E NEWS EXPRESS CASE STUDY

Business statistics project

sania

Background

- E-news Express, an online news portal, aims to expand its business by acquiring new subscribers.
- The portal displays the contents in three different languages i.e., English, French and Spanish.
- User needs to subscribe to this portal to get access to news regularly.
- The design team of the company has created a new landing page with additional features in order to grab attentions of more new users.
- It is expected that, the new landing page will convert more users to new subscribers.

EXECUTIVE SUMMARY

- The design team of the company has researched and created a new landing page that has a new outline & more relevant content shown compared to the old page.

- The objective of this use case is to explore the data and perform Statistical Analysis (at a significance level of 5%) to determine the effectiveness of the new landing page in gathering new subscribers for the news portal by answering the following questions:

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

b) 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?

c) Does the converted status depend on the preferred language?

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

DATA OVERVIEW

- **The dataset has 100 records with 6 columns .**
- **There is an integer column(user_id) and a float column(time_spent_on_the_page)**
- **There are four categorical columns(group, landing_page, converted, language_preferred)**
- **There are no missing values and duplicate values in the dataset.**
- **We have 2 groups: Control & Treatment that have 50 users each**

VARIABLES	DESCRIPTION
user_id	Unique user id of the user visiting the website
group	Whether the user belongs to control group or treatment group
landing_page	Whether the landing page is new or old
Time_spent_on_the_page	Time spent by the user on the page
converted	Whether the user gets converted to news portal or not
Language_preferred	English,Spanish and French

Key Insights based on EDA

- The average time spent by visitor on the landing page is 5.37 minutes
- Minimum and maximum time spent by visitor on the landing page is 0.19 minutes and 10.71 minutes, respectively.
- 55 users converted to new subscriber and 45 users did not convert out of 100 sample users.
- The user's preferred language: 34/100 are Spanish or French, 32/100 are English.
- Total user count on both new and old landing page is 50.

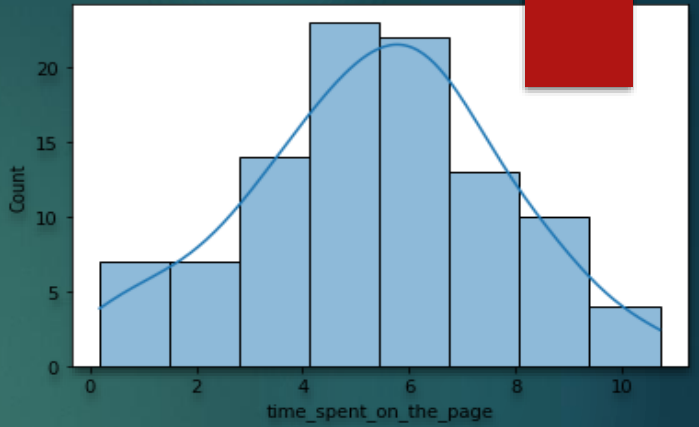
Key Insights based on EDA…..contd

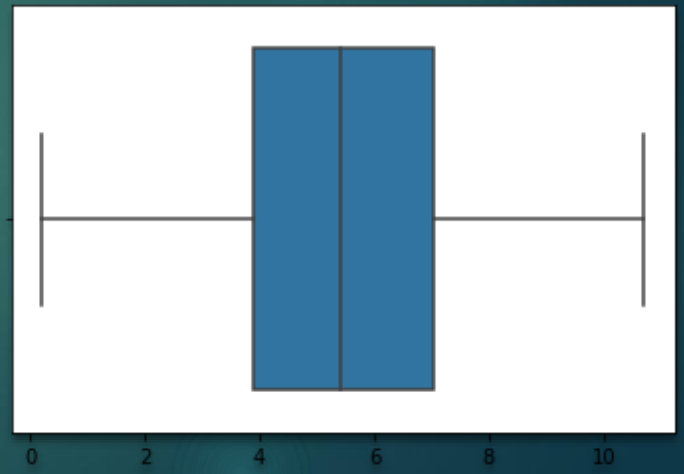
- The group that were shown the new landing page spent more time on the page than those who were shown the old page.
- Users that converted to subscribers seem to have spent a greater amount of time on the landing page as compared to those who did not convert.
- The time spent on the page seems to be very similar across the preferred languages.
- More English language readers converted to subscribers as compared to the other two languages, although the difference is negligible.

Exploratory Data Analysis

Univariate Analysis- time_spent_on_page

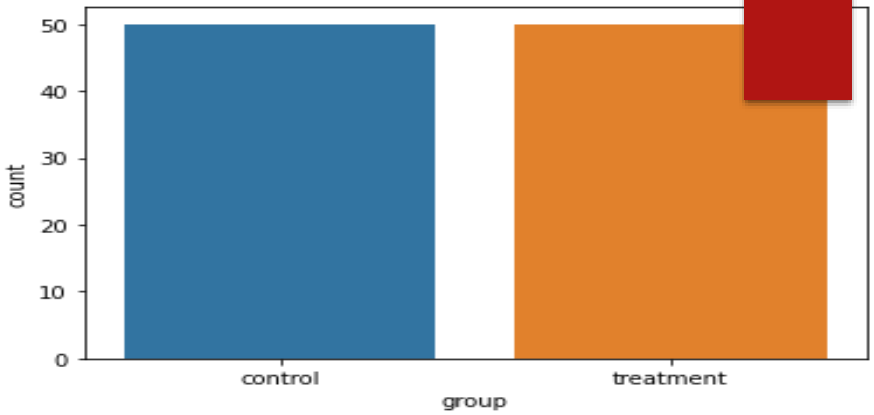
- The average time spent (5.37 minutes)is almost same as median (Q2 or 50%) time spent (5.41 minutes) on the page.
- The time spent by the user on the landing page is approximately normally distributed.
- There are no outliers in the time spent on the landing page by users.

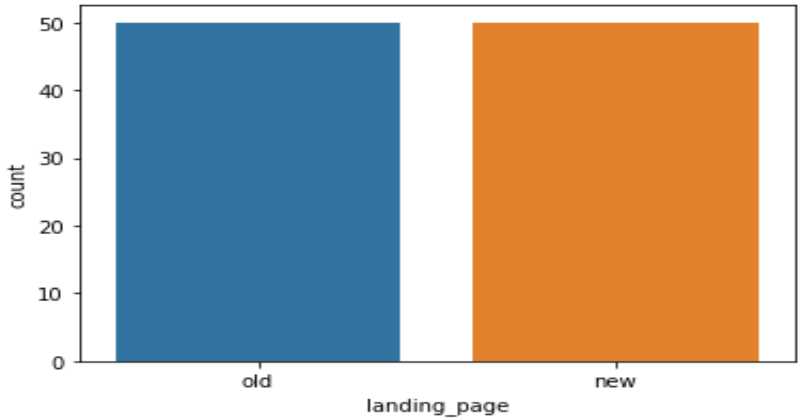




Univariate Analysis on categorical data

- The observations are distributed equally with 50 for each control group and treatment.
- The landing page has two values old and New .Total user count on both old and new is 50.

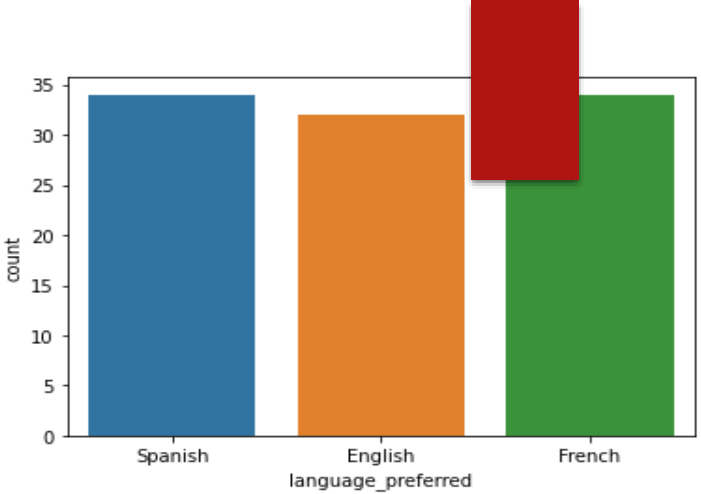


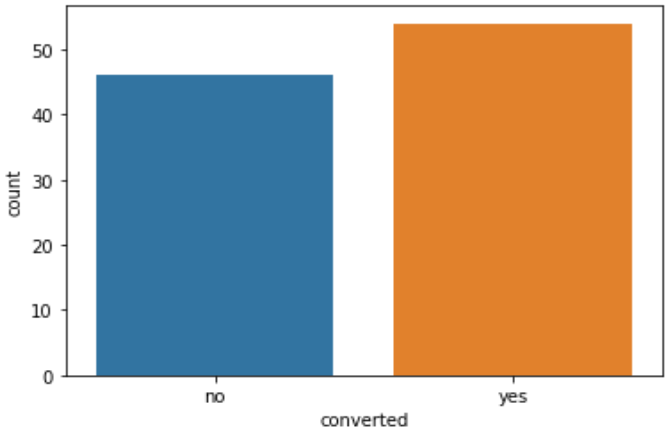


UVA on converted and language preffered

a. The column "language preferred" has 3 unique values spanish, english, and french, with both french and spanish having the highest frequencies of 34 and English 32.

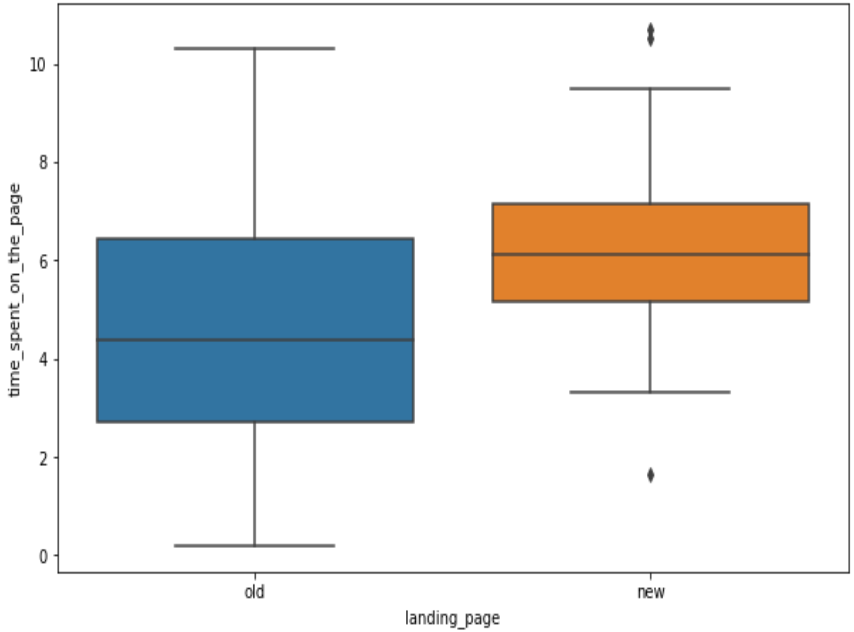
b. The feature "converted" has 2 unique values no and yes, with yes(coverted) having the highest frequency of 54 and no(converted) with 46.





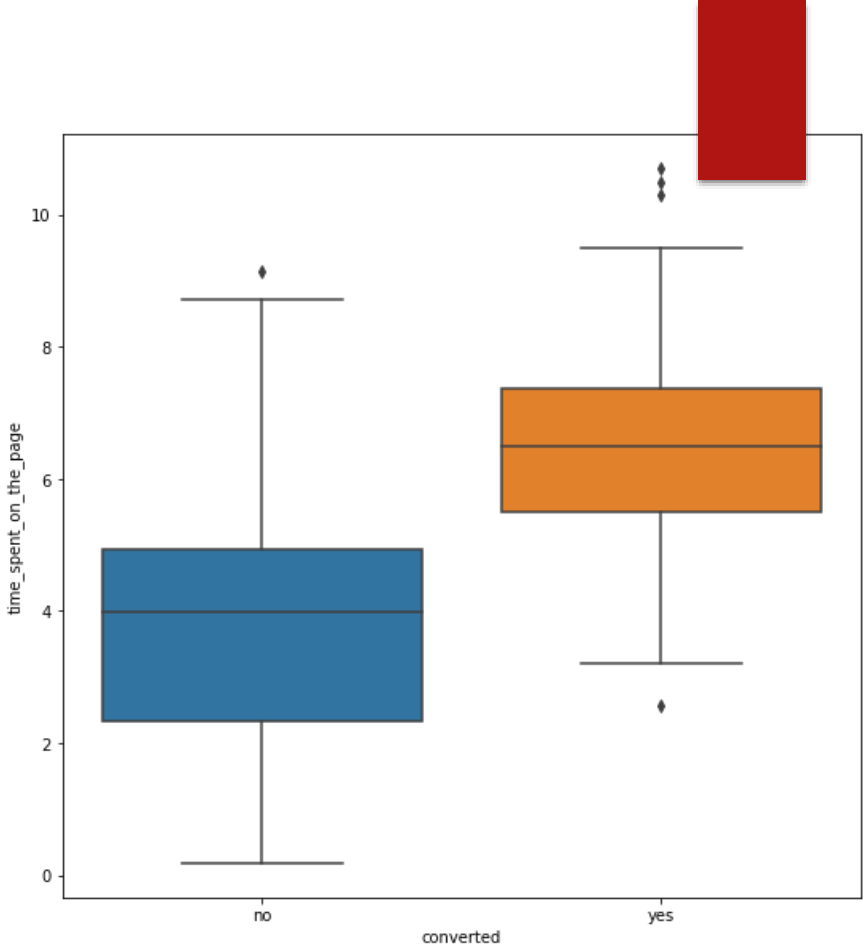
Bivariate Analysis

- On Average the user spends more time on new landing page i.e. 6.22 min than on old landing page (4.53 min).
- There are outliers that user spend less than 2 minutes and and more than 10 min on new landing page.



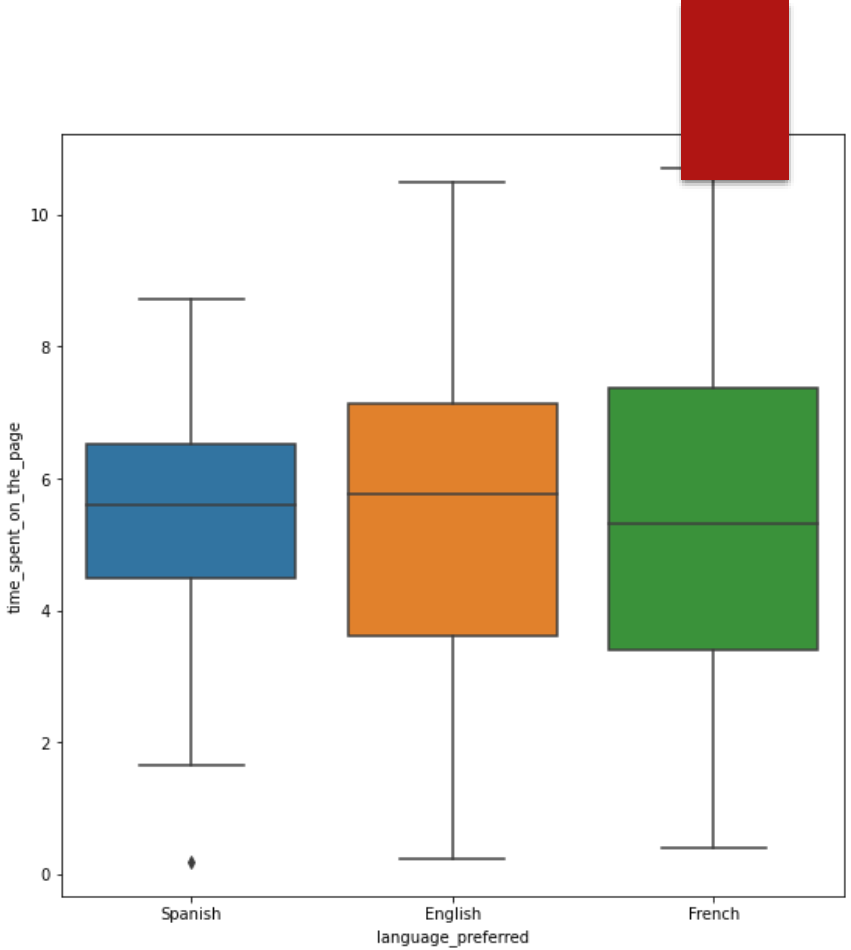
BVA on conversion status vs time spent

- The time spent by users in converted page is higher compared to those not converted.
- It is found that average time spent by users who are ultimately converted to new subscriber is 6.43 minutes and they spent at most 10.71 minutes.
- The users who were not converted spends around 2.23 to 5.4 minutes

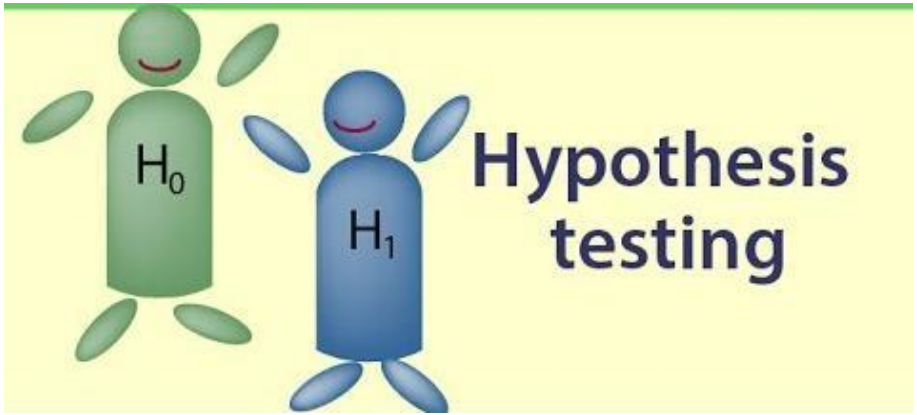


BVA on language_preferred vs time spent

The average time spent by each of English, French and Spanish are almost equal. This needs to be validated by statistically.

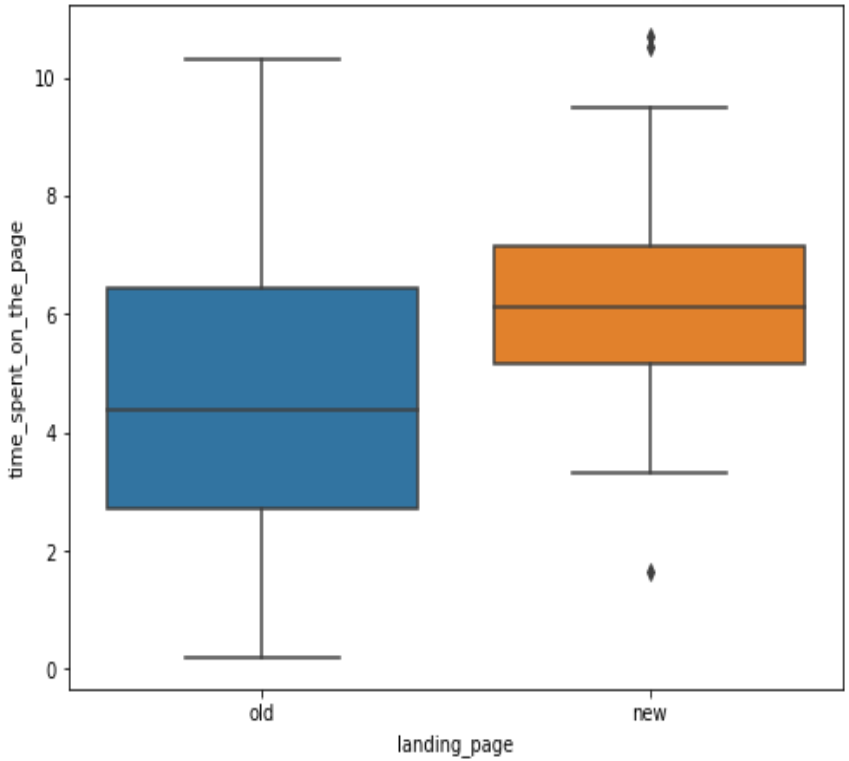



hypothesis


H_0
H_1
Hypothesis testing

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

- Average time spent by the users on old and new landing page are 4.53 and 6.22, respectively.
- On an average users spend a lot of time on new landing page.
- Let us prove the same using hypothesis testing.



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

Assumptions:

- Time spent on the page is a continuous data.
- We have already checked that the time spent on the page column is approximately normally distributed.
- The data is a random sample from the population.
- The std deviation of the population is not known. The standard dev of the sample not equal.
- As we are taking random samples from for 2 different type of users, the two samples are from 2 independent populations.

This is a **one tailed two independent sample T –Test.**

Let μ1,μ2 be the mean time spent on the new landing page and the mean time spent on the old landing page. Mathematically, the above formulated hypotheses can be written as:

H0:μ1=μ2

Ha:μ1>μ2

The p-value for the test is 0.00013. Since p-value<0.05 we have enough statistical evidence to reject Null hypothesis.

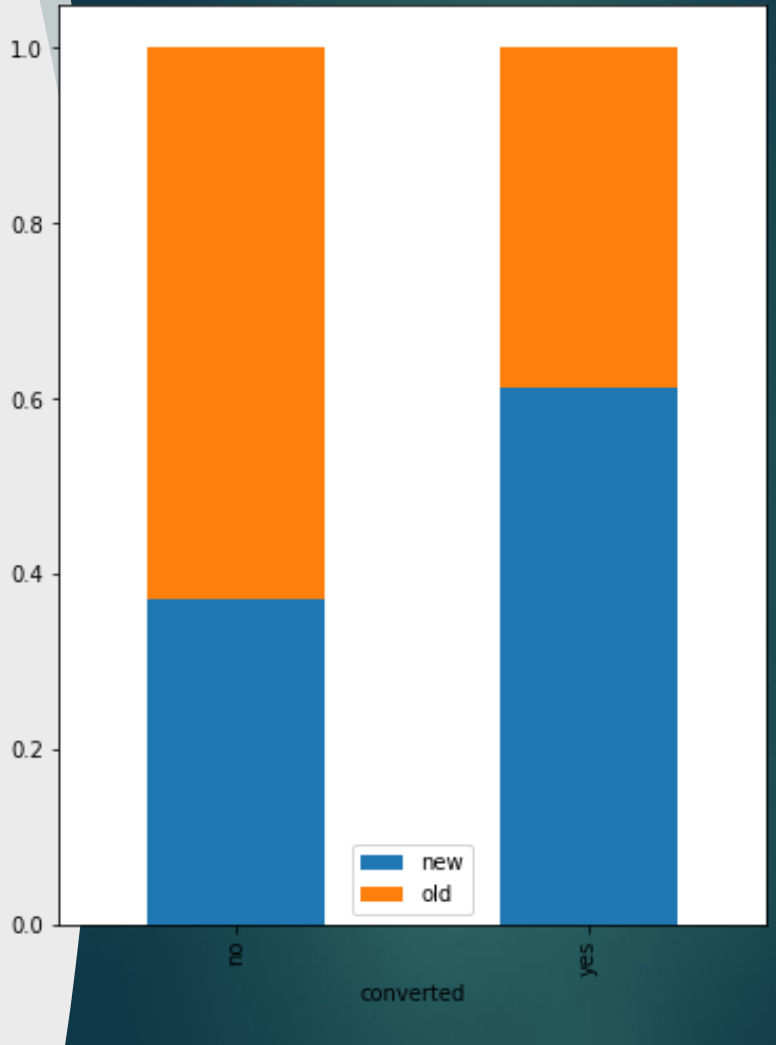
Conclusion: The user spent more time on the new landing page than the old landing page.

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

There are 50 total users in each for old and new page.

The conversion rate on the new landing page looks like greater than old page.

Now, we shall test the same using statistical methodologies.



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

- **The conversion rate on the old page is 23 out of 50 users.**
- **The conversion rate on the new page is 32 out of 50 users .**

Assumptions:

- **The user may either get converted or not converted to new subscriber. So, it is binomially distributed population.**
- **The sample is randomly selected from population.**
- **•np and n(1-p) are greater than or equal to 10. So, it can be approximated as normal population.**

This is a one tailed two proportion Z –Test.

Let p1 and p2 be the proportions for the new landing page and old landing page.

H0: p1=p2

Ha:p1>p2

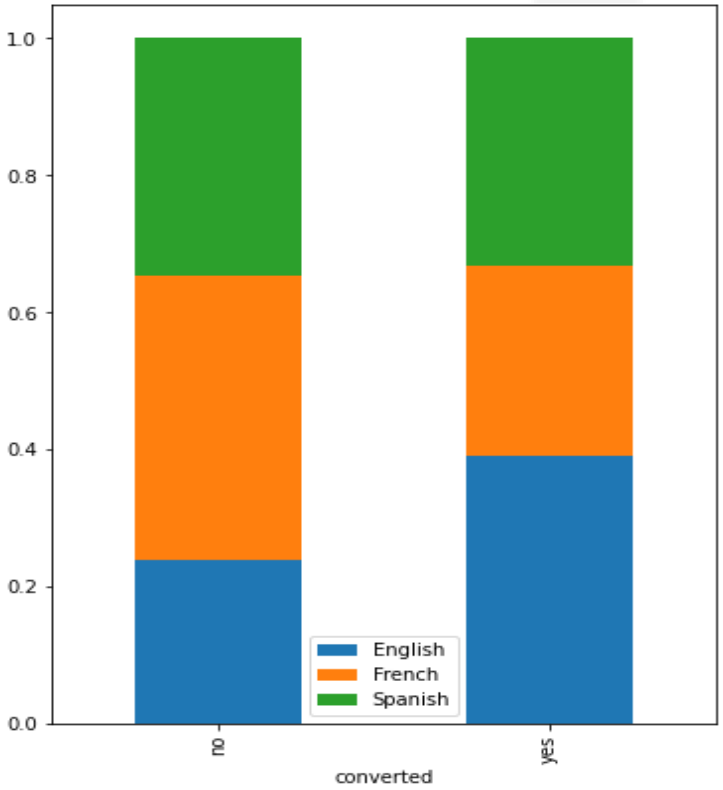
The p-value is 0.008. Since p-val < 0.05, hence we have sufficient evidence to reject null hypothesis.

- **Conclusion: The conversion rate on the new landing page is greater than old landing page.**

Does the converted status depend on the preferred language?

- Cross tabulation of Preferred Language and Converted Status.

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



Hypothesis test:

Assumptions:

- We have two categorical feature.
- Number of observation in each level of variable is more than 5.
- Random sample from the population.
- Hence, we will apply **Chi Square test of independence**.

Let us define NULL and ALTERNATIVE hypothesis

Ho:The conversion status and preferred language are independent of each other.

Ha :The conversion status and language preferred are dependent on each other.

The p-value is 0.212.

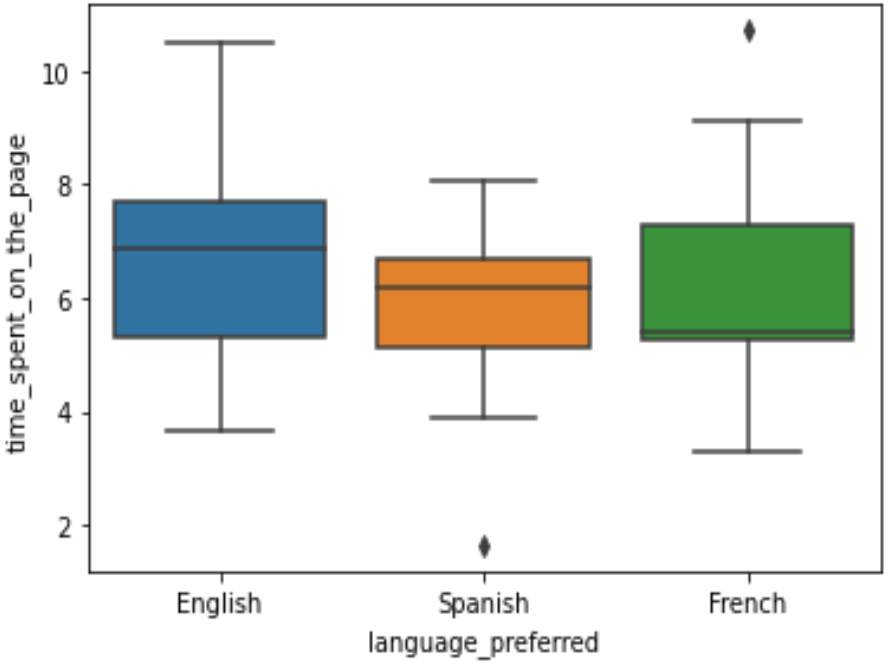
Since p-val > 0.05, hence we fail to reject null hypothesis.

Conclusion: The converted status does not depend on preferred language.

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

- Average Time spent on the new landing page by different language user

language_preferred	Mean
English	6.663750
French	6.196471
Spanish	5.835294



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

We will apply Shapiro-Wilk's test to check normality in time spent on the new landing page data.

Shapiro Wilk’s Test:

Ho: Time spent on new page follows normal distribution.
Ha: Time spent on new page does not follow normal distribution.

The p-value is 0.804.
Since p-value > 0.05 ,we fail to reject null hypothesis.
We conclude that Time spent on the new landing page is normally distributed.

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

We will apply Levene's test to check if the population variance is equal.

Levene’s Test:

Ho: All population variances are equal.
Ha: At least one variance is different.

The p-value is 0.467. p-value > 0.05. So, We fail to reject null hypothesis.
We conclude that population variance are equal.

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

Assumptions :

- ✓ Time spent on the new page follows normal distribution. This is proved by Shapiro's test.
- ✓ Population variance is equal. This is proved by Levene's test.
- ✓ The samples are randomly selected from population.
- ✓ Hence, we will apply one way ANOVA test.

Let μ1,μ2 and μ3 be the means of time spent on new page for preferred language English, Spanish and French, respectively.

Ho: μ1=μ2=μ3

Ha: At least one mean time spent on page is different from the rest.

The p-value is 0.432. p-val > 0.05, We fail to reject null hypothesis.

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

Conclusion based on Hypothesis Testing

1)The means time spent on the new landing page is more than the old landing page.

2)Users on the new page seems to have converted at a high rate than those on the old page.

3)The conversion status is independent of the preferred language.

4)The mean time spent on the new landing page by different language users is same.

Business Recommendation

- It is recommended that the company should implement the newly designed landing page as it has been statistically proven to be effective and successful in increasing the number of subscribers.
- The company has enough number preferred languages. Adding more language may not guarantee more subscribers.

THANK YOU

Great Learning

Happy Learning !