✓ Summary of the exploration (4 sentences maximum), mentioning the findings.

Here I have used python tools to explore the data,

The adoption rate is around 98%

The key feature determining adoption rate are org_id,user_id,invited_by_userid

Model accuracy in predicting adoption is 97%

✓ Preprocessing steps taken (and possible reasons for the choices made)

I have used encoder for categorical feature

Dropped the unnecessary columns

Converted timestamp to unix for machine understanding

Dropped columns having high correlation

✓ Feature engineering steps taken (if at all; and reasons for the choices made)

I have created adoption column for each user which tell about the 3 logins in 7 days

✓ Model selection (if needed; and reasons for the choice made)

Choosed Random forest model for below reasons:

Data is multicollinearity and loss function has heteroscredesticity

Model giving high accuracy , so to prevent overfitting

Multiple columns present so pruning may occur

✓ Conclusion.

The data is highly imbalanced as non adoption is only 2%

Model is giving high accuracy of 97% , so it can easily predict future user adoption rate