E-NEWS EXPRESS (Case Study)

Project 2: Business Statistics

Date: October 8th, 2022.

By: Ijeoma Ejem

Contents / Agenda

- Executive Summary pg. 3 - 4
- Business Problem Overview pg. 5
- Solution Overview pg. 6
- Data Overview pg. 7
- EDA Results pg. 8 - 13
- Hypotheses Tested and Results pg. 14 - 17
- Appendix pg. 18 - 20

Executive Summary

	Conclusion	Recommendation
1	Users who spend more time on the platform convert into subscribers more frequently than those who spend less time.	Create more interactive landing page designs with current news and captivating headlines that will engage the user.
2	More users converted into subscribers after viewing the new landing page than the older landing page.	Test more dynamic landing page designs to identify specific patterns in the user's preferences and engagement rate.
3	Tests confirm that users spent more time on the new page than on the old landing page.	Identify strengths and weaknesses between the old and new landing page designs for clearer design do's and don't for the design team guidelines.

	Conclusion	Recommendation
4	The preferred language of a user is independent of the user's conversion into a subscription.	Focus on collecting data around other factors that are could be more dependent on conversions like user's psychographic data.
5	There is no difference between time spent on the new landing when analyzed by preferred language.	Same as above.
6		To increase conversions and engagement, consider a deeper study of the company's target population through a wider array of data.

Business Problem Overview

- E-news express offered us a great opportunity to quickly get updates on the day-to-day events occurring globally through online portals.
- Traditional methods provide slower access to content and little to no interactive elements that make news more engaging for the reader.
- E-news express wants to increase its number of subscribers and conversation rate.
- The company wants to understand its audience and their interests better based on the action they take. Once this data is retrieved, to analyze the data, draw conclusions and provide business solutions to drive up user engagement.
- Company executives want to confirm via data testing that new monthly subscribers have declined compared to the previous year due to the current poorly designed webpage.
- The outline & recommended content on the current webpage do not hold the attention of the reader long enough to convince them to subscribe.

Solution Approach

- The design team has created a new landing page with more engaging content and improved outlines.
- The data science team has performed A/B testing to analyze user data from the landing pages and determine efficacy of the pages towards the desired outcome. (increased conversation rates and subscribers)
- An experiment is conducted and data is collected on a randomly selected group of 100 users divided equally into two groups, one group is tested on the current landing page and the other is tested on the new landing page.
- Based on the data collected, our tests will be able to answer the following questions:
  - Does the new landing page increase subscriptions and conversions?
  - Will the users spend more time on the new page?
  - Are conversions dependent on any other factor beyond time spent?
  - Will users with different preferred languages spend the same amount of time on the new page?

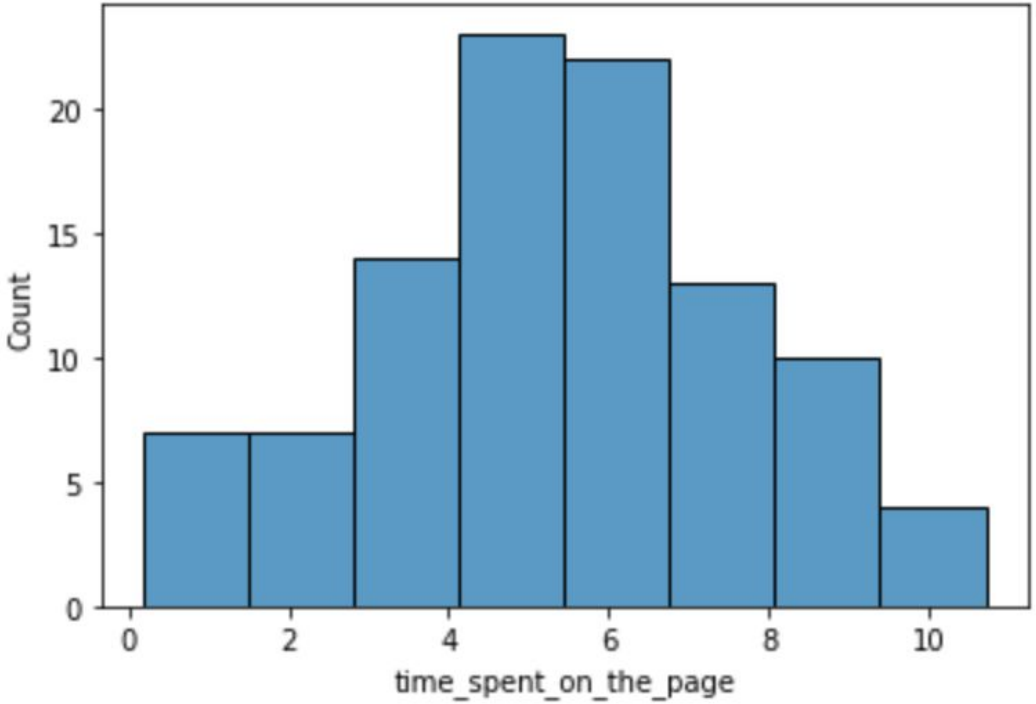
This is our sphere of interest and by testing these assumptions, we were able to deduce to what degree the new landing page created by the design team will help the company meet its objective of higher subscriber engagement and conversion rates.

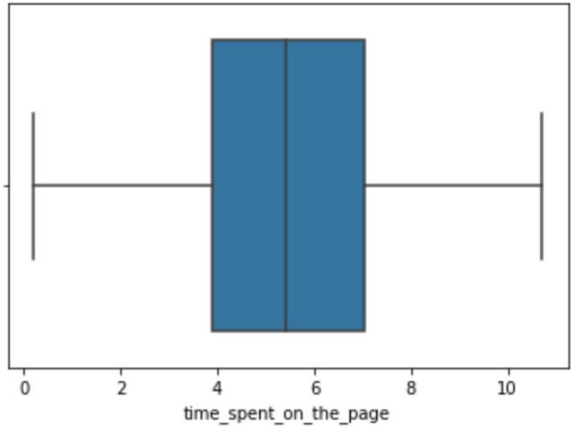
Data Overview

- Users have been divided into two equal groups of 50 users each.
- There are 100 rows (observations) and 6 columns (attributes) in total.
- The attributes we are analyzing and testing are group (control or treatment), landing page (old or new), time spent on the page, converted (yes or no), and language preferred (English, French or Spanish).
- From the statistical summary of the combined numerical and categorical data, the following is deduced;
  - Time spent on landing page has a mean of 5.3778, a median of 5.415, and a standard deviation of 2.378166.
  - The frequency of users who converted is higher than that of users who did not.
  - Spanish is the top occurring language of preference for users.
- There are no missing or duplicate values in the data.

EDA Results

UNIVARIATE ANALYSIS



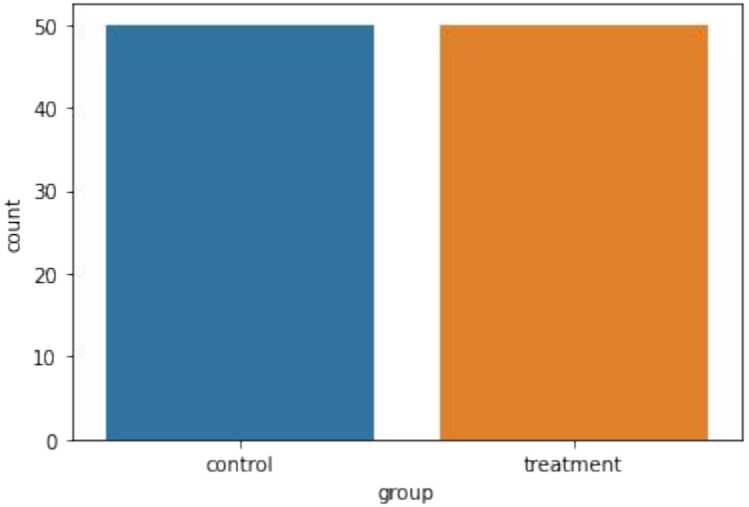


Time Spent on the Page

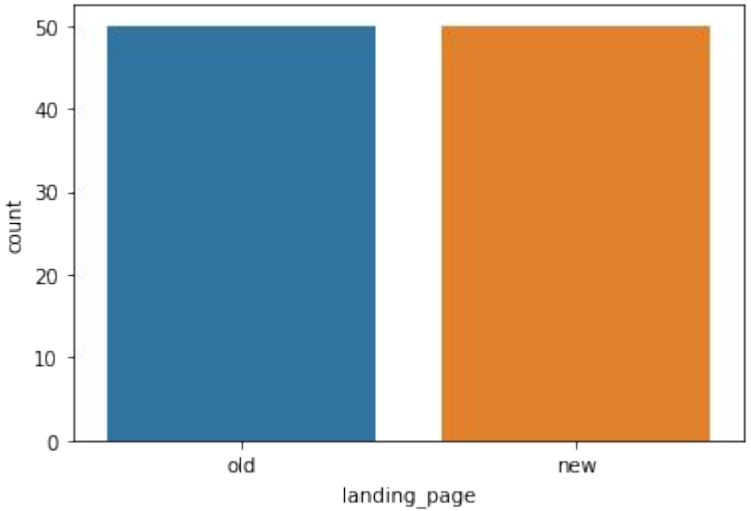
Across the sample population, the highest number of users spent between 4 and 7 minutes on both old and new pages. With the median falling at about 5.5 minutes.

EDA Results

UNIVARIATE ANALYSIS



Group

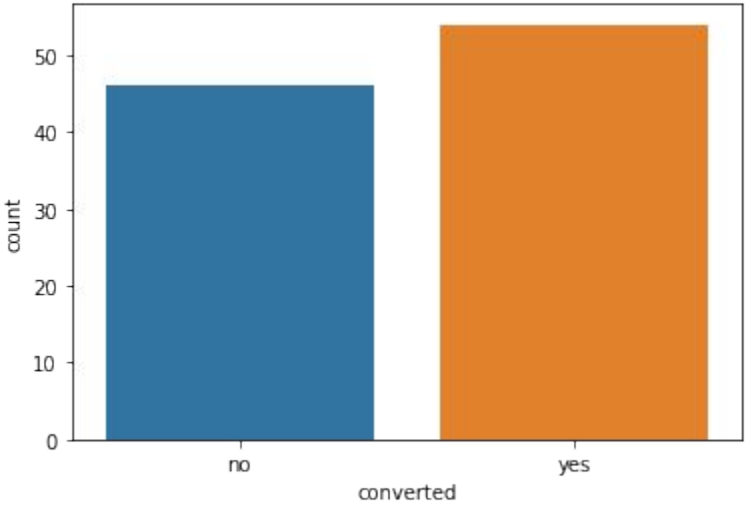


Landing Page

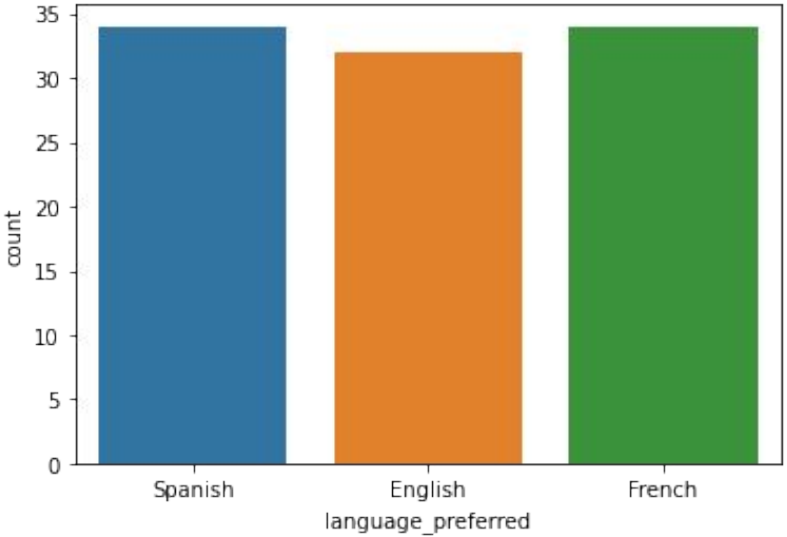
The sample population is divided equally into two groups. The histogram confirms the 50/50 split between the control group (for the old landing page) and the treatment group (for the new landing page).

EDA Results

UNIVARIATE ANALYSIS



Converted



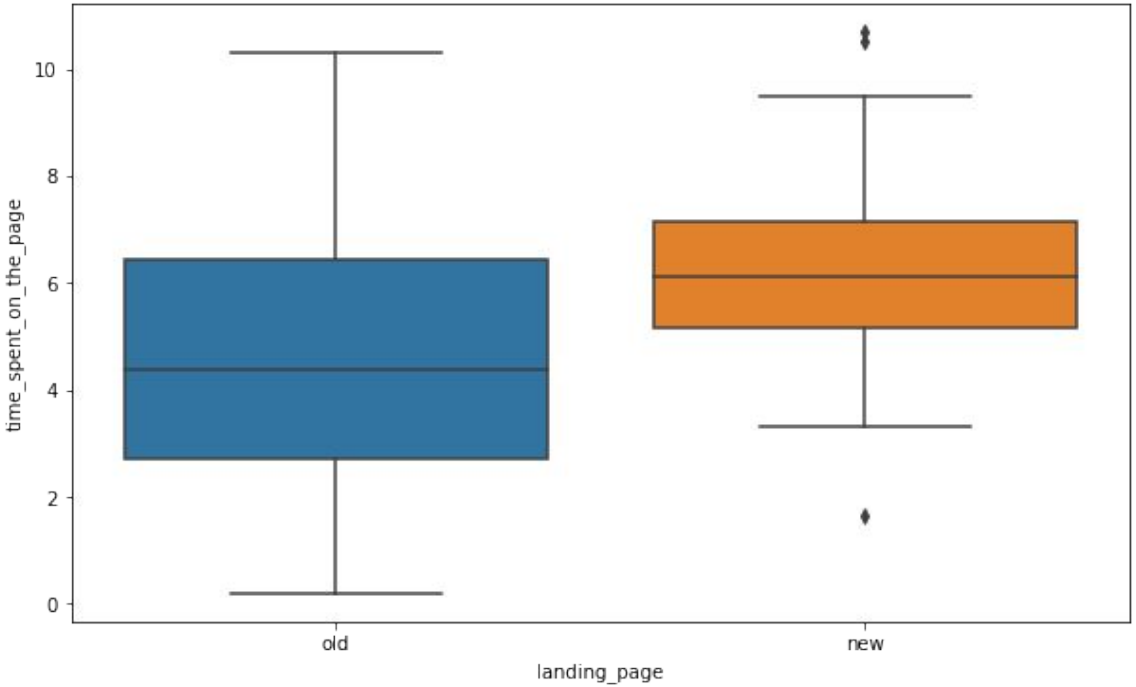
Preferred Language

As mentioned in the data overview, The total number of conversions is higher than non-conversions.

Spanish is one of the top languages and appears to have the same count as French. English is the least preferred language across this sample population.

EDA Results

BIVARIATE ANALYSIS



Landing page vs Time spent on the page

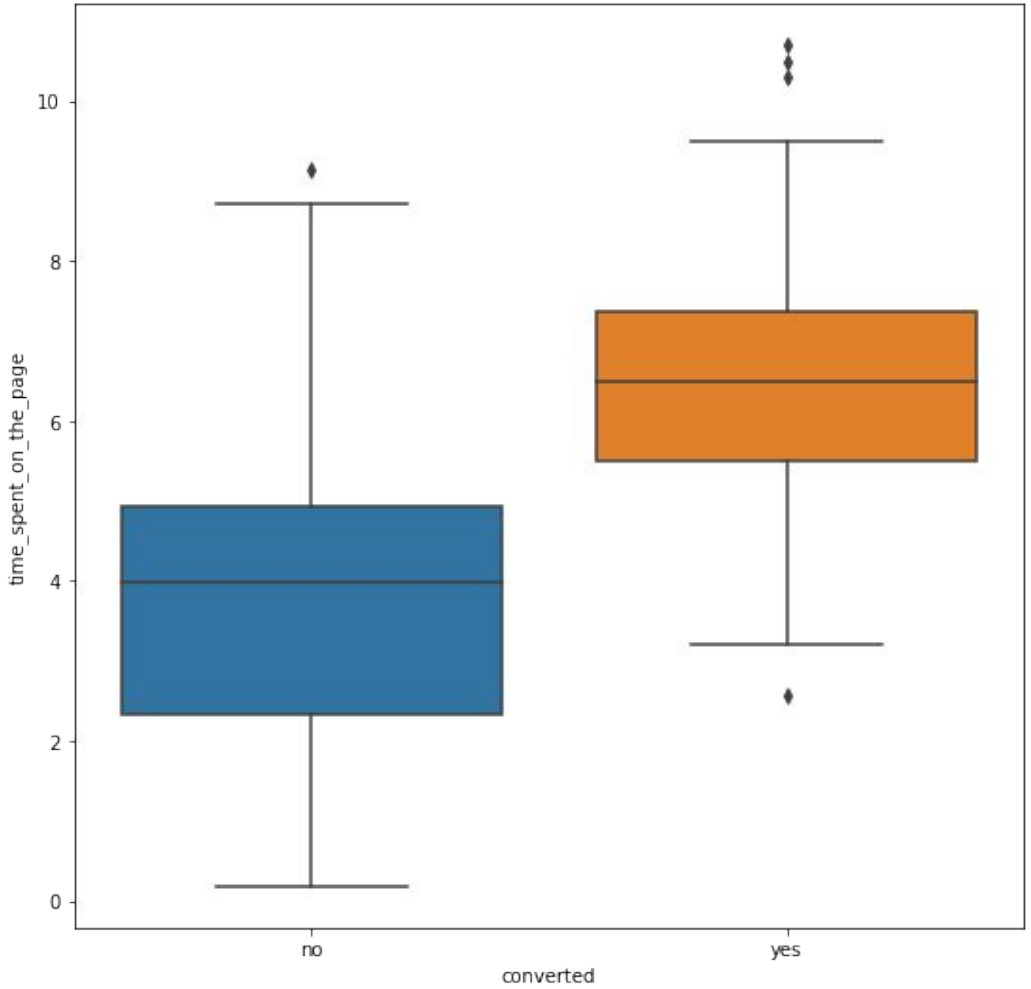
The old page has to lower minimum count of time spent and a lower median compared to the new page. However, the new page has a lower maximum count of time spent when compared to the old page. The new page has outliers on both ends of the spectrum.

EDA Results

BIVARIATE ANALYSIS

Conversion status vs Time spent on the page

There is a higher overall positive conversion rate than a negative conversion rate when compared to time spent on both pages.
This means that those who subscribed spent more time on the page than those who spent less time.



EDA Results

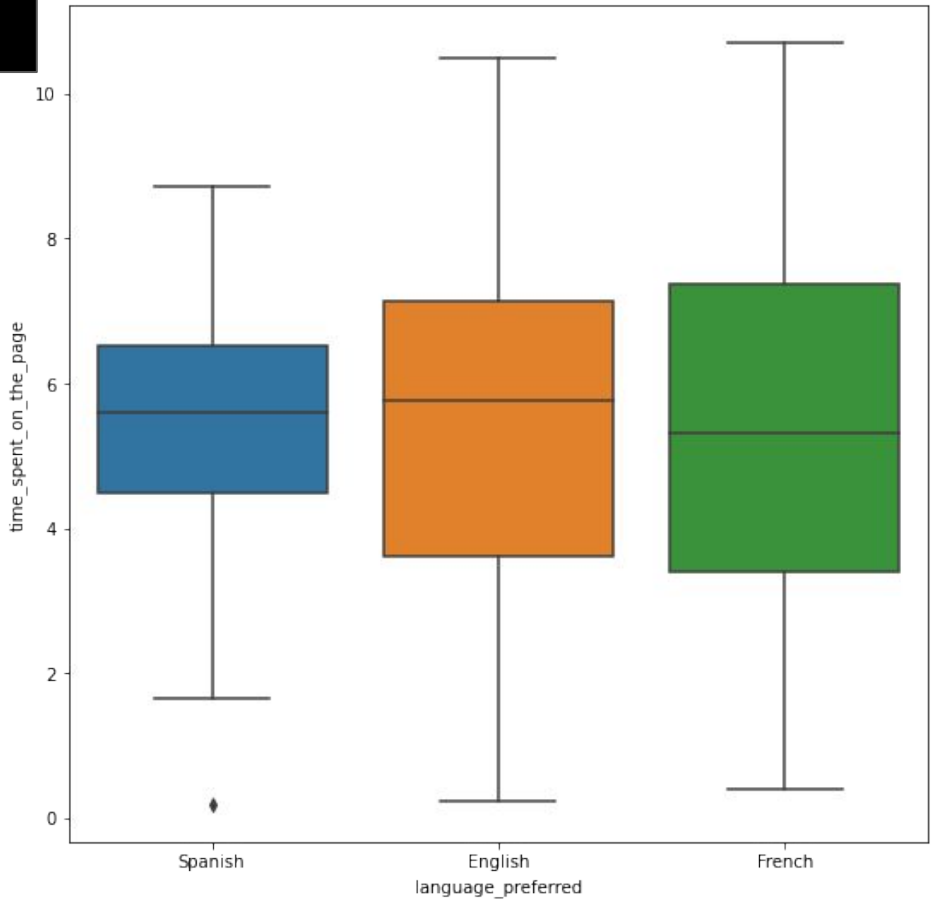
BIVARIATE ANALYSIS

Language preferred vs Time spent on the page

English and French language users spent more time on their page than Spanish language users.
However, the Median time spent for each preferred language in descending order is as follows:

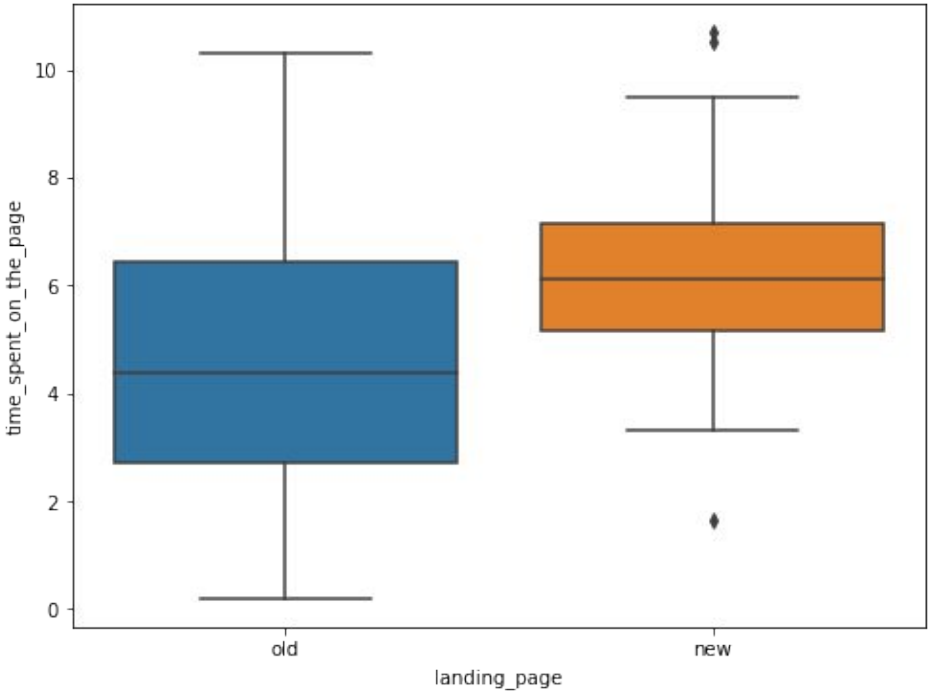
1.) English
2.) Spanish
3.) French

There doesn't appear to be a huge difference in time spent on the page across the sample population. The most substantial difference in minimum and maximum views can be observed from the Spanish language user.



Hypotheses Tested and Results

- Visual analysis performed:



- Question 1:

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

- Test Used:

Two Independent Sample T-test for Equality of Means - Unequal Std Dev

- Hypothesis tested:

$\mu 1$ and $\mu 2$ is the average time users spent on new page 1 and old page 2, respectively.

$H0$: Time spent on new landing page <= than old landing page

H_a: Time spent on new landing page > than old landing page

$$H0: \mu 1 <= \mu 2$$

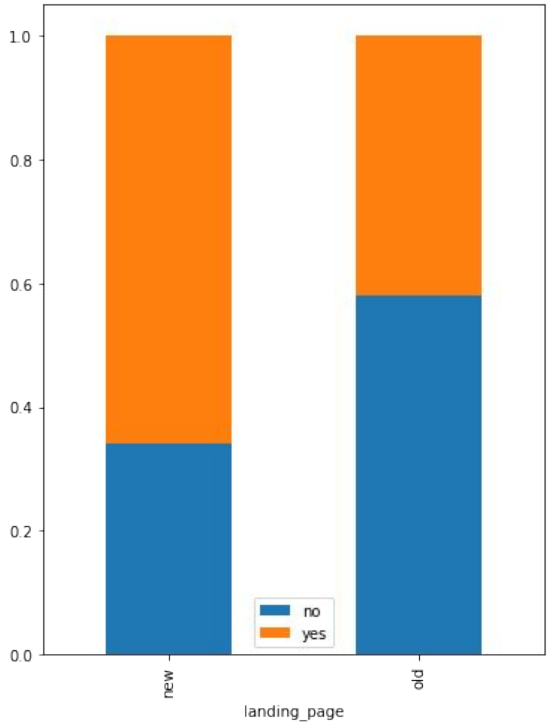
$$H_a: \mu 1 > \mu 2$$

- Test result and inference:

The p-value (0.00013) is less than the level of significance (0.05), we can reject the null hypothesis. There is enough evidence to conclude that users spent more time on average on the new landing page.

Hypotheses Tested and Results

- Visual analysis performed:



- Question 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?

- Test Used:

Two Proportion Z-test

- Hypothesis tested:

$p1$ and $p2$ are the proportions of users who converted on new page 1 and old page 2, respectively.

$H0$: Conversion rate on new landing page <= than old landing page

Ha: Conversion rate on new landing page > than old landing page

$H0$: $p1 <= p2$

Ha: $p1 > p2$

- Test result and inference:

As the p-value (0.008) is less than the level of significance (0.05), we reject the null hypothesis. There is enough evidence to prove that the conversion rate on the new landing page is higher than the conversion rate of the old page.

Hypotheses Tested and Results

- Question 3:

Does the converted status depend on the preferred language?

- Test Used:

Chi-Square Test for Independence

- Hypothesis tested:

We tested the null hypothesis,

H_0: Language preference is independent of conversions

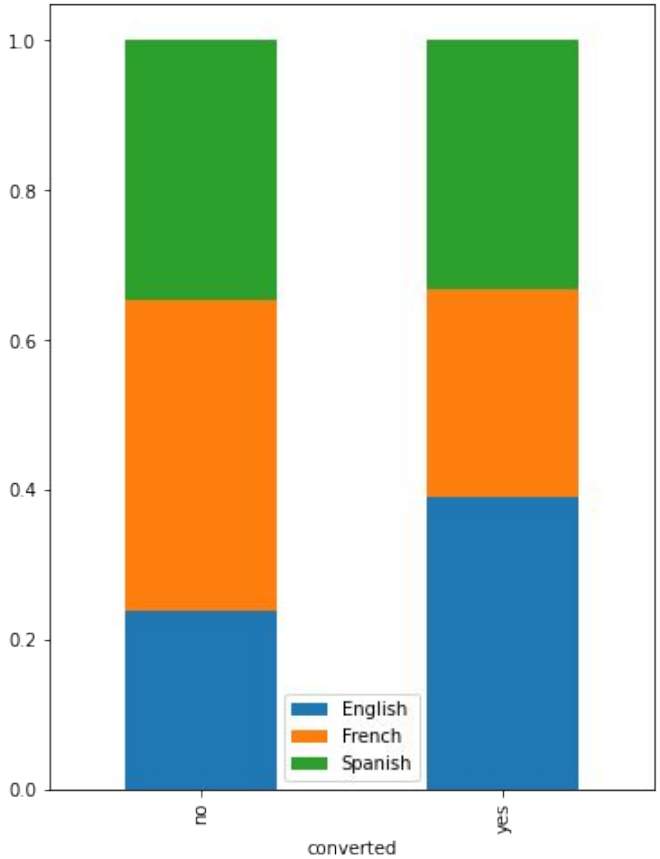
against the alternative hypothesis,

H_a: Language preference depends on conversions.

- Test result and inference:

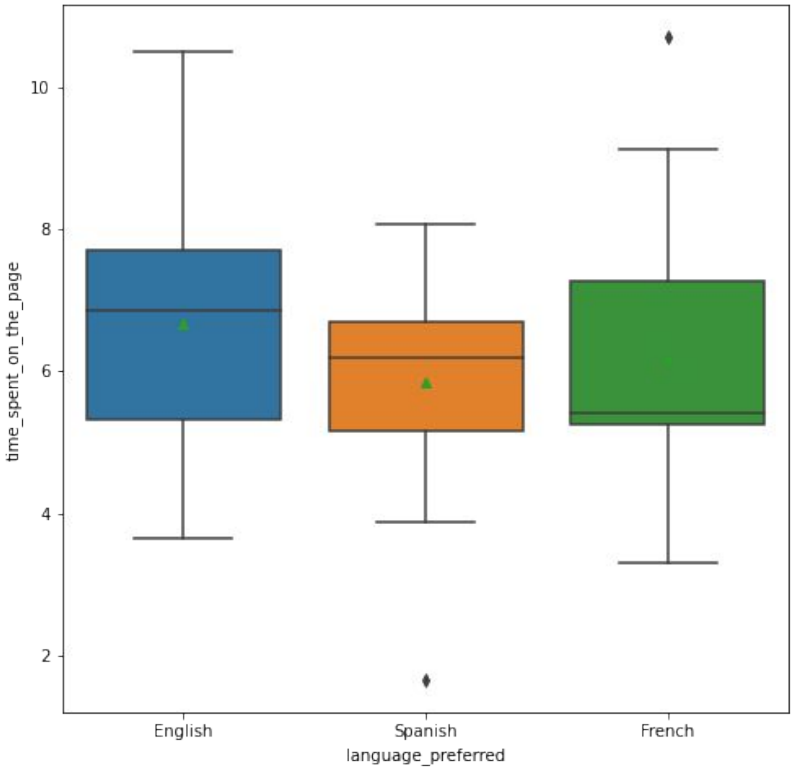
As the p-value (0.213) is greater than the level of significance (0.05), we fail to reject the null hypothesis. The test result shows lack of supporting data to prove that a user's language preference is dependent on their conversion rate.

- Visual analysis performed:



Hypotheses Tested and Results

- Visual analysis performed:



- Question 4:

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

- Test Used:

One-way ANOVA Test

- Hypothesis tested:

$\mu 1$ and $\mu 2$ and $\mu 3$ are the averages of time spent on the new page for different language users; English 1, French 2 and Spanish 3, respectively.

$H0$: $\mu 1 = \mu 2 = \mu 3$

Ha: $\mu 1 \neq \mu 2 \neq \mu 3$

$H0$: All languages are equal on average.

Ha: At least one preferred language is different from the rest.

- Test result and inference:

As the p-value (0.432) is greater than the level of significance (0.05), we fail to reject the null hypothesis. Our test denies sufficient evidence to prove that there is a difference between time spent on the new page based on the preferred languages chosen by users.

APPENDIX

P-Value Calculation

Q1:

```
from scipy.stats import ttest_ind

test_stat, p_value = ttest_ind(time_spent_new, time_spent_old, equal_var = False, alternative = 'greater')

print('The p-value is', p_value)
```

Q2:

```
from statsmodels.stats.proportion import proportions_ztest

test_stat, p_value = proportions_ztest([new_converted, old_converted] , [n_treatment, n_control], alternative ='larger')

print('The p-value is', p_value)
```

Q3:

```
from scipy.stats import chi2_contingency

chi2, p_value, dof, exp_freq = chi2_contingency(contingency_table)

print('The p-value is', p_value)
```

Q4:

```
from scipy.stats import f_oneway

test_stat, p_value = f_oneway(time_spent_English, time_spent_French, time_spent_Spanish)

print('The p-value is', p_value)
```