# Project scratch pad

## Main Idea:

Design a machine learning model that determines whether or not a customer might accept an offer given the offer details.

Find for each offer which demographics are the most probable to accept that offer.

Questions:

* A histogram showing the number of offers received by different persons divided by some demographics of these persons -> Which groups of people more probably receive offers in general.
* For each type of offer plot the numbers of people receiving that offer based on age, income .. etc to see which group of people is more likely to receive a certain offer.
* Should I remove people with None in age or just predict and impute their ages ?
* Older vs Newer users.
* We need to divide the users into segments or groups based on either their age or their incomes etc.
* We can then try to impute the missing incomes and genders based on which groups they are closer to. (divide the people with gender and income information available into age groups and check whether they have close attributes or not if so we can impute the values of the rest of the people based on these groups and their age – we don’t have none ages.
* Why not replace all hashable ids in the three tables by numbers to make it easier (don’t know If that is even necessary todo).
* Maybe do a scatter plot first based on age and income and colored based on age.