E-news Express Project Business Presentation

Contents

The presentation consists of three Four Sections:

1. Background and Business Problem Overview
2. Data Overview
3. Exploratory Data Analysis
4. Statistical analysis of business data
5. Insights and Recommendations

Background of the Business and Problem Overview

Background

E-news Express an online news portal aims to expand its business by acquiring new subscribers.

The company plans to analyze the customers interests and wants to determine if new features will be effective or not.

The design team of the company has created a new landing page.

Problem Overview

We have been assigned the task to decide whether the new landing page is more effective to gather new subscribers.
100 users has been randomly selected and divided into 2 groups.
The old landing page is served to the first group (control group) and the new landing page is served to the second group (treatment group).

We need to statistical analysis on business data and generate a set of insights and provide recommendations that will help in company to take right decisions.

Data Overview

Data	Description
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.

Key Questions to answer

- 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?

Exploratory Data Analysis

	group	landing_page	converted	language_preferred
count	100	100	100	100
unique	2	2	2	3
top	control	new	yes	French
freq	50	50	54	34

Exploratory Data Analysis

Observations

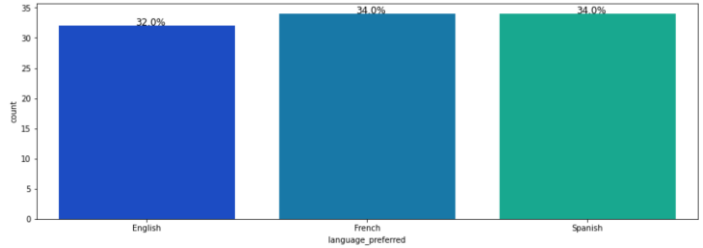
- All 6 column have 100 observations indicating that there are no missing values in it.
- The E-news Express dataset has 100 rows and 6 columns.
- E-news Express dataset has group, landing_page, converted and language_preferred as qualitative or categorical data.
- E-news Express dataset has time_spent_on_the_page as quantitative data.

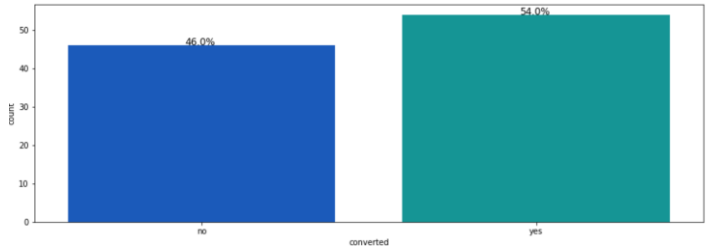
Exploratory Data Analysis

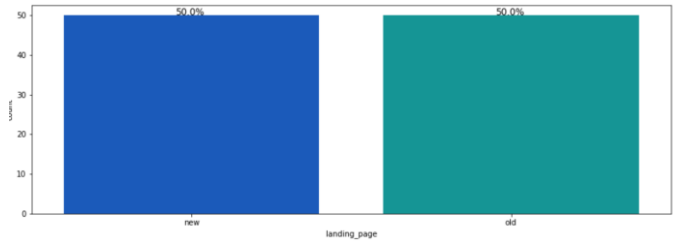
Univariate Analysis of E-news Express Data¶

Observations

- 34% Users preferred language for the page are Spanish and French.
- 32% Users preferred language for the page is English.
- 54% Users subscribed to the news portal and 46% users did not convert to the subscription.
- The population which use the old and new page are equally distributed i.e. 50%



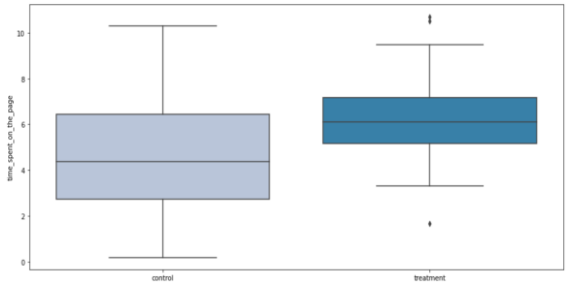


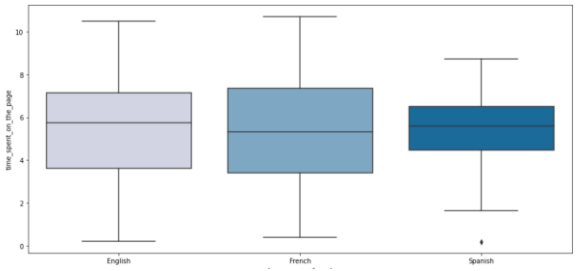


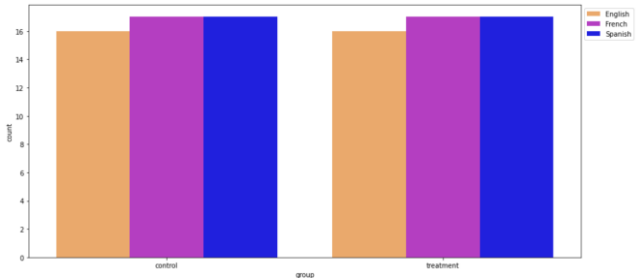
Exploratory Data Analysis

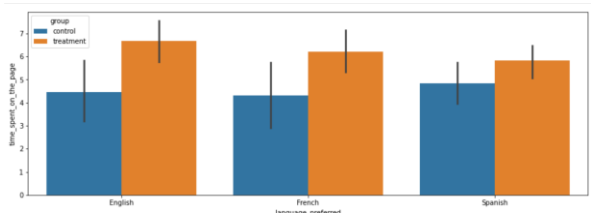
Bi Variate Analysis of E-news Express Data Observations

- The treatment group which is been served with new landing page spent more time on the page.
- The outliers exists with-in the treatment group.
- In all the language catagory English, Spanish and French users spent more time on the new landing page.
- Spanish and French users count are more than English users in both the catagory i.e. control v/s treatment.









Statistical analysis of E-News Express data

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

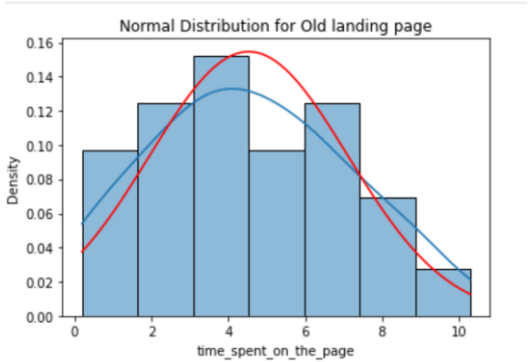
Z-test assumptions which were satisfied

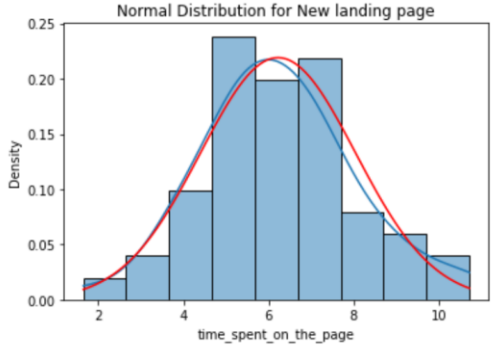
- Continuous data - The ratings are measured on a continuous scale.
- Normally distributed populations or Sample sizes > 30 - Since the sample sizes are greater than 30, Central Limit Theorem states that the distribution of sample means will be normal.
- Independent populations - As we are taking samples for two different landing pages , the two samples are from two independent populations.
- Known population standard deviation $\sigma 1$ and $\sigma 2$ - Yes, we know the population standard deviations of both the populations.
- Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample.

Statistical analysis of E-News Express data

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

Normal distribution





Conclusions

As the p-value is much less than the level of significance 0.05, we reject the null hypothesis. Thus, we have enough statistical evidence to say that users spent less time on old landing page than new landing page.

Statistical analysis of E-News Express 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?

Two Proportion Z-test assumptions which were satisfied

Binomially distributed population - Yes, a user is either converted or not-converted.

Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample.

Can the binomial distribution have approximated to normal distribution - Yes.

Conclusion

As the p-value is Less than the significance level 0.05, we can reject the null hypothesis. Thus, We can reach to the conclusion that conversion rate for the new page is greater than the conversion rate for the old page.

Statistical analysis of E-News Express data

Does the converted status depend on the preferred language?

Chi-Square Test for independence assumptions

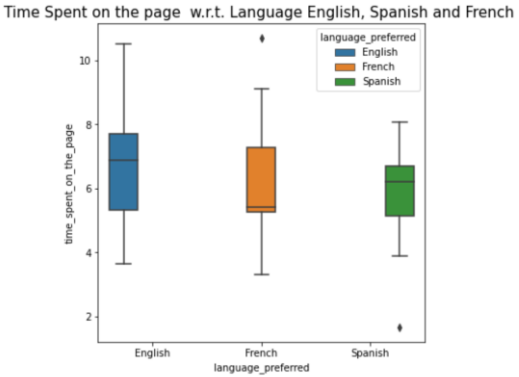
- Categorical variables – Yes
- Expected value of the number of sample observations in each level of the variable is at least 5 - Yes, the number of observations in each level is greater than 5.
- Random sampling from the population - Yes, we are informed that the collected sample is a simple random sample.

Conclusion

As the p-value is higher than the significance level, we are failed to reject the null hypothesis. Hence, we do not have enough statistical significance to conclude that conversion status is dependent on preferred language

Statistical analysis of E-News Express data

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



One Way ANOVA Test

- For testing of normality, Shapiro-Wilk's test is applied to the response variable.
- For equality of variance, Levene test is applied to the response variable.

Assumptions which were satisfied.

- The populations are normally distributed - Yes, the normality assumption is verified using the Shapiro-Wilk's test.
- Samples are independent simple random samples - Yes, we are informed that the collected sample is a simple random sample.
- Population variances are equal - Yes, the homogeneity of variance assumption is verified using the Levene's test.

Conclusion

As the p-value is much higher than the significance level, we failed to reject the null hypothesis.
Hence, we do have enough statistical significance to conclude that mean time spent on the page is not much different for the different language users.

Recommendation

Users are spending less time on old landing page and with the statistical analysis it is concluded that new landing page is more effective in getting more traction among Users.

Users who are converted to the new landing page are independent of the language and more and more users want to move to new landing page.

E-News Express will get more subscription from the Users from all languages if it will focus on moving all its users babe to new landing page.

greatlearning
Power Ahead

Happy Learning !