# E-news Express - Case Study

Jesus Torres

# Enews Expess

- Most popular news portal in India as well as Bengal.
- Find the latest and breaking news.
- Most recent and latest news about Entertainment, Business, Politics, Health, Astrology and many more.

# OBJECTIVE

Statistical analysis of business data. Explore the dataset and extract insights from the data. The idea is for you to get comfortable with doing statistical analysis in Python.

You are expected to perform the statistical analysis to answer the following questions:

1. Explore the dataset and extract insights using Exploratory Data Analysis.
2. Do the users spend more time on the new landing page than the old landing page?
3. 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?
4. Does the converted status depend on the preferred language? [Hint: Create a contingency table using the pandas.crosstab() function]
5. Is the mean time spent on the new page same for the different language users?

# DATA OVERVIEW

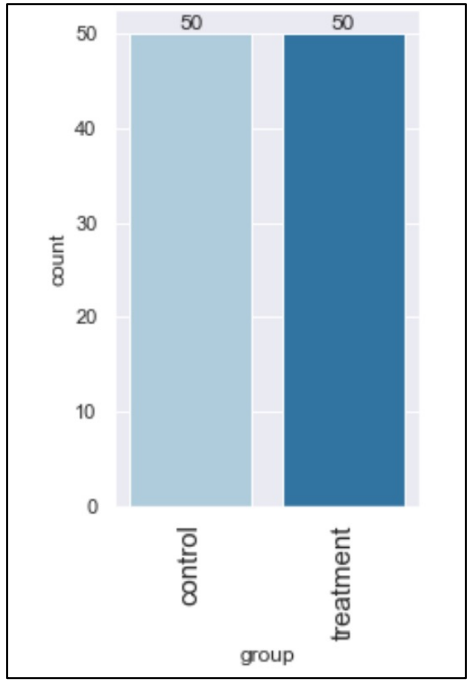
The dataset contains the different data related to each user. The detailed data dictionary is given below:

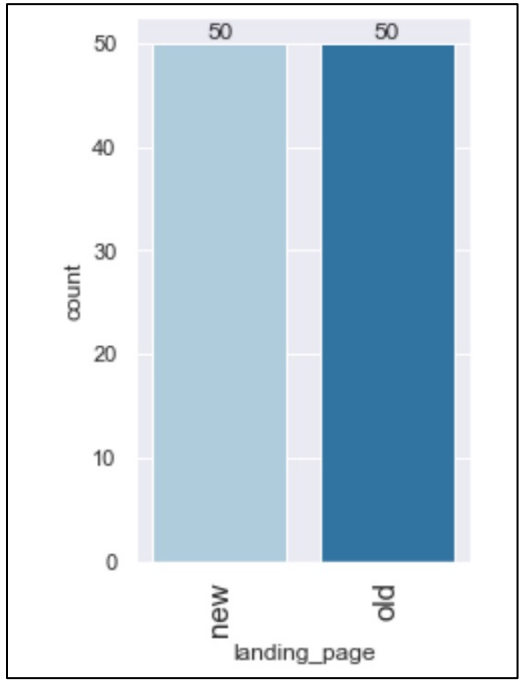
- user_id - This represents the user ID of the person visiting the website.
- group - This represents whether the user belongs to the first group (control) or the second group (treatment).
- landing_page - This represents whether the landing page is new or old.
- time_spent_on_the_page - This represents 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 | Features |
| --- | --- |
| 100 | 6 |

- There are a total of 100 observations.
- Each observation has 6 features.

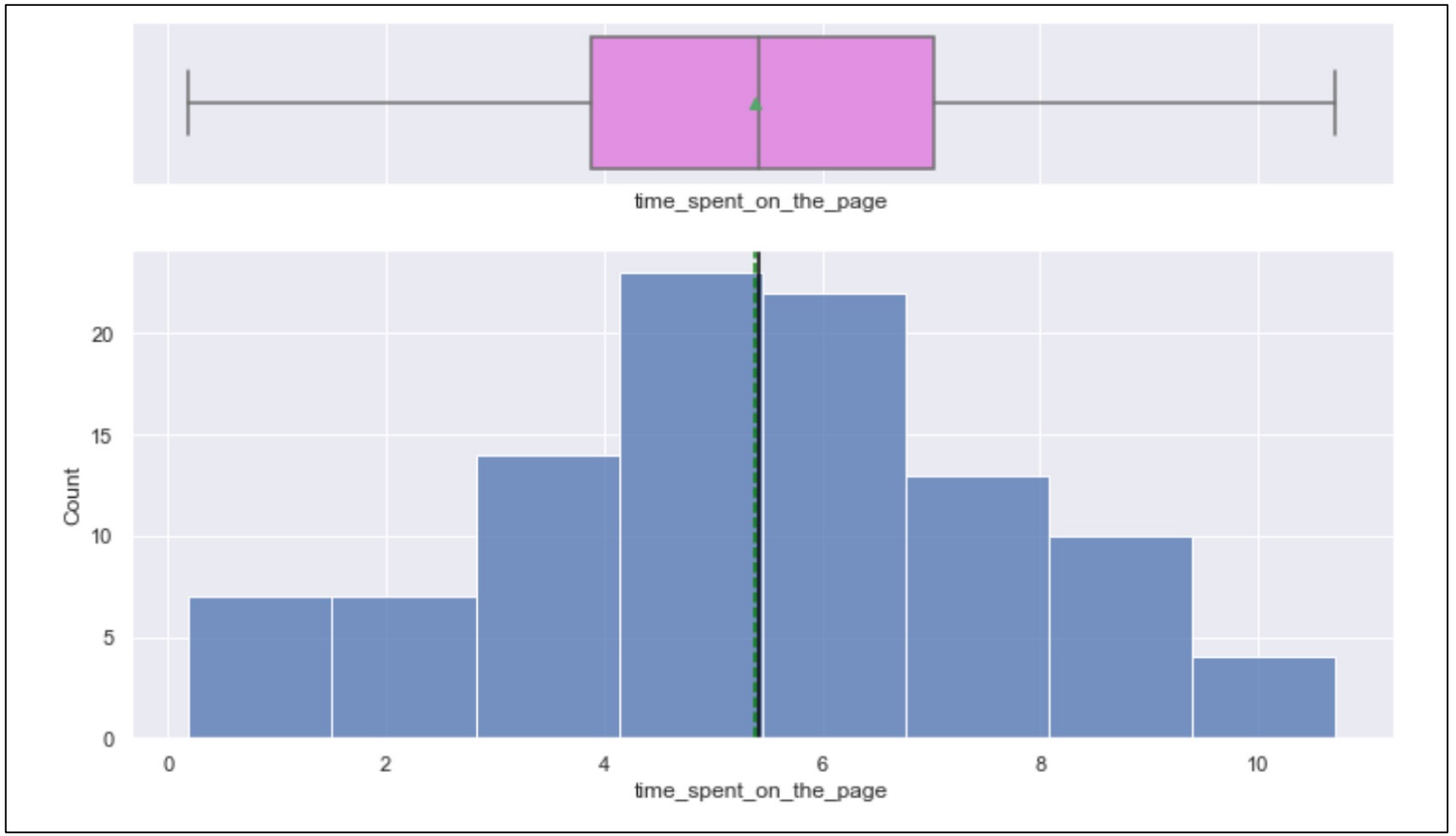
# UNIVARIATE ANALYSIS – GROUP AND LANDING PAGE




Features ‘group’ and ‘landing page’ are related. Both have two unique values and are split the same. 50 observations for each.

# UNIVARIATE ANALYSIS – TIME SPENT ON LANDING PAGE



# UNIVARIATE ANALYSIS – TIME SPENT ON LANDING PAGE

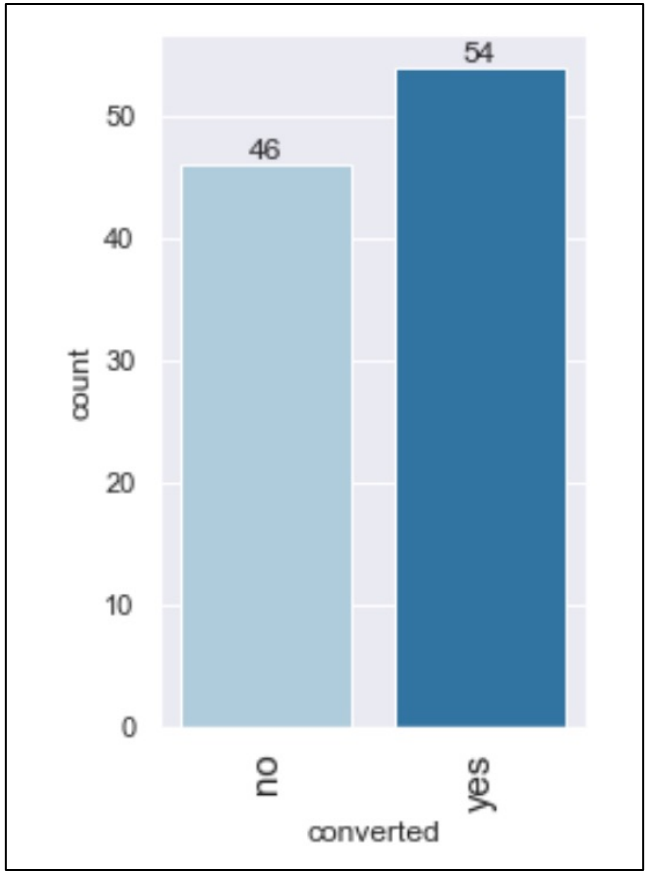
Observations:

1. The average "time spent on the page" is about 5.38 minutes
2. The standard deviation is about 2.38 minutes
3. Inter-quartile range includes:
    - Min = 0.19
    - 25% quartile = 3.88
    - 50% quartile = 5.42
    - 75% quartile = 7.02
    - Max = 10.71
4. The data seems to follow a normal distribution as seen on the barplot & histogram

# UNIVARIATE ANALYSIS – CONVERTED

Observations:

1. The feature ‘converted’ has 2 unique values
   - No – 46 observations
   - Yes – 54 observations

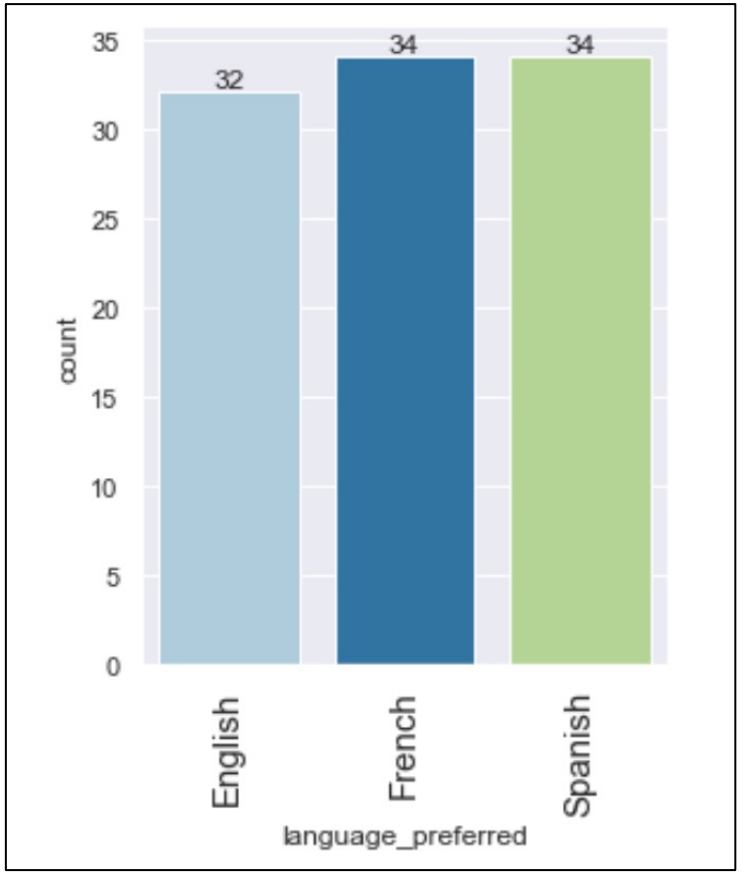


# UNIVARIATE ANALYSIS – LANGUAGE PREFERRED

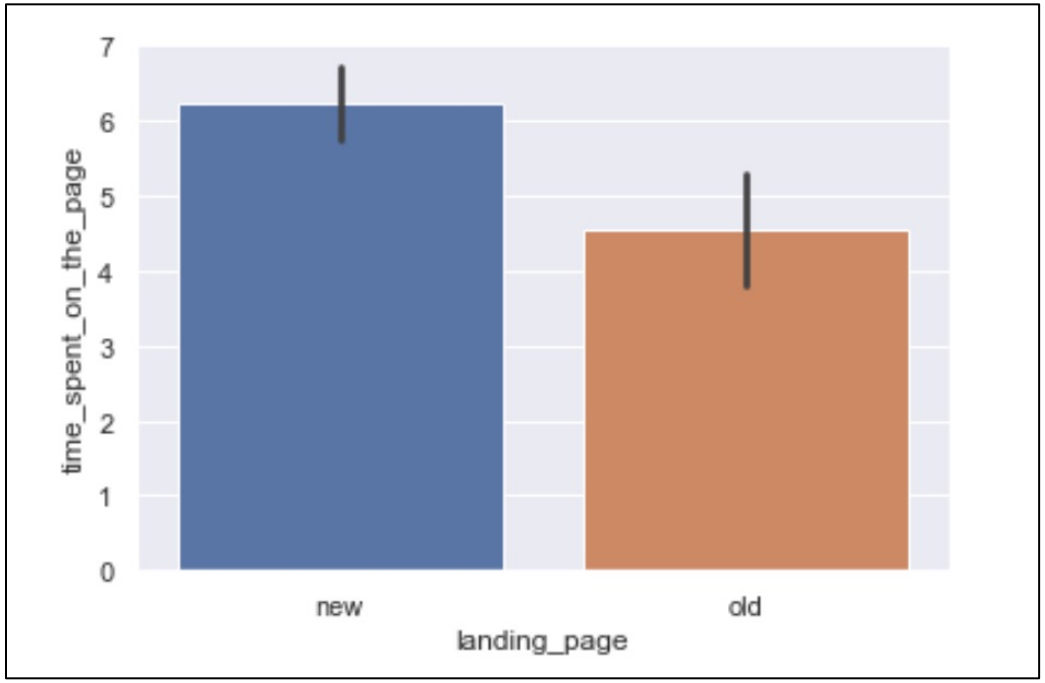
Observations:

1. The feature 'language preference' has 3 unique values
   - English – 32 observations
   - French – 34 observations
   - Spanish – 34 observations

2. Both French and Spanish have the highest frequencies

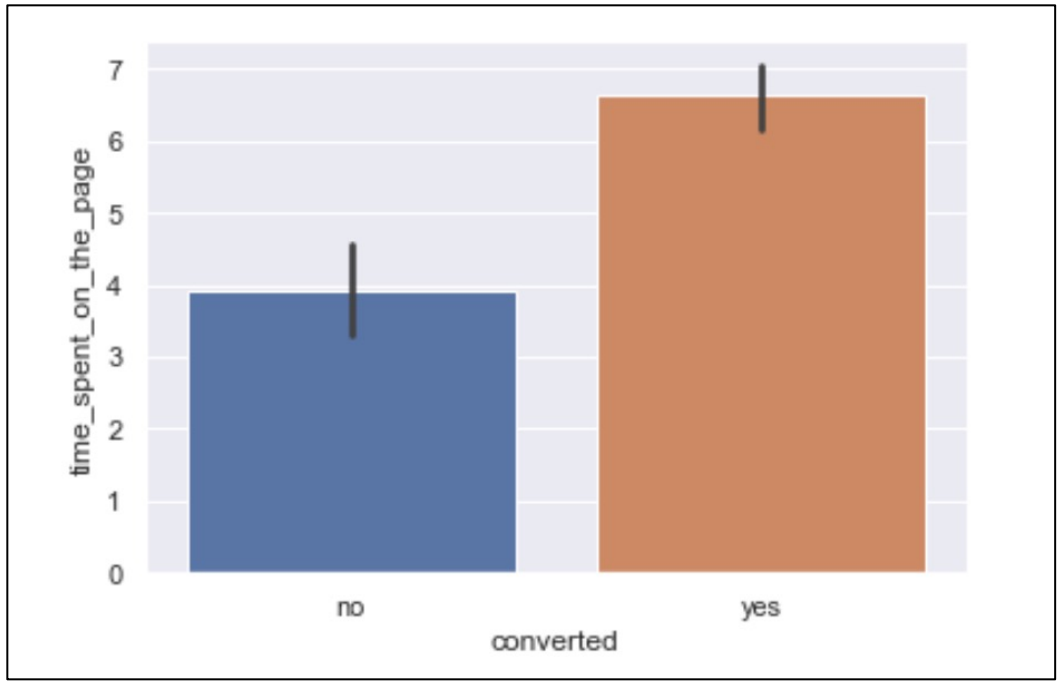


# BIVARIATE ANALYSIS – LANDING PAGE & TIME SPENT ON THE PAGE



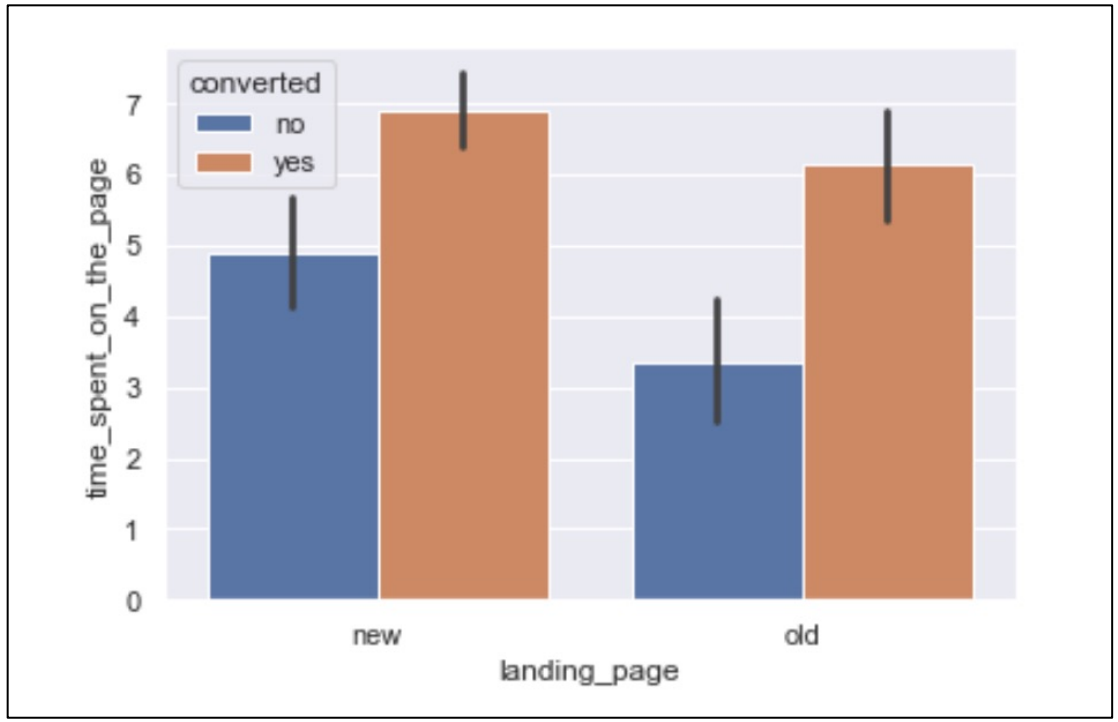

Observation: The same observation from above applies to this one too. Users exposed to the new landing page spend more time on average than those on the old landing page.

# BIVARIATE ANALYSIS – CONVERTED & TIME SPENT ON THE PAGE




Observation: Users who converted demonstrate on average a higher time spent on the page.

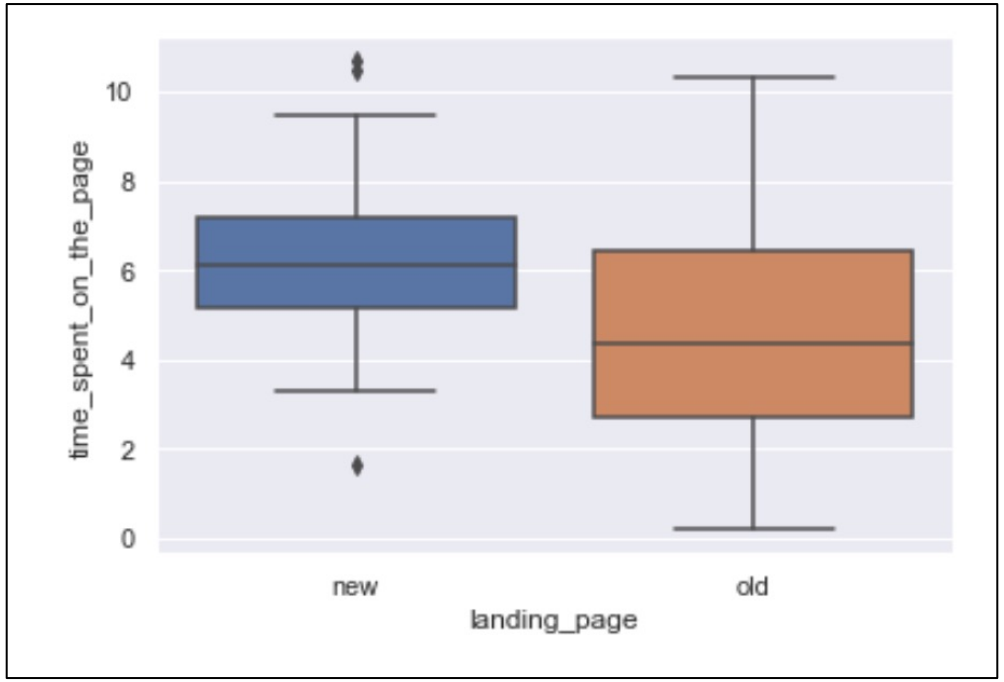
# BIVARIATE ANALYSIS – LANDING PAGE, TIME SPENT ON THE PAGE, CONVERTED




Observation: The graph demonstrates that more users are likely to convert to the new landing page despite whatever landing page they are first exposed to and those who do convert demonstrate higher time averages.

# HYPOTHESIS TESTING

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




Observations: On average users seem to spend more time on the new page than on the old page.

# HYPOTHESIS TESTING

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

$H0$: The mean time spent on the old landing page.
$Ha$: The mean time spent on the new landing page.

Let $\mu 1, \mu 2$ be the mean time spent on the old landing page vs the mean time spent on the new landing page.

Mathematically, the above formulated hypotheses can be written as:
$H0$: $\mu 1 = \mu 2$
$Ha$: $\mu 1 < \mu 2$

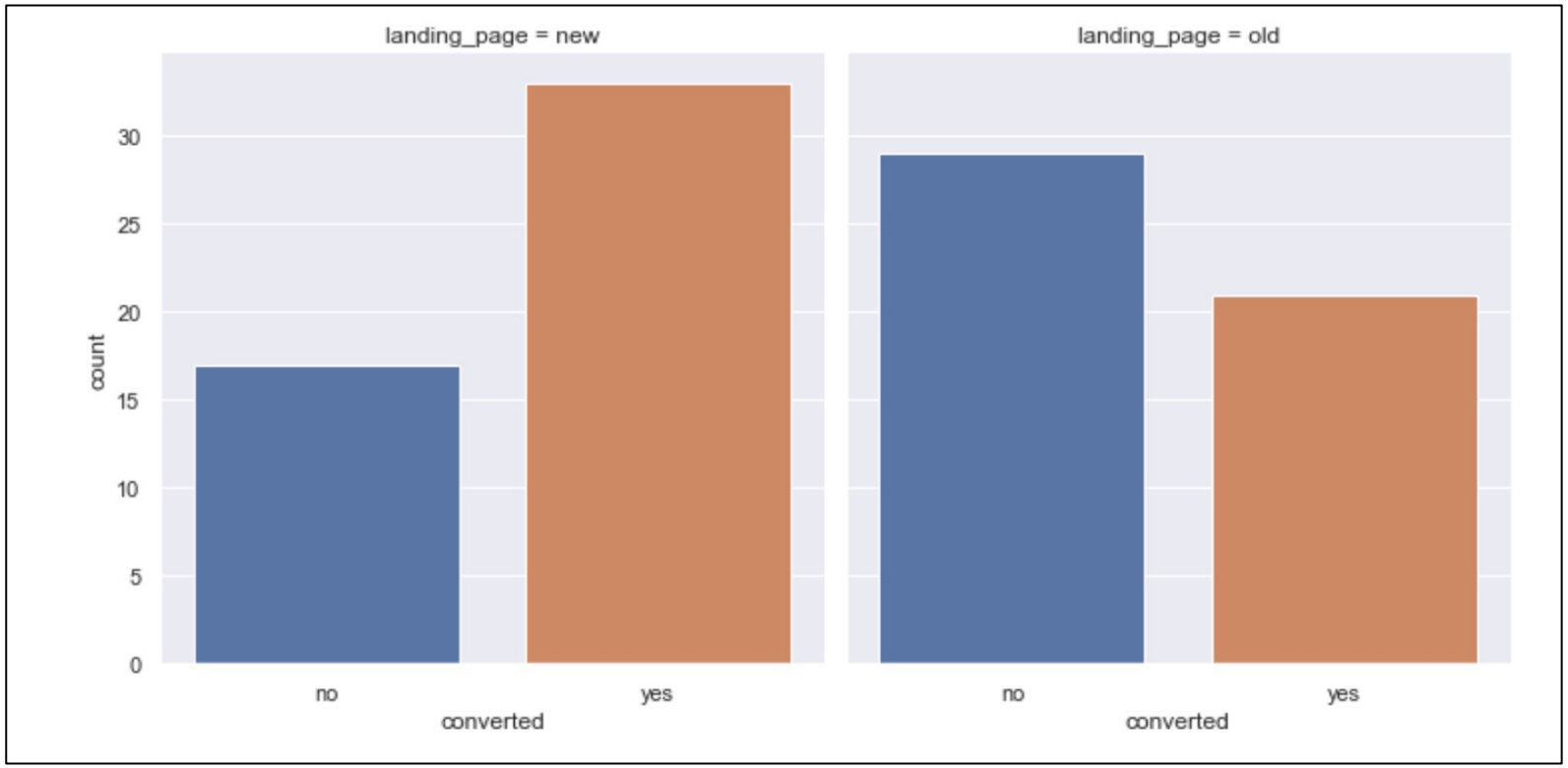
# HYPOTHESIS TESTING – RESULTS & INFERENCE

In this scenario, the p value is 0.0003 which is less than the 0.05. Hence, we reject the null hypothesis.

This means that the new landing page encourages users to spend more time significantly.

# HYPOTHESIS TESTING

Q: 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?




Observations: Users engaging with the new landing page are more likely to convert than those using the old landing page.

# HYPOTHESIS TESTING

Q: 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?

Let $p1, p2$ be the proportions of converters in old landing page and new landing page respectively.

The company will test the null hypothesis
$H0: \ p1 = p2$

against the alternate hypothesis
$Ha: \ p1 \neq p2$

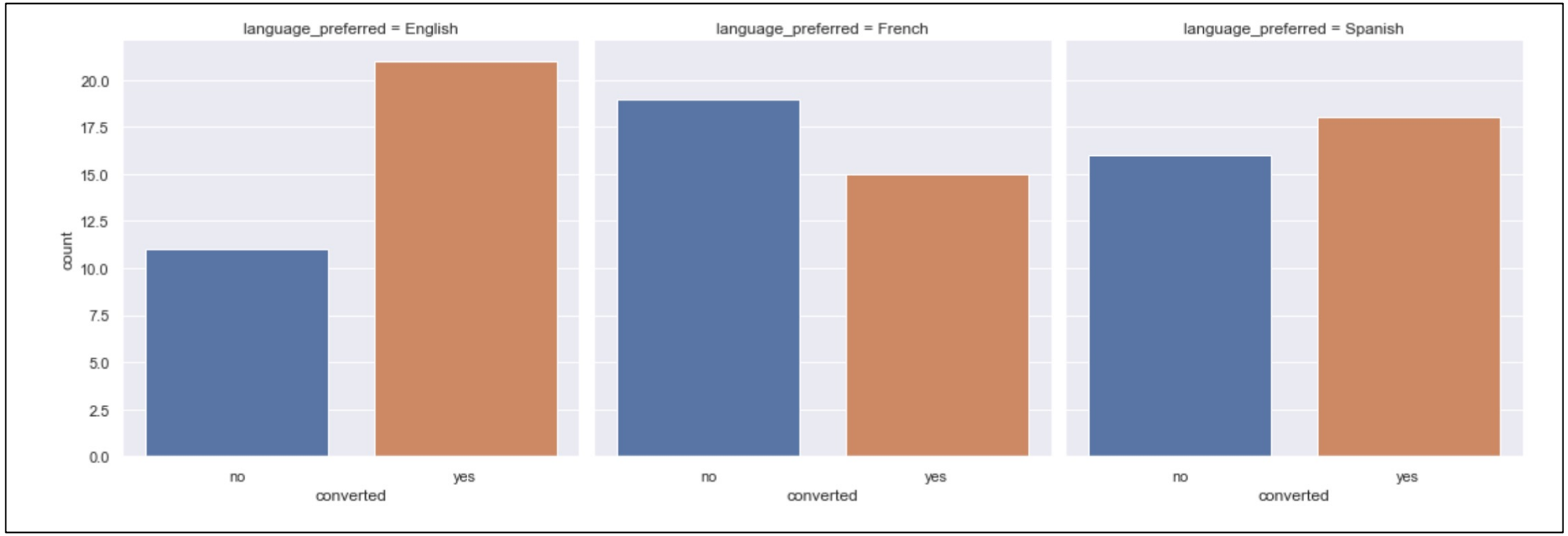
# HYPOTHESIS TESTING – RESULTS & INFERENCE

In this scenario, the p value is 0.016 which is less than the 0.05. Hence, we reject the null hypothesis.

This means that the conversion rate among the two groups, old landing page and new landing page are significantly different.

# HYPOTHESIS TESTING

Q: Does the converted status depend on the preferred language?




Observations: The converted status doesn't seem to have much of an effect on French and Spanish speaking users but does play more of a role with English speaking users.

# HYPOTHESIS TESTING

Q: Does the converted status depend on the preferred language?

We will test the null hypothesis
$H0$: Converted status is independent of language preferred.

against the alternate hypothesis
$Ha$ Converted status depends on language preferred.

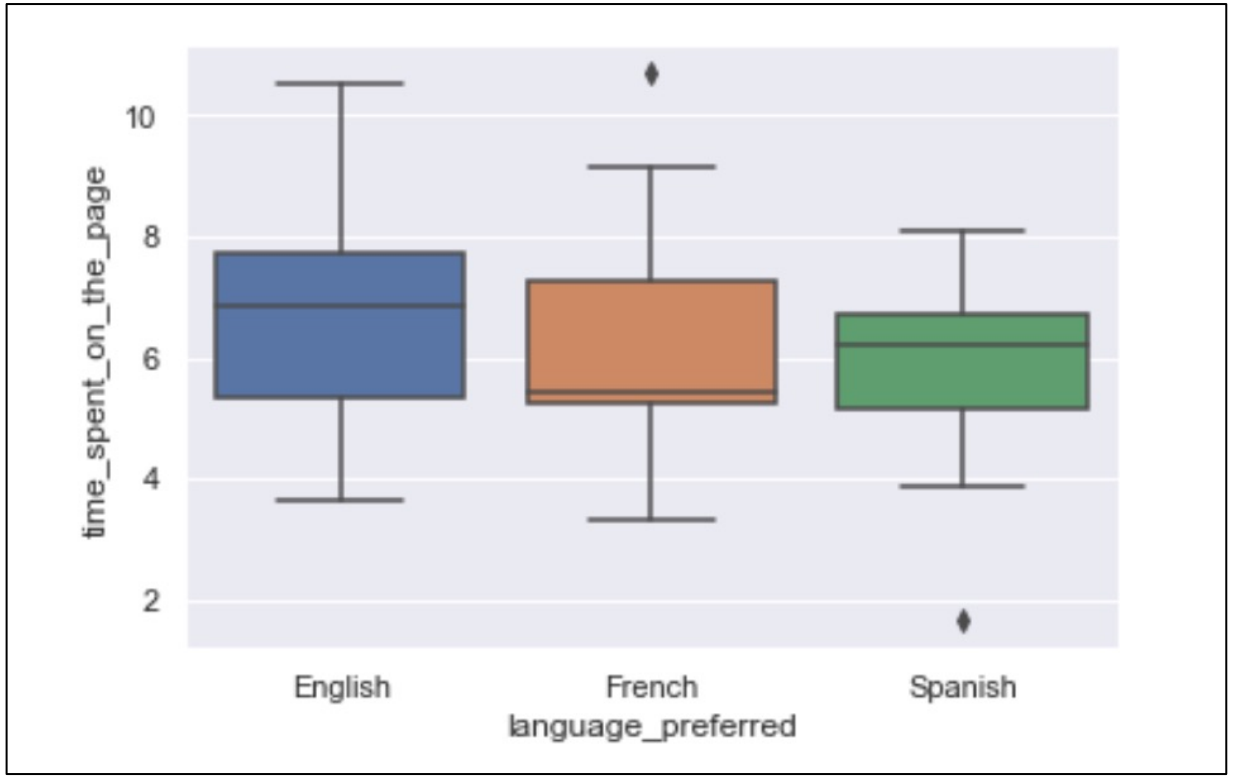
# HYPOTHESIS TESTING – RESULTS & INFERENCE

In this scenario, the p value is 0.213 which is greater than the 0.05. Hence, we fail to reject the null hypothesis.

This means that the converted status is independent of language preferred.

# HYPOTHESIS TESTING

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




Observations: The average time spent on the pages seems relatively the same among the different language preferences.

# HYPOTHESIS TESTING

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

Let $\mu 1, \mu 2, \mu 3$ be the means of time spent on the new landing page for language preferences English, Spanish and French respectively.

We will test the null hypothesis
$H0: \ \mu 1 = \mu 2 = \mu 3$

against the alternative hypothesis
$Ha$: At least one time average of the new landing page is different among the language preferences.

# HYPOTHESIS TESTING – RESULTS & INFERENCE

In this scenario, the p value is 0.432 which is greater than the 0.05. Hence, we fail to reject the null hypothesis.

This means that the time averages on the new landing page for the different languages are relatively the same.

# CONCLUSION

1. E-news users exposed to the new landing page on average spend more time than on the old landing page.
   - P-value of 0.0003
   - This means that the new landing page encourages users to spend more time significantly.

2. Users engaging with the new landing page are more likely to convert than those using the old landing page.
   - P-value of 0.016
   - This means that the conversion rate among the two groups, old landing page and new landing page are significantly different.

3. The converted status doesn't seem to have much of an effect on French and Spanish speaking users but does play more of a role with English speaking users.
   - P-value of 0.213
   - This means that the converted status is independent of language preferred.

4. The average time spent on the pages seems relatively the same among the different language preferences.
   - P-value of 0.432
   - This means that the time averages on the new landing page for the different languages are relatively the same.

# RECOMMENDATIONS

1. Fully implement the new landing page for users to use since it increases the time spent on the landing page.

2. Do not offer users anymore the option to use the old landing page since it doesn't increase time spent. In only offering the new landing page users will be more likely to adjust.

3. Language preference has no effect on time spent on the landing page and conversion rate. Continue to offer users the already provided language options.