Statistical Analysis for E-News Express

E-News Express and Business Statistics

Date: 10/07/2022

Mona Desai

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Hypotheses Tested and Results
- Appendix

Executive Summary

- **Conclusions, actionable insights:**
  - After analyzing the given sample data, I observe
    - ~45% more time spent on new landing page
    - ~20% more conversion on new landing page
    - Different language users spend different time
    - Language is independent from the conversion rate
    - Enough statistical evidence (p-value = 0.00802630 at 5% significance level) to say that the conversion of user from the new page is more than the old page
    - Enough statistical evidence (p-value = 0.000139238 at 5% significance level) to say that the mean time spent on the new landing page is more than the mean time spending on the old page.
  - We require more samples
    - To analyze language vs. conversion rate
    - To analyze language vs. old and new page
- **Recommendations:**
  - Migrate old page to new page
  - Make people spend more time on the page
  - Make more language specific content on the new page to have people spending more time

Business Problem Overview and Solution Approach

- **Problem**
  - Determine the effectiveness of the new landing page in getting new subscription for the E-News Express
- **Solution approach / methodology**
  - Perform EDA to analyze and find details from given dataset and treat if any errors in data.
  - Perform the Statistical Test
    - for an unbiased findings.
    - To find any patterns in data
    - To find any correlations, variations, and tendency among data elements

EDA Results, Data background and Content

- **Key results from EDA and Statistical Analysis**
  - Total 100 randomly selected number of people
  - 6 columns with no missing values in the dataset
  - 4 categorical and 2 numerical data type
  - Dataset has 2 groups and 3 languages
  - User spends ~5.4 average minutes on the given page
  - User spend ~4.2 mins average on old page and ~6 mins on new
  - Two equally divided groups (50 numbers) for old and new page, “control” and “treatment”
  - User who spends more time on new landing page, converts into a subscriber
  - Different language user spend different amount of time on the page
  - French language spends the longest and Spanish spends the least

Link to Appendix slide on data background check

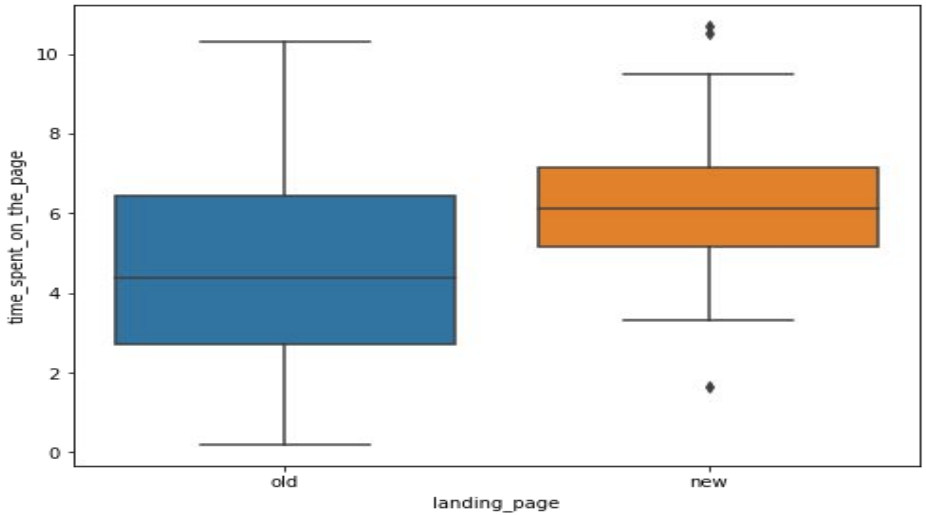
EDA Results, Data background and Content continued

- **Key results from EDA and Statistical Analysis**
  - Not enough statistical evidence (p-value = 0.212988874 at 5% significance level) to say that the conversion of user are not depend upon the different languages
  - Not enough statistical evidence (p-value = 0.46711357 at 5% significance level) to say that the mean time spent by the different language users are different

Hypotheses Tested and Results

- **Do the user spends more time on the new landing page than the existing landing page?**

Visual Analysis



User spends more time on average on New landing page vs. old

Hypotheses Tested and Results continued..

- One tailed, Two sample t-test (population standard deviation is unknown)
- The level of significance is 0.05
- Hypothesis tested
    - null and alternative Hypothesis for the claim
        - $H_0: \mu_1 = \mu_2$

          $H_a: \mu_1 > \mu_2$

          μ_1 = mean time spent on new page

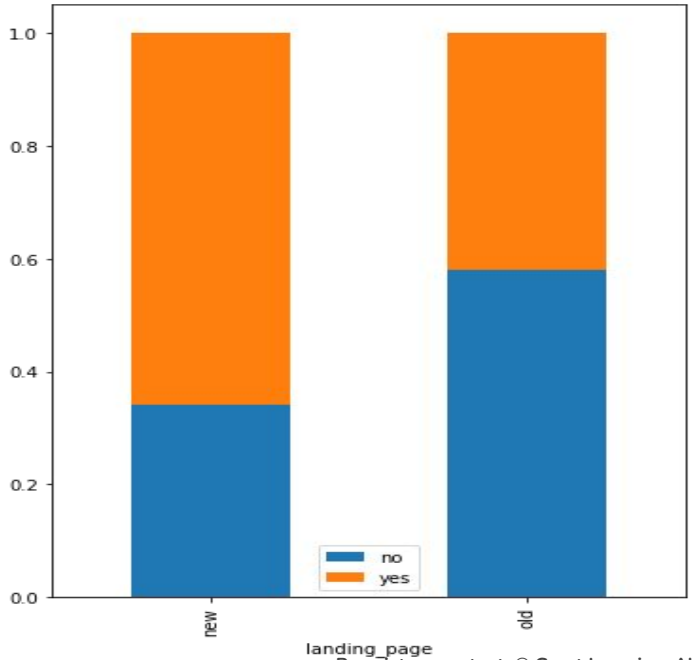
          μ_2 = mean time spent on old page
- Test result and inference.
    - p-value = 0.000139238122

      Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that the mean time spent on the new landing page is more than the mean time spending on the old page.

Hypotheses Tested and Results continued..

- **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?**

Visual Analysis



User visits the new landing page seems to get converted into more subscriptions.

Hypotheses Tested and Results continued..

- one tailed,z-test concerning two population proportion
- The level of significance is 0.05
- Hypothesis tested
  - null and alternative Hypothesis for the claim

    H_0: p1=p2

    Ha: p1>p2

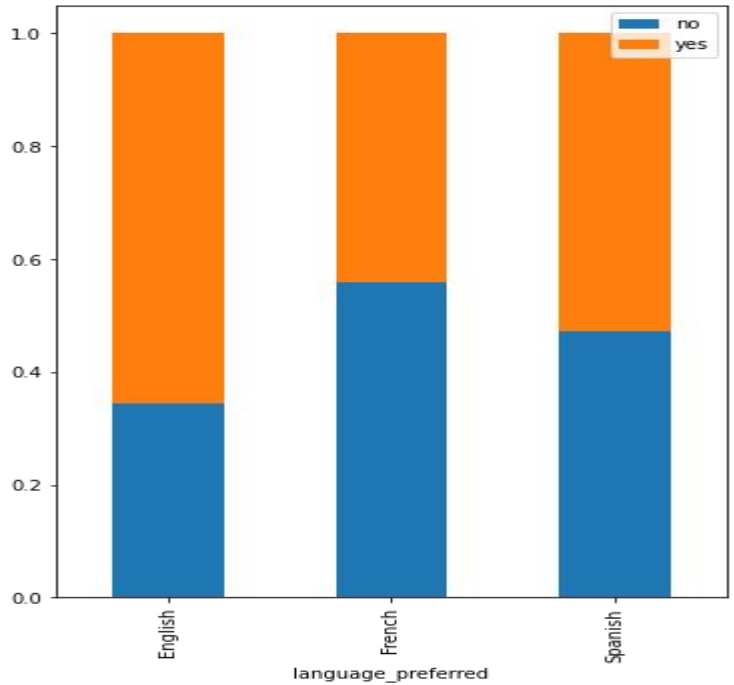
    p1 = proportion of users visiting new landing page

    p2 = proportion of users visiting old landing page
- Test result and inference.
  - p-value = 0.008026308204
  - Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that the conversion of user from the new page is more than the old page

Hypotheses Tested and Results continued..

- **Does the converted status depend on the preferred language?**

  Visual Analysis



The conversion seems to be different for different languages

Hypotheses Tested and Results continued..

- Test of Independence concerning two categorical variable -"language_preferred" and "converted"
- Chi-square test of independence
- The level of significance is 0.05
- Test of Independence
    - The contingency table

Converted	no	yes
Language_preferred		
English	11	21
French	19	15
Spanish	16	18

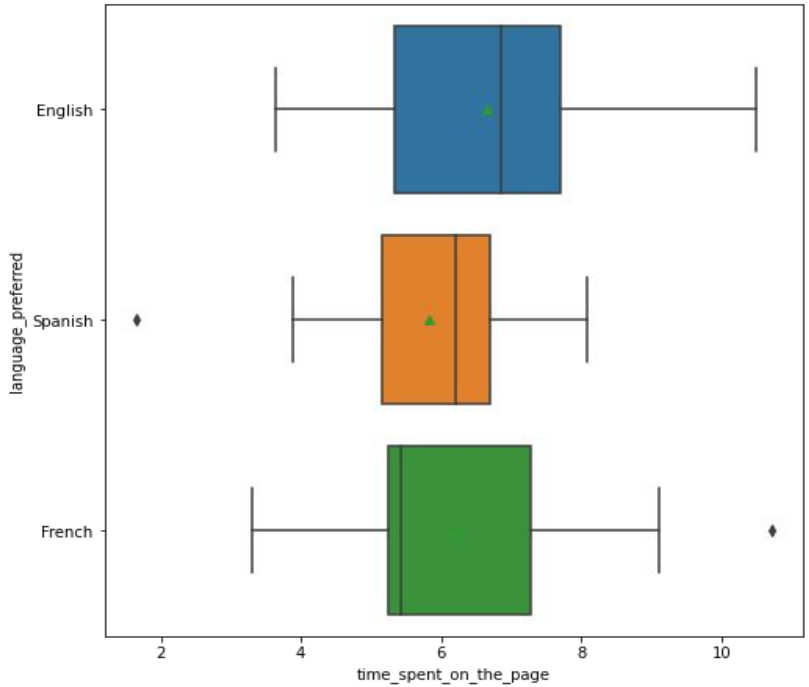
Hypotheses Tested and Results continued..

- Hypothesis tested
  - null and alternative Hypothesis for the claim
    - H_0: language and conversion are independent
    - Ha:language and conversion are dependent

- Test result and inference.
  - p-value= 0.212988874
  - Since the p-value is greater than the 5% significance level, we fail to reject the null hypothesis. Hence, we have not enough statistical evidence to say that the conversion of user are not depend upon the different languages

Hypotheses Tested and Results continued..

- **Is the time spent on the new page same for the different language users?**
  - Visual Analysis



- Average time on the new landing page by different language users

  English ~ 6.66 mins
  French ~ 6.19 mins
  Spanish ~ 5.84 mins

Hypotheses Tested and Results continued..

- ANOVA , Levene test concerning three population means
- The level of significance is 0.05
- Hypothesis tested
  - null and alternative Hypothesis for the claim
    - H_0: μ1=μ2=μ3
    - Ha: μ1≠ μ2 or μ2 ≠μ3 or μ1 ≠μ3
- Test result and inference.
  - P-value = 0.46711357
  - Since the p-value is greater than the 5% significance level, we fail to reject the null hypothesis. Hence, we have not enough statistical evidence to say that the mean time spent by the different language users are different