

E-news Express

Project 2 - Business Statistics

March 15, 2025

Submitted By:
Alex Kyeremateng Botwe

Contents / Agenda

	Topics	Page No.
	Executive Summary	5
	Business Problem Overview and Solution Approach	9
	EDA Results	12
	Hypotheses Tested and Results	20
	Appendix	28

List of Tables

No.	Name of Table	Page No.
1	Top 5 rows of the Dataset	29
2	Bottom 5 rows of the Dataset	29
3	Information on the Dataset	30
4	Statistical Summary of the Numerical Variables	32
5	Statistical Summary of the Categorical Variables	32

List of Figures

No.	Name of Figure	Page No.
1	Univariate Analysis of Time Spent on Page	12
2	Univariate Analysis of Group	13
3	Univariate Analysis of Landing Page	14
4	Univariate Analysis of Converted	15
5	Univariate Analysis of Language Preferred	16
6	Bivariate Analysis of Landing Page vs Time Spent on the Page	17
7	Bivariate Analysis of Conversion Status vs Time spent on the Page	18
8	Bivariate Analysis of Language preferred vs Time spent on the page	19
9	Visual Analysis of Time spent on the Landing Page	20
10	Visual Analysis of Conversion Rate	22
11	Visual Analysis of Converted Status vs Preferred Language	24
12	Visual Analysis of Time Spent on Landing Page vs Preferred Language	26

Executive Summary

- In a quest to expand its business by acquiring new subscribers, the design team of E-news Express, an online news portal developed a new landing page aimed at enhancing user engagement and conversion rates. This decision was based on observations made by the executives that there was a decline in new monthly subscribers, possibly due to the design and content of the old landing page. In order to test the effectiveness of the new landing page gathering new subscribers an A/B testing experiment was conducted, by randomly selecting 100 users and dividing them into two groups:
 - i. Control Group: Where users were served with the existing landing page
 - ii. Treatment Group: Where users were served with the new landing page

We analyzed data and performed hypothesis testing regarding user interactions at a significance level of 5% to determine whether the new landing page improved engagement and conversion rates.

Executive Summary (continued)

- **Business Insights:**
 - a. **Time spent on the Landing Pages:** There was a strong statistical evidence that users spend significantly more time on the new landing page as compared to the old landing page. This suggests that the new layout and content are more engaging and encourage users to stay longer.
 - b. **Conversion Rate for the Landing Pages:** There was statistically significant evidence that the conversion rate for the new landing page was higher than that of the old landing page. This indicates that the new design is more effective in converting users and will encourage visitors to subscribe more.

Executive Summary (continued)

- c. **Converted Status and Language Preference:** There was no statistical evidence to prove that preferred language significantly impacts the possibility of a user converting. Hence, language preference might not need to be prioritized when making decisions about how to improve conversion rates.

- d. **Engagement Levels by Language :** There was not enough statistical evidence to claim that the time spent on the new landing page differs significantly between the users who speak English, French, and Spanish. This therefore suggest that the language preference of users does not have any significant impact on the time they spend on the new page.

Executive Summary (continued)

- **Business Recommendations:**

- a. Roll out the new landing page to all users, since it significantly improves engagement and conversion.
- b. Gather qualitative feedback from users to understand what aspects of the pages are most engaging.
- c. Monitor user behavior after implementation to validate long-term effectiveness.
- d. Optimize the page load speed and mobile responsiveness to ensure users stay engaged.
- e. Enhance page design to increase engagement
- f. Ensuring that the user experience is consistent across language versions could lead to a uniform user satisfaction level

Executive Summary (continued)

- **Conclusions:**

To enhance overall user engagement and drive business growth for E-news Express, the next steps involve fully implementing the new landing page for all users. E-new Express should focus on boosting subscriptions while continuously testing and refining its elements. Additionally, E-new Express should tailor marketing campaigns and page content to cater to each language group, maximizing relevance and user satisfaction. It is essential to consider other performance metrics such as time spent on the page, bounce rates, and user interactions throughout this process.

Business Problem Overview and Solution Approach

- **Business Problem Overview:**

E-news Express, an online news portal, has observed a decline in new monthly subscribers compared to the previous year. Executives believe this drop may be due to the current landing page's ineffective design in engaging users and encouraging subscriptions.

To address this issue, the design team developed a new landing page with an improved layout and more relevant content. The company aims to test whether this new design increases user engagement and conversion rates.

To make a data-driven decision, the Data Science team conducted an A/B test, where users were randomly divided into two groups:

Control Group: Shown the existing landing page.

Treatment Group: Shown the new landing page.

The goal of this experiment was to analyze user interactions and determine whether the new landing page leads to increased engagement and higher subscription rates.

Business Problem Overview and Solution Approach (continued)

- **Solution Approach**

To evaluate the effectiveness of the new landing page, a **statistical analysis** was conducted using the following steps:

1. **Compare User Engagement (Time Spent on Page)**

Used a **one-tailed t-test** to check whether users spend more time on the new landing page than on the old landing page.

2. **Analyze Conversion Rates**

Applied a **Z-test for proportions** to determine if the new landing page has a significantly higher conversion rate than the old landing page.

Business Problem Overview and Solution Approach (continued)

3. Assess Impact of Language Preference on Conversion

Created a **contingency table** and performed a **Chi-square test for independence** to examine if conversion rates vary by preferred language.

4. Evaluate Time Spent by Different Language Users

Conducted an **ANOVA test** to check if different language users spend varying amounts of time on the new landing page.

EDA - Univariate Analysis

- Time Spent on the Page

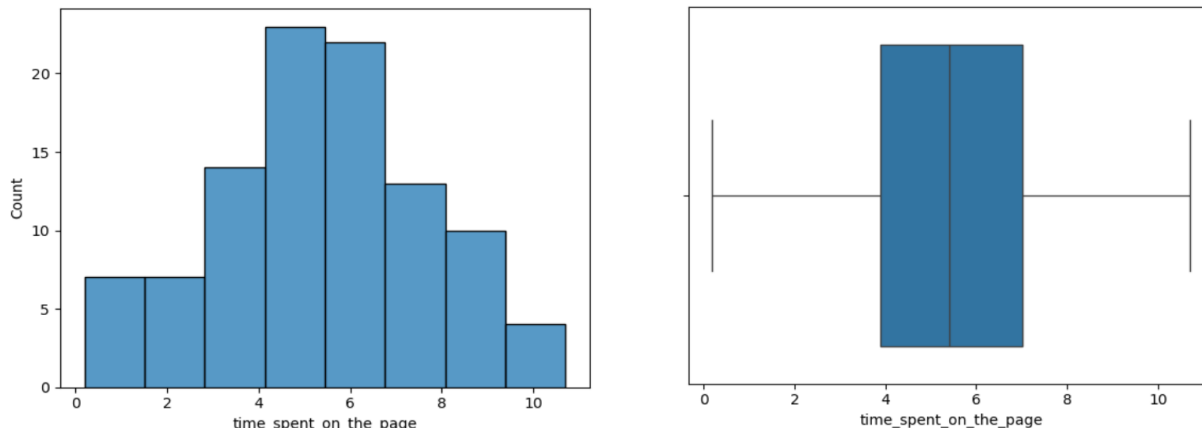


Fig 1: Univariate Analysis of Time Spent on Page

Observations:

Most users spend between 3.88 to 7.02 seconds on the page. The distribution appears roughly normal but slightly skewed.

[Link to Appendix slide on data background check](#)

- Group

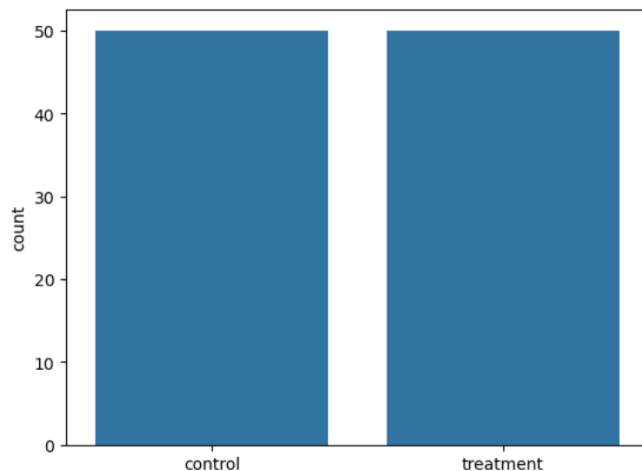


Fig 2: Univariate Analysis of Group

Observations:

The groups are evenly split with the Control Group of 50 users and the Treatment Group of 50 users. This will ensure a fair comparison between the old and new landing pages.

[Link to Appendix slide on data background check](#)

- Landing Page

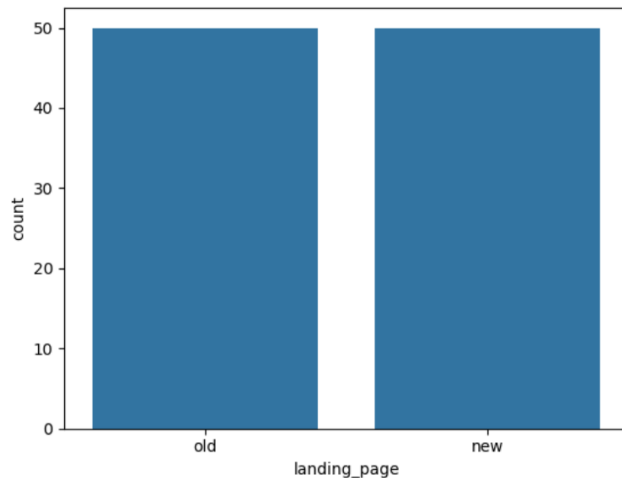


Fig 3: Univariate Analysis of Landing Page

Observations:

This is a balanced distribution since the old landing pages have 50 users and the new landing pages also have 50 users. This means the experiment is balanced and there will be no bias in the A/B testing

[Link to Appendix slide on data background check](#)

- Converted

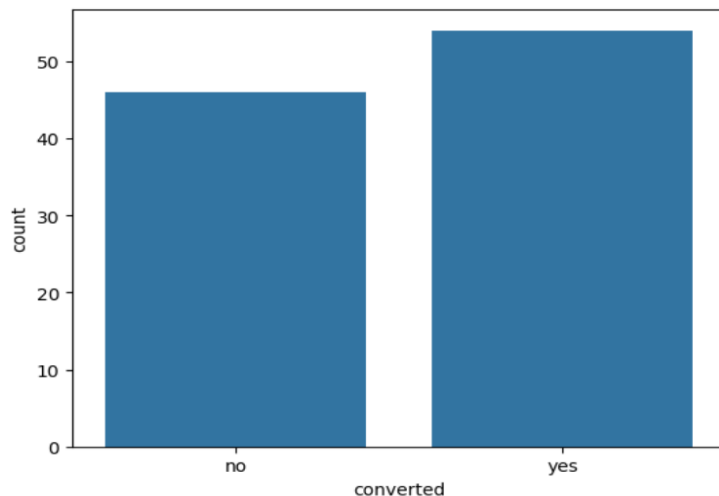


Fig 4: Univariate Analysis of Converted

Observations:

54 users converted and 46 users did not convert. This means slightly more users converted than those who did not.

[Link to Appendix slide on data background check](#)

- Language Preferred

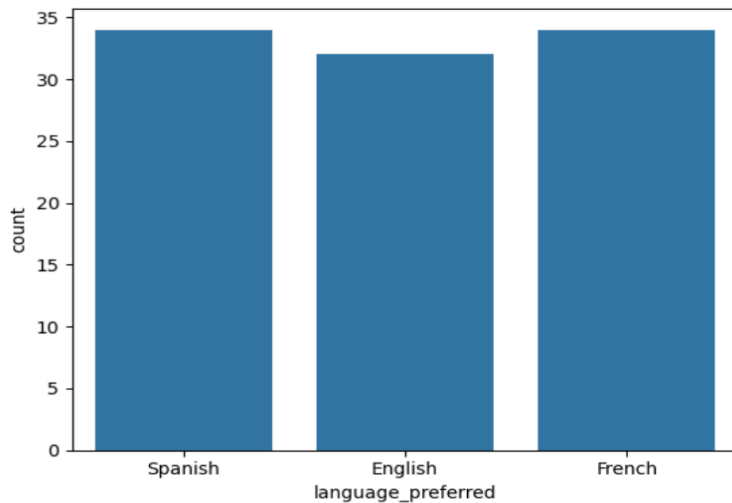


Fig 5: Univariate Analysis of Language Preferred

Observations:

The distribution is fairly balanced across languages with Spanish having 34 users, French 34 users and English 32 users. This means language preference does not likely introduce bias.

[Link to Appendix slide on data background check](#)

EDA - Bivariate Analysis

- Landing Page vs Time Spent on the Page

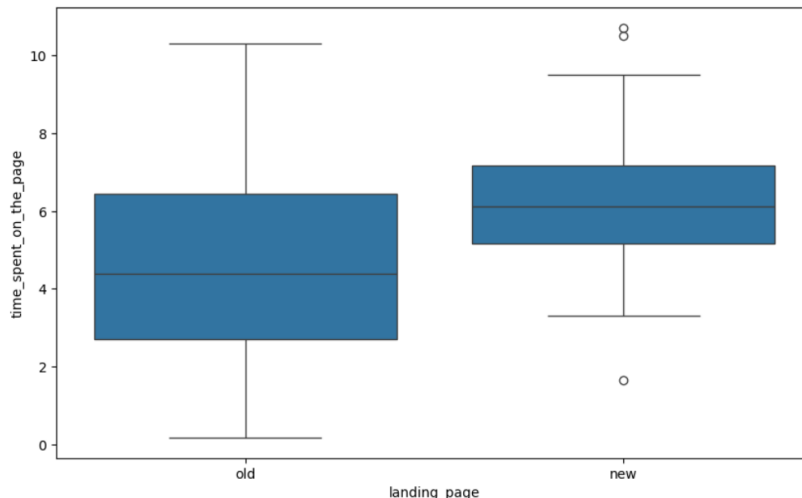


Fig 6: Bivariate Analysis of Landing Page vs Time Spent on the Page

Observations:

Users spent more time on the new landing page compared to the old landing page. There is a noticeable increase in the median time on the new landing page which suggests that users may find it more engaging or require more time to interaction.

[Link to Appendix slide on data background check](#)

- Conversion Status vs Time spent on the Page

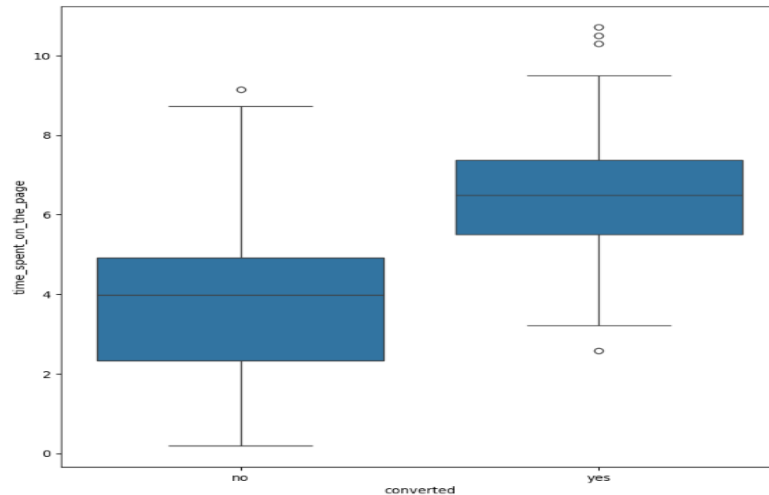


Fig 7: Bivariate Analysis of Conversion Status vs Time spent on the Page

Observations:

Users who converted tended to spend more time on the page compared to those who did not. This may indicate that either the users who explore more are more likely to convert, or that the page effectively encourages conversions over time.

[Link to Appendix slide on data background check](#)

- Language preferred vs Time spent on the page

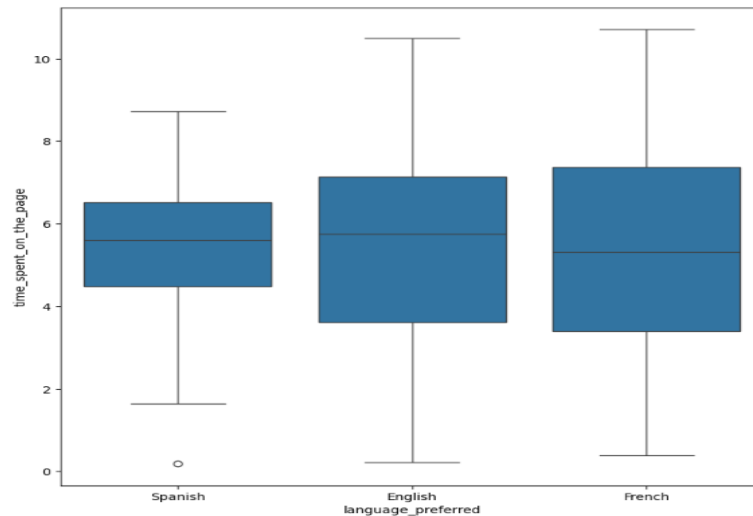


Fig 8: Bivariate Analysis of Language preferred vs Time spent on the page

Observations:

There are variations in time spent across different languages. Some languages show a higher median time spent, which might indicate differences in comprehension speed, cultural engagement, or ease of navigation in certain languages.

[Link to Appendix slide on data background check](#)

Hypotheses Tested and Results

1. Visual Analysis Performed - Time spent on the Landing Page

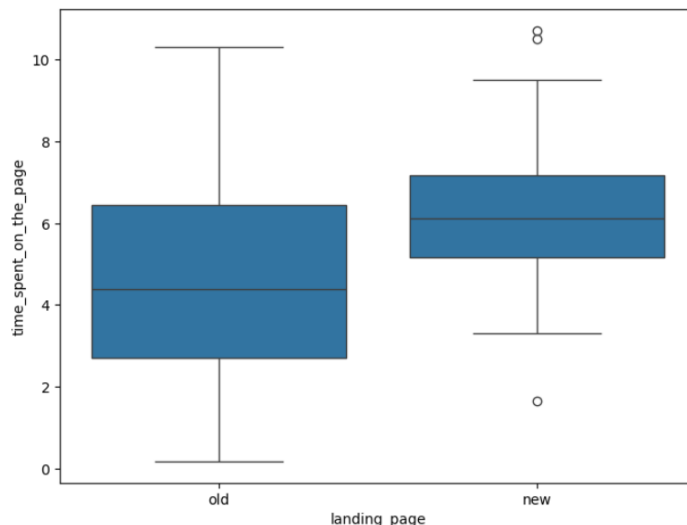


Fig 9: Visual Analysis of Time spent on the Landing Page

Analysis:

The median time spent appears higher for the new landing page indicating that users spent more time on the new landing page as compared to the old landing page. There are some outliers, indicating that a few users spent significantly more time on both pages.

[Link to Appendix slide on details of the test performed](#)

- **Hypothesis Tested**

A one-tailed t-test was performed because the population standard deviation was unknown

- **Test Result**

The p-value (0.000139) was less than the level of significance of 0.05 so we rejected the null hypothesis.

- **Inference**

There is strong statistical evidence that users spend more time on the new landing page than on the old landing page. This suggests that the new design is more engaging and thus customers prefer it.

[Link to Appendix slide on details of the test performed](#)

2. Visual Analysis Performed - Conversion Rate

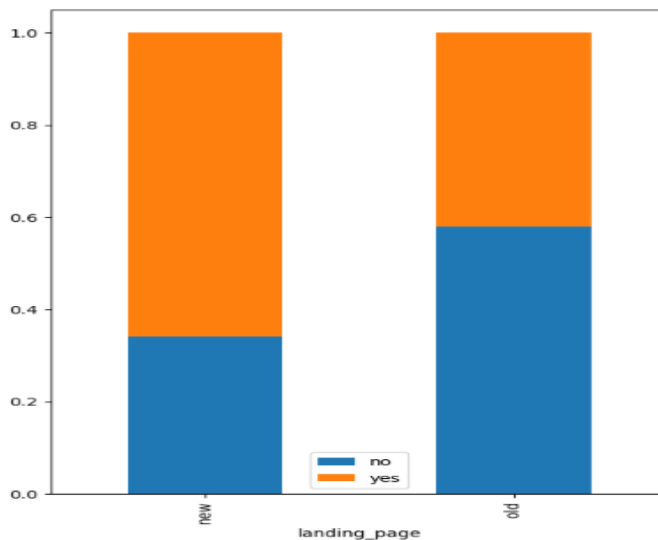


Fig 10: Visual Analysis of Conversion Rate

Analysis:

From the visual representation, the conversion rate of the new landing page seems to be higher than that of the old landing page indicating that the new page might be more effective

[Link to Appendix slide on details of the test performed](#)

- **Hypothesis Tested**

A proportion z-test, which is used to compare two population proportions was performed

- **Test Result**

The p-value of 0.01605 was less than the level of significance of 0.05 so we rejected the null hypothesis.

- **Inference**

There is statistically significant evidence that the conversion rate for the new landing page is higher than the old page. This suggests that the new page is more effective at converting users. This means the design or content of the new page are likely contributing to better performance

[Link to Appendix slide on details of the test performed](#)

3. Visual Analysis Performed - Converted Status vs Preferred Language

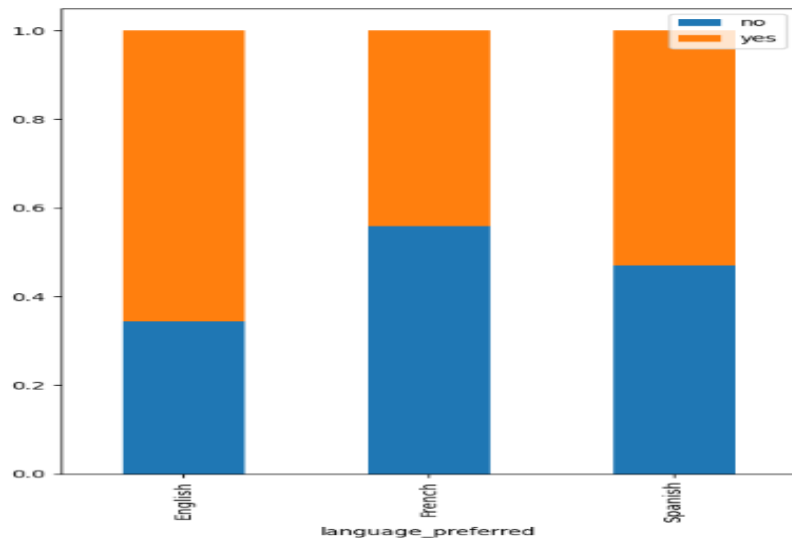


Fig 11: Visual Analysis of Converted Status vs Preferred Language

Analysis:

Conversion rates vary across different language preferences. Some languages show a higher proportion of conversion than others, suggesting that language preference might influence conversion behavior but more statistical test is required.

[Link to Appendix slide on details of the test performed](#)

- **Hypothesis Tested**

A Chi-Square test for independence was performed to determine whether conversion status depends on preferred language.

- **Test Result**

The p-value of 0.21298 was greater than the level of significance of 0.05 so we fail to reject the null hypothesis

- **Inference**

From a statistical perspective, there is no evidence to suggest that preferred language significantly impacts the likelihood of a user converting. Therefore, language preference might not need to be prioritized when making decisions about how to improve conversion rates.

[Link to Appendix slide on details of the test performed](#)

4. Visual Analysis Performed – Time Spent on Landing Page vs Preferred Language

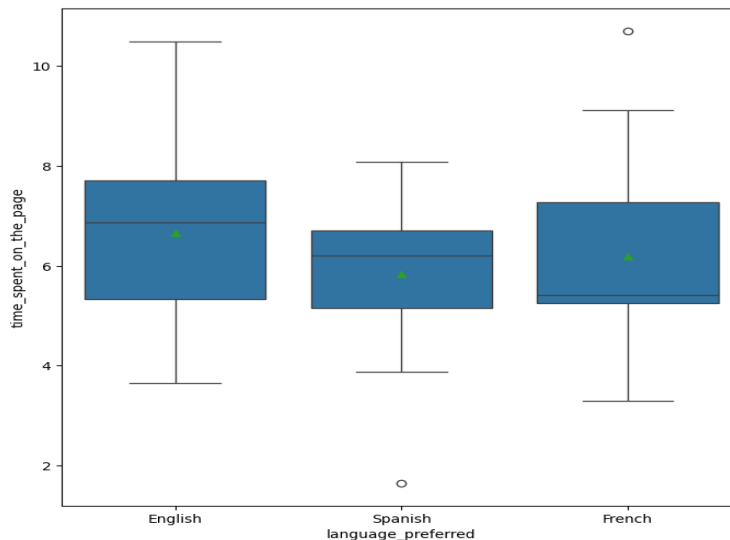


Fig 12: Visual Analysis of Time Spent on Landing Page vs Preferred Language

Analysis:

Users with different language preferences spend varying amounts of time on the landing page. Certain language groups have a higher median time spent, possibly indicating differences in engagement or ease of understanding.

[Link to Appendix slide on details of the test performed](#)

- **Hypothesis Tested**

A One-Way ANOVA (Analysis of Variance) was performed since we are comparing means across multiple independent groups

- **Test Result**

The p-value of 0.4320 was greater than the level of significance of 0.05 so we fail to reject the null hypothesis

- **Inference**

There is no statistically significant difference in the average time spent on the new landing page between the users who speak English, French, and Spanish. This therefore suggest that the language preference of users does not have a statistically significant impact on the time they spend on the new page.

[Link to Appendix slide on details of the test performed](#)

APPENDIX

Data Background and Contents

- Data Overview

The dataset consists of 100 rows and 6 columns, representing data from an A/B test. It includes user interaction details with different landing pages.

	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
0	546592	control	old	3.48	no	Spanish
1	546468	treatment	new	7.13	yes	English
2	546462	treatment	new	4.40	no	Spanish
3	546567	control	old	3.02	no	French
4	546459	treatment	new	4.75	yes	Spanish

Table 1: Top 5 rows of the Dataset

	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
95	546446	treatment	new	5.15	no	Spanish
96	546544	control	old	6.52	yes	English
97	546472	treatment	new	7.07	yes	Spanish
98	546481	treatment	new	6.20	yes	Spanish
99	546483	treatment	new	5.86	yes	English

Table 2: Bottom 5 rows of the Dataset

Data Background and Contents (continued)

- Data Background

The dataset [abtest.csv](#) was used for analyzing the effectiveness of two landing pages (old vs. new) in converting users. It includes user engagement metrics and demographic preferences, making it useful for hypothesis testing.

- Data Contents

#	Column	Non-Null Count	Dtype
0	user_id	100 non-null	int64
1	group	100 non-null	object
2	landing_page	100 non-null	object
3	time_spent_on_the_page	100 non-null	float64
4	converted	100 non-null	object
5	language_preferred	100 non-null	object

dtypes: float64(1), int64(1), object(4)
memory usage: 4.8+ KB

Table 3: Information on the Data Set

Data Background and Contents (continued)

There are three datatypes namely: int64(1), float64(1) and object64(4)

- user_id (*int*): Unique identifier for each user.
- group (*object*): Whether the user was in the "control" (old page) or "treatment" (new page) group.
- landing_page (*object*): The type of page the user was shown ("old" or "new").
- time_spent_on_the_page (*float*): Duration (in seconds) the user spent on the landing page.
- converted (*object*): Whether the user converted ("yes" or "no").
- language_preferred (*object*): The user's preferred language.

Data Background and Contents (continued)

	user_id	time_spent_on_the_page
count	100.000000	100.000000
mean	546517.000000	5.377800
std	52.295779	2.378166
min	546443.000000	0.190000
25%	546467.750000	3.880000
50%	546492.500000	5.415000
75%	546567.250000	7.022500
max	546592.000000	10.710000

Table 4: Statistical Summary of the Numerical Variables

	group	landing_page	converted	language_preferred
count	100	100	100	100
unique	2	2	2	3
top	control	old	yes	Spanish
freq	50	50	54	34

Table 5: Statistical Summary of the Categorical Variables

Data Background and Contents (continued)

```
user_id          0
group            0
landing_page     0
time_spent_on_the_page  0
converted        0
language_preferred  0
dtype: int64
```

- There are no missing values in the dataset
- There are no duplicates in the dataset

Hypothesis Testing Details

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

- Step 1: Define the null and alternate hypotheses

$$H_0: \mu_{\text{new}} \leq \mu_{\text{old}}$$

$$H_a: \mu_{\text{new}} > \mu_{\text{old}}$$

- Step 2: Select Appropriate test

This is a one-tailed t-test, as we are testing whether the mean time spent on the new page is significantly greater than that on the old page.

- Step 3: Decide the significance level

As given in the problem statement, we select $\alpha = 0.05$

Hypothesis Testing Details

- Step 4: Collect and prepare data
 - i. The sample standard deviation of the time spent on the new page is: 1.82
 - ii. The sample standard deviation of the time spent on the new page is: 2.58

- Step 5: Calculate the p-value

The p-value is 0.0001392

- Step 6: Compare the p-value with α

As the p-value is less than the level of significance α , we reject the null hypothesis

Hypothesis Testing Details

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?

- Step 1: Define the null and alternate hypotheses

$$H_0: P_{\text{new}} \leq P_{\text{old}}$$

$$H_a: P_{\text{new}} > P_{\text{old}}$$

- Step 2: Select Appropriate test

We performed a proportion z-test, which is used to compare two population proportions

- Step 3: Decide the significance level

As given in the problem statement, we select $\alpha = 0.05$

Hypothesis Testing Details

- Step 4: Collect and prepare data

The numbers of users served the new and old pages are 50 and 50 respectively

- Step 5: Calculate the p-value

The p-value is 0.01605

- Step 6: Compare the p-value with α

As the p-value is less than the level of significance α , we reject the null hypothesis

Hypothesis Testing Details

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

- Step 1: Define the null and alternate hypotheses

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

H_a : The converted status depends on the preferred language

- Step 2: Select Appropriate test

We performed a Chi-Square test for independence.

- Step 3: Decide the significance level

As given in the problem statement, we select $\alpha = 0.05$

Hypothesis Testing Details

- Step 4: Collect and prepare data

A contingency table was created showing the distribution of the two categorical variables

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

- Step 5: Calculate the p-value

The p-value is 0.21298

- Step 6: Compare the p-value with α

As the p-value is greater than the level of significance α , we fail reject the null hypothesis

Hypothesis Testing Details

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

- Step 1: Define the null and alternate hypotheses

$$H_0: \mu_{English} = \mu_{French} = \mu_{Spanish}$$

H_a : At least one mean time spent is different

- Step 2: Select Appropriate test

A One-Way ANOVA (Analysis of Variance) was performed since we were comparing means across multiple independent groups

- Step 3: Decide the significance level

As given in the problem statement, we select $\alpha = 0.05$

Hypothesis Testing Details

- Step 4: Collect and prepare data

We calculated the mean time spent on the new page for different language users

```
language_preferred
English      6.663750
French       6.196471
Spanish      5.835294
Name: time spent on the page, dtype: float64
```

- Step 5: Calculate the p-value

The p-value is 0.4320

- Step 6: Compare the p-value with α

As the p-value is greater than the level of significance α , we fail to reject the null hypothesis

E. Glossary

- User ID: A unique identifier assigned to each website visitor
- Group: Specifies whether a user is part of the control group (existing landing page) or treatment group (new landing page)
- Landing Page: The version of the webpage a user is exposed to—either the "old" or "new" design.
- Time Spent on the Page: The duration (in minutes) a user spends interacting with the landing page.
- Converted: A binary variable indicating whether a user subscribed to the news portal (1 = Yes, 0 = No).
- Language Preferred: The language selected by the user to view the landing page.

References

Great Learning. (n.d.) *Business Statistics*. **Great Learning**.

https://olympus.mygreatlearning.com/courses/124964/modules/items/6397916?pb_id=18483



Happy Learning !

