Analysis of User Adoption

Preprocessing Steps

1. Data Loading and Initial Inspection:

   - Loaded takehome_user_engagement.csv & takehome_users.csv with appropriate encoding. Converted timestamp columns to DateTime format for accurate date manipulations.

2. Handling Missing Values:

   - Inspected and dropped rows with missing values to ensure data integrity.

3. Merging Data:

   - Merged user engagement and user details data on user_id to combine usage patterns with user attributes.

Feature Engineering Steps

1. Defining Adopted Users:

   - Grouped login dates by user_id and determined if a user logged in on 3 separate days within any seven-day period to classify as adopted.

2. Encoding Categorical Features:

   - Encoded categorical variables (creation_source, opted_in_to_mailing_list, enabled_for_marketing_drip) using one-hot encoding for model compatibility.

Model Selection

A RandomForestClassifier was chosen for its robustness and ability to handle feature importance analysis. This model efficiently manages both numerical & categorical features and provides insight into the significance of different variables.

Variables Contributing to User Adoption

Based on the RandomForestClassifier model, the following variables significantly contribute to a user becoming an adopted user:

- Creation Source: The method by which the user account was created (e.g., PERSONAL_PROJECTS, GUEST_INVITE, etc.).
- Opted into Mailing List: Whether the user opted into the mailing list.
- Enabled for Marketing Drip: Whether the user is enabled for the marketing drip.
- Organization ID (org_id): The organization the user belongs to.
- Invited by User ID: The ID of the user who invited the current user.

Summary of the Exploration

This analysis investigates the factors contributing to user adoption of an application. The provided data were merged and cleaned, followed by feature engineering and exploratory data analysis (EDA). A machine learning model was built to identify predictors of user adoption. Key findings indicate that the source of account creation, opting into mailing lists, and marketing drip enablement are significant factors. Further, feature importance analysis highlights the contribution of each factor, guiding future marketing and user engagement strategies.