A/B Testing for increasing subscribers

E-news Express: Business Statistics

Contents / Agenda

- Executive Summary

- Business Problem Overview and Solution Approach

- EDA Results

- Hypotheses Tested and Results

- Appendix

Executive Summary

- Users spending more time on the new page indicates that the outline & recommended content of the new page is more likely to keep customers engaged long enough to make a decision to subscribe

- The conversion rate for the new page is greater than the conversion rate of the old page, indicating that the new page is more likely to gather new subscribers than the existing page

- The conversion status is independent of the preferred language

- The time spent on the new page does not differ with content in different languages, indicating that irrespective of the language, the outline & recommended content of the new page are engaging

- It is recommended that the news company uses the new landing page to gather more subscribers as the tests conducted support the business logic to design a page that people spend time in, conversion will follow

# Business Problem Overview and Solution Approach

- E-news Express is an online news portal that aims to expand its business by acquiring new subscribers based on their analysis of user actions to understand user interests and determine how to drive better engagement

- The executives at E-news Express are of the opinion that there has been a decline in new monthly subscribers compared to the past year because the current webpage is not designed well enough in terms of the outline & recommended content to keep customers engaged long enough to make a decision to subscribe

- 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 Data Science team experimented by displaying the old and the new landing pages to two sets of users and recording their activity. They want to test the effectiveness of the new landing page based on the data collected from the experiment.

- The task is to explore the data and perform a statistical analysis (at 5% significance) to determine the effectiveness of the new landing page in gathering new subscribers for the news portal by testing certain hypotheses

EDA Results – Time spent on the page
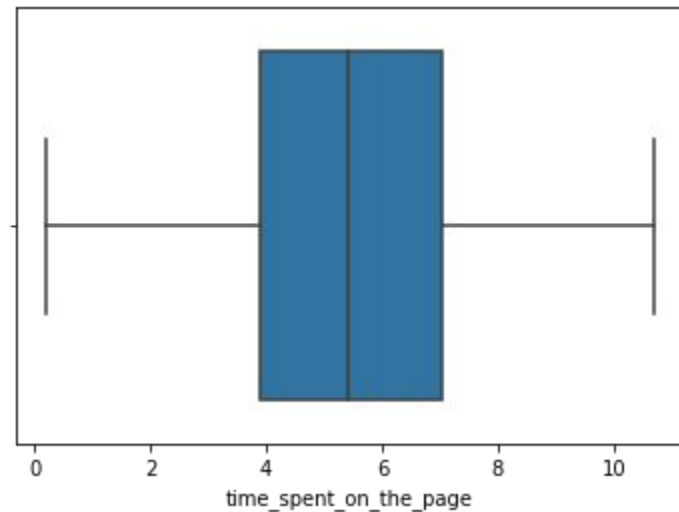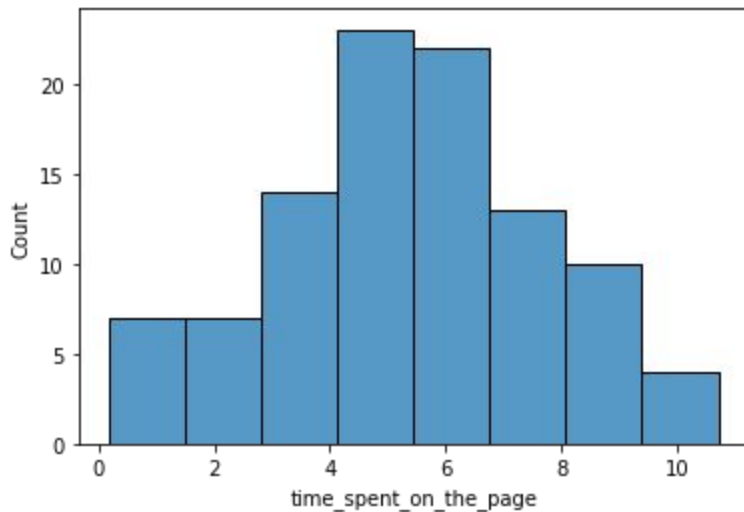
- The time spent on the page seems to have a fairly normal distribution.
- There are no outliers in this column.
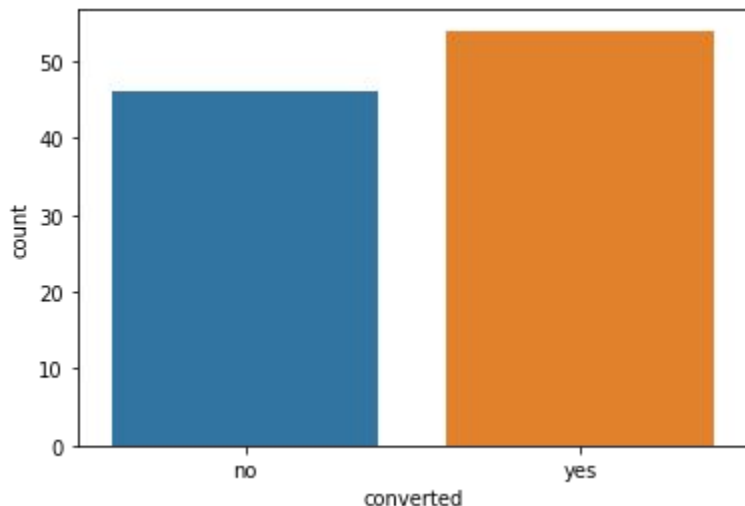
EDA Results – Group & Landing page
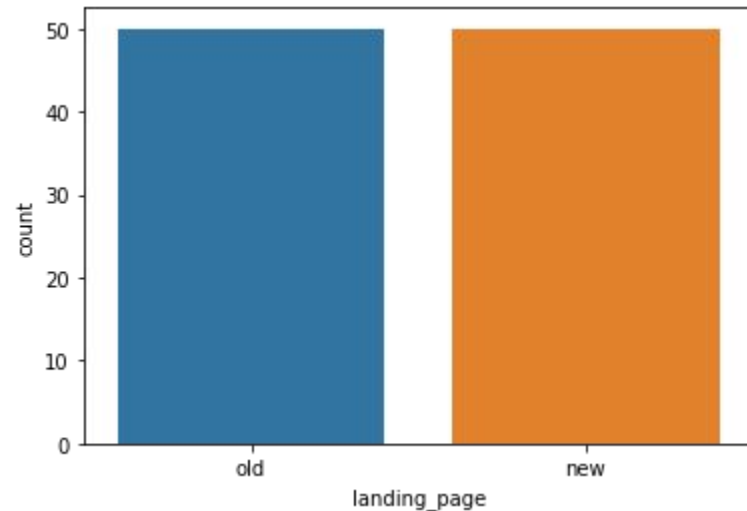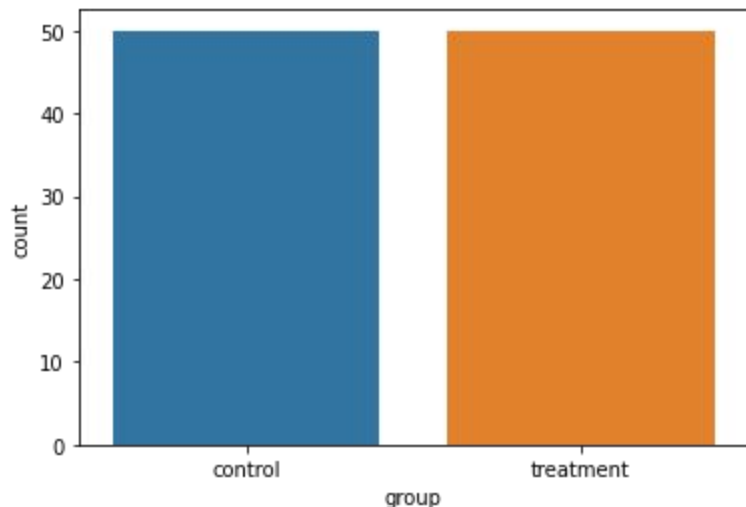
- There are 2 unique groups - control and treatment.
- The distribution of the number of users across the two groups are the same

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

# EDA Results – Converted status & Preferred language

- Overall, 54 users get converted and 46 users do not get converted after visiting the landing page
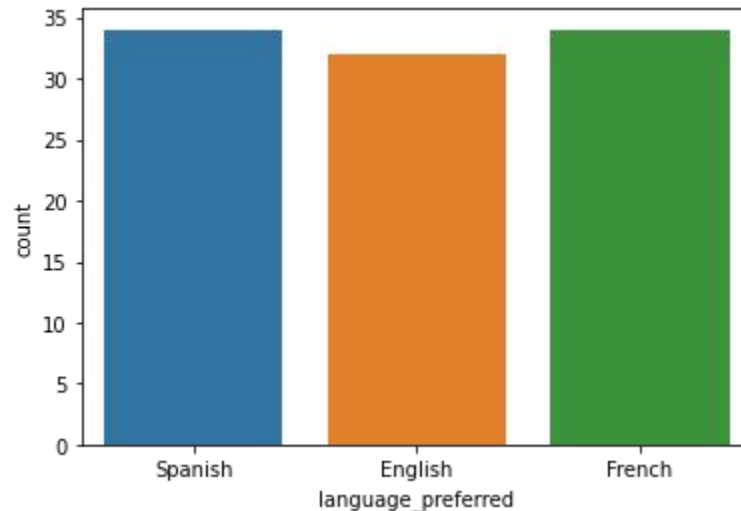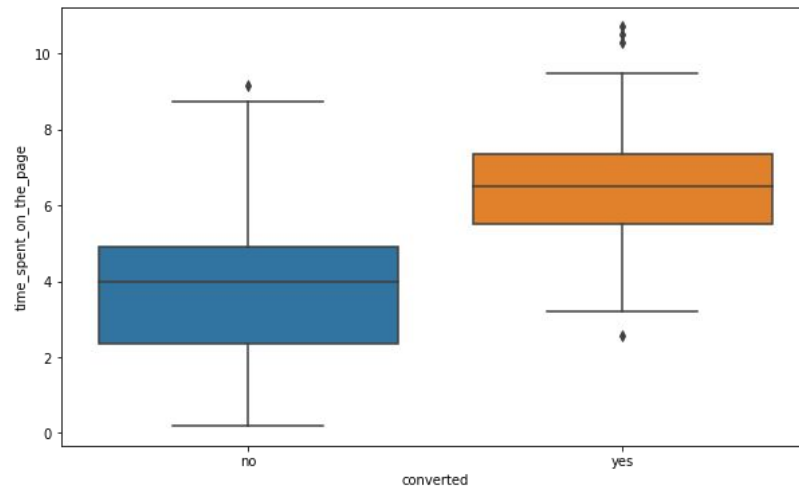
- The distribution of the number of users across various preferred languages is fairly uniform
- There are 3 unique preferred languages - English, French, and Spanish.
  Spanish and French both have been equally and mostly preferred by users

EDA Results - Bivariate Analysis

Preferred language vs Time spent on the page
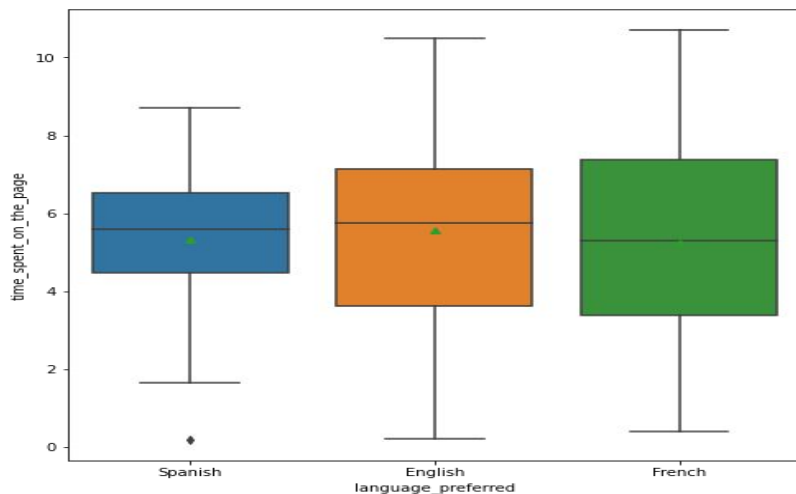
Converted status vs Time spent on the page

- The median time spent by the different language users are approximately equal.

- Overall, the users who get converted seem to spend more time on the page.
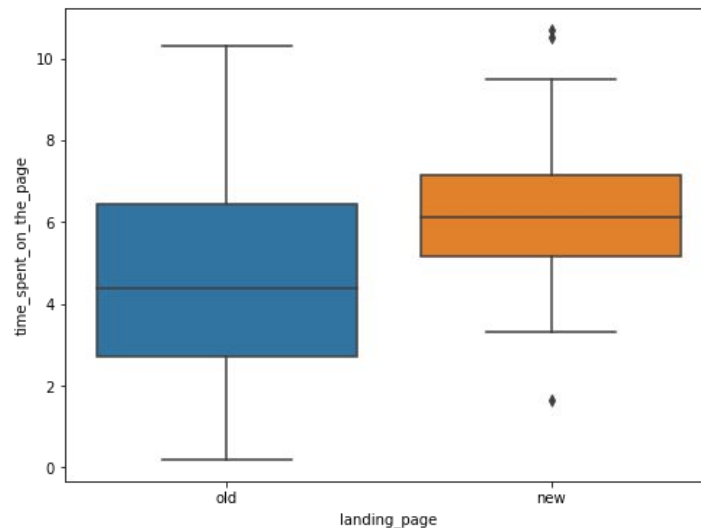
Time spent on old vs new landing page

- From the sample data, it is observed that the median time spent on the new landing page is higher than that on the old landing page

- The hypothesis that users spend more time on the new landing page than on the existing landing page was tested

- The test results showed enough statistical evidence to conclude that 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

*link to Appendix slide on details of the test performed*

Conversion rate of old vs new landing page

- From the sample data, it is observed that the number of users who get converted to subscribers (conversion rate) is more for the new page than the old page.

- The hypothesis that the conversion rate for the new page is greater than the old page was tested

- The test results showed enough statistical evidence to conclude that the conversion rate for the new page is greater than the conversion rate for the old page

*Link to Appendix slide on details of the test performed*
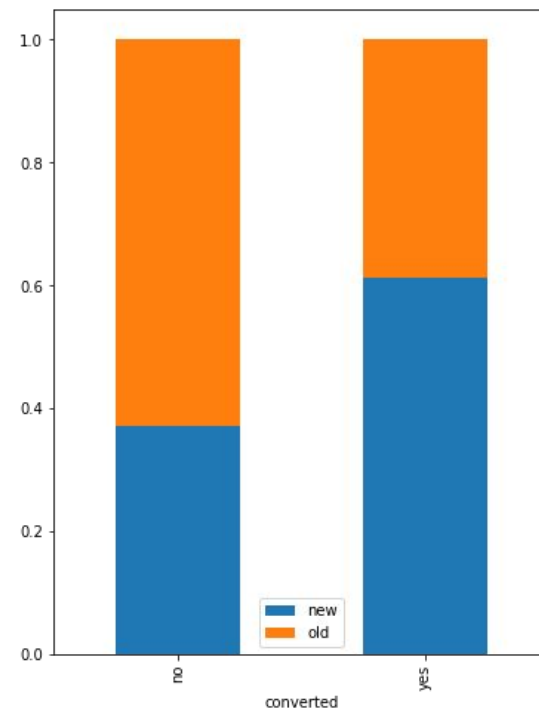
Dependence of conversion on preferred language

- From the sample data, it is observed that among the converted users, the count of English users is highest, and among the non-converted users, the count of French users is the highest.

- The hypothesis that the converted status is dependent on the preferred language was tested

- The test results showed that there isn't enough statistical evidence to say that the converted status depends on the preferred language

link to Appendix slide on details of the test performed
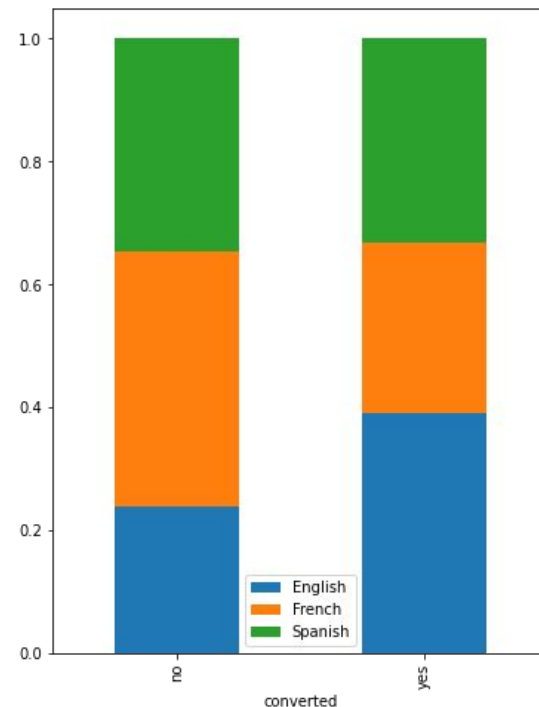
Time spent on new landing page across different languages

- From the sample data, it is observed that the median time spent on the new page by English users is a bit higher than that by French and Spanish users.

- The hypothesis that at least one of the mean times spent on the new page by English, French, and Spanish users is unequal

- The test results showed that there isn't enough statistical evidence to say that at least one of the mean times spent on the new page by English, French and Spanish users is unequal

link to Appendix slide on details of the test performed
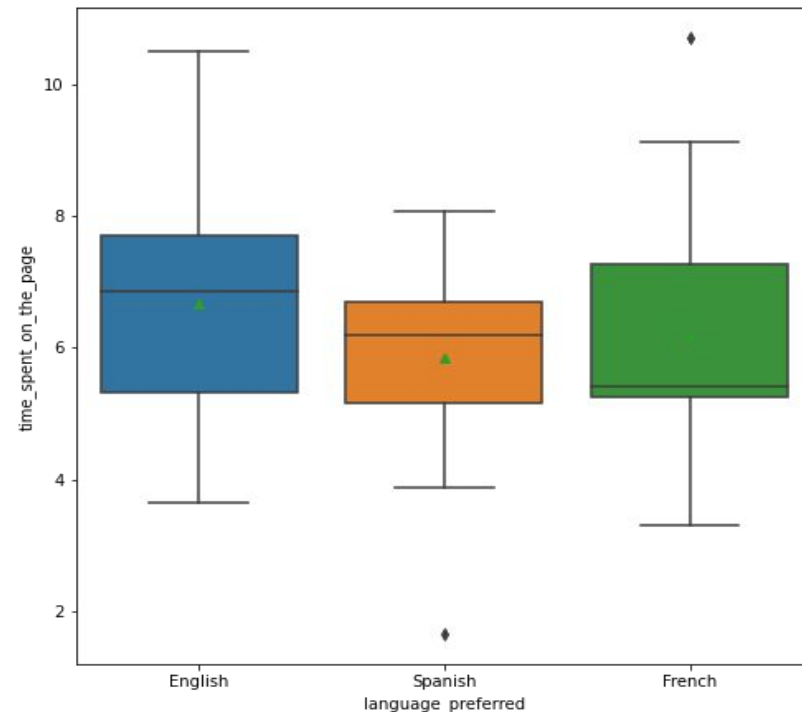
APPENDIX

Data Background and Contents

- The dataset contains 100 observations and 6 attributes

- 5 of the attributes (including user_id) are categorical in nature, while one is numerical

- There are no missing values in the dataset

- The statistical summary of the data shows us the following

  - The time spent by users on the landing page ranges approximately between 11 seconds to 11 minutes, with an average of 5.4 minutes

  - Variables *'group'*, *'landing_page'* and *'converted'* have only two categories

  - More than 50% of users are converted

  - There are three categories of languages preferred by users, among which Spanish has been preferred the most by users.

Hypothesis Testing Details

- The null hypothesis

  $H_0: \mu_1 = \mu_2$

  was tested against the alternative hypothesis

  $H_a: \mu_1 > \mu_2$

  where μ_1 and μ_2 denote the mean time spent by the users on the new and the old landing pages respectively

- As we have to test for the equality of means with unequal and unknown standard deviations from two independent populations, a two-sample independent t-test was performed

- A p-value of 0.0001 was obtained, which is lower than the significance level of 0.05, so the null hypothesis was rejected

Hypothesis Testing Details

- The null hypothesis

$$H_0: p_1 = p_2$$

was tested against the alternative hypothesis

$$H_a: p_1 > p_2$$

where p_1 and p_2 denote the conversion rate for the new and old page respectively

- As we have to test for the equality of proportions from two independent populations, a two-sample proportion z-test was performed

- A p-value of 0.0080 was obtained, which is lower than the significance level of 0.05. So, the null hypothesis was rejected

Hypothesis Testing Details

- The null hypothesis

  H_0: The converted status is independent of the preferred language.

  was tested against the alternative hypothesis

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

- As we have to test the independence between two categorical variables, a chi-square test for independence was performed

- A p-value of 0.2130 was obtained, which is higher than the significance level of 0.05 So, we fail to reject the null hypothesis

Hypothesis Testing Details

- The null hypothesis

    H_0: The mean times spent on the new page by English, French, and Spanish users are equal.

was tested against the alternative hypothesis

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

- As we have to test the equality of means from three independent populations, a one-way ANOVA test was performed

- A p-value of 0.4320 was obtained, which is higher than the significance level of 0.05 So, we fail to reject the null hypothesis

Great Learning

Happy Learning !