ENews Express

Soneeka Pratap

04/06/2022

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Hypotheses Tested and Results
- Appendix

Executive Summary

- The average time spent on the landing page is 5.38 minutes. To increase user engagement and time spent on the page, conducting a survey comparison between the new and old landing page would give more insights into the changes and features that users prefer, and are interested in for an online news portal.
- The language preferred does not seem to greatly affect the time spent on the web page, where the mean time spent on the new page is about same for the different language users. ENews Express could incorporate more language options to target other language populations/countries, for the potential to gain more subscribers. Current analysis shows this may not change the level of engagement.
- Converted status is independent of the preferred language, however the conversion rate for the new page is greater than the conversion rate for the old page.
- 54% of users converted to being a subscriber and spent around 2.5 minutes longer on average on the landing page compared to users that did not subscribe. As there is a preference for the new page with a higher amount of subscribers and longer time spent, ENews Express should perform further analysis and look into the specific interests and content that subscribers are reading.
- The above recommendations will further allow ENews Express to make further changes on their content and design of their new portal page, as the analysis shows the preference is for the new page.

Business Problem Overview and Solution Approach

Business Problem:

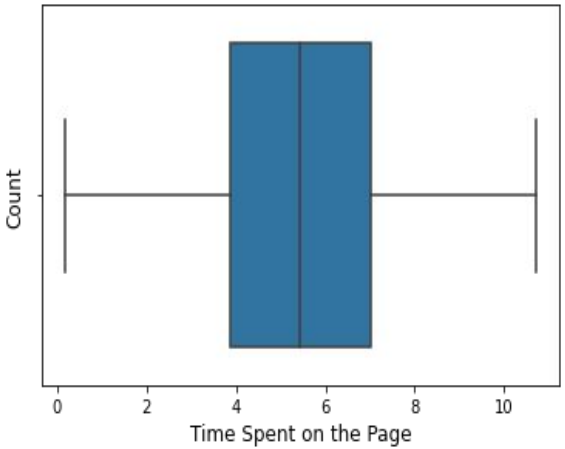
- E-news Express have an online news portal which aims to expand its business by acquiring new subscribers. The company has noticed a decline in new monthly subscribers compared to the previous year and thinks the decreased actions and levels of engagement is due to the layout and design of the current webpage.

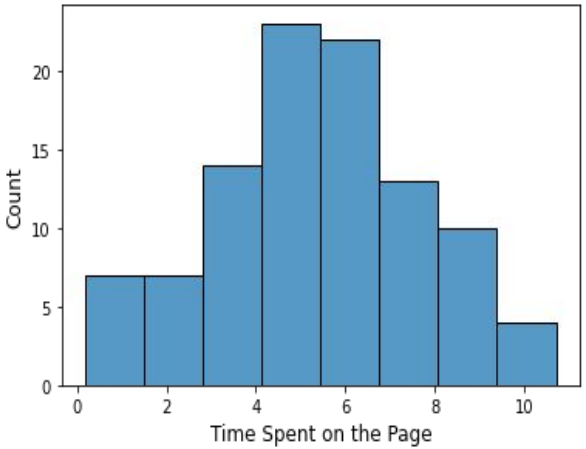
Solution Approach:

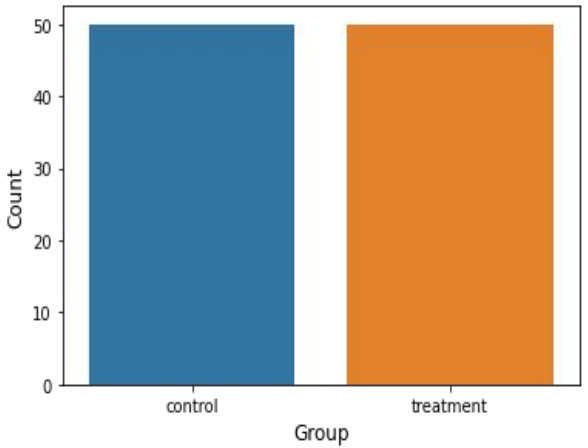
- To analyse user action and engagement levels on the website, as each visitor has different interests which results in different levels of engagement.
- To determine the effectiveness of the new landing page in gathering new subscribers for the news portal by comparing it with the old landing page, conversion rate , and different language users.

EDA Results - Univariate Analysis

- The mean time spent on the page is 5.38 minutes and is close to the median value of 5.42 minutes
- The distribution appears to follow a normal distribution with the minimum amount of time spent being 0.19 minutes, and the maximum time of 10.71 minutes
- There are 100 randomly selected users that are split into two groups
- The control and treatment groups have 50 users



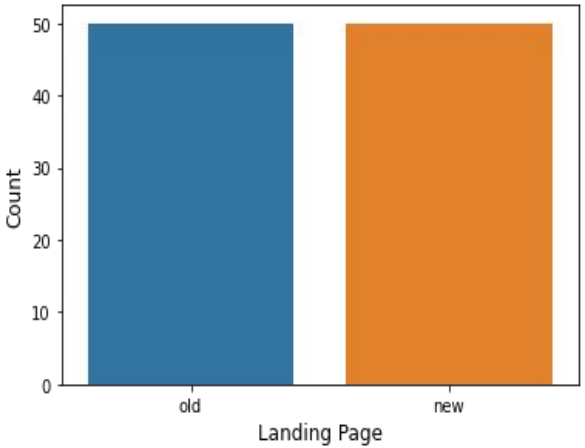


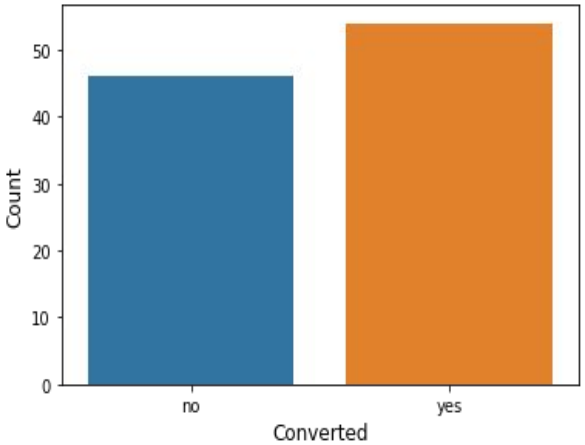


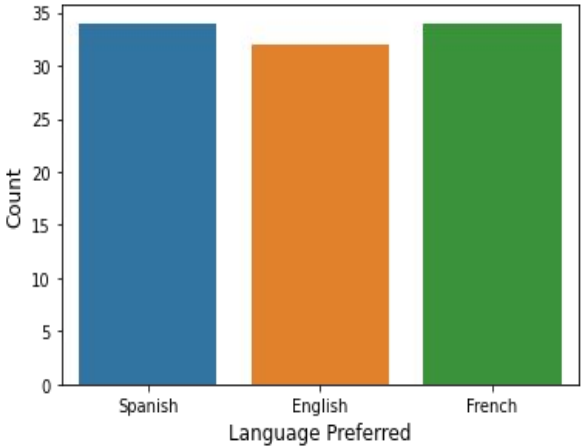
Link to Appendix slide on data background check

EDA Results - Univariate Analysis

- The control and treatment groups are assigned to the new and old landing pages, resulting in a count of 50 users for each landing page type
- 54 users have converted to a subscriber and 46 have not converted
- There are an equal amount of users using both Spanish and French as the preferred language (34 users in each), and 32 users using English as the preferred language



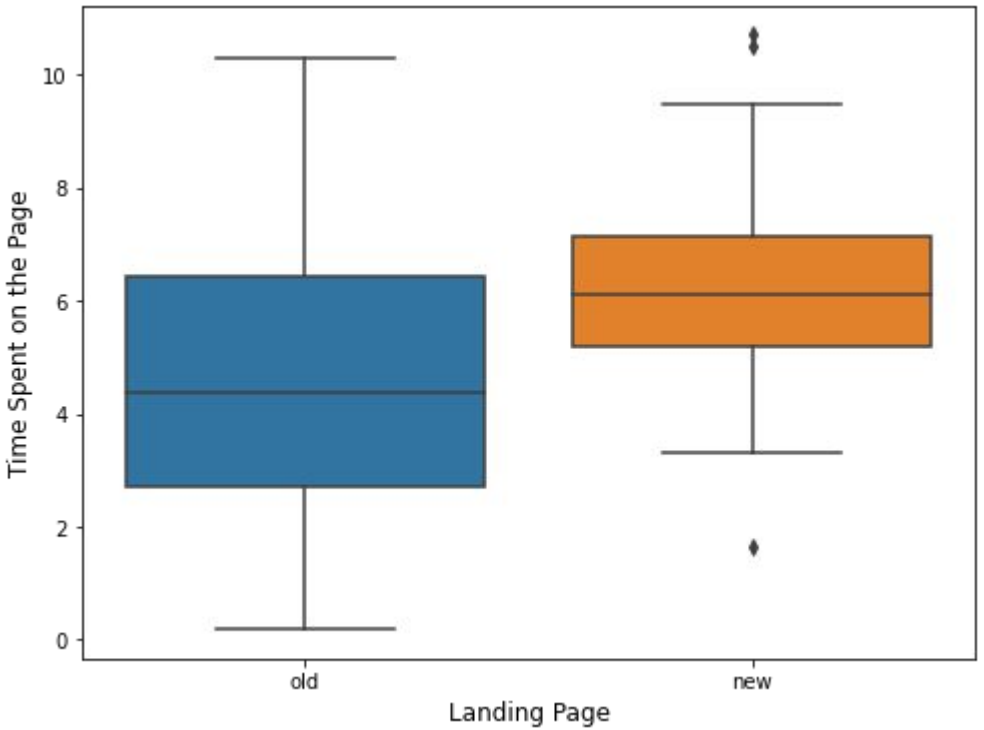




Link to Appendix slide on data background check

EDA Results - Bivariate Analysis

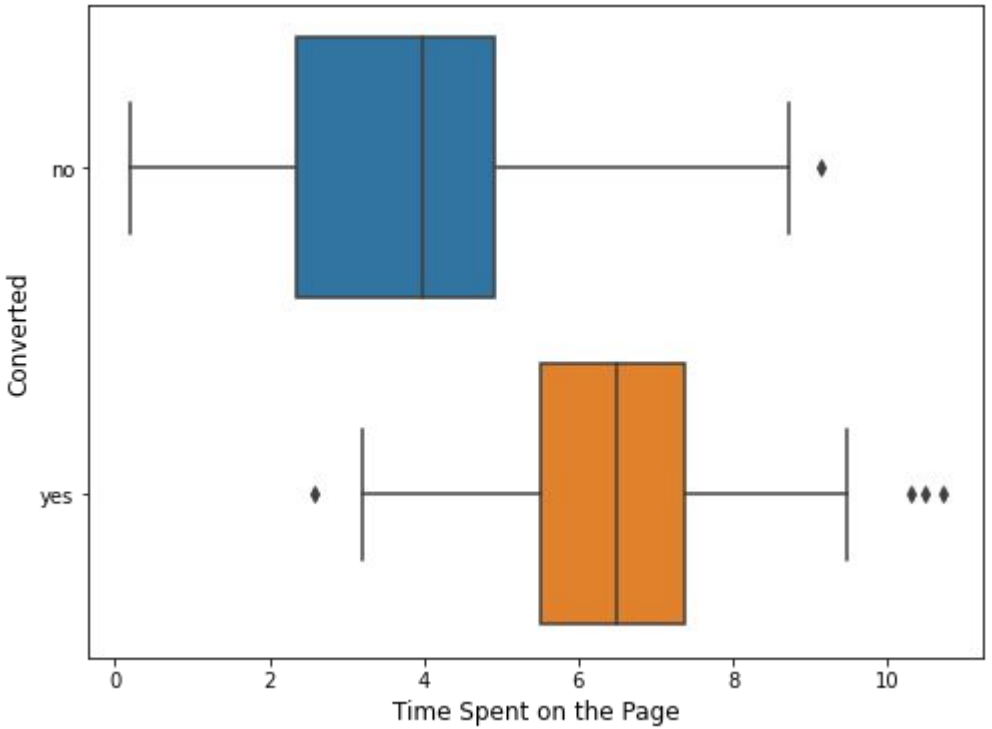
- The median time spent on the new landing page is longer than the median time spent on the old landing page
- The old landing page has more variability of time spent on the page ranging from 0.19 minutes to 10.71 minutes
- Comparing this to the new landing page, it varies from just over 3 minutes to a maximum of just over 9 minutes
- The new landing page has a few outliers on both ends of the boxplot meaning there are some shorter and some longer visits on the web page



Link to Appendix slide on data background check

EDA Results - Bivariate Analysis

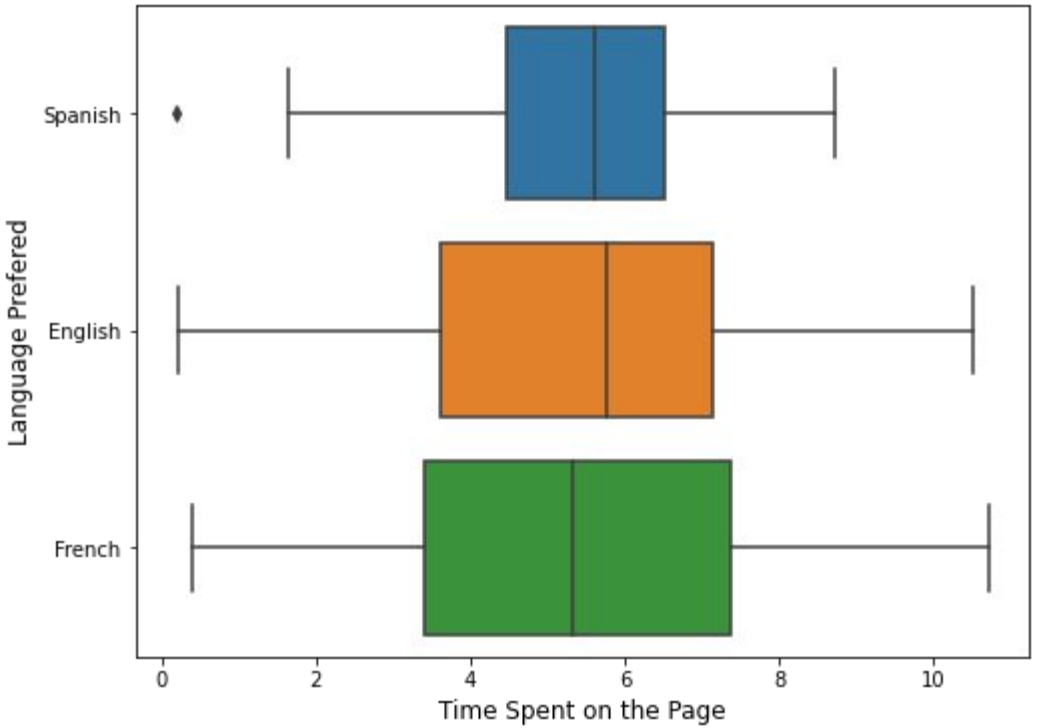
- The box plot shows the users that have converted to a being subscriber of the news portal has spent more time on the web page compared to users who did not convert to a being subscriber
- The users that did not convert to a subscriber have a lower median time spent on the page of around 4 minutes
- Users that subscribed have a median time of around 6.5 minutes
- Both converted statuses have outliers which could be dependent on the type of content that the users are interested in that are presented on the page



Link to Appendix slide on data background check

EDA Results - Bivariate Analysis

- There are three languages available on the web page - Spanish, English and French
- Spanish has an outlier
- Spanish has less variability of time being spent on the web page
- Both French and English have users spending a similar amount of time on the web page
- All three preferred languages have similar median times spent on the web page - the language preferred does not seem to greatly affect the time spent on the web page



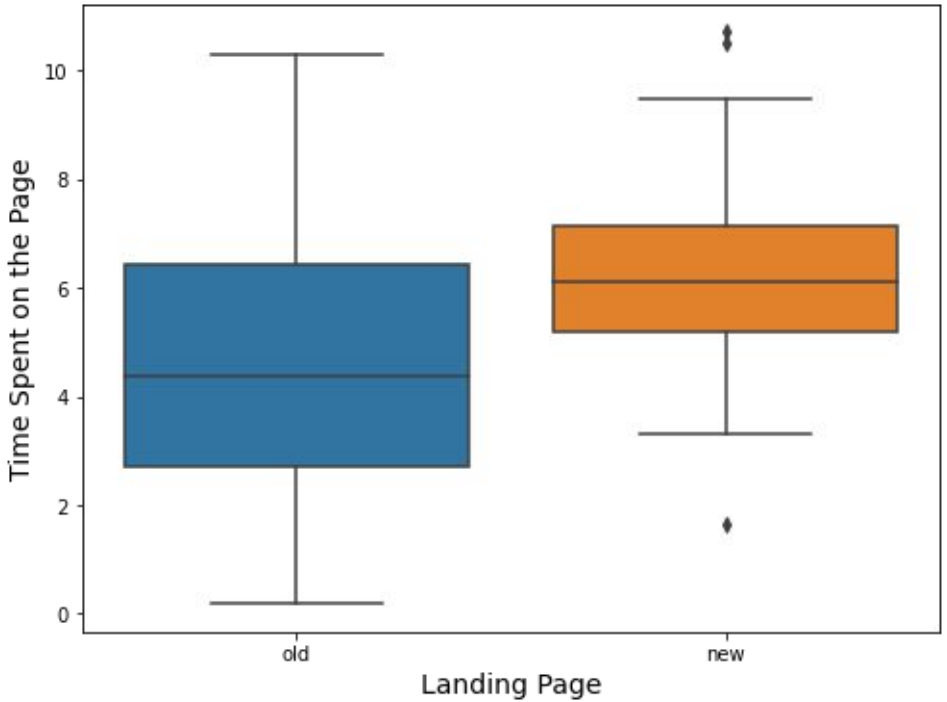
Link to Appendix slide on data background check

Hypotheses Tested and Results

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

Result:

Since the p-value is less than the 5% level of significance, we reject the null hypothesis. We have enough statistical evidence to say that the time spent on the new landing page is greater than the time spent on the old landing page.



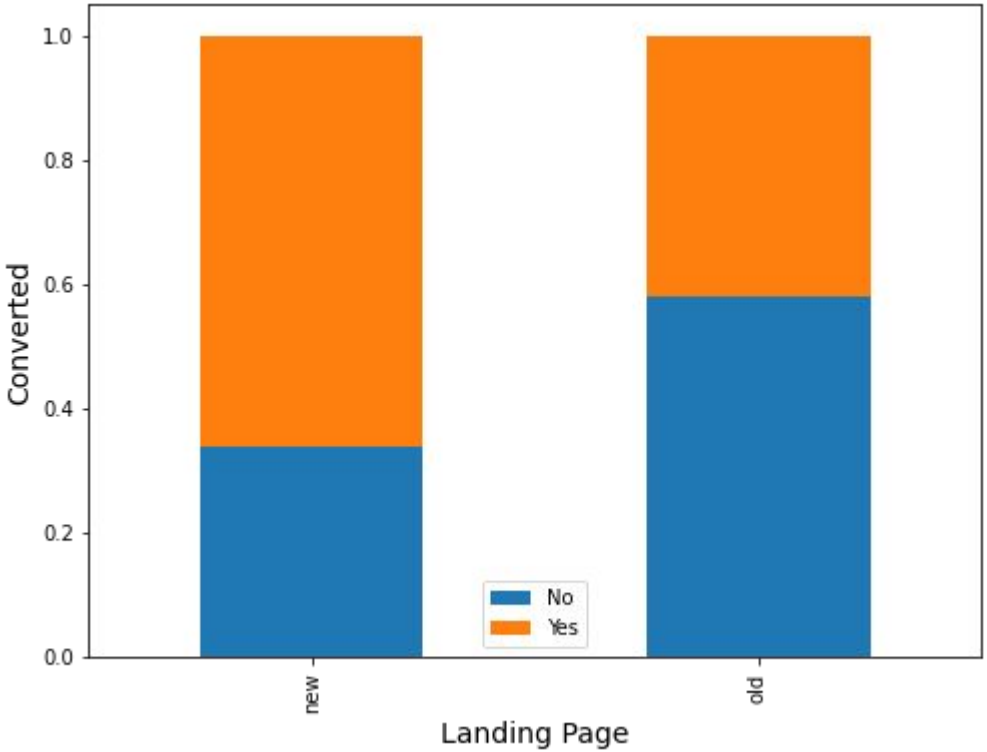
Link to Appendix slide on details of the test performed

Hypotheses Tested and Results

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

Result:

Since the p-value is less than the 5% level of significance, we reject the null hypothesis. There is enough statistical evidence to say that the conversion rate for the new page is greater than the conversion rate for the old page



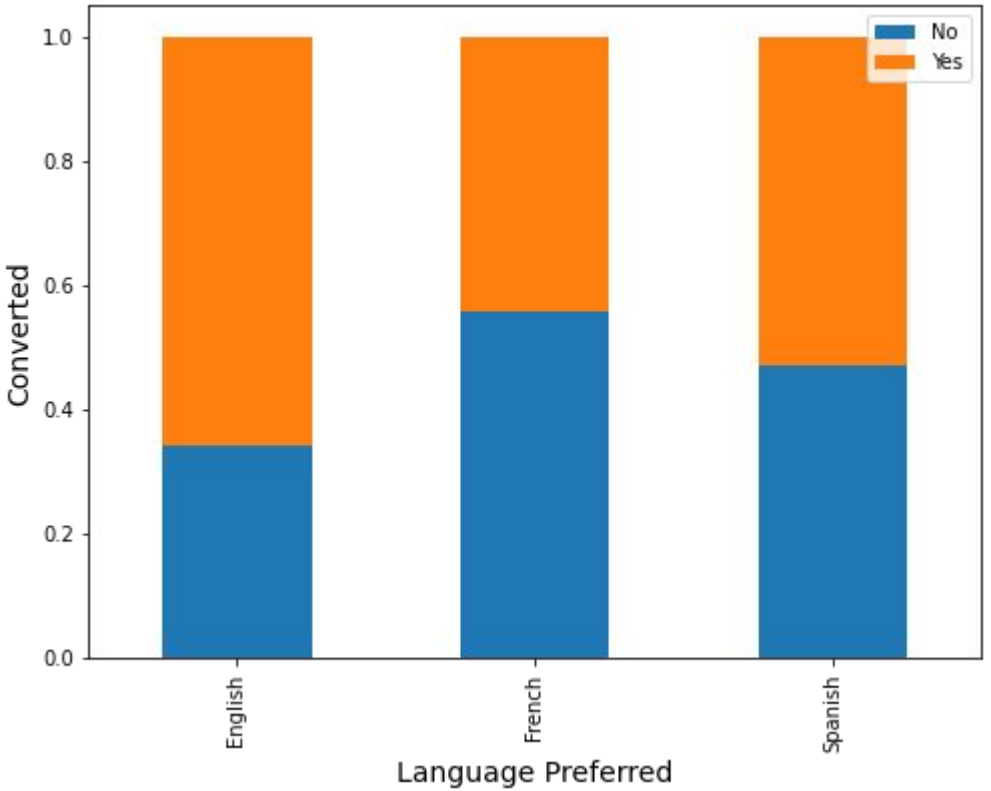
Link to Appendix slide on details of the test performed

Hypotheses Tested and Results

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

Result:

Since the p-value is greater than the 5% level of significance, we fail to reject the null hypothesis. There is not enough statistical evidence to say that the converted status depends on the preferred language.



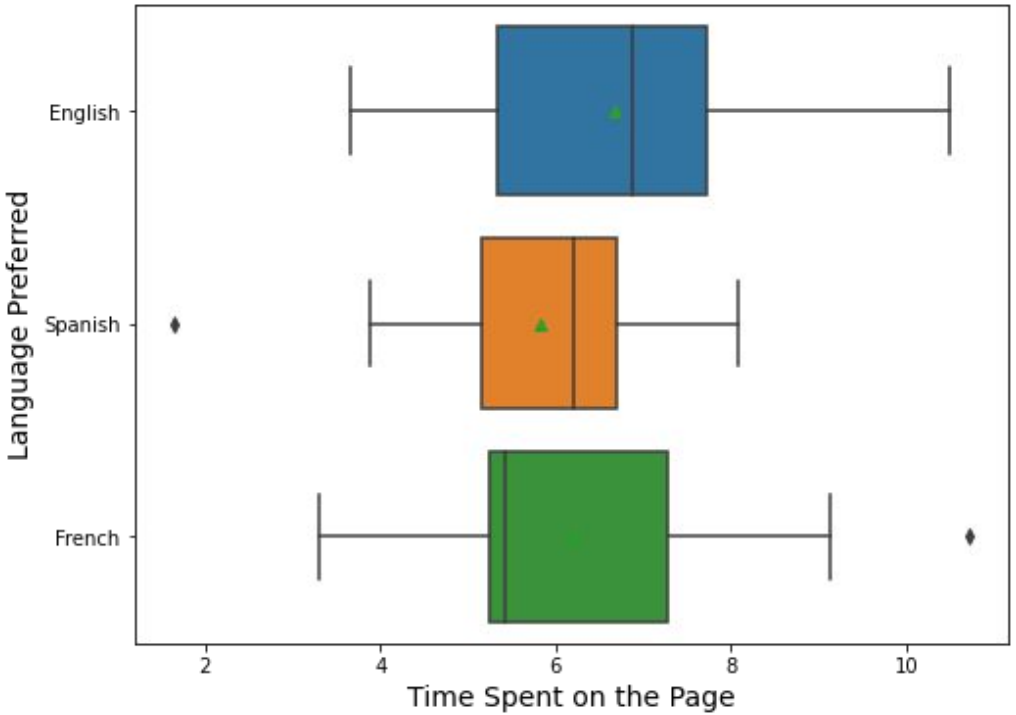
Link to Appendix slide on details of the test performed

Hypotheses Tested and Results

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

Result:

Since the p-value is greater than the 5% level of significance, we fail to reject the null hypothesis. There is enough statistical evidence to suggest that the mean time spent on the new page is the same for the different language users.



Link to Appendix slide on details of the test performed

APPENDIX

Data Background and Contents

- The data provided is from 100 randomly selected news portal users that are divided into two evenly split groups - one treatment group (50 on the new landing page) and one control group (50 on the existing landing page)
- There are a total of 100 observations with 6 columns showing information on the user ID, group type, landing page type, time spent on the page, converted status, and the language preferred
- There are no missing values or duplicate values in the data
- User_id and time_spend_on_the_page are both numerical, and the remaining four columns are categorical

	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

```
control      0.5
treatment    0.5
Name: group, dtype: float64
```

```
old    0.5
new    0.5
Name: landing_page, dtype: float64
```

```
yes    0.54
no     0.46
Name: converted, dtype: float64
```

```
Spanish    0.34
French     0.34
English    0.32
Name: language_preferred, dtype: float64
```

1. Hypothesis Testing Details

Hypotheses:

Ho: Users spend the same amount of time on the new landing page and on the existing landing page

Ha: Users spend more time on the new landing page than on the existing landing page

Test Selected:

T-test for independence

p-value:

p-value = 0.0001392381225166549

```
test_stat, p_value =  ttest_ind(time_spent_new, time_spent_old, equal_var = False, alternative =
'greater')
```

Standard deviations calculated for the new and old page:

```
time_spent_new = df[df['landing_page'] == 'new']['time_spent_on_the_page']
time_spent_old = df[df['landing_page'] == 'old']['time_spent_on_the_page']
```

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

2. Hypothesis Testing Details

Hypotheses:

Ho: The mean user conversion rate for the new page and old page are the same

Ha: The mean user conversion rate for the new page is greater than the mean conversion rate for the old page

Test Selected:

Two proportion Z-test

p-value:

p-value = 0.008026308204056278

```
test_stat, p_value = proportions_ztest([new_converted, old_converted] , [n_treatment, n_control], alternative ='larger')
```

Calculations done for the proportions of each group:

```
new_converted = df[df['group'] == 'treatment']['converted'].value_counts()['yes']
```

```
old_converted = df[df['group'] == 'control']['converted'].value_counts()['yes']
```

```
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_control, n_treatment ))
```

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

3. Hypothesis Testing Details

Hypotheses:

Ho: Converted status is independent of the preferred language

Ha: Converted status depends on the preferred language

Test Selected:

Chi2 test for independence

p-value:

p-value = 0.2129888748754345

```
chi2, p_value, dof, exp_freq = chi2_contingency(contingency_table)
```

Contingency table created for language preferred with converted status:

```
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

4. Hypothesis Testing Details

Hypotheses:

Ho: The mean time spent on the new page is the same for different language users

Ha: At least one of the mean times spent for different language users is different

Test Selected:

One-way ANOVA

p-value:

p-value = 0.43204138694325955

```
test_stat, p_value = f_oneway(time_spent_English, time_spent_French, time_spent_Spanish
```

Created a new dataframe and calculated the mean time spent for each language:

```
df_new = df[df['landing_page'] == 'new']
df_new.groupby(['language_preferred'])['time_spent_on_the_page'].mean()
```

```
language_preferred

English    6.663750

French     6.196471

Spanish    5.835294

Name: time_spent_on_the_page, dtype: float64
```

Great Learning

Happy Learning !