Business Presentation E-News Express

Contents

- Background on E-news Express Project
- Data and extract insights using Exploratory Data Analysis
- Question and Answers on the following:

  - 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

  Conclusion

Background on E-news Express Project

- E-news Express is an online news portal 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 analyze these interests and wants to determine whether a new feature will be effective or not.

- Companies often analyze users' responses to two variants of a product to decide which of the two variants is more effective.

- This experimental technique is known as a/b testing that is used to determine whether a new feature attracts users based on a chosen metric.

- The company wants to decide whether the new landing page is more effective to gather new subscribers.

Data Overview

- Brief description of data provided

  - The data provided consists of 6 columns with 100 rows of data with data corresponding to actions taken by visitors based on their interests.
  - There are a total of 100 non-null observations in each column.
  - There is one integer type, one floating type and 4 general object type.

- Brief description of significant manipulations made to raw data

  - Data type was converted from dtype object to dtype category.

Exploratory Data Analysis (EDA)
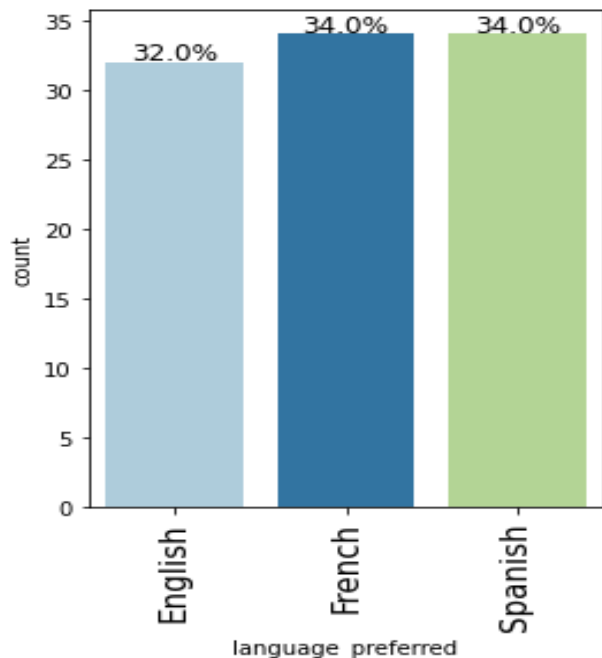
- The two languages most preferred by users are Spanish and French.
- The user count for both French and Spanish are the same with English being less than 6% drop from their number.

Exploratory Data Analysis (EDA)
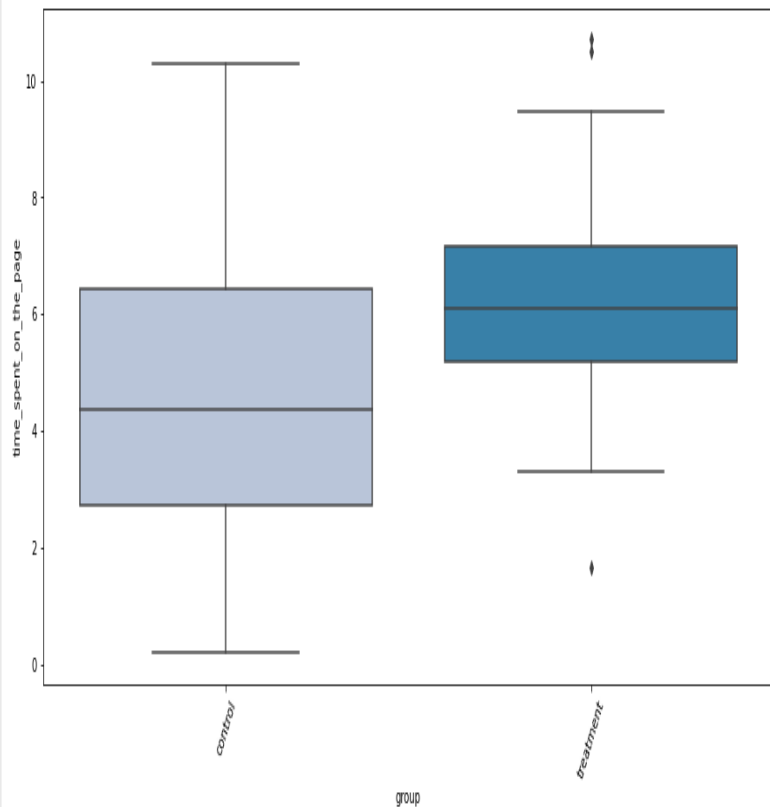
- From the graph it appears that the control group spends more time on the page, with wider quartile ranges
- Though the median appears lower than that of the treatment group there is a larger interquartile range
- With the numbers for the 25% of scores falling above the upper quartile being much larger than that of the treatment group.

Exploratory Data Analysis (EDA)

Effect on Converstion for Landing Page
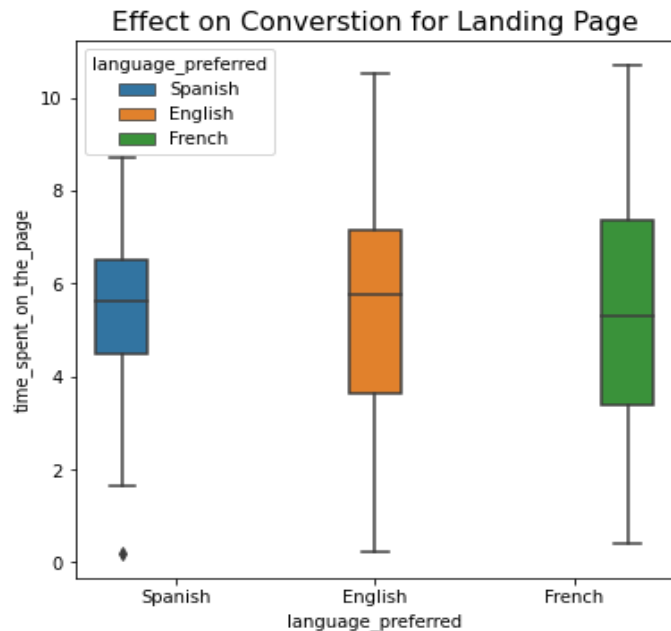
- The English and French language appear to have a higher level of conversion that Spanish.
- The Interquartile range for both English and French appears larger that of Spanish.
- There are significant outliers/whiskers in both English and French Language with French appearing to have the most outliers.

Question and statistical answers on the Data

- Do users spend more time on the new landing page than on the old landing page?
- Hypothesis test:
  - Null Hypothesis: Uold = Unew (Time spent by both old and new is the same)
  - Alternate Hypotheis: Uold<Unew (Time spent by old users is less than that of new users)
- Assumptions:
  - Samples are drawn from a normal distribution - Since the sample size is 100(which is > 30), Central Limit Theorem states that the distribution of sample means will be normal
  - Observations are from a simple random sample
- Conclusion
  - Based on the data provided the sum of time spent on the new landing page and also the average time spent on the new landing page is greater than that of the old landing page (new: sum 311.16 and mean time of 6.2232 compared to old page: sum of 226.62 and average of 4.532)
  - Users appear to spend more time on the new landing page than the old landing page

Question and statistical answers on the Data

- 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?
- Performing a two proportion Z test.

Assumptions:
   - Landing page and converted are both categorical variables
   - Random sampling from the population
   - A user can either get converted or not converted (binomally distributed population)
   - The sample was divided into two equal parts

- Based on the z test for proportions:
  stat, pval = proportions_ztest([old_conversion, new_conversion], [old, new], alternative ='smaller')
  the p-value is 0.0080263082040 56278

- As the p-value is less than the significance level 0.05, we can reject the null hypothesis

Question and statistical answers on the Data

- Does the converted status depend on the preferred language?
  - There are three languages in the data provided French, English and Spanish
  - Using the cross tab function we have the following table depicting convertions:

converted	no	yes	All
language_preferred			
English	11	21	32
French	19	15	34
Spanish	16	18	34
All	46	54	100

  - Use the chi_2 contingency to find the p-value
  - use chi2_contingency() to find the p-value
    chi2, pval, dof, exp_freq = chi2_contingency(data)

Question and statistical answers on the Data

- Does the converted status depend on the preferred language?
- Assumptions
  - The two variables are categorical variables
  - The level of significance of 0.05

- Use chi2_contingency() to find the p-value

  chi2, pval, dof, exp_freq = chi2_contingency(data)
- The p-value is 0.212988

- Given that the p-value 0.212988 is greater than the level of significance we fail to reject the null hypothesis that converted status does depend on the preferred language.
- Insight

  - The converted status depends on the preferred language

Question and statistical answers on the Data

- Is the mean time spent on the new page same for the different language users?
- Hypothesis Test
  - Ho: All means are equal Ho: u1, u2, u3
  - Ha: At least one conversion mean is different from the rest
- Now the normality and equality of variance assumptions need to be checked
- For testing of normality, Shapiro Wilks test is applied to the response variable
- For equality of variance, Levenes test is applied to the response variable
- Shapiro Test
  - w, p_value = stats.shapiro(df['time_spent_on_the_page'])
  - print('The p-value is ', p_value)
  - The p-value is 0.564
  Inference:
  - The p-value is large we fail to reject the null hypothesis that the response follows the normal distributions

Question and statistical answers on the Data

- Is the mean time spent on the new page same for the different language users?
- Levenes Test
  - Ho: All the population means are equal
  - Ha: At least one mean is different from the rest
- from scipy.stats import levene
- statistic, p_value = levene(df['time_spent_on_the_page'][df['language_preferred']=="English"],
- df['time_spent_on_the_page'][df['language_preferred']=="Spanish"],
- df['time_spent_on_the_page'][df['language_preferred']=="French"])
- The p value is 0.0651508684

Inference:

- Since the p-value is larger than the significance leve of 0.05, we fail to rejct the null hypothesis of homegenity of variances

Business Insights and Recommendations

- Actionable insights based on the results of the analysis
  - The movement from the old landing page to the new landing page has affected users in terms of conversion and time spent on the page
  - There are noticeable differences between language preferred and conversion rates and language preferred and time spent on the page
  - Conversion appears to be dependent on the language preferred
  - The user id are unique and from the data appear not repeatable thus isolation of which user id spent the most time on the page was not feasible

- Recommendations to help the organization improve its business outcomes

  - Introduction of more languages to cover more major languages spoken could increase the additional time spent on the page