E-news Express Business Presentation

Objective

Do statistical analysis and extract actionable insights from the data

We majorly focused on these problems -

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

*Consider a significance level of 0.05 for all tests.

The idea behind answering these questions is to decide whether the new page is effective enough to gather new subscribers for the news portal. We will perform the statistical analysis on the collected data to make the business decision.

Data Information

Variable	Description
user_id	This represents the user ID of the customer visiting the website.
group	This represents whether the customer 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 customer on the landing page.
converted	This represents whether the customer converted or not.
language_preferred	This represents the language chosen by the customer to view the landing page.

Observations	Variables
100	6

Note:

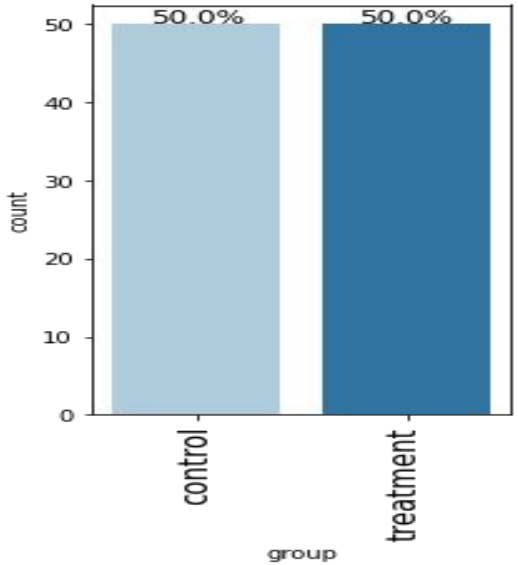
- There are no missing values in the dataset.
- The group, landing_page, converted, and language_preferred columns have been converted to category

Univariate Analysis – Time spent on the page

- The time spent on the page looks to have a fairly normal distribution.
- The mean time spent on the landing page is approximately 5 minutes.
- The median time spent on the landing page is approximately 5 minutes.

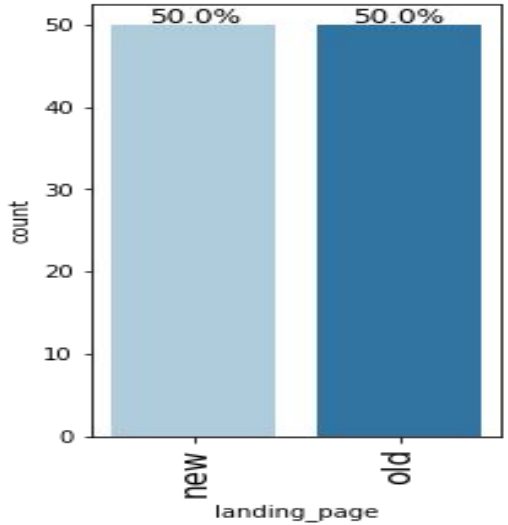
Univariate Analysis – Group & Landing page

Group



- There are 2 unique groups - control and treatment.
- The distribution of observations across groups are the same.

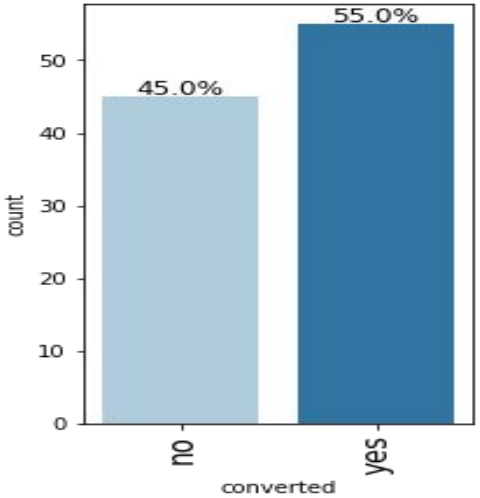
Landing page



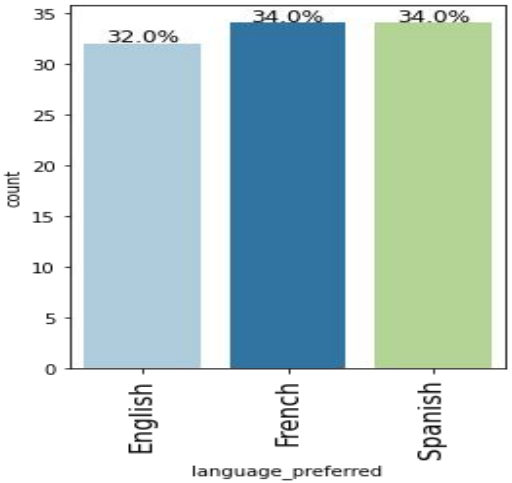
- There are 2 landing_pages - new and old.
- The distribution of observations across the two landing pages are same.

Univariate Analysis – Converted status & Preferred lang.

Converted status



Preferred language

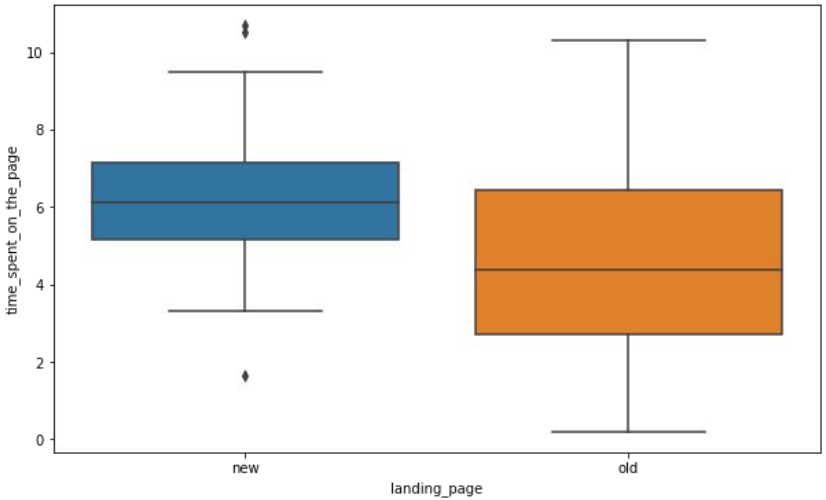


- 55% of the users get converted after visiting the landing page.
- 45% of the users do not get converted after visiting the landing page.

- There are 3 unique preferred languages - English, French, and Spanish.
- The distribution of observations across various preferred languages is fairly uniform.

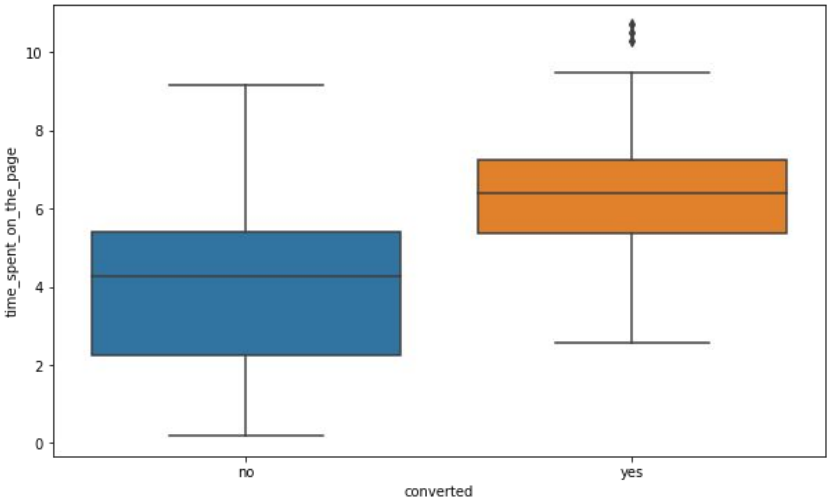
Bivariate Analysis

Landing page vs Time spent on the page



- Overall, the time spent on the new page seems to be greater than the time spent on the old page.

Converted status vs Time spent on the page



- Overall, the users who get converted seem to spend more time on the page.

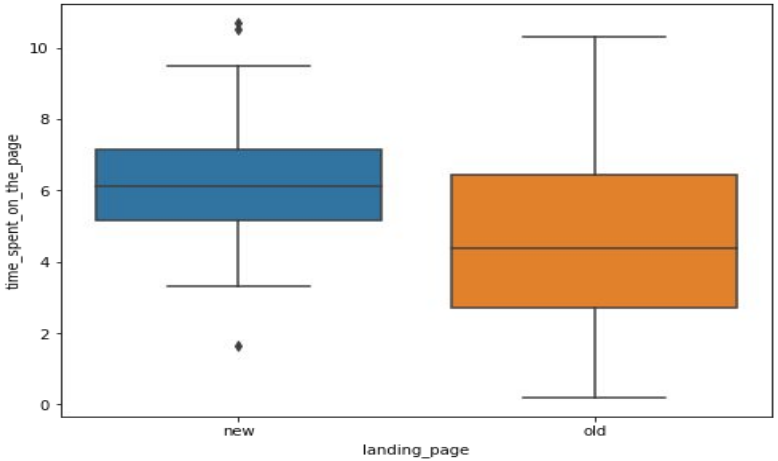
Hypothesis Testing - Time spent across new and old page

Problem: Do the users spend more time on the new landing page than the existing landing page?

- H_o: The mean time spent by the users on the new page is equal to the mean time spent by the users on the old page.
- H_a: The mean time spent by the users on the new page is greater than the mean time spent by the users on the old page.

By using the two samples independent t-test, we get the p-value as ~0.0001 that is < 0.05.

Therefore, we reject the null hypothesis that the mean time spent by the users on the new page is equal to the mean time spent by the users on the old page.



- Visually the difference between the time spent on the new page and old page is apparent. The time spent on the new page is comparatively greater than the time spent on the old page.

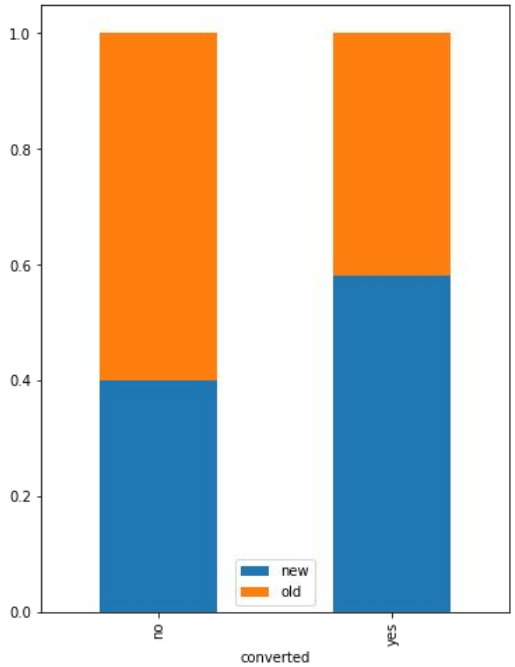
Hypothesis Testing - Conversion rate

Problem: 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?

- H_o: The conversion rate for the new page is equal to the conversion rate for the old page.
- H_a: The conversion rate for the new page is greater than the conversion rate for the old page.

By using the two sample proportions test, we get the p-value as ~0.0080 that is < 0.05.

Therefore, we reject the null hypothesis that the conversion rate for the new page is equal to the conversion rate for the old page.



- Visually, the number of users who get converted is more for the new page than the old page.

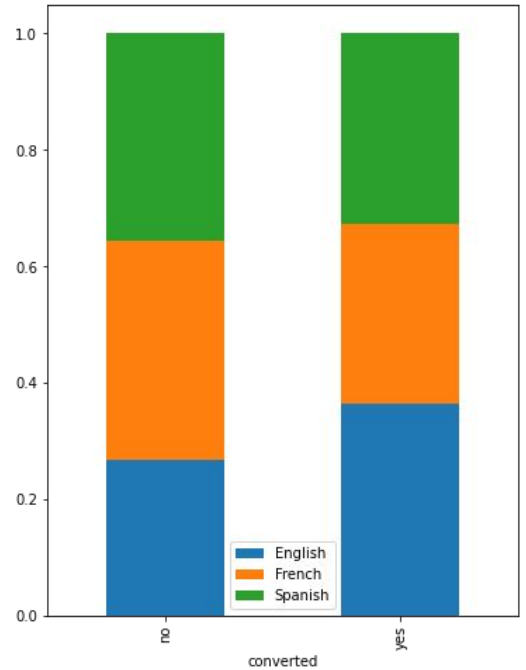
Hypothesis Testing - Converted status across languages

Problem: Is the conversion and preferred language are independent or related?

- H_o: The converted status is independent of the preferred language.
- H_a: The converted status is not independent of the preferred language.

By using the chi-square test of independence, we get the p-value as ~0.2130 that is > 0.05.

Therefore, we fail to reject the null hypothesis that the converted status is independent of the preferred language.



- The distribution of conversion status across different preferred languages looks fairly similar.

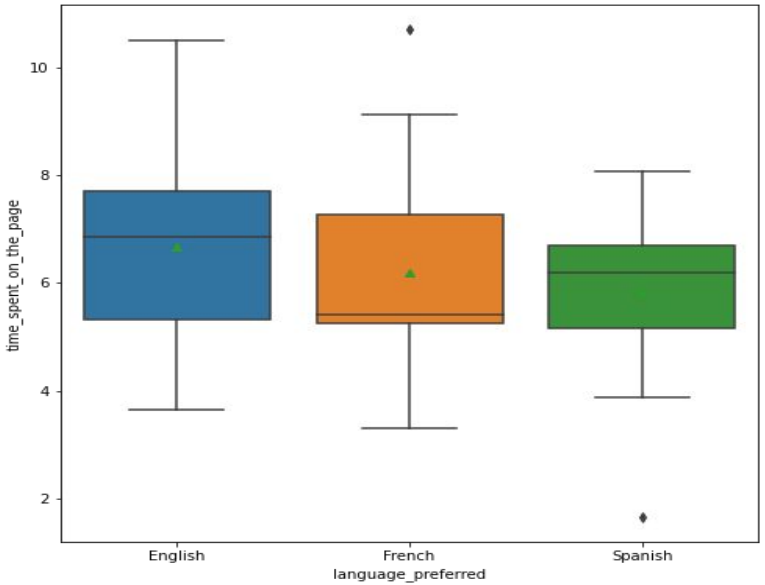
Hypothesis Testing - Time spent on the page

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

- H_o: The mean times spent on the new page by English, French, and Spanish users are equal.
- H_a: At least one of the mean times spent on the new page by English, French, and Spanish users is unequal.

By using the one-way ANOVA test, we get the p-value as ~0.4320 that is > 0.05.

Therefore, we fail to reject the null hypothesis that the mean times spent on the new page by English, French, and Spanish users are equal..



- The time spent on the new page by English, French, and Spanish users does not look drastically different. but, we need to test if this difference is statistically significant or not.

Conclusion

Based on our previous analysis, we can conclude that:

- The users spend more time on the new page compared to the old page.
- The conversion rate for the new page is greater than the conversion rate of the old page.
- The conversion status is independent of the preferred language.
- Based on the conclusions of the hypothesis tests, you can recommend that the news company should use the new landing page to gather more subscribers.
- The longer a visitor spends on a site, the more likely he/she is to convert. So, the business logic would appear to be: design a page that people spend time in, conversion will follow.

greatlearning
Power Ahead

Happy Learning !