E-news Express
Data Analysis

Contents

- Background
- Data Overview
- Exploratory Data Analysis
- Statistical Analysis
- Key Insights
- Business Recommendation

Background

- E-news Express is an online news portal aiming to expand its business by acquiring new subscribers.
- The company plans to analyze user’s interests and wants to determine whether a new feature will be effective or not.
- The design team of the company has created a new landing page.
- The task is 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).
- Various data about the customers in both groups are collected in 'abtest.csv'.

Objective

- Explore the dataset and extract insights from the data using Exploratory Data Analysis

- Perform the statistical analysis to determine:

  - Do the users spend more time on the new landing page than the old landing page?
  - Is the conversion rate 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

Variable	Description
user_id	user ID of the person visiting the website
group	first group (control) or the second group (treatment)
landing_page	new or old
time_spent_on_the_page	time (in minutes) spent by the user on the landing page
converted	whether the user gets converted to a subscriber or not.
language_preferred	language chosen by the user on landing page

Observations	Variables
100	6

Note:

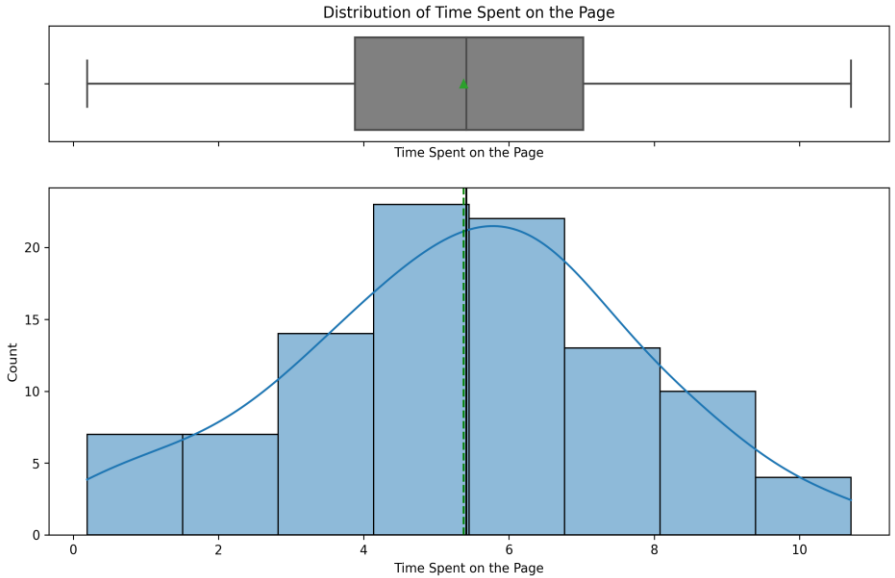
- There are no missing values in the dataset
- The variables group, landing_page, converted and language_preferred have been converted to category.

Exploratory Data Analysis (EDA)

Univariate Analysis – User id, Group & Landing page

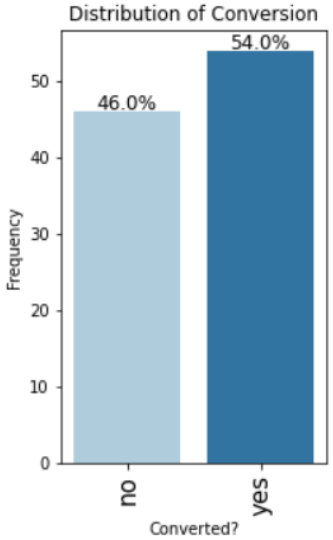
- There are 100 observations, user id is just an identifier
- There are no repeat IDs, we can say that each user is unique.
- The users are divided into two equal groups of 50 each – Control and treatment.
- The users are distributed between old and new landing page equally.

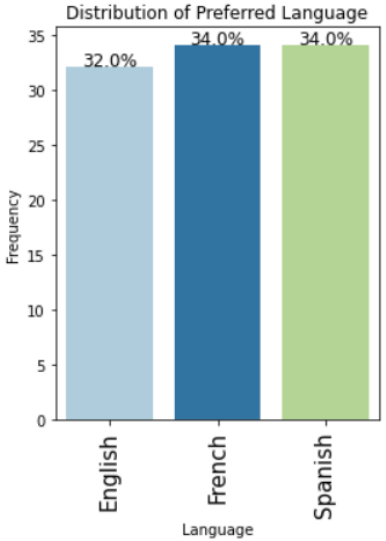
Univariate Analysis – Time spent on the landing page



- Mean and median values for this parameter are close together.
- The nature of the distribution describes the sample to be of normal distribution.
- We can assume that the population from which this sample came is also of normal distribution.

Univariate Analysis – Converted status & Preferred language

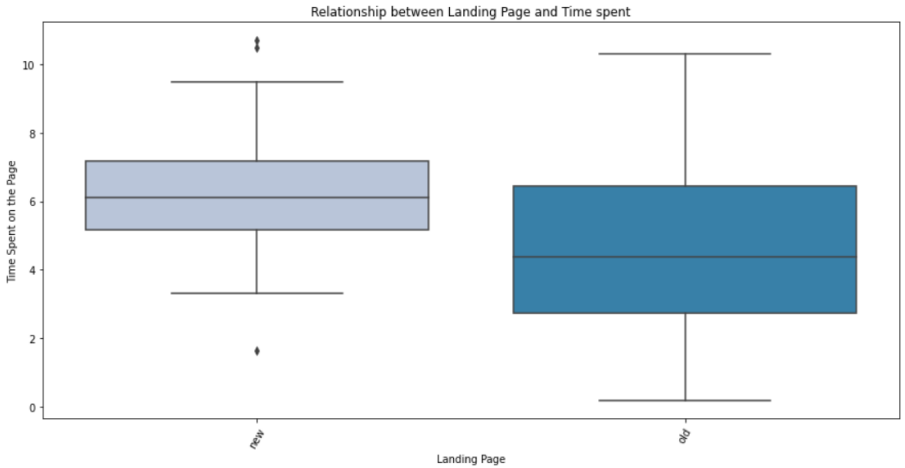




- The converted column consists of two unique values – yes and no.
- 54% of the total users chose to be subscribers after visiting the landing page irrespective of the page being old or new.
- There are equal number of users for French and Spanish.
- There is no notable difference in the number of users for the three languages that are being offered by the company.

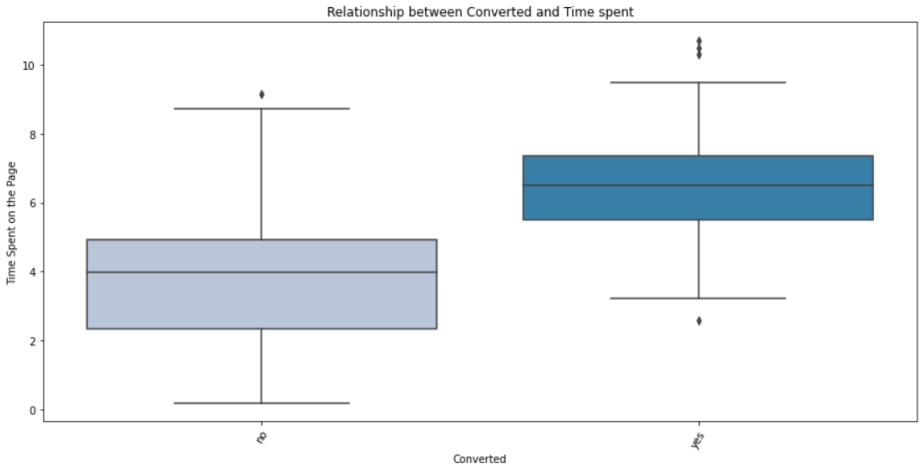
Bivariate Analysis

Landing page vs Time spent on the page



- Users spent more time on the new landing page than the old page.

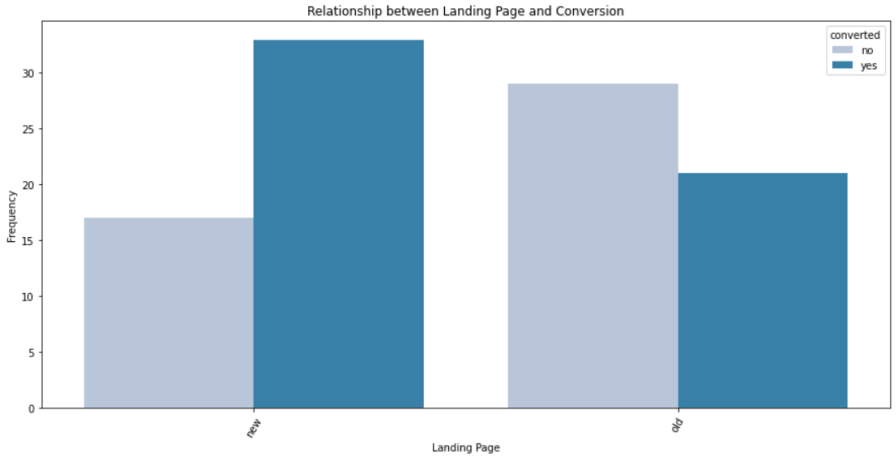
Converted vs Time spent on the page



- The average time spent by users who are converting themselves to subscribers are more than the ones who are not.

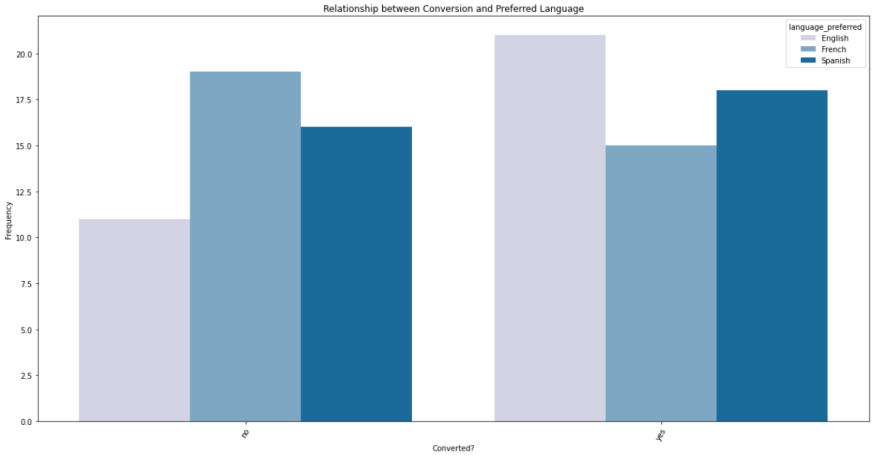
Bivariate Analysis

Landing page & Converted



- The new landing page has a higher conversion rate than the old landing page
- Most of the users that visited the new landing page were converted. However, this conversion is not reflecting for users landing on the old page.

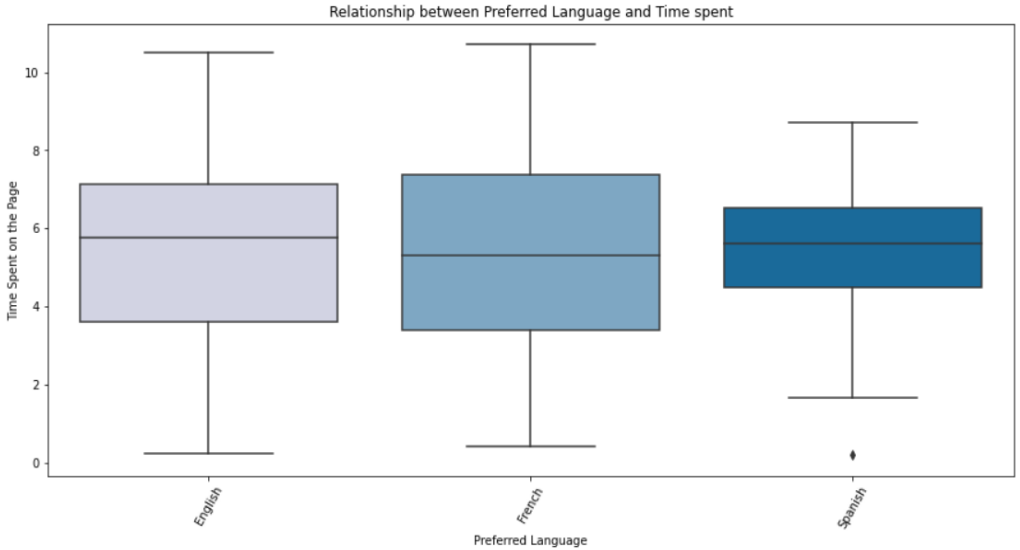
Converted & Preferred language



- Users preferring English language are the most converted ones among all the languages.
- French seem to have the least conversion rate

Bivariate Analysis

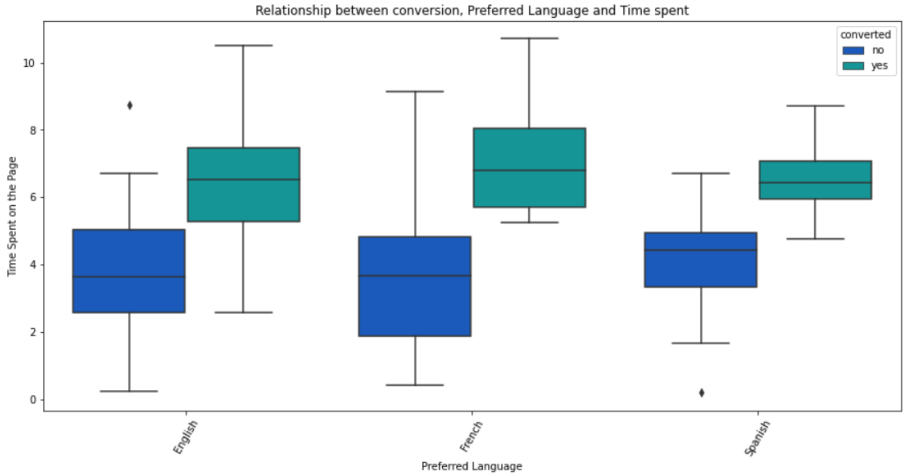
Preferred language vs Time spent on the page



- Users are spending around the same amount of time on the landing page irrespective of the language preferred.
- The mean time spent on the page for English language is the largest followed by Spanish, with French taking up the rear.

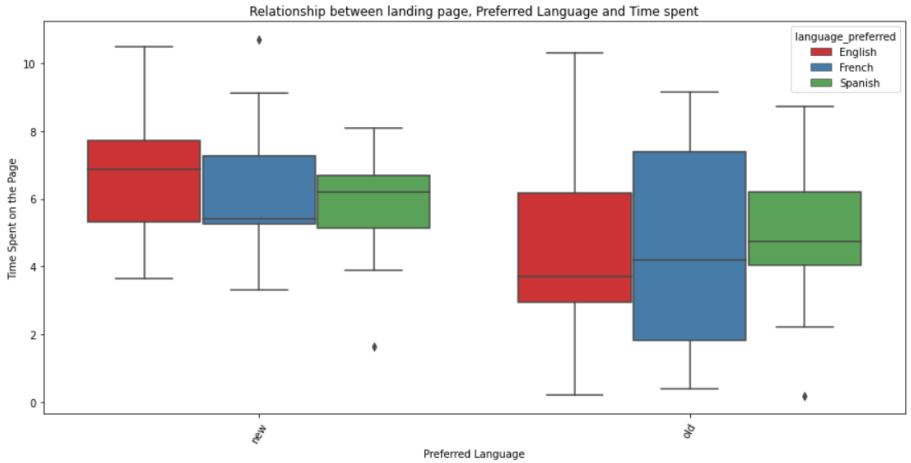
Multivariate Analysis

Conversion, preferred language & Time spent on the page



- Regardless of the language chosen, users spending more time on the landing page has become converted.
- Users who are not getting converted tend to spend lesser time on the page.

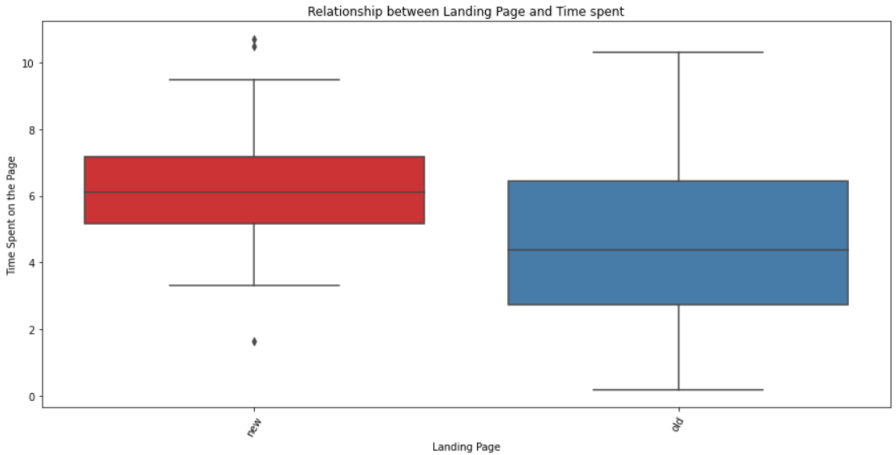
Landing page, preferred language & Time spent on the page



- Users tend to spend more time on the new landing page than old landing page, irrespective of the language they are preferring.
- We are not able to come to a specific conclusion of whether a language is preferred over the other on any landing page.

Statistical Analysis

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



Let $\mu1,\mu2$ be the mean time spent by a user on the new and old landing page.

- **Hypotheses**

$H0:\mu1=\mu2$ Null Hypothesis
$Ha:\mu1>\mu2$ Alternate Hypothesis

- **Appropriate test**

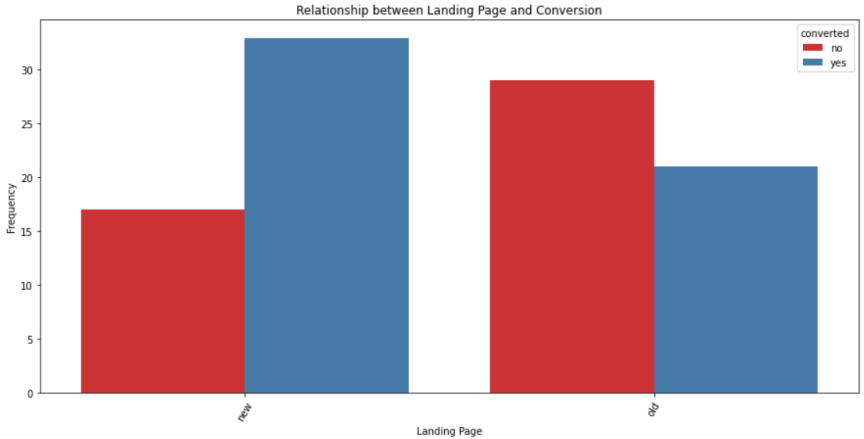
Two independent Sample T-test for Equality of Means - Unequal Standard deviation

- **P Value** 0.0000139 << significance level 0.05

Conclusion:

- As the p-value (~0.000139) is less than the level of significance, we can reject the null hypothesis.
- Hence, we do have enough evidence to support the claim that on an average, users spend more time on the new landing page.

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



Let P1,P2 be the proportions of users who visit the landing page and get converted for the new page and old page respectively.

- **Hypotheses**

$H0$:P1=P2 Null Hypothesis
Ha:P1>P2 Alternate Hypothesis

- **Appropriate test**

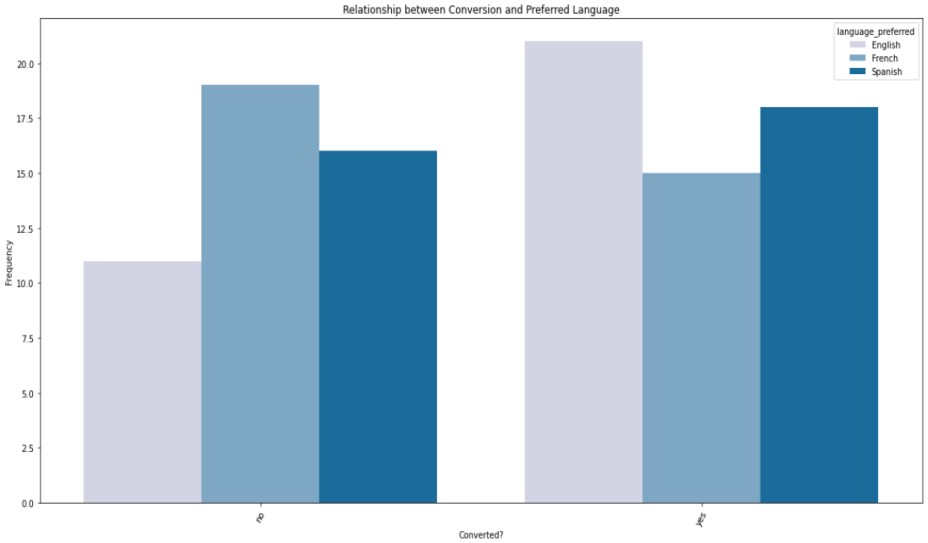
Two proportion Z test.

- **P Value** 0.0080 << significance level 0.05

Conclusion:

- As the p-value is less than the level of significance, we can reject the null hypothesis.
- We have enough statistical evidence to say that proportion of users who visit the new landing page and get converted is greater than that of old landing page.

Does the converted status depend on the preferred language?



- **Hypotheses**

$H0$: Converted status is independent of preferred language

Ha: Converted status depends on preferred language

- **Appropriate test**

Chi-square test for independence

- **Contingency table**

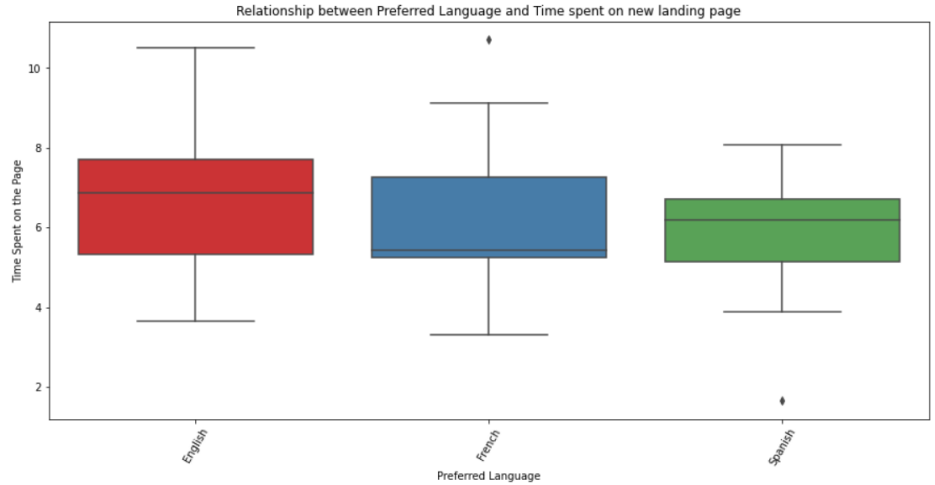
language_preferred	English	French	Spanish
converted			
no	11	19	16
yes	21	15	18

- **P Value** 0.21298 >> significance level 0.05

Conclusion:

- As the p-value is much greater than the significance level, we fail to reject the null hypothesis.
- We do not have enough statistical significance to conclude that the converted status depends on the language preferred.

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



Let $\mu1$,$\mu2$, $\mu3$ be the mean time spent on the page for French, Spanish and English respectively

- **Hypotheses**

$H0$: $\mu1=\mu2=\mu3$

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

- **Appropriate test**

One-way ANOVA Test

Normality was confirmed using the Shapiro-Wilks test

Equality of variances was confirmed using the Levene test

- **P Value** 0.4320 >> significance level 0.05

Conclusion:

- As the p-value is much greater than the significance level, we fail to reject the null hypothesis.
- We have enough statistical significance to conclude that mean time spent on the new landing page is same for different language users.

Key Insights

- When a user spend more amount of time on the landing page they tend to get converted to subscribers.
- 54% of the total users chose to be subscribers after visiting the landing page irrespective of the page being old or new.
- The users in the treatment group, who are using the new landing page, are spending more time on their page.
- The average time spent by users who are converting themselves to subscribers are more than the ones who are not.
- Users are spending around the same amount of time on the landing page irrespective of the language preferred.
- The new landing page has a higher conversion rate than the old landing page.
- The conversion rate is independent of the language preferred.

Business Recommendations

- It is statistically proven that users landing on the new page are getting converted to subscribers, we could officially replace the old landing page with the newer one.
- As we have seen, the presence of three different languages did not deter users from getting converted, there could be a chance that adding more languages could bring in more subscribers.
- Analyzing in-depth about the demographics of the users could bring in more insights for user conversion.
- This would empower us to bring in demographic specific subscription offers on the landing page which will attract more users.

greatlearning
Power Ahead

Happy Learning !