E-NEWS EXPRESS BUSINESS ANALYSIS

BY: DR. CYNTHIA OFFOHA

AGENDA

- EXECUTIVE SUMMARY
- BUSINESS PROBLEM OVERVIEW & SOLUTION APPROACH
- DATA OVERVIEW
- EDA - UNIVARIATE ANALYSIS & KEY QUESTIONS
- EDA – BIVARIATE ANALYSIS & KEY QUESTIONS
- RECOMMENDATIONS & CONCLUSIONS

EXECUTIVE SUMMARY

- Electronic News (E-News) provides great opportunity to easily access day-to-day events that occurs globally
- E-News informations are obtained from online databases and accessed utilizing variety of software then transmitted to users
- The aim for E-News which is an online new portal is to expand its business by obtaining more subscribers
  - Every visitor to the website takes certain actions according to their interest
  - The company aim to analyze these actions to understand user interests and deter how to generate better engagement
    - Thus, the goal is to access these interest and determine whether a new feature will be effective or not
- The E-news executives believe there has been a decrease in new monthly subscribers compared to previous year
- **Note** --> An experimental technique, known as A/B testing is utilized to determine whether a new feature attracts users according to the chosen metric
  - This is based on the analysis that companies often evaluate users' responses to two variants of a product to decide which of the two variants is more effective

BUSINESS PROBLEM OVERVIEW & SOLUTION APPROACH

- Company design team has created new landing page consisting of new outlines and more relevant content compared to the old page
- In order to test the effectiveness of the new landing page in acquiring new subscribers
  - A randomized experiment is conducted in selecting 100 users and dividing into equally two groups
    - Control (Frist) group --> Existing landing page
    - Treatment (Second) group --> New landing page
  - Interaction data of users in both groups with two versions of landing page was collected
- E-News data scientists have been appointed to explore the data and conduct statistical analysis ~ at 5% significance level
  - Establish the effectiveness of new landing page in acquiring new subscribers for the new portal by analyzing the following:
    - Time spent among users on the new landing page vs the existing landing page
    - The conversion rate between the new page and the old page --> If the proportion of visited users on new landing page leading to subscribers is greater than the old page
    - Preferred language on the converted status --> if users subscriptions depend on preferred language
    - Time spent on new page same for different language users

DATA OVERVIEW

VARIABLE	DESCRIPTIONS
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

- **Total of 100 Observations**
- **Total of 6 columns / features**
- **The features describes each observation**
- **No missing or null value for any 100 observations and its features**
- **All categorical variables except user_id and time spent on the page are integers**
- **Summary for time spent on page is based on old and new landing page**

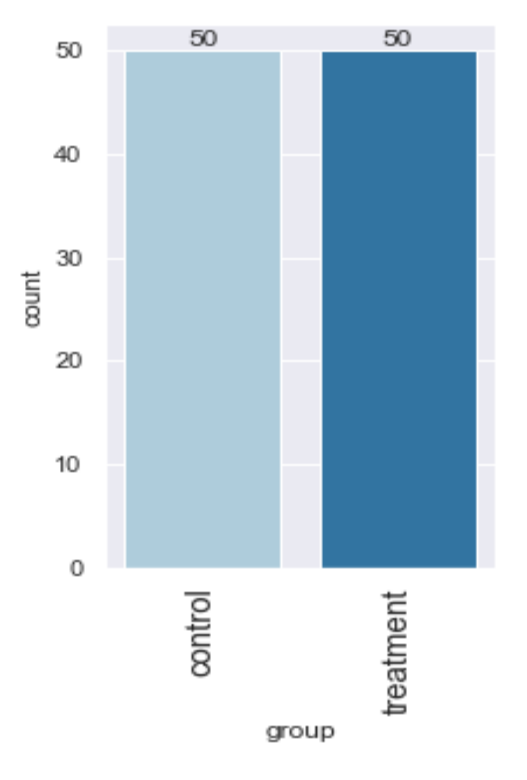
DATA STRUCTURE

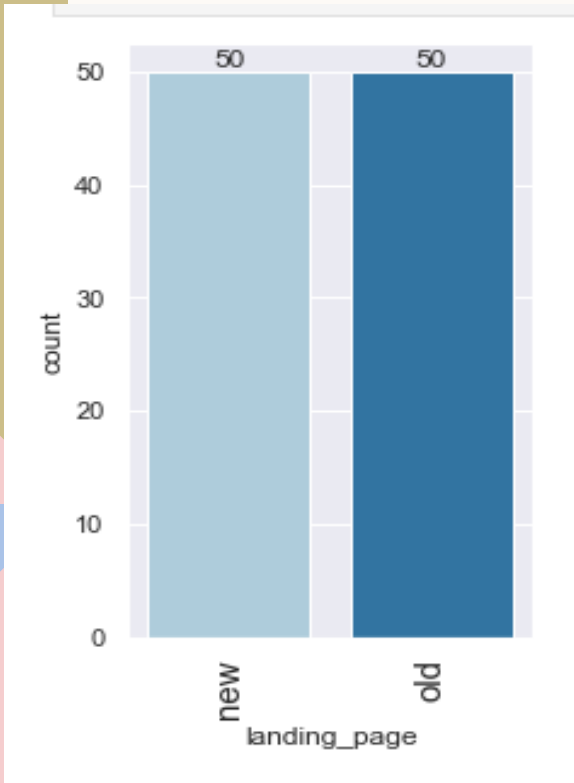
	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
0	546592	control	old	3.48	no	Spanish
1	546468	treatment	new	7.13	yes	English
2	546462	treatment	new	4.40	no	Spanish
3	546567	control	old	3.02	no	French
4	546459	treatment	new	4.75	yes	Spanish

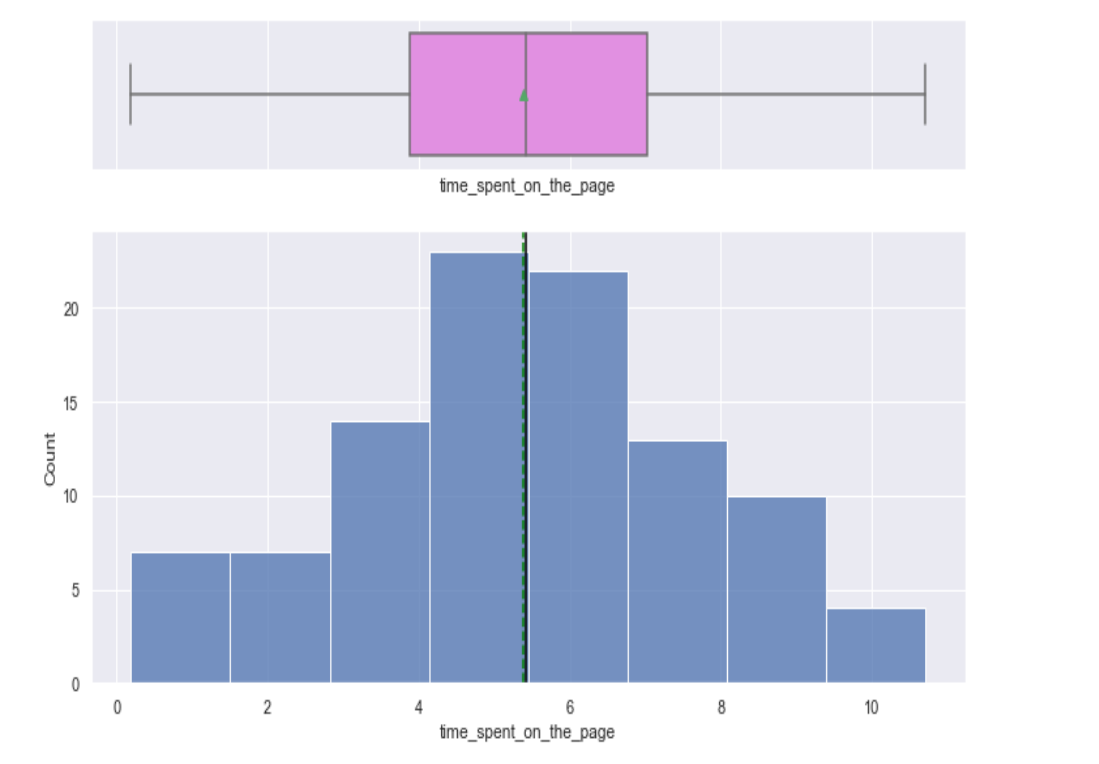
	user_id	group	landing_page	time_spent_on_the_page	converted	language_preferred
95	546446	treatment	new	5.15	no	Spanish
96	546544	control	old	6.52	yes	English
97	546472	treatment	new	7.07	yes	Spanish
98	546481	treatment	new	6.20	yes	Spanish
99	546483	treatment	new	5.86	yes	English

	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

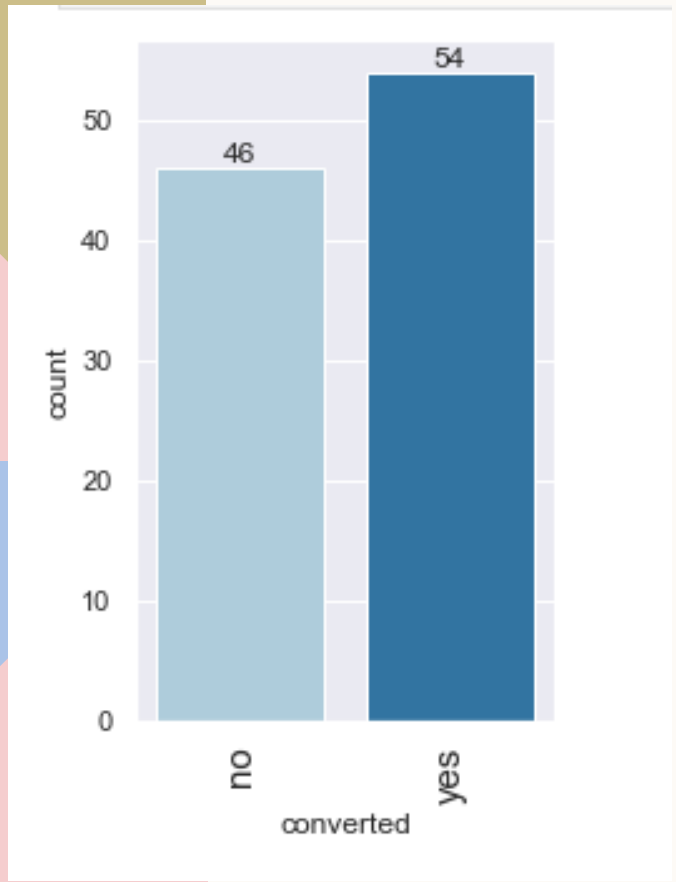
EDA: UNIVARIATE ANALYSIS

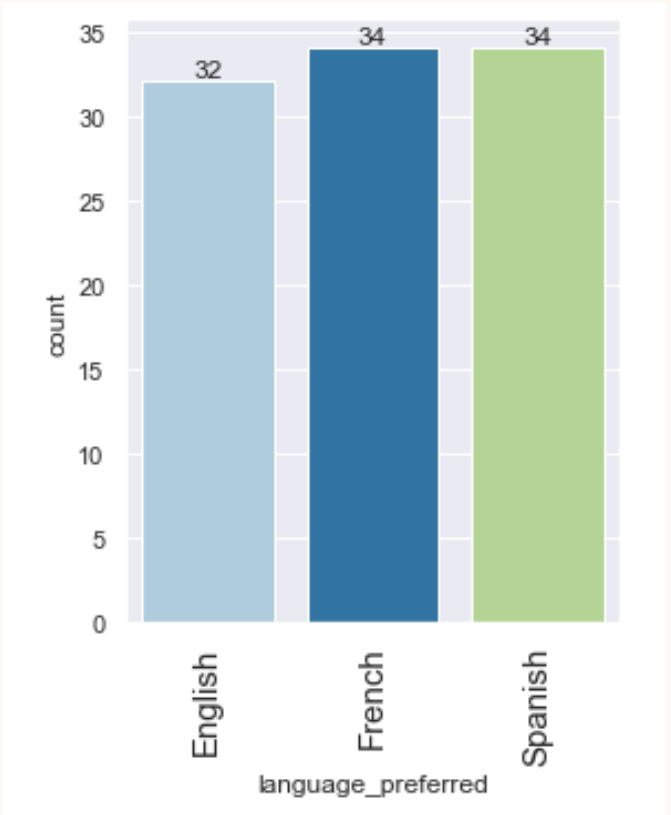






EDA: UNIVARIATE ANALYSIS

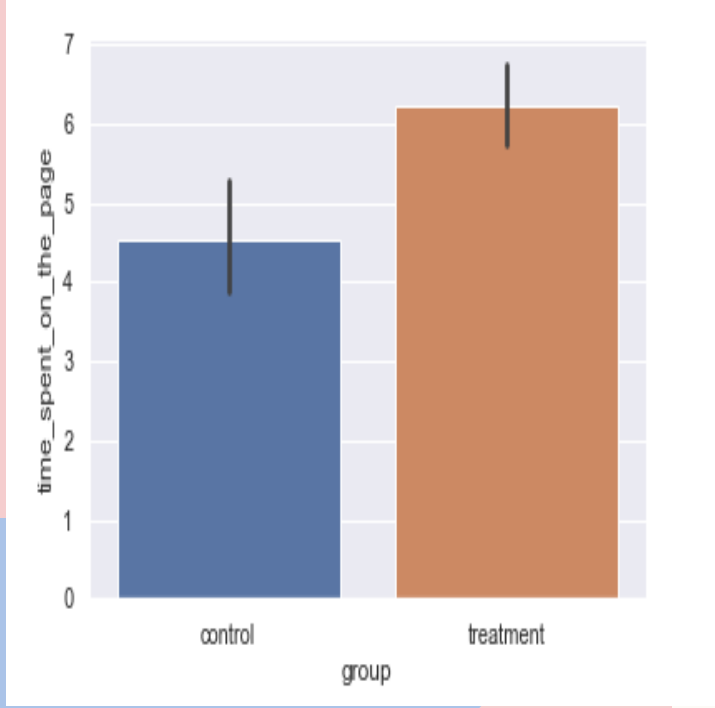


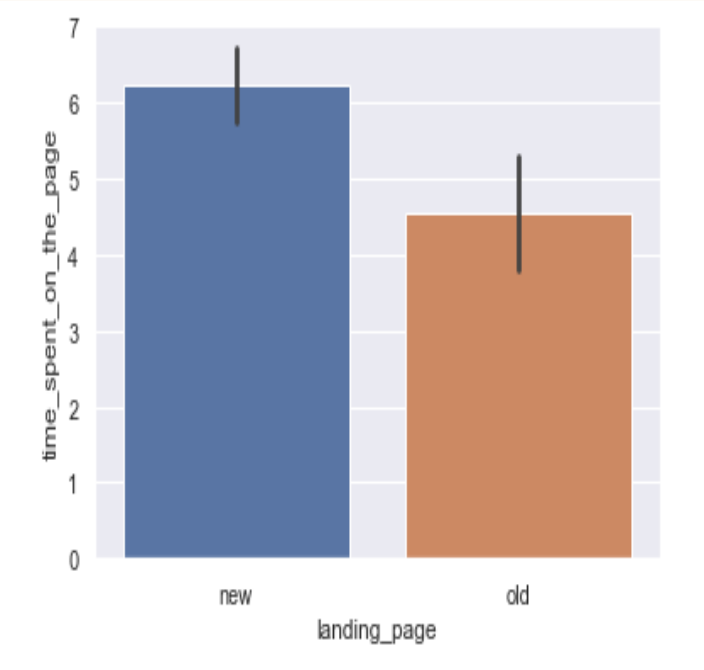


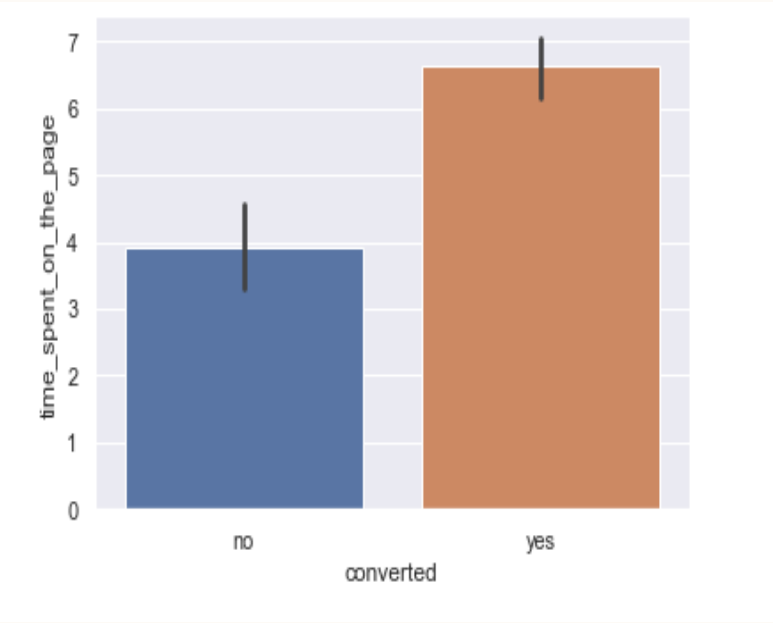
OBSERVATIONS

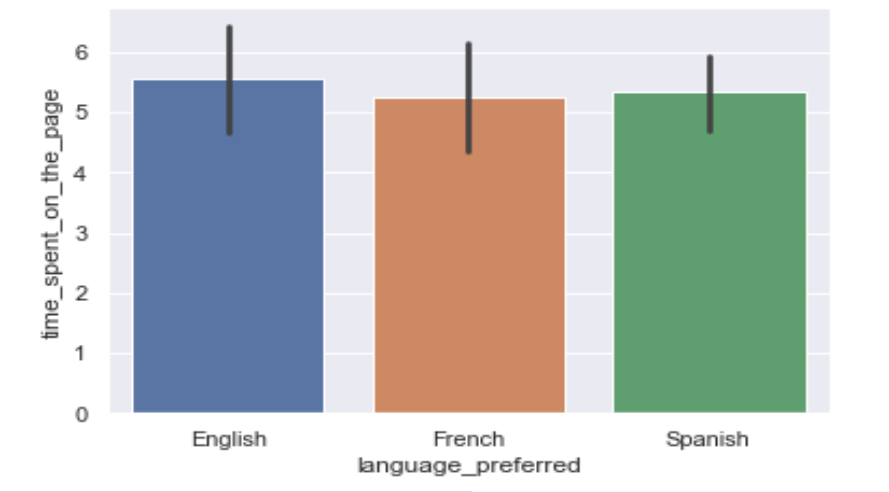
- 100 unique users
- **Group**
  - Two unique groups
    - Control group & Treatment (Tx) group
      - Each group consists of 50 users
- **Landing page**
  - Two landing pages
    - New & old
      - Consist of 50 observations each
- **Converted**
  - 2 unique values of no and yes (converted)
  - 54 users were converted while 46 users were not converted after visitation to the landing page
- **Language preferred**
  - 3 unique preferred languages
    - English, French, & Spanish
      - French & Spanish--> highest frequency of 34
- **Time spent on the landing page**
  - Average time spent on page ~ 5.38min
  - The STD ~ 2.38min
  - Inter-quartile range consists:
    - Min ~ 0.19
    - 25% quartile ~ 3.88
    - 50% quartile ~ 5.42
    - 75% quartile ~ 7.02
    - Max ~ 10.71

EDA: BIVARIATE ANALYSIS









Note:

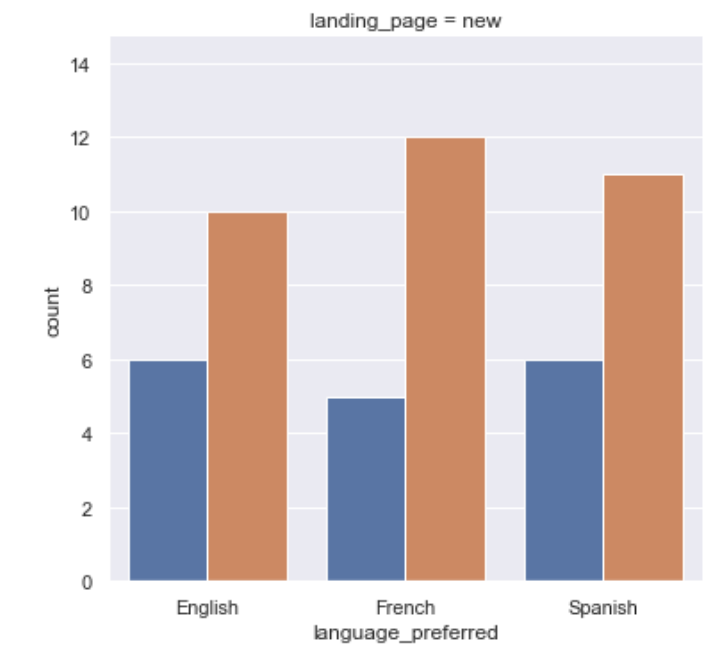
Tx group more likely to spend more time on the page

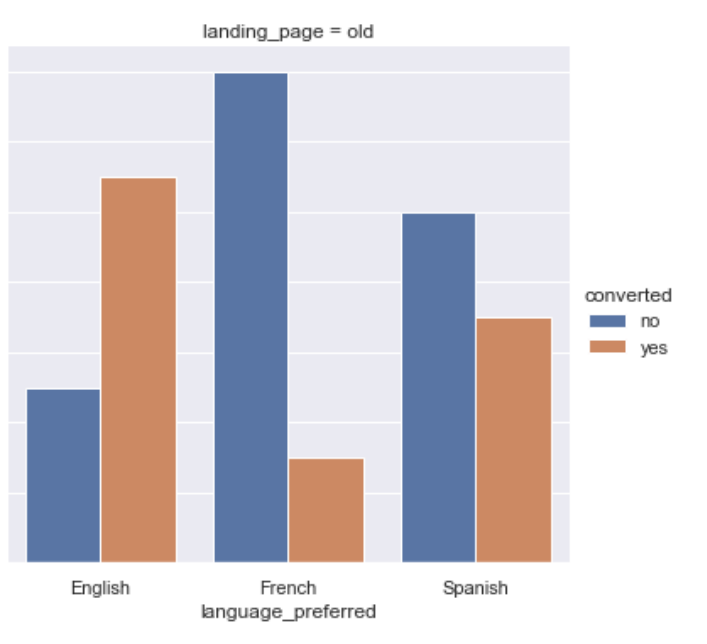
Users exposed to new landing page spend more time on average

Converted users showed a higher average time spent on the page

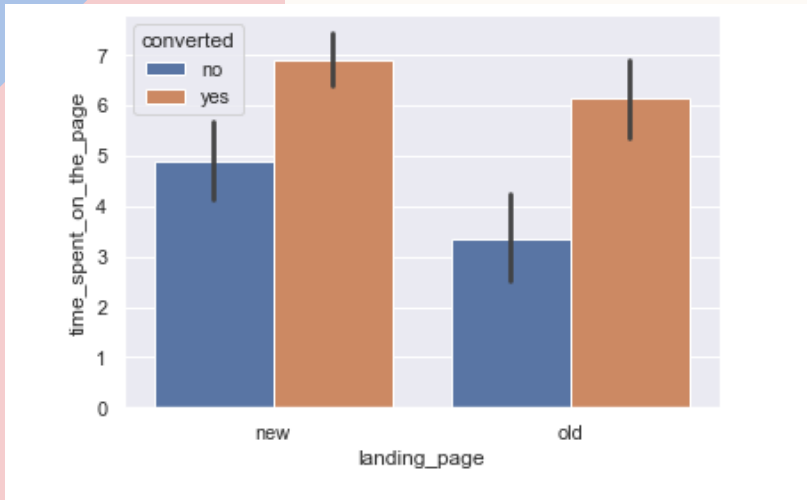
Avg. Spent on the page are relatively same among the 3 various languages (English, French, Spanish)

EDA: BIVARIATE ANALYSIS





French and Spanish users are less likely to convert to the new landing page



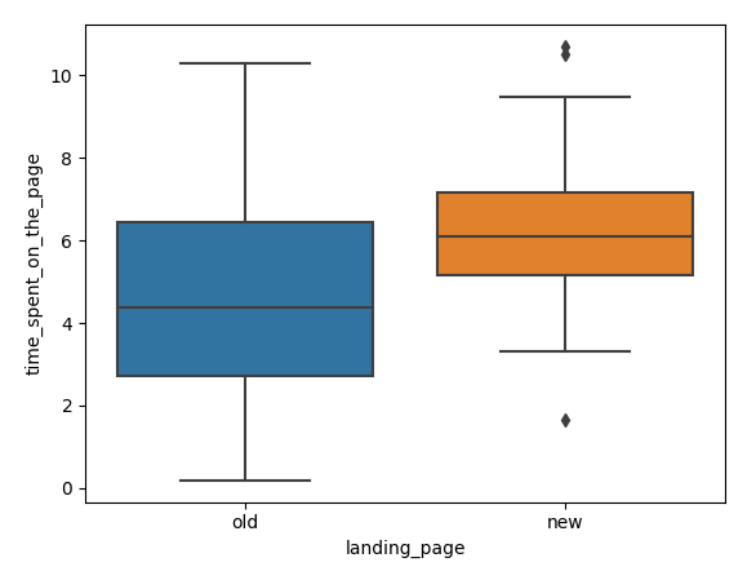
This illustrates that more users are likely to convert to the new landing page when first exposed

OBSERVATIONS

- Users spend less time on old page than the new page
- Users that converted to subscribers spent more time either on the new or old landing page
- The average time spent on the page appears to be similar for all preferred languages
- More subscribers in the treatment groups compared to the control group
- Treatment group users spend more time on the page compared to the control group users

DO USERS SPEND MORE TIME ON THE NEW LANDING PAGE THAN THE EXISTING LANDING PAGE???

VISUAL ANALYSIS



Null and alternative hypothesis defined:

H0 --> Mean time spent on old landing page = mean time spent on the new landing page

Ha --> Mean time spent on new landing page > mean time spent on old landing page.

Let $\mu 1$ and $\mu 2$ --> mean time spent on page for the old and new landing page respectively.

Mathematically hypotheses can be written as:
$H0: \mu 1 = \mu 2$
$Ha: \mu 1 < \mu 2$

Significance level --> $\alpha = 0.05$

Based upon collection & preparation of data:
- Sample STD of the time spent on the new page --> 1.82
- Sample STD of the time spent on the old page --> 2.58
- Thus, the population STD can be assumed to be unequal

Calculate the p-value with α:
- Used the ttest_ind function
- The p-value --> 0.0001392381225166549

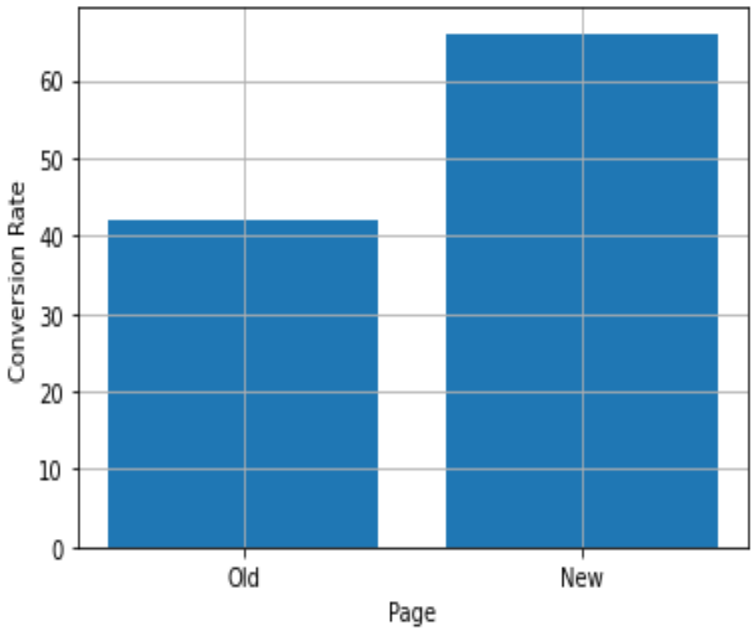
Compare the p-value with α:
- As the p-value 0.0001392381225166549 < the level of significance --> null hypothesis is rejected

OBSERVATIONS

- Users spend more time on the new page vs the old page
  - Which is proved with statistical analysis
- Null and Alternate hypothesis were determined
- It is shown to be one-tailed test consisting of two population means obtained from two independent populations
  - Two Independent sample t-test
- Inference
  - ~5% significance level: we fail to reject null hypothesis
    - There is sufficient statistical evidence to indicate that the mean time spent on new landing page is greater than old landing page

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???

VISUAL ANALYSIS



Define the null and alternative hypothesis:

Let $p1$ and $p2$ be the proportions conversion rate for old and new page, respectively

Then test the null hypothesis

$H0: p1=p2$

The proportions of conversion rate = both old and new landing pages

Against alternate hypothesis

$Ha: p1 < p2$

The proportions of conversation rate > for the new landing page

Significance level --> $\alpha = 0.05$

Based upon collection & preparation of data:

- Users that served the new pages -->50
- Users that served old pages --> 50
- Conversion rate for old website --> 42.0 %
- Conversion rate for new website --> 66.0 %

Calculate the p-value with α :

- Proportions_ztest function is utilized
- p-value --> 0.008026308204056278

Compare the p-value with α :

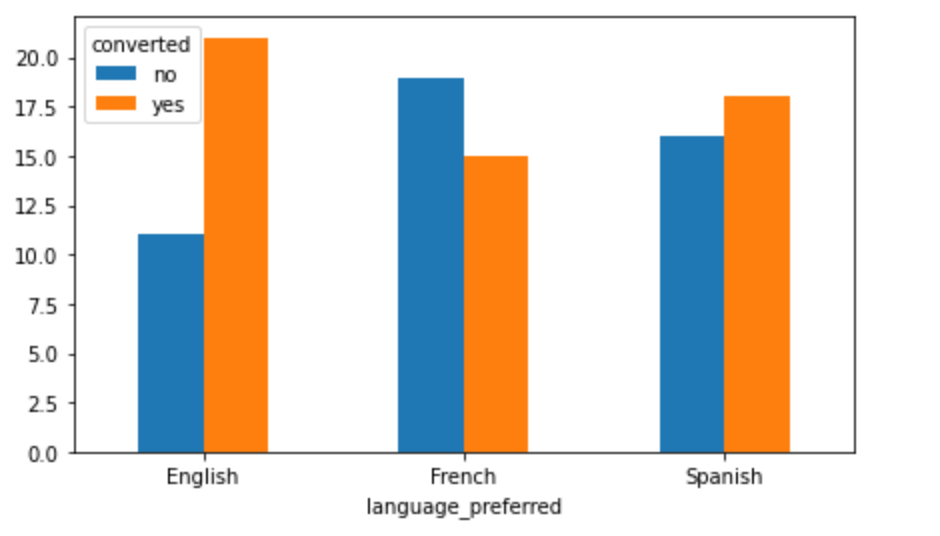
- As the p-value 0.008026308204056278 < level of significance --> Null hypothesis is rejected

OBSERVATIONS

- It shows that the conversion is greater for the new page compared to the old page
- Null and alternative hypothesis were determined
- Analysis indicates a two sample z-test
  - One-tailed test consisting of two population proportions from two independent populations
- Inference
  - P-value < significance value -->0.05
    - Null hypothesis is rejected
    - Not enough statistical significance to conclude conversion rate for new page is greater

DOES THE CONVERTED STATUS DEPEND ON THE PREFERRED LANGUAGE???

VISUAL ANALYSIS



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

PIVOT TABLE

Define the null and alternative hypothesis:

*H*0: Converted status is independent of preferred language

Ha: Converted status is dependent upon preferred language

Significance level --> $\alpha = 0.05$

Based upon collection & preparation of data:
- Pivot table generated illustrating converted status indexed by language preferred

Calculate the p-value with α:
- chi2_contingency function is utilized
- P-value --> 0.2129888748754345

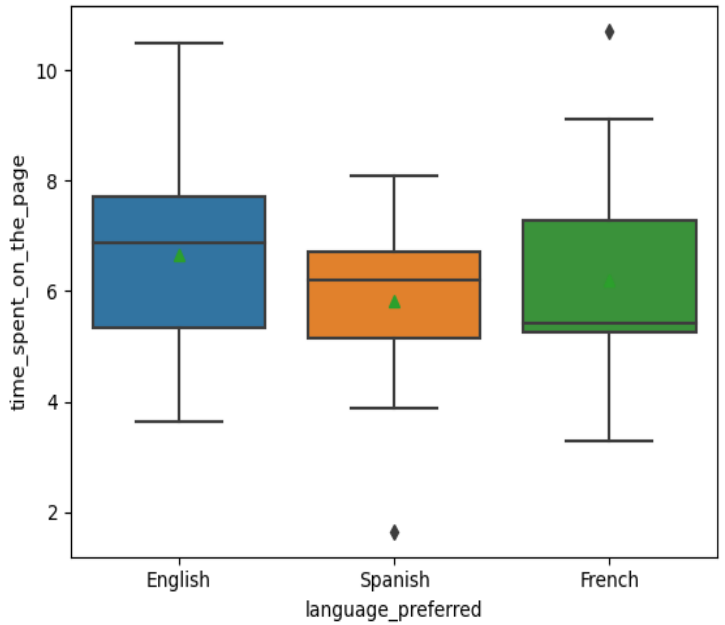
Compare the p-value with α:
- As the p-value 0.2129888748754345 > level of significance, --> Fail to reject the null hypothesis

OBSERVATIONS

- The number of subscribers, which is the converted, rely on language preferred
- Null and alternate hypothesis were set and determined
- Chi square Test of Independence was utilized
  - To analyze two categorical variables based on independence
- Inference
  - P-value > Significance level --> Fail to reject null Hypothesis
  - Not enough statistical significance to summarize that converted status is dependent on preferred language

IS THE TIME SPENT ON THE NEW PAGE SAME FOR THE DIFFERENT LANGUAGE USERS???

VISUAL ANALYSIS



Define the null and alternate hypotheses:

$\mu1, \mu2, \mu3$ --> means of time spent on page for languages preferred English, French, and Spanish respectively.

$H0: \mu1 = \mu2 = \mu3$
Ha: At least one mean time spent on page is different from the rest of the languages preferred

Shapiro-Wilk's & Levene's Test
$H0$: Time spent on the new page follows a normal distribution
Ha: Time spent on the new page does not follow a normal distribution
Assumption 1 -->Normality
Calculate the p-value with α:
- Shapiro function
- The p-value --> 0.8040016293525696

Compare the p-value with α:
- P-value > 5% significance level --> Fail to reject the null hypothesis that the response follows the normal distribution

Levene's Test
$H0$: All the population variances are equal
Ha : At least one variance is different from the rest
Assumption 1 --> Homogeneity of Variance
Calculate the p-value with α:
- Levene function utilized
- The p-value --> 0.46711357711340173

Compare the p-value with α:
- P-value > 5% significance level --> fail to reject the null hypothesis of homogeneity of variances

One Way ANOVA

Decide the significance level:
- α --> 0.05

Based upon collection & preparation of data:
- Time spent on new page determined by grouping all three languages

Calculate the p-value with α:
- f_oneway function utilized
- The p-value is 0.43204138694325955

Compare the p-value with α:
- As the p-value 0.43204138694325955 > level of significance --> fail to reject the null hypothesis.

CONCLUSION

- Two-sample independence t-test was analyzed to identify whether users spend more time on ne landing page than the previous page
  - P-value --> 0.0001 was determined from the test
  - P-value < level of significance --> Null Hypothesis rejected
    - Thus, there is significant evidence that the mean time spent on the new page > mean time spent on old page
- Two-proportion z-test analyzed to deter if conversion rate for new page > conversion rate of old page
  - P-value --> 0.008 was determined
  - P-value < level of significance --> Null Hypothesis rejected
    - Thus, there significance evidence that conversion rate of new landing page > conversion rate of old landing page
- Chi-square test of independence was investigated to analyze if there is correlation between conversion rate and preferred language
  - P-value --> 0.213 was determined
  - P-value > level of significance of 5% --> Null Hypothesis rejected
    - Thus, conversion rate and preferred language of landing page are independent of each other
- One-Way ANOVA test was performed is deter if time spent on new landing page differed based on preferred language
  - P-value --> 0.432 was populated
  - P-value > 5% level of significance --> Null Hypothesis failed to be rejected
    - Thus, mean time spent on the new landing page was relative identical

NOTE:

Average time spent on new page is greater than old page

Conversion status does not rely on language preferred

New page is converting users into subscribers

Time spent on page is not much different within language preferred

RECOMMENDATIONS

- E-News should promote new landing page because it gains more attraction than the old landing page
- More study should be conducted to better understand how new page is generating more time spent on it by the users
- Further study to investigate on how the new landing page is converting users into subscribers
- Conduct survey among the users within the control and treatment group to access the input regarding old and new landing page
- Conduct survey among subscribers to understand reasons for the conversion

THANK YOU