E-NEWS EXPRESS PROJECT

SNEHA SACHIN

PROJECT OBJECTIVE

Companies often analyze users' responses to two variants of a product to decide which of the two variants is more effective. The objective of this project for E-News Express will be to determine whether the new landing page is more effective to convert subscribers compared to old landing page.

Statistical analysis of business data shall be performed Including exploration of the dataset and extracting insights from the data to answer the following questions:

1. Explore the dataset and extract insights using Exploratory Data Analysis.
2. Do the users spend more time on the new landing page than the old landing page?
3. 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?
4. Does the converted status depend on the preferred language?
5. Is the mean time spent on the new page same for the different language users?

ABOUT DATA

Variable	Description
user-id	Unique ID of the customers.
group	Type of groups viewing the online page(Treatment & Control)
landing_page	Customers viewing on kind of page(New & Old)
time_spent_on_the_page	Time spent by the customers in viewing the pages.
Converted	Viewers who subscribed the page.
language_preferred	Languages offered to view the page(English, French & Spanish)

Observations	Variables
100	6

Notes

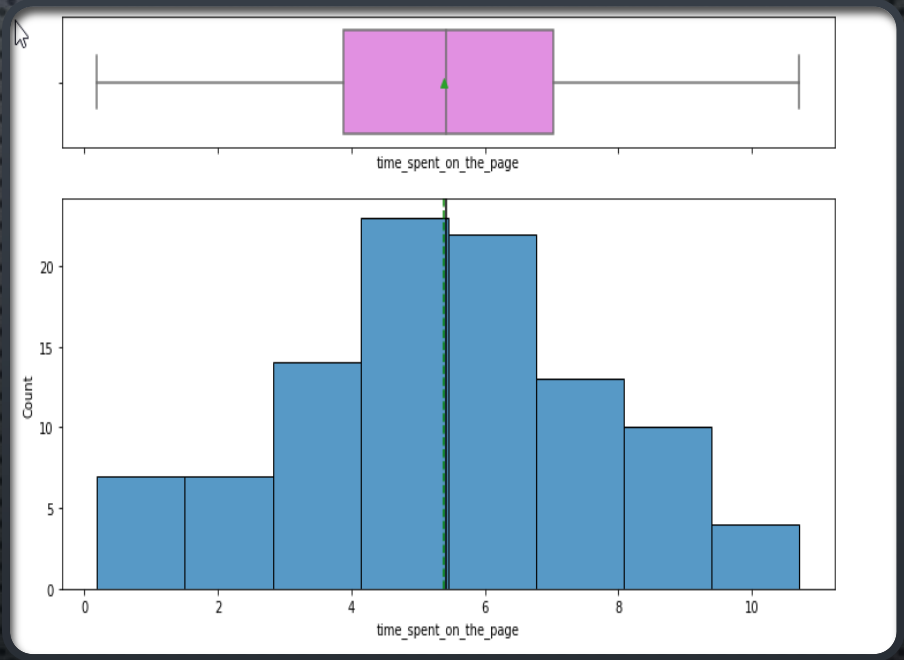
- There are no null values in the dataset.
- The group, landing_page, converted and language_preferred have been converted to category type of data.

Data Insights

- There are 100 users in the data set.
- 50 users in control group and 50 in treatment group.
- Control group is assigned to old landing page whereas treatment group is assigned to new landing page.
- 32 users prefer English whereas 34 users each prefer French and Spanish languages.
- Out 100 users in the dataset, 54 users have converted. This includes users in both control group and treatment group.

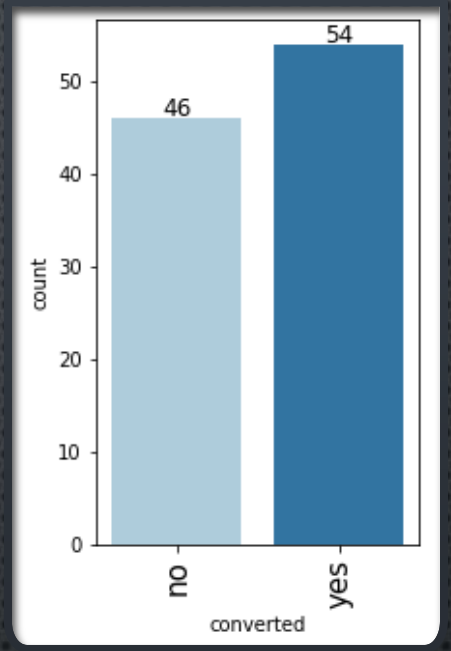
EXPLORATORY DATA ANALYSIS – TIME SPENT

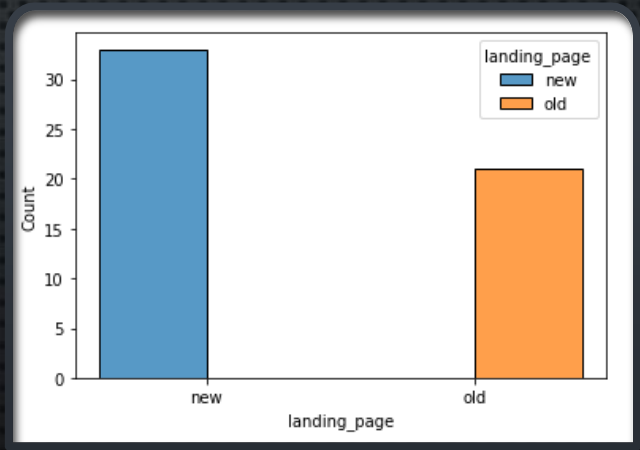
- THE TIME SPENT VARIABLE IS NORMALLY DISTRIBUTED ACROSS THE DATASET.
- THE MODE OF THE DISTRIBUTION INDICATES THAT A LARGE CHUNK OF USERS SPEND AROUND 4 TO 6 MINUTES ON THE LANDING PAGE, THE AVERAGE TIME BEING AROUND 5.37 MINUTES.
- THE PLOT ALSO TELLS THAT THERE ARE MORE USERS WHO ARE SPENDING MORE THAN 5 MINUTES ON THE LANDING PAGE COMPARED TO THE USERS SPENDING LESS THAN 5 MINUTES.



EXPLORATORY DATA ANALYSIS – CONVERSION RATE

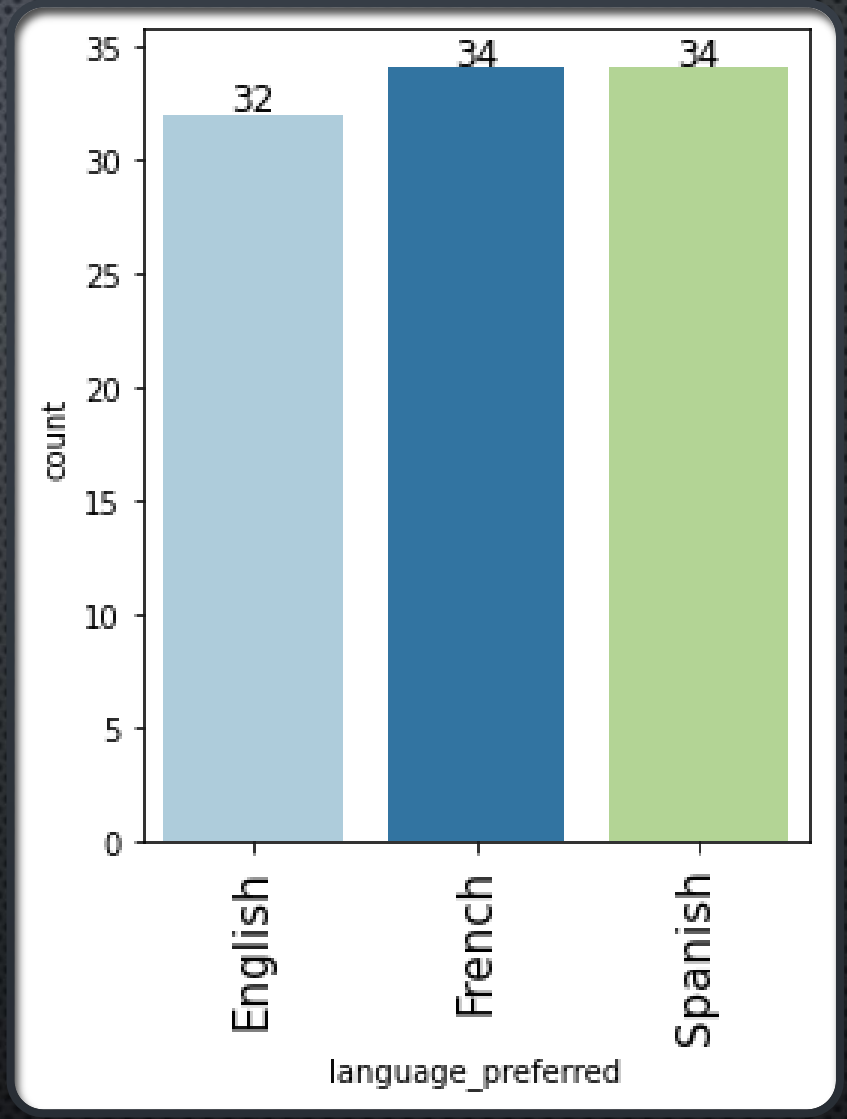
- 54% USERS CONVERTED OVERALL (OLD + NEW LANDING PAGE).
- 21% USERS CONVERTED USING OLD LANDING PAGE
- 33% USERS CONVERTED USING NEW LANDING PAGE



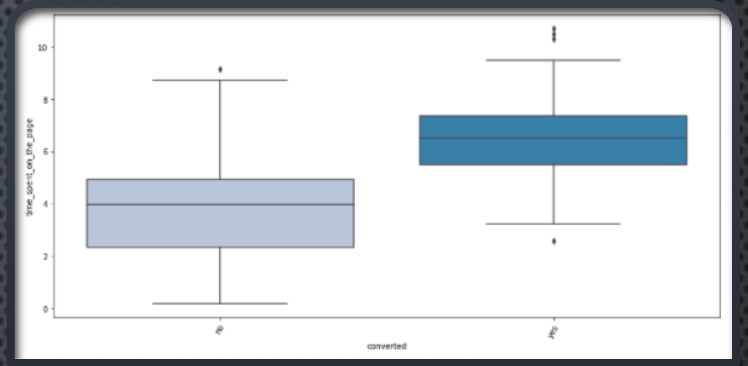


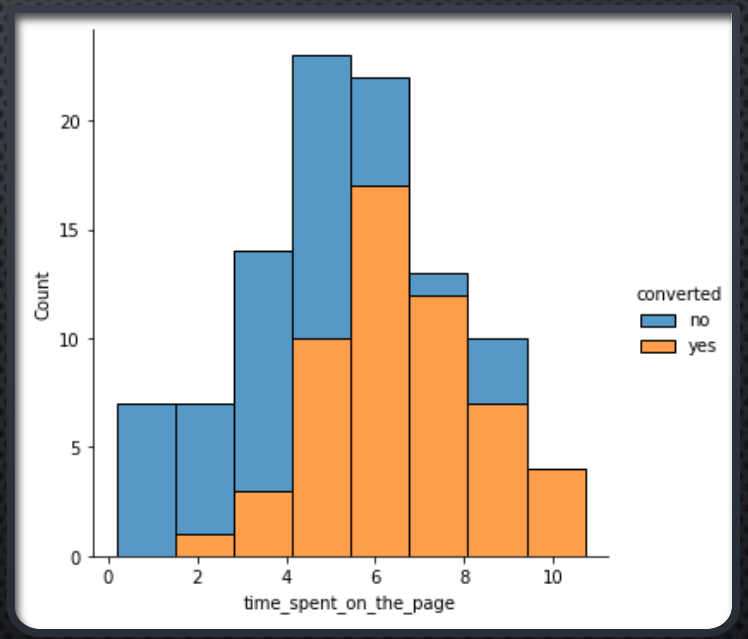
EXPLORATORY DATA ANALYSIS – LANGUAGE

- From the overall dataset, there are 3 unique language groups in the dataset which is English, Spanish and French.
- There are equal readers in Spanish and French language at 34 users, however the English readers is slightly less compared to the other two at 32 users.



EDA – TIME SPENT VS CONVERSION





- As shown, the users tend to convert when they spend more time on landing page.
- The average time spent by the users who did not convert is 3.91 minutes whereas the average time spent by the users who did convert is 6.62 minutes.

EDA – CONVERSION VS LANGUAGE

- The overall conversion rate is 54% however
- The conversion rate for English readers went from 34% from the old landing page to 31% in the new landing page
- The conversion rate for French readers significantly jumped from 9% from the old landing page to 35% in the new landing page
- There was also significant jump for the Spanish readers from 21% from the old landing page to 32% in the new landing page
- Hence the new landing page is successfully making the readers subscribe for it

	Overall	Landing Page	
		Old	New
Conversion Rate			
English	32	11 (34%)	10 (31%)
French	34	3 (9%)	12 (35%)
Spanish	34	7 (21%)	11 (32%)
Overall	**100**	**21**	**33**

EDA – TIME SPENT VS LANGUAGE

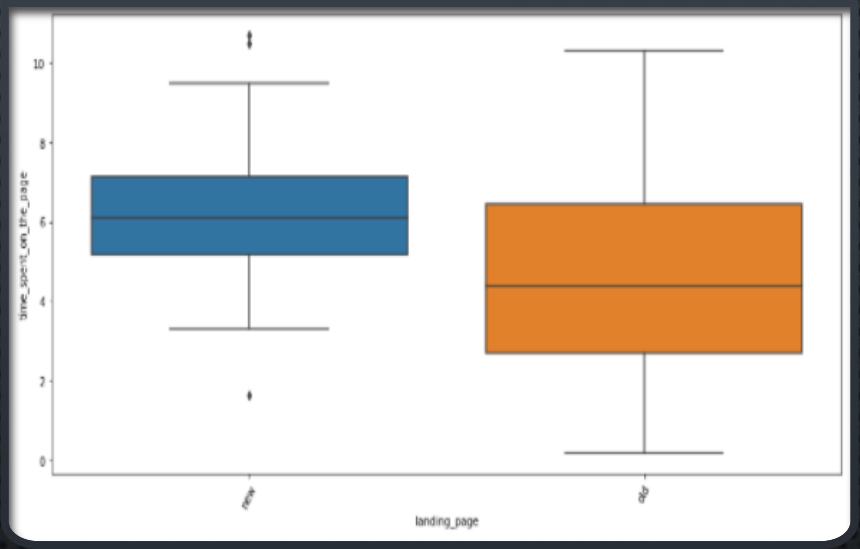
- The chart shows the **MEAN** time spent for overall as well as old and new landing page
- **English** – Overall spent 5.55 minutes. However, time spent increased from 4.45 minutes to 6.66 minutes from old to new landing page (additional 2.21 minutes average).
- **French** – overall spent 5.25 minutes. However, time spent increased from 4.31 minutes to 6.19 minutes from old to new landing page (additional 1.88 minutes average).
- **Spanish** - overall spent 5.33 minutes. However, time spent increased from 4.82 minutes to 5.83 minutes from old to new landing page (additional 1.01 minutes average).

	Overall	Landing Page	
		Old	New
Time Spent on the Page (Mean)			
English (mins)	5.55	4.45	6.66
French (mins)	5.25	4.31	6.19
Spanish (mins)	5.33	4.82	5.83
Overall (mins)	**5.37**	**4.53**	**6.22**

HYPOTHESIS TESTING

ARE USERS SPENDING MORE TIME ON NEW LANDING PAGE?

time_spent_on_the_page							
count	mean	std	min	25%	50%	75%	max
50.0	6.2232	1.817031	1.65	5.175	6.105	7.1600	10.71
50.0	4.5324	2.581975	0.19	2.720	4.380	6.4425	10.30



We will test the null hypothesis. Let $\mu 1$ be the time spent on the new landing page. Let $\mu 2$ be the time spent on the old landing page.

- $H0{:}\mu 1 = \mu 2$ which is users spend equal amount of time in the old and the new landing page, against the alternate hypothesis which is users spend greater amount of time on the new landing page,
- $Ha{:}\mu 1 > \mu 2$.

We are going to use the **two independent sample t-test with the level of significance as 0.05.**

- The p-value from analysis is 0.000139238
- As the p-value is less than the level of significance, we can reject the null hypothesis. We do not have enough evidence to support the claim that the users spend equal amount of time on the new and the old landing page. Hence, we can **conclude that users spend more time on the new landing page.**

IS CONVERSION RATE GREATER ON NEW LANDING PAGE?

Our Null hypothesis will be that the rate of conversion for the new page is equal to the rate of conversion on the old page.

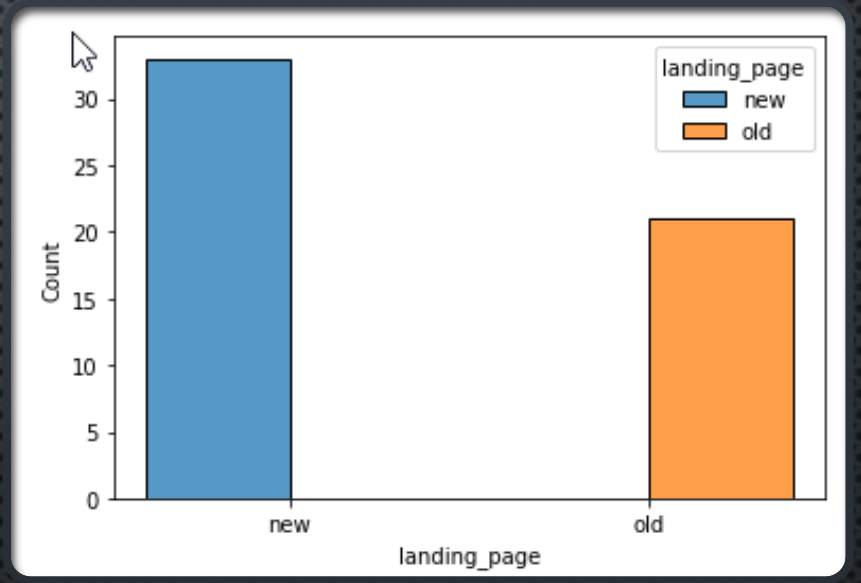
- $H0{:}p1{=}p2$

against the alternate hypothesis which is the rate of conversion for the new page is higher than the rate of conversion on the old page.

- $Ha{:}p1{>}p2$

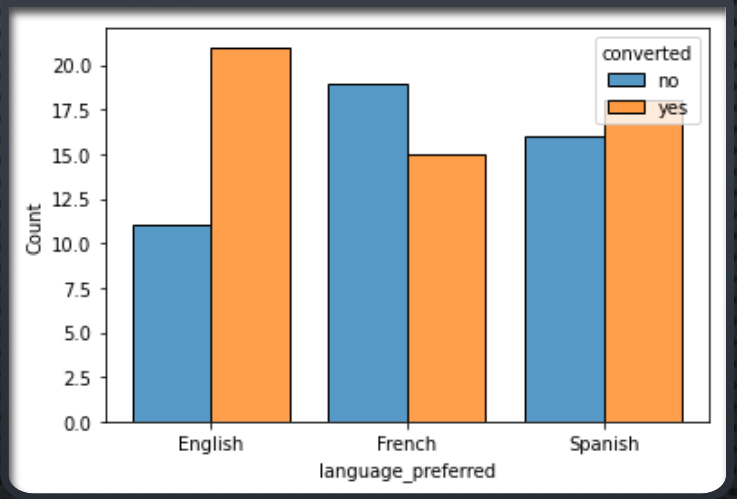
We are going to use the **Two proportions z-test with the level of significance set to 0.05.**

- **The p-value is 0.03522**
- As the p-value is less than the level of significance, we can reject the null hypothesis. We do not have enough evidence to support the claim that the conversion rate for the new page is the same as the conversion rate for the old page. Hence, we can **conclude that the rate of conversion for the new page is higher than the rate of conversion on the old page.**



IS CONVERSION STATUS DEPENDENT ON LANGUAGE?

	Overall	Landing Page	
		Old	New
Conversion Rate			
English	32	11 (34%)	10 (31%)
French	34	3 (9%)	12 (35%)
Spanish	34	7 (21%)	11 (32%)
Overall	**100**	**21**	**33**



OUR NULL HYPOTHESIS IS THAT THE CONVERTED STATUS IS INDEPENDENT OF THE PREFERRED LANGUAGE.

- $H0$: CONVERTED STATUS IS INDEPENDENT OF PREFERRED LANGUAGE. AGAINST THE ALTERNATE HYPOTHESIS WHICH IS THE CONVERTED STATUS DEPENDS ON THE PREFERRED LANGUAGE.
- Ha: CONVERTED STATUS DEPENDS ON PREFERRED LANGUAGE.

SINCE IT INVOLVES CATEGORICAL VARIABLES AND A CONTINGENCY TABLE, WE ARE GOING TO USE **CHI SQUARE TEST OF INDEPENDENCE** WITH THE LEVEL OF SIGNIFICANCE SET TO 0.05.

- THE P-VALUE IS 0.79708
- AS THE P-VALUE IS GREATER THAN THE LEVEL OF SIGNIFICANCE, WE FAIL TO REJECT THE NULL HYPOTHESIS. THEREFORE, WE CAN **CONCLUDE THAT WE DO NOT HAVE ENOUGH EVIDENCE TO SUPPORT THAT CONVERSION STATUS IS DEPENDENT ON PREFERRED LANGUAGE.**

IS MEAN TIME SPENT IN NEW LANDING PAGE DIFFERENT FOR DIFFERENT LANGUAGE USERS?

	Overall	Landing Page	
		Old	New
Time Spent on the Page (Mean)			
English (mins)	5.55	4.45	6.66
French (mins)	5.25	4.31	6.19
Spanish (mins)	5.33	4.82	5.83
Overall (mins)	**5.37**	**4.53**	**6.22**

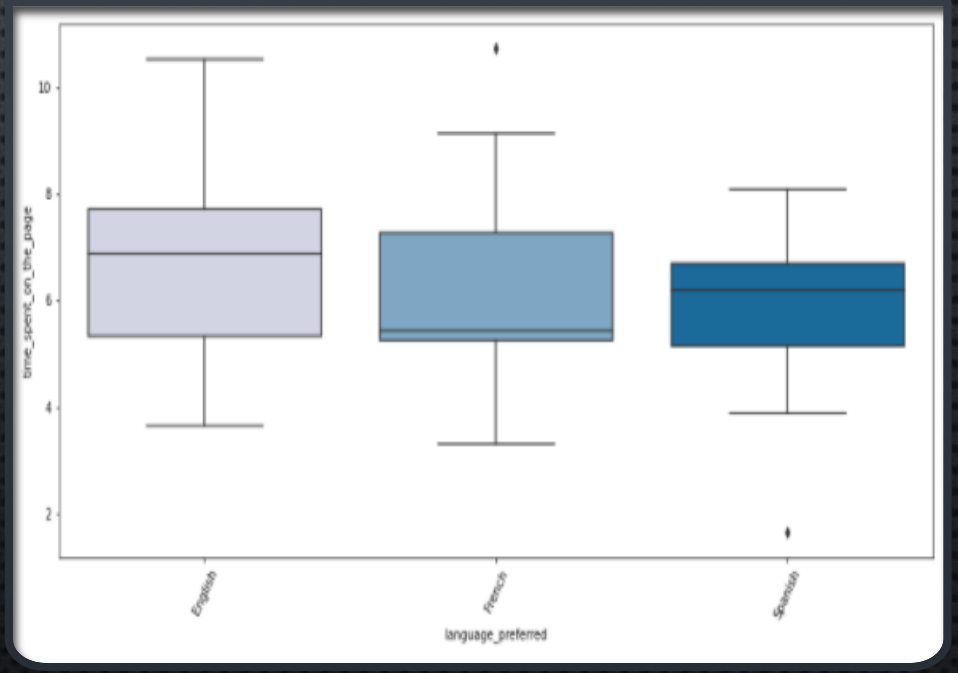
LET $\mu1,\mu2,\mu3$ BE THE MEAN TIME SPENT ON THE NEW PAGE BY ENGLISH, FRENCH AND SPANISH USERS RESPECTIVELY.

WE WILL TEST THE NULL HYPOTHESIS BY STATING THAT THE MEAN TIME SPENT ON THE NEW PAGE BY ALL LANGUAGE USERS IS EQUAL

- $H0:\mu1=\mu2=\mu3$ AGAINST THE ALTERNATIVE HYPOTHESIS WHICH IS ATLEAST ONE MEAN TIME IS UNEQUAL.
- Ha: AT LEAST ONE MEAN IS DIFFERENT FROM THE REST.

WE ARE GOING TO USE **ONE WAY ANOVA F TEST** WITH THE LEVEL OF SIGNIFICANCE SET TO 0.05.

- THE P-VALUE IS 0.43204
- AS THE P-VALUE IS GREATER THAN THE LEVEL OF SIGNIFICANCE, WE FAIL TO REJECT OUR NULL HYPOTHESIS. HENCE, WE HAVE ENOUGH EVIDENCE TO SUPPORT THE CLAIM THAT THE MEAN TIME SPENT ON THE NEW PAGE IS EQUAL ACROSS THE DIFFERENT LANGUAGE USERS.



CONCLUSIONS AND RECOMMENDATIONS

- USERS SPEND MORE TIME ON THE NEW LANDING PAGE COMPARED TO THE OLD LANDING PAGE.
- THE CONVERSION RATE SEEMS TO BE HIGHER FROM THE GROUP USING THE NEW LANDING PAGE.
- THERE IS NOT ENOUGH EVIDENCE TO SUPPORT THAT CONVERSION STATUS IS DEPENDENT ON PREFERRED LANGUAGE.

IN CONCLUSION, WE CAN SAY THAT THE **NEW LANDING WOULD BE SUCCESSFUL TO CONVERT MORE SUBSCRIBERS** COMPARED TO OLD LANDING PAGE.