Summary of the Exploration –

* We explored two datasets: one containing user engagement data and the other containing user information.We cleaned the data by handling null values and creating new features.
* We identified users who logged in at least three times within a seven-day window as adopted users.
* We analyzed the relationship between various features and user adoption.
* The decision tree analysis revealed that the "inactive\_duration" feature is the most important factor in predicting user adoption.

Preprocessing steps taken –

* Cleaned null values in df2.
* Removed duplicate entries for same date and id.
* Grouped df1 by user\_id and checked for 3 logins in a 7 day window.
* Left outer joined df2 with df3 to map all IDs and the number of login.
* Filled null values with 0 for consecutive login column.
* Converted creation\_time column to utc timetstamp.
* Calculated the inactive\_duration for each id.
* Label encoded the creation\_source column.

Feature engineering steps taken –

* Created a new feature 'number\_of\_consecutive\_login' to identify users who logged in at least three times within a seven-day window.
* Identified the adopted and non-adopted users by identifying 3 logins in a 7 day window and left outer joined with user engagement data.
* Converted the 'creation\_time' column to a UTC timestamp.
* Calculated the 'inactive\_duration' feature by subtracting the last session creation time from a fixed date.
* Label encoded the 'creation\_source' feature.
* One-Hot Encoded the 'creation\_source' and 'org\_id' features.

Model selection –

The model selected was a decision tree. Decision trees are simple to understand and interpret, making them a good choice for this problem. Additionally, decision trees can handle both categorical and numerical features, which is important for this problem as we have both types of features.

Conclusion –

The analysis of user engagement and user information revealed that inactive duration is a crucial factor in predicting user adoption. Adopted users tend to have a shorter inactive duration compared to non-adopted users. Additionally the number of consecutive login parameter can give further clarity to business decision making in future