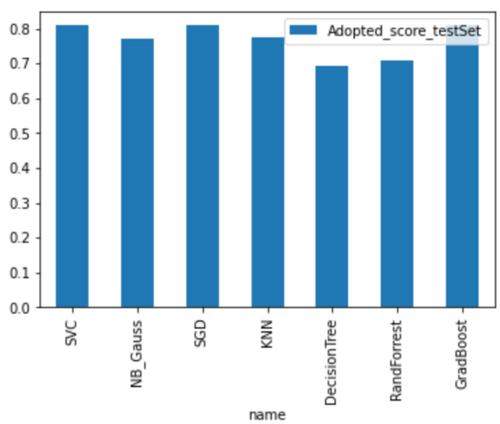
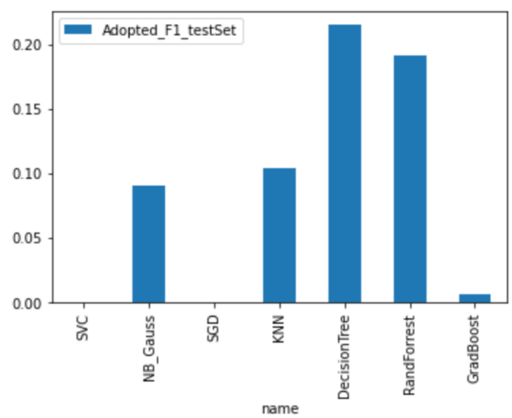
After testing the data set with various Machine Learning models, it was clear that the features provided for each user do not have adequate predictive power to identify user adoption.

Process Followed:

- Import Data
- Feature Engineering
    - User ‘Adopted’, User invited, Number of users invited by this user, Length of time in system
- Normalize Features
- Generate Classification Models (SVC, GaussianNB, SGD, KNN, Decision Tree, Random Forest, Gradient Boosting)
- Score Classification Models (Average Precision, F1 Score, Model Score Function)
- Determine Feature Importances

Results:





Typical Classification Report

	precision	recall	f1-score	support
Not Adopted	0.83	0.89	0.86	1761
Adopted	0.20	0.12	0.15	376
accuracy			0.76	2137
macro avg	0.51	0.51	0.50	2137
weighted avg	0.72	0.76	0.73	2137

Conclusion

The features in the data provided produced models that were unable to accurately predict user adoption. Classification reports of several models show that the false positives and the false negatives of the models outweigh the true positives for predicted Adoption. Further metrics should be gathered on users such as age, gender, race, income, education, marital status, number of children, employment status, or geographic location.

For what it is worth, the best predictive variable is creation time, but this feature is not independent of the output variable of Adoption, which is three separate days of logins in a seven day period. Users who have a later creation time are clearly less likely to have Adopted. However, this is not useful to the business who are looking for ways to influence their customers to Adopt, and user sign up date is not a feature they can alter for existing customers.