variables to find insights
- Investigate erroneous data in number of sessions, driving_days and activity_days

User Churn project

Prepared for waze leadership team

OVERVIEW

The waze data team is currently developing a ML model that will help reduce user churn rate on the app. User churn rate is the no of users who have uninstalled or stopped using the app. We are providing a binomial regression model to the team as it can be a useful tool for predicting binary outcomes such as user churn or not churned. This report details the outcome of the model.

PROJECT STATUS

Milestone-5-Regression Modeling

Target Goal- Apply user data to build a binomial logistic regression model.

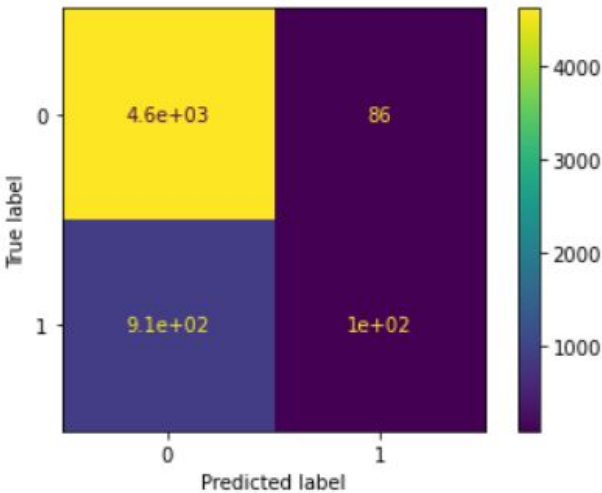
Methods:

- Create new features that might be of interest to stakeholders.
- Check for multicollinearity
- Build Model for regression

NEXT STEPS

- Utilize key insights from this model to guide for next steps.
- This model should not be use to make a business decision due to not providing strong results. More data is needed and variables tested regarding user churn to effectively predict user churn. We need to focus on user profile and types of user we are trying to target.

KEY INSIGHTS



Note: Churned: 1 and retained=0

- They key metrics for binomial regression are Accuracy,Precision and Recall.
- Precision for the model is 53% in forecasting user churned which is fairly low. There is a 9% recall which means only 9% of churned users are identified.
- Activity_days had the most correlation with the user churn.

User Churn Project- Two sample T-test

Subtitle

Overview

The Waze data team is currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn on the Waze app. As part of the effort to improve retention, Waze wants to learn more about users' behavior. **This report offers information on the project status and results of Milestone 4, which impact the future development of the overall project.**

Objective

Target Goal: Develop a two-sample hypothesis test to analyze and determine

whether there is a statistically significant difference between mean number of rides and device type – Android vs. iPhone.




Impact: Statistical tests, such as the one conducted for Milestone 4, enable the Waze data team to make inferences about the populations from which the data was drawn and help them learn more about their user base.

Results

- Based on the calculations, drivers who use an iPhone to interact with the application have a higher number of drives on average.
- **The t-test results concluded there is not a statistically significant difference in mean number of rides between iPhone users and Android users.**

Next Steps

- We may need to run additional A/B tests on different variables to see if is any statistically significance
 - Might be helpful to tweak interface or change marketing strategies to see if it impacts churn rate.
- 
- A decorative graphic at the bottom of the page consisting of a red triangle on the left and a blue triangle on the right, both pointing towards the center.