A/B Testing Analysis

Problem Statement: Evaluating the Effectiveness of Localized Translations in Improving Conversion Rates

Background
A multinational company operating in Latin American (LATAM) countries recently implemented a localized translation feature for its platform, aimed at enhancing user experience and improving conversion rates. Spain, which already had a localized version, serves as a baseline for high conversion rates. An A/B test was conducted to evaluate the impact of the new translations on user behavior in other LATAM countries.

Loading Required Libraries

```
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
```

```
library(rpart)
library(ggplot2)
```

Importing Data

```
user <- read.csv("user_table.csv")
test <- read.csv("test_table.csv")
```

Check for Duplicate IDs

```
length(unique(test$user_id)) == length(test$user_id) # Check duplicates in test table
```

```
## [1] TRUE
```

```
length(unique(user$user_id)) == length(user$user_id) # Check duplicates in user table
```

```
## [1] TRUE
```

Check Missing IDs

```
length(user$user_id) - length(test$user_id) # Check for missing IDs between tables
```

```
## [1] -454
```

Comment: Some IDs are missing in the `user` table. When joining, we ensure not to lose the IDs in the `test` table.

Merging Data

```
data <- merge(test, user, by = "user_id", all.x = TRUE) # Keep all IDs in the test table
data$date <- as.Date(data$date)
summary(data)
```

```
##     user_id             date               source             device
##  Min.   :      1   Min.   :2015-11-30   Length:453321      Length:453321
##  1st Qu.: 249816   1st Qu.:2015-12-01   Class :character   Class :character
##  Median : 500019   Median :2015-12-03   Mode  :character   Mode  :character
##  Mean   : 499938   Mean   :2015-12-02
##  3rd Qu.: 749522   3rd Qu.:2015-12-04
##  Max.   :1000000   Max.   :2015-12-04
##
##  browser_language   ads_channel          browser            conversion
##  Length:453321      Length:453321      Length:453321      Min.   :0.00000
##  Class :character   Class :character   Class :character   1st Qu.:0.00000
##  Mode  :character   Mode  :character   Mode  :character   Median :0.00000
##                                                           Mean   :0.04958
##                                                           3rd Qu.:0.00000
##                                                           Max.   :1.00000
##
##       test            sex                 age          country
##  Min.   :0.0000   Length:453321      Min.   :18.00   Length:453321
##  1st Qu.:0.0000   Class :character   1st Qu.:22.00   Class :character
##  Median :0.0000   Mode  :character   Median :26.00   Mode  :character
##  Mean   :0.4764                      Mean   :27.13
##  3rd Qu.:1.0000                      3rd Qu.:31.00
##  Max.   :1.0000                      Max.   :70.00
##                                      NA's   :454
```

Country Conversion Analysis

Check Conversion Rates by Country

```
data_conversion_country <- data %>%
  group_by(country) %>%
  summarize(conversion = mean(conversion[test == 0])) %>%  # Old version
  arrange(desc(conversion))

head(data_conversion_country)
```

```
## # A tibble: 6 x 2
##   country     conversion
##   <chr>            <dbl>
## 1 Spain           0.0797
## 2 <NA>            0.0776
## 3 El Salvador     0.0536
## 4 Nicaragua       0.0526
## 5 Costa Rica      0.0523
## 6 Colombia        0.0521
```

Comment: Spain converts much better than the rest of LATAM countries.

Hypothesis Testing

Exclude Spain and Perform Proportion Test

```
data_test <- subset(data, country != "Spain")
prop.test(table(data_test$conversion, data_test$test), correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  table(data_test$conversion, data_test$test)
## X-squared = 54.491, df = 1, p-value = 1.562e-13
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03524553 -0.02042671
## sample estimates:
##    prop 1    prop 2
## 0.4607531 0.4885892
```

Perform T-Test

```
t.test(data_test$conversion[data_test$test == 1], data_test$conversion[data_test$test == 0])
```

```
##
##  Welch Two Sample t-test
##
## data:  data_test$conversion[data_test$test == 1] and data_test$conversion[data_test$test == 0]
## t = -7.3539, df = 385258, p-value = 1.929e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.006181421 -0.003579837
## sample estimates:
##  mean of x  mean of y
## 0.04341116 0.04829179
```

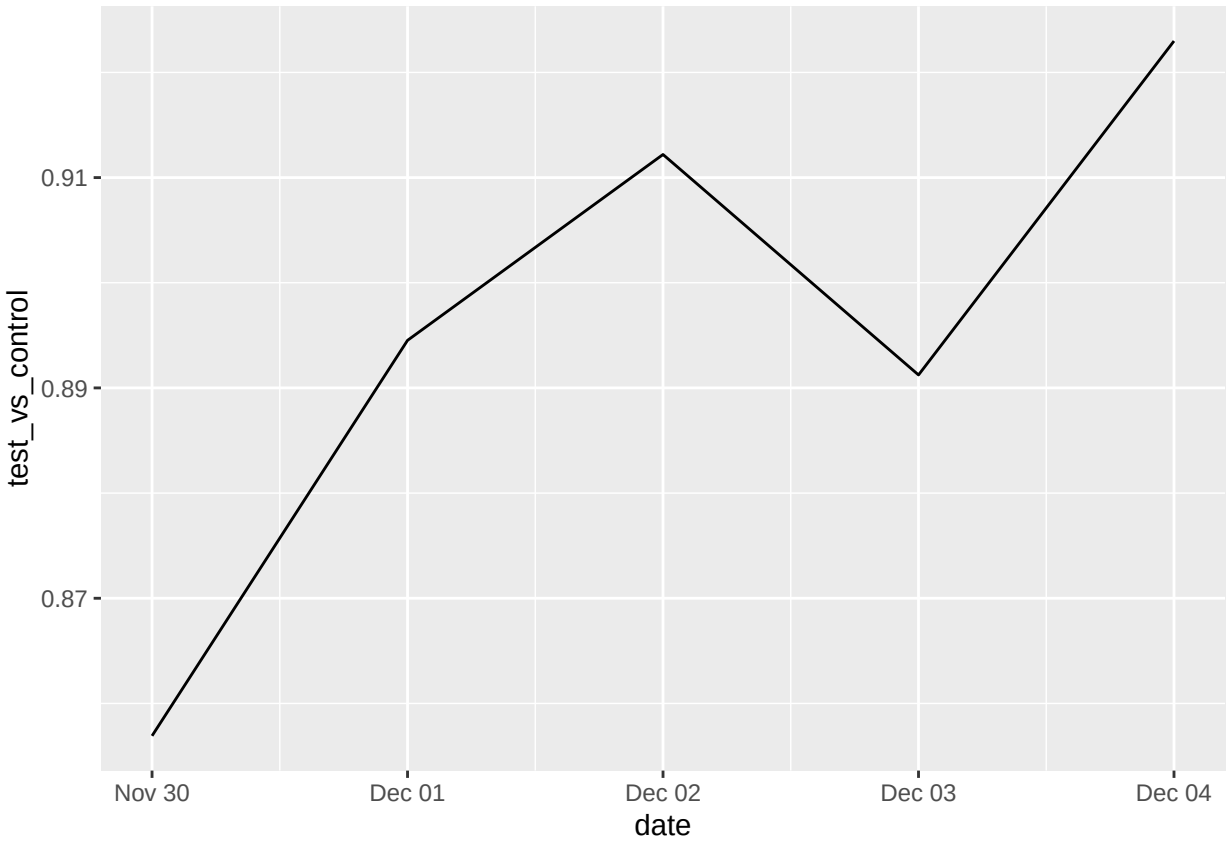
Comment: Test users are converting at a lower rate than control users. Possible reasons include insufficient data or biases in experiment setup.

Plot Test-Control Ratios Over Time

```
data_test_by_day <- data_test %>%
  group_by(date) %>%
  summarize(test_vs_control = mean(conversion[test == 1]) / mean(conversion[test == 0]))

qplot(date, test_vs_control, data = data_test_by_day, geom = "line")
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Observations: 1. Test consistently performs worse than control. 2. Experiment ran for only 5 days; longer testing is recommended to capture weekly patterns.

Check Randomization Using Decision Tree

```
head(data_test)
```

```
##   user_id       date source device browser_language ads_channel     browser
## 1       1 2015-12-02    SEO    Web               EN        <NA>      Chrome
## 2       2 2015-11-30    SEO Mobile               ES        <NA> Android_App
## 3       3 2015-12-03    SEO Mobile               ES        <NA>  Iphone_App
## 4       5 2015-11-30    Ads    Web               ES    Facebook      Chrome
## 5       8 2015-12-03    Ads Mobile               ES      Google Android_App
## 6      11 2015-12-03    Ads    Web               ES       Yahoo      Chrome
##   conversion test sex age   country
## 1          0    0   M  38     Chile
## 2          0    0   M  27  Colombia
## 3          0    1   M  18 Guatemala
## 4          0    1   M  22 Argentina
## 5          0    1   M  19 Venezuela
## 6          0    1   F  28  Colombia
```

```
tree <- rpart(test ~ ., data = data_test[, -8],
              control = rpart.control(minbucket = nrow(data_test) / 100, maxdepth = 2))
tree
```

```
## n= 401085
##
```

```
## node), split, n, deviance, yval
##       * denotes terminal node
##
## 1) root 401085 99692.820 0.5379757
##   2) country=Bolivia,Chile,Colombia,Costa Rica,Ecuador,El Salvador,Guatemala,Honduras,Mexico,Nicaragu
##   3) country=Argentina,Uruguay 50867  7894.097 0.8079108 *
```

Comment: Randomization appears perfect for some countries but biased for others like Argentina and Uruguay.

Country-Specific Analysis

```
data_test_country <- data_test %>%
  group_by(country) %>%
  summarize(
    p_value = prop.test(table(conversion, test), correct = FALSE)$p.value,
    conversion_test = t.test(conversion[test == 1], conversion[test == 0])$estimate[1],
    conversion_control = t.test(conversion[test == 1], conversion[test == 0])$estimate[2]
  ) %>%
  arrange(p_value)

data_test_country
```

```
## # A tibble: 16 x 4
##    country     p_value conversion_test conversion_control
##    <chr>         <dbl>           <dbl>              <dbl>
##  1 Mexico        0.166          0.0512             0.0495
##  2 El Salvador   0.248          0.0479             0.0536
##  3 Chile         0.303          0.0513             0.0481
##  4 Argentina     0.322          0.0137             0.0151
##  5 Colombia      0.424          0.0506             0.0521
##  6 Honduras      0.472          0.0475             0.0509
##  7 Guatemala     0.572          0.0486             0.0506
##  8 Venezuela     0.574          0.0490             0.0503
##  9 Costa Rica    0.688          0.0547             0.0523
## 10 Panama        0.705          0.0494             0.0468
## 11 Bolivia       0.719          0.0479             0.0494
## 12 Peru          0.772          0.0506             0.0499
## 13 Nicaragua     0.780          0.0542             0.0526
## 14 Uruguay       0.883          0.0129             0.0120
## 15 Paraguay      0.884          0.0492             0.0485
## 16 Ecuador       0.962          0.0490             0.0492
```

Comment: Argentina and Uruguay show low conversion rates in both test and control groups, demonstrating Simpson's paradox.

Remove Argentina and Uruguay, Reanalyze

```
data_test <- subset(data, !(country %in% c("Spain", "Argentina", "Uruguay")))
unique(data_test$country)
```

```
##  [1] "Chile"       "Colombia"    "Guatemala"   "Venezuela"   "Paraguay"
##  [6] "Mexico"      "Peru"        "Bolivia"     "Ecuador"     "Panama"
## [11] "Honduras"    "Nicaragua"   "El Salvador" "Costa Rica"  NA
```

```r
# Proportion Test
prop.test(table(data_test$conversion, data_test$test), correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  table(data_test$conversion, data_test$test)
## X-squared = 0.10119, df = 1, p-value = 0.7504
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.006342567  0.008800236
## sample estimates:
##    prop 1    prop 2
## 0.5013422 0.5001134
```

```r
# T-Test
 t.test(data_test$conversion[data_test$test == 1], data_test$conversion[data_test$test == 0])
```

```
##
##  Welch Two Sample t-test
##
## data:  data_test$conversion[data_test$test == 1] and data_test$conversion[data_test$test == 0]
## t = 0.3181, df = 350651, p-value = 0.7504
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001212034  0.001681678
## sample estimates:
##  mean of x  mean of y
## 0.05042113 0.05018631
```

Conclusion: After removing Argentina and Uruguay, the localized translation does not significantly impact conversion rates. While not a success, it confirms that the test did not negatively affect the overall performance.