The Colab notebook is found [here](https://github.com/BTExpress1/relax/blob/af4237c570146c5cf6dd2613677810eda1556e37/Relax.ipynb).

# Introduction

Relax Inc. makes productivity and project management software. We have two data sets that contain user signups and user activities. Relax considers users who log in three times or more in at least one seven-day period as adopters of the platform. This analysis aims to determine which factors help Relax predict future adopters.

# Analysis

After examining the two datasets, I decided to merge them into a single Pandas DataFrame, resulting in about 208 thousand rows. 18.16% of the users are adopters. Of all the features, the invited by user was the only one with null values. This is expected behavior since not all users sign up because of an invite. To make this feature useful, I converted it to a Boolean feature where 1 represents invited.

During data preparation, I dropped features from X that were not useful or could cause leakage and overfitting. The target feature y is a new Boolean column representing whether a user is an adopter (the 18.16%).

I chose Random Forest as the model to predict future adopters. The merged DataFrame is large, and I found no sequential dependency in the data. Therefore, I felt comfortable with the chosen algorithm. After splitting the data 80/20 for train/test, the results are promising, as seen in the validation metrics (screenshot below).

A screenshot of a computer

AI-generated content may be incorrect.

# Conclusion

At first glance, an accuracy of 99.5% suggested deeper analysis of the results to ensure overfitting is not present. A 90% class 0 recall confirms the model is not memorizing. The confusion matrix shows 208 non-zero false positives. F1 Score and ROC AUC highlight a balanced performance. The chart below shows the top list of feature importance. In conclusion, we have a good model, but I recommend future improvement with hyperparameter optimization.

A graph with numbers and text

AI-generated content may be incorrect.