Engagement Test

Goal

Many sites make money by selling ads. For these sites, the number of pages visited by users on each session is one of the most important metric, if not **the most important** metric.

Data science plays a huge role here, especially by building models to suggest personalized content. In order to check if the model is actually improving engagement, companies then run A/B tests.

It is often data scientist responsibility to analyze test data and understand whether the model has been successful. The goal of this project is to look at A/B test results and draw conclusions.

Challenge Description

The company of this exercise is a social network. They decided to add a feature called: Recommended Friends, i.e. they suggest people you may know.

A data scientist has built a model to suggest 5 people to each user. These potential friends will be shown on the user newsfeed. At first, the model is tested just on a random subset of users to see how it performs compared to the newsfeed without the new feature.

The test has been running for some time and your boss asks you to check the results. You are asked to check, for each user, the number of pages visited during their first session since the test started. If this number increased, the test is a success.

Specifically, your boss wants to know:

- Is the test winning? That is, should 100% of the users see the Recommended Friends feature?
- Is the test performing similarly for all user segments or are there differences among different segments?
- If you identified segments that responded differently to the test, can you guess the reason? Would this change your point 1 conclusions?

Data

We have 2 tables downloadable by clicking **here**.

The 2 tables are:

```
"user_table" - info about each user sign-up date
```

Columns:

- **user_id** : the Id of the user. It is unique by user and can be joined to user id in the other table
- **signup_date** : when the user joined the social network

```
    "test_table" - data about the test results. For each user, we only consider
her first session since the date when the test started. That is, if the test
started on Jan 1, and user 1 visited the site on Jan, 2 and Jan, 3, we only care
about how many pages she visited on Jan, 2.
```

Columns:

- **user_id** : the Id of the user
- **date** : the date of the first session since the test started
- **browser** : user browser during that session
- **test**: 1 if the user saw the new feature, 0 otherwise
- **pages_visited**: the metric we care about. # of pages visited in that session

Example

```
    Let's check the first row:
```

head(test_table,1)

Column Name	Value	Description
user_id	600597	this is the user
date	2015-08-13	her first session since the test started was on Aug, 13.
browser	IE	she used Internet Explorer to visit the site
test	0	she was in the control group, i.e. didn't see the new feature
pages_visited	2	she visited 2 pages during that session and then left the site

---

```
Let's check when the user Id from the previous table joined the site:
```

subset (user_table, user_id== 600597)

Column Name	Value	Description
user_id	600597	this is the user we care about
signup_date	2015-01-19	she created her account on Jan, 19.