E-news Express Project

PGP - Data Science & Business Analytics

September 14, 2024

Leslieane Beltran

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Hypotheses Tested and Results
- Appendix

Executive Summary

- E-news Express is evaluating a new landing page design to boost subscriber acquisition by comparing it with the old page through A/B testing.
- Users spent significantly more time on the new landing page, with a p-value of 0.0003, and had a higher conversion rate compared to those using the old page.
- This indicates an improvement in user engagement and subscription rates. Additionally, language preference does not significantly affect either the time spent on the page or conversion rates, as shown by the p-values of 0.213 and 0.432.
- Based on these results, E-news Express should go ahead with the new landing page because it keeps users engaged longer and leads to more subscriptions. They should stop using the old landing page since it's less effective. Also, they should keep the current language options because they don't affect how users interact with the page.

Business Problem Overview and Solution Approach

- E-news Express is facing a problem with declining new subscriber numbers, which is attributed to the potential ineffectiveness of its current landing page design.
- E-news Express used an A/B testing approach, dividing users into two groups, a control and treatment group.
- The control group includes users interacting with the old landing page and the treatment group includes users interacting with the new landing page.
- The data collected on the users included; time spent on the page, conversion rates, and language preferences, user group, and unique ID.
- Statistical analyses were conducted on the collected data, including a paired t-test, z-test, chi-square test, and ANOVA test.
- The analysis concluded that the new landing page increased user engagement and conversion rates. Based on these results, it is recommended that E-news Express fully adopt the new design, discontinue the old page, and keep current language options as they do not affect user behavior.
-

Data Overview

The dataset contains the the following variables:

• user_id: This represents the user ID of the person visiting the website.

• group: This represents whether the user belongs to the first group (control) or the second group (treatment).

• landing_page: This represents whether the landing page is new or old.

• time_spent_on_the_page: This represents the time (in minutes) spent by the user on the landing page.

• converted :This represents whether the user gets converted to a subscriber of the news portal or not.

• language_preferred: This represents the language chosen by the user to view the landing page

	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
95	546446	treatment	new	5.15	no	Spanish
96	546544	control	old	6.52	yes	English
97	546472	treatment	new	7.07	yes	Spanish
98	546481	treatment	new	6.20	yes	Spanish

Last 5 rows of the dataset

	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
0	546592	control	old	3.48	no	Spanish
1	546468	treatment	new	7.13	yes	English
2	546462	treatment	new	4.40	no	Spanish
3	546567	control	old	3.02	no	French
4	546459	treatment	new	4.75	yes	Spanish

First 5 rows of the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 6 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   user_id                 100 non-null     int64
 1   group                   100 non-null     category
 2   landing_page            100 non-null     category
 3   time_spent_on_the_page  100 non-null     float64
 4   converted               100 non-null     category
 5   language_preferred      100 non-null     category
dtypes: category(4), float64(1), int64(1)
memory usage: 2.6 KB
```

There are 100 records in the dataframe. There are 6 attributes. There are 4 category types, 1 float type and 1 integer type.

	user_id	time_spent_on_the_page
count	100.000000	100.000000
mean	546517.000000	5.377800
std	52.295779	2.378166
min	546443.000000	0.190000
25%	546467.750000	3.880000
50%	546492.500000	5.415000
75%	546567.250000	7.022500
max	546592.000000	10.710000

- A total of 100 entries are included in the dataset
- The average "time spent on the page" is about 5.38 minutes
- The standard deviation is about 2.38 minutes

Summary of dataset

```
Summary for column: group
group
control      50
treatment    50
Name: count, dtype: int64


Summary for column: landing_page
landing_page
old     50
new     50
Name: count, dtype: int64


Summary for column: converted
converted
yes     54
no      46
Name: count, dtype: int64


Summary for column: language_preferred
language_preferred
Spanish     34
French      34
English     32
Name: count, dtype: int64
```

Summary of Categorical Variables

group

count	100
unique	2
top	control
freq	50

dtype: object

converted

count	100
unique	2
top	yes
freq	54

dtype: object

language_preferred

count	100
unique	3
top	Spanish
freq	34

dtype: object

user_id

count	100
unique	100
top	546592
freq	1

landing_page

count	100
unique	2
top	old
freq	50

dtype: object

Summary of Numerical Value

	time_spent_on_the_page
count	100.000000
mean	5.377800
std	2.378166
min	0.190000
25%	3.880000
50%	5.415000
75%	7.022500
max	10.710000

Sample size is 100, where the mean time spent on a page by the users is 5.3 minutes, with a minimum of 0.19, and a maximum of 10.71 minutes.

- There are no missing values in the dataset
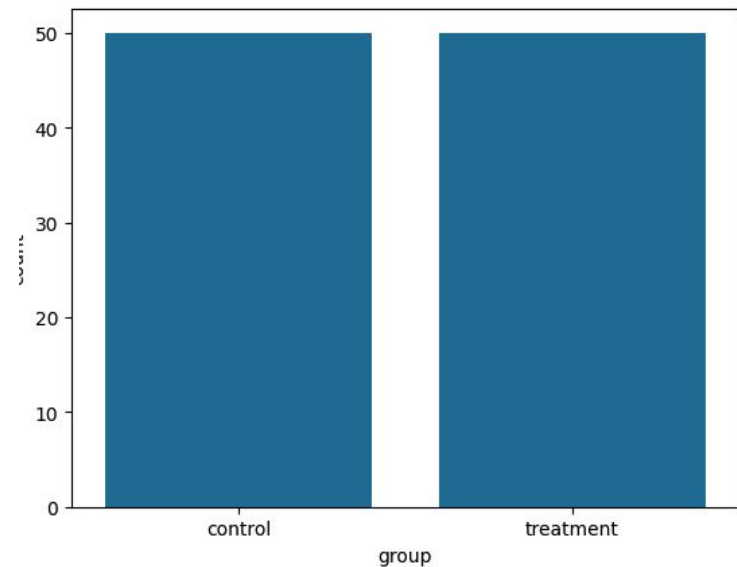
- There are no duplicates in the data set

EDA Results - Univariate Analysis

TIME SPENT ON PAGE
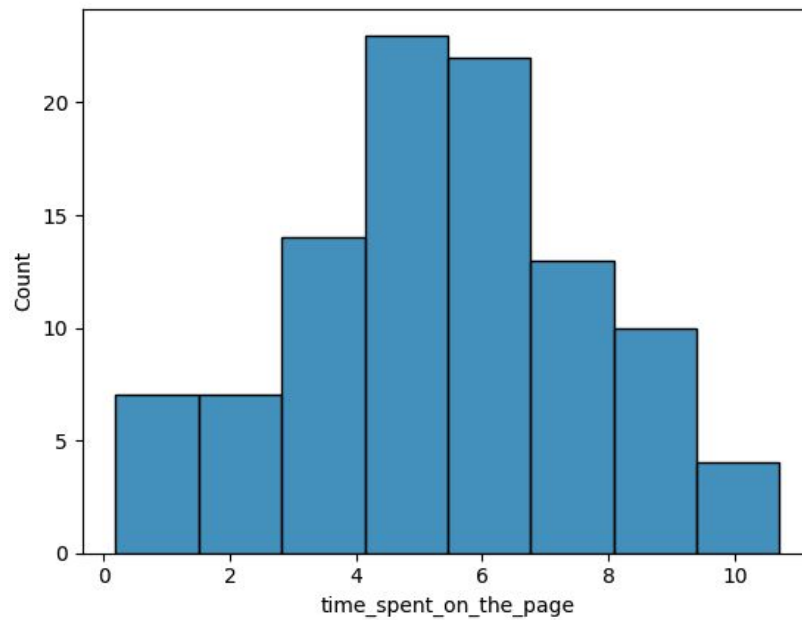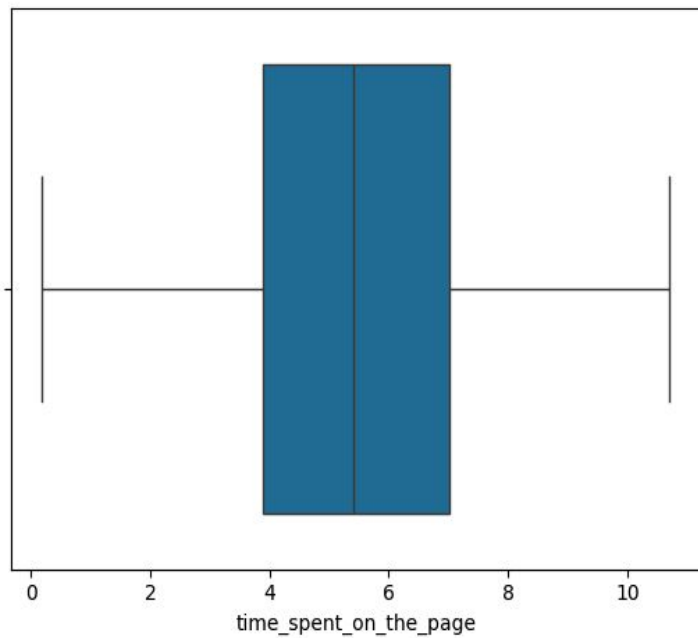
Observation

- Normal distribution
- The average "time spent on the page" is about 5.38 minutes
- The standard deviation is about 2.38 minutes
- Interquartile range includes: • Min = 0.19 • 25% quartile = 3.88 • 50% quartile = 5.42 • 75% quartile = 7.02 • Max = 10.71

GROUP


```
                        count
            group
            control      50
            treatment    50

dtype: int64
```

Half of the users belongs to the first group (control) and half of users belong the second group (treatment)
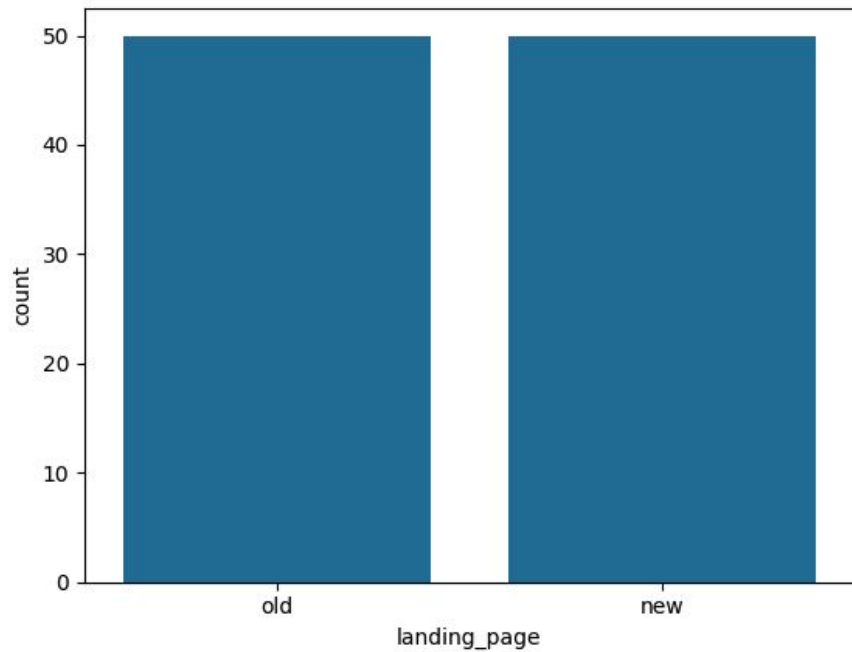
LANDING PAGE

	count
landing_page	
old	50
new	50

dtype: int64

The users are split evenly between the new and old landing page.

CONVERTED

	count
converted	
yes	54
no	46

dtype: int64
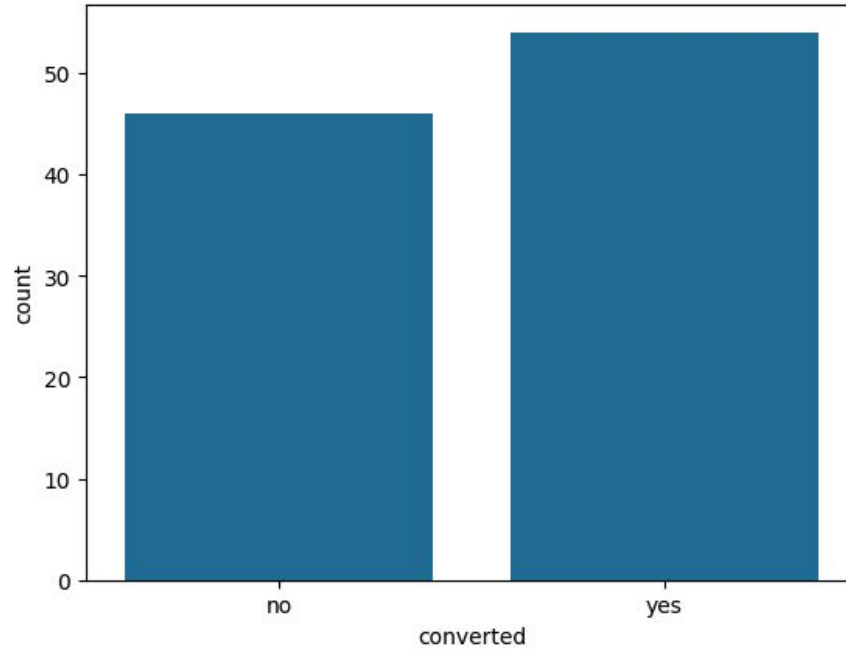
The Converted variable has two unique values.

LANGUAGE PREFERRED
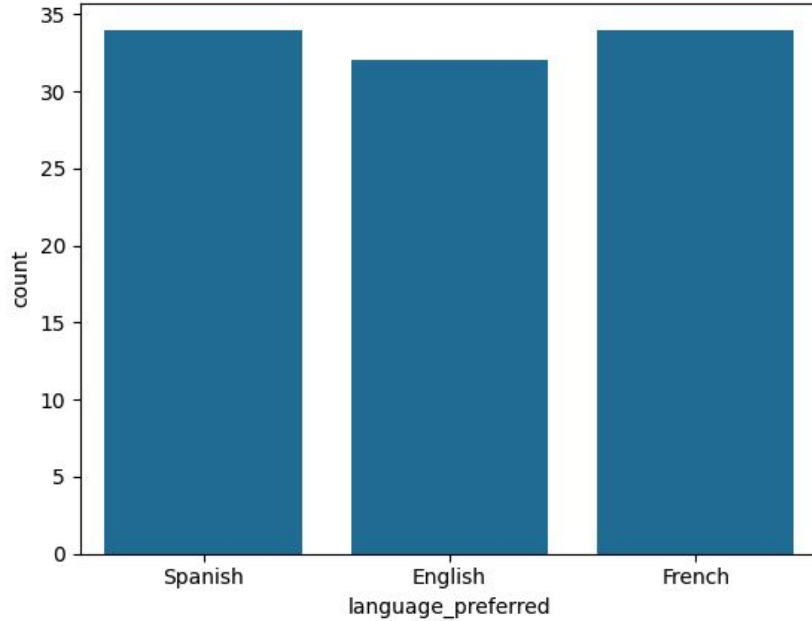
	count
language_preferred	
Spanish	34
French	34
English	32

dtype: int64

- Three unique values including English, French, & Spanish.
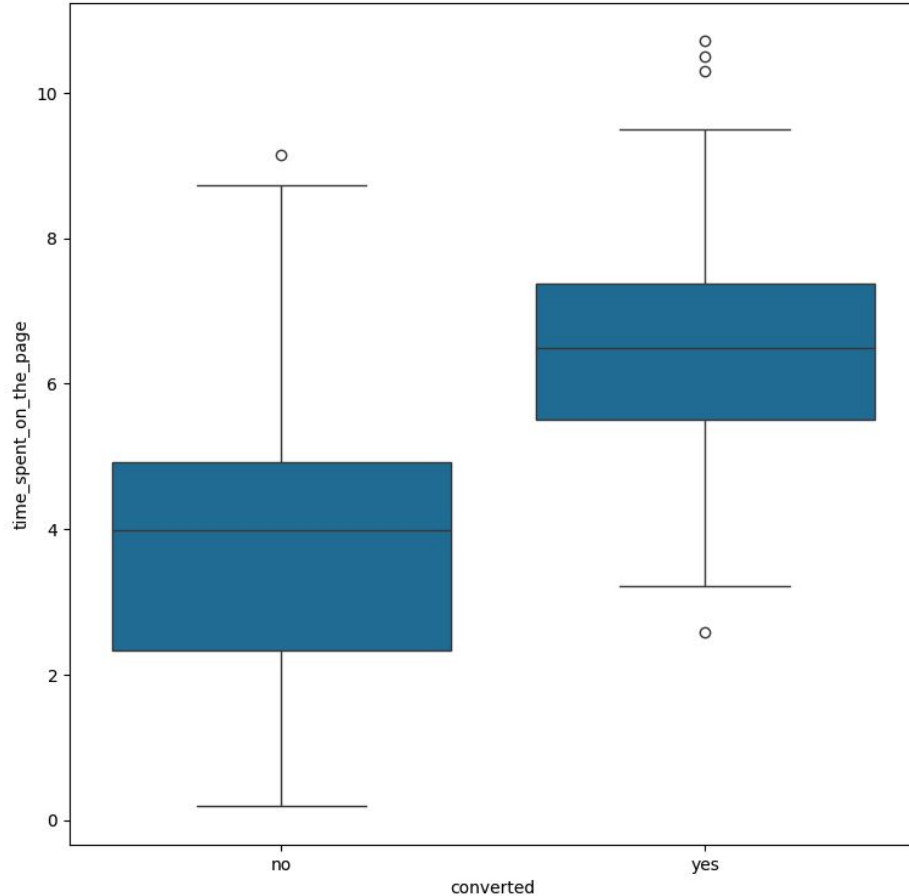- French and Spanish have the highest frequencies

EDA Results - Bivariate Analysis
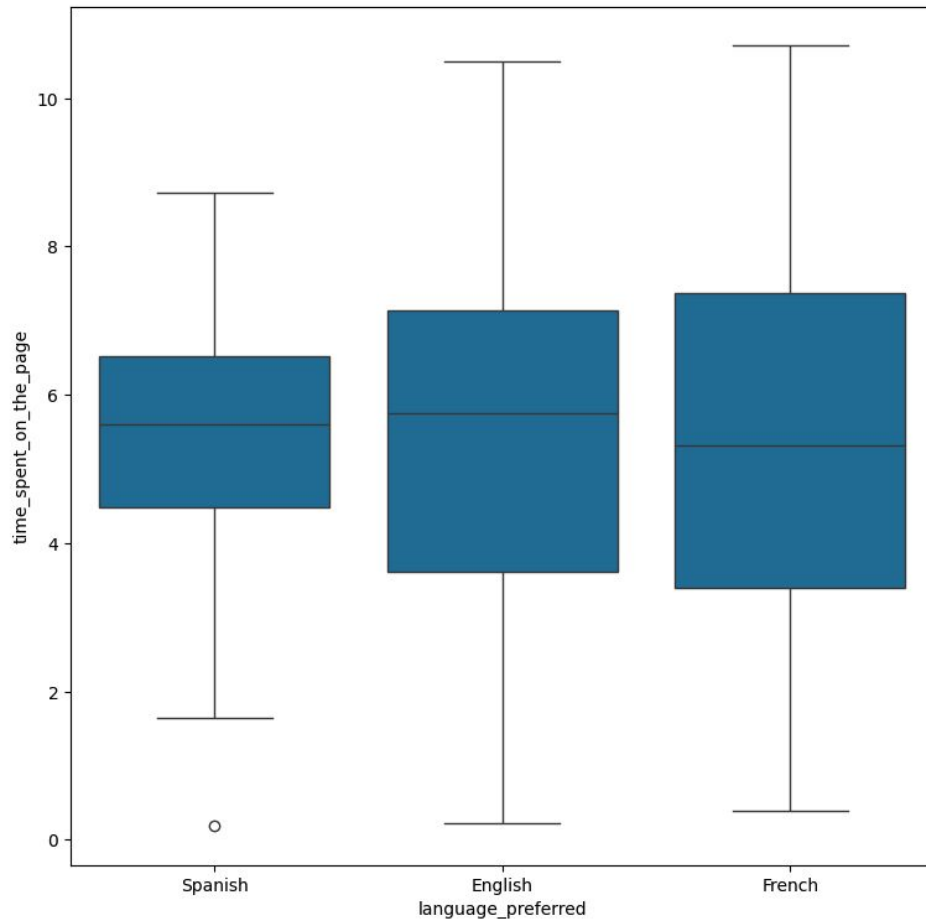
Landing page vs Time spent on the page
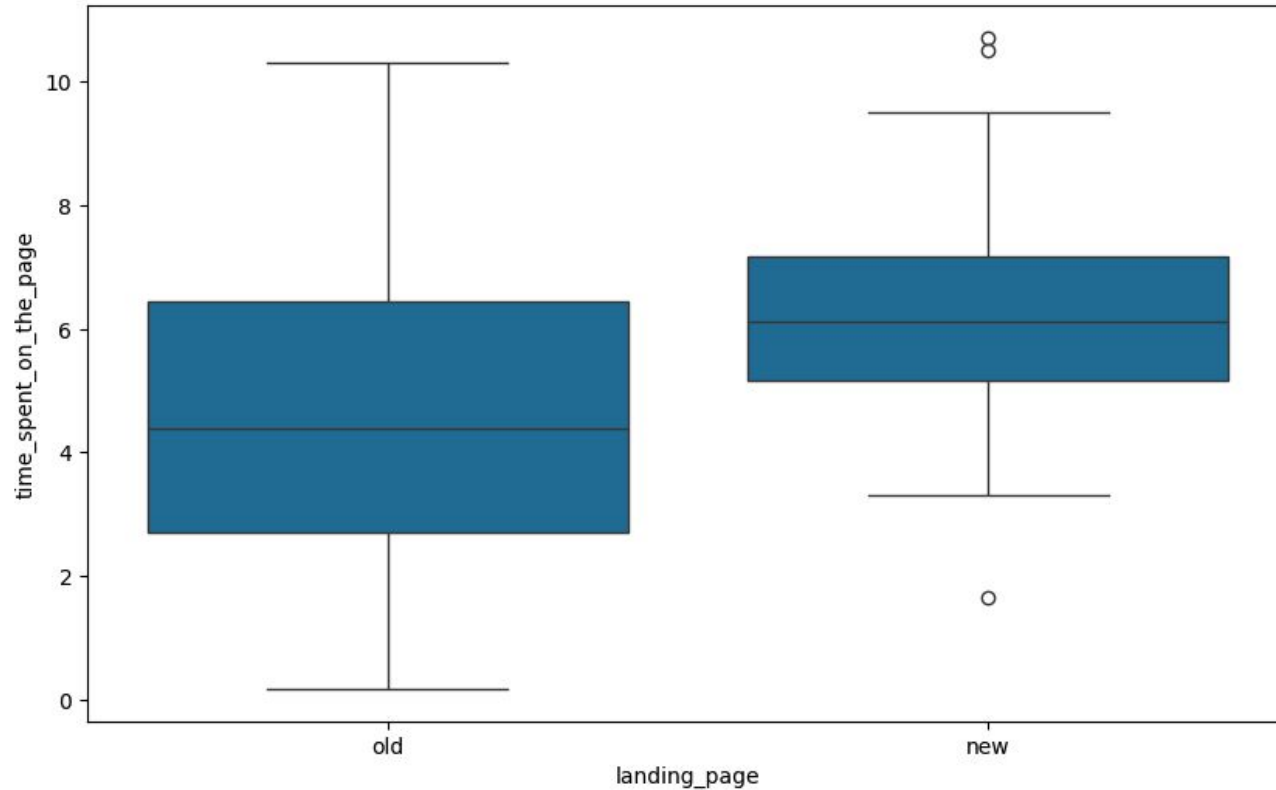
Users exposed to the new landing page spend more time on average than those on the old landing page.

Conversion status vs Time spent on the page

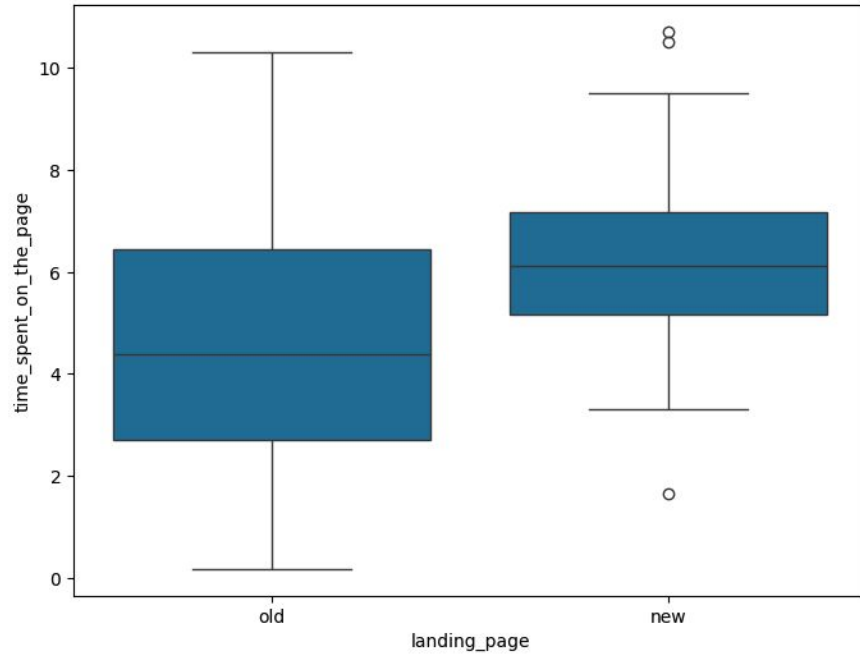Users who converted demonstrate on average a higher time spent on the page.

Language preferred vs Time spent on the page

There is not much difference on how much time is spent on page based on language preference.

Hypotheses Tested and Results

Q1. Do the users spend more time on the new landing page than the existing landing page?

Users seem to
spend more
time on the new
landing page in
comparison to
the old landing
page

Q1 - Hypothesis Testing Details

H_0: The mean time spent on the old landing page.

H_a: The mean time spent on the new landing page

Type of test: paired t-test.

P value is 0.0003.
0.0003 less than 0.05.
Hence, we reject the null hypothesis. This means that the **new landing page** encourages users to spend **more time** significantly.

2. Is the conversion rate (the proportion of users who visit the landing page and get converted) for the new page greater than the conversion rate for the old page?

Comparison of Conversion Rates by Landing Page
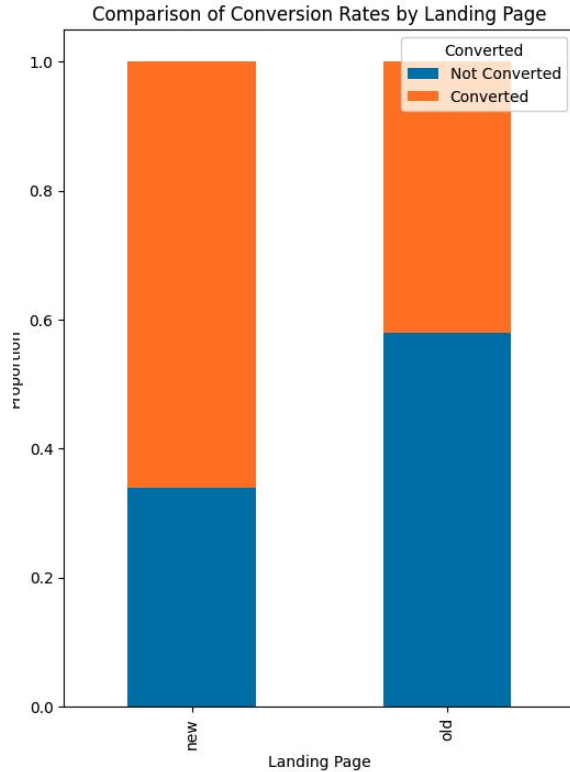
Users who visit the new landing page are more likely to convert than those using the old landing page.

Q2 - Hypothesis Testing Details

$H0$: Landing page has no effect on conversion
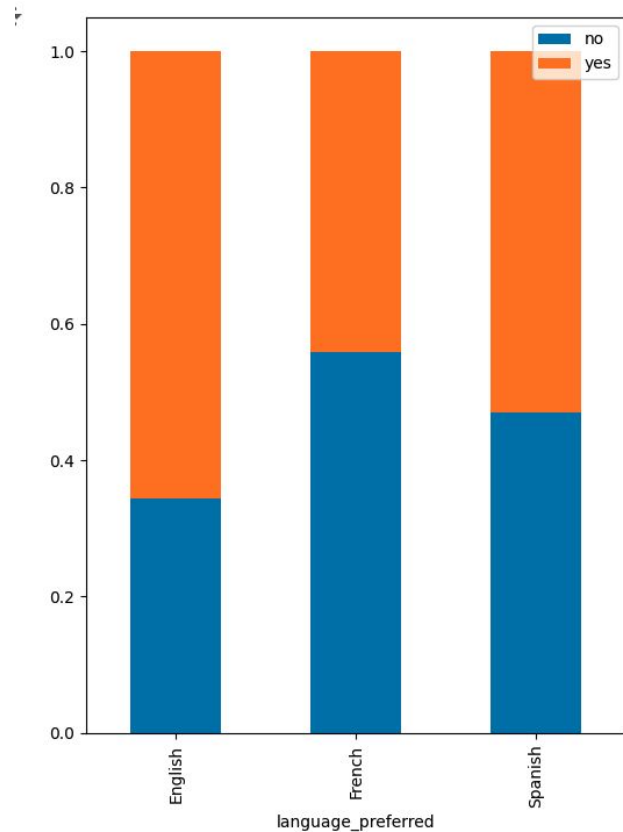
Ha: Landing page has an effect on conversion

Type of test: Z-test

p value is 0.016

 0.016 is less than the 0.05. Hence, we reject the null hypothesis.

Meaning, landing page **has an effect** on conversion

3. Does the converted status depend on the preferred language?

The converted status seems to have more of an effect on English speaking users vs Spanish & French speakers.

Q3 - Hypothesis Testing Details

$H0$: Language has no effect on conversion

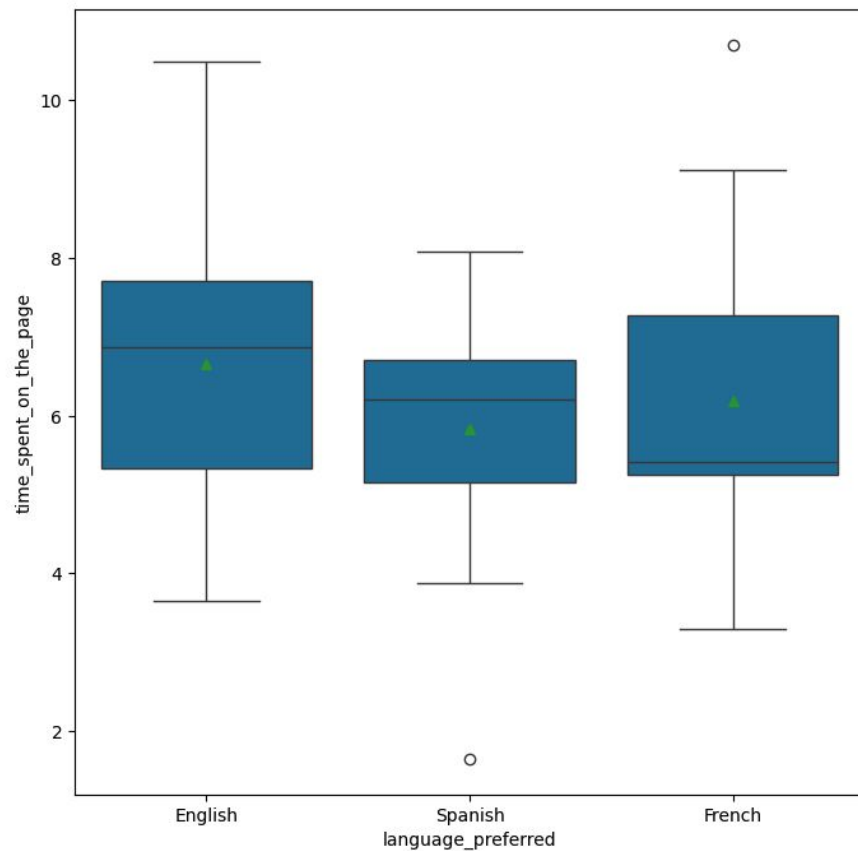Ha: Language has an effect on conversion

Type of test: Chi Square test

P value is 0.213

0.213 is greater than the 0.05. Hence, we fail to reject the null hypothesis.

Language has **no effect** on the conversion.

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

The average time spent on the pages seems relatively the same among the different language preferences.

Q4 - Hypothesis Testing Details

$H0$:Preferred language has no effect on time spent

Ha:Preferred language has an effect on time spent

Type of test: ANOVA

P value is 0.432

0.432 is greater than the 0.05. Hence, we fail to reject the null hypothesis.

Preferred language has **no effect** on time spent on the page

Conclusion & Business Recommendations

Conclusion

- The analysis revealed that the new landing page significantly increases user engagement and conversion rates compared to the old page.
- Users spent more time on the new page and were more likely to subscribe.
- Language preference did not significantly affect time spent or conversion rates.

Business Recommendations

- E-news Express should fully transition to the new landing page to capitalize on its improved user engagement and higher conversion rates.
- The old landing page should be discontinued to avoid shortcomings.
- The current language options should be kept the same as they do not seem to impact user behavior.