

E-news Express Project

Contents

1. Business Problem Overview and Solution Approach

- ☐ Background
- ☐ Objective

2. Data Overview

- ☐ Data Description
- ☐ Significant manipulations made to raw data

3. EDA

- ☐ Univariate Analysis
- ☐ Bivariate Analysis
- ☐ Multivariate Analysis

4. Key Questions Answered through Hypothesis Testing

5. Business Insights and Recommendations

- ☐ Conclusion
- ☐ Business Recommendations

Business Problem Overview and Solution Approach

- **Background:**

An online news portal aims to expand its business by acquiring new subscribers. Every visitor to the website takes certain actions based on their interest. The company plans to analyze these interests and wants to determine whether a new feature will be effective or not. Companies often analyze users' 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.

I am hired as a Data Scientist in E-news Express. The design team of the company has created a new landing page. I have been assigned the task to decide whether the new landing page is more effective to gather new subscribers. 100 users are randomly selected 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).

- **Objective:**

Explore the dataset and extract insights from the data. Perform the statistical analysis of business data 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?
- Is the mean time spent on the new page same for the different language users?

Data Overview

Data Description:

- 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 | Variables | Duration |
|--------------|-----------|----------|
| 100 | 6 | 1 month |

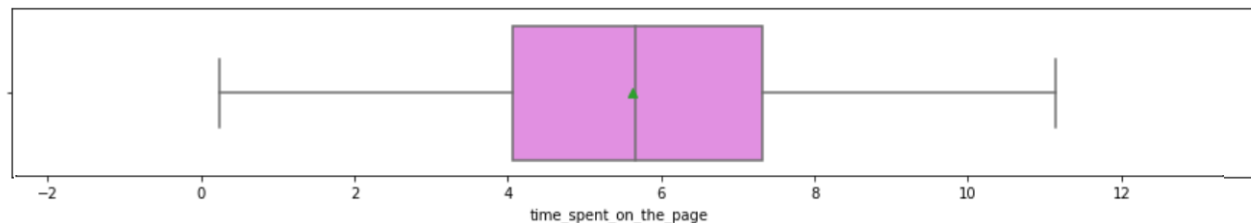
Significant manipulations made to raw data

- All columns have 100 observations. Hence there are no missing values.
- group, landing_page, converted, language_preferred are object datatype i.e String data type in Python.
- user_id is a Numerical variables of type int.
- time_spent_on_the_page is a float datatype variable.
- 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.
- group, landing_page, converted, language_preferred should be changed to Categorical variables as they have limited set of values.

EDA

- **Univariate Analysis of Numerical Variables.**

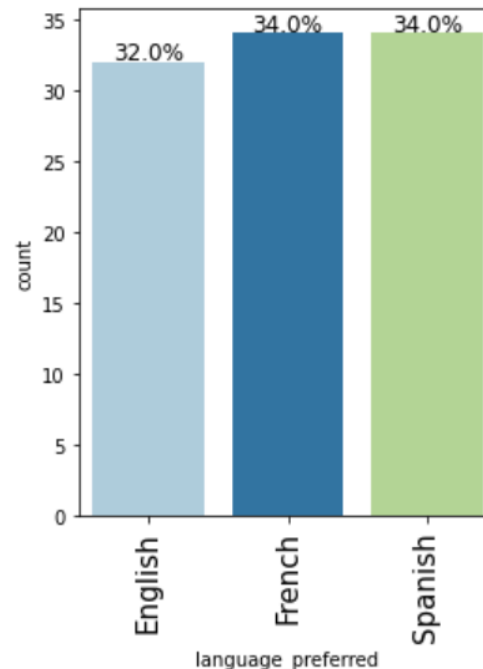
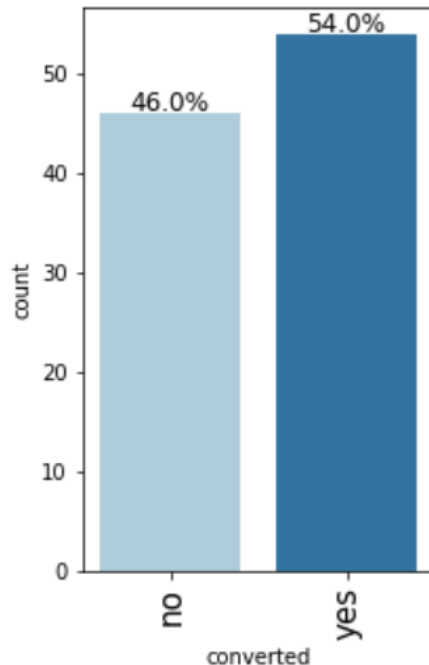
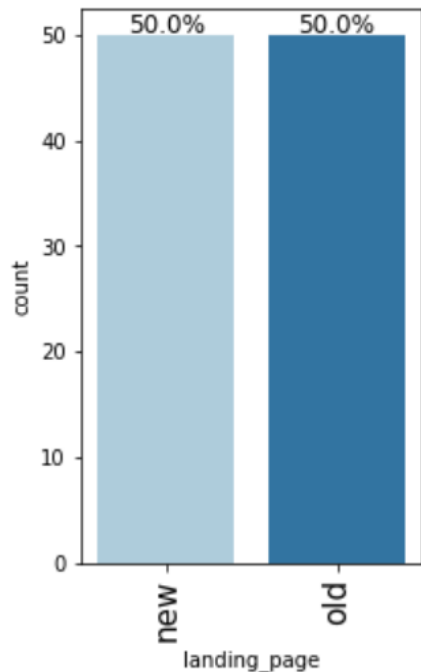
time_spent_on_the_page



Observations:

- It is an almost symmetrical graph with mean and median having almost equal values of 5.38 and 5.41 minutes respectively.
- Mode is around 0.2 to 0.3 minutes.
- Minimum time spent is 0.19 minutes and maximum is 10.71 minutes.
- Very few users have spent time less than 3 minutes and more than 9 minutes.

■ Univariate Analysis of Categorical Variables.



Observations:

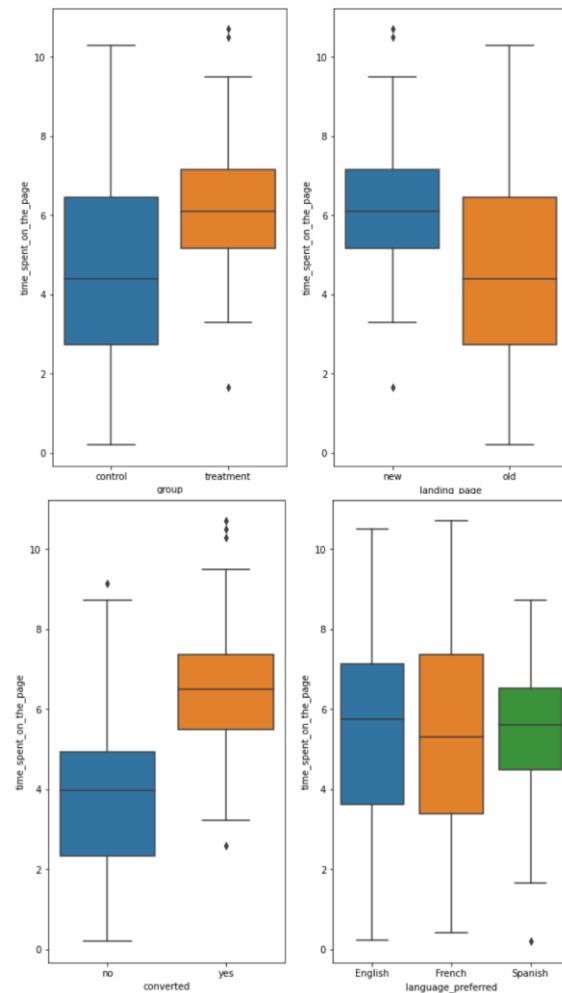
- * 50 users have used the old landing page and 50 users have used the new landing page.
- * More number of users which is 54 out of 100 were converted as subscriber of the news portal.
- * 34 out of 100 users prefer French language and another 34 users prefer Spanish language. The remaining 32 users prefer English language.

Bivariate Analysis

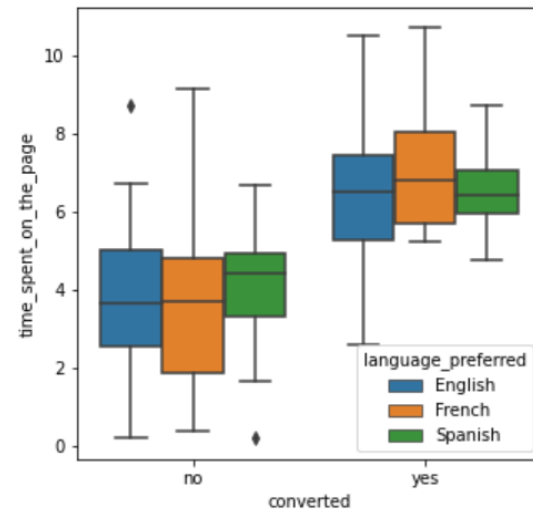
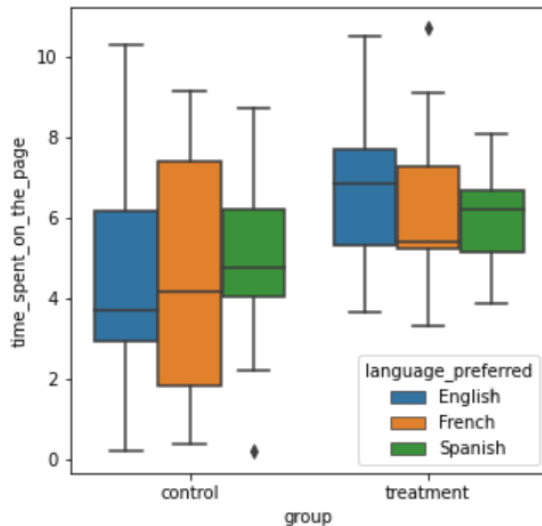
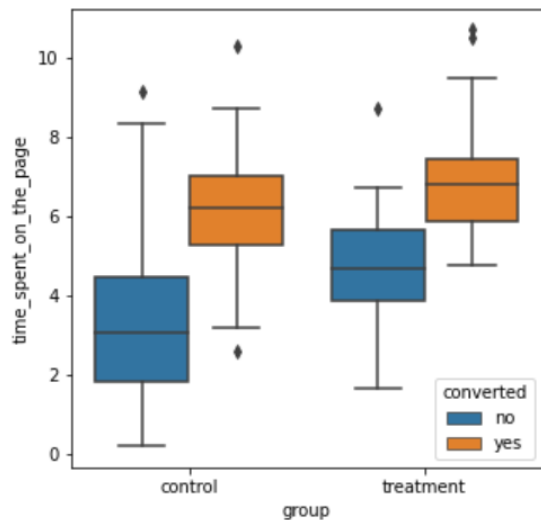
Relationship of time_spent_on_the_page with Categorical Variables

Observations:

- Most of the users in the control group spent 3 to 6.5 minutes on the page. Whereas most of the users in the treatment group spent 5 to 7 minutes on the page.
- The median time spent on the page by the treatment group is around 6 minutes whereas for the control group the median time is only 4.5 minutes.
- There are some outliers for the treatment group with time exceeding 10 minutes and also with time less than 3 minutes.
- The data for landing_page has the same results as the data for group. With users spending more time on the new page (median 6) than the old page (median 4.5)
- The users who did not convert to subscribers spent mostly 2 to 5 minutes with median 4 minutes.
- The users who got converted to subscribers spent mostly 5 to 7 minutes with median 6.5 minutes.
- There are few outliers for the converted users with time spent more than 10 minutes.
- For users who prefer English, median time spent is 5.75 minutes.
- For users who prefer French, median time spent is 5.25 minutes.
- For users who prefer Spanish, median time spent is 5 minutes.



Multivariate Analysis



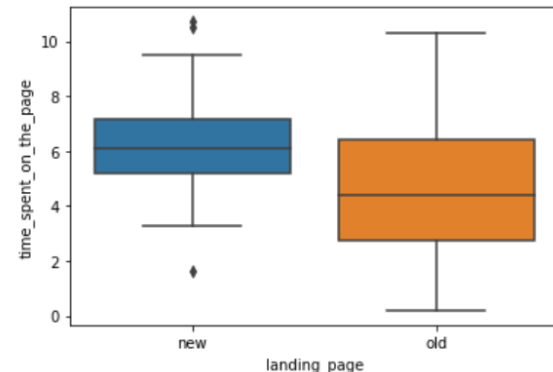
Observations

- Users from the treatment group have spent more time on the page than users in the control group.
- In both the groups, users who are converted have spent more time than who were not converted.
- For all three languages, time spent by users in the treatment group is more than the time spent by them in the control group.
- Most Users who converted have spent more time on the page than users who did not convert.
- For all three languages, time spent by users who are converted is more than the time spent by those not converted.

Key Questions Answered through Hypothesis Testing

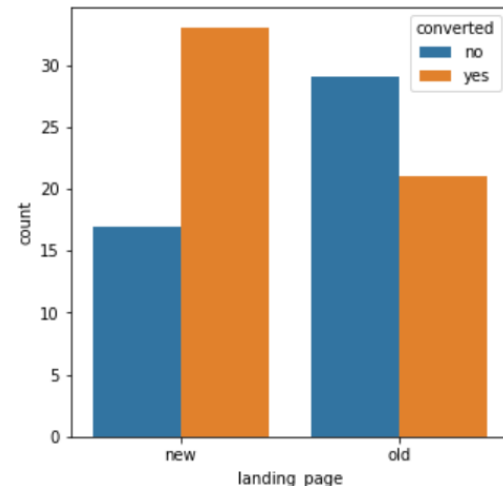
1. Do the users spend more time on the new landing page than the old landing page?

- This is a case of a one-tailed test for testing the means from two Sample. Since the standard deviation is known, we choose Two Independent Sample Z-test.
- As the p-value $7.634426865454902e-05$ is much less than the level of significance 0.05, we reject the null hypothesis. Thus, we have enough statistical evidence to say that users spend more time on the new landing page as compared to the old landing page.



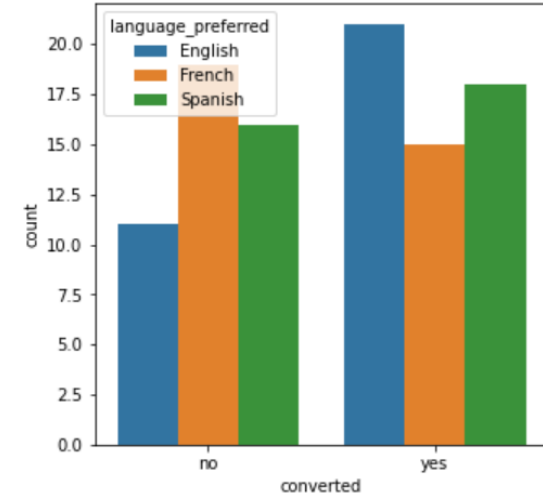
2. Is the conversion rate (the proportion of users who visit the landing page and get converted) for the new page greater than the conversion rate for the old page?

- As we need to compare the proportions from two independent samples, we use the Two Proportion Z-test.
- As the p-value 0.016052616408112556 is less than the level of significance 0.05, we reject the null hypothesis. Thus, we have enough statistical evidence to say that the conversion rate for the new page is greater than the conversion rate for the old page.



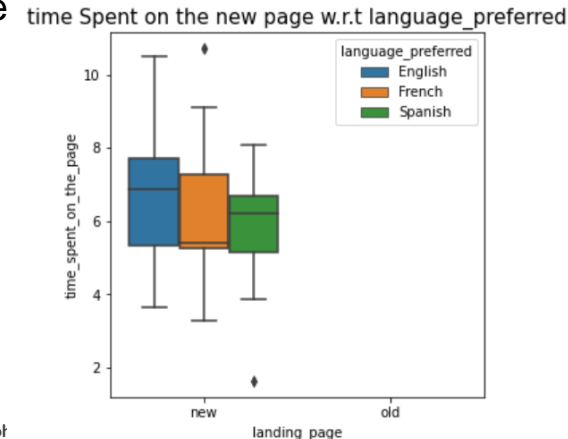
3. Does the converted status depend on the preferred language?

- As we need to Check whether the categorical variables from a population are independent, we use the Chi-Square Test for Independence.
- Since p-value of the test 0.21298887487543447 is large, we fail to reject the null hypothesis. Hence we conclude that the converted status is independent of the preferred language.



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

- As there are more than 2 independent samples, we use the One-way ANOVA Test provided normality and equality of variance assumptions are verified.
- For testing of normality, Shapiro-Wilk's test is applied to the response variable.
- For equality of variance, Levene test is applied to the response variable.
- Since the p-value 0.43204138694325955 is larger than the 5% significance level, we fail to reject the null hypothesis which says that the mean time spent on the new page is same for the different language users.



Business Insights and Recommendations

Conclusion :

We analyzed a dataset of 100 Users among which 50 users have spent time on the new landing page and 50 users have spent time on the old landing page. The main factors determining if the users get converted to subscribers are the mean time spent on the page and language preferred. We have been able to conclude that:

- Most of the Users who got converted to Subscribers have spent more time on the page than users who did not convert to subscribers.
- Users from the treatment group or the ones who used the new landing page have spent more time on the page than users in the control group or the old landing page.
- The above statement stands true even when we filter the users with their preferred language. Hence whichever language the user prefers, they spent more time on the page if they are using the new landing page.
- The highest Average time spent of 7.358 minutes is observed for users with English preference who are on the treatment group and got converted to subscriber.
- The lowest Average time spent of 1.776 minutes is observed for users with English preference who are on the control group and did not get convert to subscriber.

- The highest difference in means is the one we got for English users. Hence, customers who prefer English seem to like the new landing_page more than the old landing_page.
- Users who prefer English are the majority of users who got converted.
- Users who prefer Spanish are the majority of users who did not convert.
- In the control group, majority of the users which is around 29 out of 50 did not convert to subscribers.
- In the treatment group, majority of the users which is around 33 out of 50 converted to subscribers.
- User_id does not have any effects on the analysis.

Recommendation to business

- English Users seem to like the New landing page more than the Spanish and French users. Hence, the new landing page should be made more friendly for the Spanish and French users as well.
- The new landing page definitely converts more users to subscribers. So plans should be made to improve the new landing page such as news from the current day but many years ago, attractive headlines etc. to attract more users and convert them to subscribers.
- More attractive and exciting contents such as word puzzle games that provide prizes for the winners, challenges, trivia etc. to be added on the new page that can attract users and they spent more time on the page. As more time spent is proven to convert more users to subscribers.

greatlearning
Power Ahead

Happy Learning !

