E_NEWS EXPRESS BUSINESS EXPANSION PROJECT

By: Sushma Rao

Description and Background

- The online News portal wants to expand its business by acquiring new subscribers.
- Every visitor to the website takes certain actions based on their interest.
- The company wants to determine whether a new feature will be effective or not.
- Company has introduced a new variant, and it wants to analyze user's responses to two variants of a product to decide which of the two variants is more effective.
- The Company wants to determine whether the new feature attracts users based on a chosen metric.

Objective

- The main objective of this project is to compare the two variants of a product and to decide which is more effective .
- The design team of the company has designed a new landing page, and the task is to decide whether the new landing page is more effective to gather new subscribers.
- An experimental technique called A/B testing is used to determine the desired results.

An A/B test is an example of a statistical hypothesis testing ,it is a process whereby a hypothesis is made about the relationship between two data sets, and those datasets are then compared against each other to determine if there is statistically significant relationship or not.

Data Overview of the Sample

Data Dictionary

User ID	This represents the user ID of the person visiting the website.
Group	This represents whether the user belongs to the first group (control) or the second group (treatment).
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.

Statistical Analysis

- Our goal here is to identify the true conversion rate of the population ,which is impossible !!
- While running an A/B test ,we are making a hypothesis that the variation B will converse at a higher rate for all of our overall population than variation A will.
- We Display them to a sample instead and observe what happens :

Here we have a sample data set ,which has two categories in the group attribute ,which shows the CONTROL group, which used an old landing page and the TREATMENT group which used the new landing page.

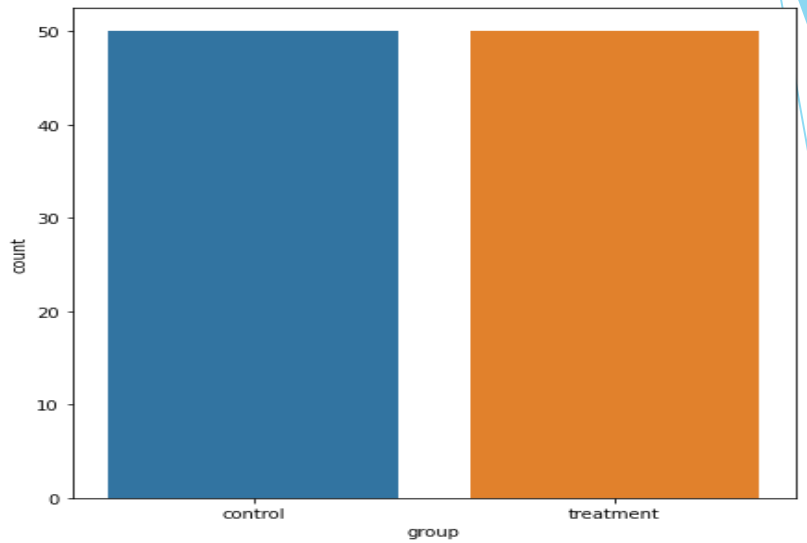
  - *Divide the data set into two, one for the control group and the other for the treatment group , and then perform the hypothesis testing at a significance level of 5%.*

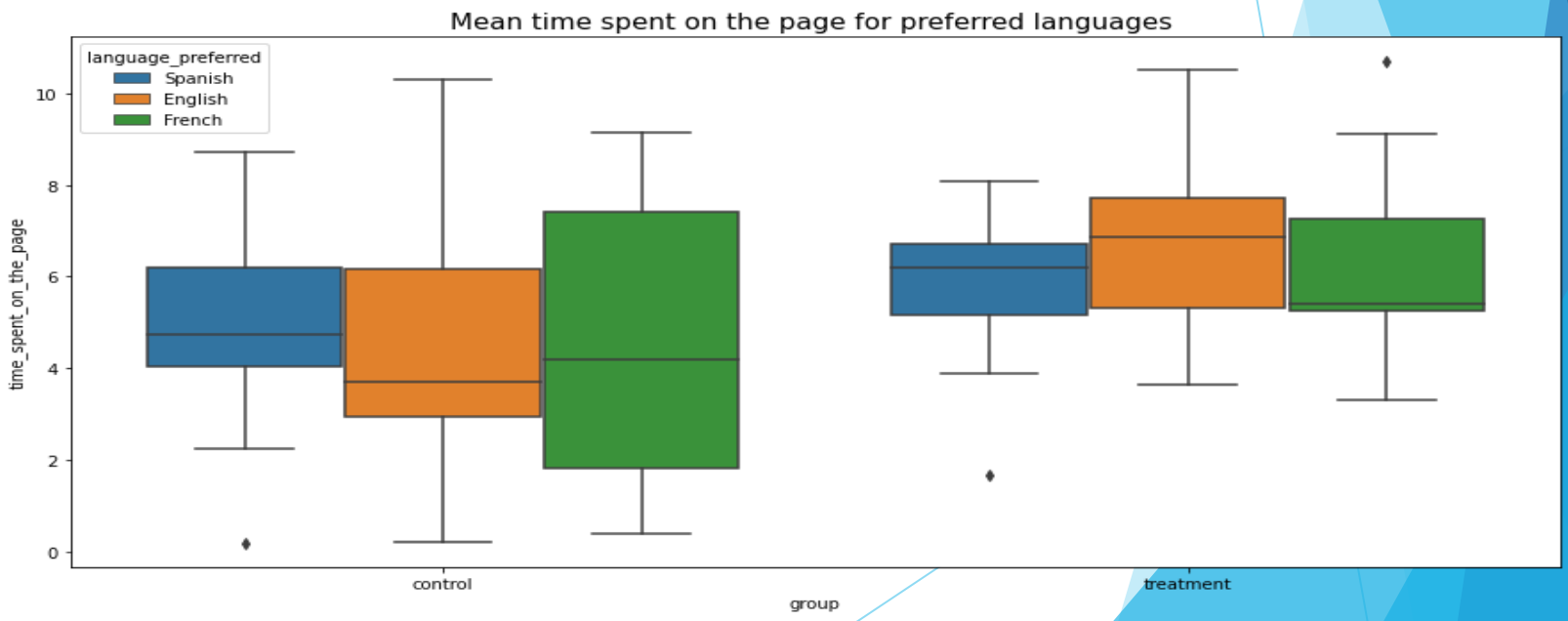
- If Variation A (original landing page, control group) has a better conversion rate with our sample of visitors ,then no further actions need to be taken.

- If Variation B has a better Conversion rate ,then we need to determine whether the improvement was statistically large enough for us to conclude that the change would be reflected in the larger population and thus change to page B.

Exploratory Data Analysis Univariate and Bivariate Analysis

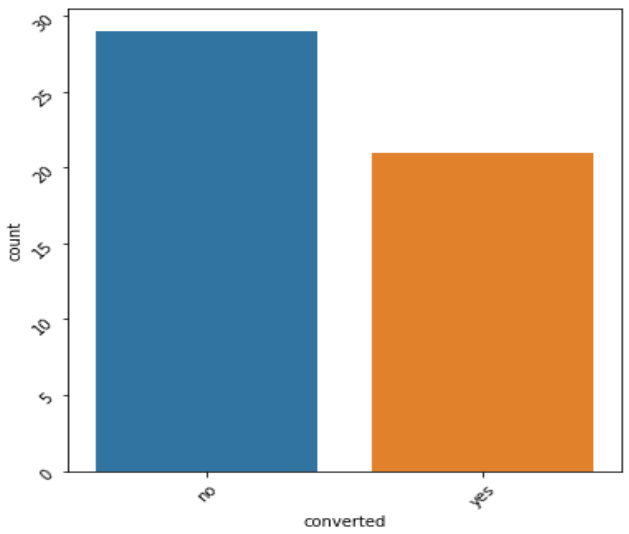
- The given data set has two groups, control group and the treatment group, both with equal number of values.
- The chart shows that the mean time spent on the new landing page(treatment group) is greater than the old landing page(control group).

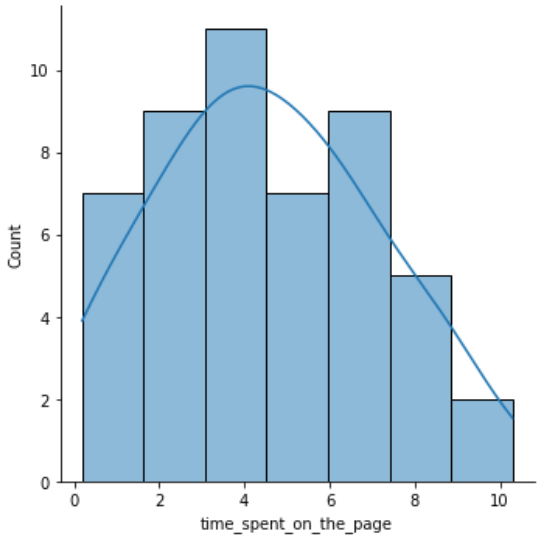




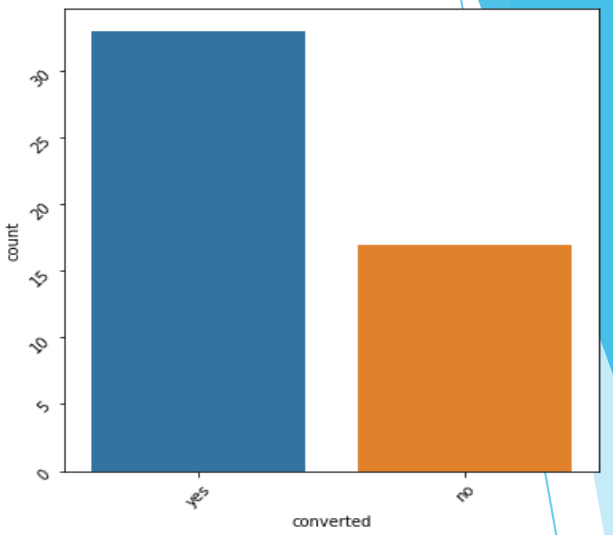
Conversion Rate (2_Proportion Z test)

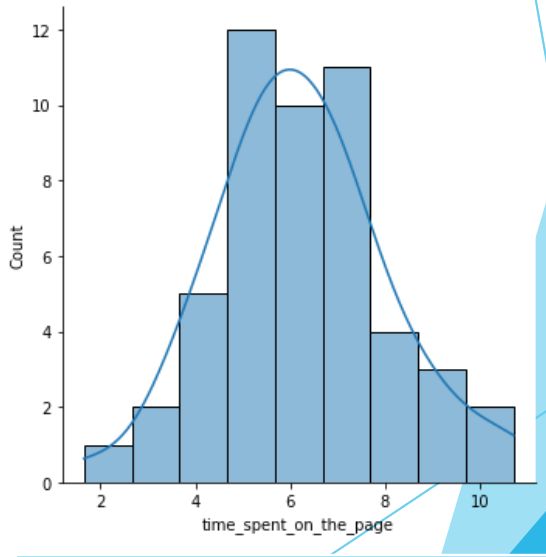
- Null Hypothesis: The conversion rate is same for both the pages.
- Alternate Hypothesis: The proportions are not same.
- The P value is less than the level of significance (0.05),hence we reject the null hypothesis.
- The conversion rate is high in the case of treatment group(new landing page).
- The treatment group distribution is normal.





CONTROL GROUP

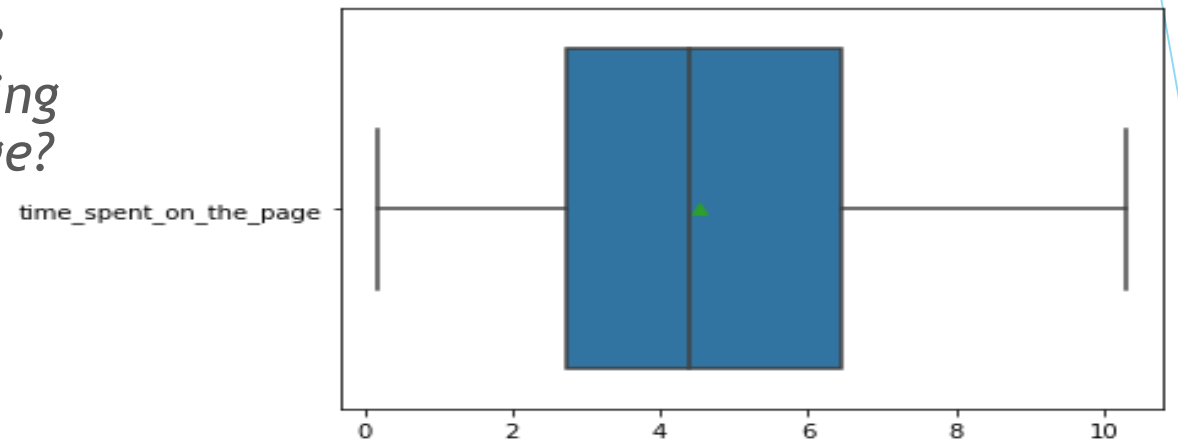




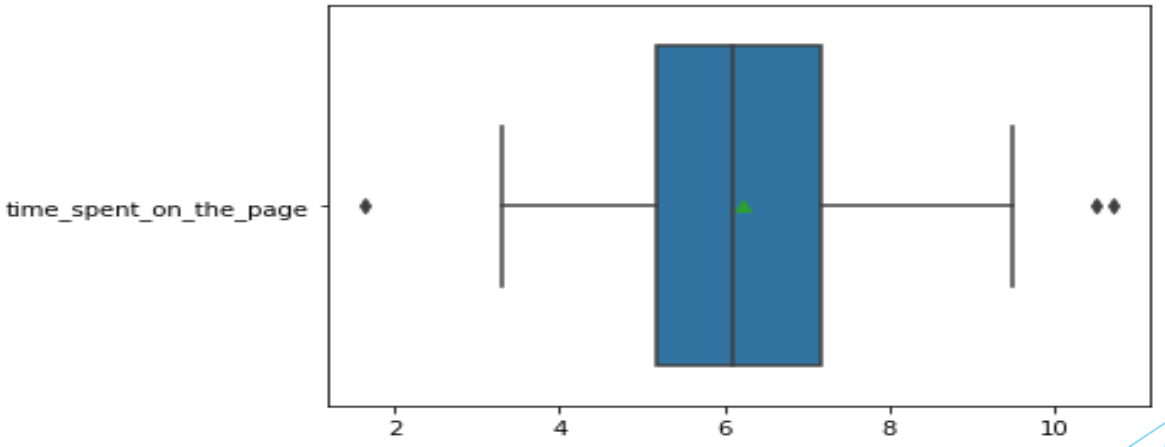
TREATMENT GROUP

2-Sample T test, to Compare time spent on the new landing page and the old landing page?

- Null Hypothesis: The mean time spent on both the pages is same.
- Alternate Hypothesis: The mean time spent on the new landing page is greater than the old page.
- The p-value (0.0001) is less than the level of significance(0.05), we can reject the null hypothesis. Hence, we do have enough evidence to support the claim that the mean time spent on the new landing page is more than that of the old landing page.



Control group



Treatment group

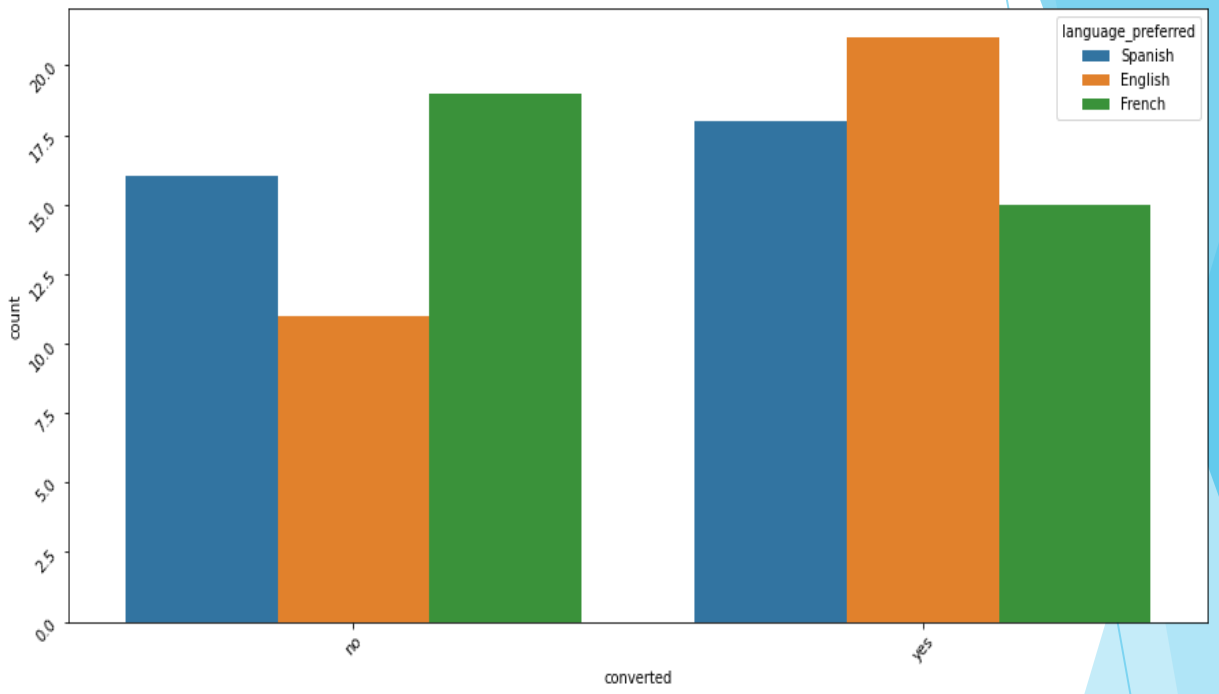
Chi-Square Test for Independence(converted status depends on the preferred language or not)

Null Hypothesis: Converted status is independent on the preferred language.

Alternate hypothesis: Converted status is dependent on the preferred language.

The P-Value 0.212 is greater than the level of significance(0.05), we fail to reject the null hypothesis.

we do not have enough evidence to say that converted status depends on the language preferred.



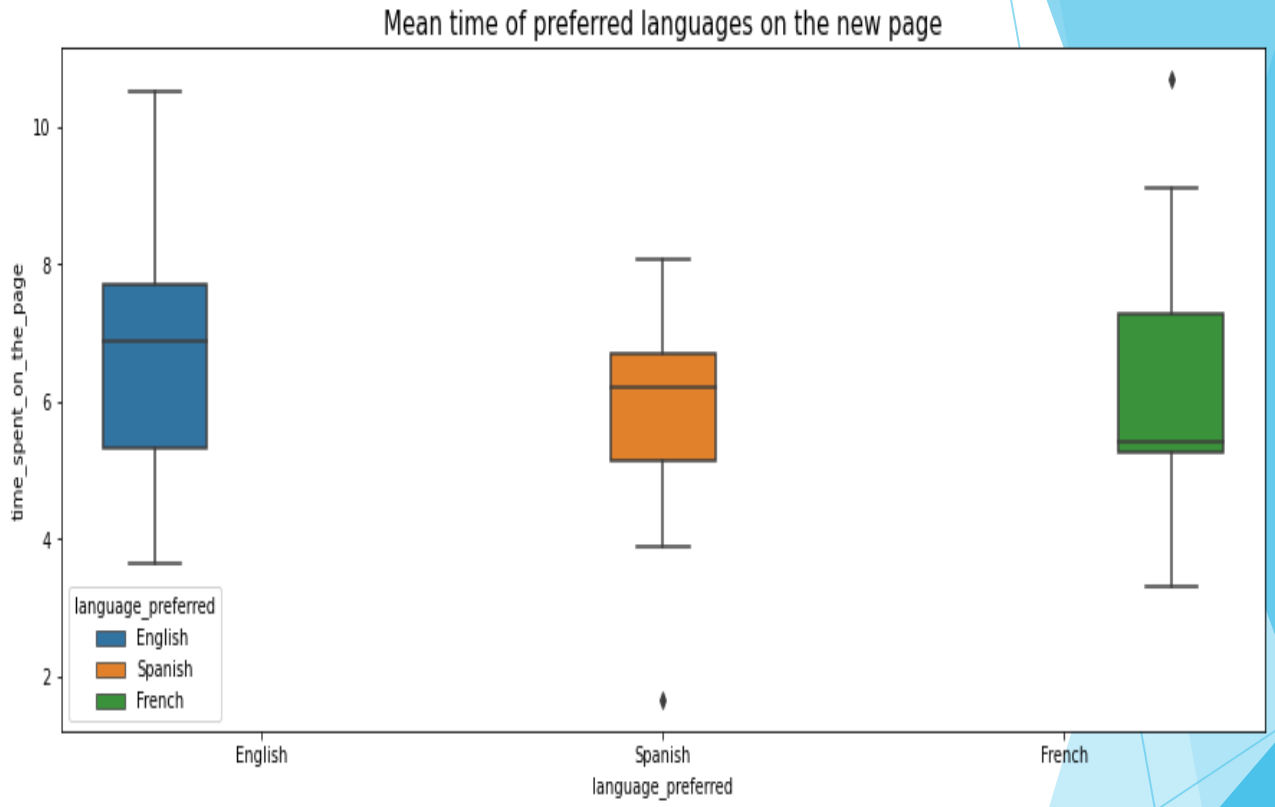
ONE-WAY ANOVA TEST
Analysis of the mean time spent on the new page for different language users.

Null Hypothesis: The mean time spent on the new landing page is same for all the preferred languages.

Alternative Hypothesis: at least one of the mean time is different.

The p-value is greater than the significance level, we can not reject the null hypothesis. Hence, we do not have enough evidence to conclude that at least one mean time spent on the page by one of the 3 languages is different from the rest at 5% significance level.

hence, we can say the mean time spent on the new page for different language users is almost the same.



Conclusions Based on the Statistical Analysis and Hypothesis testing

- The goal of the online E-news Express company is to expand their business by acquiring new customers. The design team of the company has designed a new landing page. The task of the Data Scientist is to do statistical analysis based on the sample data set provided , to know the performance of the new landing page .The aim is to give insights on which page performs better.

- The data set is divided into Variation A (control group, old landing page) and Variation B (treatment group, new landing page). Various hypothesis testing are done, based on which the following conclusions can be drawn:

- 1.The Average or mean time spent on the New landing page is greater than the old landing page.

- 2.The Conversion rate of the New landing page is more than the old landing page.

- 3.The converted Status is independent of the Language preferred.

- 4.The average or the mean time spent on the new landing page for different languages (Spanish, French, English) is almost the same .

Recommendation to Business

New Landing page is more effective than the old Landing page based on all the Statistical Analysis done .

The New Variation created by the design team of the E_NEWS EXPRESS can definitely help in expanding their Business by attracting more subscribers.

Thank you