

E-news Express - Case Study

Background

"E-News Express" is an online news portal aims to expand its business by acquiring new subscribers. To help with this goal the design team of the company has created a new landing page. Executives of the company are interested to know if this new landing page is more effective to gather new subscribers.

To help make this inference the Data Scientist team –

- Randomly selected 100 users, divided them in 2 equal groups.
- Served old landing page to control group and New landing page to treatment group.
- Gathered various data attributes of these users.
- They have performed the required data analysis to help answer the key business questions.

Objective

To explore and visualize the user's data gathered from the randomly chosen sample, extract insightful and/or actionable information and perform statistical analysis to help answer some of the key questions business has.

- **We will focus to find -**
 - User behavior in new landing page vs old page.
 - Perform statistical analysis on the sample data
 - Draw reasonable conclusion on the following key questions
 1. Do the users spend more time on the new landing page ?
 2. Has the conversion rate improved in new landing page?
 3. Does the converted status depend on the preferred language?
 4. Is the average time spent on the new page same for the different language users?

Data Overview

The gathered dataset contains various user attributes for control and treatment group of samples

Variable	Description
user_id	Represents the user ID of the person visiting the website
group	Represent whether user belong to control or treatment group
landing_page	Represent which landing page was served
time_spent_on_the_page	Represent the time spent on the page in minutes
converted	Represent user gets converted to a subscriber or not
language_preferred	Represent language chosen by the user to view the landing page

Observations	Variables
100	6

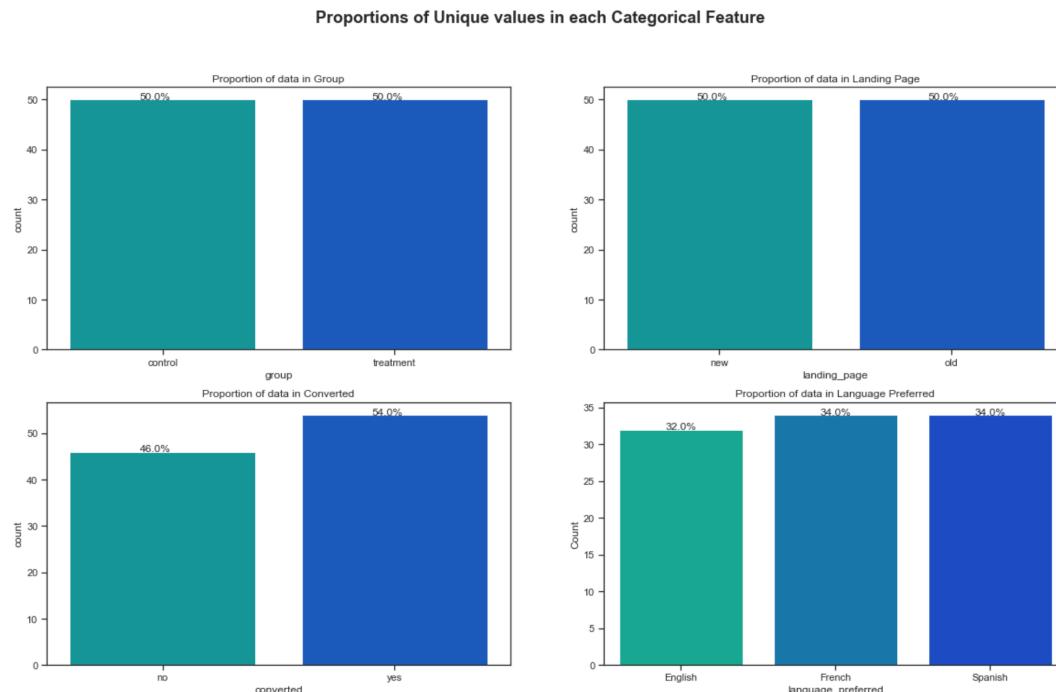
Notes:

1. Dataset looked clean and consistent with the data dictionary provided.
2. There were no missing values across all variables.

Exploratory Data Analysis – Univariate

Group, Landing Page, Converted status, Preferred Language

The data gathered for sample users has 4 categorical feature. Below is a snapshot of how their proportions are

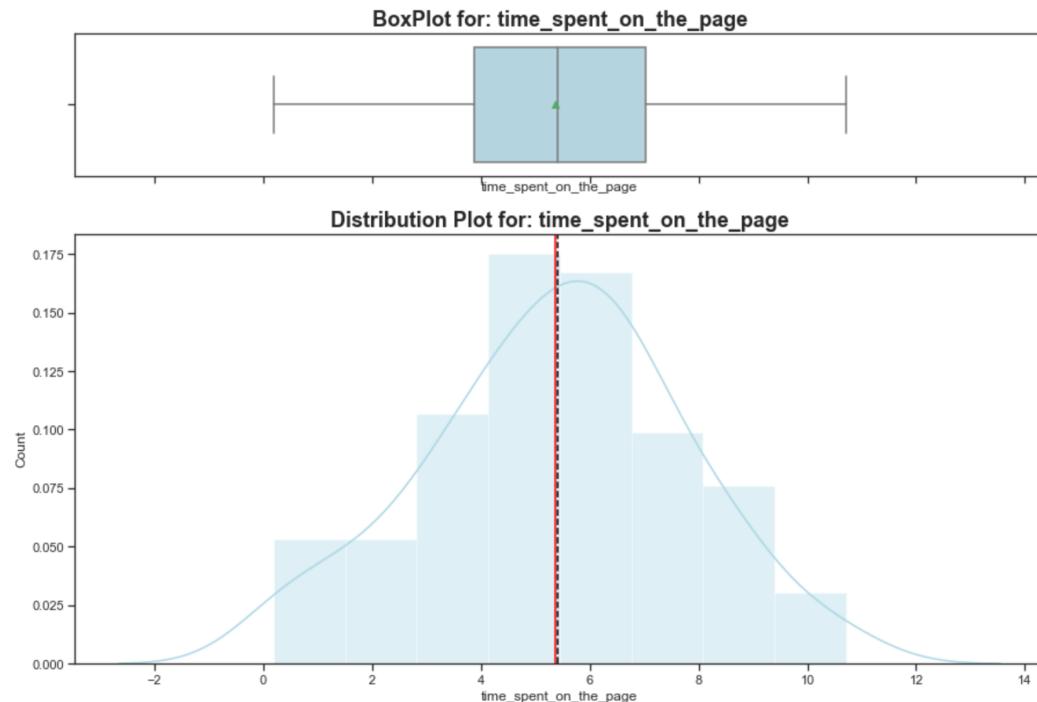


Observations:

- Proportions of groups and landing pages are inline with experiment performed.
- 54% of users from the sample converted to subscribers and 46% hasn't subscribed
- 34% of sample users chosen preferred language as French, so as Spanish, whereas 32% has chosen English

Exploratory Data Analysis – Time Spent on Page

Below snapshot provides the distribution of the time spent by users on the landing pages



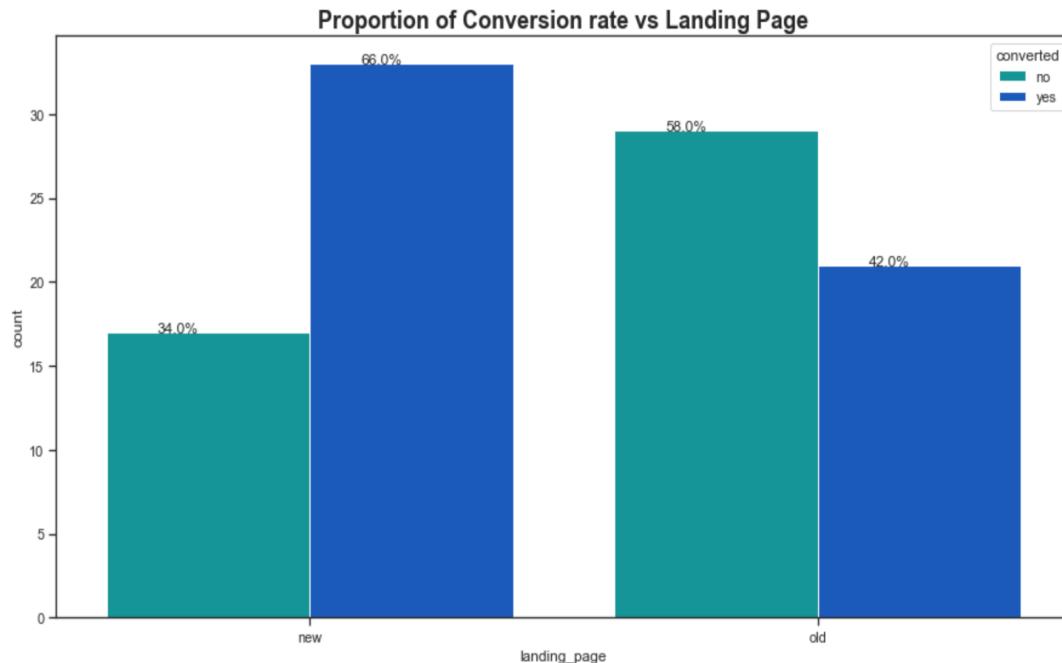
Observations:

- Users has spent time as little as less than 1 minute, and as high as ~11 minutes.
- Time spent for sample users tends to show sign of normally distributed. With sample size >30 we can consider the data normally distributed based on Central limit theorem
- No outliers can be observed

Exploratory Data Analysis – Bivariate

Converted status by Landing page

Below snapshot provides the proportions of converted status of the users on each the landing page

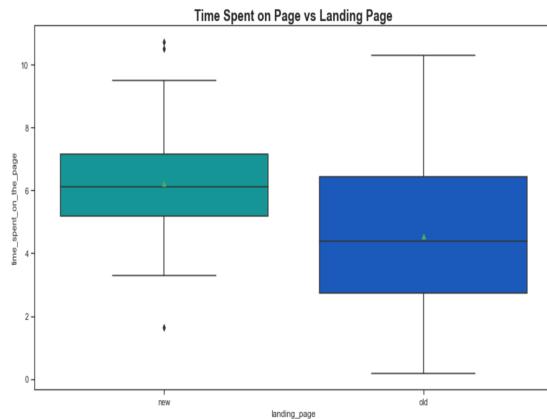


Observations:

- New landing page has 66% of conversion rate compared to old landing page which has 42%
- New landing page has brought down the non converted user proportions to 34% from 58% in old landing page
- Further analysis has been done to check if the conversion rate is significantly increased in new landing page

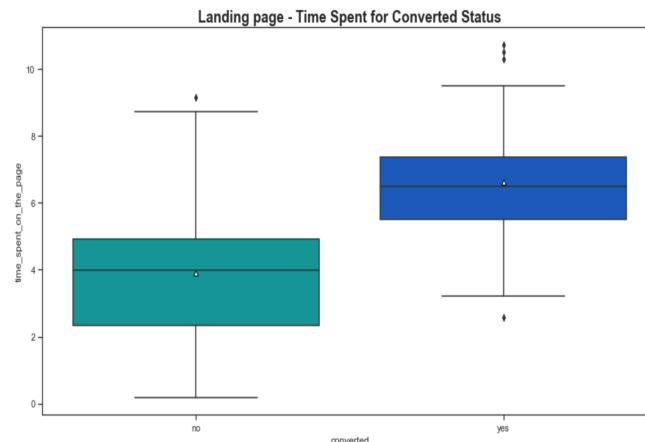
Exploratory Data Analysis –

Time Spent with respect to landing page, language preferred & converted status



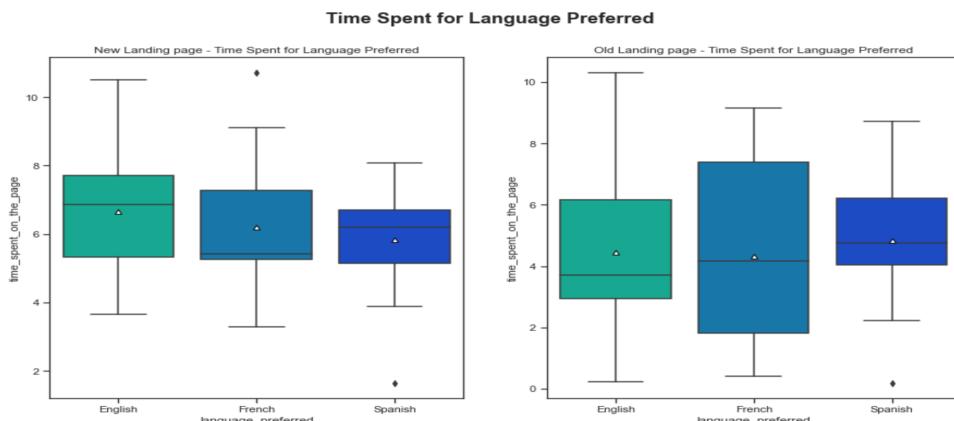
Observations:

- Average time spent by the users in new landing page is higher than the old landing page



Observations:

- Users who became subscribers spent more than ~2 mins in the landing page
- Average time spent is higher for the users became subscribers



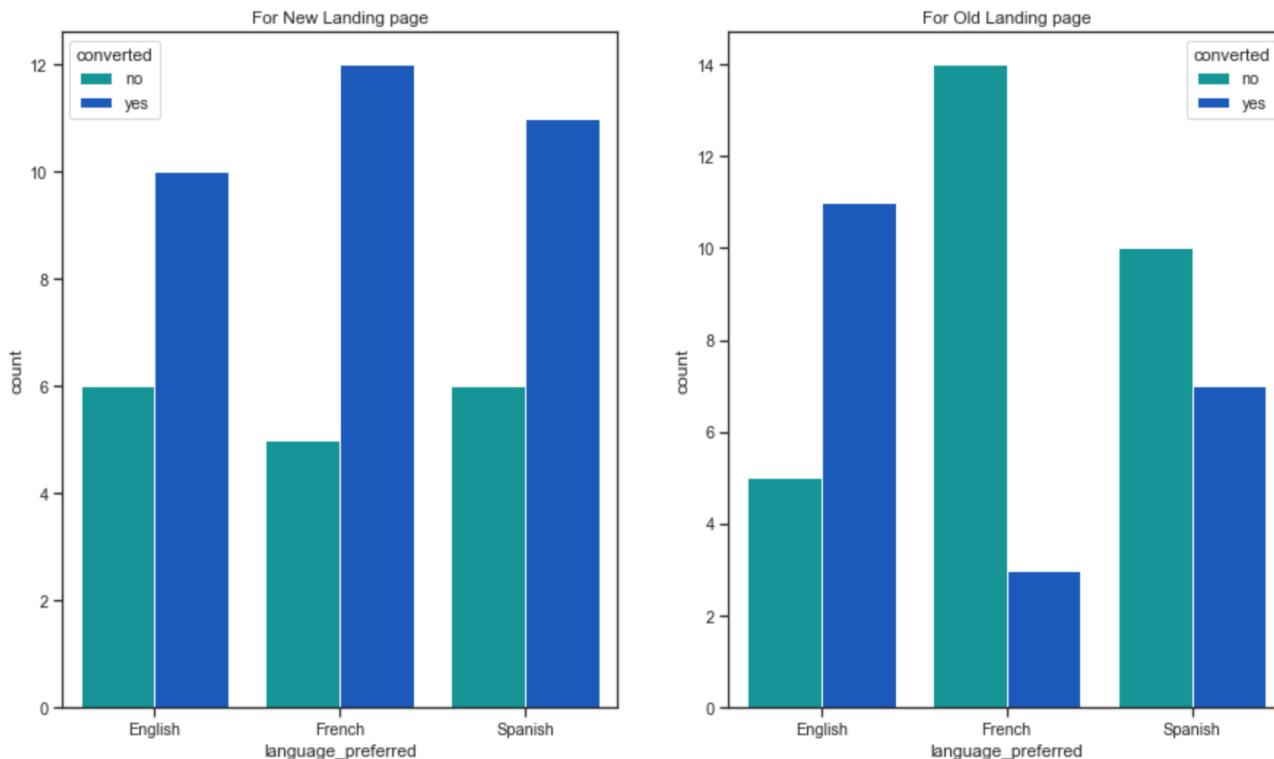
Observations:

- Average time spent on the new landing page by the subscribers are around ~6 minutes compared to old landing page of ~4 minutes
- In new landing page average time spent slightly differs based on the preferred language.
- Users with preferred language as English has higher average spent time

Exploratory Data Analysis –

Converted status based on Preferred language

Converted Status based on Language Preferred



Observations:

- Users of new landing page are subscribing more regardless of their language choice
- Except users with preferred language as English for old landing page, nonsubscribers counts are more than subscribers

Statistical Testing – Key Questions & Assumptions we were told

- **Key questions which we will answer -**

1. Do the users spend more time on the new landing page ?
2. Has the conversion rate improved in new landing page?
3. Does the converted status depend on the preferred language?
4. Is the average time spent on the new page same for the different language users?

- **Assumptions which were considered or given -**

1. The experiment we have performed, is based on randomly chosen 100 users, so the data present in dataset considered random sample.
2. Sample observations are independent of each other.
3. Data distribution follows Normal distribution
4. For all Hypothesis testing, significance level (α) has been considered as 5%

Statistical Testing – Answering Questions

- Do the users spend more time on the new landing page ?**

To answer this question, we have formed the Hypothesis as below –

Let μ_1, μ_2 be the mean time spent on page of New Landing Page and Old Landing Page groups, respectively.

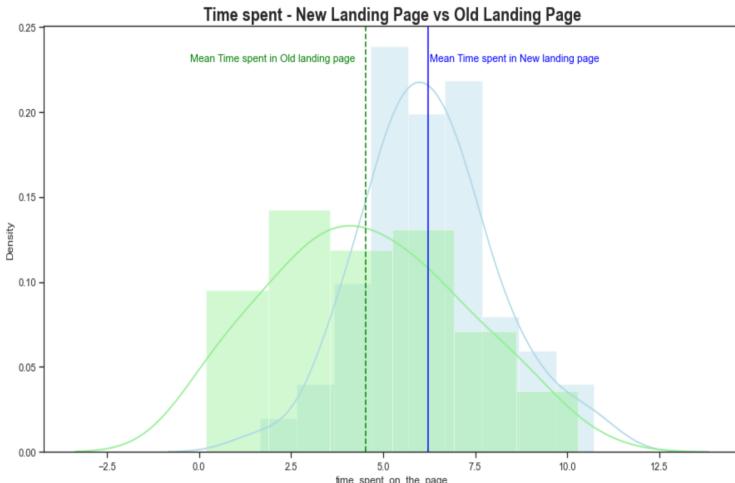
We will test the null hypothesis

$$H_0: \mu_1 = \mu_2$$

against the alternate hypothesis

$$H_a: \mu_1 > \mu_2$$

Visual Representation of Mean time difference



Test Performed

We have considered T-Test Independent to test the Hypothesis and following assumptions are validated –

- Time spent is captured on continuous scale
- We were told population data is normally distributed
- Samples are independent and from different groups
- Population standard deviations are different for these sample groups, concluded based on samples standard deviation
- Samples are random sample

We have taken p-value approach and calculated p-value is ~0.0001

Conclusion

With the p-value ~0.0001 being less than significance level 5%, we can reject the Null Hypothesis of the average time spent in new page being equal to old page. Hence, the –

Conclusion is.... **There is enough statistical evidence to confirm users are indeed spending more time in new landing page than old.**

Statistical Testing – Answering Questions contd..

- Is the conversion rate for the new page greater than the conversion rate for the old page?

To answer this question, we have formed the Hypothesis as below –

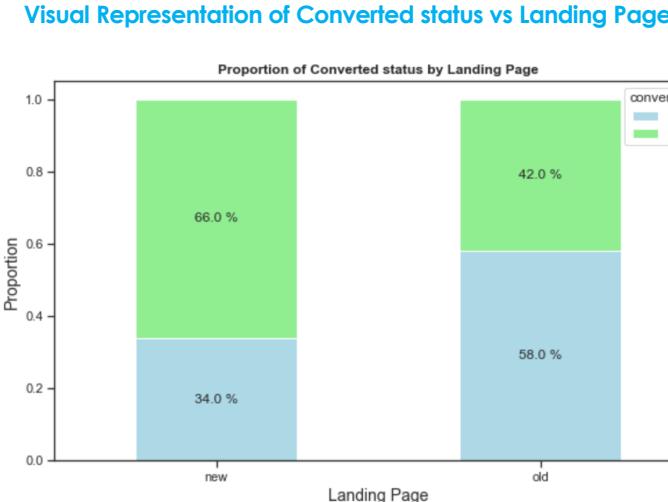
Let p_1, p_2 be the conversion rate of new landing page and old landing page respectively.

We will test the null hypothesis

$$H_0: p_1 = p_2$$

against the alternate hypothesis

$$H_a: p_1 > p_2$$



Test Performed

We have considered Proportions Z Test to test the Hypothesis and following assumptions are validated –

- Binomially distributed population, as converted status is either yes or no
- Samples are random sample
- Binomial distribution approximated to normal distribution, with standard check of sample(n) * proportion (p) and sample (n) * (1- proportion) should be ≥ 10
- For our samples $n_1p_1 = 33$, $n_1(1-p_1) = 17$, and $n_2p_2 = 21$, $n_2(1-p_2) = 19$ (n_1, p_1 new page sample and proportion, n_2, p_2 is for old page)

We have taken p-value approach and calculated p-value is ~ 0.008

Conclusion

With the p-value ~ 0.008 being less than significance level 5%, we can reject the Null Hypothesis of conversion rate of new page is equal to the conversion rate of old page. Hence, the –

Conclusion is.... **There is enough statistical significance to confirm new landing page has higher conversion rate than old page.**

Statistical Testing – Answering Questions contd..

Does the converted status depend on the preferred language?

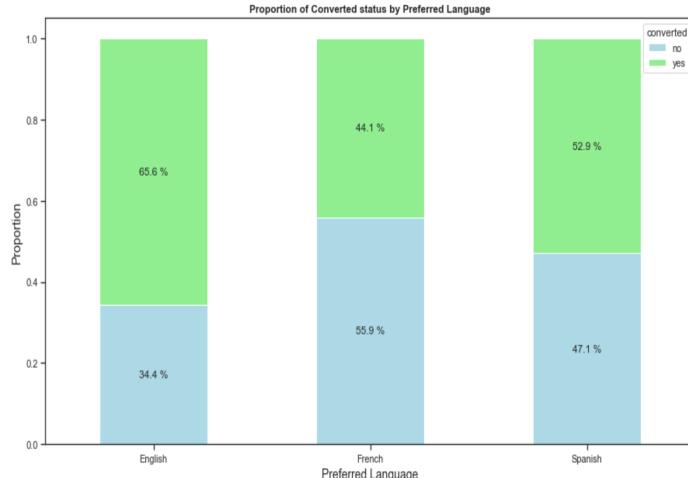
To answer this question, we have formed the Hypothesis as below –

We will test the null hypothesis

H_0 : Converted status is independent of preferred language.
against the alternate hypothesis

H_a : Converted status depends on preferred language.

Visual Representation of Converted status vs Preferred Language



Test Performed

We have considered Chi-square Independence Test to test the Hypothesis and following assumptions are validated –

- Conversion status and Preferred language are categorical variables
- Samples are random sample
- Number of observations in each level is greater than 5

We have taken p-value approach and calculated p-value is ~0.21

Conclusion

With the p-value ~0.21 being way greater than significance level 5%, we fail to reject the Null Hypothesis of converted status is independent of preferred language. Hence, the –

Conclusion is.... **There is not enough statistical evidence to confirm that converted status has any dependency on preferred language**

Statistical Testing – Answering Questions contd..

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

To answer this question, we have formed the Hypothesis as below –

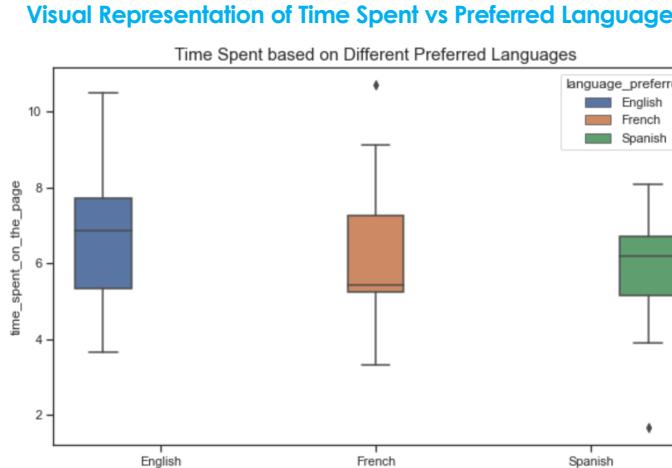
Let μ_1, μ_2, μ_3 be the means of time spent on the page for languages French, Spanish and English, respectively.

We will test the null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3$$

against the alternative hypothesis

H_a : At least one mean time spent is different from the rest.



Test Performed

We have considered One-way ANOVA Test to test the Hypothesis and following assumptions are validated –

- Normality test of population for new page has been validated using Shapiro Wilk's test
- Samples are independent and random sample
- Homogeneity of variance for new page population has been validated using Levene's test

We have taken p-value approach and calculated p-value is ~0.43

Conclusion

With the p-value ~0.43 being way greater than significance level 5%, we fail to reject the Null Hypothesis of average time spent on the new page for different language user is same. Hence, the –

→ Conclusion is.... **There is not enough statistical evidence to confirm that average time spent on the new page differs for different language users**

Conclusion

After all the analysis we have been able to conclude –

- New landing page significantly improved the conversion rate.
- Average usage of new landing page has been improved compared old landing page
- We were not able to conclude that preferred language has any dependency on the conversion rate, however the new page shows trends of French and Spanish users are also subscribing more compared to old page
- Average time spent on the new page doesn't differ based on user's preferred language.
- Subscribers spend more time on the page than nonsubscribers.

Recommendation

- Conversion rate is higher, new page should be rolled out to all users.
- Average usage has been increased in new page, ensure the momentum continues by adding more relevant and engaging contents.
- Even though the conversion rate has increased, there are still over ~30% of nonsubscribers in new page. A quick survey to gather reason of non subscribing and further analysis on that could point to further improvement leading to converting some of that population.
- French and Spanish users have started subscribing more in the new page, even though we couldn't conclude that preferred language has any impact on becoming subscribers, it must be ensured this momentum continues.
- Gathering more attributes such as content specific time spent, user's geo location and doing further analysis on them may lead to new findings of further improvement in conversion rate and usage.

greatlearning
Power Ahead

Happy Learning !

