E-news Express

The world news in one click

Business Statistics

EXPAND BUSINESS BY ACQUIRING NEW SUBSCRIBERS

Objectives.

1. Determine whether the new landing page is more effective to gather new subscribers.
2. Perform uni-variate and multi-variate analyses;
3. Perform the statistical analysis and visual analysis;
3. Generate a set of insights and recommendations that will help the company in to expand its business by acquiring new subscribers.

Data Overview

Columns	Description	Dtype
user_id	User ID of the person visiting the website.	Int64
group	This represents whether the user belongs to the first group (control) or the second group (treatment).	category
landing_page	This represents whether the landing page is new or old.	category
time_spent_on_the_page	This represents the time (in minutes) spent by the user on the landing page.	Int64
converted	This represents whether the user gets converted or not.	category
language_preferred	This represents the language chosen by the user to view the landing page.	category

Note:

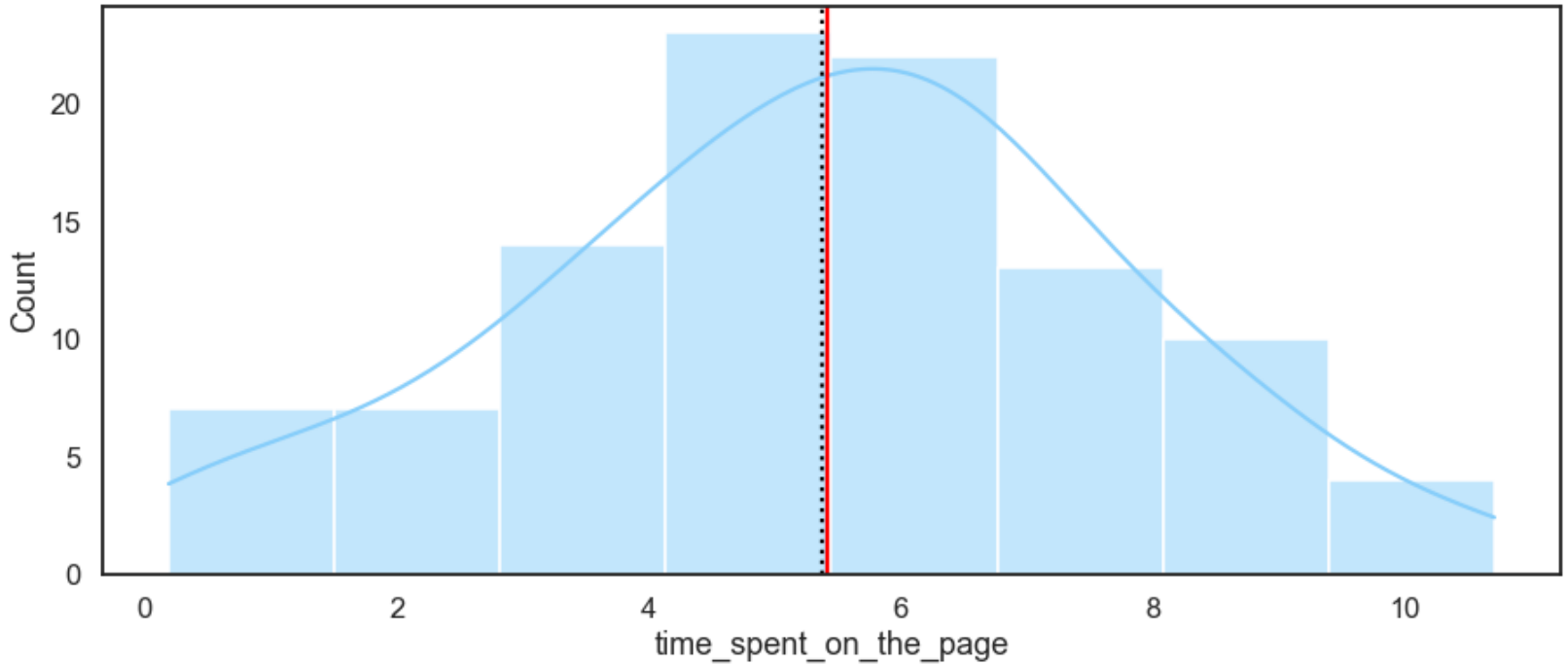
1. There are no missing values in the data;
2. group, landing_page, converted and language_preferred should be categorical variables
3. Group control is equal to landing page Old and,
   Group treatment is equal to landing page New.

3.1 In the project we are going to refer to New an Old landing page.

Observations	Variables
100	6

EDA – Uni numerical variable

Users Time Spent on the page - Distribution:

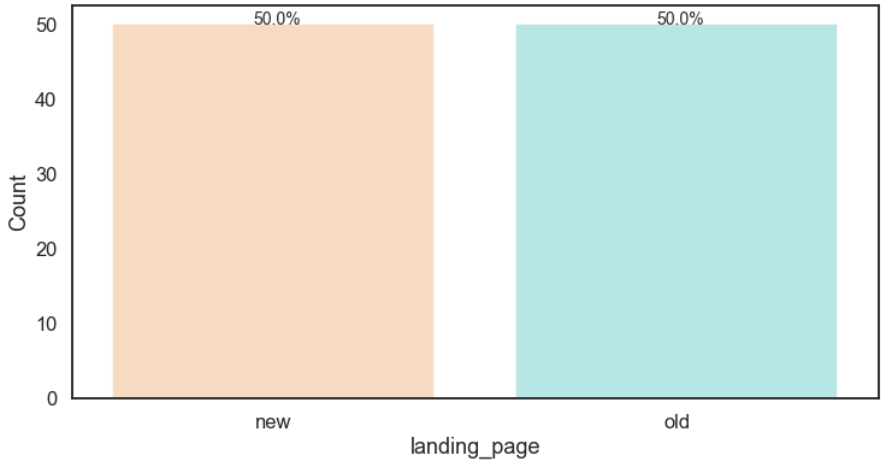


Observations

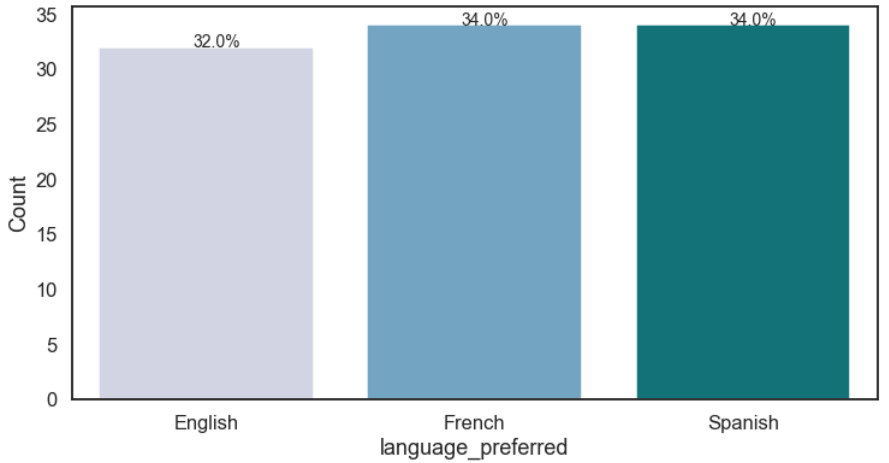
* The histogram shows that data is approximately normally distributed.
* Mean and median it is close to each other.

EDA – Uni categorical variable

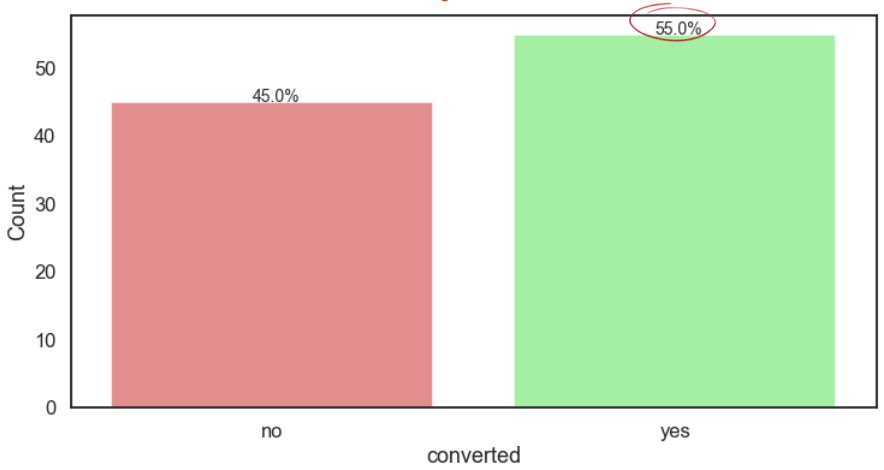
Users distributed by Landing Page:



Users distributed by Language preferred:



Users distributed by Converted status:



Observations

•There is 50 users in each group/landing page;

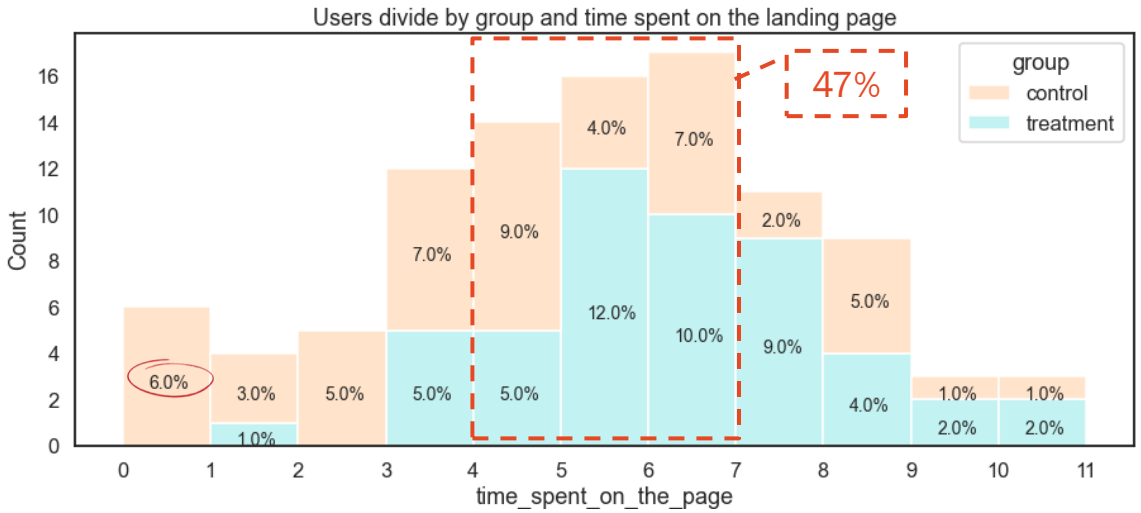
•Language preferences its almost equal distributed between the three languages.

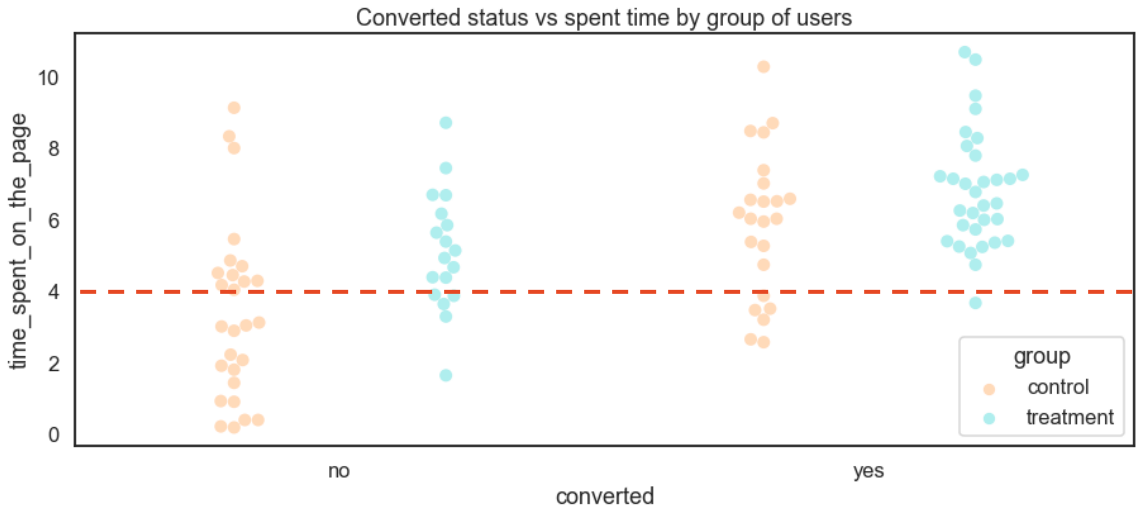
•There is 32 users that prefers English as language (32% of data),

• 34 % users prefers to visualize the landing page in French and the same is true for Spanish.

•55% of users became subscribers to the landing page.

EDA – Bi-variate Multivariate





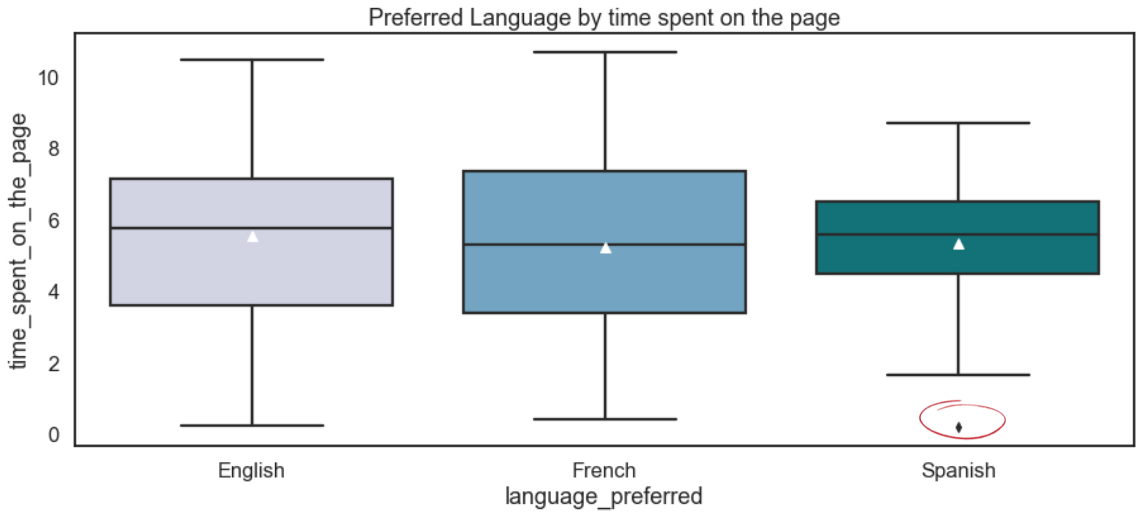
Observations

- 6% of users spent less than 1 min in the Control/Old page;
- Treatment users spent at least more than 1 min on the page;
- 8% of Control users spent 1 to 3 min on the page VS 1% of Treatment users;
- 47% of users spent between 4 to 7 min on the page (27% new page/treatment);
- Users on treatment group (New landing page) is more likelihood to spend more time on the page.

Observations

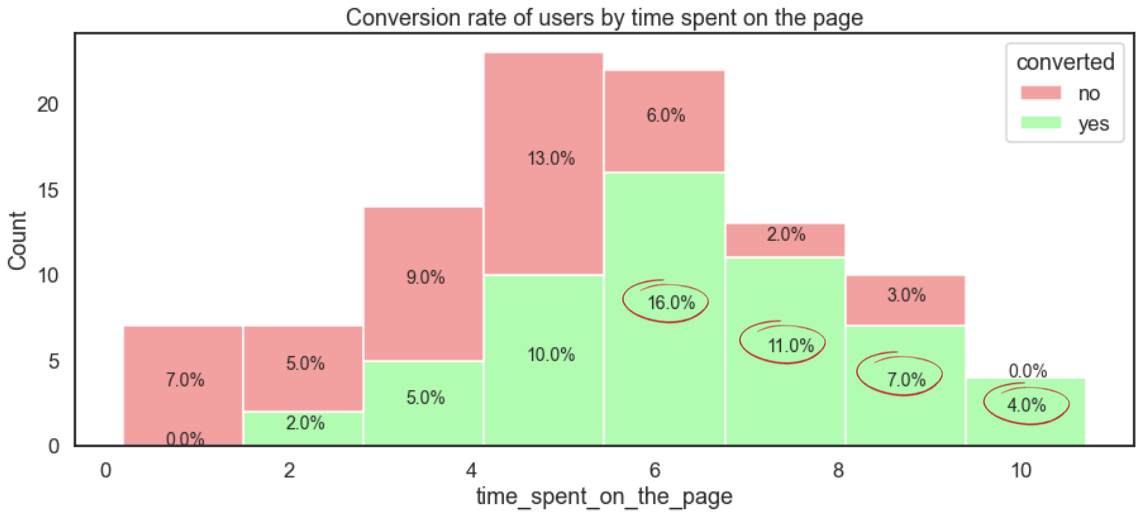
- More treatment users / New landing page have been converted.
- Users landing on the new page spent more time on the page compared to old page
- We can see a relation between users that spent more time on the page an got converted .

EDA – Bi-variate Multivariate



Observations

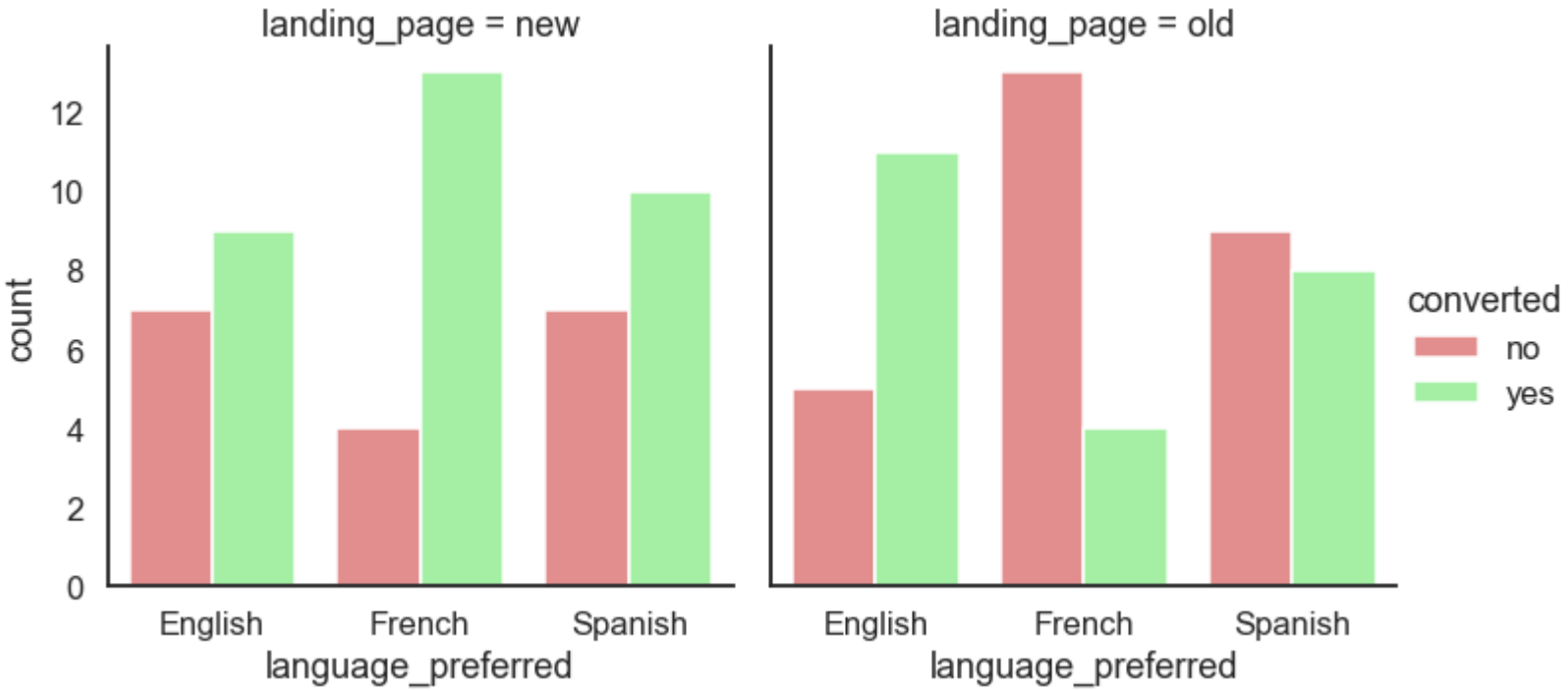
- Mean time spent on page is almost equal for all language preferred.
- Spanish users has small variation between time spent on page (concentrated between 4 to 7 min) and, 1 outlier.
- French and English users are more distributed between time spent on the page (greater variation).



Observations

- Greater the time spent on the page, greater the probability to get convert (subscribe).

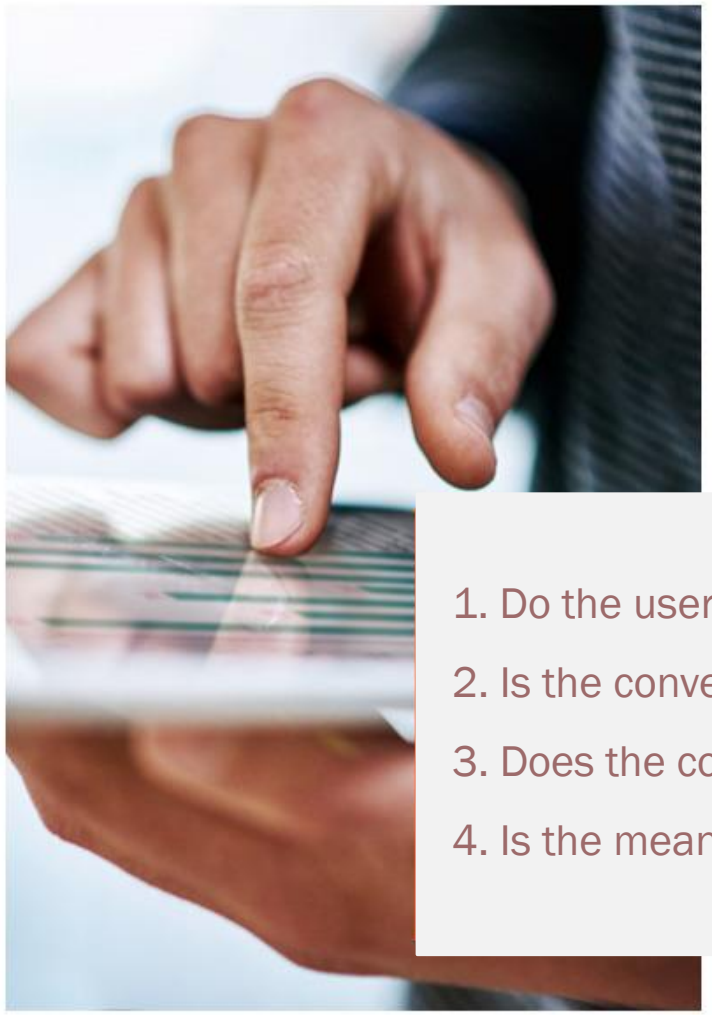
EDA – Bi-variate Multivariate



	English		French		Spanish		
	New	Old	New	Old	New	Old	Total
No	7	5	4	13	7	9	45
Yes	9	11	13	4	10	8	55
All	16	16	17	17	17	17	100

Observations

- French users get the opposite conversion status per landing page, showing preference for the New landing page (treatment group).
- English users got more conversion status on the Old landing page (control group). Needs more analysis to understand preference
- Spanish users doesn't show much variation between converts status and landing page; it also needs more analysis to understand preference



Key Questions:

1. Do the users spend more time on the new landing page than the old landing page?
2. Is the conversion rate for the new page greater than for the old page?
3. Does the converted status depend on the preferred language?
4. Is the mean time spent on the new page same for the different language users?

Time spent between landing page

Two sample t-test (independent)

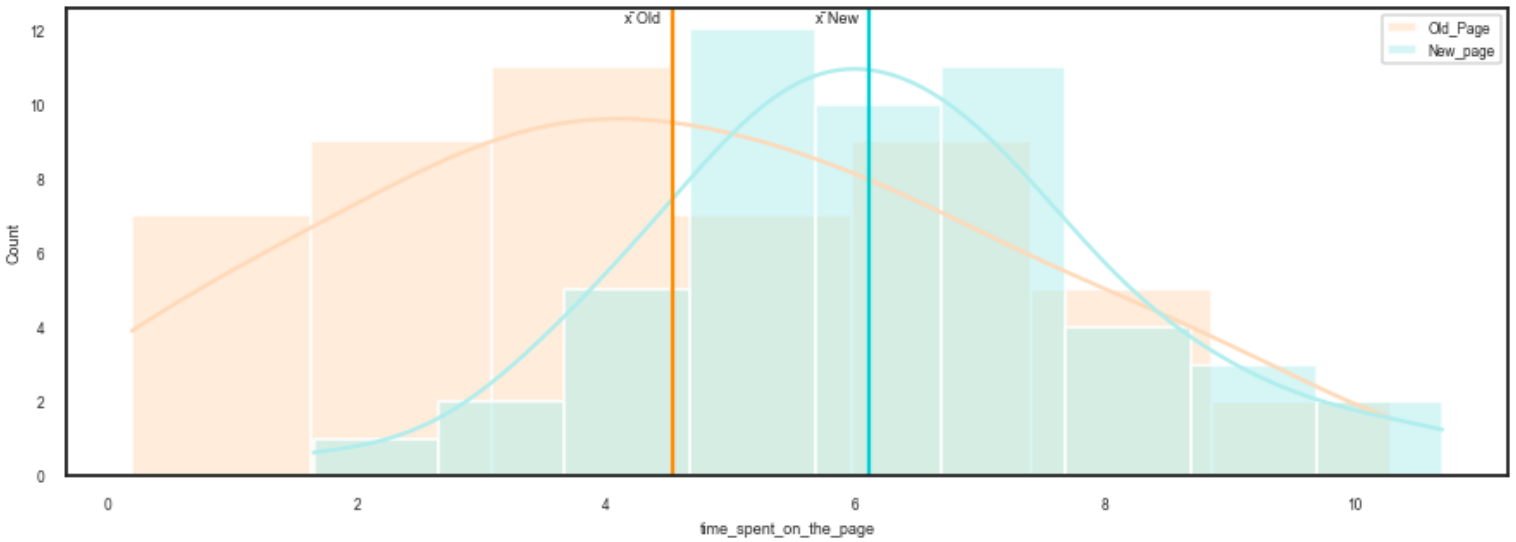
Assumptions:

- ✓ Normally distributed
- ✓ Random sampling from the population
- ✓ Continuous data (spent time on page)
- ✓ Independent populations
- ✓ Unequal population standard deviations

Right tailed t-test.

$H_0 : \mu_{new} = \mu_{old}$ (New landing page did not raise the spent time on page)

$H_a : \mu_{new} > \mu_{old}$ (New landing page did raise the spent time on page)



SPENT TIME ON:	Mean	Std	Observations	T-Stats	P-Value	Significance level (α)
New page	6.223	1.82	50	3.7868	0.00014	0.05
Old page	4.532	2.58	50			

P-value < α
(~0.00014) < 0.05

We have stronger evidence to:
Reject the null hypothesis

There is less than 0.01% of chance of mean spent time on New Landing page to be equal a mean spent time on Old Landing .

The spent time on the new landing page is greater compared to the old.

Converted rate greater for new page

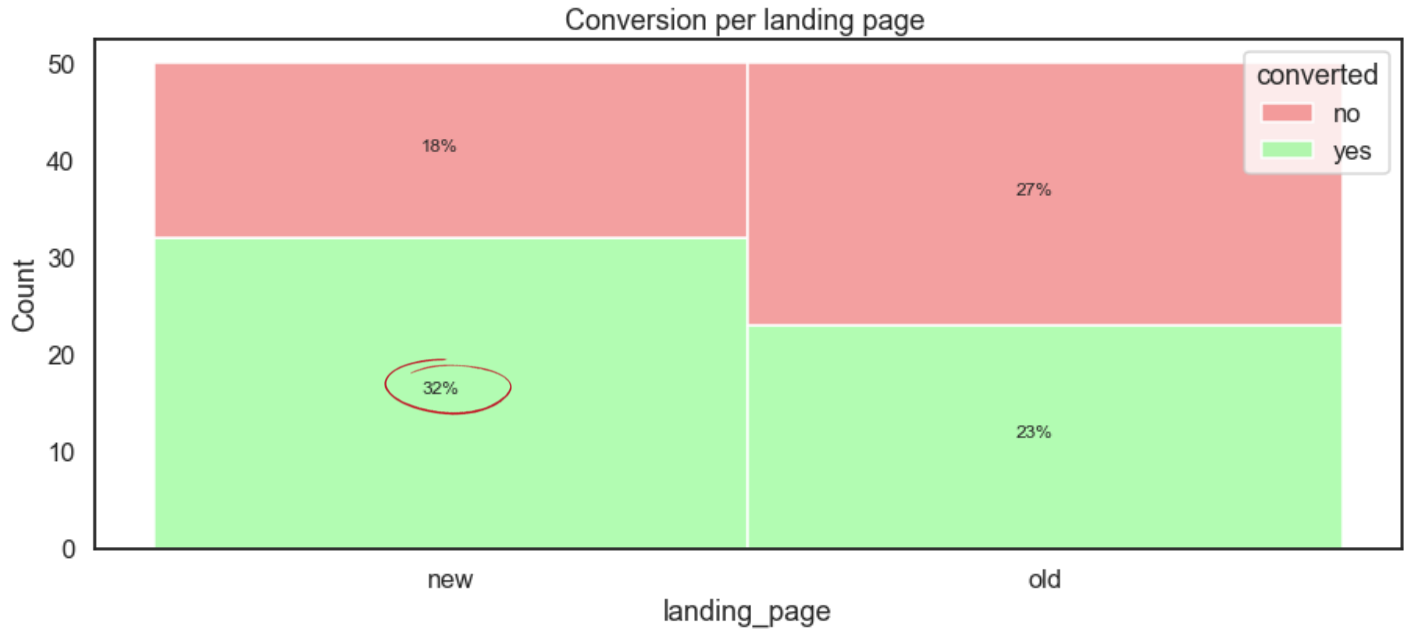
Two proportion Z-Test:

- ✓ Random sampling from the population
- ✓ Binomial distribution approximated to normal distribution

Right *tailed t-test.*

$H_0 : p_{new} = p_{old}$ (Converted rate did not raise for New landing page)

$H_a : p_{new} > p_{old}$ (Converted rate did raise for New landing page)



CONVERTED RATE	Converted Count	Sample size	T-Stats	P-Value	Significance level (α)
New page	32	50	1.8091	0.0352	0.05
Old page	23	50			

P-value (~0.0352) < α 0.05

We have evidence to:
Reject the null hypothesis

and conclude that Converted rate did raise for New landing page at 0.05 significance level.

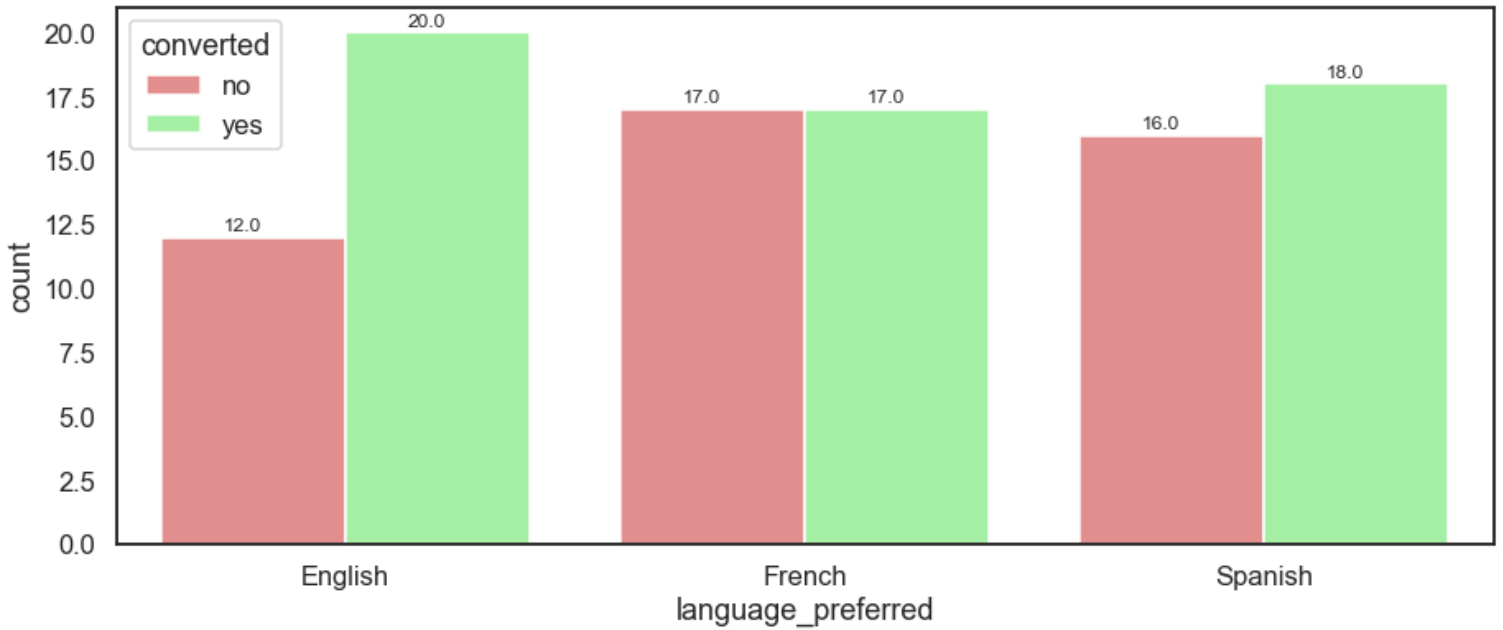
Converted status depend on the preferred language

Chi-Square Test for independence:

- ✓ Categorical Variables
- ✓ Random sampling from the population
- ✓ Expected value of the number of sample observations in each level of the variable is at least 5

H_0: Converted status and preferred language are independent

H_a: Converted status and preferred language are not independent



Language preferred	English		French		Spanish		
Converted status	Observed	Expected	Observed	Expected	Observed	Expected	Total
No	12	14.4	17	15.3	16	15.3	45
Yes	20	17.6	17	18.7	18	18.7	50
All	32	32	34	34	34	34	100

Degrees of freedom	Chi-Value	P-value	Significance level (α)
2	1.1289	0.5687	0.05

P-value (~0.5687) > α 0.05

We fail to Reject the null hypothesis

A chi-square test of independence was performed to examine the relation between converted status and the user preferred language.

We have strong evidence (more than 50%) that the converted status and language are independent.

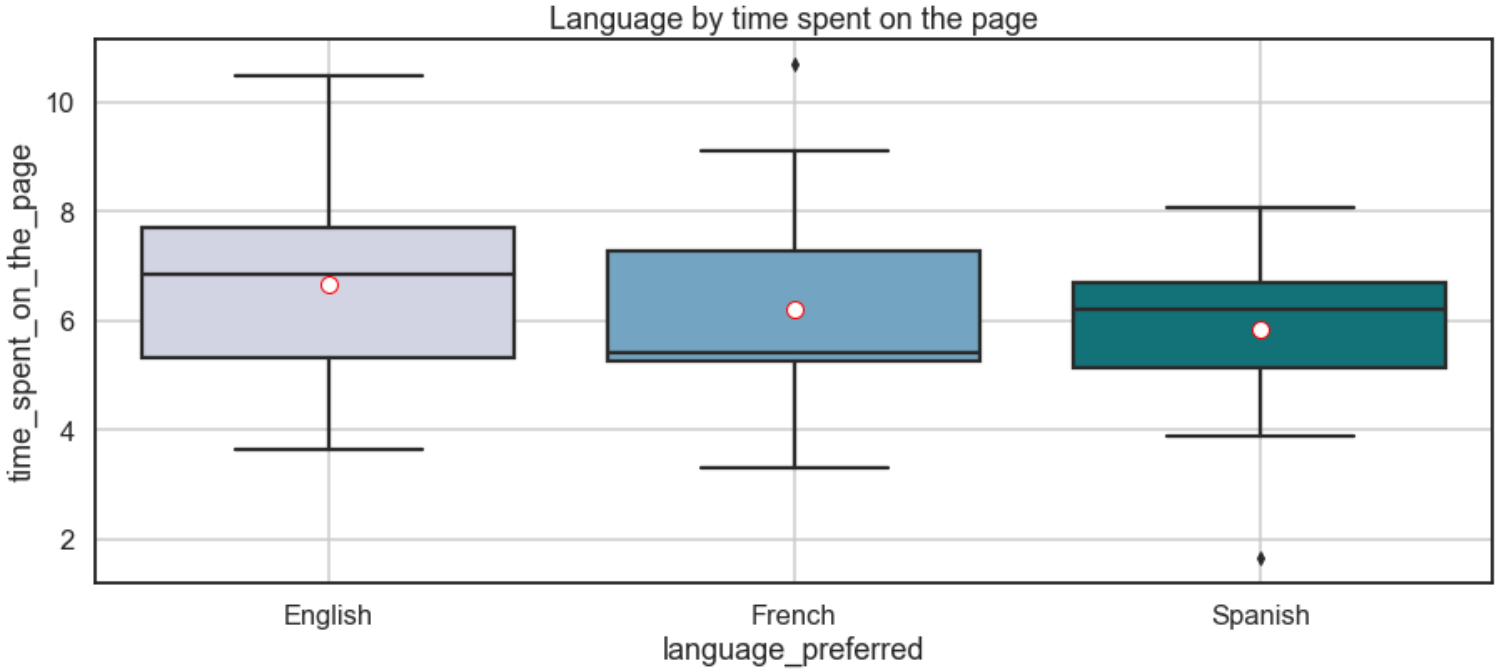
Mean time spent on New landing page same for different language users

Analysis of variance (ANOVA):

- ✓ Normally distribution
- ✓ Random sampling from the population
- ✓ Population standard deviation equal.

$H_0 : \mu_{English} = \mu_{Spanish} = \mu_{French}$

H_a: At least one of these means is not the same



F-ratio value	P-value	Significance level (α)
0.8544	0.4320	0.05

P-value (~0.4320) > α 0.05
We fail to Reject the null hypothesis

This means that we have strong evidence that the mean time spent on the new page are the same for different language users

Business Insights and Recommendations

Insights

1. Users of the New landing page usually spent more time on the page (*Slide 6):*
    i. *Control/Old page:* 6% of users spent less than 1 min on the page, and 8 % of users less than 3 min.
    ii. *Treatment/New page:* 1% of users spent between 1 to 2 min and other users spent more than 3min
2. New landing page (treatment group), has a greater converted rate compare to the old page.
    i. There is a relation between users that spent more time on the page an got converted (slide6/7).
3. Language preferences its almost equal distributed between the two types of pages.
4. Converted status is independent of language preferred;
5. The mean spent time on page doesn't change for users with different language preference.

Recommendations for further investigate:

1. Deep analyze on the relationship between language preferred Vs preference landing page (Slide8):
    i. New landing page and Old landing page has similar numbers of users preferred language, with different converted preference (French seems to like New landing page, English seems to like the Old)
2. Users who spent more time on page, got converted, we should analyze if this users after became subscribers, if they change the spent time (Paired T-test, same users, before and after subscribe).
3. Why does New Landing page make users spent at least more than 1 min on the page?
4. What does make users spent more time on the page?
    i. Analyze profile of users that spent more than average on the page.

Amanda Mendonca