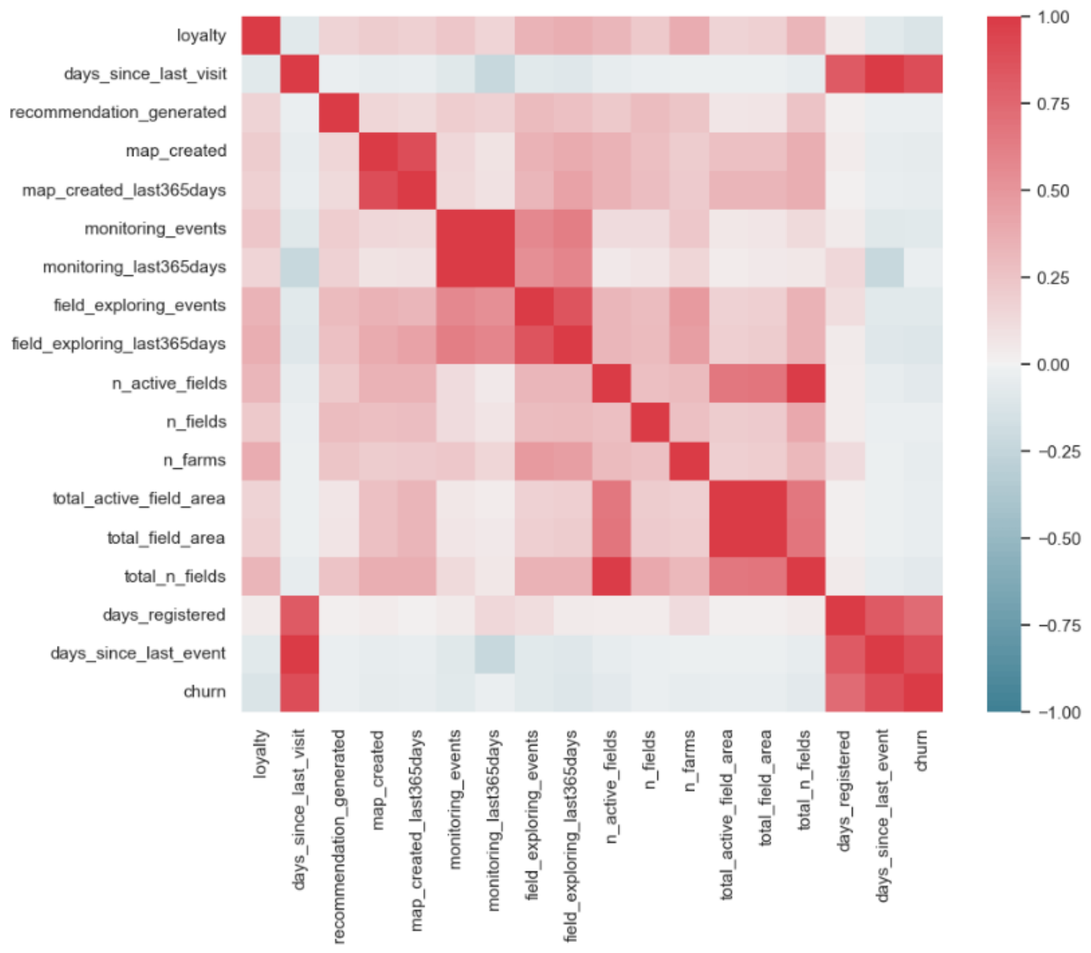
Churn Model: Analyzing User Inactivity

Considering the Loyalty farmer database passed in the case, we will see a churn modeling based on its characteristics and next steps

Factors Contributing to Churn Rate



Loyalty users have higher retention rates compared to non-loyalty users for each churn reason. Specifically, the **field_exploring_events** feature has the highest average value for both churned and retained loyalty users, indicating its importance in retaining loyalty users. On the other hand, non-loyalty users have lower utilization rates across all churn reasons, suggesting the need to improve engagement and utilization of the app features for this user group.

Model Performance

Five models were utilized, and the winner was chosen based on its f1-score, as the objective is to capture the likelihood of churn with maximum accuracy.

After analyzing the feature importance, the following variables were used: 'loyalty', 'recommendation_generated', 'map_created', 'monitoring_events', 'field_exploring_events', 'n_farms', and 'total_field_area':

	column_name	importance
6	recommendation_generated	0.71
3	map_created	0.15
4	n_farms	0.07
2	field_exploring_events	0.03
5	loyalty	0.02
0	total_field_area	0.01
1	monitoring_events	0.00

The models employed were:

- **Logistic Regression** (accuracy: 54.4% and f1-score: 29.4%)
- **LightGBM** (accuracy: 55.7% and f1-score: 35.4%)
- **Decision Tree Classifier** (accuracy: 55.1% and f1-score: 34.9%)
- **Random Forest Classifier** (accuracy: 55.5% and f1-score: 35.6%)
- **KNeighbors Classifier** (accuracy: 48.9% and f1-score: 58.3%)

Based on the f1-score, the selected model was the **KNeighbors Classifier**.

User Churn Prediction Model and Action Plan

Improving Performance and Data Capture

Problem with the Model

Despite achieving relevant accuracy and f1-score, the model did not perform well due to variables that exhibited leakage.

Leakage refers to the inclusion of variables that contain information about the target variable, leading to an overestimation of the model's performance.

Variables that shows leakage: **days_since_last_visit**, **days_since_last_event**, **days_registered**

These variables were created based on the data variables in the given file, where:

- **days_since_last_visit** is the current date minus the date of the last visit to the member
- **days_since_last_event** is the current date minus the date of the last event done in the app
- **days_registered** is the current date minus the date of user registration.

Improving Data Capture

- To develop a high-performing churn prediction model, it is necessary to conduct thorough research and enhance data capture from the database.
- The goal is to identify and capture better variables that are more indicative of churn behavior.
- By improving data capture, we can ensure that the model incorporates relevant and reliable predictors of user churn.

Action Plan

1. **Refine Feature Selection**: Identify and remove variables with leakage to enhance the model's accuracy and generalizability.
2. **Conduct In-depth Research**: Collaborate with domain experts to understand the underlying factors that contribute to user churn and identify potential predictive variables.
3. **Data Exploration and Collection**: Explore and extract additional meaningful variables from the existing database to augment the feature set.
4. **Implement Regular Model Updates**: Establish a system for continuous monitoring and updating of the churn prediction model to adapt to evolving user behaviors and trends.
5. **Validate and Fine-tune the Model**: Employ robust validation techniques, such as cross-validation and testing on holdout datasets, to ensure the model's reliability and effectiveness.
6. **Develop Churn Prevention Strategies**: Utilize the insights gained from the model to create personalized retention strategies targeting at-risk users. This can include targeted marketing campaigns, proactive customer support, or customized incentives.