

## ab\_conversions

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

```
# Checks control and treatment users
result <- table(df$group, df$landing_page)
print(result)
```

```
##
##           new_page old_page
## control      1928   145274
## treatment  145311    1965
```

```
library(dplyr)
```

```
##
## 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
```

```
# Removes users that saw the wrong page
df_cleaned <- df %>%
  filter(group == "control" & landing_page == "old_page" |
         group == "treatment" & landing_page == "new_page")
head(df_cleaned)
```

```
##   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
```

```
# Checks for duplicate rows
duplicates <- df_cleaned[duplicated(df_cleaned),]
print(duplicates)
```

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

```
# Double checking for duplicates
if (any(duplicated(df_cleaned))) {
  print("There are duplicates in the cleaned data.")
} else {
  print("No duplicates found in the cleaned data.")
}
```

```
## [1] "No duplicates found in the cleaned data."
```

```
# Checks proportions of users who converted by group
conversion_table <- table(df_cleaned$group, df_cleaned$converted)
conversion_proportions <- prop.table(conversion_table, margin = 1)

print(conversion_proportions)
```

```
##
##              0          1
## control  0.8796137 0.1203863
## treatment 0.8811928 0.1188072
```

```
# Chi-squared test
chi_test_result <- chisq.test(conversion_table)
print(chi_test_result)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  conversion_table
## X-squared = 1.7054, df = 1, p-value = 0.1916
```

The p-value of 0.1916 is greater than the common significance level of 0.05. As a result, there isn't enough evidence to reject the null hypothesis. In this context, the null hypothesis suggests that there is no significant association between the 'group' and 'converted' variables, indicating that the observed data is consistent with the assumption of no association.