**Write up: Relax Challenge.**

Data: We have two sets of data, one giving the login times and the second with all the features of the user id.

Data wrangling & EDA:

After setting the data types right, we sort the login times and map the number of users that have more than three separate logins. Next, we check for every user, if the logins are in a seven day period by parsing the time period into seven day periods.

Once we have the count of the users with three logins of three separate days in a seven day period, we create a new variable - ‘Adopted user’ and classify the user as one.

We then perform EDA on the two datasets. Our objective is to find a predictor of adopted users.

1. We perform Chi Square tests on the following categorical variables:

(opted\_in\_to\_mailing\_list),(adopted\_user)

(Creation\_source), (adopted\_user)

(enabled\_for\_marketing\_drip),(adopted\_user)

In all the three tests, the p-value is higher than 0.05, hence we accept the null hypothesis that none of these variables are correlated to adopted\_user.

1. We next check the login data for adopted and unadopted users. We create two subsets- one for adopted and the other for unadopted, to arrive at the following results.

Text

Description automatically generatedChart, box and whisker chart

Description automatically generated

A box plot for the adopted users reveals that the median is far lower than the average and we have quite a few outliers. So we analyze the top 20 adopted users with the highest logins.

Chart, bar chart

Description automatically generated

For unadopted users, we find.

Text

Description automatically generated with medium confidence

While the minimum number of logins for unadopted and adopted is not a huge difference, we notice the maximum and the avg logins for these two sets of users has a huge difference.

The adopted user logins in on an average 89 times more than the unadopted user.

The highest usage by an adopted user is 39 times more than the highest usage by an unadopted user.

While this activity of login itself could be an indicator of adoption and a predictor of adopted users. We also notice that the max and average are affected by the outliers as the box plot indicates that the median is below 100.So the login number in adopted users, has a high variance. Yet, we clearly see a difference in the number of logins, which can be used as a predictor.