

The background is a deep blue gradient with a starry space texture. On the left side, there are several overlapping circular patterns. One large circle has a scale from 140 to 260 in increments of 10, with tick marks. Other circles are smaller and some have dashed outlines. Arrows indicate a clockwise direction of movement for these circular elements.

E-NEWS EXPRESS STATISTICAL ANALYSIS

BY: SYEDA AMBREEN KARIM BOKHARI

CONTENTS

- Background
- Business Problem Overview and Solution Approach
- Data Overview
- Exploratory Data Analysis
- Hypothesis Testing
- Conclusions & Recommendations

BACKGROUND

E-News Express is an online news portal which aims to expand its business by acquiring new subscribers. Every visitor to the website takes certain actions based on their interest. The company plans to analyse these interests and wants to determine whether a new feature will be effective or not.

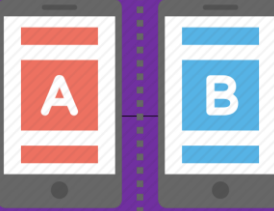
The logo for Enews Express, featuring the word "Enews" in black and "Express" in red, both in a serif font.

BUSINESS PROBLEM

The design team of the company has created a new landing page and wants to decide whether the new landing page is more effective to gather new subscribers.

Task: Perform A/B testing on the landing page

100 randomly selected Users were divided equally into two groups:

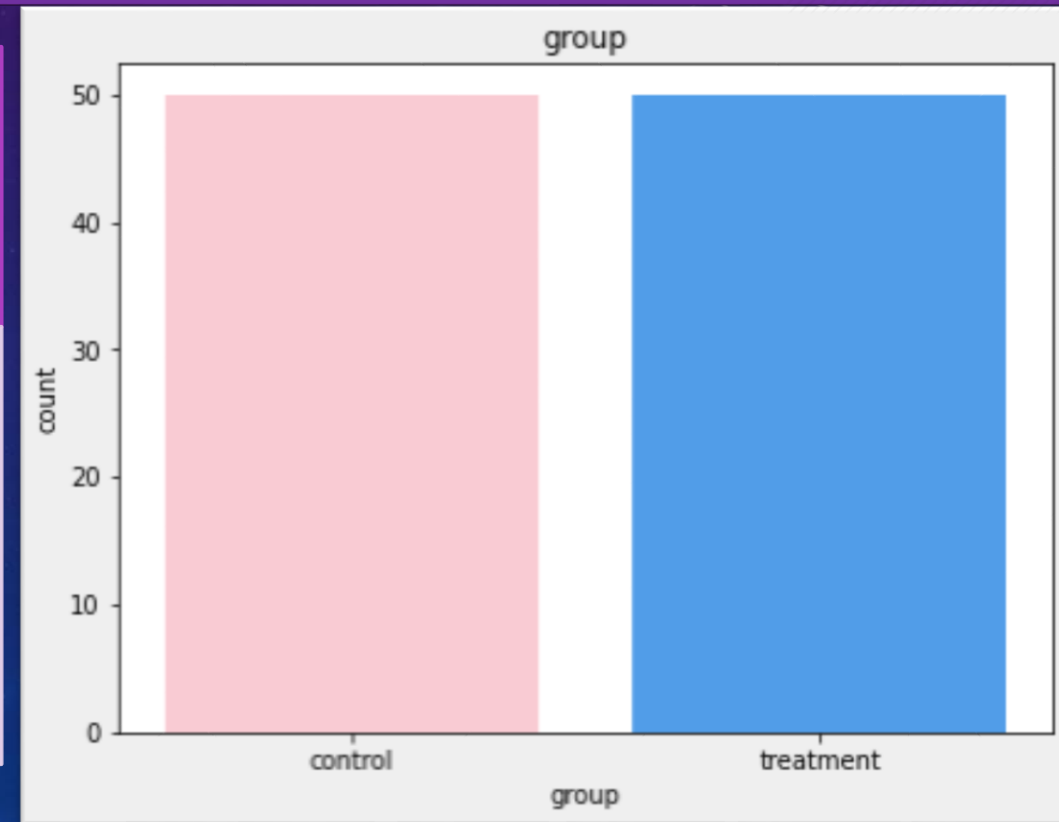


Old landing
page

- Control Group.
- 50 Users

New landing
page

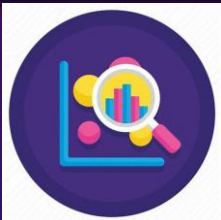
- Treatment Group
- 50 Users



SOLUTION APPROACH

Perform the statistical analysis to answer the following questions using the collected data.

Perform Exploratory Data Analysis.



- Perform statistical summary on data
- Perform Univariate , Bivariate and Multivariate analysis

Perform Hypothesis testing to answer these questions:



- Do the users spend more time on the new landing page than the 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?
- Does the converted status depend on the preferred language?
- Is the mean time spent on the new page same for the different language users?

DATA OVERVIEW

File abtest.csv contains the customer data with following specifications:

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

Observations	Variables
100	6
<p><u>Note:</u></p> <ol style="list-style-type: none">There are no missing values in the data.time_spent_on_the_page is the only numeric variable.The following variables were changed to category data type:<ol style="list-style-type: none">user_idgrouplanding_pageconvertedlanguage_preferred	

EXPLORATORY DATA ANALYSIS: TIME SPENT ON THE PAGE(IN MINUTES)

All 100 users statistics of time_spent_on_the_page(in minutes)

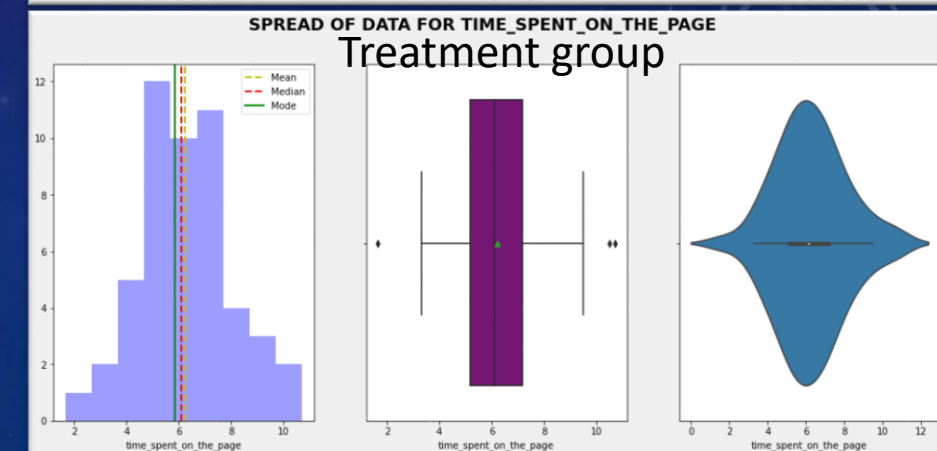
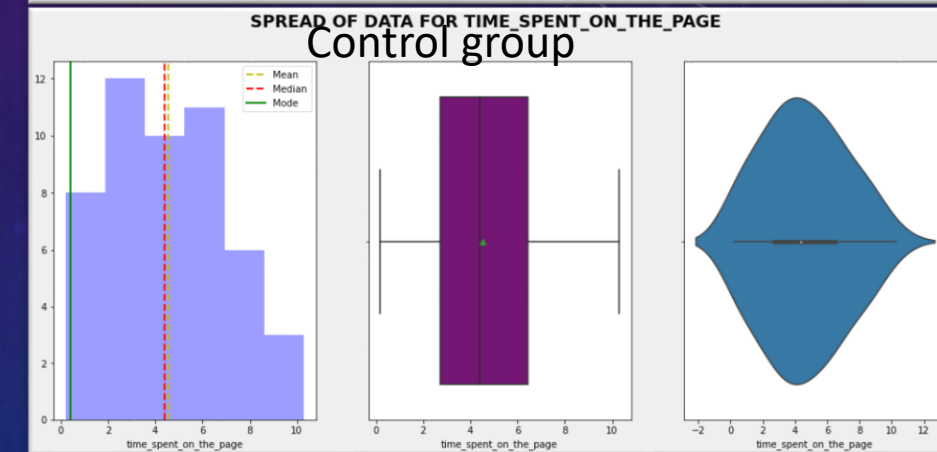
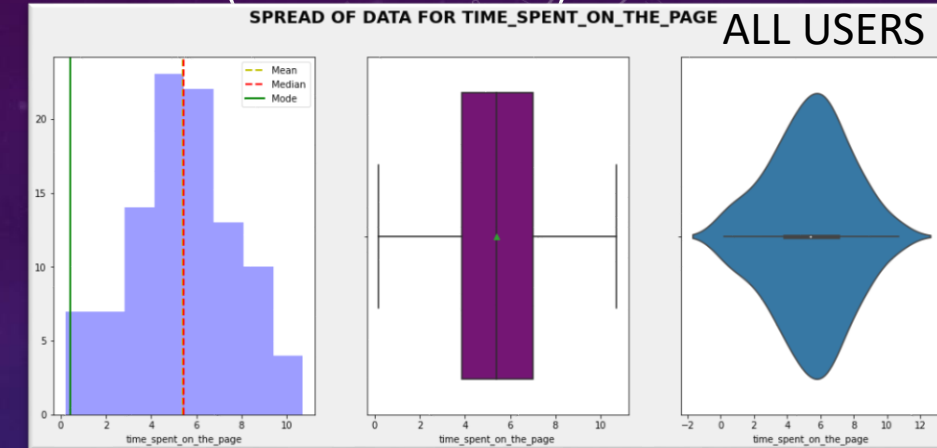
- Mean time spent on page by a user = 5.378
- Median = 5.415 , Multiple modes
- Standard deviation = 2.378, Variance = 5.6556
- Maximum time spent=10.710 ,Minimum time = 0.190
- 50% of users spent between 3.880 - 7.023

50 Control group (old page)users statistics

- Mean: 4.532, Median: 4.38, multi-modal
- 50% users spent:2.27- 6.443
- Standard deviation: 2.582, Variance: 6.667
- Minimum: 0.19, Maximum: 10.3

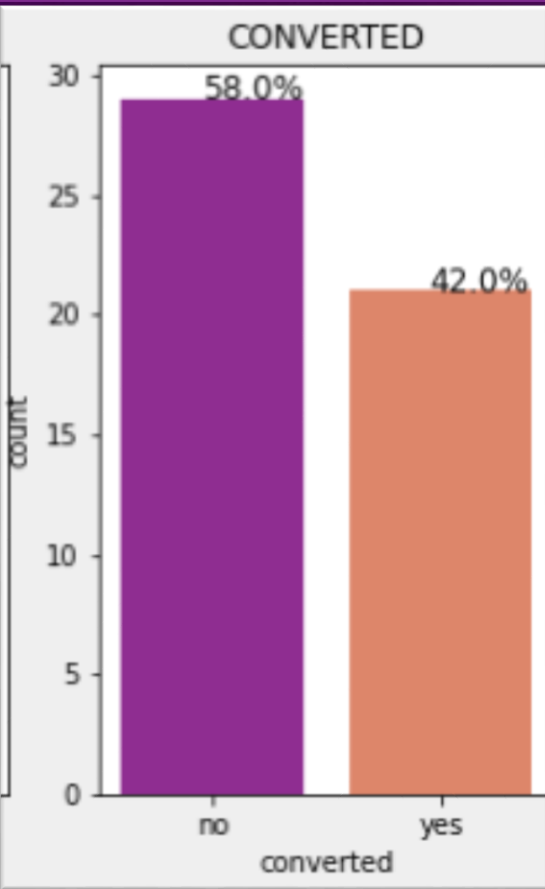
50 Treatment group (new page)users statistics

- Mean: 6.2232, Median: 6.105
- 50% users spent: 5.175- 7.16
- Standard deviation: 1.82, Variance: 3.3016
- Minimum: 1.65, Maximum: 10.71



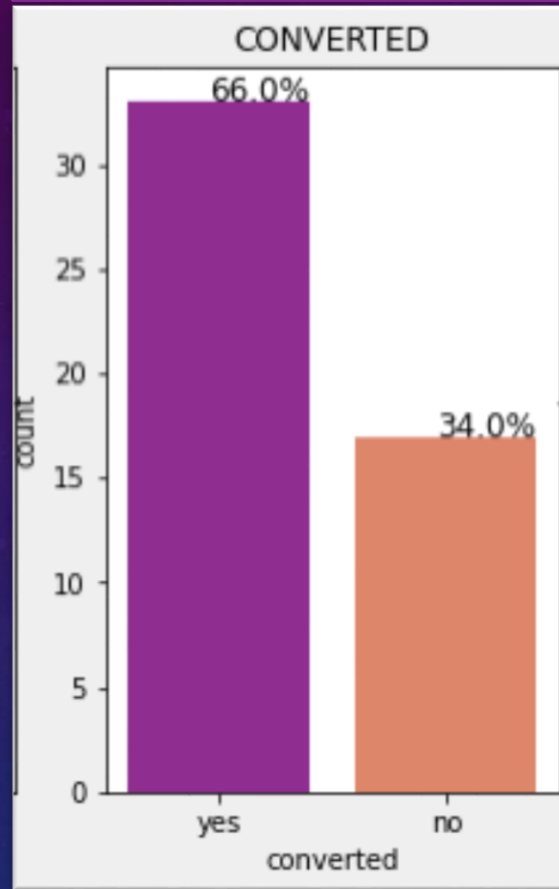
EXPLORATORY DATA ANALYSIS: CONVERTED USERS PERCENTAGE COUNT

Control Group



- Out of 50 users
- 58% are converted
- 42% not converted

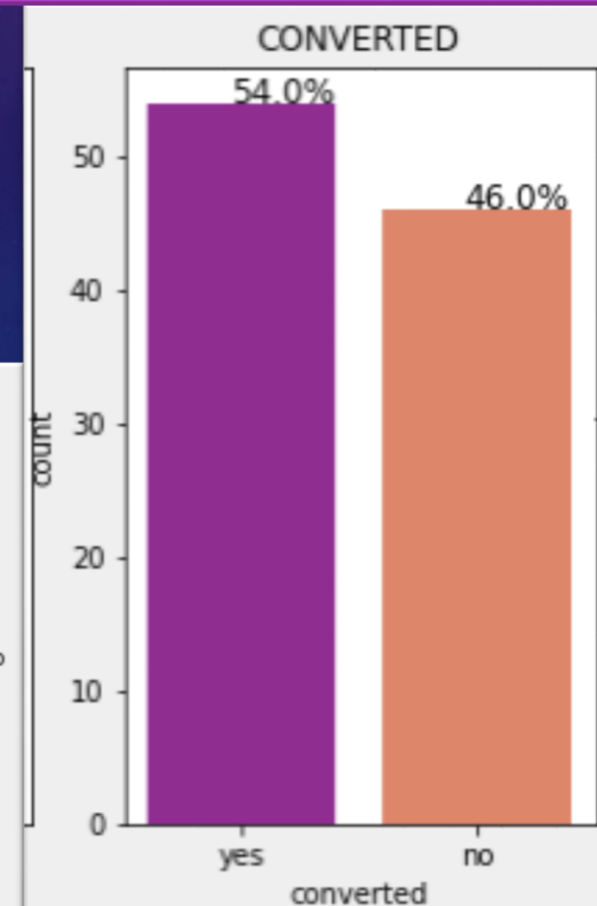
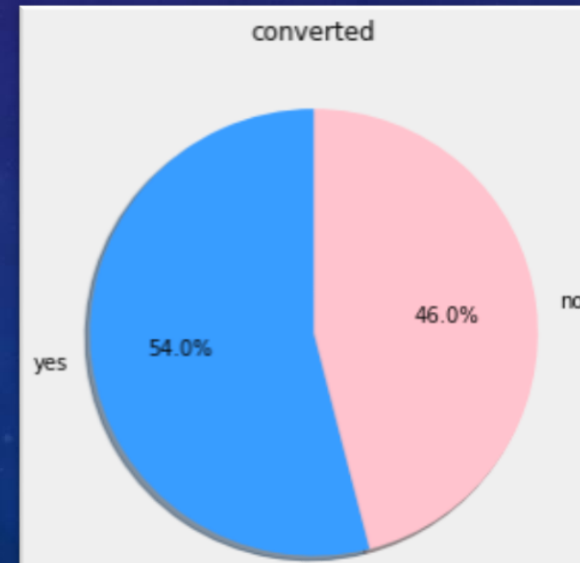
Treatment Group



- Out of 50 users
- 66% are converted
- 34% not converted

All 100 users

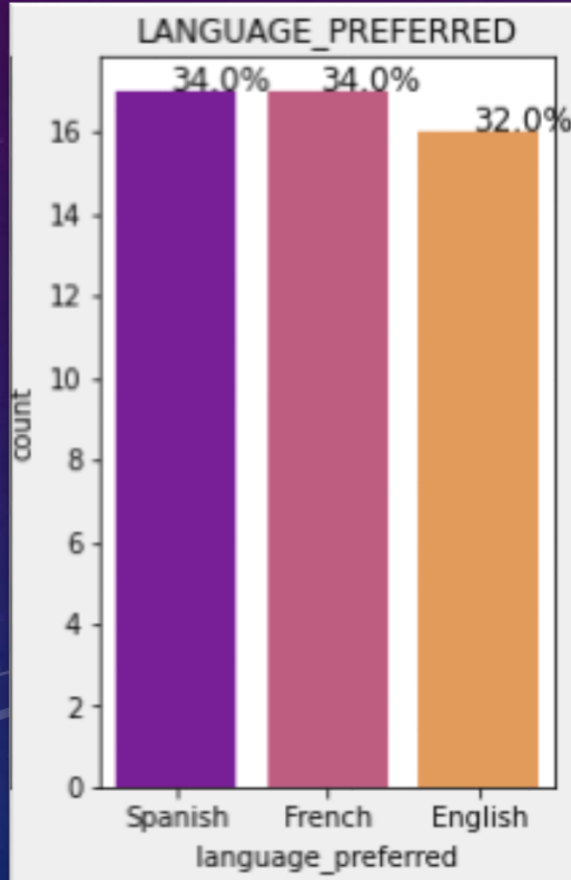
- Out of 100 users
- 54% are converted
- 46% not converted



EXPLORATORY DATA ANALYSIS: CONVERTED USERS

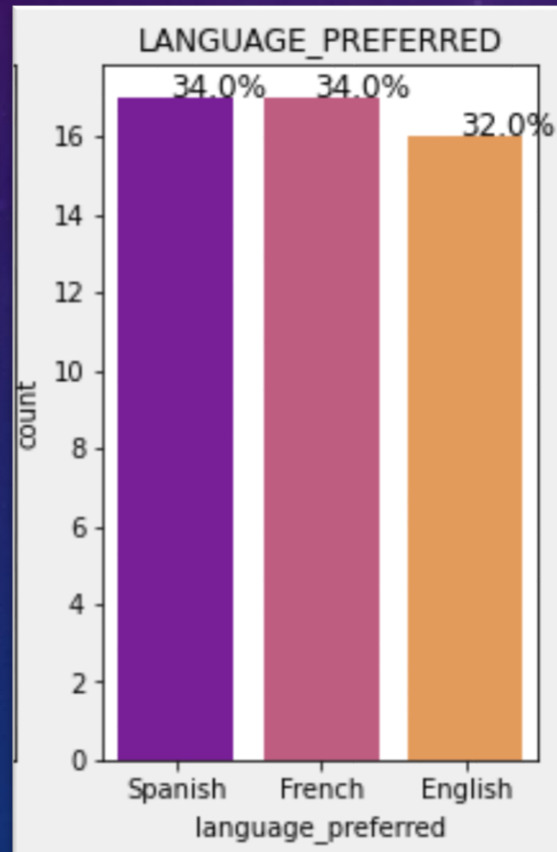
Control Group

- Out of 50 users
- 34% preferred Spanish
- 34% preferred French
- 32% preferred English



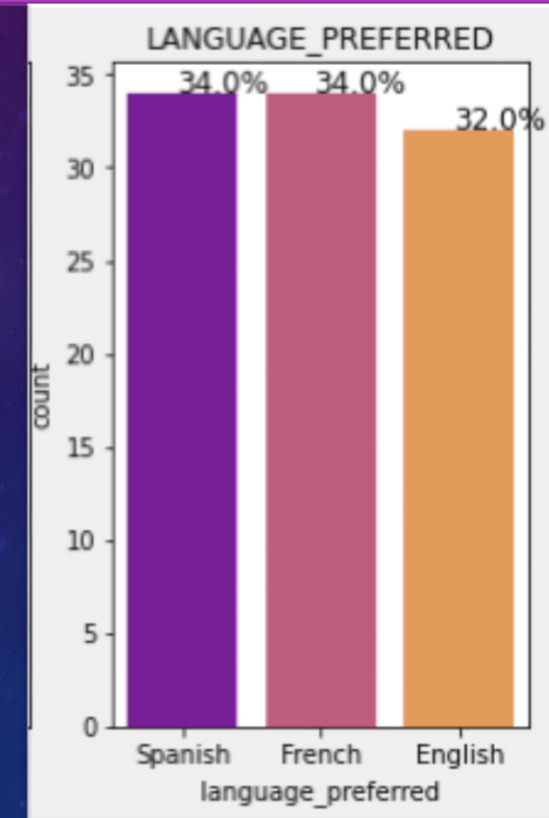
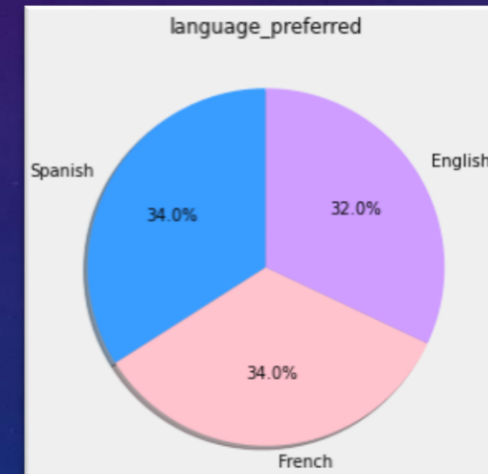
Treatment Group

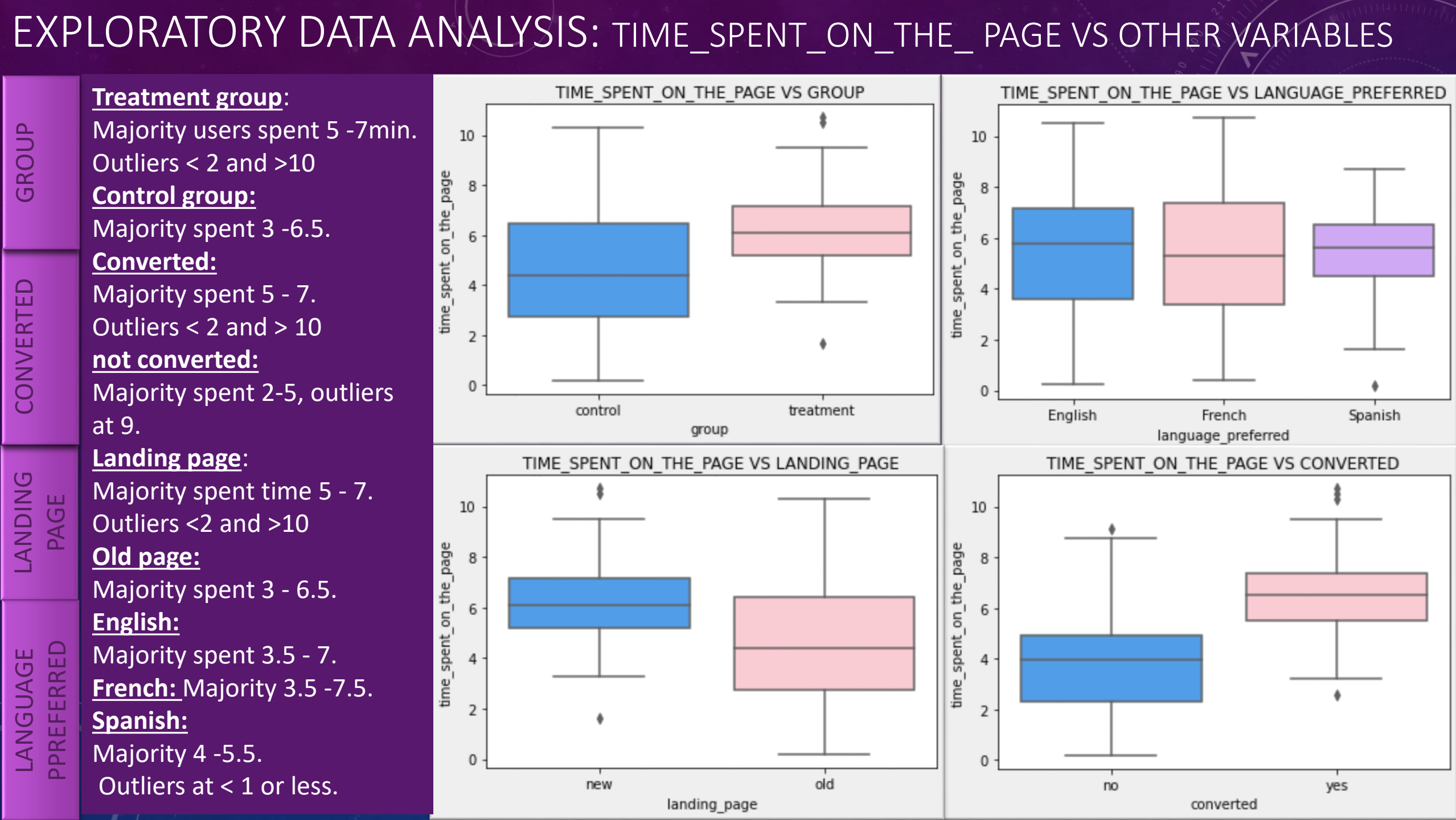
- Out of 50 users
- 34% preferred Spanish
- 34% preferred French
- 32% preferred English



All 100 users

- Out of 100 users
- 34% preferred Spanish,
- 32% preferred English
- 34% preferred French





EXPLORATORY DATA ANALYSIS: TIME_SPENT_ON_THE_PAGE VS OTHER VARIABLES

ENGLISH

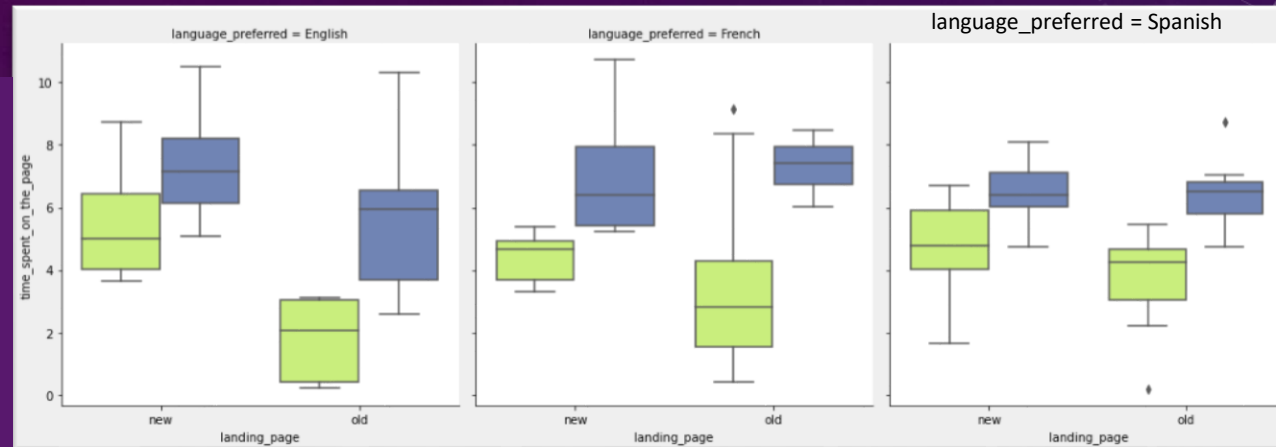
English:

New page: Majority spent 3-7 mins.
Majority converted than not.

Converted spent 6-8 mins
Not converted spent 4-6.5

Old page: Majority spent 2-7 mins.
Majority converted than not.

Converted spent 3.5-6.5
Not converted spent 0.5-3



FRENCH

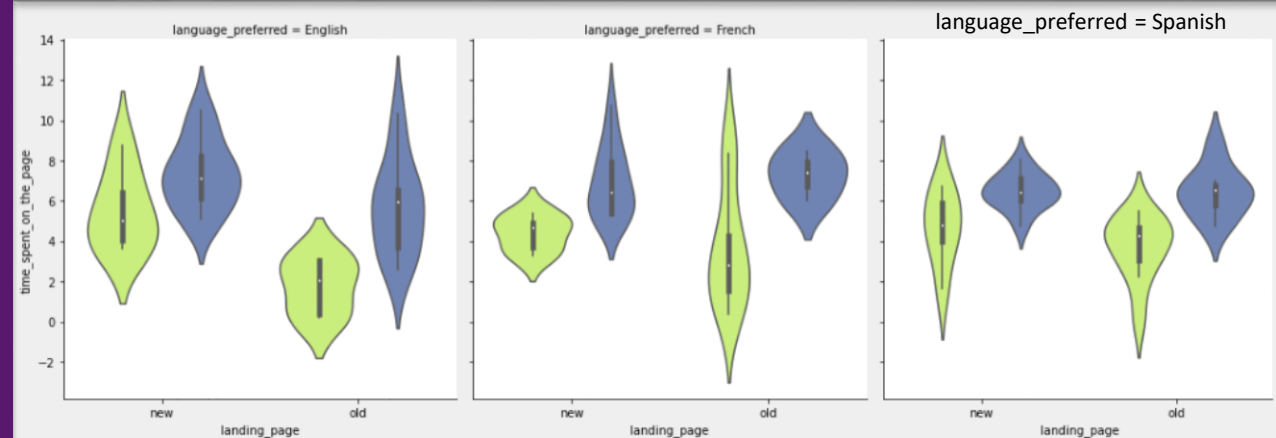
French:

New page: Majority spent 4-8 mins.
Majority converted than not.

Converted spent 5-8
Not converted spent 3-5

Old page: Majority spent 2-5 mins.
Majority not converted than converted.

Converted spent 7-8
Not converted spent 1-4.5



SPANISH

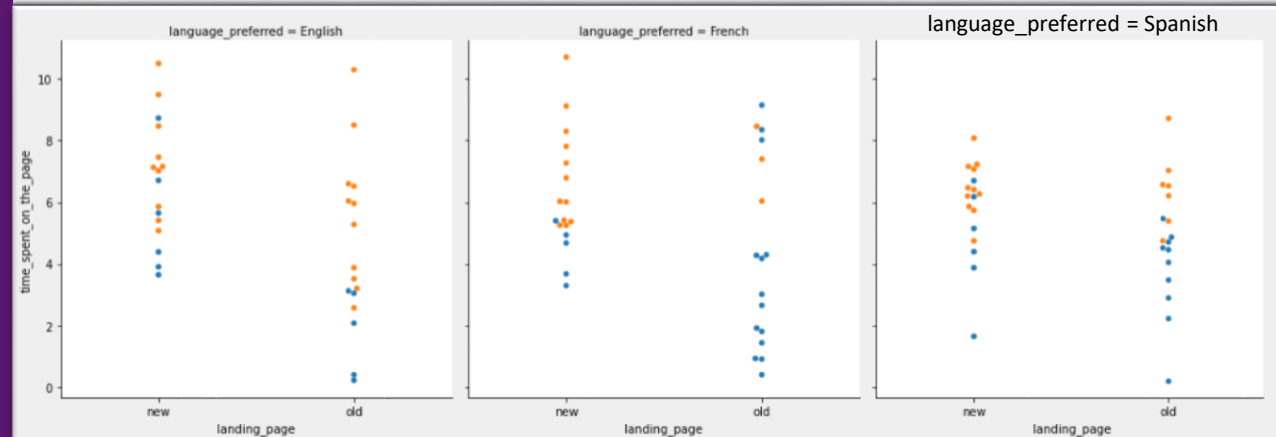
Spanish:

New page: Majority spent 4-8 mins.
Majority converted than not.

Converted spent 6-7
Not converted spent 4-6

Old page: Majority spent 2-7 mins.
Almost equally converted and not converted.

Converted spent 6-7
Not converted spent 3-4.5



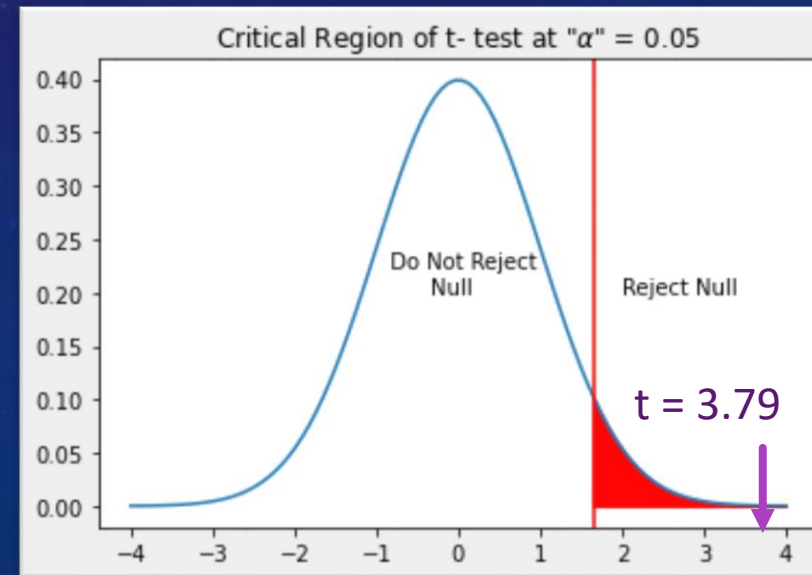
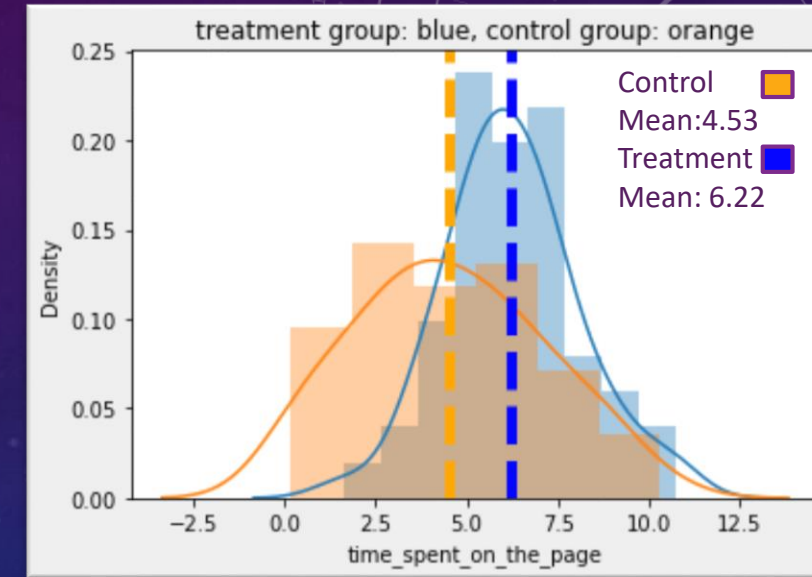
HYPOTHESIS TESTING: TWO SAMPLE T TEST FOR INDEPENDENT SAMPLES

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

- Let μ_1, μ_2 be the mean time spent on the page by treatment group and control group respectively. We will test
- Null hypothesis: time spent by user on the new and old landing page is same.
 $H_0: \mu_1 = \mu_2$ against
- Alternate hypothesis: time spent by user on new page is greater than old page.
 $H_a: \mu_1 > \mu_2$
- level of significance $\alpha = 0.05$
- Two Independent Random Samples: 1. control and 2. treatment groups.
- Distributions are normal as sample size > 30 (50 users for each group)
- Data meets all assumptions of two independent sample t-test for equality of means.
- Mean time spent by treatment group: 6.22 is higher than mean time spent by control group: 4.53

Results:

- p-value is 0.00013 which is very less than the α 0.05
- test statistic is 3.79 which also lies in the critical region.
- So we can reject the null hypothesis that mean time spent on the page by old and new users is same.
- We can conclude that users spend more time on new page than old page.



HYPOTHESIS TESTING: TWO SAMPLE PROPORTION Z-TEST

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

- Let p_1, p_2 be the proportion of users visiting the new and old page who got converted, respectively. We will test
- Null hypothesis: The conversion rate of users of both new and old pages is same.

$$H_0: p_1 = p_2 \text{ against}$$

- Alternate hypothesis: The conversion rate of users of new page is greater than old page.

$$H_a: p_1 > p_2$$

- level of significance $\alpha = 0.05$
- Two Independent Random Samples: 1. control and 2. treatment groups.
- Distributions are normal as sample size > 30 (50 users for each group)
- Binomially distributed population : converted or not converted.
- Data meets all assumptions of **two sample proportion z-test** for finding proportion of converted out of both the treatment and control group.

- n is sample size.
- Both np_1 and $n(1-p_1)$ are greater than 10
- Both np_2 and $n(1-p_2)$ are greater than 10

$$np_1 = 50 \cdot \frac{33}{50} = 33 \geq 10$$

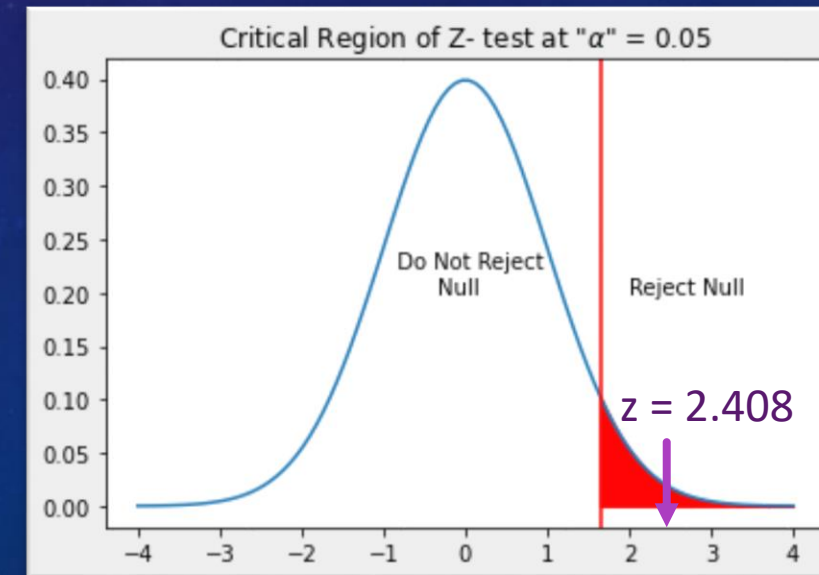
$$n(1 - p_1) = 50 \cdot \frac{50-33}{17} = 17 \geq 10$$

$$np_2 = 50 \cdot \frac{21}{50} = 21 \geq 10$$

$$n(1 - p_2) = 50 \cdot \frac{50-21}{29} = 29 \geq 10$$

Results:

- p-value is 0.00802 which is very less than the alpha 0.05
- Test statistic is 2.408 which lies in the rejection region.
- so we can reject the null hypothesis that conversion rate of users of old and new page is same, rather new page users have a higher conversion rate.
- We can conclude that conversion rate of new page is greater than old page.



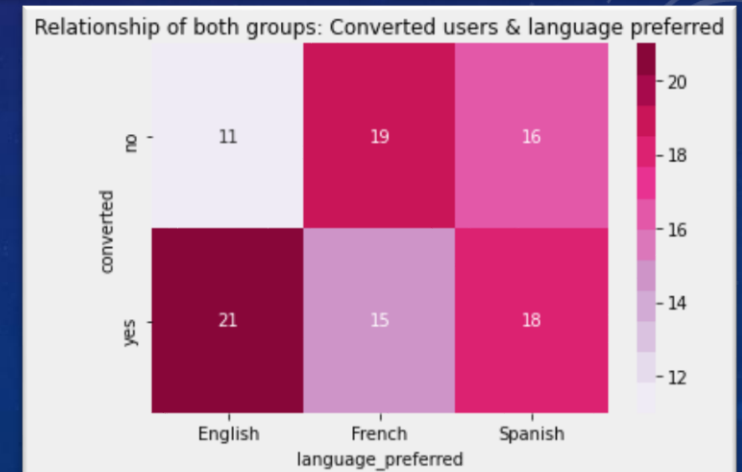
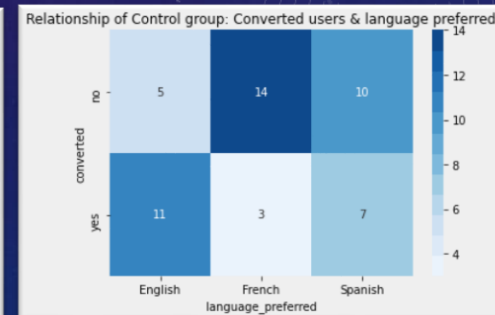
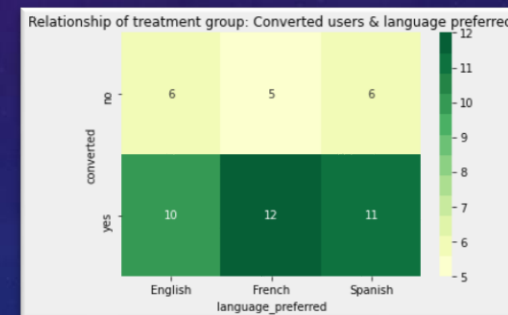
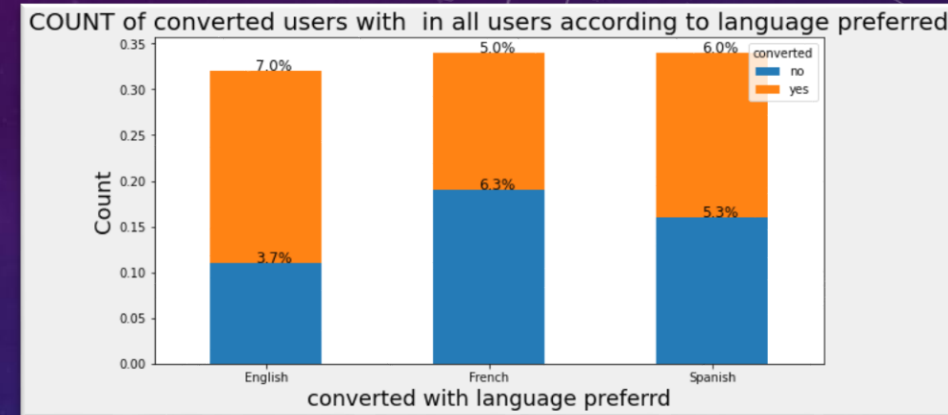
HYPOTHESIS TESTING: χ^2 test of Independence

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

- We will test the dependency of converted status on preferred language.
- Null hypothesis H_0 : the converted status is independent of preferred language.
against
- Alternate hypothesis H_a : The converted status depends on preferred language.
- level of significance $\alpha = 0.05$
- Two Independent Random Samples: 1. control and 2. treatment groups.
- The number of observations in each level is greater than 5.
- Categorical variables
- Data meets all assumptions of **χ^2 test for contingency** for finding dependency between converted status and preferred language.

Results:

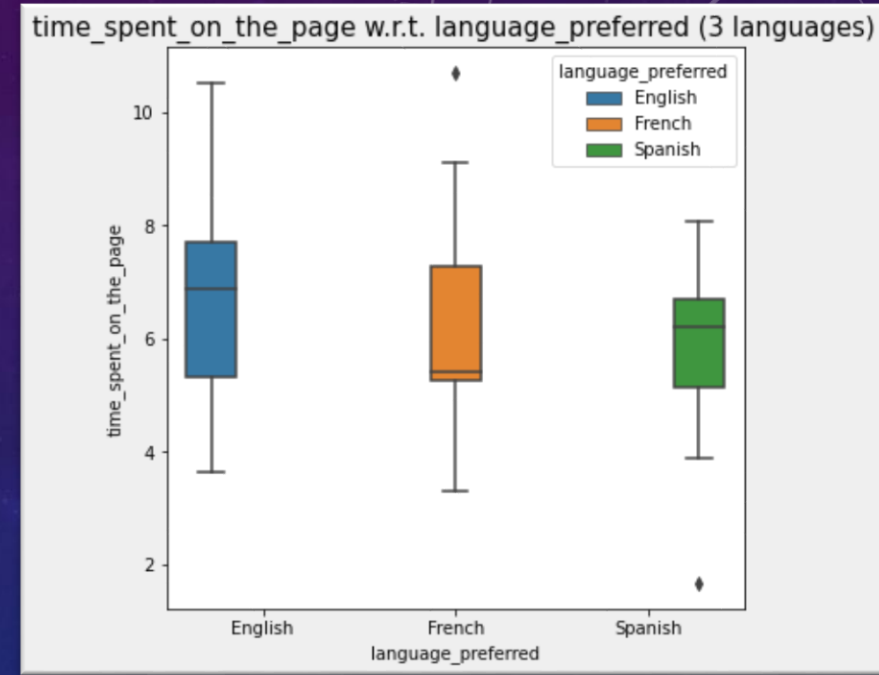
- p-value is 0.213 which is greater than the alpha 0.05
- Test statistic is 3.093 which does not lie in the rejection region,
- so we fail to reject the null hypothesis that conversion rate of users is independent of preferred language.
- We can conclude that converted status is independent of preferred language.



HYPOTHESIS TESTING: MULTIPLE COMPARISON TEST (TUKEY HSD)

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

- Let μ_1, μ_2, μ_3 be the mean time spent on the new page by English, French and Spanish users respectively. We will test
- Null hypothesis: time spent by on the new page by English, French and Spanish users is same.
 $H_0: \mu_1 = \mu_2 = \mu_3$ against
- Alternate hypothesis: time spent by on the new page by English, French and Spanish users is changed for at least one language user..
 $H_a: \mu_1 \neq \mu_2$ or $\mu_2 \neq \mu_3$ or $\mu_1 \neq \mu_3$
- level of significance $\alpha = 0.05$
- Distributions are normal as sample size > 30 (50 users for each group)
- Data meets all assumptions of **Multiple Comparison test (Tukey HSD)** for equality of means.
- Mean time spent by users of English 6.664, French 6.1965 Spanish 5.835



- level of significance $\alpha = 0.05$

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
English	French	-0.4673	0.7259	-2.0035	1.069	False
English	Spanish	-0.8285	0.401	-2.3647	0.7078	False
French	Spanish	-0.3612	0.816	-1.874	1.1516	False

Results:

- As the p-values (p-adj column) for comparing the mean time_spent_on_the_page for all language pairs, all the p-adj value are more than the significance level: $\alpha = 0.05$,
- we failed to reject the null hypothesis of equality of all population means
- Thus, we can say that the mean time spent by different language users is similar.

CONCLUSION:

- Insights about New page vs Old page:
- Most (50%) users spent at between 5-7 (mean time: 6.22)mins time on the page, which is higher as compared to old page users, 50% of whom spent between 2-6 mins (mean time: 4.53) on the page.
 - Users have higher conversion rate 66% as compared to old page conversion rate:42%.
 - Language preference does not seem to play much role in user conversion in both the pages: Spanish:34%, French:34%, English 32%
- Mean time that users spend on the old vs new page is different, time users spent on new page is higher.
- New page users do have higher conversion rate.
- Conversion rate is independent of preferred language
- Means of all populations of different languages users is similar.

RECOMMENDATIONS:

- New landing page seems to get higher conversion rate and better response from the users:
- Most (50%) users spent at between 5-7 mins time on the page which is higher than the old page.
- Users have higher conversion rate 66% as compared to old page conversion rate:42%.
- Language preference does not seem to play much role in user conversion in both the pages:
Spanish:34%, French:34%, English 32%
- Mean time spent by different Language users is similar.
- New landing page is more effective to gather new subscribers.
- Company should continue with their new landing page.



THE END

Prepared by: Syeda Ambreen Karim Bokhari.