E-News Express

Analysis of the new landing pages.

Background

- E-News is an online news portal that aims to expand its business by acquiring new subscribers. Every visitor to the website takes certain actions based on their interest. The company plans to analyze these interests and wants to determine whether a new feature will be effective or not. Companies often analyze users' responses to two variants of a product to decide which of the two variants is more effective. This experimental technique is known as a/b testing is used to determine whether a new feature attracts users based on a chosen metric

Objective

E-News wants to decide whether the new landing page is more effective to gather new subscribers. To achieve the later one is requested to fulfill the following points:

✓ Explore the dataset and extract insights.

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

✓ 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?

✓ Does the converted status depend on the preferred language?

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

Data Dictionary

The following elements compose the data base:

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.

Exploring the dataset

Data Overview

The shape of the database consists of 6 columns and 100 rows.

The original base had the following types of variables:

- int64 – user_id
- float64 – time_spent_on_the_page
- object – group, landing_page, converted and language_preferred. These object type variable were transformed later to a category type.

Finally, the database does not contain any missing value that requires a transformation

	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

Data Overview

The five-point summary for the variable **time_spent_on_the_page** is the following:
- Minimum: **0.19** minutes
- Percentile 25: **3.88** minutes
- Median: **5.42** minutes
- Percentile 75: **7.02** minutes
- Maximum: **10.71** minutes

This study consist of 100 unique users.

The **group** variable consists of the control and treatment group, each one has 50 elements.

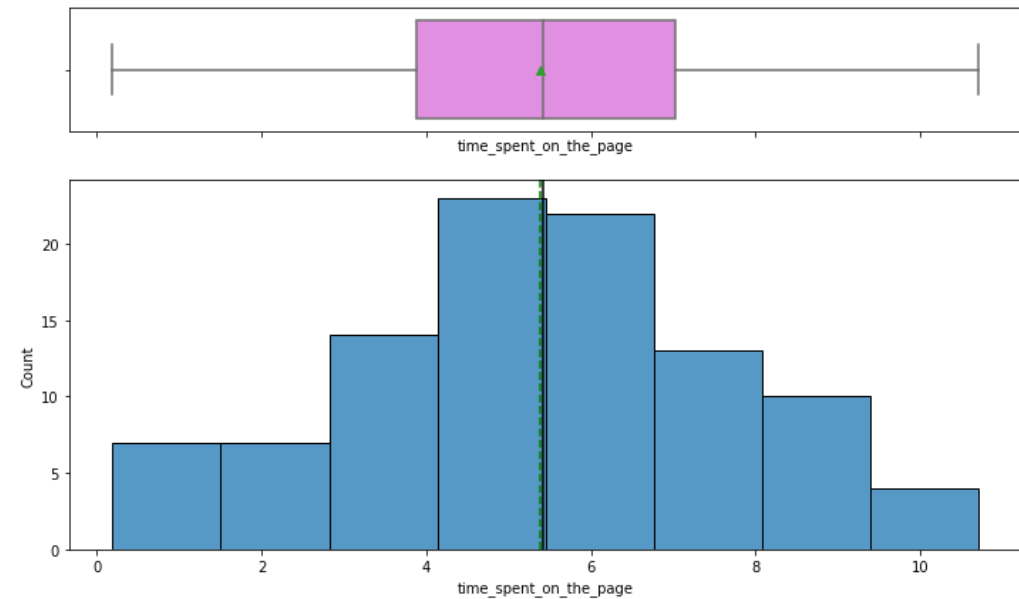The **landing_page** variable consists of the "new" and "old" groups each with 50 elements.

The **converted** variable has 54 subscribed users and 46 users who didn't.

The language_preferred by the users are English with 34 users, Spanish with 34 users, and French with 32 users.

Univariate Analysis

Time spent on the page

- The time spent on the page tends to form normal distribution around the mean of 5.37 mins of time spent on the landing page.

- The distribution of the time spent on the page doesn't show any outlier value, as all the time is within the normal distribution range.
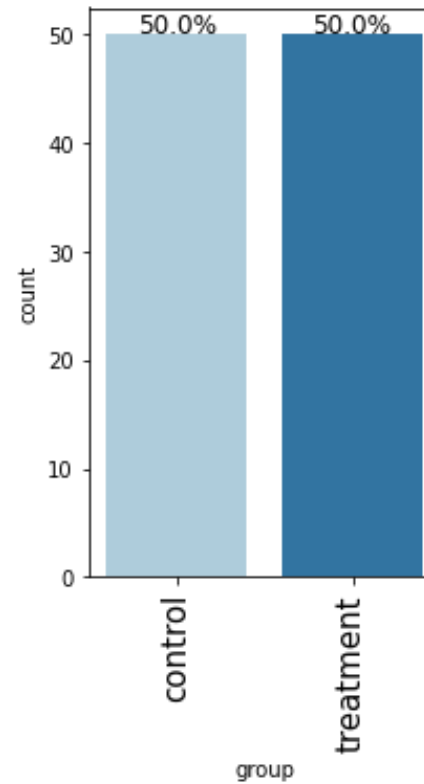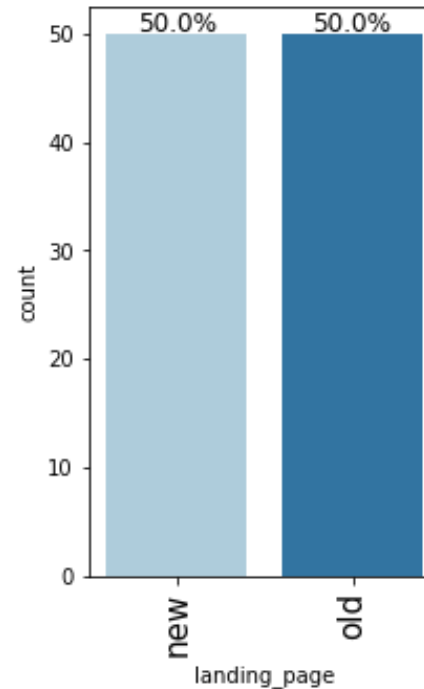
Univariate Analysis

Group

- The Group variable shows an equal distribution between the control and treatment groups. This distribution helps for the analysis of the two populations.

Univariate Analysis

Landing_page

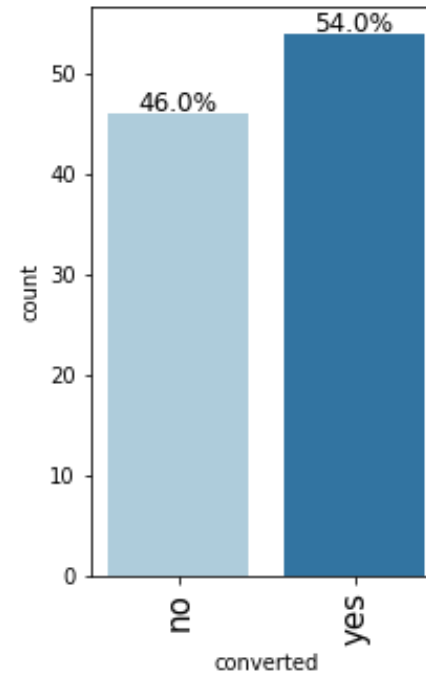- The landing_page variable shows an equal distribution between the new and old, corresponding to the respective group as shown in the previous variable.

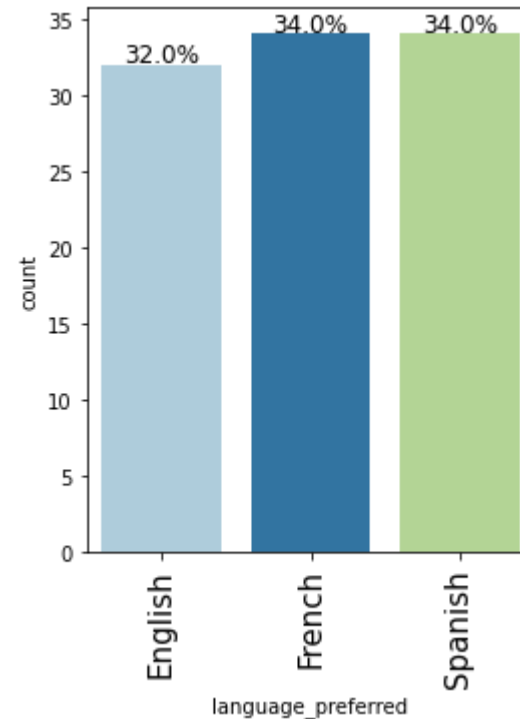# Univariate Analysis

Converted

- The converted variable shows that 54% of the users decided to subscribe to the service of E-News.
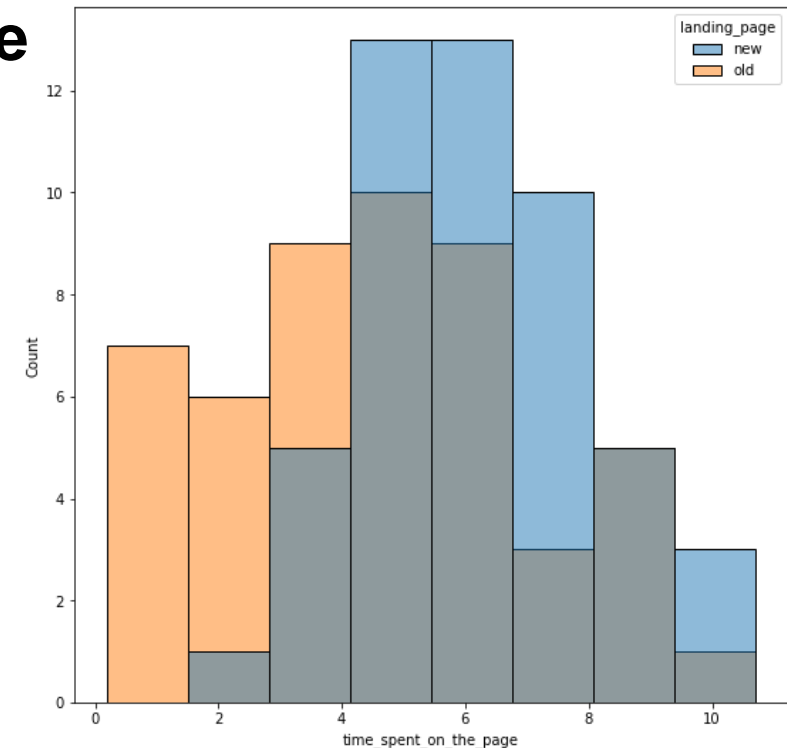
Univariate Analysis

Language preference

- The langue preference variable shows an almost equal distribution between the preferred languages for the site.

Bivariate Analysis
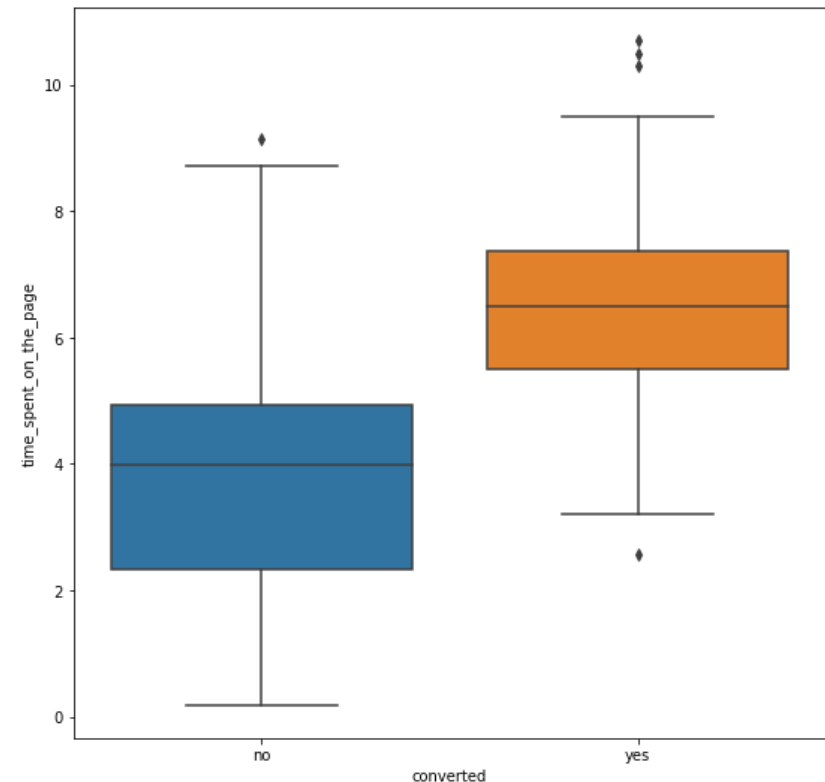
Landing page vs Time spent on the page

- The time spent on the page tends to form normal distribution around the mean of 5.37 minutes of time spent on the landing page.

- The comparison between the old and the new landing page shows that the users tend to stay more time on the new page rather than the old one.

Bivariate Analysis

Conversion status vs Time spent on the page

- The conversion to a subscribed user correlates with the time spent on the page. The converted status with a no shows some users as an outlier with the time spend and not being subscribed. On the other hand, some users on the yes converted status show that has outliers on the bottom and top of the box.

Bivariate Analysis

Preferred language vs Time spent on the page

- The means between the spend time and the preferred language shows no significant difference between each other. The variance between them shows a little bit of difference but most of them show that are at a similar level. Also, the only outlier is in the Spanish language, the minimum value of 0.19 minutes.
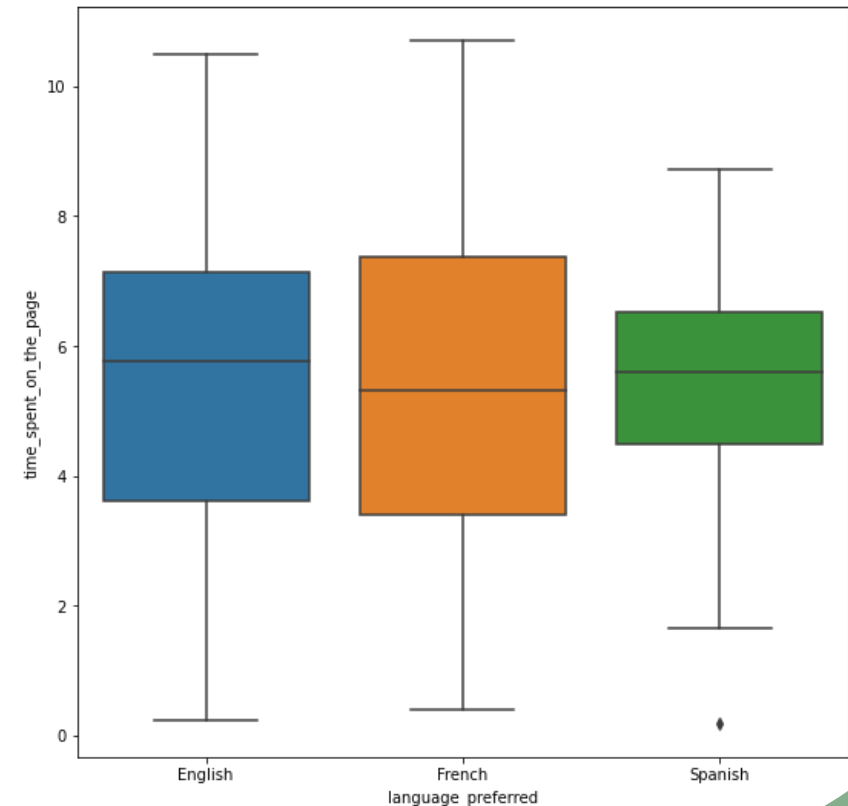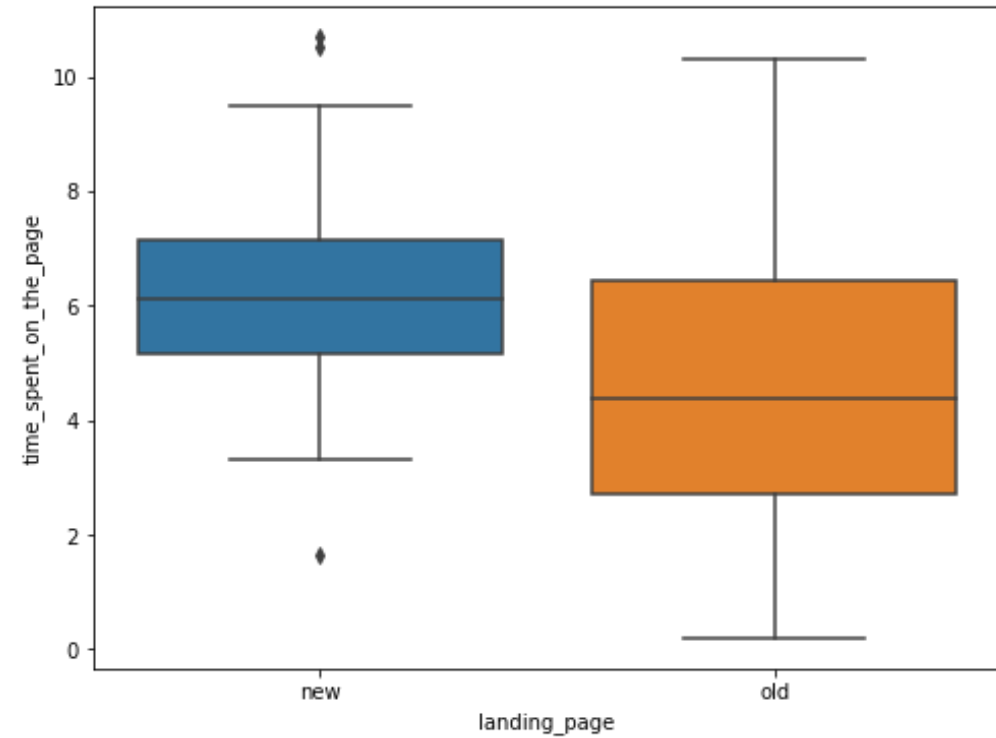
Do the users spend more time on the new landing page than the existing landing page?

Visual Analysis

The comparison between the spent time in the landing pages shows that the average time spend in the newest one is over a minute.

But this time difference is significant enough to consider the impact of the newest landing page?
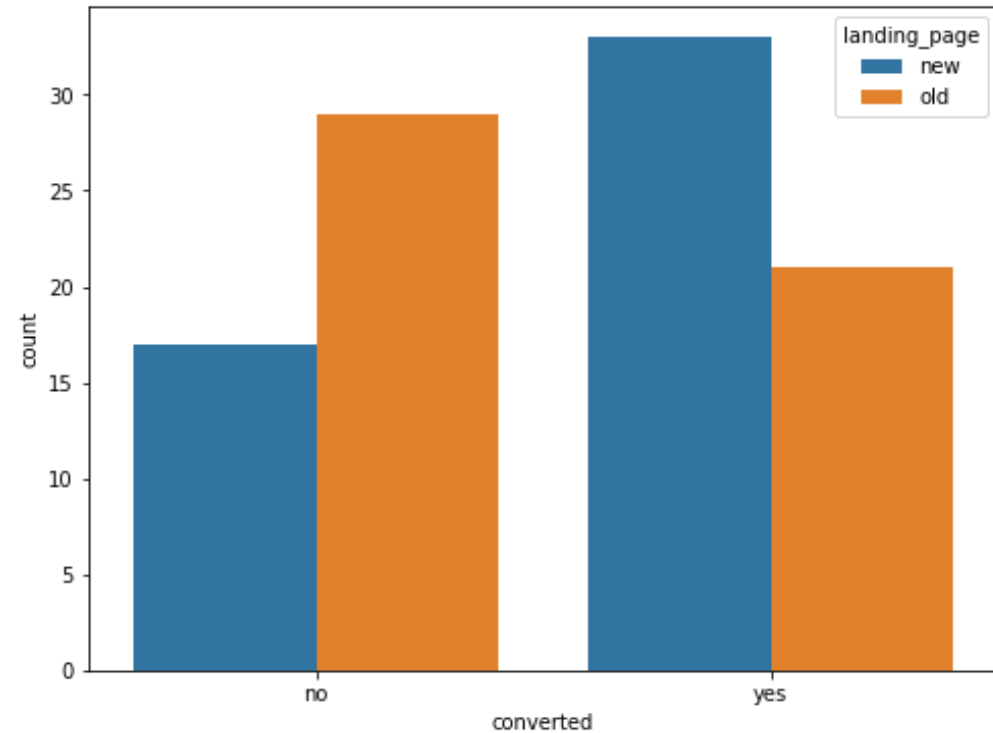
Test - Results

- In order to test the previous statement, the following expiring has been designed:

- H0: μ0 = μ1. Being μ1 is the new landing page and μ0 is the old one.

- H1: μ1 > μ0.

- As we don't know the parameters of the populations, we proceed to use a Test for equality of means. For this test, we are going to use an α = 0.05. Also, the standard deviation of the new page is 1.82 mins and of the old page is 2.58 mins. Based on this test the equality of variations will be false for the evaluation of this test.

- For this test the data base was divided by the groups classified by the landing_page variable, the new and old landing page.

- The resulting p-value of this test is **0.0001**. This level of significance is less than the significance level previously defined, therefore the null hypothesis is **rejected**.

- Based on these results it can be inferred that the meantime in the new landing page is greater than the one on the old page.

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

Visual Analysis

The comparison between the converted in the landing pages shows that the more users get converted in the newest one over the old one.

Is this difference significant enough to consider the impact of the newest landing page?
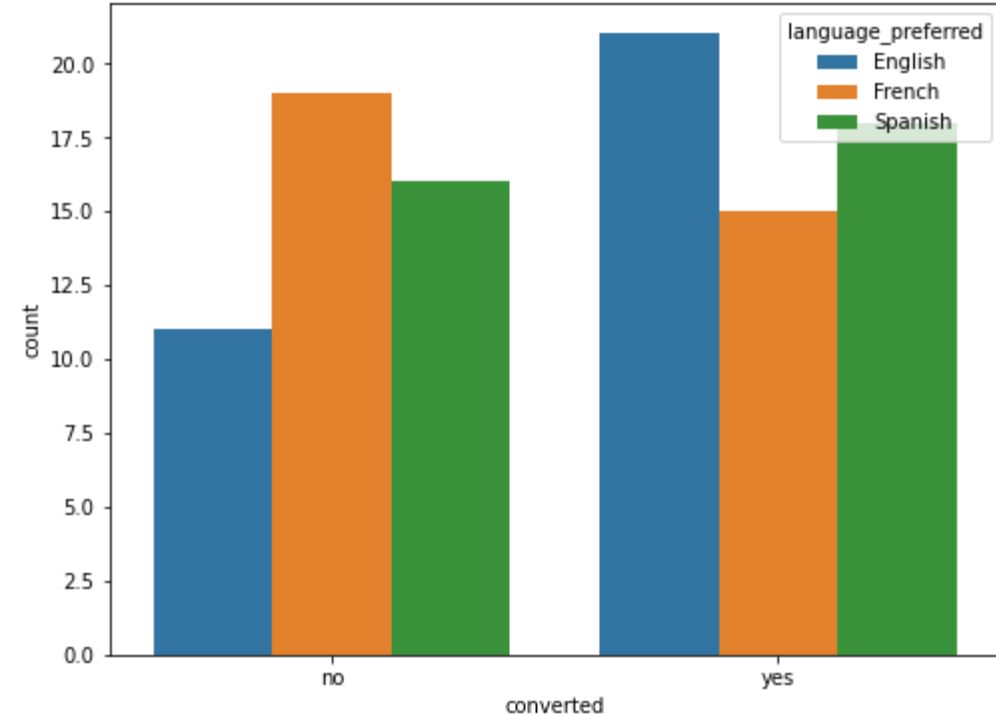
Test - Results

- In order to test the previous statement, the following expiring has been designed:

- H0: p0 = p1. Being p1 is the proportion of converted users in the new landing page and p0 is the proportion of the old one.

- H1: p1 > p0.

- As we want to compare the proportions between groups a Test for Two Proportions will be used. For this test, we are going to use an $\alpha = 0.05$. Also, the total number of elements in each group is 50 users.

- The data for this test was gathered by the converted users in each group. These values would give the proportion for the converted population on each sample group.

- The resulting p-value of this test is 0.0161. This level of significance is less than the significance level previously defined, therefore the null hypothesis is rejected.

- Based on these results it can be inferred that the proportion of converted users from the new landing page is significantly larger than those of the older page. Therefore, the new landing page is fulfilling its objective of

Does the converted status depend on the preferred language?

Visual Analysis

At a first glance, it seems to be that the English speakers are the ones with the highest conversion rate. The French and Spanish speakers seem to be on similar levels of conversion rates.

Is there a significant difference between the languages to consider the language as a relevant factor to convert users?

Test - Results

- In order to test the previous statement, the following expiring has been designed:

- H0: the preferred langue is independent.

- H1: the preferred language is not independent.

- This is a test to be realized with a Chi-Square test of independence concerning the two independent categorical variables converted and Language_preferred.

- To prepare the data a contingency table based on the language_preferred and converted giving the results in the following table.

- The resulting p-value of this test is 0.2130. This level of significance is above the significance level previously defined, s the hypothesis is accepted.

- Based on these results it can be inferred that the preferred language is independent of the converted users from the new landing page are significantly larger than those of the older page. Therefore, the users that are converted to subscribed users are independent of their preferred language.

| language_preferred	English	French	Spanish
converted			
no	11	19	16
yes	21	15	18

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

Visual Analysis

At a first glance, it seems to be that the English speakers are the ones with the highest spent time on the landing page. conversion rate. But the French and Spanish languages show a similar level of spent time.

Is there a significant difference between the languages to consider one of the languages as has more time spent on the landing page?
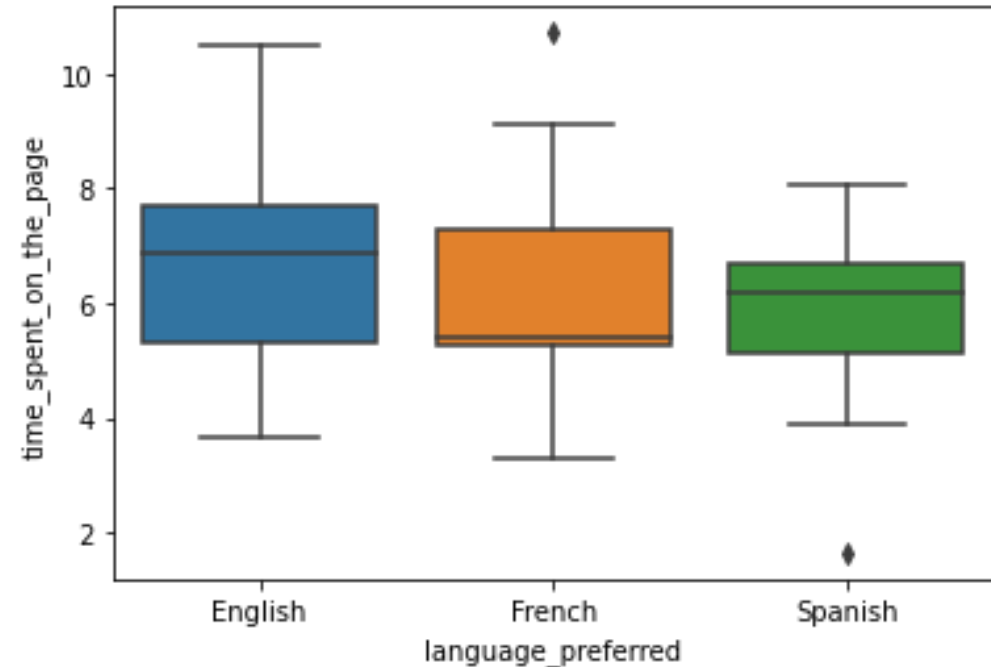
Test - Results

- In order to test the previous statement, the following experiments had been designed with an $\alpha = 0.05$:

- H0: The means of the time spent on the landing are the same.

- H1: The mean of at least one language on the landing page is different.

- To realize this, an ANOVA test must be realized. The depend variable is time_spent_on_the_page and the factor is the language_preferred.

- Before realizing the ANOVA test, a Shapiro-Wilk's test will be used for to evaluate if the distributions follow a normal distribution.

- H0: Time spent on the new page follows a normal distribution.

- H1: Time spent on the new page does not follow a normal distribution

- Later, a Levene's test will be used to determine if all the population variances are equal.

- H0: All the population variances are equal

- H1: At least one variance is different from the rest

Test - Results

- For the Shapiro-Wilk's test the p-value is 0.8040, since this value is above the 5% of significance previously defined the three languages follow a normal distribution.

- For Levene's test the p-value is 0.4671, since this value is above the 5% of significance previously defined the three languages have a similar variance in their distributions.

- Finally for the ANOVA test the p-value is 0.4320, since this value is above the 5% level of significance it is concluded the means of the time spent on the landing are the same.

Conclusions and Recommendations

Conclusions and Recommendations

- Based on the collected information of both groups it's recommended to use the new landing page. The new one helps to retain during a longer time the users.

- Following the previous results of the time spent on the new page, it shows that the users in the new group tend to be converted in a higher proportion compared to the old group.

- With the information of these two conclusions, it's recommended to maintain the design of the new landing page.

- With the language of the content presented it shows that the users don't have a preferred language. Therefore, for every language, a similar level of quality must be applied to the news the portal gives as it had been shown that users are well distributed.

- To support the previous claim an ANOVA test was realized. As it has been shown that the users in the three languages have a similar mean of spent time, the standard deviation of the time, and follow a normal distribution.

- Finally, it's recommended to invest in personnel who understand every language as they can help to reach a desired quality on the conte of the news.