Project Title E-news Express Project

Module: Business Statistics

Date: 13-04-2025

The E-news Express project analyzed user interactions with old and new landing pages to determine effectiveness in gathering new subscribers.

This statistical analysis at a 5% significance level explored

1. Time spent on pages - **One-tailed Independent Samples t-test**

1. Conversion rates - **One-tailed Two-Proportion Z-test**

2. Language preferences - **Chi-Square Test of Independence**

3. User behaviour to guide business decisions for the online news portal - **One-Way ANOVA**

Alan McGirr

Contents / Agenda

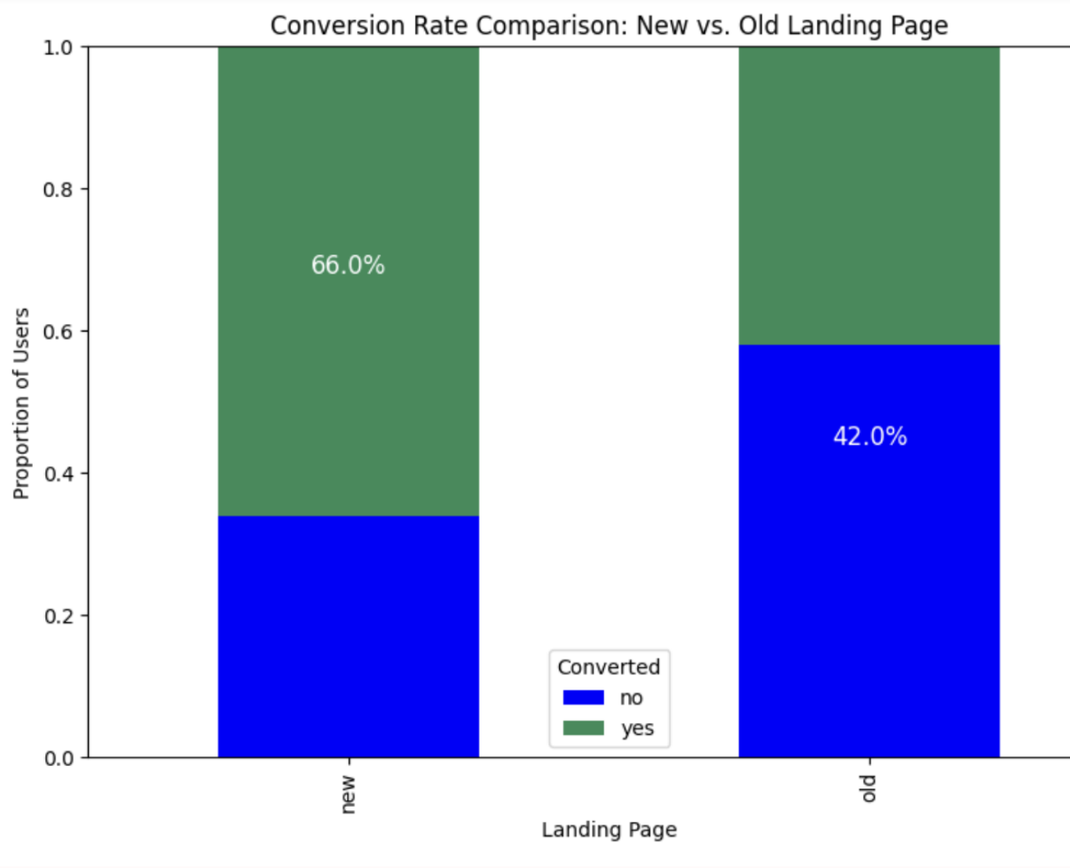
Executive Summary

New Landing Page Outperforms the Old

The new landing page demonstrates a significantly higher conversion rate (66%) compared to the old landing page (42%), with a statistically significant difference (p-value = 0.008, α = 0.05).

Language Preference Does Not Impact Conversion

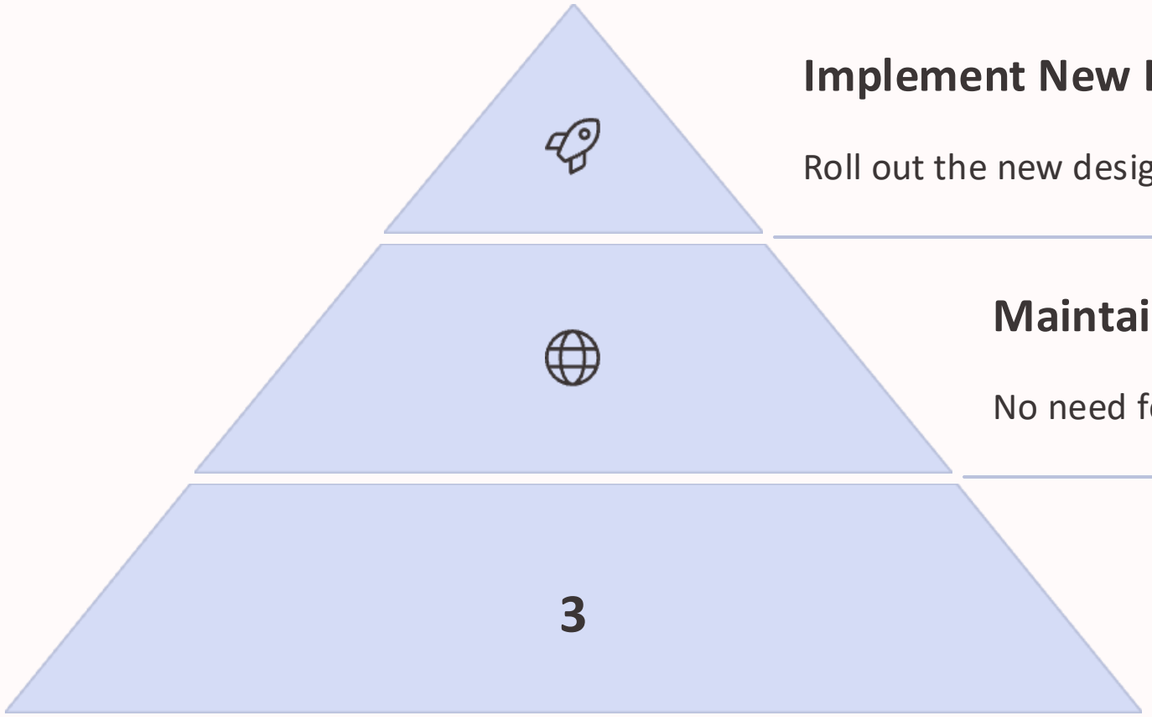
There is no statistically significant relationship between a user's preferred language (English, French, Spanish) and their conversion status (p-value = 0.213, α = 0.05).



Time Spent on New Page Consistent Across Languages

The mean time spent on the new landing page is similar across English (6.66 minutes), French (6.20 minutes), and Spanish (5.84 minutes) users, with no statistically significant difference (p-value = 0.432, α = 0.05).

Business Recommendations



Implement New Landing Page

Roll out the new design across all user segments

Maintain Unified Approach

No need for language-specific optimizations

Continue Monitoring Performance

Track metrics to ensure sustained improvement

Since the new landing page performs better overall and user engagement (time spent) and conversion rates do not significantly vary by language, E-news Express should prioritize a full implementation of the new page across all user segments. Efforts should focus on further optimizing the new page's design to sustain and enhance its conversion performance.

Further Analysis Opportunities

User Demographics

Investigate how age, location, and other demographic factors might influence conversion rates and engagement

Page Elements

Analyze which specific elements of the new page design contribute most to improved performance

Outlier Analysis

Explore why some users spend significantly more or less time on the page to identify potential UX improvements

Device Compatibility

Examine if the new page performs equally well across different devices and screen sizes

Conclusion

Proven Effectiveness

The new landing page significantly outperforms the old one in both engagement time and conversion rate

Continuous Improvement

Further analysis can identify additional optimization opportunities

Language Consistency

User behavior is consistent across language preferences, simplifying implementation strategy

Clear Path Forward

E-news Express should proceed with full implementation of the new landing page design

Data Dictionary

user_id

Unique 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

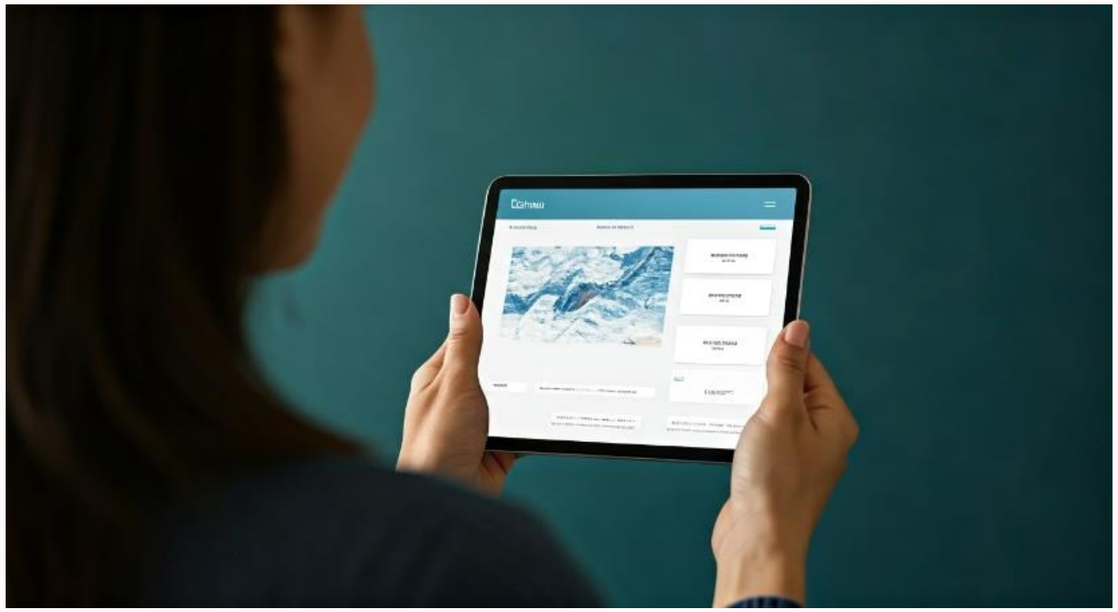
language_preferred

Language chosen by the user to view the landing page

Business Context

The advent of e-news, or electronic news, portals has offered us a great opportunity to quickly get updates on the day-to-day events occurring globally. The information on these portals is retrieved electronically from online databases, processed using a variety of software, and then transmitted to the users.

There are multiple advantages of transmitting new electronically, like faster access to the content and the ability to utilize different technologies such as audio, graphics, video, and other interactive elements that are either not being used or aren't common yet in traditional newspapers.



E-news Express, an online news portal, aims to expand its business by acquiring new subscribers. With every visitor to the website taking certain actions based on their interest, the company plans to analyze these actions to understand user interests and determine how to drive better engagement.

Key Questions

Do users spend more time on the new landing page?

Analyze if the redesigned page keeps visitors engaged longer

Is the conversion rate higher for the new page?

Determine if the new design leads to more subscriptions

Does conversion depend on preferred language?

Investigate if language preference affects subscription decisions

Is time spent consistent across language users?

Examine if different language users engage similarly with the new page

Solution Approach / Methodology

To address the key questions, a structured statistical analysis approach was employed, combining exploratory data analysis (EDA), visualizations, and hypothesis testing.

Solution - General Approach

Data Preparation

The dataset was filtered and segmented as needed for each question. For example, Question 2 used the full dataset (control and treatment groups), while Questions 3 and 4 focused on specific subsets (e.g., new page users for Question 4).

Exploratory Data Analysis (EDA)

Descriptive statistics (e.g., means, proportions) and visualizations (e.g., bar charts, boxplots) were used to understand the data distributions and identify patterns before formal testing.

Hypothesis Testing

Appropriate statistical tests were selected based on the data types (categorical or continuous) and the number of groups being compared. Each test was conducted at $\alpha = 0.05$, with p-values used to make decisions about the null hypotheses.

Business Recommendations

Conclusions from the statistical tests were translated into actionable insights for E-news Express to improve its landing page strategy.

Solution - Detailed Methodology

Exploratory Data Analysis (EDA)

Univariate and bivariate visualizations (histograms, boxplots, bar charts)

Summary statistics and initial insights

Data Preparation

Filtered data for new vs. old page groups

Grouped users by preferred language

Extracted conversion status and time spent values

Statistical Hypothesis Testing

Independent t-test: To compare average time spent between old and new landing pages

Two-proportion z-test: To compare conversion rates between page versions

Chi-square test: To test independence between conversion status and language

One-way ANOVA: To compare mean time spent across language groups (on the new page)

All tests were performed after checking key assumptions (normality and equal variances using Shapiro-Wilk and Levene's tests)

Experimental Design

Random Selection

100 users were randomly selected and divided equally into two groups

Control vs Treatment

The existing landing page was served to the first group (control group) and the new landing page to the second group (treatment group)

Data Collection

Data regarding user interaction with both versions of the landing page was collected for analysis

Data Overview

Dataset Shape

100 rows (users) and 6 columns (variables)

Data Quality

- No missing values
- No duplicate records
- Clean data ready for analysis

Key Statistics

Time spent on page:

- Mean: 5.38 minutes
- Minimum: 0.19 minutes
- Maximum: 10.71 minutes
- Standard deviation: 2.38 minutes

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
user_id	100.0	NaN	NaN	NaN	546517.0	52.295779	546443.0	546467.75	546492.5	546567.25	546592.0
group	100	2	control	50	NaN	NaN	NaN	NaN	NaN	NaN	NaN
landing_page	100	2	old	50	NaN	NaN	NaN	NaN	NaN	NaN	NaN
time_spent_on_the_page	100.0	NaN	NaN	NaN	5.3778	2.378166	0.19	3.88	5.415	7.0225	10.71
converted	100	2	yes	54	NaN	NaN	NaN	NaN	NaN	NaN	NaN
language_preferred	100	3	Spanish	34	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Univariate Analysis General Observations

Balanced groups and pages:

- The group and landing page are evenly split (50/50), so the experiment was well designed.

Balanced language preference:

- There are roughly equal numbers of users in Spanish, French, and English—this will help ensure fairness when comparing languages later.

Conversion Rate is slightly skewed:

- More users said “yes” to being converted (54%) than “no” (46%). This could be good news for the business, but it needs hypothesis testing.

Time spent distribution is fairly symmetric:

- Based on the histogram and boxplot, most users spend 3 to 8 minutes on the page, with very few outliers.

No clear visual class imbalance:

- This supports fair hypothesis testing.

Group Distribution – Balanced Groups & Pages

Random Selection

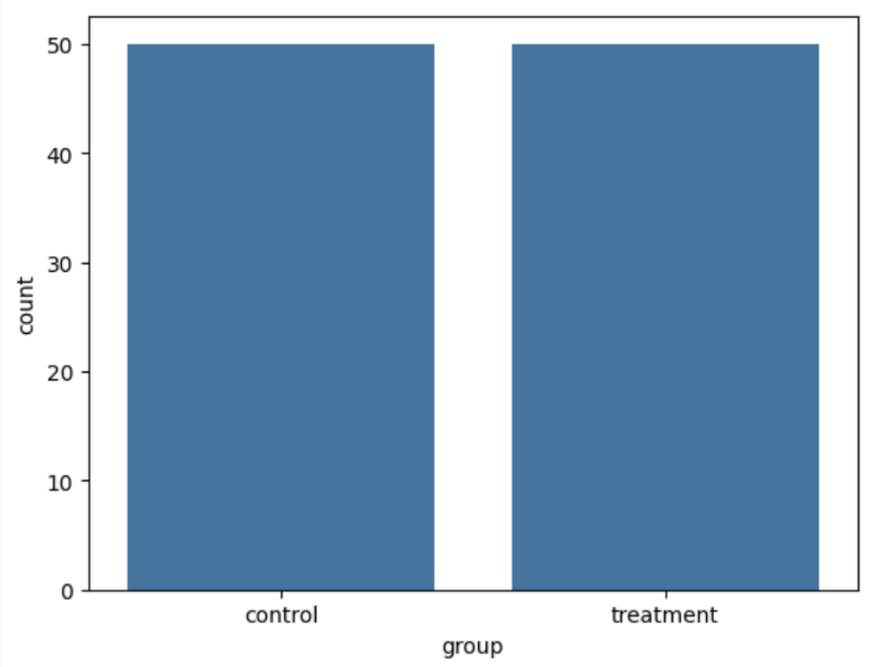
100 users were randomly selected and divided equally into two groups

Control vs Treatment

The existing landing page was served to the first group (control group) and the new landing page to the second group (treatment group)

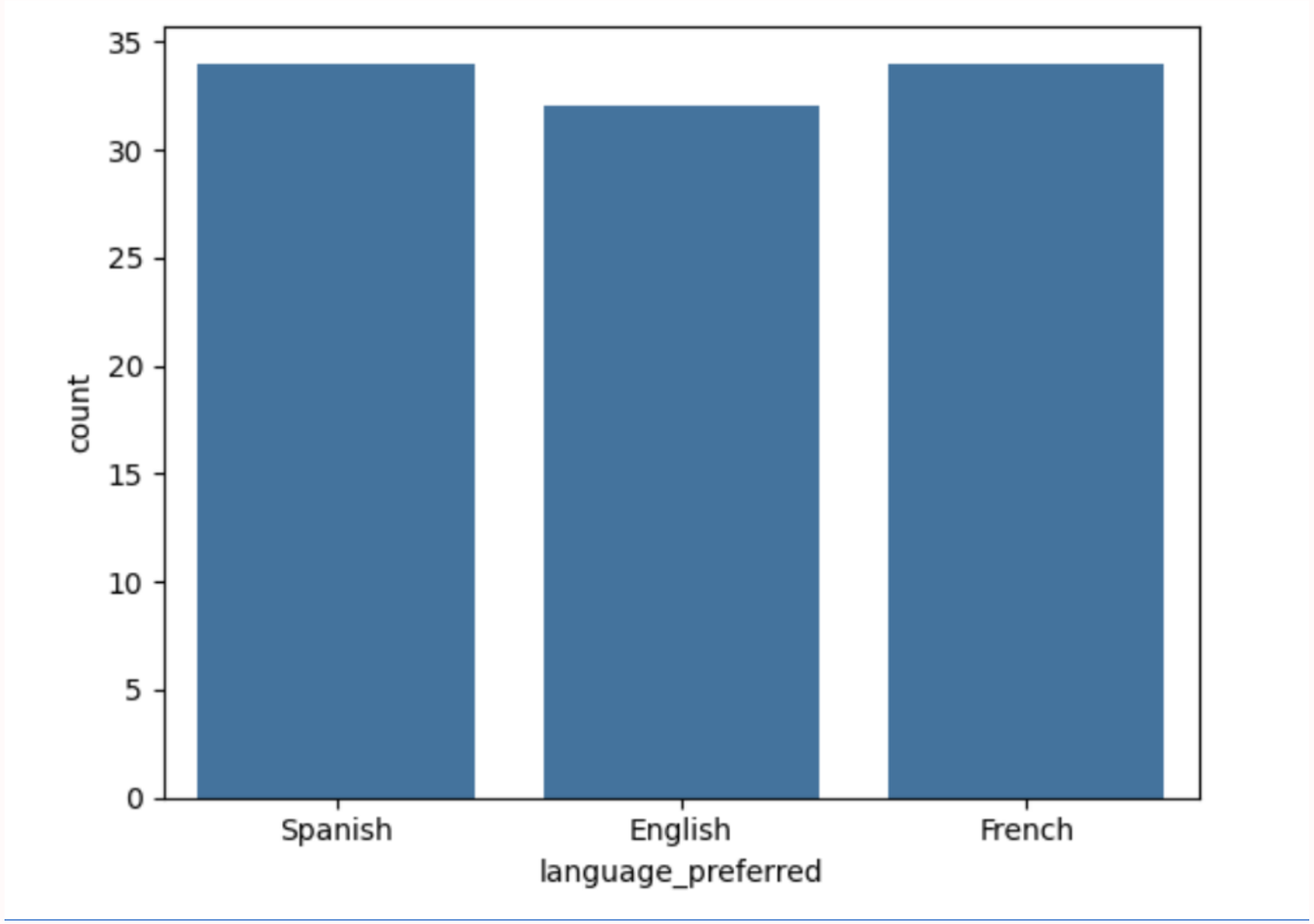
Data Collection

Data regarding user interaction with both versions of the landing page was collected for analysis



The experiment was designed with a perfect 50/50 split between the control group (old landing page) and the treatment group (new landing page), ensuring a balanced representation for statistical comparison.

Balanced Language Preference Distribution

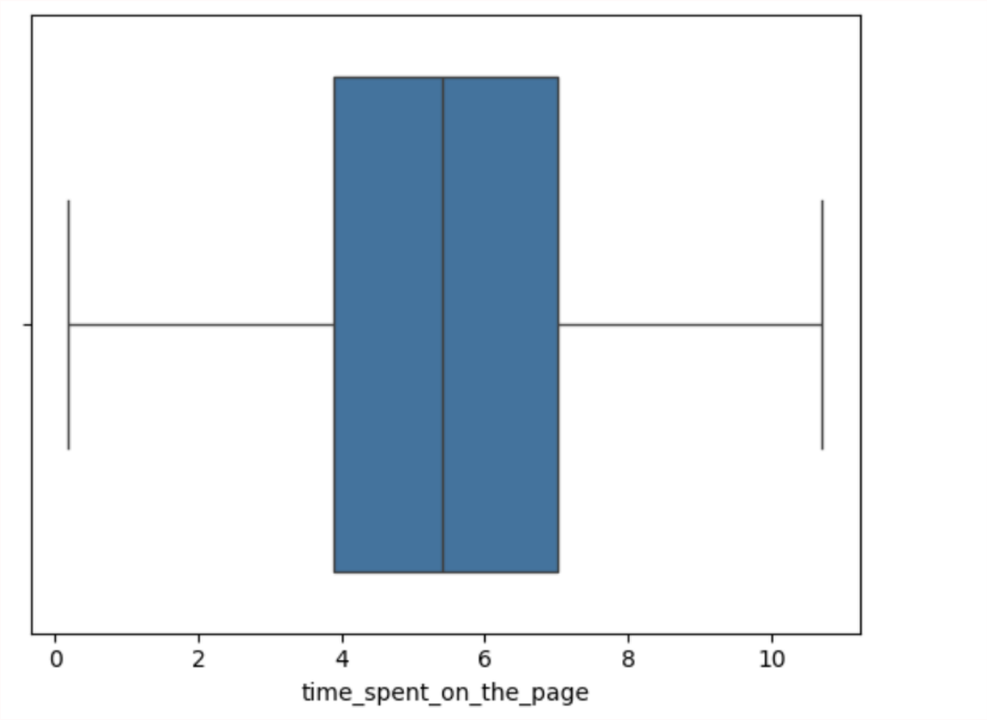


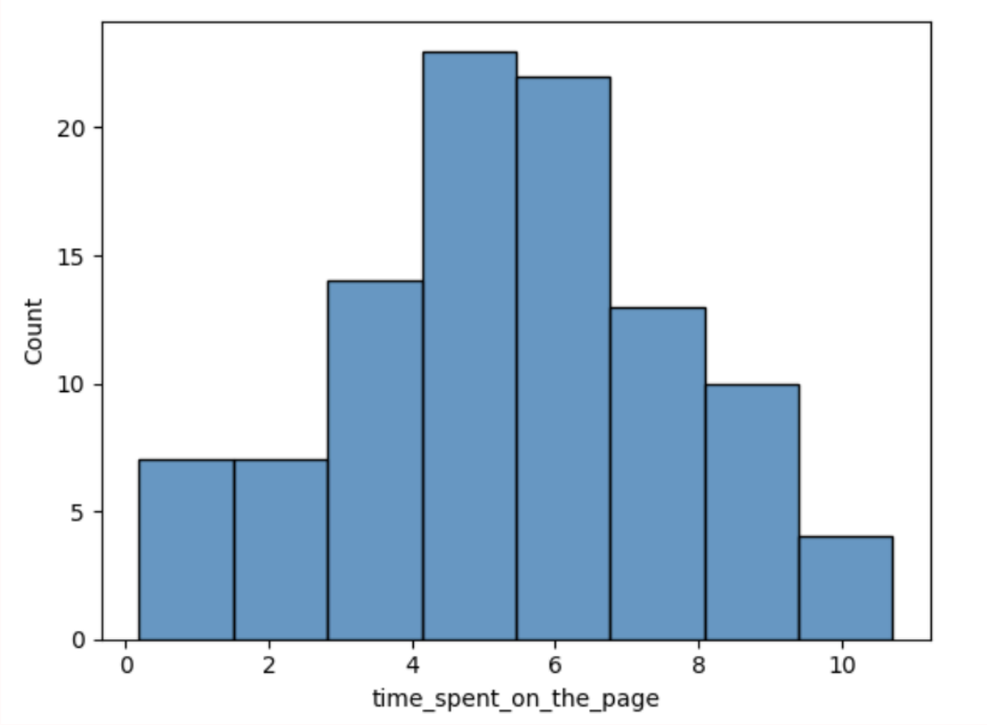
The distribution of language preferences among users is relatively balanced, with Spanish and French each representing 34% of users, while English accounts for 32%. This balanced distribution helps ensure fair comparison across language groups in our analysis.

Time Spent Distribution

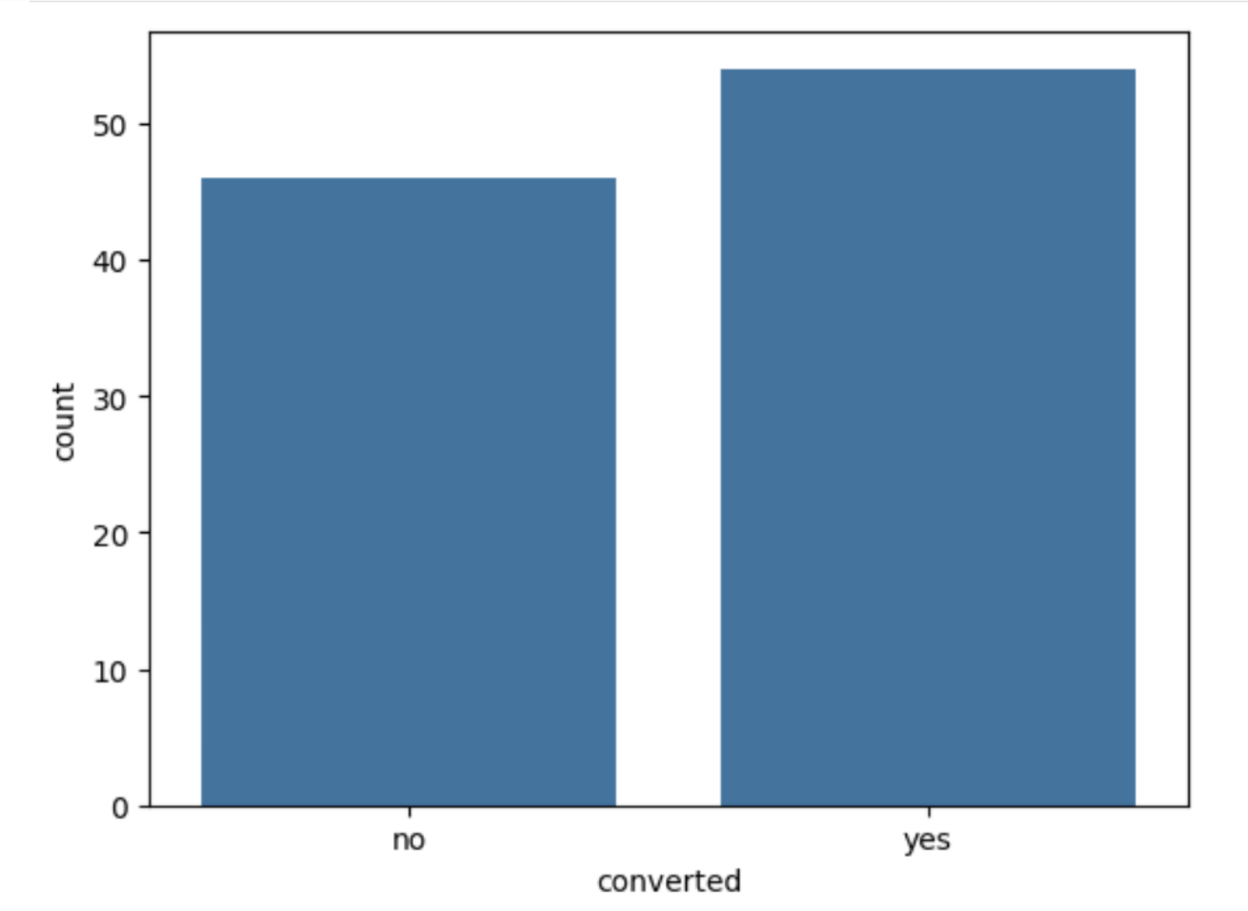
The distribution of time spent on the page reveals two distinct user behaviors: a smaller group of users who spend a short time (1-2 minutes) and a larger group who spend a moderate time (5-6 minutes), with a median of around 5 minutes.

This bimodal pattern suggests that E-news Express has a mix of users—those who quickly skim the page and leave and those who engage more deeply with the content. The lack of extreme outliers indicates that the time spent by users, even at the extremes, is within the expected range.



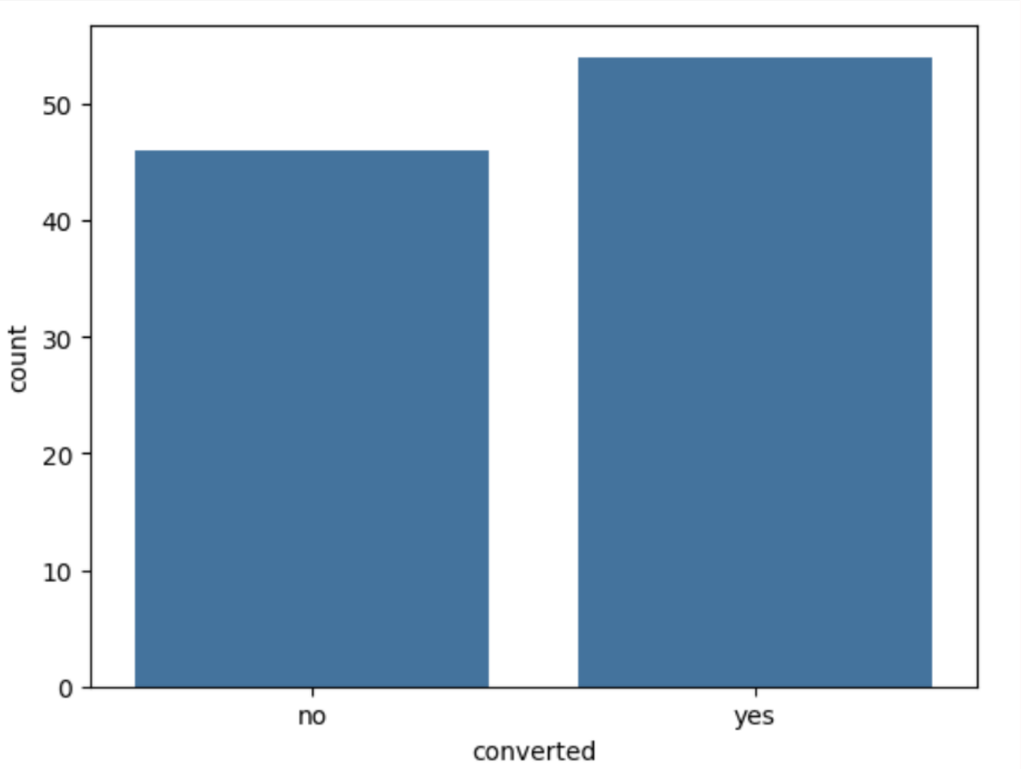


Time Spent by Landing Page Type



Users spent significantly more time on the new landing page (an average of 6.23 minutes) than on the old landing page (an average of 4.52 minutes). This visual comparison suggests that the new design is more engaging, keeping users on the page longer.

Conversion Status Distribution - Conversion Rate is slightly skewed



54%

Converted

Users who subscribed

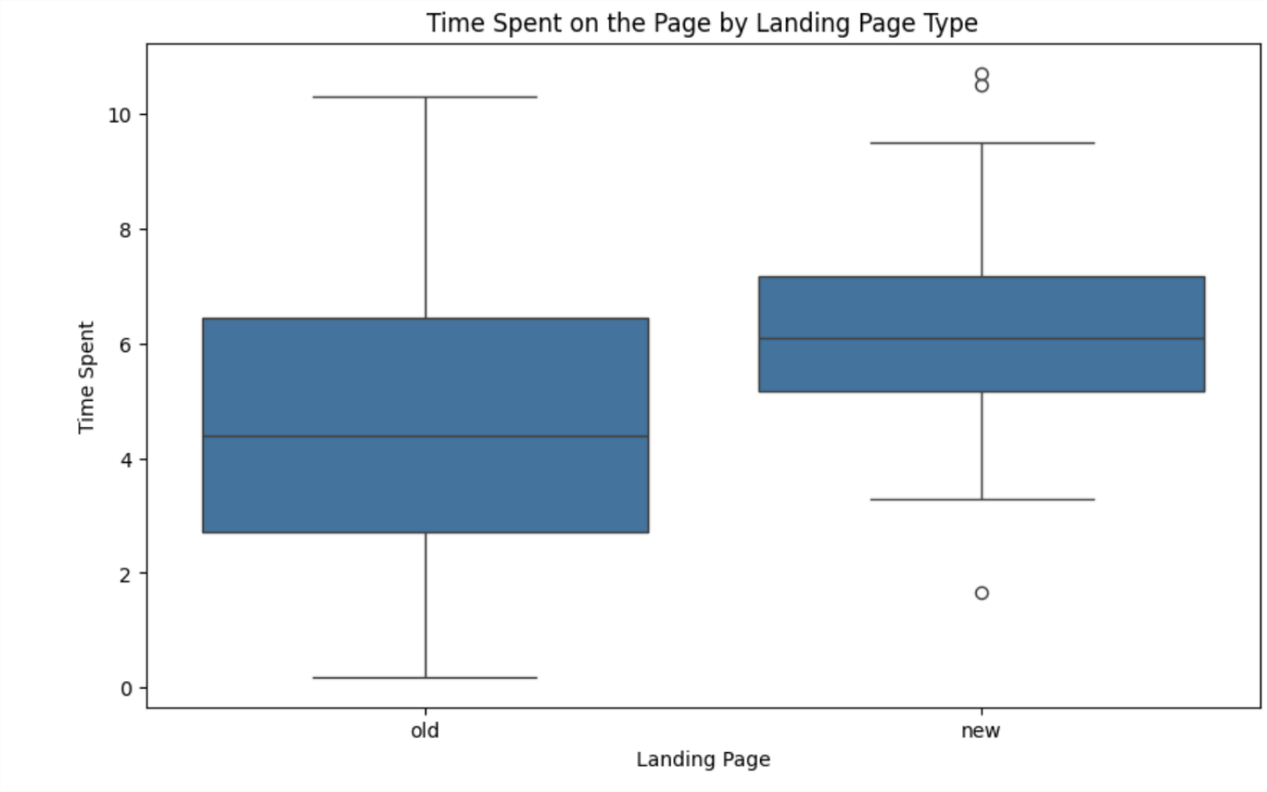
46%

Not Converted

Users who did not subscribe

Overall, the experiment showed a positive conversion rate with more users subscribing (54%) than not subscribing (46%). This suggests generally effective landing pages, but further analysis is needed to determine which version performs better.

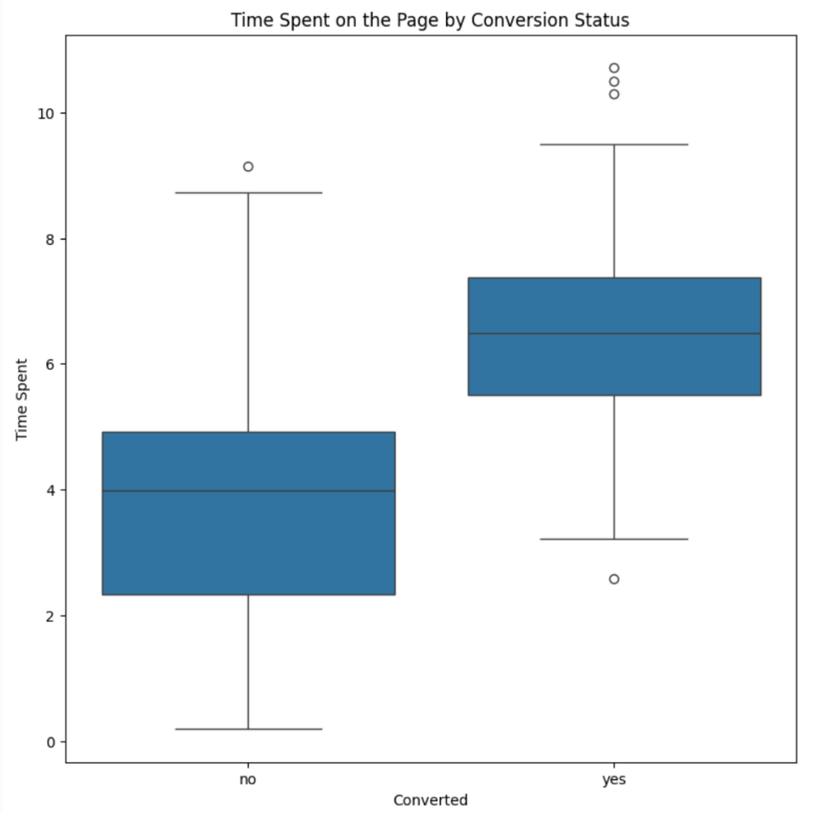
Bivariate Analysis - Landing Page vs Time Spent on the Page



Observation	Interpretation
Higher Median for New Page	Users tend to spend more time on the new landing page (median shifted upward).
Less Spread in New Page	The interquartile range (IQR) is narrower, indicating more consistent engagement on the new page.
Outliers	A few users spent unusually high or low time (dots above/below the whiskers), which is normal in user behavior data.
Higher Overall Distribution	The new page's entire box is shifted upward compared to the old — not just the median, but also Q1 and Q3.

Overall conclusion. The new landing page appears to increase average user engagement and makes user behavior more consistent. This visual supports the t-test result which showed that the increase in time spent was statistically significant.

Bivariate Analysis - Conversion Status vs Time Spent on the Page



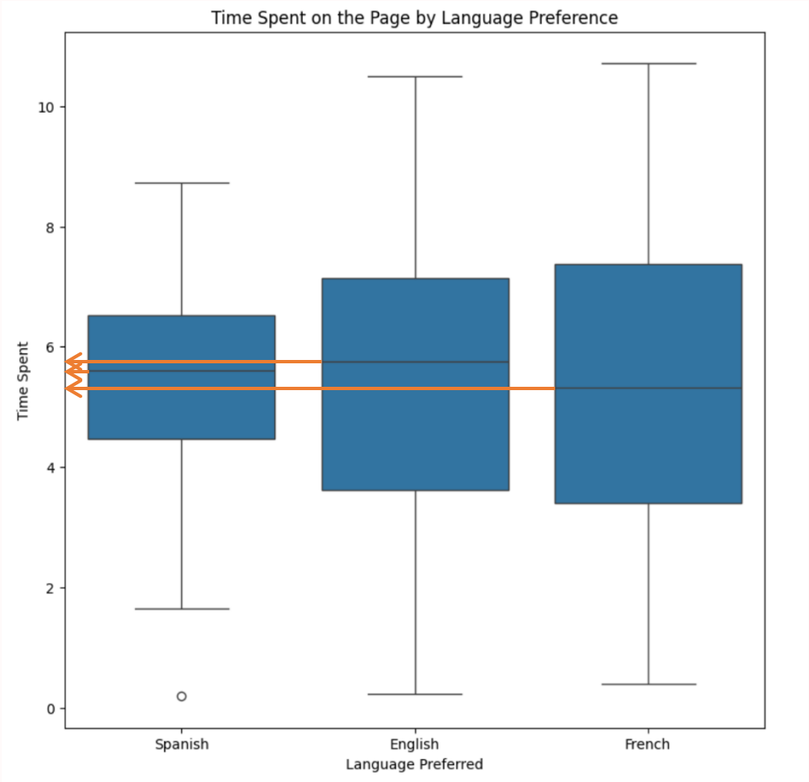
Observation	Interpretation
Higher Median for Converters	Users who converted spent more time on the page (median approx. 6.8) than those who didn't (Approx. 4.0).
Greater Upper Range in 'Yes' Group	Converted users not only spent more time, but also had more high-end engagement (Q3 is higher).
Tighter IQR in 'Yes' Group	Time spent by converters is more concentrated, suggesting a typical engagement pattern before conversion.
Outliers in Both Groups	Normal behavior — but doesn't distort the conclusion due to the robust medians and IQR.

Overall Conclusion: There is a positive association between time spent on the page and the likelihood of conversion. Users who spend more time on the landing page are more likely to convert.

This supports potential strategies like:

- Optimising content to hold attention longer.
- Tracking scroll depth or click interactions as part of a conversion funnel.

Bivariate Analysis - Language Preferred vs Time Spent on the Page

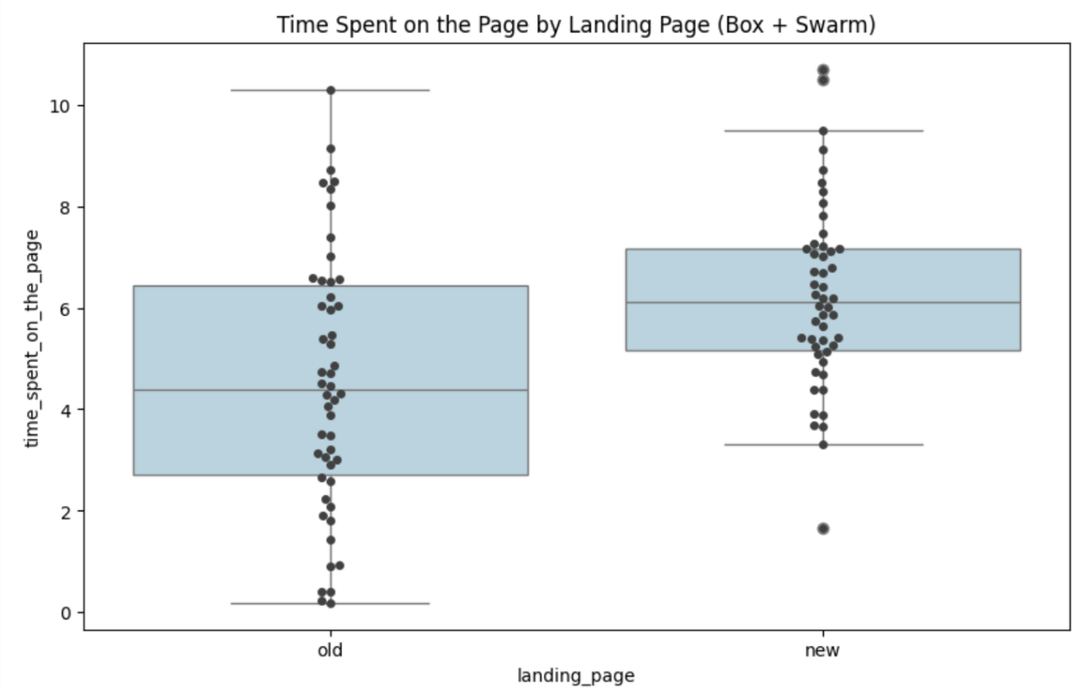


All three language groups have **similar median time spent** (5.5–5.8 minutes).

- The distributions overlap significantly, and no group stands out clearly.
- French and English have more variability, while Spanish shows more consistent behavior.

Language	Median	Spread	Outliers	Observation
Spanish	Slightly higher median (5.6)	Narrower spread	Few low outliers	Users are more consistent in their time spent
English	Median 5.8	Widespread	Outliers on both ends	Broad range in user engagement
French	Median slightly lower (5.4–5.5)	Similar spread to English	Outliers	Slightly more variability in behavior

Bivariate Analysis - Swarm plot overlay on Boxplot for clarity



Feature	Observation	Interpretation
Median	Higher for new page	On average, users spend more time on the new landing page
IQR (box width)	Narrower in the new group	Time spent is more consistent across users on the new page
Outliers	More in old page	Greater variability in user engagement for the old version
Swarm (dots)	New page has more tightly packed dots around the center	Most users cluster around the 6-minute mark for the new page, while the old page is more spread out
Overall shape	The new page group is shifted upward	Suggests better engagement with the new design

This box + swarm plot clearly supports the idea that the new landing page performs better in terms of user engagement. It visually shows a higher central tendency and tighter clustering in the new group, but we will still conduct statistical analysis to prove this.

Q1 Hypothesis Test: Do the users spend more time on the new landing page than the existing landing page?

Null Hypothesis (H_o)

The average time spent on the new page is less than or equal to that of the old page

Test Selection

Independent Samples t-Test (One-Tailed) with unequal variances

Results

p-value = 0.000139, which is less than $\alpha = 0.05$

Null Hypothesis (H_o): (The Status Quo)

The average time spent on the new page is less than or equal to that of the old page.

Alternative Hypothesis (H_1): (Claim)

The average time spent on the new page is greater than that of the old page.

Link to Appendix slide on details of the test performed

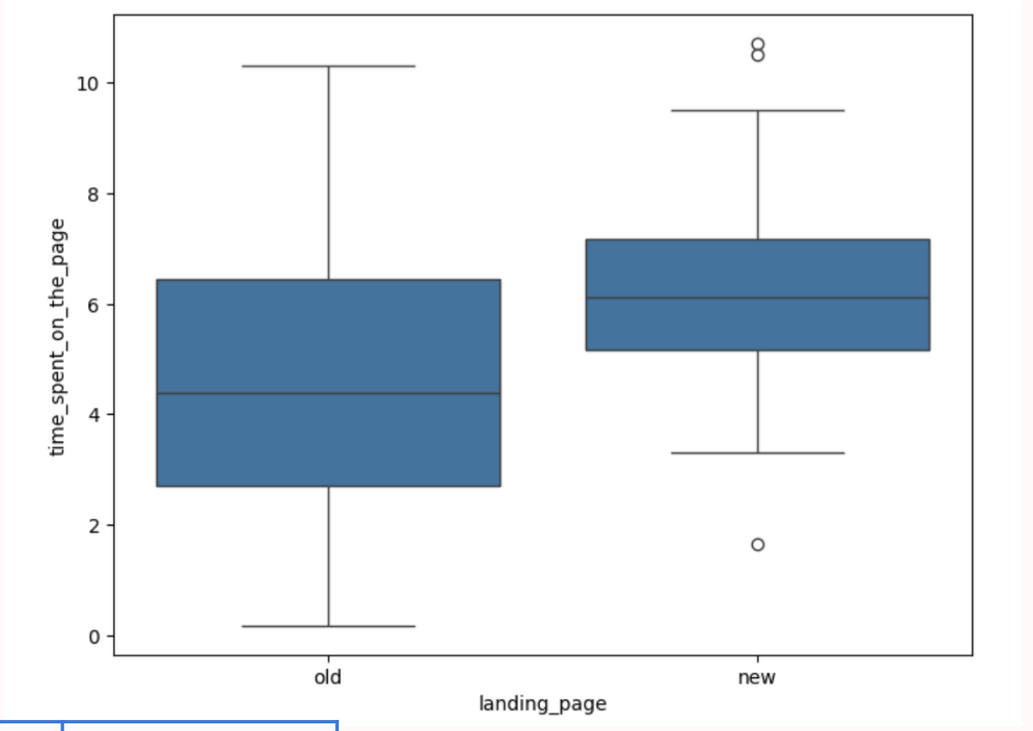
Q1 – Hypothesis Test: Do the users spend more time on the new landing page than the existing landing page?

This is a one-tailed test concerning the difference between two population means from two **independent** groups (users who saw the old landing page vs. those who saw the new one).

Since the population standard deviations are **unknown** and the sample sizes are relatively small, we use the:

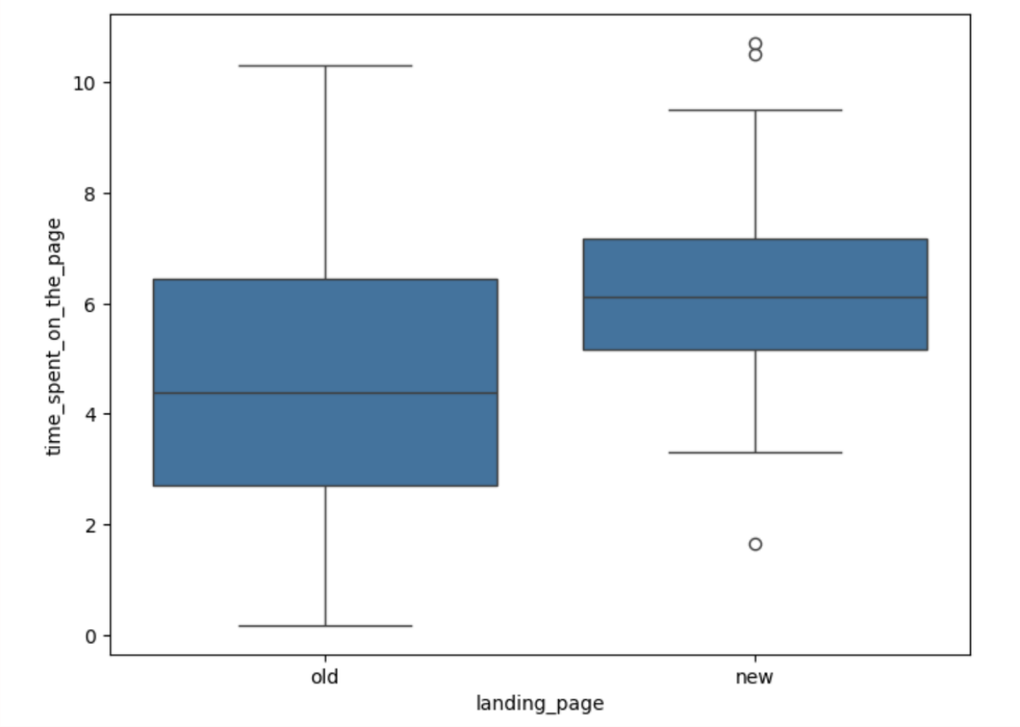
- **Independent Samples t-Test (One-Tailed)**

This test will help determine whether the **mean time spent on the new landing page** is **significantly greater** than that on the old page.



Significance of the test	Assumptions met? Yes	Test Statistic Distribution
Test for population mean $H_0 : \mu = \mu_0$	• Continuous data • Normally distributed population and sample size < 30 • Unknown population standard deviation • Random sampling from the population	t distribution (The test is also known as One-sample t-test)

Time Spent Analysis Conclusion



Remember that if the P Value is low, the Null hypothesis must go.

Statistical Finding

We reject the null hypothesis. There is statistically significant evidence that users spend more time on the new landing page compared to the old one (p-value = 0.000139 < α = 0.05).

Business Insight

The new landing page is more engaging and effective at holding user attention. This supports moving forward with its broader rollout to enhance user engagement.

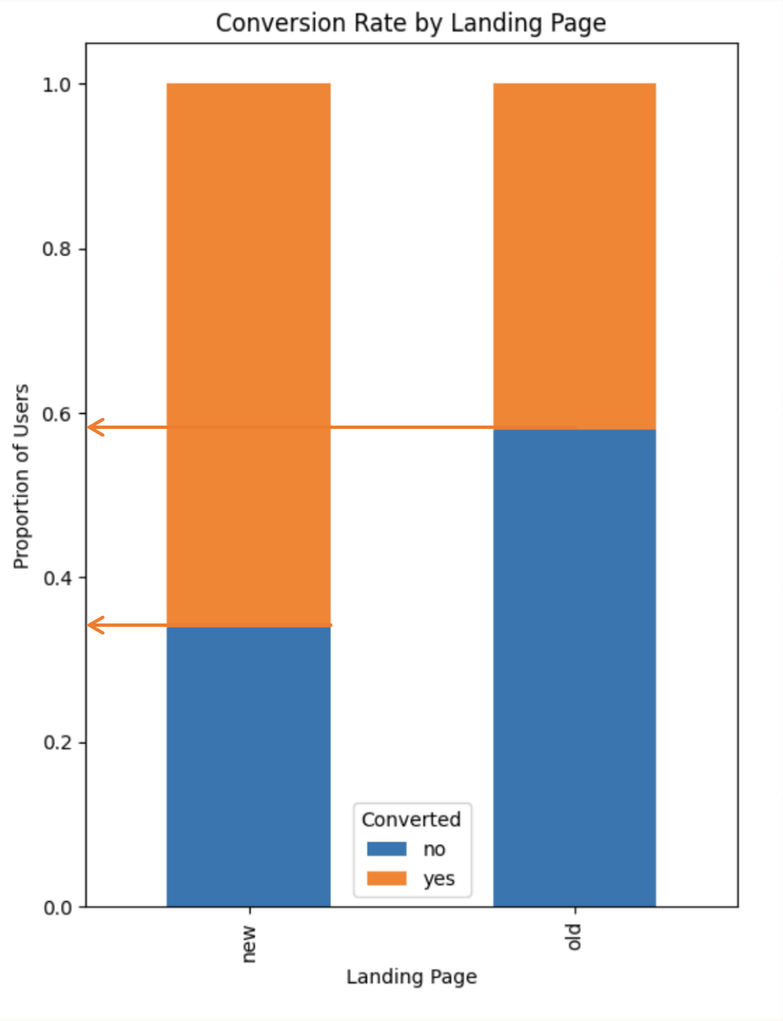
Standard Deviations

New page: 1.82 minutes

Old page: 2.58 minutes

The lower standard deviation for the new page indicates more consistent user engagement.

Q2 Hypothesis Test: 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?



The new landing page shows a substantially higher conversion rate (66%) compared to the old landing page (42%). This visual comparison suggests the new design is more effective at converting visitors to subscribers. Which will be formally tested.

Selecting appropriate test:

We are comparing the conversion rates (proportions of users who were converted) between two independent groups — users who saw the new landing page and those who saw the old one.

Since the outcome is binary (Yes or No), and we are comparing two independent proportions, the appropriate statistical test is:

Two-Proportion Z-Test (One-Tailed)

This will allow us to test if the conversion rate for the new page is significantly greater than that for the old page.

Assumptions are met?

- Binomially distributed populations
- Independent populations
- Random sampling from the populations
- When both mean (np) and n(1-p) are greater than or equal to 10. The binomial distribution can be approximated by a normal distribution

Hypothesis Test: Conversion Rate

Null Hypothesis (H_0)

There is no improvement in conversion rate on the new page

Test Selection

Two-Proportion Z-Test (One-Tailed)

Results

Z-statistic = 2.41, p-value = 0.0080, which is less than α = 0.05

Null Hypothesis (H_0): (The Status Quo)

There is no improvement in conversion rate on the new page

Alternative Hypothesis (H_1): (Claim)

The new page has a higher conversion rate than the old page..

Link to Appendix slide on details of the test performed

Conversion Rate Analysis Conclusion

Statistical Finding

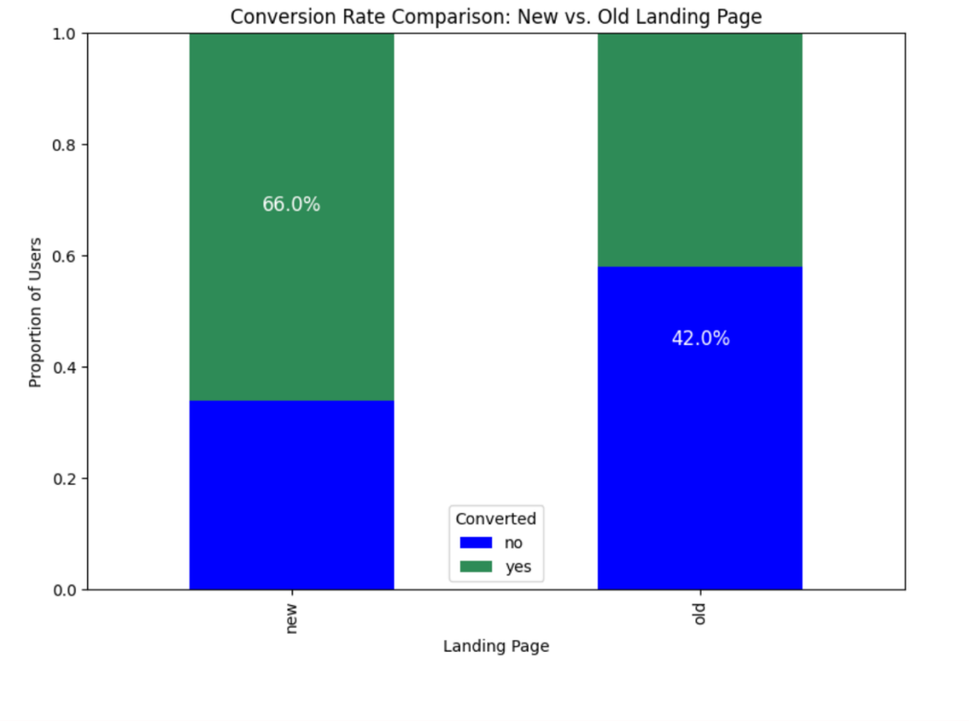
We reject the null hypothesis. There is statistically significant evidence that the conversion rate for the new landing page is greater than that for the old landing page (p-value = 0.0080 < α = 0.05).

Business Recommendation

The new landing page is performing significantly better in terms of conversions. E-news Express should consider a wider rollout of the new landing page to enhance user engagement and increase subscription rates.

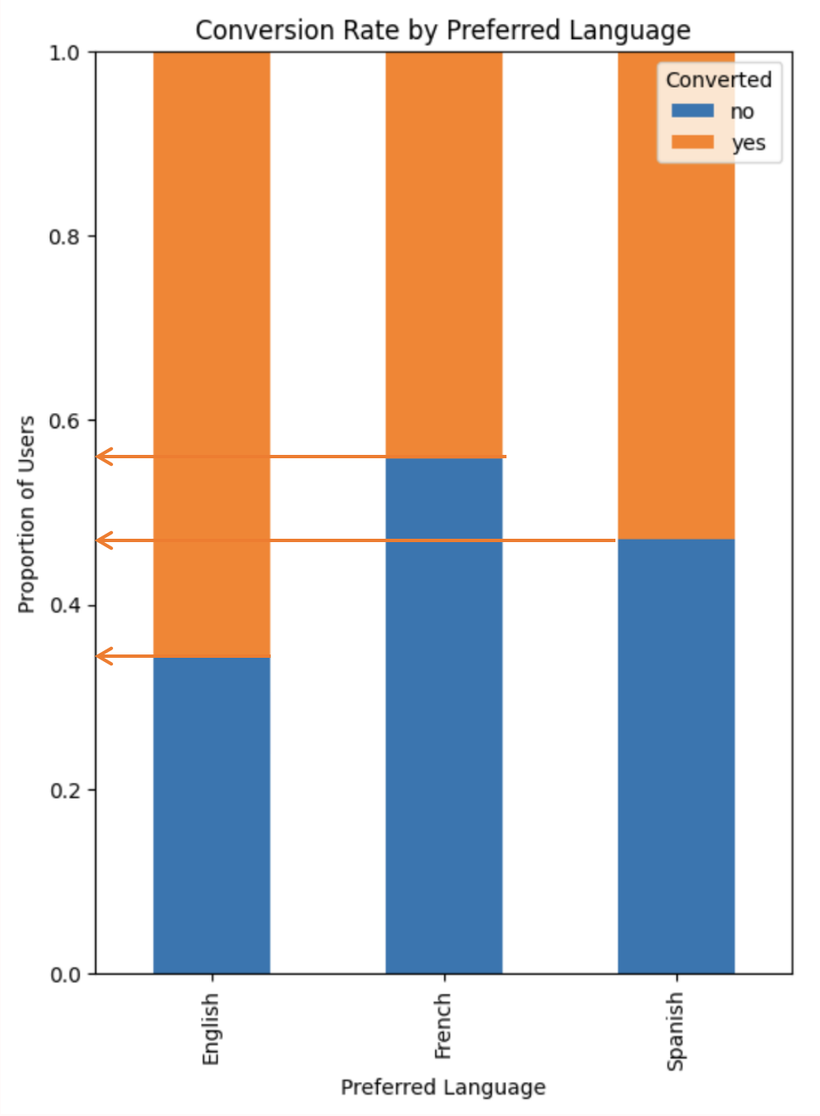
Impact Measurement

The new page shows a 24 percentage point improvement in conversion rate, which represents a 57% increase relative to the old page's performance.



Remember that if the P-Value is low, the Null hypothesis must go.

Q3 – Hypothesis Test: Is the conversion and preferred language are independent or related?



The chart shows variation in conversion rates across language preferences:

Each bar represents a preferred language group, broken down by the proportion of users who converted (yes) and did not convert (no).

- English appears to have the highest conversion rate (more orange, less blue).
- French shows the lowest conversion rate (more blue, less orange).
- Spanish is somewhere in between, leaning toward a low conversion rate. There are visible differences in conversion rates across the language groups, which suggests a possible dependency between converted and language_preferred.

But — to confirm whether these differences are statistically significant, we need to do some tests.

Hypothesis Test: Language vs Conversion

Null Hypothesis (H_0)

Conversion status is independent of preferred language

Test Selection

Chi-Square Test of Independence

Results

p-value = 0.213, which is greater than $\alpha = 0.05$

Null Hypothesis (H_0): (The Status Quo)

Conversion status is independent of preferred language.

Alternative Hypothesis (H_1): (Claim)

The new page has a higher conversion rate than the old page..

Link to Appendix slide on details of the test performed

Contingency Table: Language vs Conversion

Significance of the test	Assumptions met? Yes	Test Statistic Distribution
In a contingency table **H_0 : The row and column variables are independent**	• Categorical variables • Expected value of the number of sample observations in each level of the variable is at least 5 • Random sampling from the population	Chi Square distribution (The test is also known as **Chi-square test of independence**)

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

This contingency table shows the distribution of conversion status across different language preferences.
While there appear to be some differences in conversion rates (English: 66%, Spanish: 53%, French: 44%), we need to test if these differences are statistically significant.

Step 5 Calculate the p-value:

```
# Import the chi2_contingency function from scipy.stats
from scipy.stats import chi2_contingency

# Run the Chi-Square test
chi2, p_value, dof, exp_freq = chi2_contingency(contingency_table)

print('The p-value is', p_value)
```

```
The p-value is 0.2129888748754345
```

Language vs Conversion Analysis Conclusion

Statistical Finding

We fail to reject the null hypothesis.

There is no statistically significant relationship between a user's preferred language and whether or not they converted (p-value = 0.213 > α = 0.05).

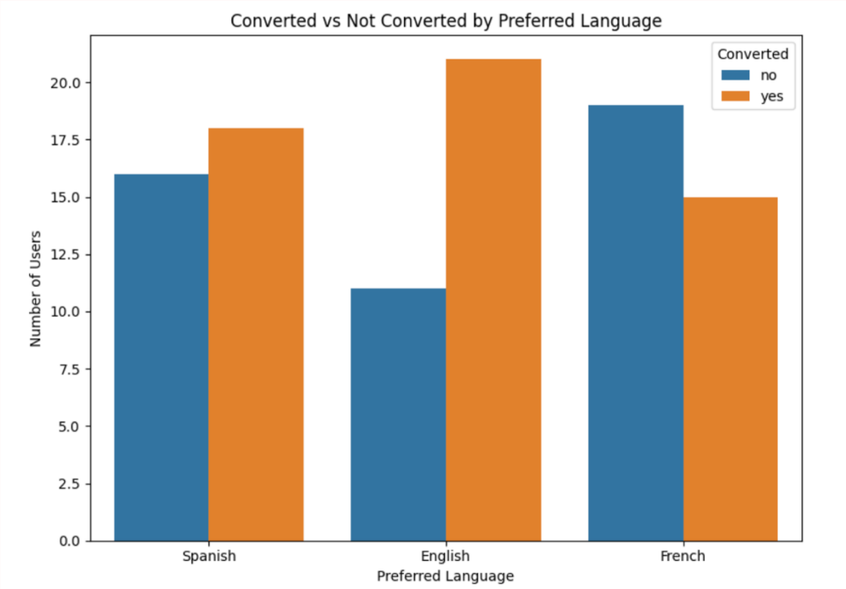
Business Recommendation

Resources should focus on improving overall page effectiveness rather than tailoring landing pages by language based on conversion data alone.

The apparent differences in conversion rates across languages are not statistically significant.

Further Consideration

While conversion rates don't significantly differ by language, other metrics like user satisfaction or specific content engagement might still vary and could be worth investigating separately.

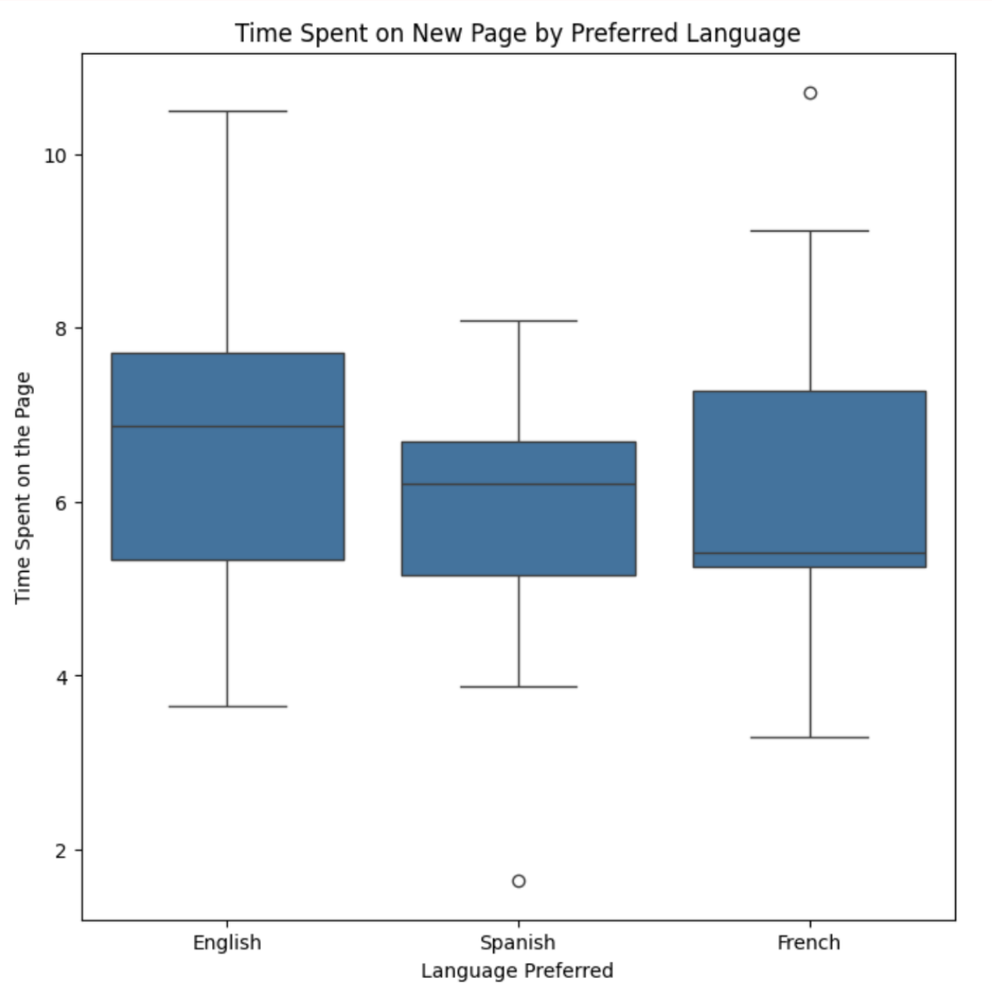


Visual Insight

The grouped bar chart shows some variation in conversion rates across languages — with English users converting more frequently than French or Spanish users.

However, this **visual difference is not statistically significant**, as confirmed by the chi-square test (p = 0.213).

Q4 – Hypothesis Test: Is the time spent on the new page same for the different language users?



The boxplot visualisation shows the distribution of time spent on the new landing page across language groups.

- While English users appear to spend more time on average, there is considerable overlap in the distributions.
- Spanish users show the least variability (tightest box).
- While French users show more spread with one high outlier.

Numerical insights on average

- English (6.66 minutes), followed by French (6.20 minutes) and Spanish (5.84 minutes) users.

While there are visible differences, statistical testing is needed to determine if these differences are significant.

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

Null Hypothesis (H_0)

Mean time spent is the same across all language groups

Test Selection

One-Way ANOVA (Analysis of Variance)

Results

F-statistic = 0.8544, p-value = 0.432, which is greater than $\alpha = 0.05$

Null Hypothesis (H_0): (The Status Quo)

Mean time spent is the same across all language groups

Alternative Hypothesis (H_1): (Claim)

At least one group has a mean time spent that is different.

Link to Appendix slide on details of the test performed

Selecting the Appropriate Test

We are comparing the mean time spent on the new landing page across three independent language groups (English, French, and Spanish).

The response variable (time_spent_on_the_page) is continuous, and the grouping variable (language_preferred) is categorical with more than two groups.

Therefore, the appropriate test is:

One-Way ANOVA (Analysis of Variance)

This test will help us determine whether there is a statistically significant difference in mean time spent across the different language groups.

We will proceed to check the assumptions of ANOVA (normality and equal variance) before conducting the test.

Significance of the test	Assumptions met? We will need to verify	Test Statistic Distribution
Test for means for more than two populations H_0 : All population means are equal	• The populations are normally distributed • Samples are independent simple random samples • Population variances are equal	F distribution (The test is also known as One-way ANOVA F-test)

Check Assumptions

Shapiro-Wilk test

“Shapiro-Wilk’s test is applied to the response variable to validate the normality assumption. The Shapiro-Wilk test examines if the population from which the sample is drawn follows a normal distribution or not.”

I applied the Shapiro-Wilk test to assess whether the ”time_spent_on_the_page” variable is normally distributed within each language group

Language	p-value	Result
English	0.886	Normal
French	0.321	Normal
Spanish	0.090	Normal

All p-values are greater than 0.05, so the normality assumption for ANOVA is satisfied.

Levene’s test

“Levene's test is applied to the different groups to validate the common variance assumption. It checks whether different samples are drawn from a population with a common variance.”

The next step is to use Levene’s test to check the second key ANOVA assumption: equal variances across groups.

Levene’s Test was used to assess the assumption of homogeneity of variances across language groups.

p-value = 0.467

Since the p-value is greater than 0.05, we conclude that the variances are equal. The assumption of equal variance is satisfied.

Time Spent by Language Analysis Conclusion

```python
from scipy.stats import f_oneway

# Run the ANOVA test
f_stat, p_value = f_oneway(english_time, french_time, spanish_time)

print("ANOVA F-statistic:", round(f_stat, 4))
print("ANOVA p-value:", p_value)
```

```
ANOVA F-statistic: 0.8544
ANOVA p-value: 0.43204138694325955
```

Statistical Finding

We fail to reject the null hypothesis. There is no statistically significant difference in the mean time spent on the new landing page across English, French, and Spanish users (p-value = 0.432 > α = 0.05).

Business Recommendation

All language groups appear to engage similarly with the new landing page. No localization changes are needed based on time-spent behavior, as the differences observed are not statistically significant.

Assumption Verification

The ANOVA assumptions were verified: normality was confirmed with Shapiro-Wilk tests (all p-values > 0.05) and equal variance was confirmed with Levene's test (p-value = 0.467 > 0.05).

Key Findings Summary

Time Spent Analysis

Users spend significantly more time on the new landing page compared to the old one (p-value = 0.000139 < α = 0.05)

2 Conversion Rate Analysis

The new landing page has a significantly higher conversion rate (66%) than the old page (42%) with p-value = 0.008 < α = 0.05

Language Impact on Conversion

No statistically significant relationship between language preference and conversion status (p-value = 0.213 > α = 0.05)

Language Impact on Time Spent

No statistically significant difference in time spent on the new page across language groups (p-value = 0.432 > α = 0.05)

Considerations on the Limitations of the study

Sample Size

The study used 100 users (50 per group), which may limit the statistical power to detect smaller effects, especially in subgroup analyses

Time Frame

The analysis represents a snapshot in time and may not account for seasonal variations in user behavior

Limited Variables

The study focused on a few key metrics and may not capture all factors influencing user decisions

Geographic Considerations

The analysis did not account for potential regional differences in user behavior beyond language preference

APPENDIX

Question 1 Hypothesis Testing Details

Step 1: Define the null and alternate hypotheses

Hypothesis: Time Spent on New vs. Old Landing Page

Business Question:
Do users spend more time on the new landing page than the old one?

Null Hypothesis (H_0):
The average time spent on the new page is less than or equal to that of the old page.

Alternative Hypothesis (H_1):
The average time spent on the new page is greater than that of the old page.

Step 2: Select Appropriate test

This is a one-tailed test concerning the difference between two population means from two **independent** groups (users who saw the old landing page vs. those wh saw the new one).

Since the population standard deviations are **unknown** and the sample sizes are relatively small, we use the:

Independent Samples t-Test (One-Tailed)

This test will help determine whether the **mean time spent on the new landing page** is **significantly greater** than that on the old page.

Step 3: Decide the significance level

As given in the problem statement, we select α = 0.05.

Step 4: Collect and prepare data

```
# Create subsetted data frame for new landing page users
time_spent_new = df[df['landing_page'] == 'new']['time_spent_on_the_page']

# Create subsetted data frame for old landing page users
time_spent_old = df[df['landing_page'] == 'old']['time_spent_on_the_page']

# Print sample standard deviations
print('The sample standard deviation of the time spent on the new page is:', round(time_spent_new.std(), 2))
print('The sample standard deviation of the time spent on the old page is:', round(time_spent_old.std(), 2))
```

```
The sample standard deviation of the time spent on the new page is: 1.82
The sample standard deviation of the time spent on the old page is: 2.58
```

Observation: Collect and Prepare Data

The sample standard deviation of time spent on the **new landing page** is **1.82**, and on the **old landing page** is **2.58**.

Since the difference in standard deviations is significant, we assume **unequal variances** for the two groups and set `equal_var = False` in our t-test.

Step 5: Calculate the p-value

```
# Import ttest_ind from scipy.stats
from scipy.stats import ttest_ind

# Perform one-tailed t-test (note: use equal_var based on step 4)
test_stat, p_value = ttest_ind(time_spent_new, time_spent_old, equal_var=False, alternative='greater')

print('The p-value is', p_value)
```

```
The p-value is 0.0001392381225166549
```

Step 6: Compare the p-value with α

The p-value is 0.000139

significance level α = 0.05

Since 0.000139 < 0.05 = True

We reject the null hypothesis.

Step 7: Draw Inference

The p-value (0.00014) is significantly less than the significance level (α = 0.05), so we **reject the null hypothesis**.

Conclusion:
There is statistically significant evidence to suggest that users **spend more time** on the **new landing page** compared to the old one.

Business Insight:
The new landing page is more engaging and may be more effective at holding user attention. This supports moving forward with its broader rollout.

Question 2 Hypothesis Testing Details

Step 1: Define the null and alternate hypotheses

H_0 : There is no improvement in conversion rate on the new page.

H_a : The new page has a higher conversion rate than the old page.

Step 2: Select the Appropriate Test

We are comparing the **conversion rates** (proportions of users who were converted) between **two independent groups** — users who saw the new landing page and those who saw the old one.

Since the outcome is **binary** (Yes or No), and we are comparing **two independent proportions**, the appropriate statistical test is:

Two-Proportion Z-Test (One-Tailed)

This will allow us to test if the **conversion rate for the new page** is **significantly greater** than that for the old page.

Assumptions are met?

Binomially distributed populations

Independent populations

Random sampling from the populations

When both mean (np) and n(1-p) are greater than or equal to 10.

The binomial distribution can be approximated by a normal distribution

Step 3: Decide the Significance Level

As given in the problem statement, we select a **significance level of**

alpha = 0.05

This means we require at least 95% confidence to conclude that the new landing page performs better than the old one.

Step 4: Collect and Prepare the Data

```
# Calculate the number of converted users in the treatment group
new_converted = df[df['group'] == 'treatment']['converted'].value_counts()['yes']

# Calculate the number of converted users in the control group
old_converted = df[df['group'] == 'control']['converted'].value_counts()['yes']  # Completed

# Total number of users in each group
n_control = df['group'].value_counts()['control']
n_treatment = df['group'].value_counts()['treatment']

print('The numbers of users served the new and old pages are {0} and {1} respectively'.format(n_treatment, n_control))
```

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

Step 5: Calculate the P Value:

```
# Import the required function for a two-proportion z-test
from statsmodels.stats.proportion import proportions_ztest

# Calculate the p-value for a one-tailed test (new > old)
test_stat, p_value = proportions_ztest([new_converted, old_converted], [n_treatment, n_control], alternative='larger')

print('The p-value is', p_value)

#Z-Statistic printed to see Standard Deviation
print('Z-statistic:', test_stat)
```

```
The p-value is 0.008026308204056278
Z-statistic: 2.4077170617153842
```

Step 6 Compare P Value with Level of significance:

The Z-statistic is **2.41**, and the corresponding **p-value is 0.0080**, which is **less than the significance level (α = 0.05)**.

```
# print the conclusion based on p-value
if p_value < 0.05:
    print(f'As the p-value {p_value} is less than the level of significance, we reject the null hypothesis.')
else:
    print(f'As the p-value {p_value} is greater than the level of significance, we fail to reject the null hypothesis.')
```

```
As the p-value 0.008026308204056278 is less than the level of significance, we reject the null hypothesis.
```

Step 7 Conclusions:

Conclusion:

We reject the null hypothesis. There is statistically significant evidence to suggest that the **conversion rate for the new landing page** is **greater** than that for the old landing page.

Business Recommendation:
The new landing page is performing better in terms of conversions. It is recommended that a wider rollout of the new page be considered to improve overall user engagement and success metrics.

Question 3 Hypothesis Testing Details

Step 1: Define the null and alternate hypotheses

H_0 : Conversion status is **independent** of preferred language

H_a : Conversion status is **dependent** on preferred language

Step 2: Select Appropriate test

Chi-Square Test of Independence

Used when:

```
Categorical variables

The expected value of the number of sample observations in each level of the v

|Random sampling from the population
```

Step 3: Decide the Significance Level

As given in the problem statement, we select a significance level of

alpha = 0.05

Step 4 Collect and Prepare the data:

```
# Create a contingency table showing the distribution of converted vs language_preferred
contingency_table = pd.crosstab(df['language_preferred'], df['converted'])

contingency_table
```

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

Step 5 Calculate the p-value:

```
# Import the chi2_contingency function from scipy.stats
from scipy.stats import chi2_contingency

# Run the Chi-Square test
chi2, p_value, dof, exp_freq = chi2_contingency(contingency_table)

print('The p-value is', p_value)
```

```
The p-value is 0.2129888748754345
```

Step 6: Compare the p-value with alpha

```
p-value = 0.213

Significance level alpha = 0.05
```

```
# print the conclusion based on the p-value
if p_value < 0.05:
    print(f'As the p-value {p_value} is less than the level of significance, we reject the null hypothesis.')
else:
    print(f'As the p-value {p_value} is greater than the level of significance, we fail to reject the null hypothesis.')
```

```
As the p-value 0.2129888748754345 is greater than the level of significance, we fail to reject the null hypothesis.
```

7: Conclusion

We fail to reject the null hypothesis.

Conclusion:

```
There is no statistically significant relationship between a user's preferred language and whether or not they
converted.
The data suggests that conversion status is independent of language preference.
```

Business Recommendation:

```
There is no need to tailor landing pages by language based on conversion data alone.
Resources can be focused on improving overall page effectiveness across all language segments.
```

Question 4 Hypothesis Testing Details

Step 1: Define the null and alternate hypotheses

We are testing whether the **mean time spent on the new landing page** differs across **language groups**.

H_0 : Mean time spent is the same across all language groups

H_a : At least one group has a mean time spent that is different

Step 2: Select the Appropriate Test

We are comparing the **mean time spent on the new landing page** across **three independent language groups** (`English`, `French`, and `Spanish`).

The response variable (`time_spent_on_the_page`) is **continuous**, and the grouping variable (`language_preferred`) is **categorical with more than two groups**.

Therefore, the appropriate test is:

One-Way ANOVA (Analysis of Variance)

This test will help us determine whether there is a **statistically significant difference in mean time spent** across the different language groups.

We will proceed to check the assumptions of ANOVA (normality and equal variance) before conducting the test.

Step 3: Decide the Significance Level

As specified in the question, we will use a significance level of:

alpha = 0.05

This means we require at least 95% confidence to conclude that **mean time spent differs across language groups**.

Step 4: Check Assumptions

```
from scipy.stats import shapiro

# Subsets of time spent by language group
english_time = df_new[df_new['language_preferred'] == 'English']['time_spent_on_the_page']
french_time = df_new[df_new['language_preferred'] == 'French']['time_spent_on_the_page']
spanish_time = df_new[df_new['language_preferred'] == 'Spanish']['time_spent_on_the_page']

# Shapiro-Wilk test for normality
print("English:", shapiro(english_time))

print("French:", shapiro(french_time))

print("Spanish:", shapiro(spanish_time))
```

```
English: ShapiroResult(statistic=0.9731259942054749, pvalue=0.8863451480865479)
French: ShapiroResult(statistic=0.9402673244476318, pvalue=0.3213688135147095)
Spanish: ShapiroResult(statistic=0.9072974920272827, pvalue=0.09006832540035248)
```

Step 4a: Check Normality with Shapiro-Wilk Test

I applied the Shapiro-Wilk test to assess whether the `time_spent_on_the_page` variable is normally distributed within each language group.

Language	p-value	Result
English	0.886	Normal
French	0.321	Normal
Spanish	0.090	Normal

All p-values are greater than 0.05, so the normality assumption for ANOVA is satisfied.

Step 4b: Levene's Test for Equal Variance

Step 4b is Levene's test to check the second key ANOVA assumption: equal variances across groups.

```
from scipy.stats import levene

# Levene's test for homogeneity of variances
stat, p_value = levene(english_time, french_time, spanish_time)

print("Levene's Test p-value:", p_value)
```

```
Levene's Test p-value: 0.46711357711340173
```

Step 4b: Check Equal Variance with Levene's Test

Levene's Test was used to assess the assumption of homogeneity of variances across language groups.

p-value = 0.467

Since the p-value is greater than 0.05, we conclude that the variances are equal. The assumption of equal variance is satisfied.

Step 5: Calculate the p-value

```
from scipy.stats import f_oneway

# Run the ANOVA test
f_stat, p_value = f_oneway(english_time, french_time, spanish_time)

print("ANOVA F-statistic:", round(f_stat, 4))
print("ANOVA p-value:", p_value)
```

```
ANOVA F-statistic: 0.8544
ANOVA p-value: 0.43204138694325955
```

Step 6: Compare the p-value with α

```
# print the conclusion based on p-value
if p_value < 0.05:
    print(f'As the p-value {p_value} is less than the level of significance, we reject the null hypothesis.')
else:
    print(f'As the p-value {p_value} is greater than the level of significance, we fail to reject the null hypothesis.')
```

```
As the p-value 0.43204138694325955 is greater than the level of significance, we fail to reject the null hypothesis.
```

Step 7: ANOVA Conclusion

Conclusion:
There is **no statistically significant difference** in the mean time spent on the new landing page across English, French, and Spanish users.

Business Recommendation:
All language groups appear to engage similarly with the new landing page. No localization changes are needed based on time-spent behavior.

Type Markdown and LaTeX: α^2