greatlearning
Power Ahead

Business Presentation

Contents

Business Problem Overview and Approach

Data Overview

EDA

Hypothesis Tests

Business Insights and Recommendations

Business Problem Overview and Solution Approach

Core business idea

- Analyze the interests of the users and determine if the new feature will acquire new suscribers

Problem to tackle

- Identify the user segments for the user groups control and treatment.
- Identify deficiencies if any in the current target segmentation

Financial implications

- Expand the business by acquiring new subscribers.

Data Overview

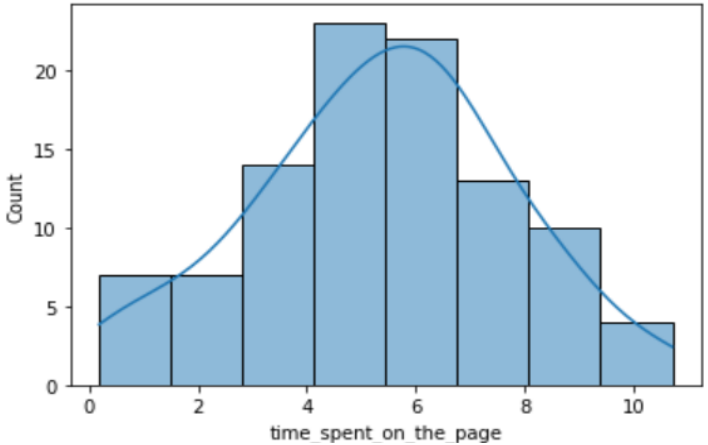
Variable	Description
1. user_id	This represents the user ID of the person visiting the website.
2. group	This represents whether the user belongs to the first group (control) or the second group (treatment).
3. landing_page	This represents whether the landing page is new or old.
4. time_spent_on_the_page	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.

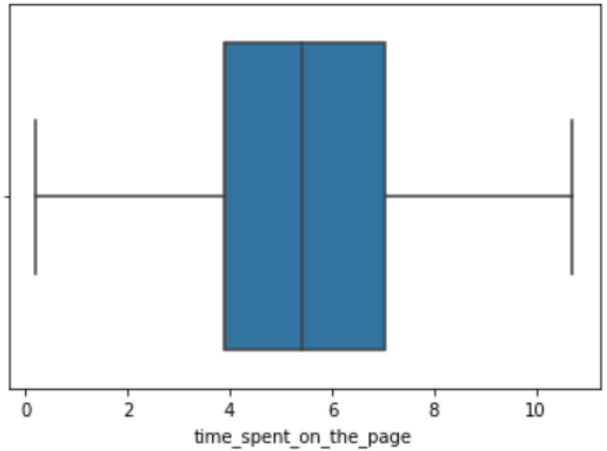
Observations	100
Variables	6

Note:

- There are no missing values.
- time_spent_on_the_page is the only numerical variable and the rest are object variable. All the object variables have been converted to categorical variables.

Exploratory Data Analysis – time_spent_on_the_page

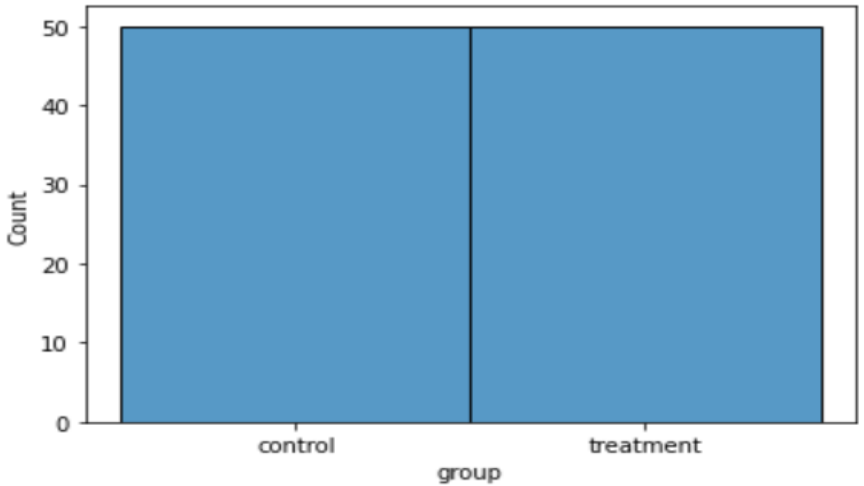




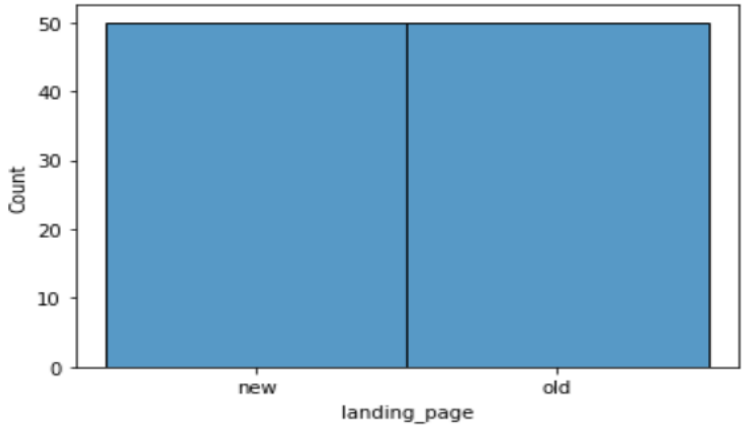
- Most of the users use the page between 4 -6.5 minutes.
- Mean time_spent_on_the_page is 5.4 minutes and the median is 5.4 minutes.

- 1. time_spent_on_the page doesnot have any outliers.

EDA – Categorical Variables– Group, landing_page

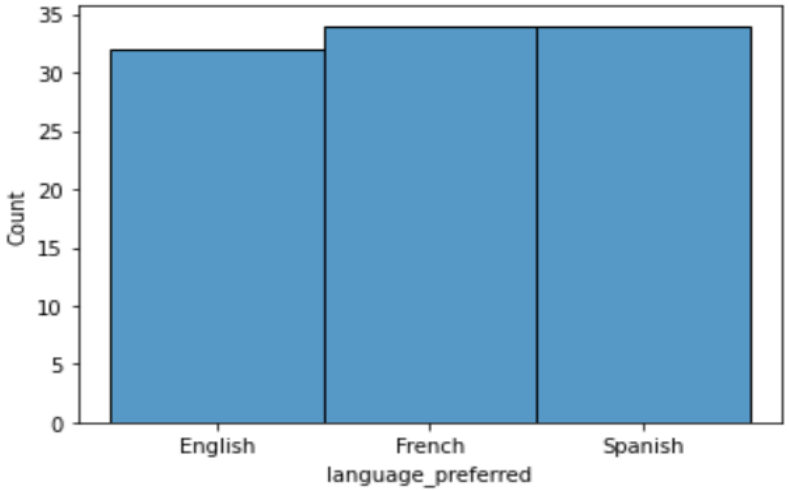


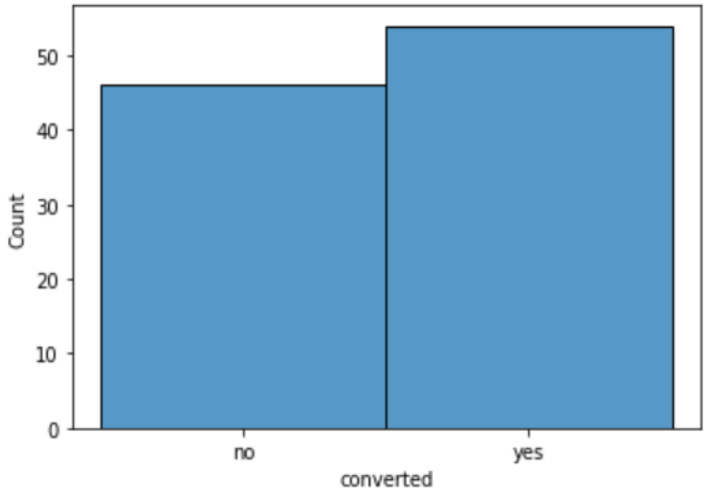
- **Observation**: 50% of the customers belong to the group control and the rest belong to the group treatment.



- **Observation**: The landing_page of 50% of the users is old and the rest is new.

EDA – Categorical Variables – language_preferred, converted





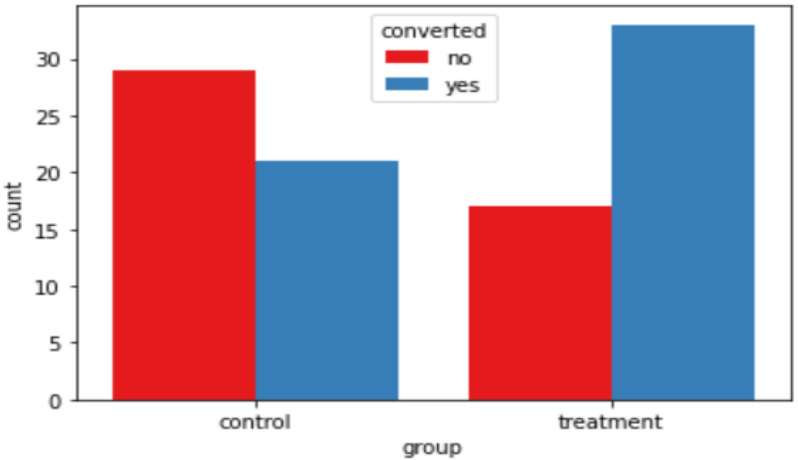
- **Observation**:

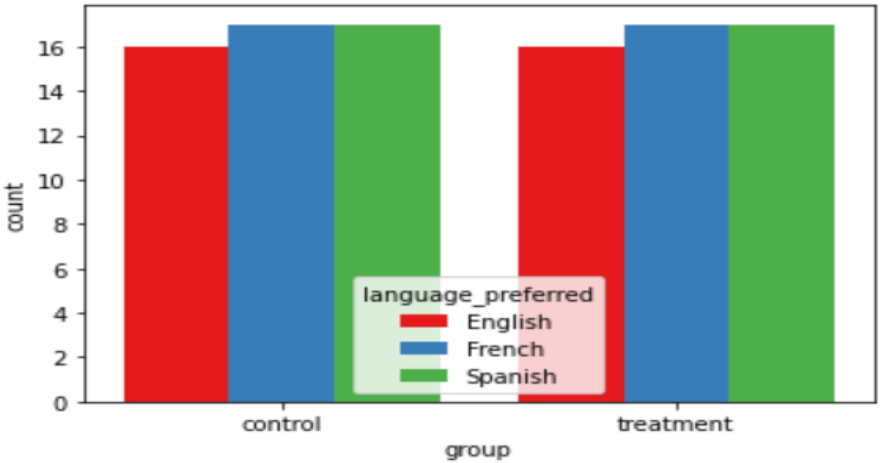
Most of the user either prefer Spanish or French

- **Observation**:

Out of 100, 54% percent of the users have subscribed.

EDA – group count by converted, language_preferred





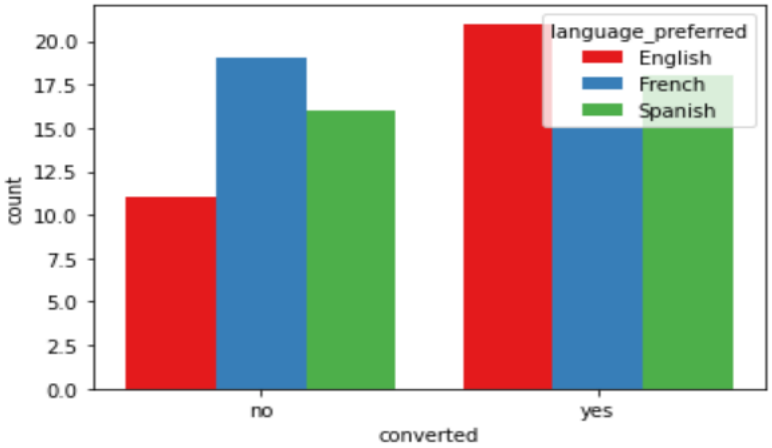
Observation:

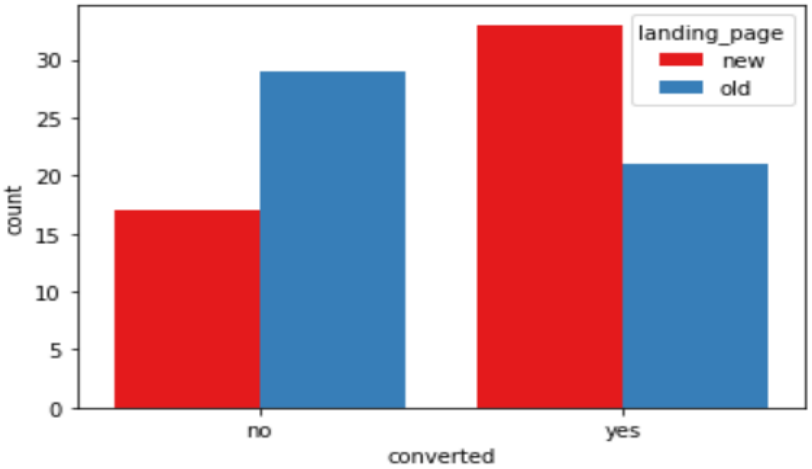
- There are more users from treatment group who converted to subscriber of the news portal.
- There are more user from control group who did not convert to a subscriber.

Observation:

- Number of users who choose French, Spanish and English are same among both the groups - control and treatment.

EDA - converted count by language_preferred, landing_page





- **Observations**:

The users who converted to subscriber mostly preferred English.

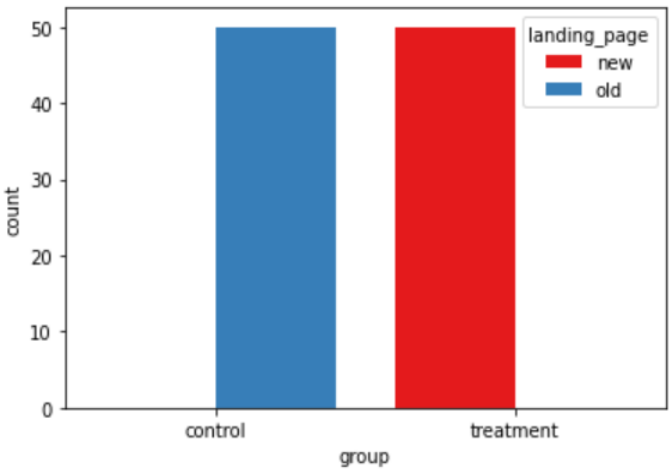
The users who didnot convert to subscriber mostly preferred French language.

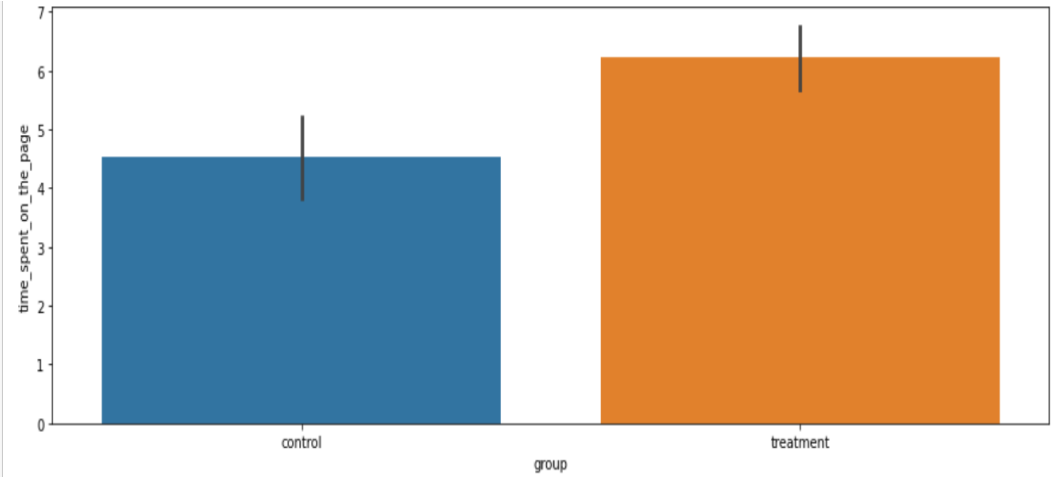
Observation

1.The users who converted to subscriber mostly preferred new_landing_page.

2.The users who didnot convert to subscriber preferred old_landing_page.

EDA – group count by landing_page, Time_spent_on_the page Vs group





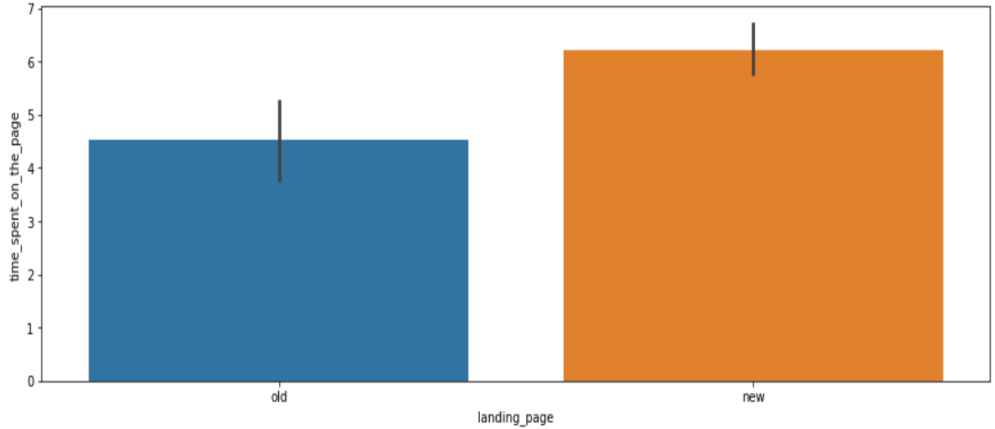
Observation

- Control group preferred old_landing_page.
- treatment group preferred new_landing_page.

Observations

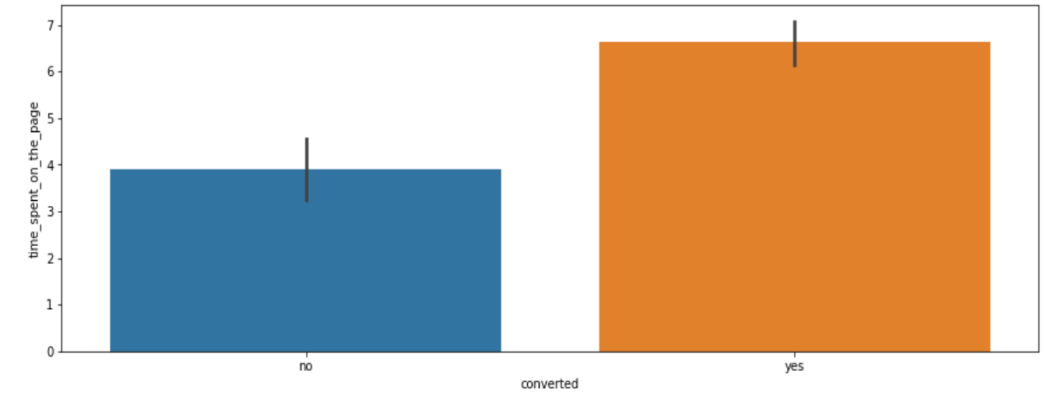
1.The mean time spent on the page by treatment group users is greater than control group users.

EDA - Time_spent_on_the_page Vs landing_page, Time_spent_on_the_page Vs converted



Observations

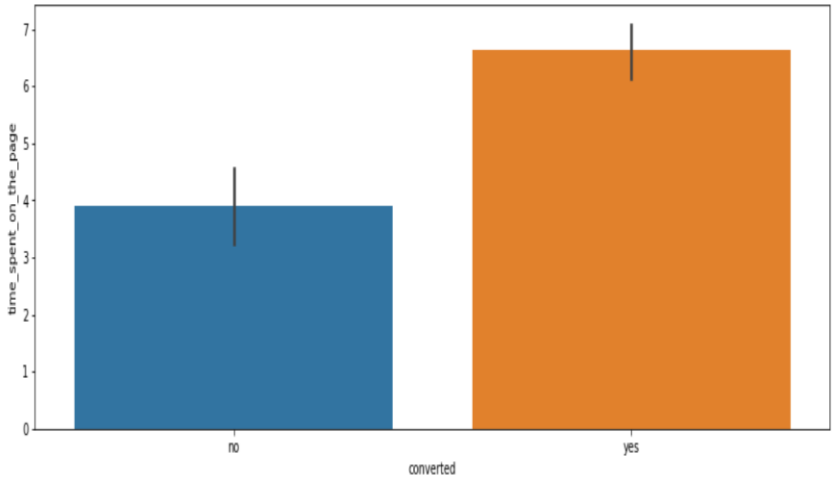
The mean time spent on the new page is greater than the mean time spent on the old page.

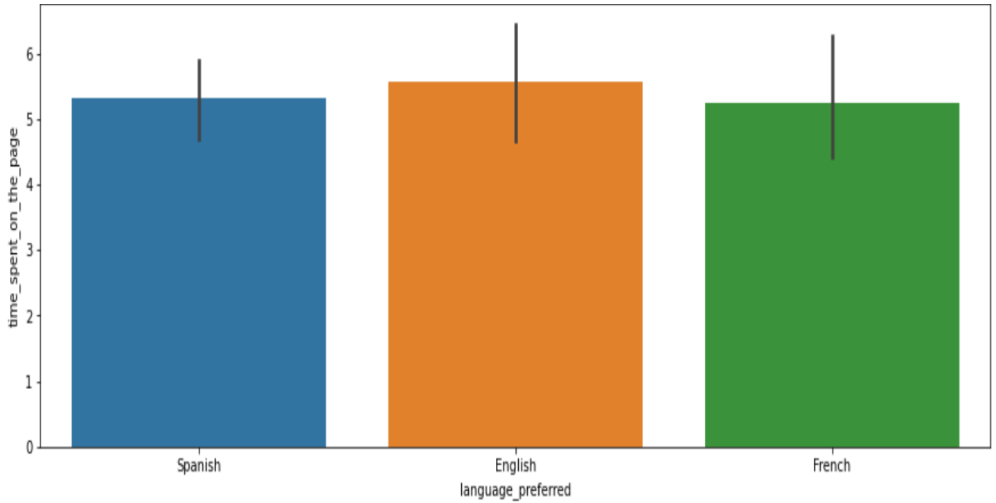


Observations

- The user who has subscribed for the news portal spends more time than the user who didnot subscribe.

EDA - Time_spent_on_the_page Vs converted, Time_spent_on_the_page Vs language_preferred





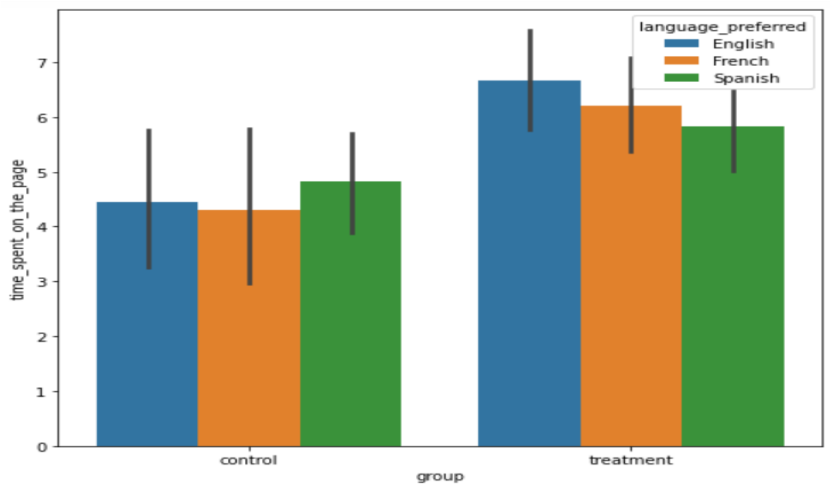
Observations

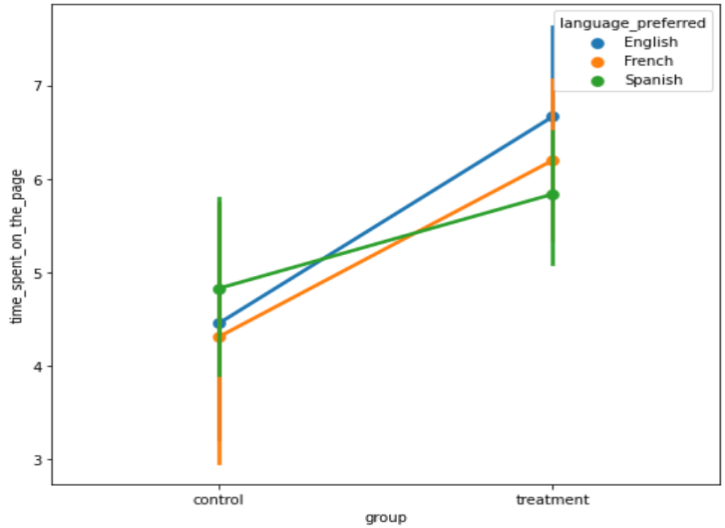
- The user who has subscribed for the news portal spends more time than the user who didnot subscribe.

Observations

1.The users who prefer English slightly spend more time than the users who prefer Spanish or French

Correlation between time_spent_on_the_page, group and language_preferred





- **Observations**

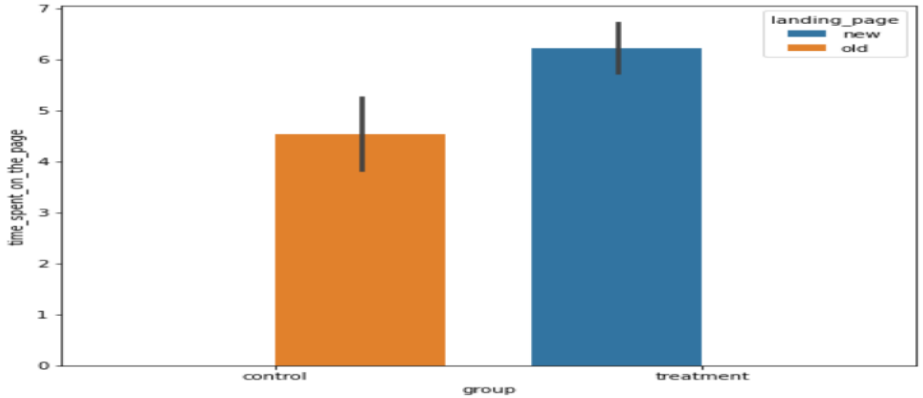
Among treatment group users, those who prefer English spend more time on the page and those who prefer Spanish spend lesser time.

Among control group users, those who prefer Spanish spend more time on the page and those who prefer English spend lesser time.

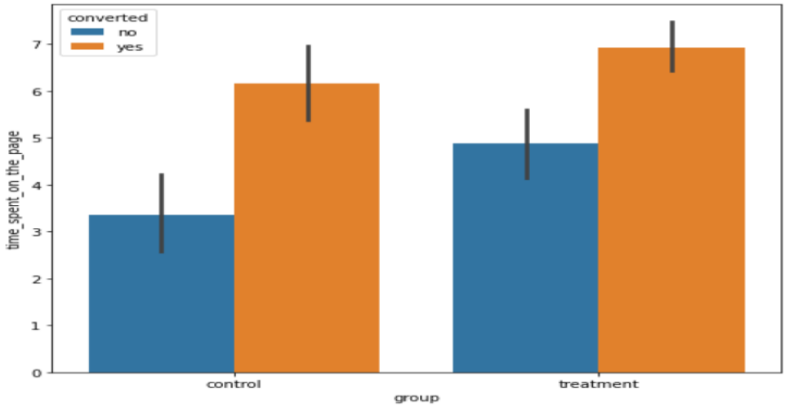
Correlation between time_spent_on_the_page, group and landing_page

- **Observations**

Treatment group spends more time on the page than control group as all the treatment group users prefer new landing page.



Correlation between time_spent_on_the_page, converted and group



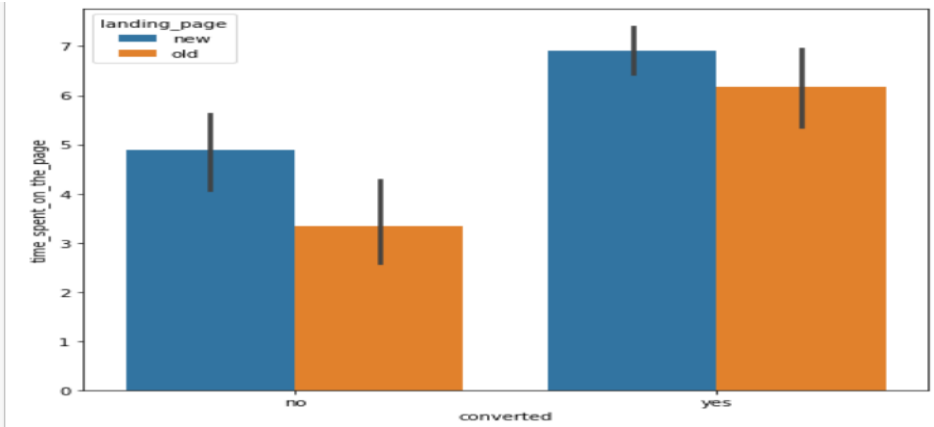
Observations

1.Among both control and treatment users, time spent on the page is more among the users who have subscribed

Correlation between time_spent_on_the_page, converted and landing_page

- ***Observations***

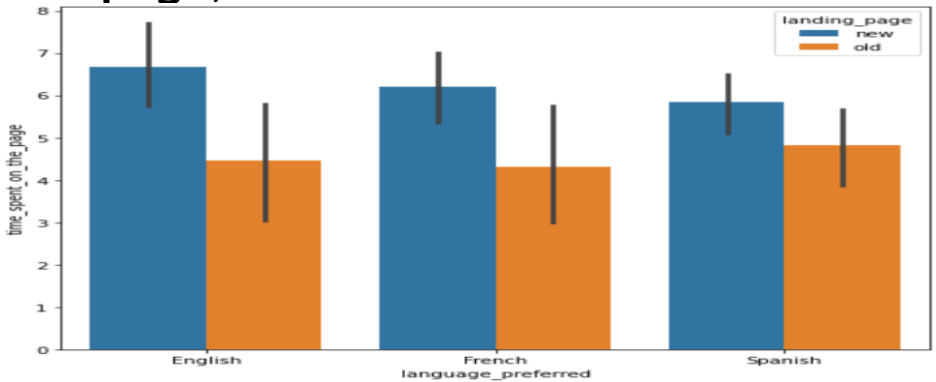
Among both subscribers and non-subscriber, most of the users preferred new_landing_page. The time spent on new landing page is more than the old landing page



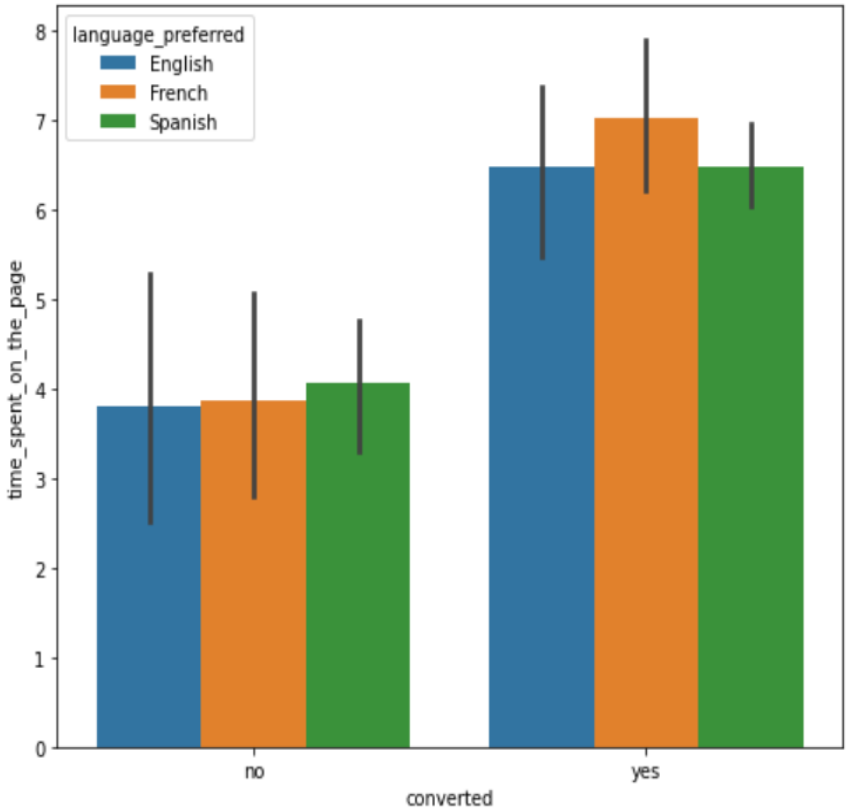
Correlation between time_spent_on_the_page, converted and language_preferred

Observation

1. time_spent_on_new_landing_page is more irrespective of the language preferred



Correlation between time_spent_on_the_page, converted and language_preferred



Observations

1.Subscribers spend more time than non-subscribers.

2. Among subscribers, users who prefer French spend more.

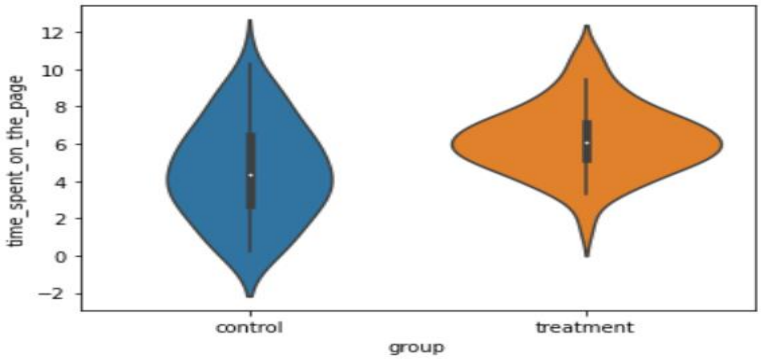
3. Among non - subscribers, users who prefer Spanish spend more time.

4.Among subscribers, users who prefer English and Spanish spend equal amont of time as there is only slight difference.

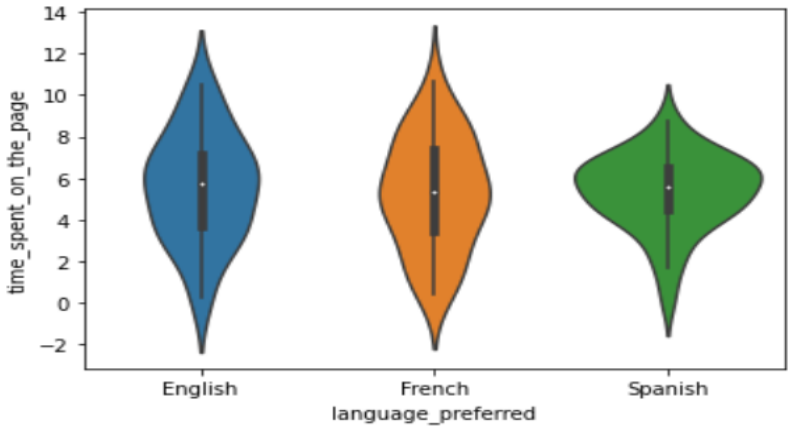
group Vs time_spent_on_the_page

- **Observation**

The time spent on page by the control group varies from -2 to 12 showing more variation than the treatment group.



language_preferred vs time_spent_on_the_page



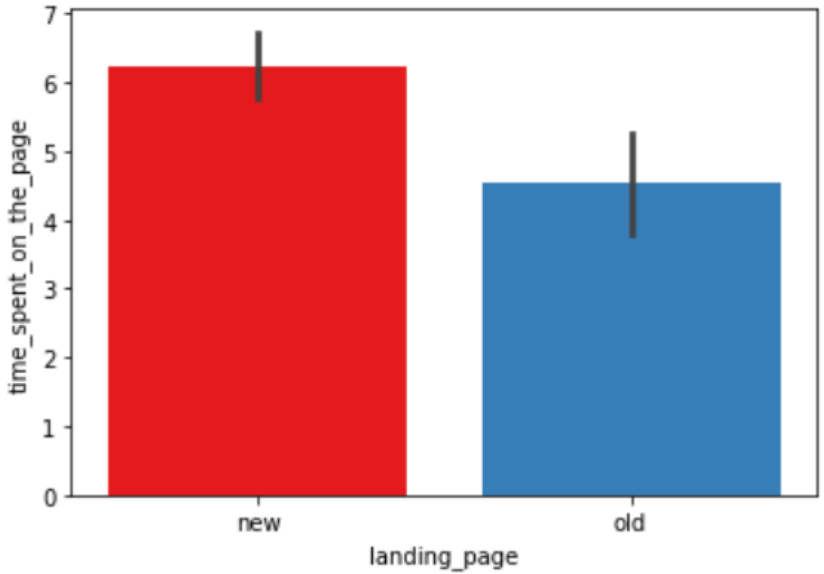
Observations

Users who prefer English and French spend more time on the Page than users who prefer Spanish

Hypothesis Tests

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

Visual Analysis



Observation

1.Time spent on the page on new landing page is more than the old landing page.

2.The mean time spent on the new page is 6.1 minutes and the mean time spent on the old page is 4.5 minutes.

Hypothesis test (contd)

Significance of the test	Assumptions	Test Distribution
Test for the equality of two population means $H0: \mu1=\mu2$	1. Continuuous data - time_spent_on_the_page 2. Normally distributed population or sample size > 30 3. Independent populations 4. Known population standard deviations σ 1 and σ 2 5. Random Sampling from the population	Standard Normal Distribution **Two independent sample z-test**

We will test the null hypothesis

$$H_0 : \mu_1 = \mu_2$$

against the alternate hypothesis

$$H_a : \mu_1 > \mu_2$$

Insight

As the p_value (7.6896e-05) which is much less than the level of significance 0.05. We reject the null hypothesis. Thus, from the statistical evidence we can say that time spent on new landing page is more than time spent on old landing page.

Hypothesis test

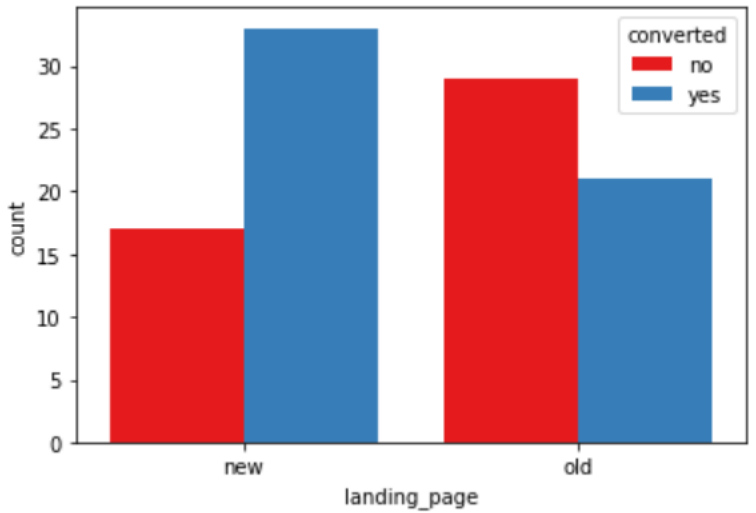
- Is the conversion rate for the new page greater than the conversion rate for the old page.

Visual Analysis

Observation

1.Among the users who prefer new landing page, there are more users who converted to subscriber.

2.Among the users who prefer old landing page, there are less subscribers.



Hypothesis test (contd)

Significance of the test	Assumptions	Test Distribution
Test for the equality of two population means $H0$Test for the equality of proportions $H_0:p1 = p2:\mu1=\mu2$	Binomially distributed and Independent populations Random sampling from the populations. When both mea(np) and n(1-p) are greater than or equal to 10, binomial distribution can be approximated by a normal distribution	Test Statistic Distribution **Two proportions z-test**

Let p_1, p_2 be the conversion rates of new landing and old landing pages

The manufacturer will test the null hypothesis

$$H_0 : p_1 = p_2$$

against the alternate hypothesis

$$H_a : p_1 > p_2$$

Hypothesis test (contd.)

$$np_1 = 50 \cdot \frac{33}{50} = 33 \geq 10$$

$$n(1 - p_1) = 50 \cdot \frac{50-33}{50} = 17 \geq 10$$

$$np_2 = 50 \cdot \frac{21}{50} = 21 \geq 10$$

$$n(1 - p_2) = 50 \cdot \frac{50-21}{50} = 29 \geq 10$$

- **Insight**

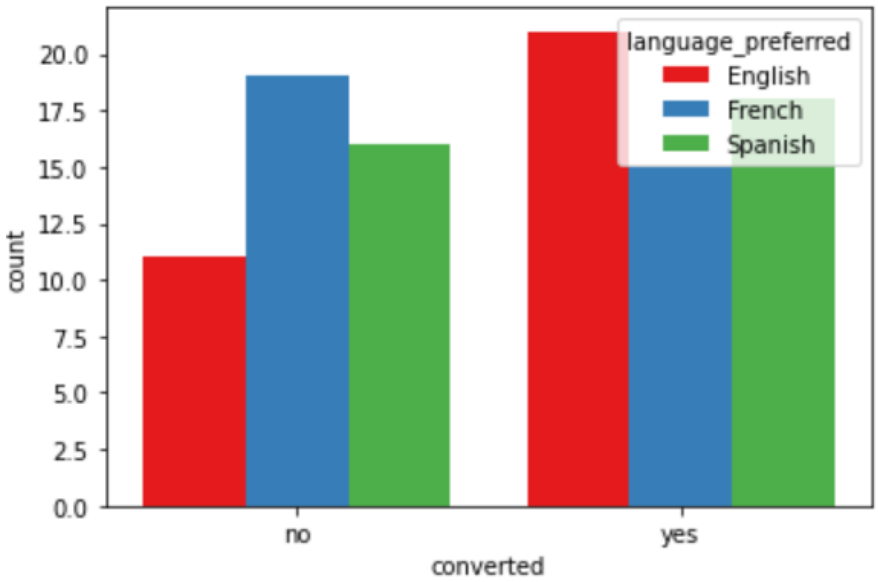
p_value = 0.016

As the p_value is less than the level of significance 0.05, we reject the null hypothesis. Thus,from the statistical analysis we can say conversion rate for the new page greater than the conversion rate for the old page.

Hypothesis test

- Does the converted status depend on the preferred language?

Visual Analysis



Observation

1.There are more subscribers who prefer to use English language.
2.There are more non-subscribers who prefer French.

Contingency Table

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

Hypothesis test(contd.)

Significance of the test	Assumptions	Test Distribution
In the contingency table, H0: The row and column variables are independent.	Categorical Variables. Expected value of the number of observations in each level of the variable is at least 5. Random sampling from the population	Test Statistic Distribution **Chi-square test of Independence.**

We will test the null hypothesis.

> $H_0: Converted status is independent preferred language

against the alternative hypothesis

> $H_a: Converted status depends on preferred language

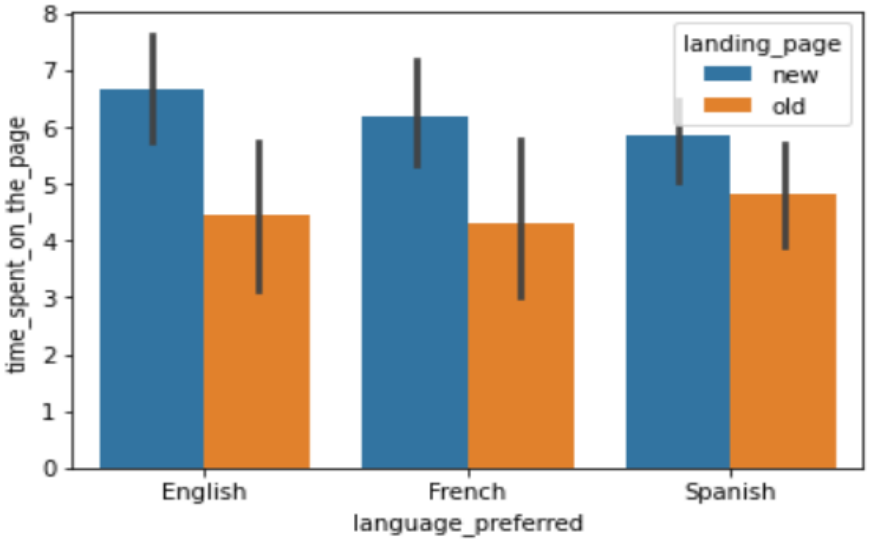
Insight

p_value = 0.2129
P value is greater than the level of significance 0.05. We fail to reject the null hypothesis. Hence, we have enough statistical significance to conclude that Converted status is independent of the language_preferred at 5% significance level.

Hypothesis Test

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

Visual Analysis



Observation

1.Among the users who prefer English, French and Spanish, the users prefer to use new_landing_page.

2.The time spent on the new landing page is more than old landing page irrespective of the language preferred.

Hypothesis test(contd.)

Significance of the test	Assumptions	Test Distribution
$H_0: Mean time spent on the new page is same for different language users. against the alternative hypothesis $H_a: Mean time spent on the new page is different for different language users.	The population are normally distributed. Samples are independent simple random samples.	Test Statistic Distribution **One-way ANOVA F-test**

Hypothesis test(Contd.)

Shapiro- Wilk's test

We will test the null hypothesis

> $H_0: time_spent_on_the_page follow normal distribution

against the alternative hypothesis

> $H_a: time_spent_on_the_page doesnot follow normal distribution

p_value = 0.80
Since, p_valus is larger than 0.05 we fail to reject the null hypothesis follow normal distribution.

Levene's test

p_value = 0.467
Since, p_value is slightly larger than 0.05, we fail to reject the null hypothesis of homogeneity of variances

Finding p_value using f_oneway

Hypothesis test(Contd.)

- p_value = 0.432

```
 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
-------------------------------------------------
```

From the above pairwise comparison, looking at the reject column we can say that the mean_time_spent_on_the_new_page is not so different between (English and French), (English and Spanish) and (French and Spanish)

Business Insights and Recommendations

- Use the new landing page as it is making users spend more time on the portal than the old landing page.
  Also, it has been found that more time spent is directly related to the more subscriptions.

- Among the different language users, Spanish users spent time the least indicating the new landing page is counter effective in spanish. Hence considering revising the landing page for non-english users to improve subscriptions even more.

greatlearning
Power Ahead

Happy Learning !