A/B Testing in R

The goal of this experiment is to determine if a company should change to a newly designed page or stick with the existing one.

You can find the data used here on Sadiq Alreemi's Github page.

The dplyr library is used in this experiment

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Let's load the data in

```
df<- read.csv('ab_data.csv')
```

quick structure of the data

```
str(df)
```

```
## 'data.frame':    294478 obs. of  5 variables:
##  $ user_id     : int  851104 804228 661590 853541 864975 936923 679687 719014 817355 839785 ...
##  $ timestamp   : chr  "2017-01-21 22:11:48.556739" "2017-01-12 08:01:45.159739" "2017-01-11 16:55:06
##  $ group       : chr  "control" "control" "treatment" "treatment" ...
##  $ landing_page: chr  "old_page" "old_page" "new_page" "new_page" ...
##  $ converted   : int  0 0 0 0 1 0 1 0 1 1 ...
```

how exactly does the data look

```
head(df)
```

```
##   user_id                   timestamp     group landing_page converted
## 1  851104 2017-01-21 22:11:48.556739   control     old_page         0
## 2  804228 2017-01-12 08:01:45.159739   control     old_page         0
## 3  661590 2017-01-11 16:55:06.154213 treatment     new_page         0
## 4  853541 2017-01-08 18:28:03.143765 treatment     new_page         0
## 5  864975 2017-01-21 01:52:26.210827   control     old_page         1
## 6  936923 2017-01-10 15:20:49.083499   control     old_page         0
```

Filter data for control group and old landing page

```
control_old <- df %>%
  filter(group == "control", landing_page == "old_page")
```

```
head(control_old)
```

```
##   user_id                   timestamp   group landing_page converted
## 1  851104 2017-01-21 22:11:48.556739 control     old_page         0
## 2  804228 2017-01-12 08:01:45.159739 control     old_page         0
## 3  864975 2017-01-21 01:52:26.210827 control     old_page         1
## 4  936923 2017-01-10 15:20:49.083499 control     old_page         0
## 5  719014 2017-01-17 01:48:29.539573 control     old_page         0
## 6  644214 2017-01-22 02:05:21.719434 control     old_page         1
```

Filter data for treatment group and new landing page

```
treatment_new <- df %>%
  filter(group == "treatment", landing_page == "new_page")

head(treatment_new)
```

```
##   user_id                   timestamp     group landing_page converted
## 1  661590 2017-01-11 16:55:06.154213 treatment     new_page         0
## 2  853541 2017-01-08 18:28:03.143765 treatment     new_page         0
## 3  679687 2017-01-19 03:26:46.940749 treatment     new_page         1
## 4  817355 2017-01-04 17:58:08.979471 treatment     new_page         1
## 5  839785 2017-01-15 18:11:06.610965 treatment     new_page         1
## 6  929503 2017-01-18 05:37:11.527370 treatment     new_page         0
```

check where the treatment group do not get new page or the control group do not get the old page

```
 treatment_new %>%
  filter(group == "treatment", landing_page == "old_page")
```

```
## [1] user_id      timestamp    group        landing_page converted
## <0 rows> (or 0-length row.names)
```

```
control_old %>%
  filter(group == "control", landing_page == "new_page")
```

```
## [1] user_id      timestamp    group        landing_page converted
## <0 rows> (or 0-length row.names)
```

now we can calculate the conversion rate for each group

```
conversion_rate_control <- mean(control_old$converted)
conversion_rate_treatment <- mean(treatment_new$converted)
```

perform the independent t-test

H0 (null hypothesis) - there is no significant difference between the conversion rate of the control and treatment groups H1 (alternative hypothesis) - there is a significant difference between the conversion rate of the control and treatment groups

```
t_test_result <- t.test(control_old$converted, treatment_new$converted)
```

display the results

```
cat("Conversion Rate (Control):", conversion_rate_control, "\n")
```

```
## Conversion Rate (Control): 0.1203863
```

```
cat("Conversion Rate (Treatment):", conversion_rate_treatment, "\n")
```

```
## Conversion Rate (Treatment): 0.1188072
```

```r
cat("p-value:", t_test_result$p.value, "\n")
```

```
## p-value: 0.1896542
```

we can already see that the p-value we got is greater than 0.05. However, to make this interesting;

```r
if (t_test_result$p.value < 0.05) {
  cat("Statistically significant - reject the null hypothesis\n")
} else {
  cat("Not statistically significant - fail to reject the null hypothesis\n")
}
```

```
## Not statistically significant - fail to reject the null hypothesis
```

From the experiment above, we can recommend the company stick with the existing page as there is no significant difference between thhe conversion rate of both pages.