

# E-NEWS EXPRESS BUSINESS STUDY

- By Arsalaan Saiyed

# BACKGROUND



E-News is an online news portal and it aims to expand its business by acquiring new subscribers. Visitors on their website take certain actions based on their interests.

The company plans to analyze these interests and wants to determine whether a new feature will be effective or not.

The design team of the company has created a new landing page. We need to determine whether the new landing page is more effective to gather new subscribers than the existing landing page.

100 users were selected at random and divided equally into two groups. The old landing page is served to the first group (control group) and the new landing page is served to the second group (treatment group).

## Sample Data:

	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

# OBJECTIVE

Companies often analyze user's 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 that is used to determine whether a new feature attracts users based on a chosen metric. The objective of this document is to display the extracted insights from the data and conduct statistical analysis on the business data.

## Analysis Methodology:

- Understanding the raw data
- Univariate Exploratory analysis
- Multivariate Exploratory analysis
- Statistical Analysis to answer some key questions



Following the analysis we showcase key findings, insights and have recommended some improvements that can drive business decision that help boost revenue and enhance customer experiences.

# DATA INFORMATION

## *Observation:*

Upon studying the raw data we observed the following:

- There are 100 unique users.
- There are 2 unique groups - control and treatment. Each group consists of 50 users.
- There are 2 landing pages - new and old.
- Overall, 54 users get converted and 46 users do not get converted after visiting the landing page.
- There are 3 unique preferred languages - English, French, and Spanish.

## *Note:*

- *There is no missing data.*
- *Refer to the univariate analysis on the next page that gives a graphical representation of the above observations.*

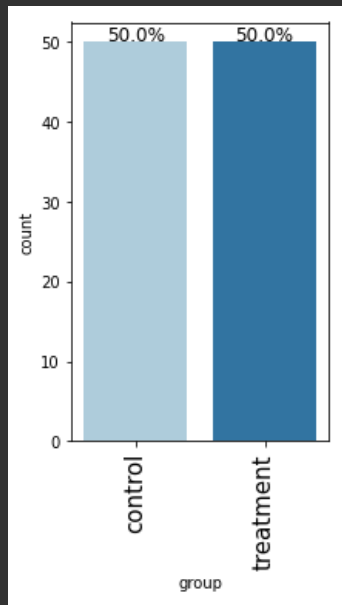
Observations	Variables
100	6

Variables	Description
user_id	User ID of the person visiting the website
group	Represents whether the user belongs to the first group (control) or the second group (treatment).
landing_page	Represents whether the landing page is new or old
time_spent_on_the_page	Time (in minutes) spent by the user on the landing page
converted	Represents whether the user gets converted to a subscriber of the news portal or not

# DATA INFORMATION

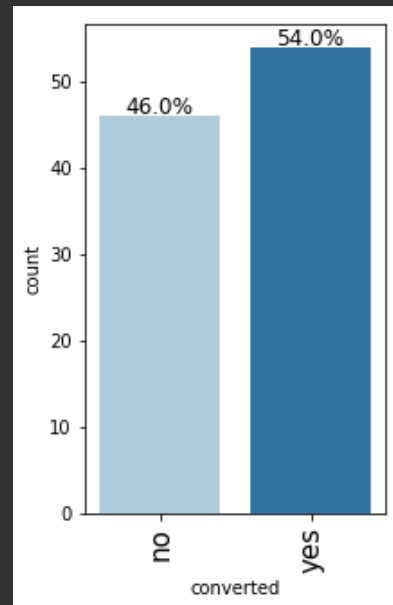
## UNIVARIATE ANALYSIS – CATEGORICAL DATA

Group



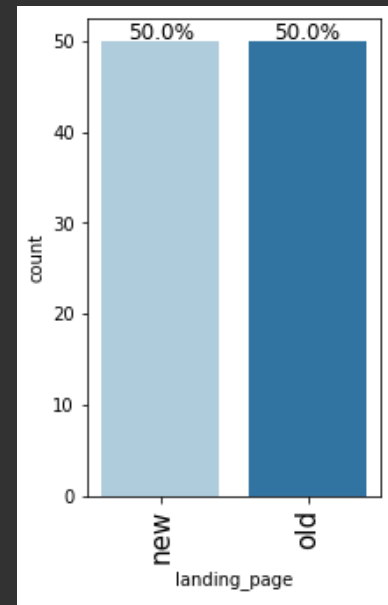
Control Group – 50  
Treatment Group – 50

Converted



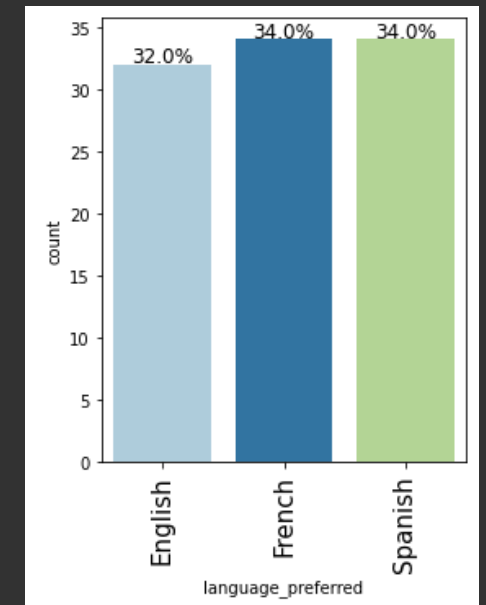
Converted Visitors – 54  
Not Converted Visitors – 46

Landing Page



New Page – 50  
Old Page – 50

Language Preference



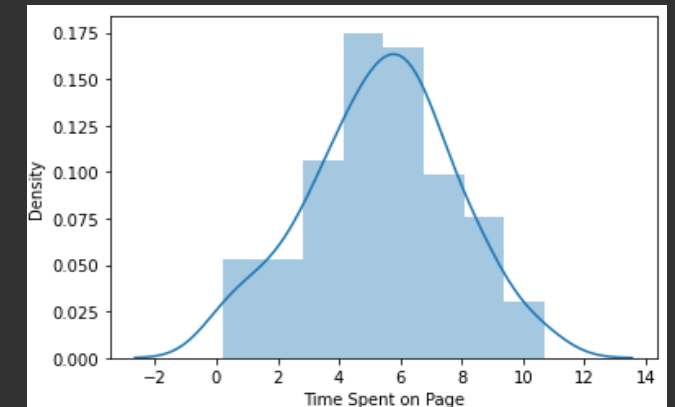
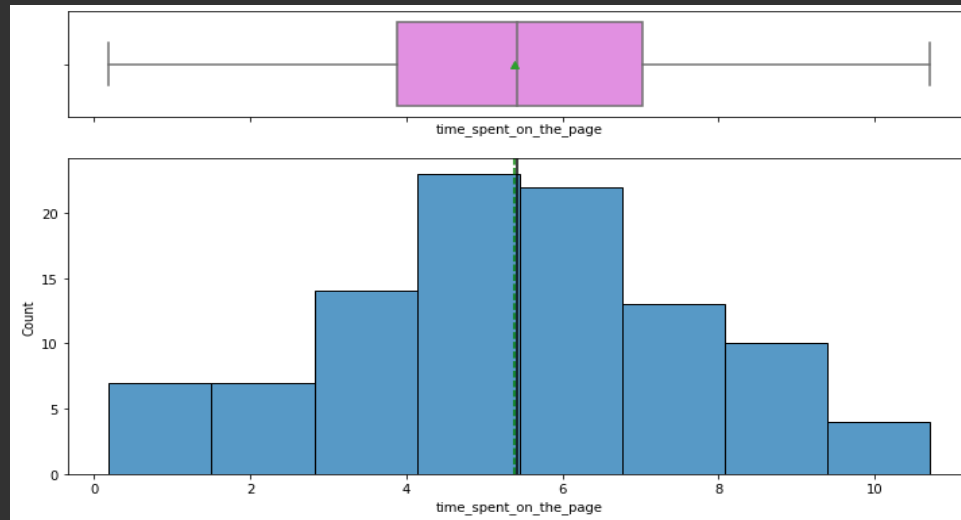
English Preferred Visitors – 32  
French Preferred Visitors – 34  
Spanish Preferred Visitors – 34

# EXPLORATORY DATA ANALYSIS – TIME SPENT ON PAGE UNIVARIATE ANALYSIS – NUMERIC DATA

## *Observation:*

Below is the 5 point statistical summary of our numeric value the time spent on the landing page

Stat	Time Spent on Page
Count	100
Mean	5.38
Std.dev	2.38
Min	0.19
25%	3.88
50%	5.42
75%	7.02
Max	10.71



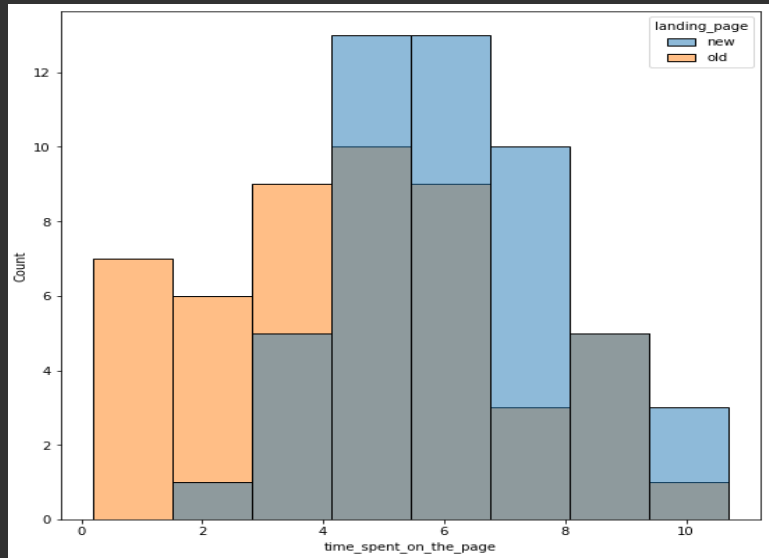
## **Note:**

- Using density plot we can observe that this variable follows a normal distribution.
- There is a negligible left skewness to variable.

# EXPLORATORY DATA ANALYSIS

## TIME SPENT ON PAGE - BIVARIATE ANALYSIS – NUMERIC & CATEGORICAL DATA

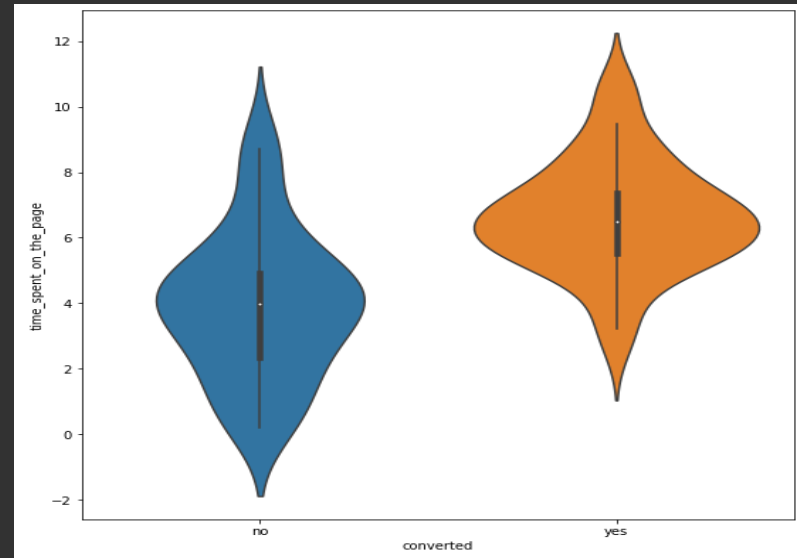
Landing Page



### Observation:

- Mean time spent on new landing page is 6.2 minutes.
- Mean time spent on old landing page is 4.5 minutes

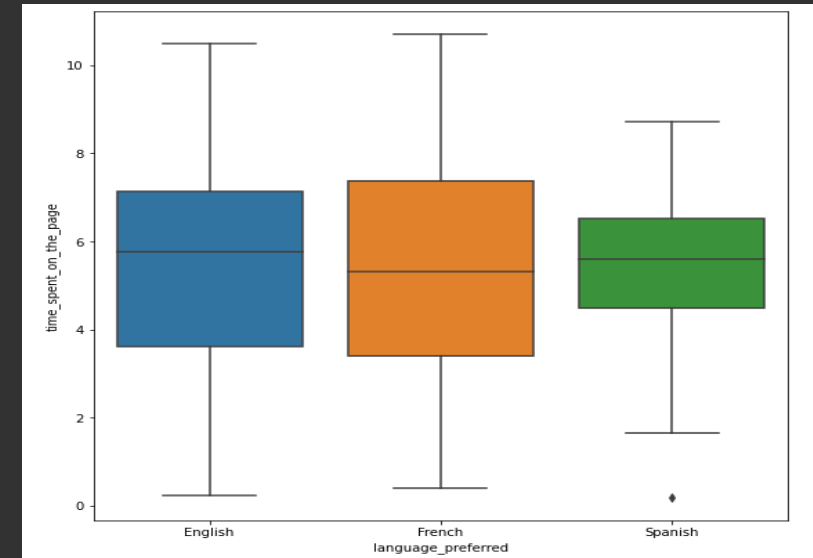
Converted



### Observation:

- Mean time spent when a visitor was not converted was  $\approx 4$  minutes
- Mean time spent when a visitor was converted was  $\approx 7$  minutes

Language Preference



### Observation:

- Mean time spent by visitors of all 3 language preferences was  $\approx 5$  minutes.

# STATISTICAL ANALYSIS – QUESTION 1

## DO THE USERS SPEND MORE TIME ON THE NEW LANDING PAGE THAN THE EXISTING LANDING PAGE?

### Method:

Significance of the test	Assumptions	Test Statistic Distribution
Let $\mu_1, \mu_2$ be the mean time spent on landing page new and existing respectively. We will test the null hypothesis $H_0: \mu_1 = \mu_2$ against the alternate hypothesis $H_a: \mu_1 > \mu_2$	<ul style="list-style-type: none"><li>Continuous data</li><li>Normally distributed populations</li><li>Independent populations</li><li>Equal population standard deviations</li><li>Random sampling from the population</li></ul>	t-distribution  (The test is one-tailed and is also known as <b>Two sample independent t-test</b> )

### Significance Level:

As given in the problem statement, we select  $\alpha = 0.05$ .

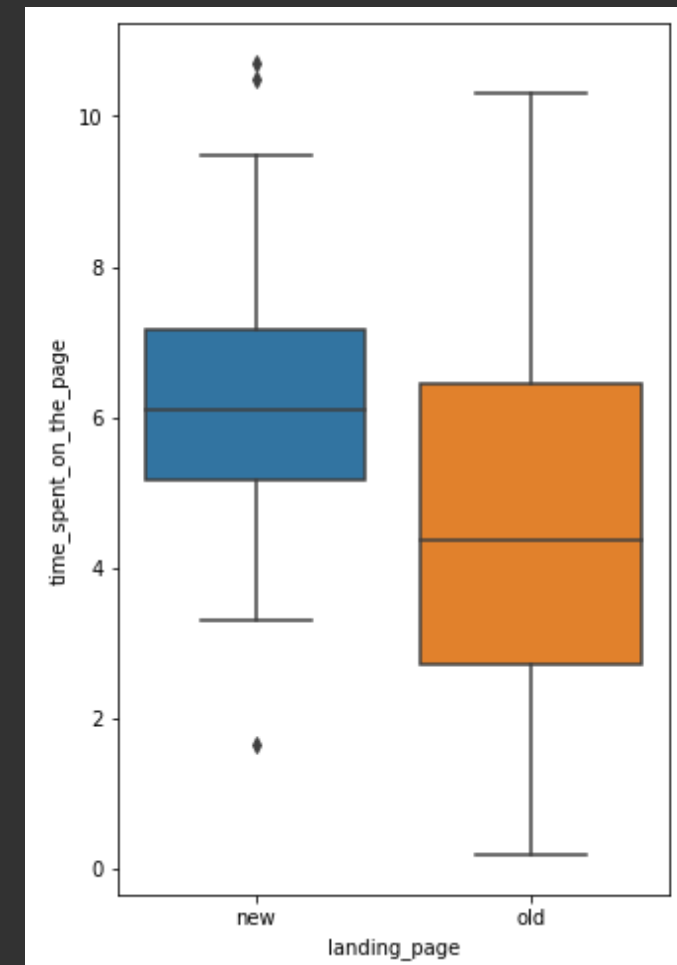
### P-value:

$p = 0.00013$

### Inference:

As the p-value 0.00013 is less than the level of significance, we reject the null hypothesis.

Therefore, we have strong statistical evidence to say, **YES** there is greater significant difference on the time spent on the new landing page compared to that of the old landing page at a 5% significance level.





# STATISTICAL ANALYSIS – QUESTION 2

## IS THE CONVERSION RATE FOR THE NEW PAGE GREATER THAN THE CONVERSION RATE FOR THE OLD PAGE?

### Method:

Significance of the test	Assumptions	Test Statistic Distribution
Let $p_1, p_2$ be the proportions of conversions for treatment group and control group respectively. We will test the null hypothesis  $H_0: p_1 = p_2$  against the alternate hypothesis $H_a: p_1 > p_2$	<ul style="list-style-type: none"><li>Binomially distributed populations</li><li>Independent populations</li><li>Random sampling from the populations</li><li>When both mean (<math>np</math>) and <math>n(1-p)</math> are greater than or equal to 10, the binomial distribution can be approximated by a normal distribution</li></ul>	Standard Normal Distribution  (The test is one-tailed and is also known as <b>Two proportions z-test</b> )

### Significance Level:

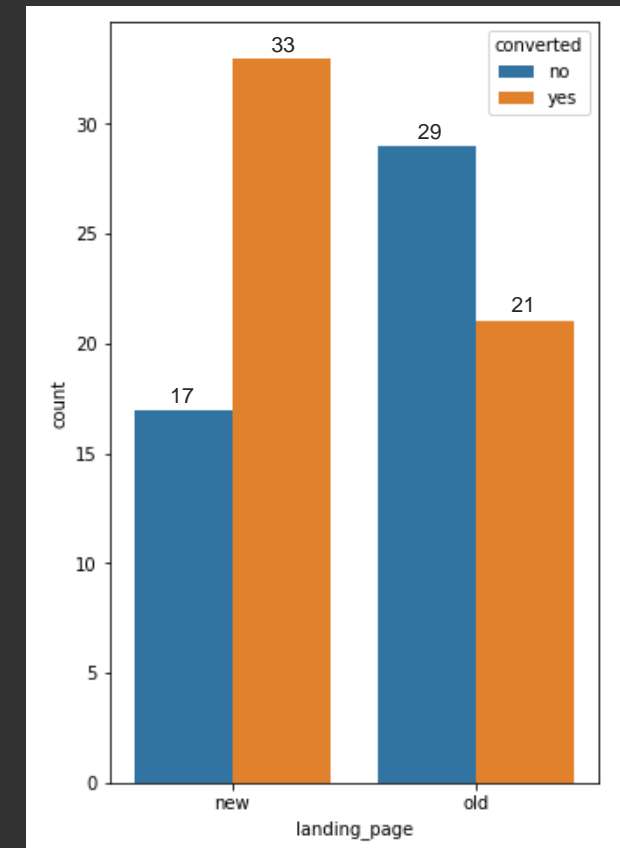
As given in the problem statement, we select  $\alpha = 0.05$ .

### P-value:

$p = 0.008$

### Inference:

As the p-value 0.008 is less than the level of significance, we reject the null hypothesis. Therefore, we have strong statistical evidence to say, **YES** there is greater significant difference in the conversion rate of the new landing page compared to that of the old landing page at a 5% significance level.



# STATISTICAL ANALYSIS – QUESTION 3

## DOES THE CONVERTED STATUS DEPEND ON THE PREFERRED LANGUAGE?

### Method:

Significance of the test	Assumptions	Test Statistic Distribution
We will test the null hypothesis <b><math>H_0</math>: Conversion is independent of language preference.</b>  against the alternate hypothesis <b><math>H_a</math>: Conversion depends on language preference.</b>	<ul style="list-style-type: none"><li>• Categorical variables</li><li>• Expected value of the number of sample observations in each level of the variable is at least 5</li><li>• Random sampling from the population</li></ul>	Chi Square distribution  (The test is also known as <b>Chi-square test of independence</b> )

### Significance Level:

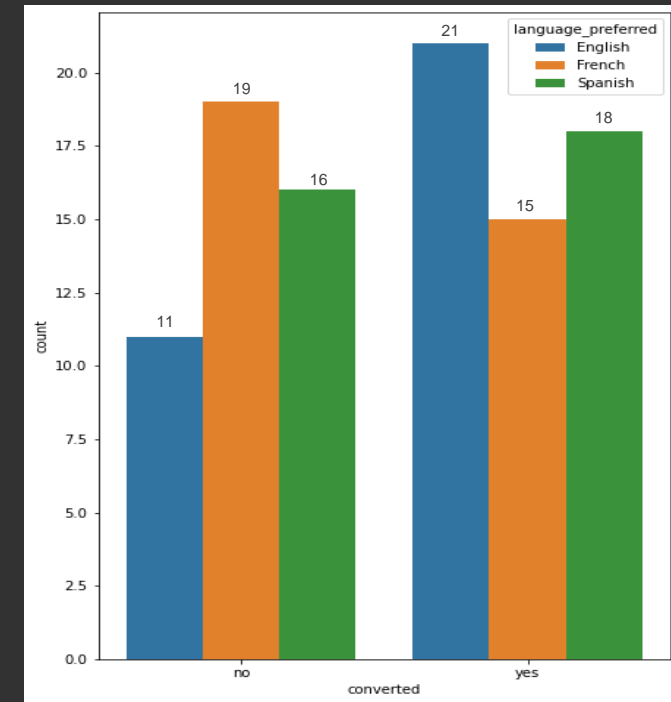
As given in the problem statement, we select  $\alpha = 0.05$ .

### P-value:

$p = 0.21298$

### Inference:

As the p-value 0.21298 is greater than the level of significance, we fail to reject the null hypothesis. Therefore, we have strong statistical evidence to say, **YES** the conversion rate is independent of the language preferences at a 5% significance level.



Contingency Table

Language Preference	Converted	
	No	Yes
English	11	21
French	19	15
Spanish	16	18

# STATISTICAL ANALYSIS – QUESTION 4

## IS THE TIME SPENT ON THE NEW PAGE SAME FOR THE DIFFERENT LANGUAGE USERS?

### Method:

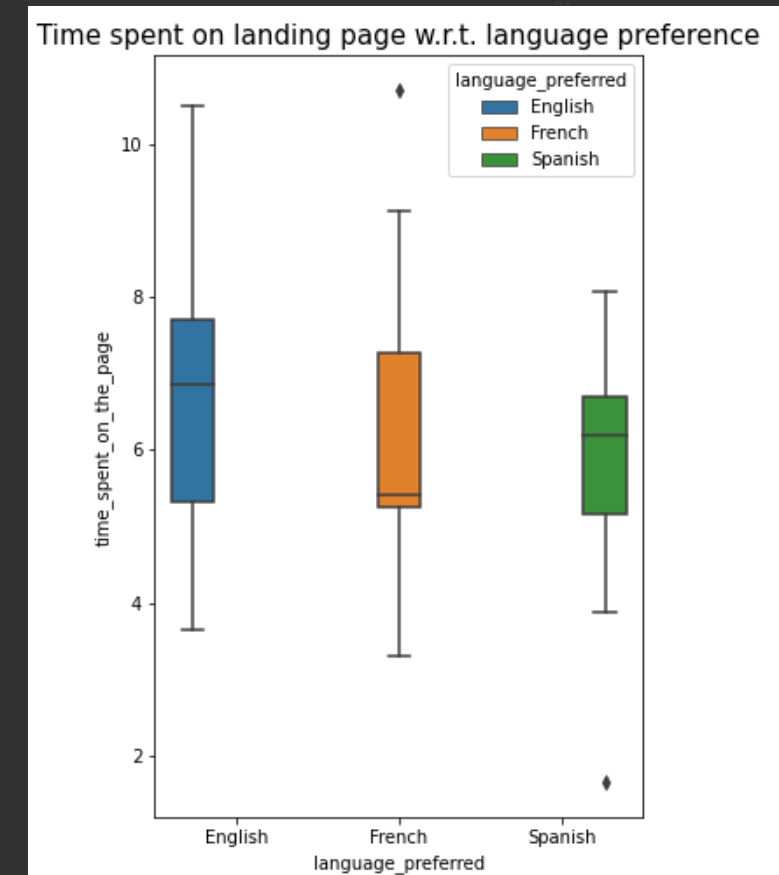
Significance of the test	Criteria/Assumptions	Test Statistic Distribution
<p>Let <math>\mu_1, \mu_2, \mu_3</math> be the means of time spent on the new landing page for languages English, French &amp; Spanish respectively.</p> <p>We will test the null hypothesis</p> <p><b><math>H_0: \mu_1 = \mu_2 = \mu_3</math></b></p> <p>against the alternative hypothesis</p> <p><b><math>H_a</math>: At least one of the means is different from the rest.</b></p>	<ul style="list-style-type: none"><li>The populations are normally distributed</li><li>Samples are independent simple random samples</li><li>Population variances are equal</li></ul>	<p>t-distribution</p> <p>(The test is one-tailed and is also known as <b>Two sample independent t-test</b>)</p>

### Significance Level:

As given in the problem statement, we select  $\alpha = 0.05$ .

We need to check the assumptions of normality and equality of variance for the three groups:

- For testing of normality, **Shapiro-Wilk's test** is applied to the response variable.
- For equality of variance, **Levene's test** is applied to the response variable.



# STATISTICAL ANALYSIS – QUESTION 4

## IS THE TIME SPENT ON THE NEW PAGE SAME FOR THE DIFFERENT LANGUAGE USERS?

### Shapiro-Wilk's test:

We will test the null hypothesis

**$H_0$ : Time spent on the new page follows a normal distribution**

against the alternative hypothesis

**$H_a$ : Time spent on the new page does not follow a normal distribution**

### Significance Level:

As given in the problem statement, we select  $\alpha = 0.05$ .

### P-value:

$p = 0.497$

### Inference:

As the p-value 0.497 is greater than the level of significance, we fail to reject the null hypothesis that the response follows the normal distribution.

### Levene's test:

We will test the null hypothesis

**$H_0$  : All the population variances are equal**

against the alternative hypothesis

**$H_a$  : At least one variance is different from the rest**

### Significance Level:

As given in the problem statement, we select  $\alpha = 0.05$ .

### P-value:

$p = 0.194$

### Inference:

As the p-value 0.194 is greater than the level of significance, we fail to reject the null hypothesis of homogeneity of variances.

# STATISTICAL ANALYSIS – QUESTION 4

## IS THE TIME SPENT ON THE NEW PAGE SAME FOR THE DIFFERENT LANGUAGE USERS?

Let's test whether the assumptions are satisfied or not:

- The populations are normally distributed - Yes, the normality assumption is verified using the Shapiro-Wilk's test.
- Samples are independent simple random samples - Yes, we are informed that the collected sample is a simple random sample.
- Population variances are equal - Yes, the homogeneity of variance assumption is verified using the Levene's test.

Since all the assumptions are satisfied we carry out the ANOVA f-test to find the p-value

P-value:

$p = 0.7325$

Inference:

As the p-value 0.7325 is greater than the level of significance, we fail to reject the null hypothesis that all the means of the time spent on the new landing page for all three languages is the same. Therefore, we have enough statistical evidence to say, *time spent on the new landing page is same for all languages* at a 5% significance level.

Multiple Comparison test (Tukey HSD):

In order to identify for which language's conversion mean is different from other languages, the null hypothesis is

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

against the alternative hypothesis

$H_a: \mu_1 \neq \mu_2$  or  $\mu_1 \neq \mu_3$  or  $\mu_2 \neq \mu_3$

Significance Level:

We select  $\alpha = 0.05$ .

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
English	French	-0.3058	0.8482	-1.7123	1.1006	FALSE
English	Spanish	-0.2273	0.9	-1.6338	1.1792	FALSE
French	Spanish	0.0785	0.9	-1.3065	1.4635	FALSE

Inference:

As the p-values (refer to the p-adj column) for comparing the mean time spent for all the possible pairs of languages is significantly more than the confidence level, the null hypothesis of equality of all population means can not be rejected.

*Thus, we can say that the mean time spent on the new landing page for all the three languages are not significantly different from one another.*

# STATISTICAL ANALYSIS – QUESTION 5

## DOES MORE TIME SPENT ON PAGE MEAN A HIGHER CHANCE OF CONVERSION?

### Method:

Significance of the test	Criteria/Assumptions	Test Statistic Distribution
Let $\mu_1, \mu_2$ be the mean time spent on converted visitors and non converted visitors  We will test the null hypothesis <b><math>H_0: \mu_1 = \mu_2</math></b>  against the alternate hypothesis <b><math>H_a: \mu_1 &gt; \mu_2</math></b>	<ul style="list-style-type: none"><li>Continuous data</li><li>Normally distributed populations</li><li>Independent populations</li><li>Equal population standard deviations</li><li>Random sampling from the population</li></ul>	t-distribution  One-tailed  <b>Two sample independent t-test</b>

### Significance Level:

As given in the problem statement, we select  $\alpha = 0.05$ .

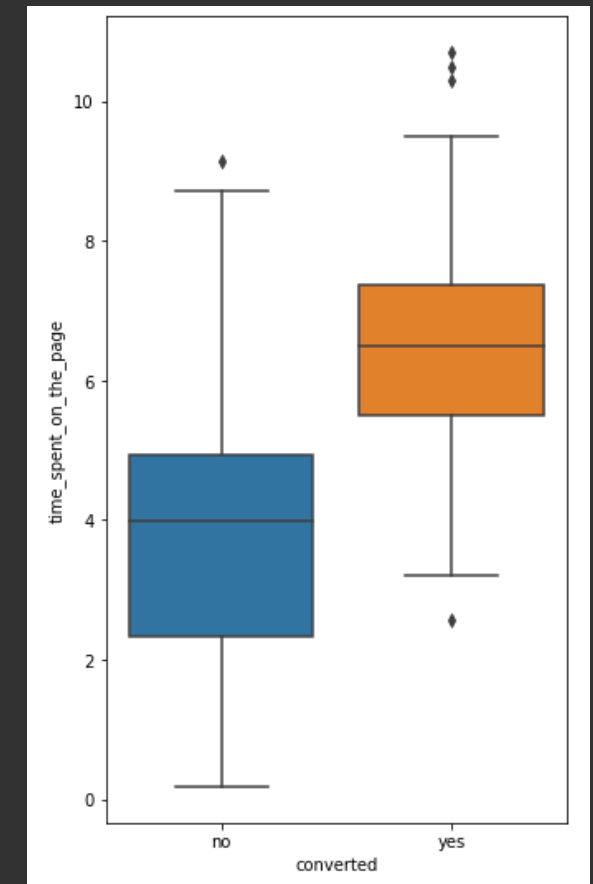
### P-value:

$p = 2.953e-10$

### Inference:

As the p-value  $2.953e-10$  is less than the level of significance, we reject the null hypothesis.

Therefore, we have strong statistical evidence to say, **YES** there is greater significant difference on the time spent by visitors that are converted compared to those which are not converted at a 5% significance level.



# CONCLUSION

Through exploratory data analysis and statistical analysis we can conclude the following:

- In the random sample we observe an approximate difference of 2 minutes between the average time spent on the new landing page than that of the old one. This is a greater significant difference on the time spent on the new landing page compared to that of the old landing page at a 5% significance level.
- In the 50-50 sample for the new and old landing page the conversion is 33 and 21 subscribers earned respectively. This is a significant difference in the conversion rate of the new landing page compared to that of the old landing page at a 5% significance level.
- The conversion rate is independent of the language preferences at a 5% significance level.
- The mean time spent on the new landing page for all the three language preferences are not significantly different from one another.
- An additional test was carried out to test whether more time spent on landing page meant higher conversion rate. The test concluded that it was the case and there was a greater significant difference on the time spent by visitors that are converted compared to those which are not converted at a 5% significance level.

# RECOMMENDATIONS

## **Recommendation 1:**

Switch to the new landing page as it yields significantly higher conversions and as seen through our statistical analysis a significantly higher mean time spent on it. We can say with much confidence that deploying the new landing page will bring in more subscribers for Enews and consequently drive their business up.

## **Recommendation 2:**

As seen more time spent on page suggests a higher chance on conversion. Hence, implement changes that increase time spent on the landing page, such as:

- Offer more captivating and interactive content
- Provide suggestions to keep people reading
- Incorporate videos into Content Strategy
- Improve website design and readability of content makes a huge difference on keeping your audience engaged
- Increase engagement with the visitors by asking them to sign up, entice them with a news-letter discounts when you sign-up and ask for feedback.
- Use exit-intent pop-ups.

## **Recommendation 3:**

Enews also needs to bring in more visitors i.e. increase website traffic. They can do so in various ways like:

- Making the website mobile friendly
- Advertise - Paid search, social media advertising, and display advertising are all excellent ways of attracting visitors
- Write Irresistible Headlines, invite others guests to post articles on Enews and foster a sense of community.
- Make use of GSC (google search console) to improve Enews's SEO(search engine optimization).