**Seedbox Data Science Application Test**

In this test, we would like you to write up an analysis about a mock experiment we've run at Seedbox. We aim to measure your current knowledge in Data Cleaning, Exploratory Data Analysis and Drawing Conclusions from observations.

**Experimentation Details**

We recently ran an A/B test on the cancellation page of our subscription service. Before running the test, members where able to cancel using a simple web form. The experiment aims measure the impact of forcing members to phone-in to our customer service line in order to cancel.

Information about the test:

* control group can cancel using a web form
* test group can only cancel by calling in
* Users were randomly assigned to a group when they go to the websites cancel page for the first-time
* The distribution probability between both groups is uneven (you can see this like an unfair coinflip)
* We've recorded additional transactions generated after users were randomized
* REBILLs are Transactions recurring payments that were processed
* CHARGEBACKs or REFUNDs transactions represent payments that were cancelled

**Information About the Data**

You will find the required data-sets for this analysis in the current git repo. This is split in the following 2 csv files:

* [testSamples.csv](https://github.com/seedboxtech/datasciencetest/blob/master/testSamples.csv)
* [transData.csv](https://github.com/seedboxtech/datasciencetest/blob/master/transData.csv)

In testSamples.csv, you will find a list of unique users that were randomized in the A/B test.

* sample\_id : is the unique identifier for the sample
* test\_group : is the group in which the sample was placed, 0= control group, 1=test group

In transData.csv, you will find a list of transactions generated by randomized users after their randomization:

* transaction\_id : is the unique identifier for the transaction
* sample\_id : is a foreign key that links transactions to test samples
* transaction\_type : is the transaction type for a transaction, can be REBILL, CHARGEBACK or REFUND
* transaction\_amount : is the amount generated for a transaction, this can be a negative value

**Analysis Requirements**

In this analysis we would like you to answer the following questions:

1. What is the approximate probability distribution between the test group and the control group
2. Is a user that must call-in to cancel more likely to generate at least 1 addition REBILL?
3. Is a user that must call-in to cancel more likely to generate more revenues?
4. Is a user that must call-in more likely to produce a higher chargeback rate(CHARGEBACKs/REBILLs)?

Technical Requirements:

1. Analysis must be coded in R or Python
2. Analysis must be submitted to a github repository
3. Analysis must be written in markdown format
4. Please include at least 1 vizualization with your analysis
5. Please use statistical significance tools to answer the questions we've asked
6. Include the code you used to perform the analysis in the github repository
7. Send us the link to your repository once you've completed the analysis