greatlearning
Power Ahead

E-News Express
Project 2

Objective

To collect insight into what influences conversion of users of old vs new landing pages.

We will be focusing on asnwering the following questions -

- Do the users spend more time on the new landing page than the old 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 mean time spent on the new page same for the different language users?
- Is the new landing page more effective to gather new subscribers than the old landing page?

Data Information

The data contains the following information:

Variable	Description
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.

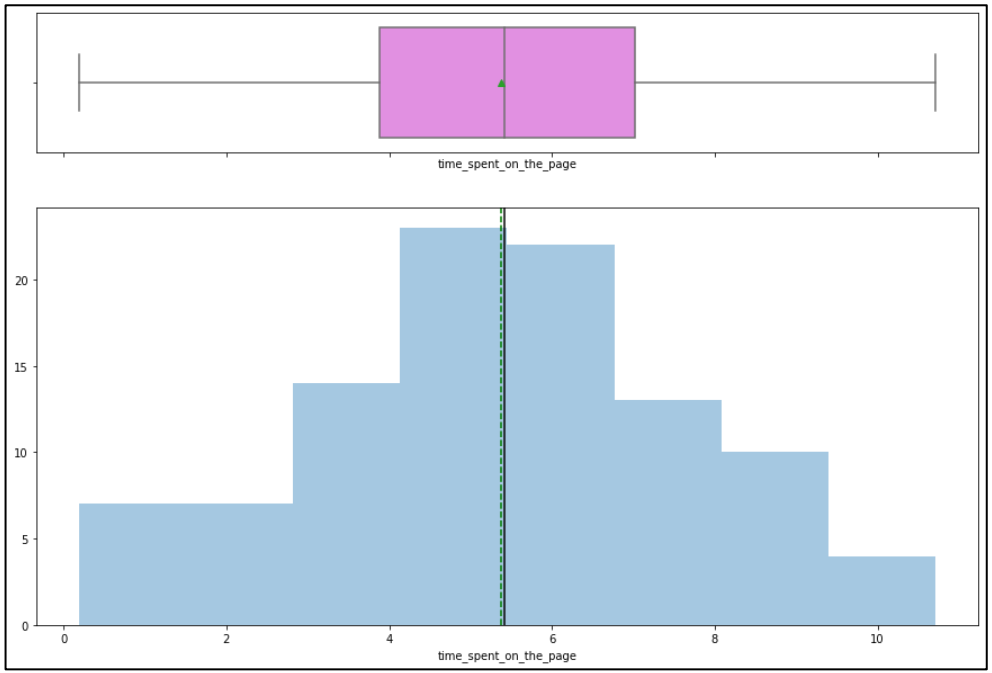
Observations	Variables
100	6

Note:

- There are no missing values in the dataset
- Group, landing page, converted, and language preferred have been converted to category data types

Exploratory Data Analysis

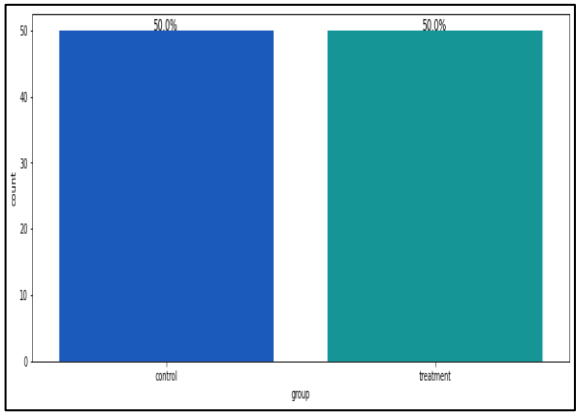
Time Spent on the Page



- As expected, based on the summary data above, the time spent on the page is uniformly distributed.
- The mean and the median are very close to the same value around the center of the dataset.
- There are no outliers present in the dataset.

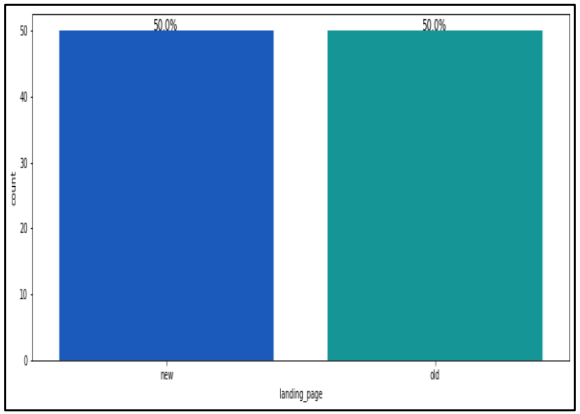
Exploratory Data Analysis – Time Spent on Page, Landing Page, Converted

Time Spent on the Page



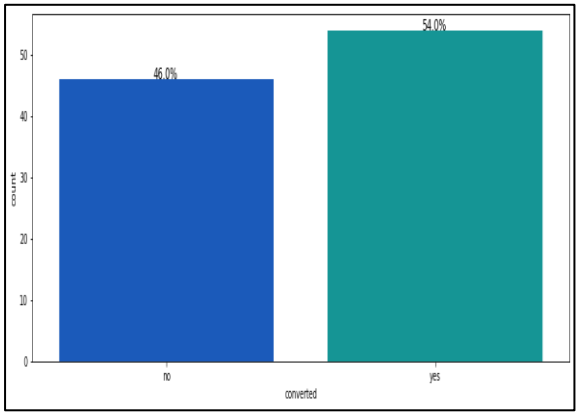
- The sample data is equally distributed between control and treatment groups

Landing Page



- The sample data is equally distributed between new and old landing pages

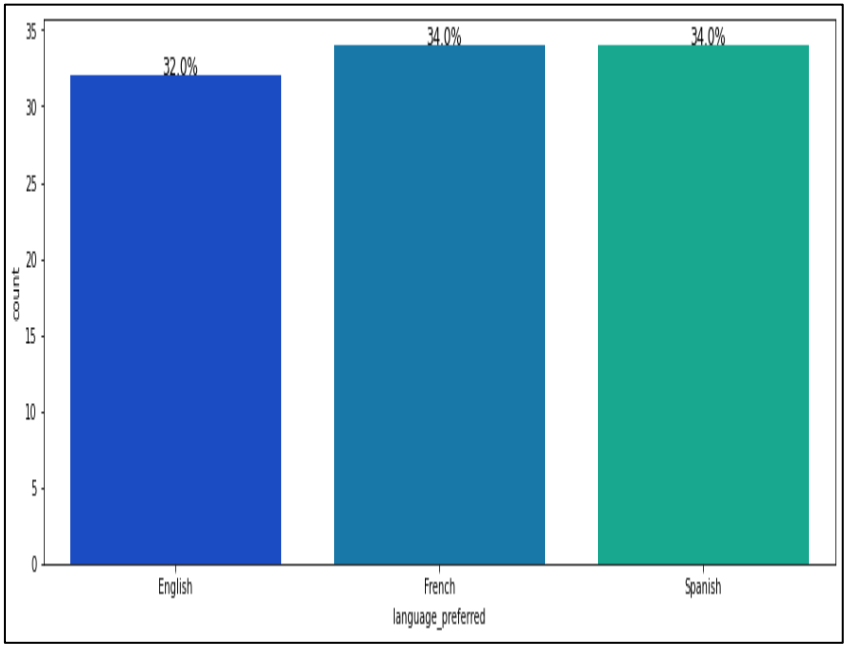
Converted



- 54% of users converted to subscription while 46% did not.

Exploratory Data Analysis – Language Preferred

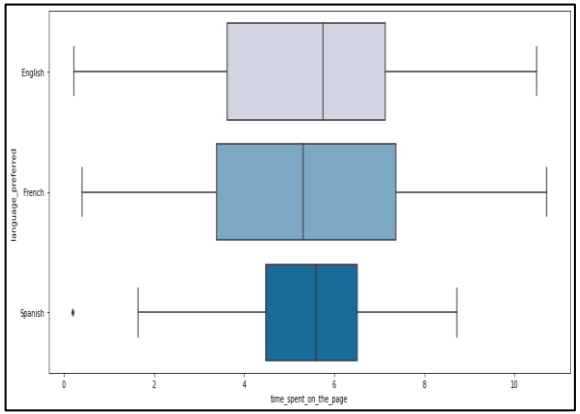
Language preferred



- 34% of users preferred Spanish.
- An additional 34% of users set their language to French.
- 32% of users prefer English.

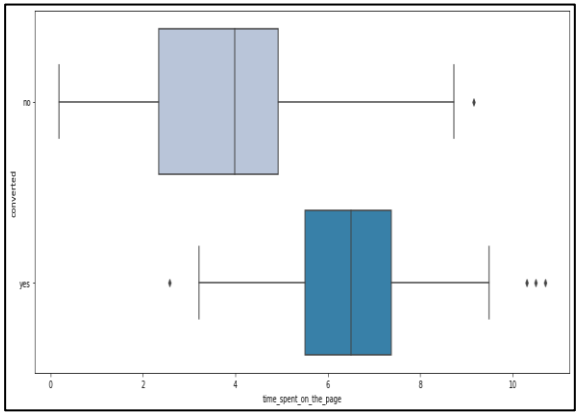
Exploratory Data Analysis – Time Spent on the Page and Language Preferred, Converted, Group

Language Preferred



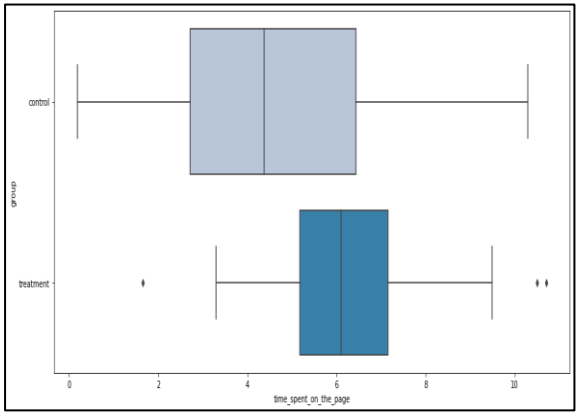
- Users who prefer Spanish spend a more consistent amount of time on the site with one outlier.
- Users who prefer English will on average spend more time on the site.
- Users who prefer French vary more on the amount of time spent on the site.

Converted



- Users who spent more time on the site on average also converted to subscribe.

Group

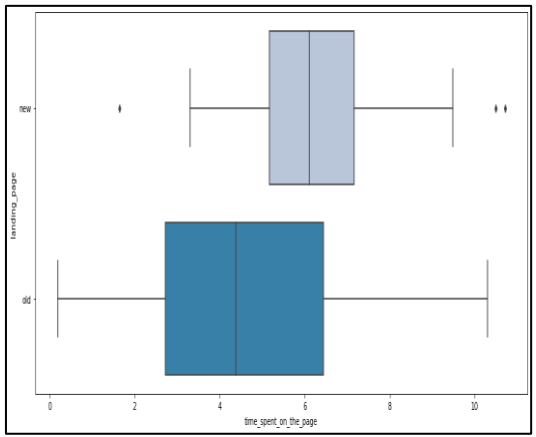


- Users in the treatment group on average spend more time on the site.
- Users in the control group have a wider range of time spent on the site.

Exploratory Data Analysis – Time Spent on the Page, Landing Page

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

Landing Page




```
The mean time spent on the new landing page is 6.2232
The mean time spent on the old landing page is 4.532400000000001
The standard deviation of time spent on the new landing page is 1.82
The standard deviation of time spent on the old landing page is 2.58
```

Defining the null and alternative hypothesis:

> H_0 : Users spend an equal amount of time on the new landing page that they do on the old page

> H_a : Users spend more time on the new landing page than the old landing page

Testing whether T-test assumptions are satisfied or not

- Continuous data - Yes, time spent on the landing page is measured on a continuous scale.
- Normally distributed populations - Yes, the populations are assumed to be normal.
- Independent populations - As we are taking random samples for two different groups, the two samples are from two independent populations.
- Unequal population standard deviations - As the sample standard deviations are different, the population standard deviations may be assumed to be different.
- Random sampling from the population - Yes, we assume that the collected sample a simple random sample.

We can use two sample T-test for this problem.

- As the p-value (~0.00013) is less than the level of significance, we can reject the null hypothesis.
- Hence, we can conclude that users spend more time on the new landing page than the old landing page.

Observation on Conversion

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

Defining the null and alternative hypothesis:

> H_0 : The conversion rate for the new page is equal to the conversion rate for the old page

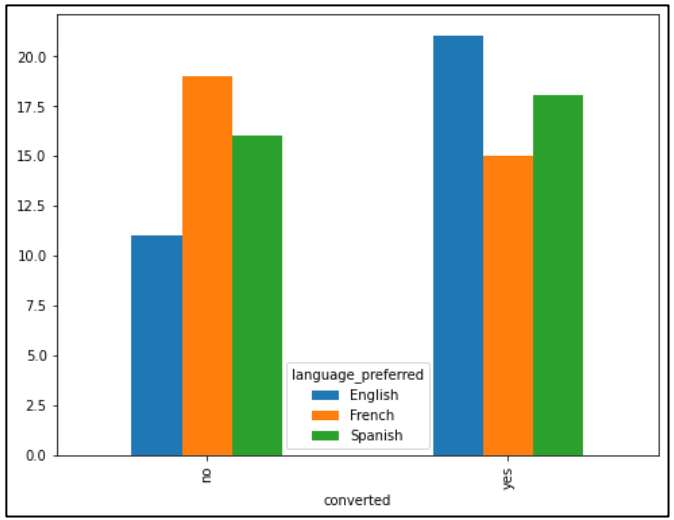
> H_a : The conversion rate for the new page is greater than the conversion rate for the old page

The formulated hypotheses are concerned with proportions. A test of proportions can be used to analyse the hypotheses and draw a conclusion. Therefore we can use a proportions Z test for this problem.

- The numbers of new and old page conversions are 33 and 21 respectively.
- The total numbers of new and old page visits are 50 and 50 respectively.
- The proportions of new and old page conversions are 0.17, 0.24 respectively.
- The p-value is 0.008.
- As the p-value 0.008026308204056278 is less than the level of significance, we reject the null hypothesis.
- The conversion rate for the new page is greater than the conversion rate for the old page.

Observation on Language Preferred and Converted

Does the converted status depend on the preferred language?



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

Defining the null and alternative hypothesis:

> H_0 : The converted status is dependent on the preferred language

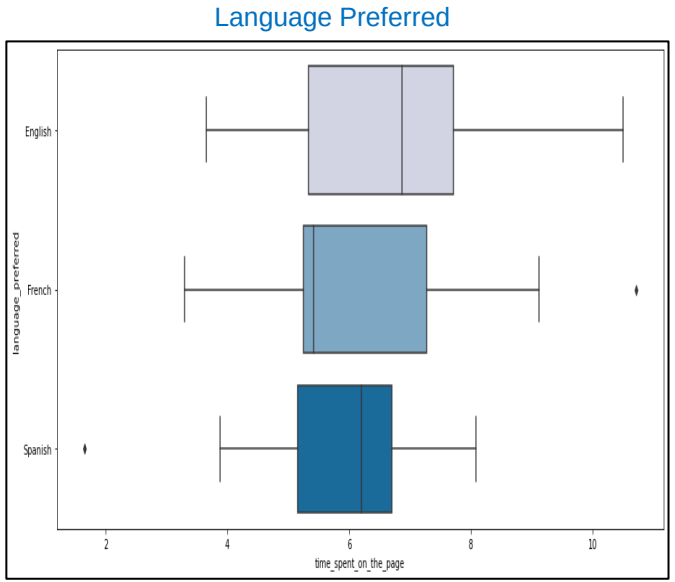
> H_a : The converted status is NOT dependent on the preferred language

The formulated hypotheses can be tested using a Chi-square test of independence of attributes, concerning the two categorical variables, converted or not converted.

- Since the p-value (~0.212) is > 0.05, we fail to reject the null hypothesis.
- Hence, we do not have enough statistical evidence to say that the convert status is not independent on the preferred language.

Exploratory Data Analysis – Time Spent on the New Page and Language Preferred

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




```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
====================================================
 group1  group2 meandiff p-adj   lower  upper  reject
----------------------------------------------------
English  French  -0.4673 0.7259 -2.0035  1.069  False
English Spanish  -0.8285  0.401 -2.3647 0.7078  False
 French Spanish  -0.3612  0.816  -1.874 1.1516  False
----------------------------------------------------
```

Defining the null and alternative hypothesis:

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_a : The mean time spent on the new page is different for the different language users.

- For multiple comparison testing, we will use Tukey HSD

- As the p-values (refer to the p-adj column) for comparing the mean time spent on the page are all greater than the significance level, the null hypothesis of equality of all population means can be rejected.
- Thus, we can say that the mean time spent on the new page is not dependent on the language preference of the user.

Conclusion

After all the analysis, we have been able to conclude:

- The sample data is equally distributed between the control and treatment as well as new and old landing page visits.
- On average users spend more time on the new landing page than the old landing page.
- There are more users in the data set who converted to subscribed users (54% converted compared to 46%% who did not).
- There was an equal distribution between users who prefer French and Spanish while there was 2% less users who prefer English.
- On average users who prefer English spend more time on the page followed by users who prefer Spanish.
- On average users who spend more time on the new page also converted to subscribe.
- Users in the treatment group on average spent more time on the page.
- Users spend more time on the new landing page than the old landing page.
- The conversion rate for the new page is greater than the conversion rate for the old page.
- We do not have enough statistical evidence to say that the convert status is not independent on the preferred language.
- The mean time spent on the new page is not the same for the different language users; English language users on average spend more time on the new page.

Recommendations

After all the analysis, we suggest the following recommendations:

- The new landing page design holds users for more time than the old landing page based on the mean time spent between the two.
- Additionally, the conversion rate supports the success of the new landing page in converting users to subscribers.
- Based on these two factors and inferenced based on the p-value, we can say that the new page is more successful at converting users to subscribers.
- More English users converted to subscribers; therefore E-News should focus on English speaking countries.
- E-news should focus their attention to the new landing page for higher conversion rate of subscriptions.

greatlearning
Power Ahead

Happy Learning !