Summary of the Exploration:

The data exploration revealed distinct patterns in user engagement and adoption. Key findings include significant variances in the number of logins and account age days, which were linked to user adoption rates. The analysis underscored the importance of initial user engagement in predicting long-term platform adoption.

Preprocessing Steps:

Data Loading and Encoding: Data was loaded from CSV files with specific encoding to handle any special characters properly.
Timestamp Conversion: Timestamps were converted into proper datetime objects to facilitate time-based analysis.
Handling Missing Data: Missing values were filled with zeros, a choice made to maintain dataset integrity without introducing bias from mean/median imputation.

Feature Engineering Steps:

Time Features: Extracted day of the week, month, and year from the creation time to capture potential cyclical patterns in user sign-up behavior.
Interaction Features: Created a new feature 'marketing_interaction' by multiplying two boolean flags, which might reveal the combined effect of marketing strategies on user adoption.
Dummy Variables: Transformed 'creation_source' into dummy variables to quantify the impact of different signup sources on adoption.

Model Selection:

A Gradient Boosting Classifier was chosen for its robustness in handling imbalanced datasets and its effectiveness in capturing complex patterns through ensemble learning of decision trees. The use of a GridSearchCV helped in tuning hyperparameters like the number of estimators, max depth, and learning rate to optimize model performance.

Conclusion:

The analysis effectively identified key factors influencing user adoption and provided actionable insights through visualizations and predictive modeling. Adoption appears strongly influenced by the initial user engagement and marketing interactions. Future strategies could focus on enhancing early user experience and personalized marketing to increase adoption rates. The results and insights have been encapsulated in various visualizations and a detailed performance report of the predictive model, aiding in strategic decision-making.

Key Variables and Their Influence on User Adoption

1. User Engagement Metrics

Total Logins: There's a positive correlation between the total number of logins and user adoption. Frequent logins suggest increased engagement with the platform.

Activity Duration: The length of time between a user's first and last login can indicate their long-term interest in the services provided.

2. Account Characteristics

Account Age: The age of the account from creation to the present day. Older accounts tend to have higher adoption rates, possibly due to the accumulation of value over time.

Creation Source: The method through which the user signed up.

3. Marketing Interactions

Opted into Mailing List: Whether the user has opted into receiving marketing emails.

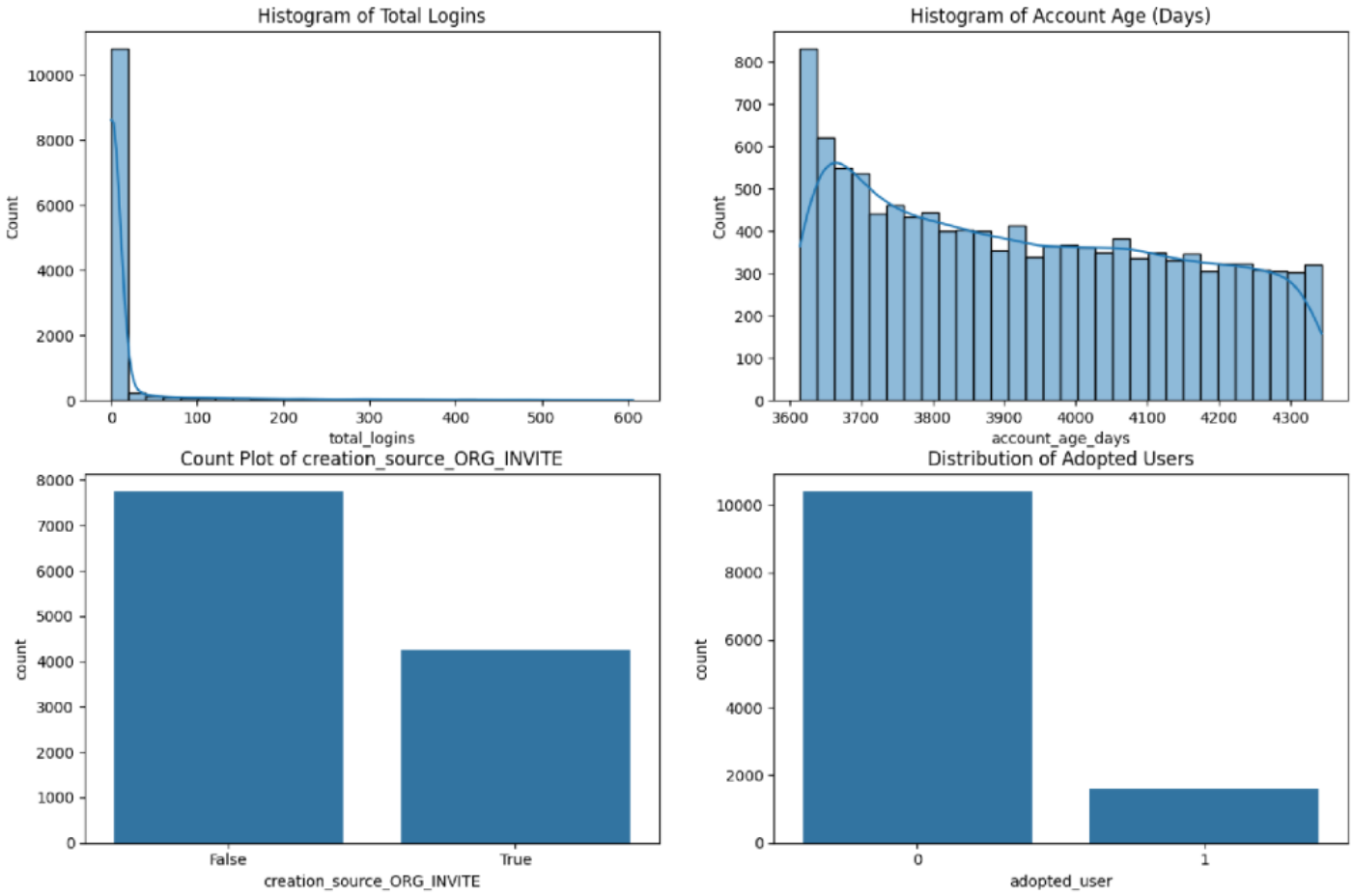
Enabled for Marketing Drip: Whether the user is part of a regular marketing email drip. These features might increase user engagement by keeping the platform top-of-mind.

Data Visualization

Histograms of Login Counts: Show the distribution of login counts among users to highlight how adopted users typically behave.

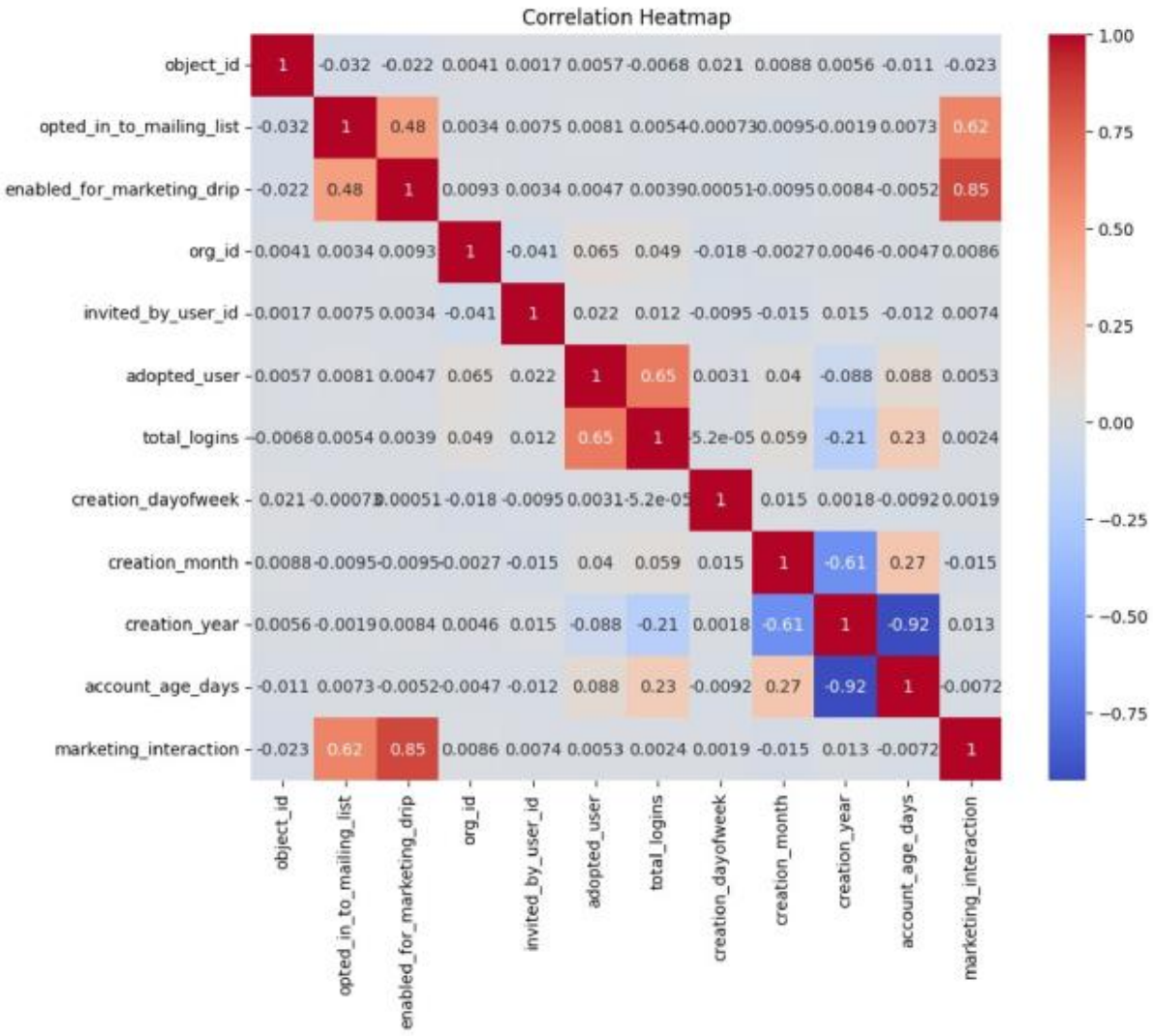
Bar Charts of Creation Sources: Illustrate which signup methods are most likely to lead to adoption.

Correlation Heatmap: Display the relationships between numeric variables to show how factors like account age and logins correlate with adoption.



Histogram & Bar Charts for Analysis

Correlation Heatmap




```
Feature Importances: [3.84478856e-03 1.54121723e-05 8.63452085e-06
3.23388525e-03
 2.85161249e-03 9.79265421e-01 3.46181412e-04 1.42054842e-03
 8.55740815e-05 8.85620766e-03 7.17346170e-05]
```

Results Accuracy of Model

```
Data saved to CSV at: final_users_data.csv
Best parameters: {'classifier__learning_rate': 0.1,
'classifier__max_depth': 3, 'classifier__n_estimators': 100}
Best cross-validation score: 0.98
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      2593
           1       0.95      0.93      0.94       407

    accuracy                           0.98      3000
   macro avg       0.97      0.96      0.97      3000
weighted avg       0.98      0.98      0.98      3000
```