Writeup Over Findings in the process and approach used

- Started with the loading of datasets from takehome_user and taken_user_engagemnt files.
- It is observed that there are around 1656 active users (meaning users which are activated in the last 7 days) in the dataset which was around 14% of the total users.
- Now I found out how long the user has been active and named that field as usage length
- After removing unnecessary fields, now we have a small feature set and sample size, so the best ML models for these types of datasets are Knn and Random Forest Classifiers from previous experience.
- After training with Random Forest we get accuracy of 97% and after plotting feature importance according to Random Forest Classifier usage_lenght in most important feature.
- Below plot shows the Empirical Distribution function with respect to Length of Usage

Length of Usage b/w Adopted and Not Adpoted Users

-