E-news Express-Project

Dharitri Bollini

E-NEWS EXPRESS

- Most popular news portal in globally.
- Latest and breaking news in E-News.
- Most recent and latest news about Entertainment, Business, Politics, Health, Astrology, Travel and many more.

OBJECTIVE

Statistical analysis of business data. Explore the dataset and extract insights from the data.
You are expected to perform the statistical analysis to answer the following questions:

- Explore the dataset and extract insights using Exploratory Data Analysis.
- 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? [Hint: Create a contingency table using the pandas. Crosstab() function]
- Is the mean time spent on the new page same for the different language users?

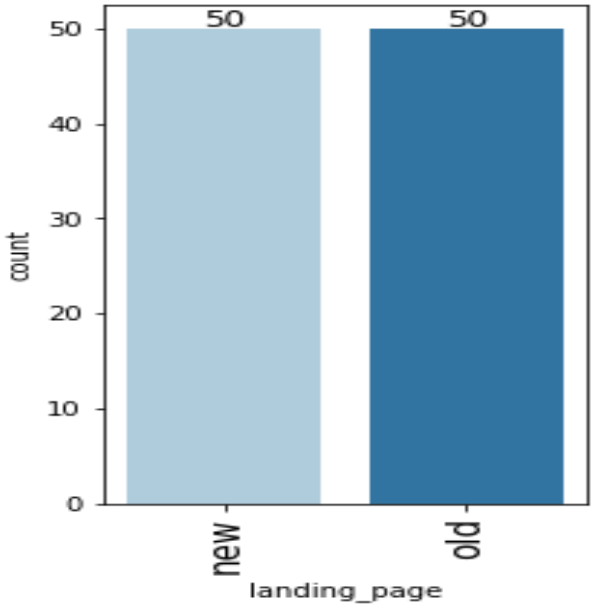
DATA OVERVIEW

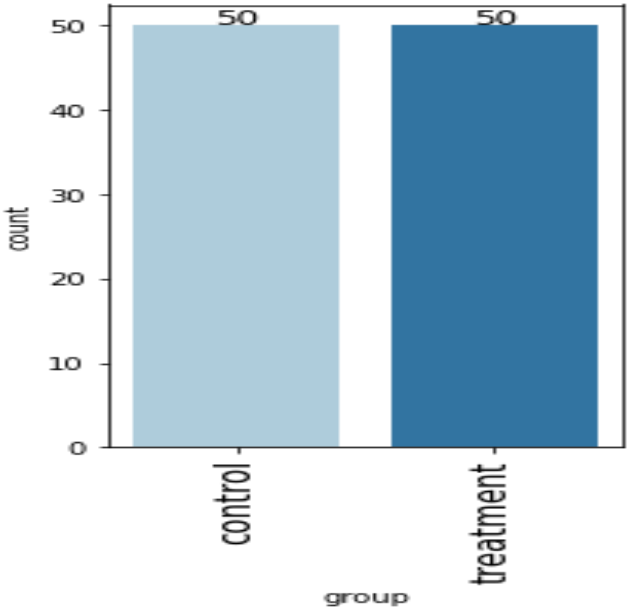
The detailed data dictionary is given below:

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

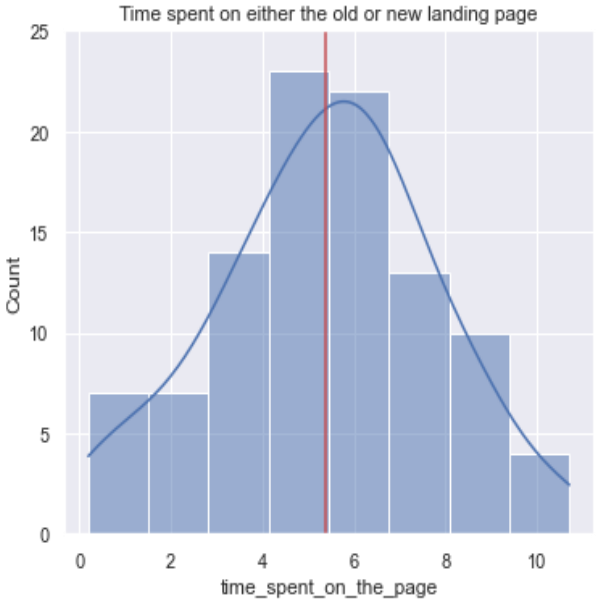
UNIVARIATE ANALYSIS-GROUP AND LANDING PAGE

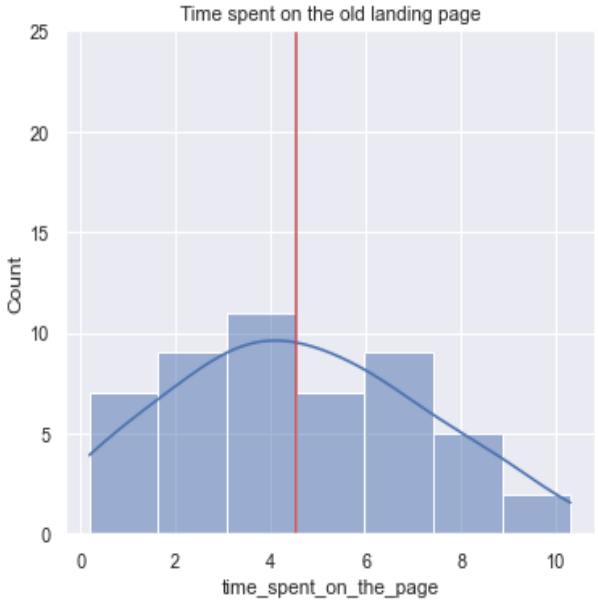
Features ‘group’ and ‘landing page’ are related. Both have two unique values and are split the same. 50 observations for each.

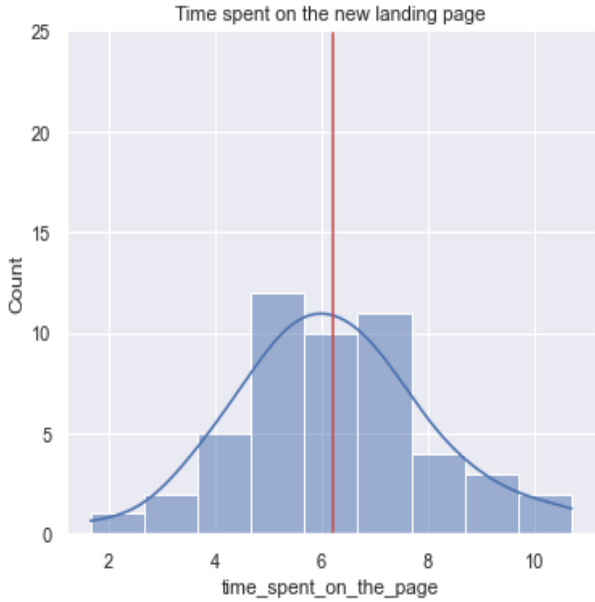




UNIVARIATE ANALYSIS-TIME SPENT ON LANDING PAGE







UNIVARIATE ANALYSIS-TIME SPENT ON LANDING PAGE

OBSERVATIONS

- Total & Average time spent on the old landing page across 50 users (Control group):

  226.32 mins & 4.53 mins

- Total & Average time spent on the old landing page across 50 users (Treatment group):

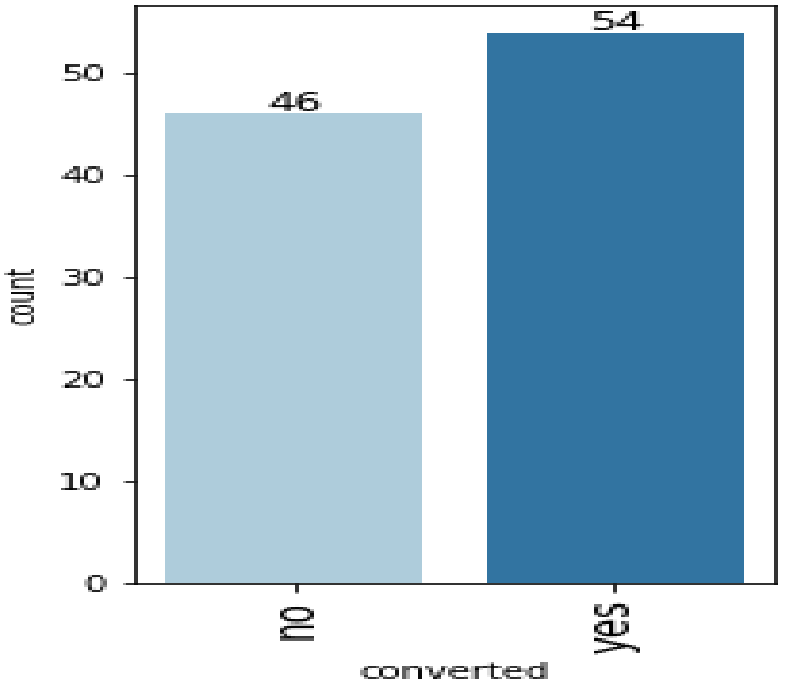
  311.16 mins & 6.22 mins

- Data for time spent on the new landing page follows more Normal distribution than time spent on the old landing page.
- Total across all users and average time spent on the new landing page is higher than the old landing page.

UNIVARIATE ANALYSIS-CONVERTED

The Feature ‘converted’ has 2 unique values

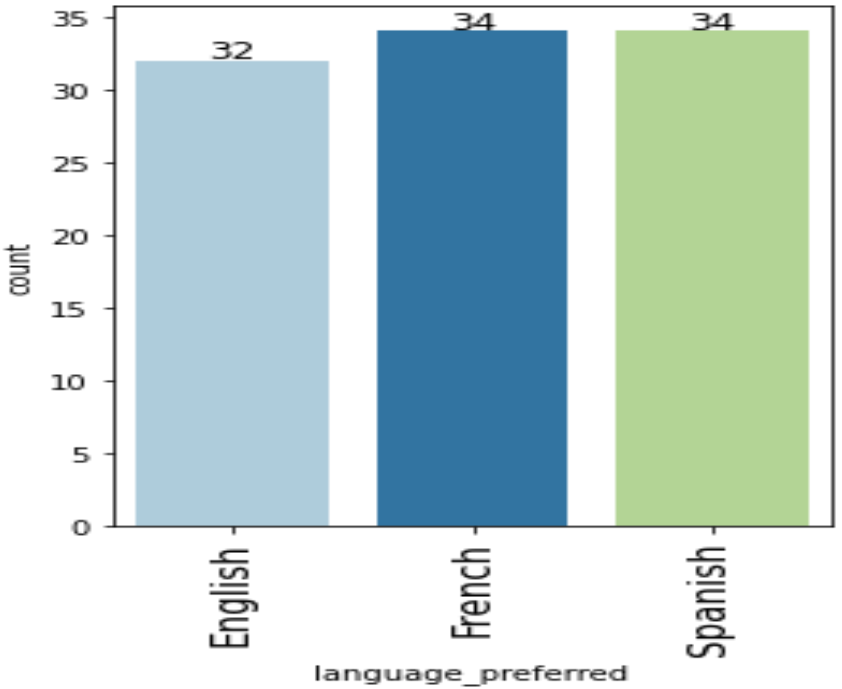
- NO-46 observations
- YES- 54 observations



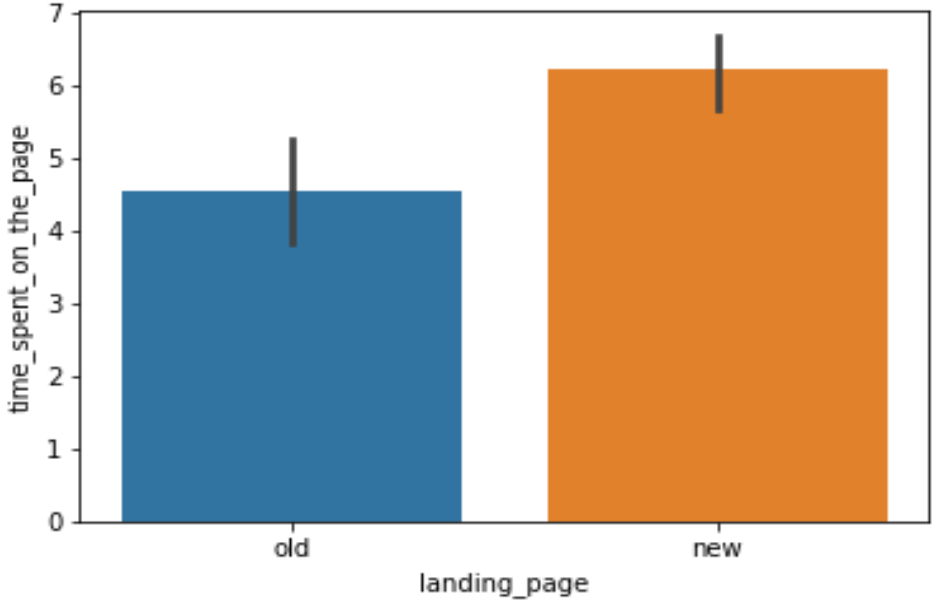
UNIVARIATE ANALYSIS-LANGUAGE PREFERRED

Feature ‘language preferred’ has 3 unique values

- English-32 observations
- French-34 observations
- Spanish-34 observations
- Both French and Spanish have highest users .

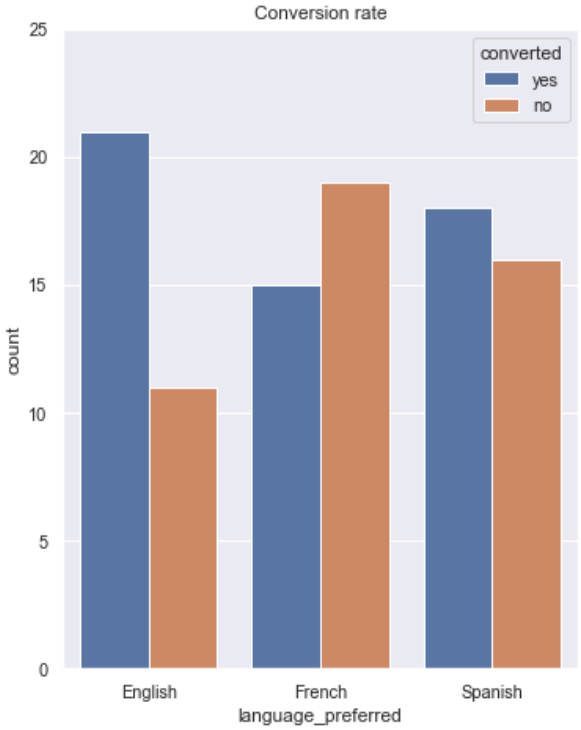


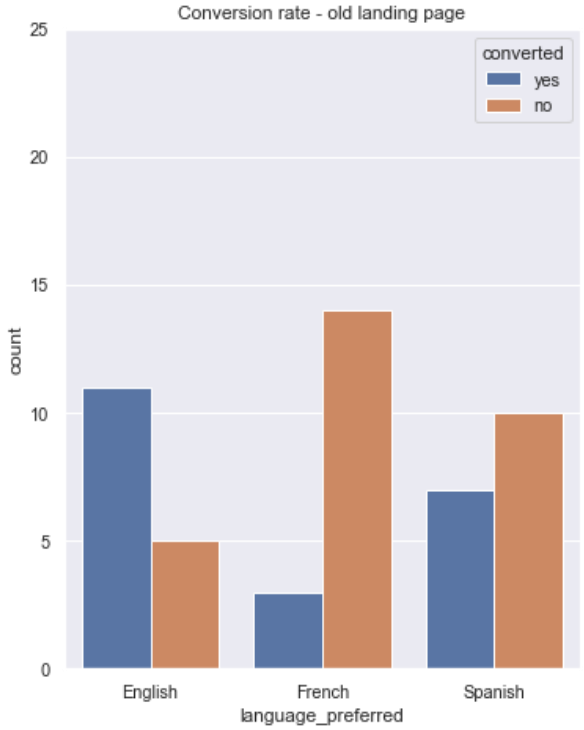
BIVARIATE ANALYSIS-LANDING PAGE & TIME SPENT ON THE PAGE

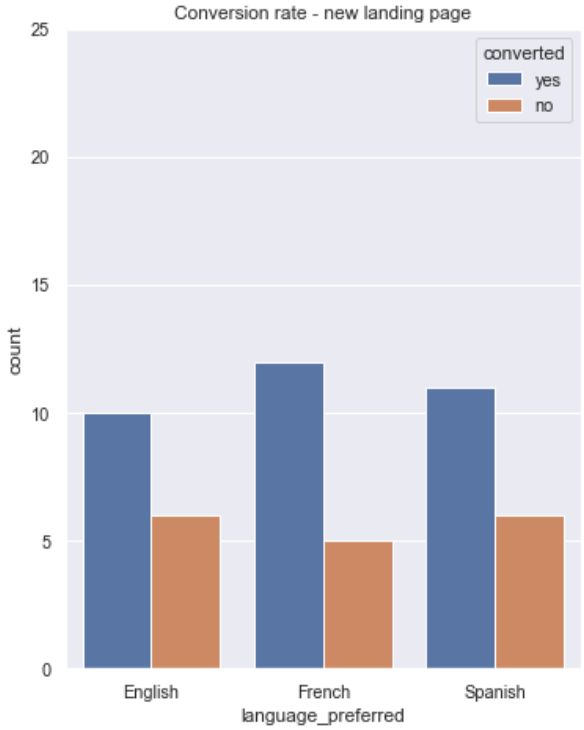


Observation: Users exposed to the new landing page spend more time on average than those on the old landing page.

BIVARIATE ANALYSIS-CONVERTED,LANDING,TIME SPENT ON THE PAGE







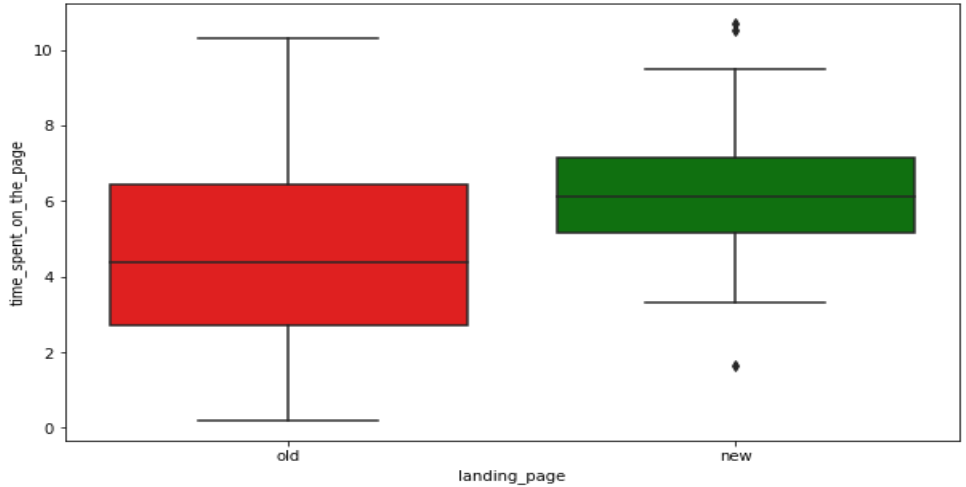
BIVARIATE ANALYSIS-CONVERTED,LANDING,TIME SPENT ON THE PAGE

OBSERVATIONS

- Change in converted –New landing page vs old landing page:
- English-10.00, French- 75.00,Spanish-36.36
- Change in conversion for new landing page vs old landing page is -10% for English. This means users that prefer English converted marginally more when on the old landing page.
- Change in conversion for new landing page vs old is 75% for French and 36.36% for Spanish. users that prefer these languages converted more when on the new landing page.
- Change in conversion on the new landing page is more among users that prefer French.

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 landing page than on the old landing page

HYPOTHESIS TESTING - RESULTS

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

H0 : Null Hypothesis - The mean time spent on the new landing page is the same as the mean time spent on the old landing page.

Ha : Alternate Hypothesis - The mean time spent on the new landing page is higher than the mean time spent on the old landing page.

Let μ1 and μ2 be the mean times spent on the new and old landing pages.

H0 : $\mu1 = \mu2$

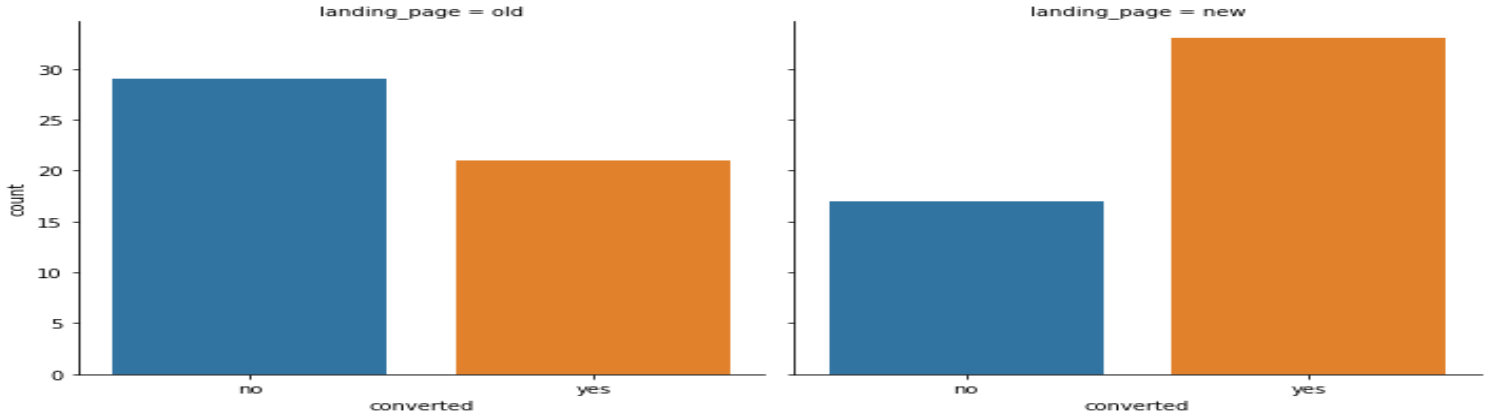
Ha : $\mu1 > \mu2$

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 than on 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?



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

HYPOTHESIS TESTING -RESULTS

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?

H0 : Null Hypothesis - The conversion rate (the proportion of users who visit the landing page and get converted) for the new page is the same as the conversion rate for the old page.

Ha : Alternate Hypothesis - The conversion rate for the new page is greater than the conversion rate for the old page.
Let p1 and p2 be the conversion rates for the new and old landing pages.

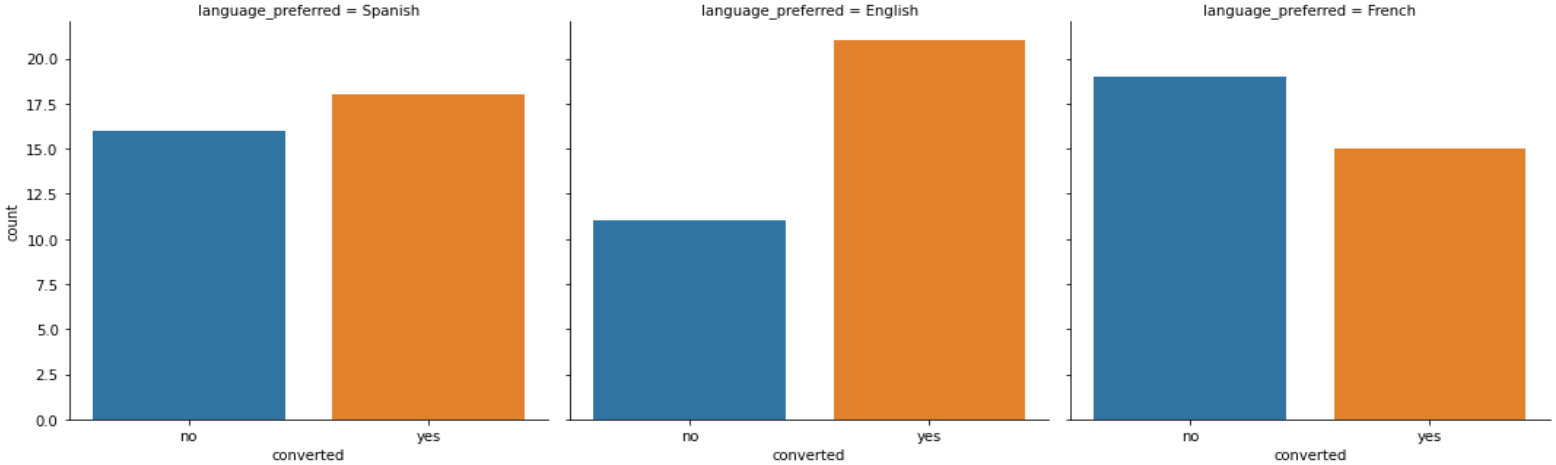
- H0 : p1 = p2
- Ha : p1 > p2

In this scenario, the p value is 0.008 which is less than the 0.05. Hence, we reject the null hypothesis.

This means that the conversion 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 - RESULTS

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

H0 : Null Hypothesis - Converted status is independent of the preferred language.

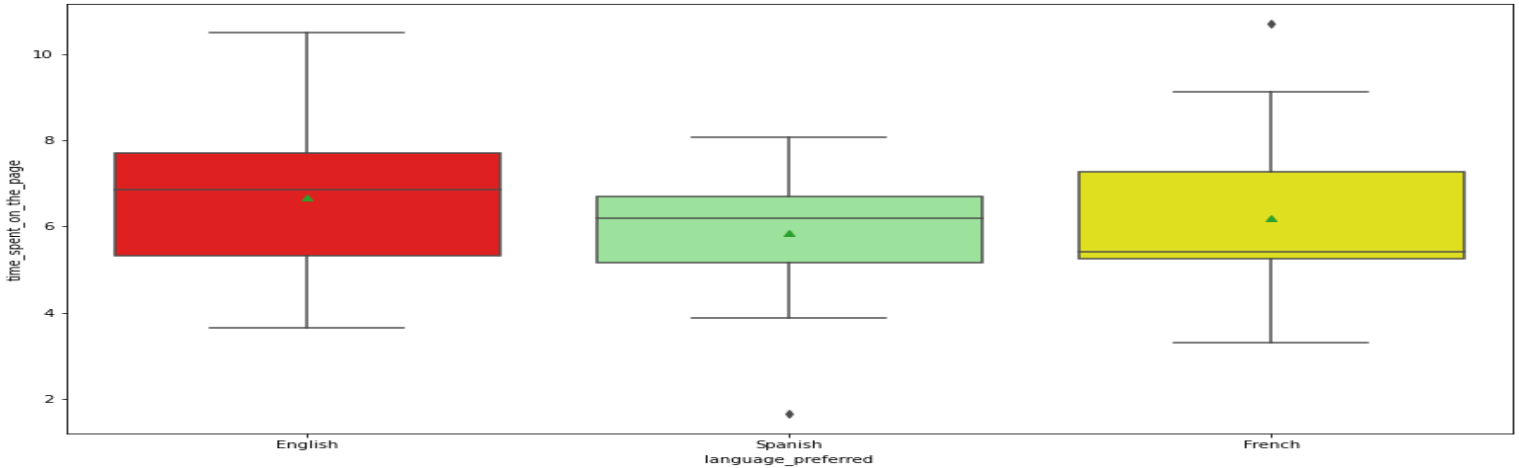
Ha : Alternate Hypothesis - Converted status is dependent on the preferred language.

In this scenario, the p value is 0.21 which is greater than the 0.05. Hence, we cannot reject the null hypothesis.

This means that the conversion 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-RESULTS

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

Let μ1, μ2, μ3 be the mean time spent on the new landing page for the different language users - English, French and Spanish respectively.

- H0: Null Hypothesis- μ1 = μ2 = μ3
- Ha: Alternate Hypothesis- At least one of μ1, μ2, or μ3 is different among the language preferences.

In this scenario, the p value is 0.432 which is greater than the 0.05. Hence, we cannot reject the null hypothesis.

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

SUMMARY-EXPLORATORY DATA ANALYSIS

- Total (across all users) and average time spent on the new landing page (311.16 & 6.22 mins) is higher than on the old landing page (226.62 & 4.53 mins).

- % conversion rate is higher for new landing page (66%) when compared to old landing page (42%).

- There are approximately equal number of English, French and Spanish preferring users (32, 34 and 34) in the dataset. % change in conversion for new landing page vs. old landing page is -10% for English, 75% for French and 36.36% for Spanish. Conversion when on the new landing page is very strong among users preferring French and marginally poor for users preferring English.

- The median time spent on the new landing page are significantly more than old landing page for all users irrespective of preferred language. The mean time spent on the new landing page are 6.66mins, 6.19mins and 5.83 mins for users that speak English, French and Spanish (~approximately equal).

SUMMARY-STATISTICAL ANALYSIS

- We statistically found enough evidence to support the claim that mean time spent on the new landing page is higher than mean time spent on the old landing page using a two independent sample T-test for equality of means - unequal std dev.

- We statistically found enough evidence to support the claim that conversion rate for the new landing page is higher than conversion rate for the old landing page using a two proportion Z-test.

- We did not find enough evidence to support the claim that conversion rate is dependent on the preferred language using a chi-square test for independence.

- We did not find enough statistical significance to conclude that at least one mean time spent on the new landing page for different preferred languages is different from the rest at 5% significance level using One-way ANOVA Test. Normality and equality of variances assumptions for ANOVA were tested using Shapiro-Wilk’s test and Levene's test respectively.

SUMMARY-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 0.008
- 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 0.21
- 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) The new landing page was found to be statistically better in time spent by users on the page as well as the conversion rate of a user as a subscriber when compared to the old landing page. Hence, the new landing page is effective from business standpoint.

2) While we did not find enough statistical evidence from the dataset that conversion rate is dependent on the preferred language, exploratory data analysis of the dataset revealed that French and Spanish preferring users had a higher conversion rate on new landing page while English preferring users had a marginally poorer conversion rate on the new landing page.

3) Do not offer users anymore the option to use the old landing page since it doesn't increase time spent.

4) Language preference has no effect on time spent on the landing page and conversion rate. This needs to be communicated to the design team as it appears even though mean time spent by English preferring users on the new landing page is higher than old landing page, and similar to the Spanish and French preferring users, conversion rate was marginally poorer on the new landing page. It requires further data gathering.