

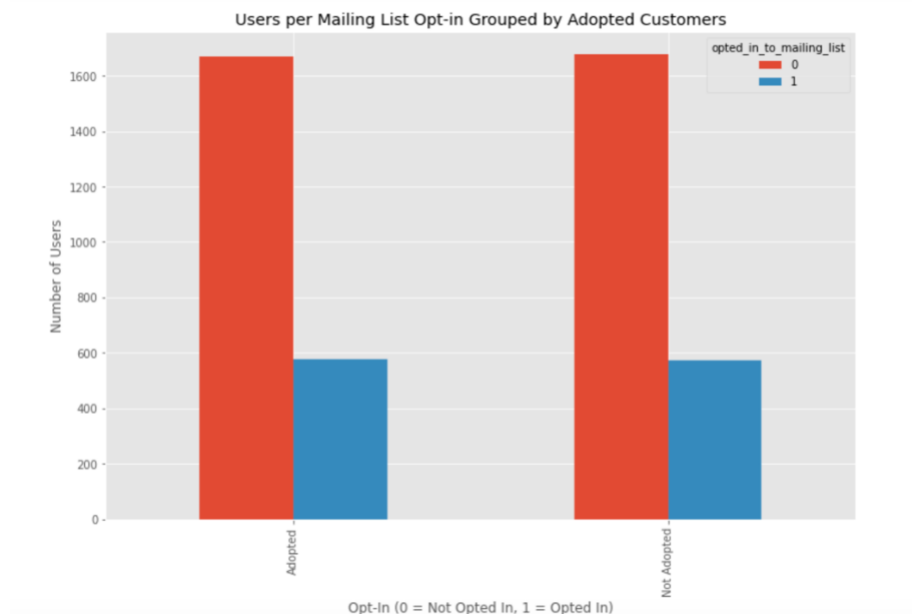
Practice Challenge 2

Relax Challenge Findings

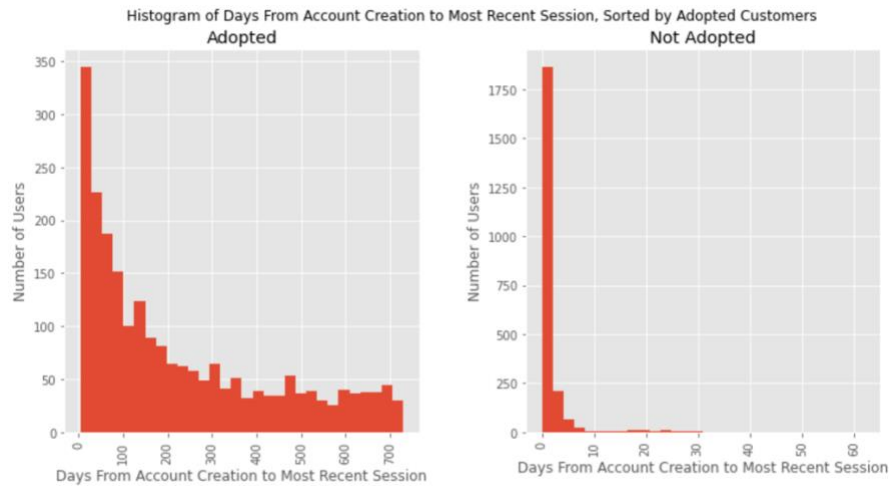
In this challenge, I used predictive modeling to determine which factors lead to user adoption, with adoption being defined as a user having logged into the product on at least three separate days in any seven-day period. I used rolling windows of seven days to find the user IDs of all individuals who met this criteria from the log-in data, and appended this information to the user information table as my dependent variable, assigning 1 for users who had been adopted and 0 for users not considered adopted.

In my exploration, I noticed that the unadopted users outnumbered the adopted users by more than four times, so I randomly sampled from a subset from the unadopted users to match the number of adopted users and create a dataset with an equivalent number of adopted and unadopted classes.

In further exploration I found that features like signing up for the mailing list or having marketing drip enabled did not make a difference in user adoption, having roughly the same distribution in both classes.



Account lifetime, on the other hand, defined as the number of days between the account creation date and the most recent log-in, was a strong indicator for user adoption. The longer the account lifetime, the more likely it was that the user was retained. Note that a long user lifetime does not presuppose that a user has logged in at least three times in a seven-day period. The user could, in theory, have logged in once a month, for example.



Although, as we can see from the distributions, there is certainly a relationship between users who have logged in more recently and those who have met the adoption criteria.

Due to the small dataset size (4496 rows total), I used Linear Support Vector Classification as my first model, achieving an accuracy score of 96.55% on the training data, and 96.44% on the testing data. The confusion matrix for the test data is as follows:

	Predicted Adopted (1)	Predicted Not Adopted (0)
Actual Adopted (1)	424	12
Actual Not Adopted (0)	20	444

The most important feature contributing to the split was unsurprisingly account lifetime (member days).

