E-News Express A/B Testing Analysis

By Emilie Helen Wolf

Content

1. Problem Statement and Solution Approach
2. Data Overview
3. Exploratory Data Analysis
4. Insights of Inferential Statistics
5. Recommendations

Problem Statement and Solution Approach

Expand the business by attracting and acquiring new subscribers based on their interests

1. Create a new landing page and perform A/B Testing
2. Collect and explore the data from the two random samples
3. Analyze users' responses and infer whether or not the new landing page is effective

Dataset File

abtest.csv

Dataset Columns

1. **user_id** - This represents the user ID of the person visiting the website.
2. **group** - This represents whether the user belongs to the first group (control) or the second group (treatment).
3. **landing_page** - This represents whether the landing page is new or old.
4. **time_spent_on_the_page** - This represents the time (in minutes) spent by the user on the landing page.
5. **converted** - This represents whether the user gets converted to a subscriber of the news portal or not.
6. **language_preferred** - This represents the language chosen by the user to view the landing page.

- The randomly selected control and treatment groups are divided equally
  - 50 users in each sample
  - 16 English, 17 French, 17 Spanish

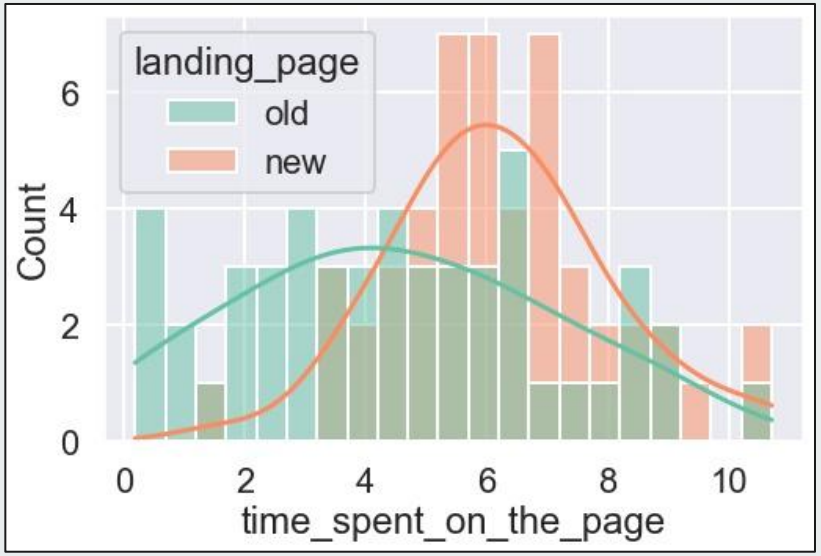
- 1 continuous variable (time spent on page) and 4 discrete variables (converted, language, landing page, group)

- All 100 randomly selected users have unique user IDs (thus independent)

Data is tidy and complete with no manipulations needed other than splitting the two groups

Exploratory Data Analysis

Univariate Plots



Histogram of Time Spent on Page

Old page had a mean time of **4.5 minutes** with a standard deviation of 2.6 minutes.

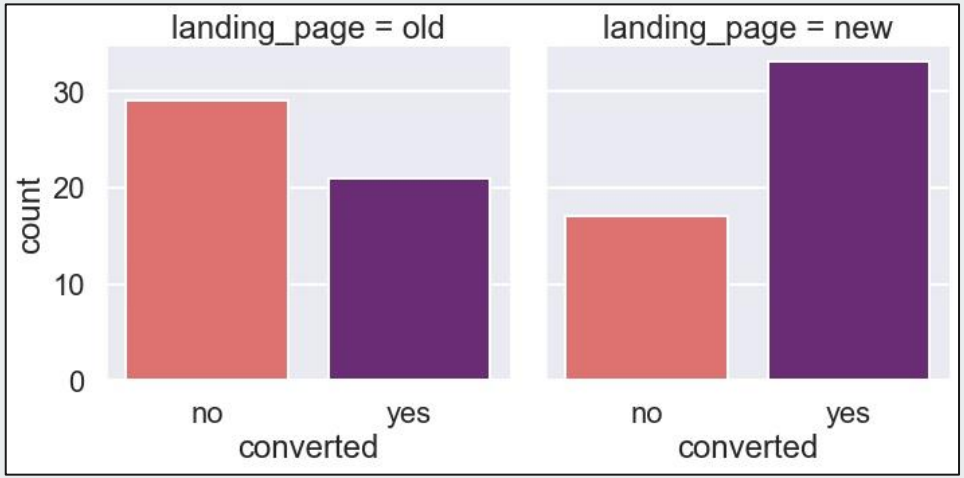
New page had a mean time of **6.2 minutes** with a standard deviation of 1.8 minutes.

Both resemble the **Normal Distribution**.

Histogram of Converted Status

Old page conversion rate: **0.42**

New page conversion rate: **0.66**

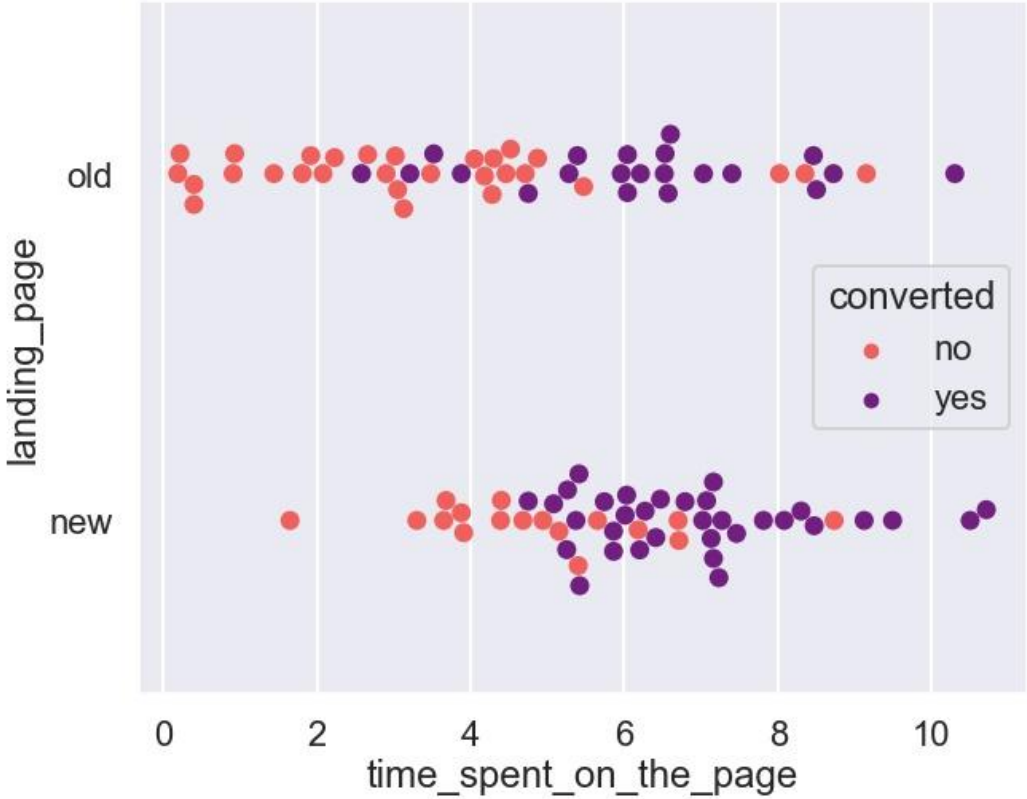


Exploratory Data Analysis

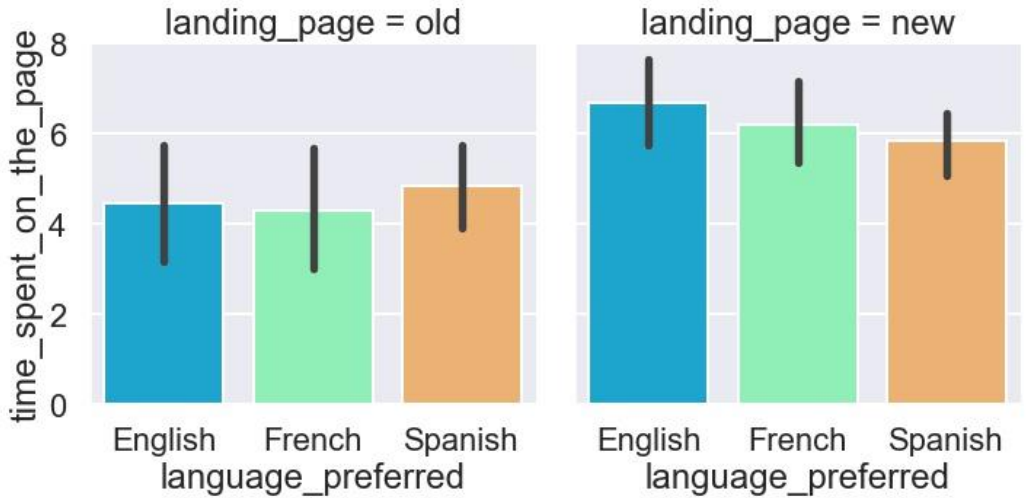
Multivariate Plots

Time Spent on Page and Conversion Status

On both the old and new landing pages, most conversions appear to have come from users who spent more time on the page.



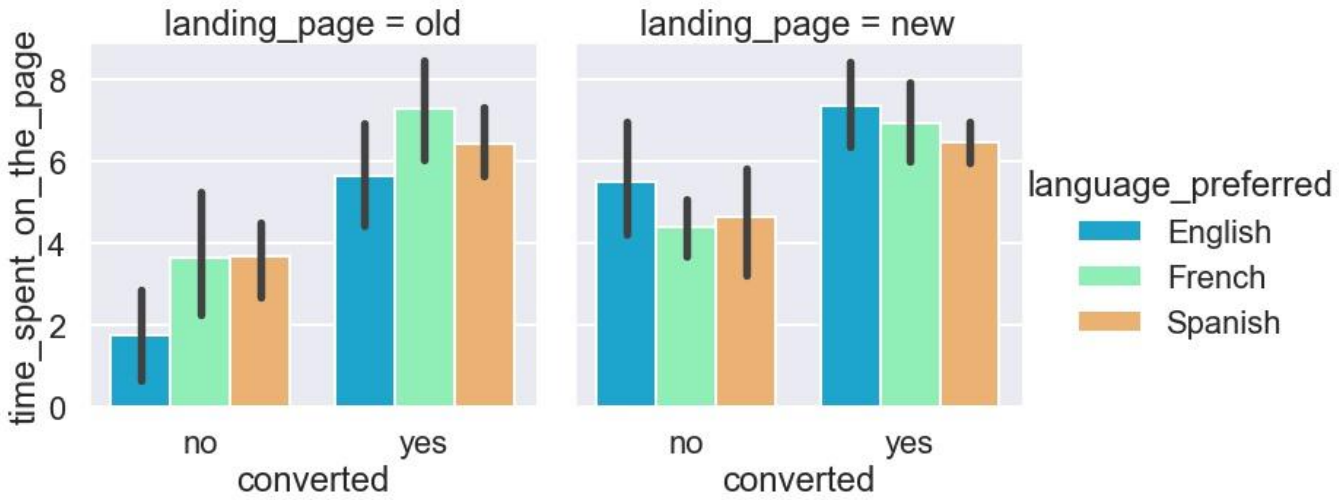
Average Time Spent on Page by Language Preferred



The average times spent on the page in both groups seems fairly even across all 3 languages.

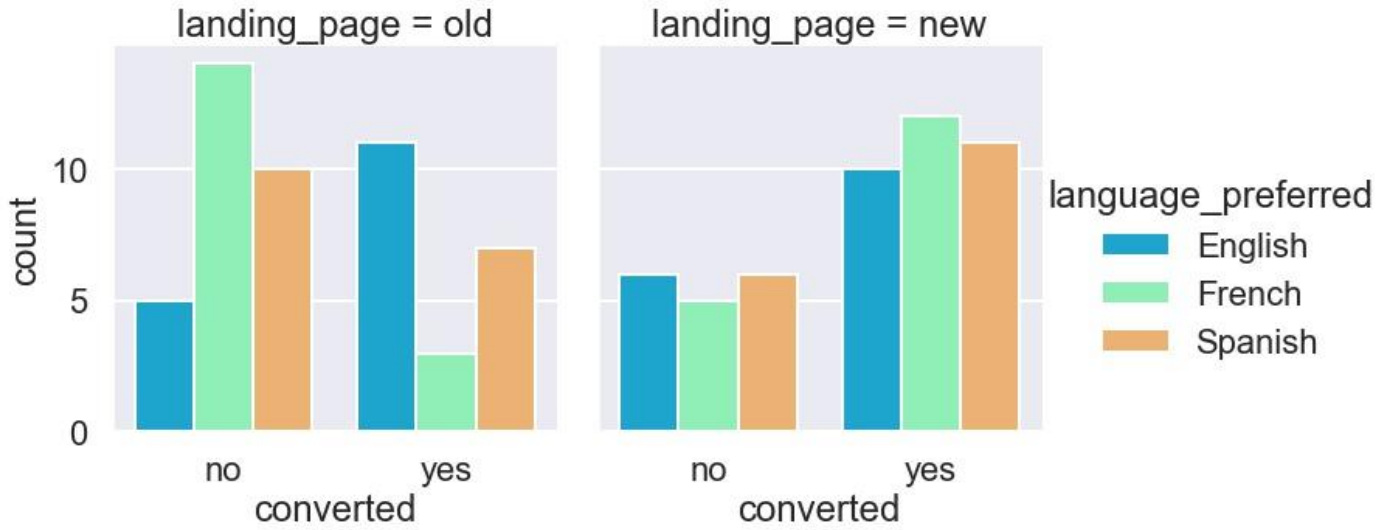
Average Time split by Language and Conversion

On the old page, English readers who did not convert had the shortest times.



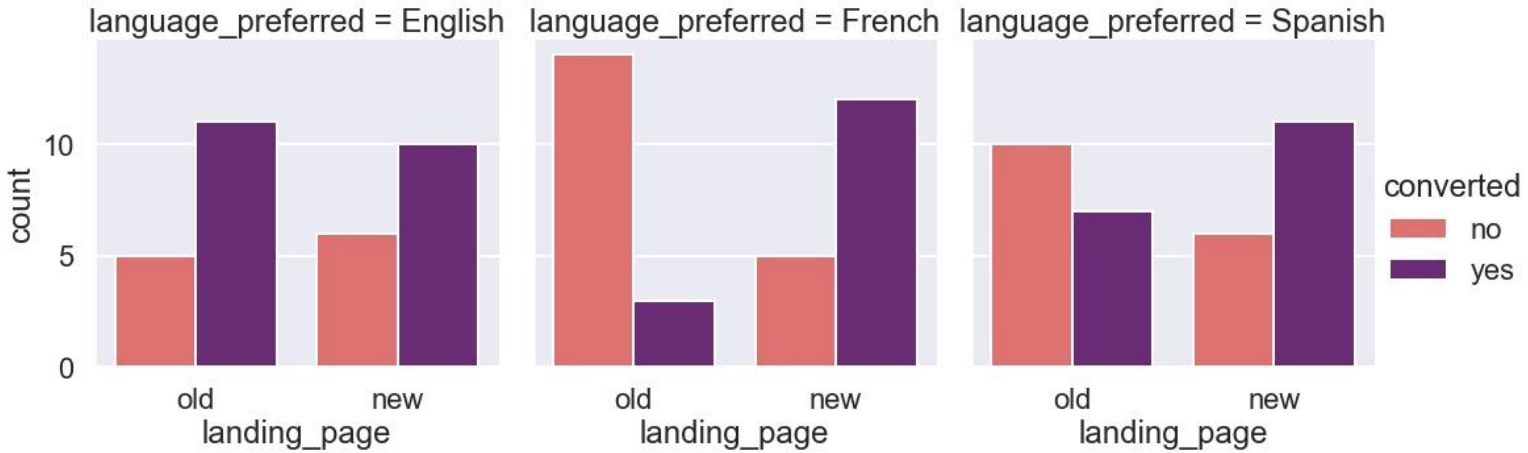
Conversion Status split by Language

Compared to the old page, the rate of conversion on the new page looks fairly even across all languages.



Conversion Rate Comparison for each Language

English readers converted at nearly the same rate on both pages.
French readers converted at a much greater rate on the new page.
Spanish readers converted at a somewhat greater rate on the new page.

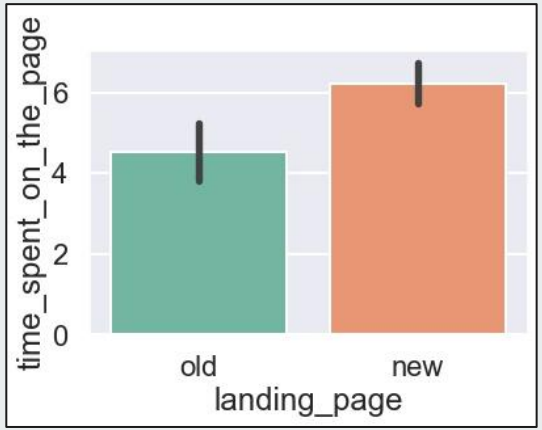


Time Differences for Conversion Status for each Language

English users had the largest difference in average times.
French users who converted were observed to have spent less time on the new page.

Insights of Inferential Statistics

Do the users spend more time on the new landing page than the old landing page?



Null Hypothesis: The average time spent on each page is equal.

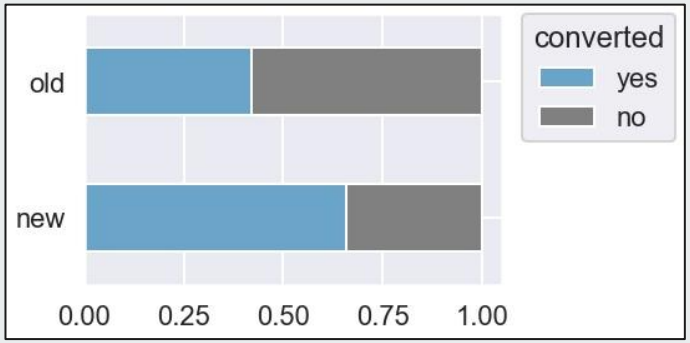
Test Performed: One-Tailed Two Independent Sample T Test

Significance Level = 0.05

Results: We have estimated that the probability of observing data this extreme, assuming the null hypothesis is still true, is `0.00013`.

YES, there is enough statistical evidence from the samples to conclude that users spend more time on the new landing page.

Is the conversion rate for the new page greater than the conversion rate for the old page?



Null Hypothesis: The rate of conversion on each page is equal.

Test Performed: One-Tailed Two Proportion Z Test

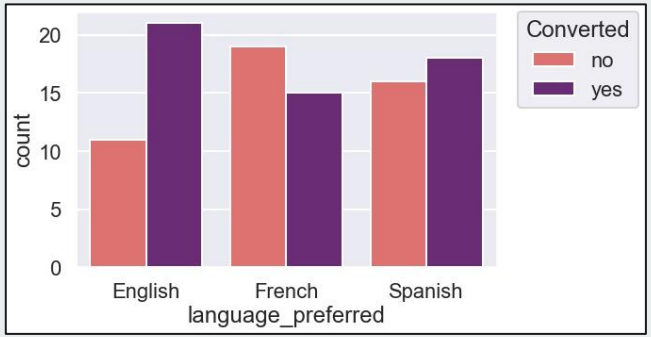
Significance Level = 0.05

Results: We have estimated that the probability of observing data this extreme, assuming the null hypothesis is still true, is `0.00802`.

YES, there is enough statistical evidence from the samples to conclude that the conversion rate is greater on the new page.

Does the converted status depend on the preferred language?

For this test, both groups were combined for a total of 100 data points.



Null Hypothesis: Conversion status is independent of preferred language

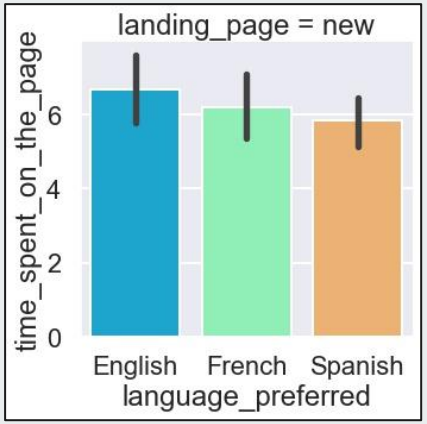
Test Performed: Chi Square Test for Independence

Significance Level = 0.05

Results: We have estimated that the probability of observing data this extreme, assuming the null hypothesis is still true, is `0.21299`.

NO, there is NOT enough statistical evidence from the samples to conclude that converted status depends on preferred language.

Is the mean time spent on the new page same for the different language users?



Null Hypothesis: The mean time on the new page for each language is equal.

Test Performed: One Way ANOVA

Significance Level = 0.05

Results: We have estimated that the probability of observing data this extreme, assuming the null hypothesis is still true, is `0.43204`.

YES, there is enough statistical evidence from the samples to conclude that the mean time spent on the new page is the same for the different language users.

Recommendations

Based on the sample data and the statistical models used, we may accept that the new landing page is effective.

- Users spend more time on the new page.
- A higher percentage of users subscribe on the new page.
- The new page is equally read by all 3 language preferences.
- The subscription rate does not depend on language preference.

In conclusion, designing a landing page that appeals to more languages will potentially multiply E-news Express subscriptions.

Thank You