E-News Express Business Presentation

Contents

1. Business Problem Overview and Solution Approach
2. Data Overview
3. Exploratory Data Analysis (EDA) – Univariate Analysis
4. Exploratory Data Analysis (EDA) – Correlation Matrix
5. Exploratory Data Analysis (EDA) – Bivariate and Multivariate Analysis
6. Business Insights and Recommendations
7. Statistical analysis to answer the below questions:
    - 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?

Business Problem Overview and Solution Approach

- Core business idea - E-news Express, 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 using an experimental technique is known as a/b testing.
- Problem to tackle - The design team of the company has created a new landing page. The assigned task is to decide whether the new landing page is more effective to gather new subscribers. Suppose you randomly selected 100 users and divided them 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). Perform the statistical analysis to answer the questions using the collected data.

Data Overview

- Data description - Various data about the customers in control group and treatment group are collected. There are 100 rows and 6 columns in the dataset. There are no null values in the data.

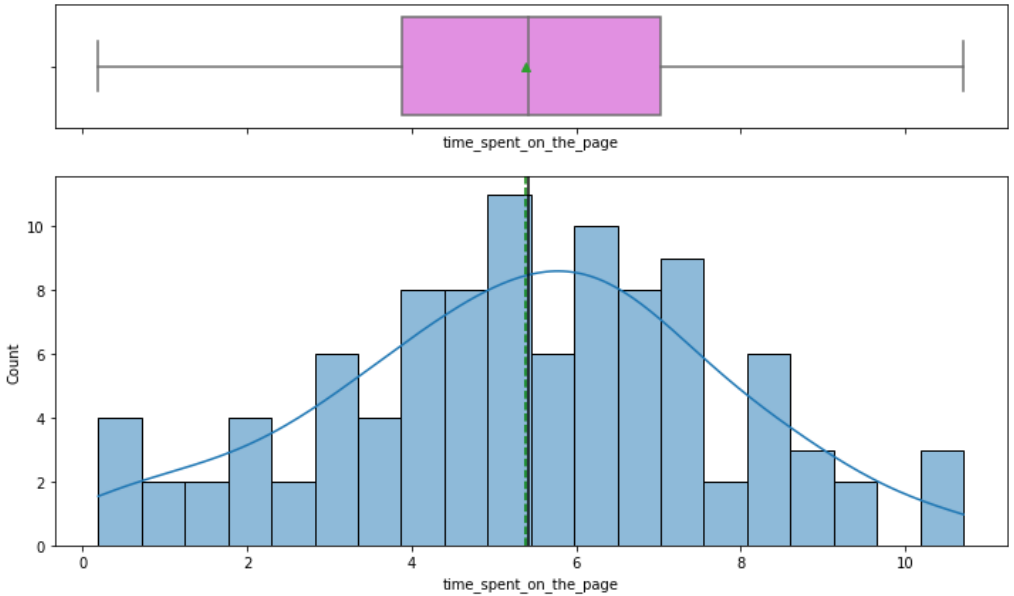
Variable	Description
user_id	user ID of the person visiting the website
group	whether the user belongs to the first group (control) or the second group (treatment)
landing_page	whether the landing page is new or old
time_spent_on_the_page	time (in minutes) spent by the user on the landing page
converted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

Exploratory Data Analysis (EDA)

- Univariate Analysis – time spent on the page field

Observations

1. The box plot indicates that there are no data outliers.
2. Since the mean and median are almost equal, The histogram does not have any skewness.
3. About 68% of the values of 'time spend on the page' field ranges from 4 to 7 minutes.

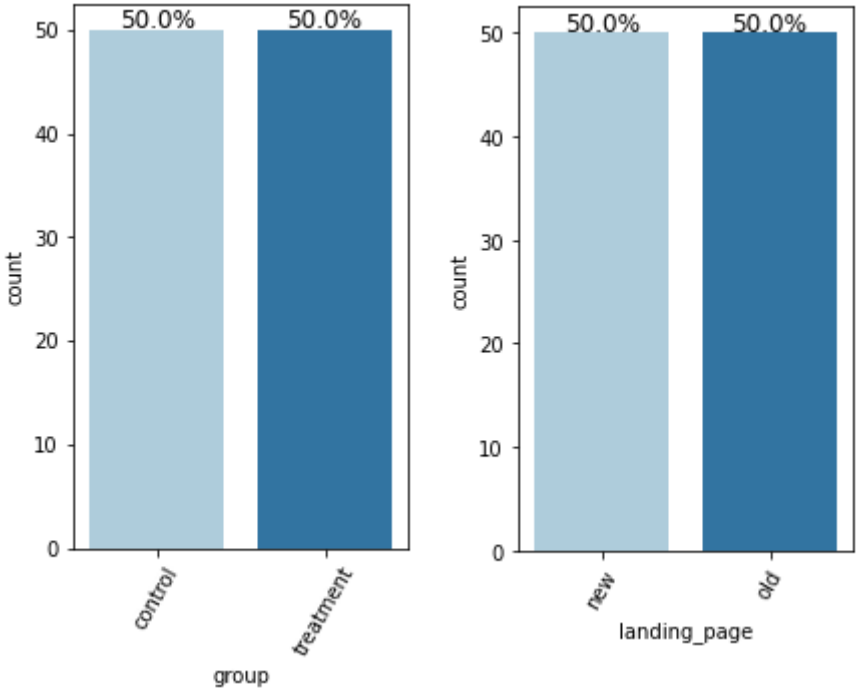


Exploratory Data Analysis (EDA)

- Univariate Analysis – group and landing page.

Observations

1. The field group has two unique values - control and treatment.
2. There are 50 observations for both control and treatment groups in this dataset
3. The field landing_page has two unique values - old and new.
4. There are 50 observations for both old and new landing pages in this dataset.

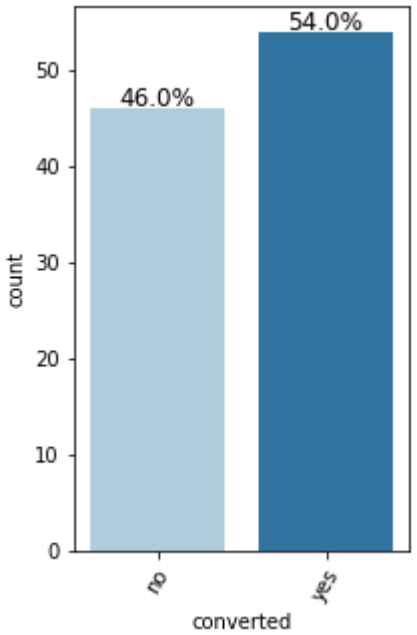


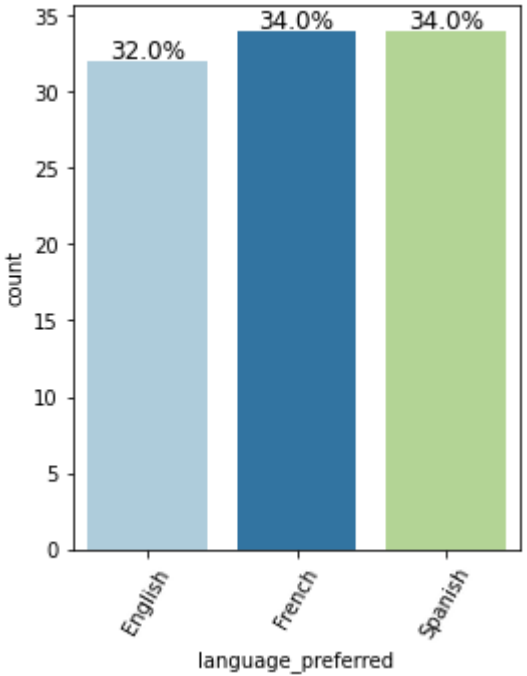
Exploratory Data Analysis (EDA)

- Univariate Analysis – converted and language preferred

Observations

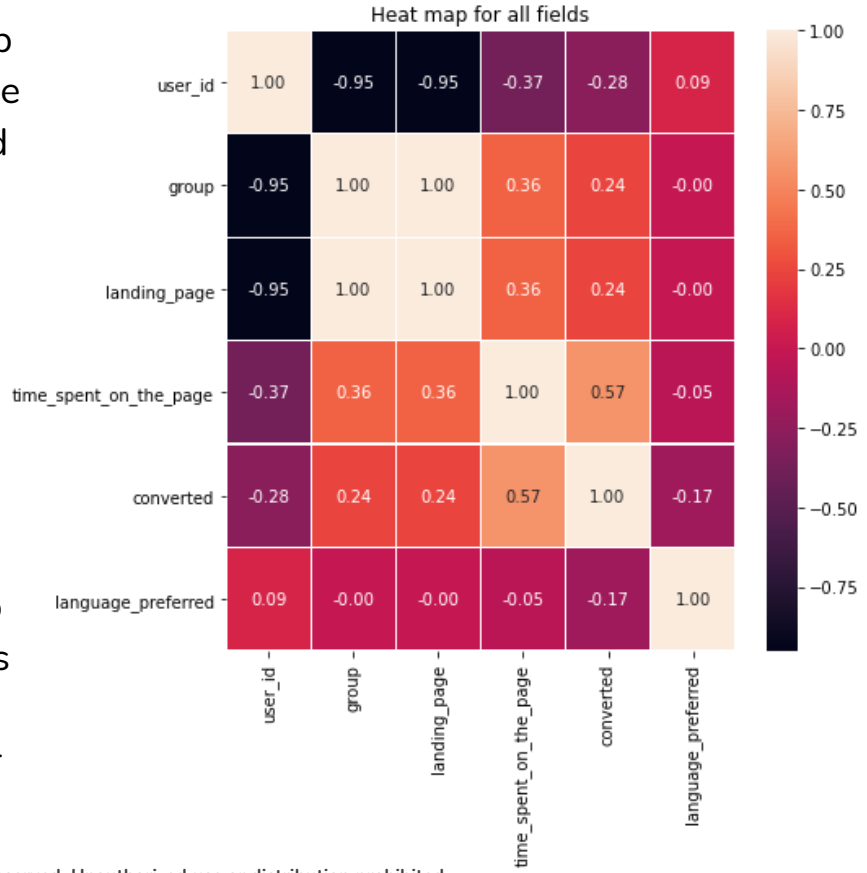
1. The field converted has two unique values - yes and no.
2. There are 54 observations with converted status 'yes' and 46 observations with converted status 'no‘
3. The field language_preferred has three unique values - French, Spanish and English.
4. There are 34 observations each for preferred language French and Spanish and 32 observations for English.





Exploratory Data Analysis (EDA) – Correlation Matrix

- There is a one-to-one correlation between group and landing page. This is not significant as all the users in control group gets old landing page and all users in treatment group gets new landing page.
- This is good correlation (0.57) between time spent on page and converted to subscriber. So we can say that more the time spent on the page, there is a better chance to become a subscriber.
- This is good correlation (0.36) between time spent on page and landing page (and group). So depending on landing page (old or new), there is an increase in time spent on page.
- It appears like the language is not a major factor for users to be converted to subscriber.



Exploratory Data Analysis (EDA) – Bivariate Analysis

- Bivariate Analysis – landing page and time spent on the page

Observations

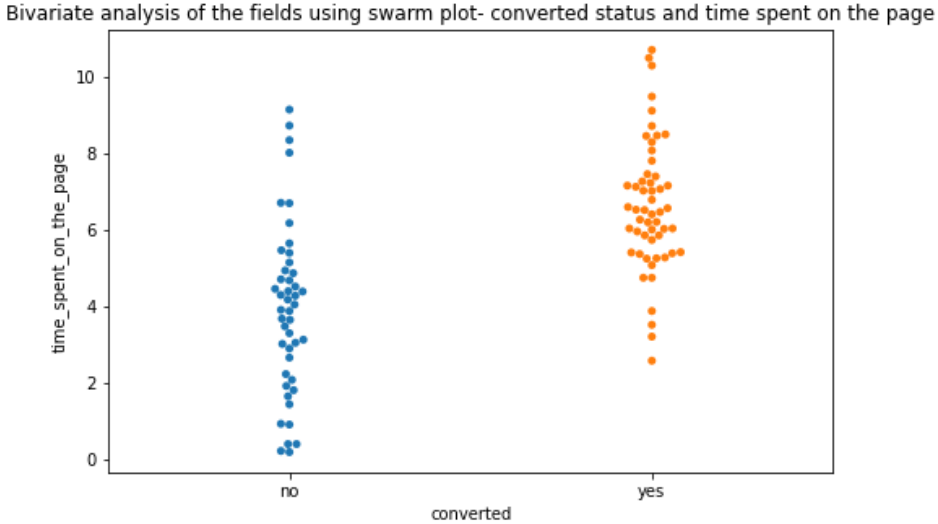
1. It is evident that the users spent more time on the new landing page.
2. The users landed on the old page tend to spent a time from less than a minute to more than 10 minutes.
3. The users landed on the new page have consistently spent more than 3.5 minutes.

Exploratory Data Analysis (EDA) – Bivariate Analysis

- Bivariate Analysis – converted status and time spent on the page

Observations

1. The more the time spent on the page, there is a greater chance to getting converted to a subscriber.
2. The users who did not get converted to subscriber have less mean time spent on the page.

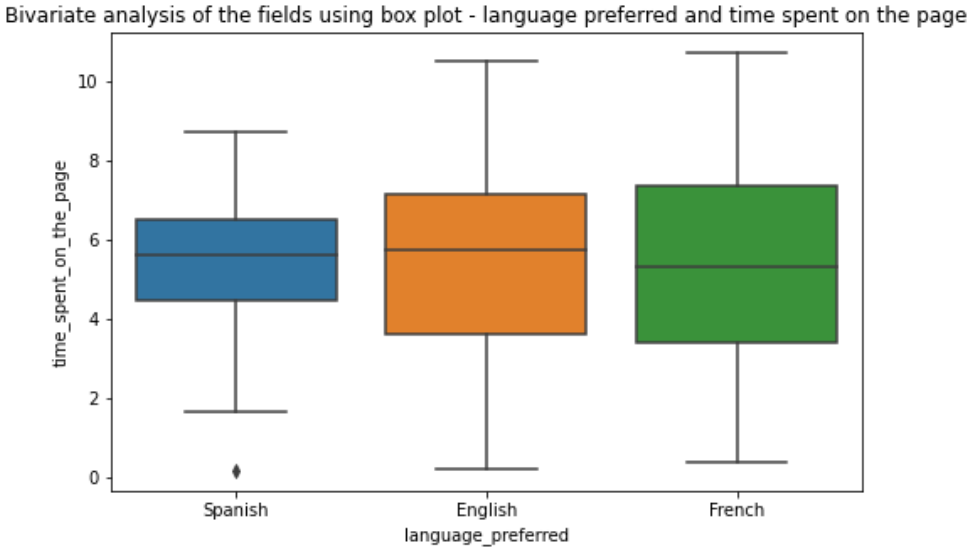


Exploratory Data Analysis (EDA) – Bivariate Analysis

- Bivariate Analysis – language preferred and time spent on the page

Observations

1. The mean time spent on the page is almost same for all three languages - Spanish, English and French.
2. The Interquartile range (IQR) for Spanish is lower compared to others, meaning the values for Spanish is in a closer range. Note that Spanish has an outlier value.
3. The Interquartile range (IQR) for English and French is higher than Spanish.

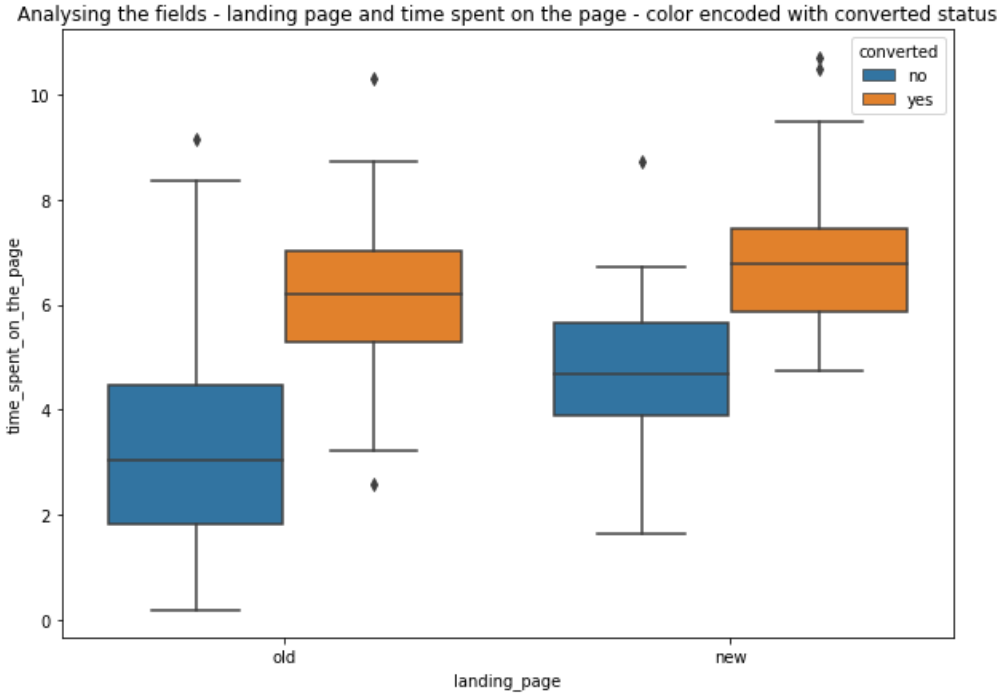


Exploratory Data Analysis (EDA) – Multivariate Analysis

- Multivariate Analysis – landing page, time spent on the page and converted

Observations

1. Regardless of the type of landing page, that more the time spent on the page, there is a better chance to become a subscriber.
2. The mean time spend on new landing page is higher than the old landing page.

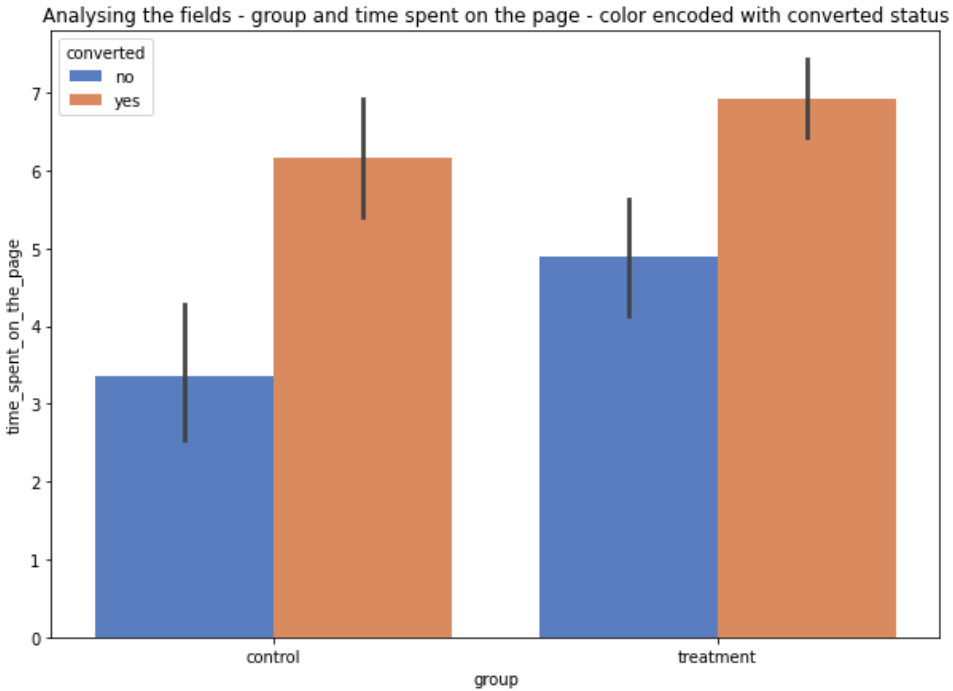


Exploratory Data Analysis (EDA) – Multivariate Analysis

- Multivariate Analysis – group, time spent on the page and converted

Observations

1. Regardless of the type of group, that more the time spent on the page, there is a better chance to become a subscriber.
2. The mean time spend on treatment group's landing page is higher than the control group's landing page.

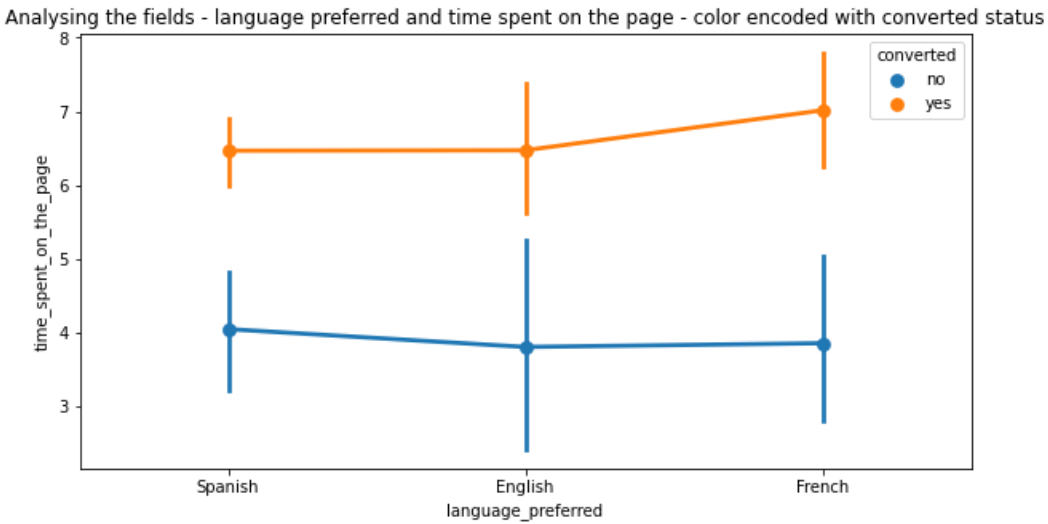


Exploratory Data Analysis (EDA) – Multivariate Analysis

- Multivariate Analysis – language preferred, time spent on the page and converted

Observations

1. Again, no surprises here. Regardless of the language, the users who spent more time on a page has a better chance to becoming a subscriber.
2. The French language users who convert to subscribers have a higher mean time spent on the page. Looks like they liked the features of the new page more than other language customers.

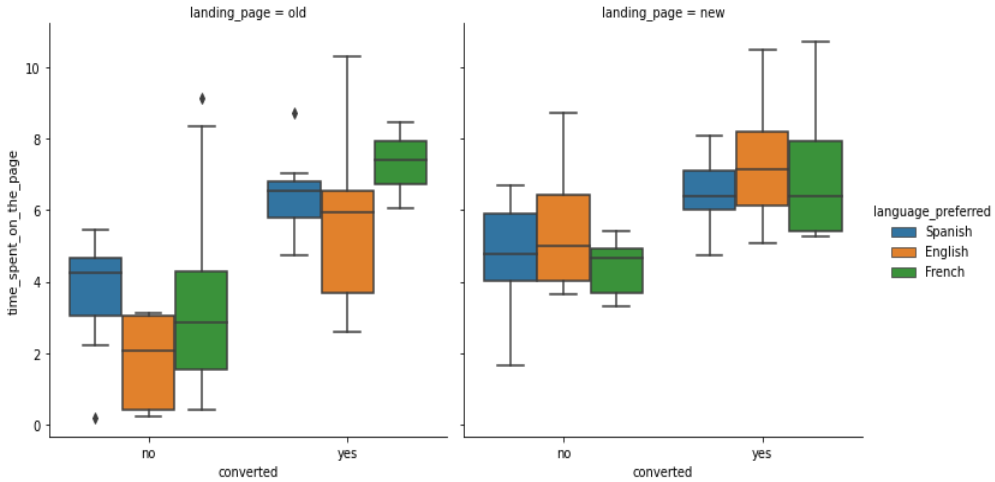


Exploratory Data Analysis (EDA) – Multivariate Analysis

- Multivariate Analysis – landing page, language, time spent on the page and converted

Observations

1. For the old landing page, the time spent on page for converted users is very high for users who did not convert. This is true for all languages.
2. For the new landing page, the time spent on page for converted users is marginally high for users who did not convert. This is true for all languages.
3. This proves that the new page is doing a better job compared to the old landing page across languages.



Exploratory Data Analysis (EDA) – Multivariate Analysis

- Multivariate Analysis – landing page, language, no of users and converted

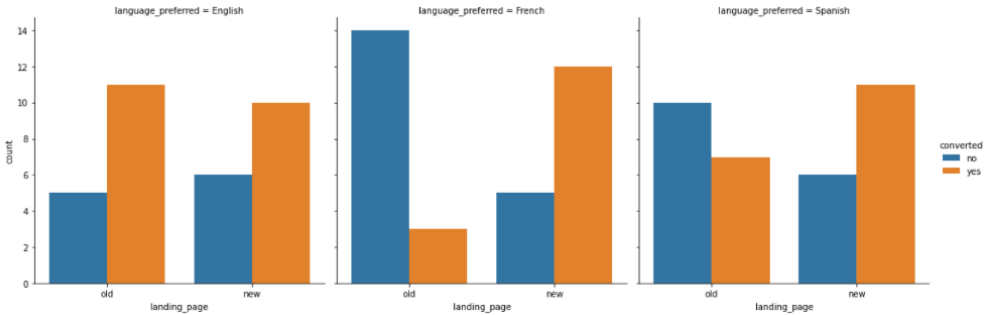
Observations

For English, there is a good chance to converting to subscriber regardless of the landing page.

- Out of the 16 users landed in old page, 11 of them converted to subscribers.
- Out of the 16 users landed in new page, only 10 of them converted to subscribers.

For French, there is a good chance to converting to subscriber if the user lands in new landing page.

- Out of the 17 users landed in old page, only 3 of them converted to subscribers.
- Out of the 17 users landed in new page, 12 of them converted to subscribers.



For Spanish, there is a good chance to converting to subscriber if the user lands in new landing page.

- Out of the 17 users landed in old page, only 7 of them converted to subscribers.
- Out of the 17 users landed in new page, 11 of them converted to subscribers.

Business Insights and Recommendations

1. Overall, the new landing page design is promising. The users landed on the new page has a great chance of becoming a subscriber. The E-News Express Design team should focus to making the page more user friendly.
2. The time spent on the page is a major factor for becoming a subscriber. This trend can be seen on old landing page as well. The E-News Express Design team should add more features in the page, so that the user spends more time on it and eventually becomes a subscriber.
3. The preferred language does not seem to have a big impact when it comes to user subscription. For English, one can say that the almost 65% of users have subscribed regardless of the landing page. But on a deeper analysis, we can infer that these converted users spend an average of 5 minutes in the landing page and that could be a reason for the subscription.

Business Insights and Recommendations (continued...)

4. It is impossible to ignore that the French language users who converted to subscribers spent higher than average time on the landing page. Looks like they liked the features on the new page more than other language customers. This implies the importance of designing pages taking into account the nature of the target population rather than a one-size-fits-all approach.
5. E-News Express Design team should continue to improve the design of landing pages and continue testing with an expanded user group to yield better results.
6. E-News Express should consider better marketing approaches to promote user subscription. Offering discounts and shopping coupons will encourage users to subscribe for E-News Express.
7. E-News Express should consider expanding the existing market and provide more language options. This could lead to more user subscription.

Statistical Analysis – Hypothesis Testing

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

Writing the null and alternate hypotheses

Let μ_o, μ_n be the mean time spend on old landing page and new landing page respectively.

We will test the null hypothesis

$$H_0 : \mu_o = \mu_n$$

against the alternate hypothesis

$$H_a : \mu_o < \mu_n$$

Insight

As the p-value(~0.0001) is much less than the level of significance, we can reject the null hypothesis. Hence, we have significant evidence to conclude that the users spend more time on the new landing page than the old landing page at 0.05 significance level.

Testing whether the T-test assumptions are satisfied or not

- Continuous data - Yes, the usage time is measured on a continuous scale.
- Normally distributed populations - Yes, we are informed that the populations are assumed to be normal.
- Independent populations - As we are taking random samples for two different type of landing pages, the two samples are from two independent populations.
- Unequal population standard deviations - As the sample standard deviations are different, the population standard deviations may be assumed to be different.
- Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample

Statistical Analysis – Hypothesis Testing

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

Writing the null and alternative hypothesis

Let p_o, p_n be the proportions of users converted to subscribers in old and new landing pages respectively.

We will test the null hypothesis

$$H_0 : p_o = p_n$$

against the alternate hypothesis

$$H_a : p_o < p_n$$

Testing whether the Z-test assumptions are satisfied or not

- Binomially distributed population - Yes, a user is either converted to a subscriber or not.
- Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample.
- Can the binomial distribution approximated to normal distribution - Yes. The standard thing is to check whether np and n(1-p) are greater than or equal to 10. Here, n and p refer to the sample size and sample proportion respectively.

Insight

As the p-value(~0.016) is much less than the level of significance, we can reject the null hypothesis. Hence, we have significant evidence to conclude that 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 at 0.05 significance level.

Statistical Analysis – Hypothesis Testing

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

Writing the null and alternative hypothesis ¶

We will test the null hypothesis

> H_0 : Language preference is independent of converted status.

against the alternate hypothesis

> H_a : Language preference depends on converted status.

We can find the p-value using the Chi-Square Test for Independence

Testing whether the assumptions for Chi-Square Test for Independence are satisfied or not

- Categorical variables - Yes
- Expected value of the number of sample observations in each level of the variable is at least 5 - Yes, the number of observations in each level is greater than 5.
- Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample.

Insight

As the p-value(~0.212) is much greater than the level of significance (0.05), we failed to reject the null hypothesis. Hence, we have significant evidence to conclude that the language preference is independent of converted status at 0.05 significance level.

Statistical Analysis – Hypothesis Testing

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

Let's write the null and alternative hypothesis

Let μ_f, μ_s, μ_e be the means of time spent on new landing page for fuel type French, Spanish and English respectively.

We will test the null hypothesis

> $H_0 : \mu_f = \mu_s = \mu_e$ The mean time spent on new landing page for fuel type French, Spanish and English is equal

against the alternative hypothesis

> H_a : At least one of the mean time spent on new landing page is different from the rest.

Testing whether the assumptions for One-way ANOVA Test are satisfied or not

- The populations are normally distributed - Yes, the normality assumption is verified using the Shapiro-Wilk's test.
- Samples are independent simple random samples - Yes, we are informed that the collected sample is a simple random sample.
- Population variances are equal - Yes, the homogeneity of variance assumption is verified using the Levene's test.

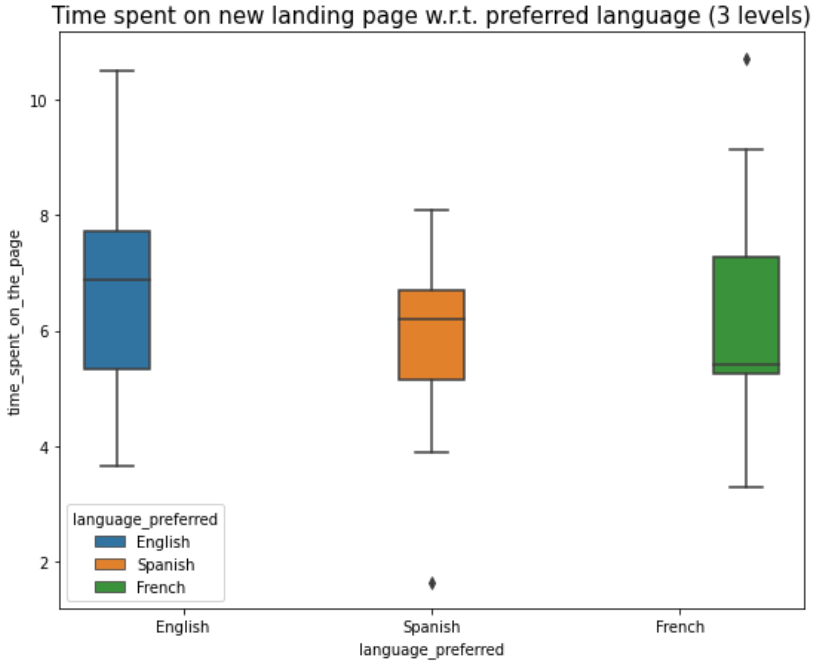
Statistical Analysis – Hypothesis Testing

Insight

As the p-value (~0.432) is much greater than the significance level, we failed to reject the null hypothesis. Hence, we do have enough statistical significance to conclude that the mean time spent on the new page is same for the different language users at 5% significance level.

If we want to drill down further and verify the above conclusion, Multiple comparison tests can be used to see the differences between all pairs of means.

- As the p-values for comparing the mean time spent on the new page for the pair English-French, English-Spanish and French-Spanish is greater than the significance level, the nul hypothesis of equality of all population means cannot be rejected.
- Thus, we can say that the mean time spent on the new page for English-French, English-Spanish and French-Spanish is similar.



greatlearning
Power Ahead

Happy Learning !