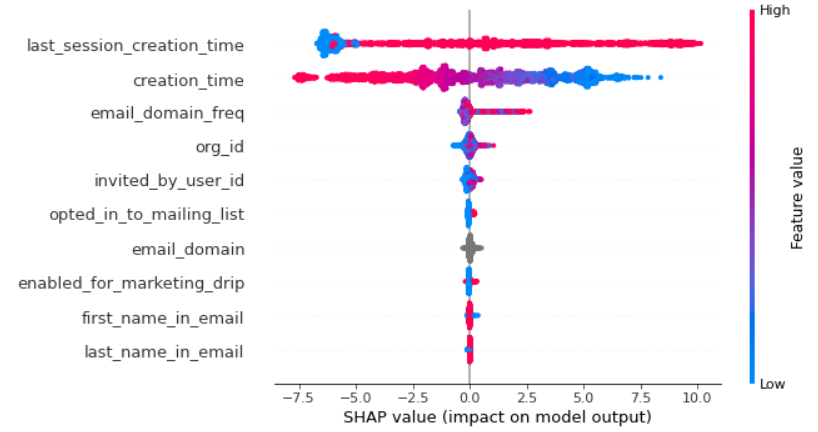
After exploratory analysis, I decided to investigate feature importance using CatBoost, because there are many categorical features here and CatBoost is particularly designed to accommodate those. Naturally, there are many other options that could be examined with more time.

A simple CatBoost model achieves 97.5% accuracy in predicting adopted users. Feature importance results can be seen in the following graph:



From this we see that by far the most important features are last\_session\_creation\_time followed by creation\_time. This graph (along with a simple calculation of the correlation itself, -0.088) suggests that creation time is negatively correlated with adoption, which is a concerning trend and bears further analysis. However, this could just be due to more recent users not having as much data yet to show they’re adopting.

Email\_domain\_freq, which is just the count of how many times that email domain appears in the data, and org\_id have some small effect. Further investigation with the goal of developing targeted marketing could be fruitful. Data matching org\_id to industry might be helpful here.

Last\_session\_creation\_time is the most influential factor. This suggests that adopted users tend to keep using the product and therefore will have used it recently. However, it’s hard to say that’s useful for prediction in the future. One note is that the big blob at the left should be users who never logged in – unsurprisingly, they do not adopt. This is a significant number of users, so investigating why they register but never log in could be worthwhile.

Opted\_in\_to\_mailing\_list and enabled\_for\_marketing\_drip have marginal effects. If their goal is to encourage adoption, success appears very limited.