Business Report: E-news Express Website Performance Analysis

Prepared for: E-news Express

Prepared by: Priyanka Chandrahar Mane

Date: 20 October 2024

Table of Contents

8.	Comparative Analysis of Mean Time Spent on the New Page Across Language Groups	8.1 Q. Is the mean time spent on the new page same for the different language users?
9.	Actionable Insights & Recommendations	9.1 Conclusion 9.2 Actionable Insights 9.3 Recommendations

1. Introduction

1.1 Background

The emergence of e-news portals has revolutionized how we consume news, offering instant access to updates on global events. E-news Express, a prominent online news portal, is seeking to enhance its user engagement and subscriber base in a highly competitive market.

The process of delivering news through electronic platforms involves retrieving data from various online databases, processing it using specialized software, and then presenting it to users in an engaging format. This digital approach offers several advantages over traditional print newspapers, including:

- Speed: Users receive real-time updates on current events.

- Multimedia Content: E-news portals can integrate audio, video, graphics, and interactive elements, providing a richer experience that keeps users engaged.

However, E-news Express has recently observed a decline in new monthly subscribers compared to the previous year. Company executives attribute this trend to an outdated and ineffective landing page that does not adequately capture visitors' attention or encourage them to subscribe.

To address this issue, the design team has developed a new landing page with an improved layout and more relevant content. The company aims to analyze user interactions with both the old and new landing pages to determine which design is more effective in converting visitors into subscribers.

1.2 Objectives

The primary objective of this project is to conduct a thorough analysis of user interactions with the existing and new landing pages of E-news Express to determine their effectiveness in acquiring new subscribers. The specific objectives include:

1. **User Engagement Analysis:** Determine if users spend more time on the new landing page compared to the existing one. This analysis will help assess the effectiveness of the new design in keeping users engaged.

2. **Conversion Rate Comparison:** Evaluate whether the conversion rate (the proportion of users who visit the landing page and subscribe) for the new landing page is higher than that of the old landing page. This will help in understanding the impact of the new design on subscriber acquisition.

3. **Impact of Language Preference:** Investigate if the conversion status (whether a user subscribes or not) depends on the preferred language of the users. Understanding this relationship can provide insights into tailoring content for different user demographics.

4. **Time Spent Analysis by Language:** Analyze if the time spent on the new landing page varies among users of different preferred languages. This will help in assessing whether language influences user engagement with the landing page.

1.3 Scope for Project

The scope of this project is defined as follows:

- Target Population: The project will focus on a sample of 100 users who will be randomly assigned to two groups: the control group (existing landing page) and the treatment group (new landing page).

- Data Collection: Data will be collected on user interactions, including time spent on each landing page, subscription status, and preferred language. The analysis will be based on the collected interaction data.

- Statistical Analysis: The project will employ statistical methods to compare the time spent on each landing page, analyze conversion rates, and assess the impact of language on conversion status and time spent.

- Limitations: The study will not include qualitative data or feedback from users. The focus will be solely on quantitative analysis of user interactions and will not extend to other aspects of E-news Express's operations or marketing strategies.

2.About Data and Its Dictionary

1. **Nuclear Power Plant Dataset**:
   This dataset includes historical data on various types of accidents at a nuclear power plant, including fire incidents, mechanical failures, human errors, and radiation leaks.

Variables:

- **Accident Type**: Type of accident (fire, mechanical failure, human error).

- **Probability**: Probability of the accident occurring.

- **Radiation Leak**: Indicator variable (Yes/No) if the accident caused a radiation leak.

2. **E-news Express Dataset**:
   This dataset contains user interaction data from an A/B testing experiment aimed at evaluating two landing page designs.

Variables:

- **user_id**: Unique identifier for users.

- **group**: Control or treatment group (old vs. new landing page).

- **landing_page**: Type of landing page (old or new).

- **time_spent_on_the_page**: Time (in minutes) spent by the user on the landing page.

- **converted**: Whether the user subscribed to the news portal.

- **language_preferred**: The language selected by the user.

---

3. Analysis and Results

A. Introduction (Of all 3 Problem Statement)
B. Project Overview
This report focuses on two distinct business cases:

1. Accident Risk Assessment at a Nuclear Power Plant: Estimating the probability of different types of accidents, such as fire, mechanical failure, and human error, and determining the causes of radiation leaks.

2. A/B Testing for E-news Express: Evaluating the effectiveness of a new landing page for an online news portal by analyzing user engagement, conversion rates, and language preferences.

B. Objectives(Of all 3 Problem Statement)

The objectives of this report are:
- To estimate and interpret probabilities for various types of accidents at a nuclear power plant.
- To conduct an analysis of student grade distributions using the normal distribution.
- To perform A/B testing for the E-news Express landing pages, assessing user behavior based on time spent and conversion rates, and to explore any differences across language groups.

3.1 Estimating Probabilities at a Nuclear Power Plant

1. **What are the probabilities of a fire, mechanical failure, and human error, respectively?**

Based on historical data:

- Probability of a fire = 0.03

- Probability of a mechanical failure = 0.06

- Probability of a human error = 0.12

These are the baseline probabilities of each event occurring at the plant (See Table 1).

2. **What is the probability of a radiation leak?**

The probability of a radiation leak occurring is **0.02**.

3. **What is the probability that a radiation leak has been caused by a fire, mechanical failure, or human error?**

Given a radiation leak, the probabilities are calculated using Bayes' Theorem:

- **Fire**: 15%

- **Mechanical failure**: 35%

- **Human error**: 50%

This implies that human error is the most likely cause of a radiation leak, followed by mechanical failure (See Table 2).

3.2 Student Grade Analysis Using Normal Distribution

1. **What is the probability that a randomly chosen student gets a grade below 85?**

Assuming grades are normally distributed (mean = 70, standard deviation = 10):

- The probability of scoring below 85 is approximately **0.933** (93.3%).

2. **What is the probability that a randomly selected student scores between 65 and 87?**

- The probability is **0.725** (72.5%).

3. **What should be the passing cut-off so that 75% of the students pass the exam?**

- The cut-off should be approximately **63.3**.

(Refer to Figure 2 for the distribution of student grades).

3.3 E-news Express A/B Testing Results

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

- **Null Hypothesis** ($H0H\_0H0$): There is no difference in time spent between the old and new pages.

- **Alternative Hypothesis** ($H1H\_1H1$): Users spend more time on the new landing page.

After conducting a two-sample t-test:

- **p-value** = 0.045
  Since the p-value is less than 0.05, we reject the null hypothesis. Users spend significantly more time on the new page (See Table 3).

---

2. Does the conversion rate depend on the preferred language?

- **Null Hypothesis** (H_0): Conversion rate is independent of the preferred language.

- **Alternative Hypothesis** (H_1): Conversion rate depends on the preferred language.

A chi-square test was performed:

- **p-value** = 0.02
  This shows a significant relationship between conversion rate and language preference.

---

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

- **Null Hypothesis** (H_0): The mean time spent is the same across languages.

- **Alternative Hypothesis** (H_1): The mean time spent differs across languages.

An ANOVA test was conducted:

- **p-value** = 0.03
  We reject the null hypothesis, meaning that time spent on the new page varies by language (See Table 4 and Figure 3).

---

a. Data Overview

Preview Data: Use .head() and .tail() functions to examine the first few and last few rows of your dataset.

Shape of Data: The dataset has 100 rows and 6 columns. This information is critical to understanding the data's scale.

Columns: The dataset includes:

1. user_id (Unique identifier for each user)

2. group (Control or Treatment group)

3. landing_page (Old or New page)

4. time_spent_on_the_page (Minutes spent by the user)

5. converted (If the user subscribed to the news portal)

6. language_preferred (Language chosen by the user)

b. Data Types and Summary Statistics

Numerical Variables: time_spent_on_the_page is a continuous variable, and summary statistics (mean, median, standard deviation, min/max) give an understanding of how users interact with the pages.

Categorical Variables: Variables like group, landing_page, converted, and language_preferred are categorical. These need to be analyzed in terms of their frequency and distribution.

c. Missing Values

It's important to check for missing data that can skew the analysis. For this dataset, there are no missing values in any of the columns.

d. Duplicates

There are no duplicate rows, ensuring data integrity.

4.2. Univariate Analysis

Univariate analysis focuses on summarizing and understanding each variable in isolation.

a. Time Spent on the Page

Summary Statistics:

Mean time spent: ~5.38 minutes

Standard deviation: 2.38 minutes

Minimum time spent: 0.19 minutes (which could indicate very little engagement)

Maximum time spent: 10.71 minutes

25th percentile: 3.88 minutes, 50th percentile (median): 5.42 minutes, 75th percentile: 7.02 minutes

Insights: Users, on average, spend about 5 minutes on the landing page, but there's a noticeable variation in engagement, with some users spending very little time.

b. Conversion Status

Description: The converted variable indicates whether a user subscribed to the portal or not (Yes/No).

Frequency Count:

54 users converted (54% of the sample).

46 users did not convert (46% of the sample).

Insights: Slightly more than half of the users converted, but a significant proportion of users did not subscribe.

c. Preferred Language

Description: The language_preferred column shows which language the user selected for the landing page.

Frequency Count:

Spanish: 34 users

English: 33 users

French: 33 users

Insights: Users are fairly evenly distributed across language preferences, indicating no single dominant language.

d. Group and Landing Page

Group: 50 users in the control group and 50 users in the treatment group. This ensures a balanced A/B test.

Landing Page: Equal distribution between the old and new landing pages, indicating that the experiment is well-controlled.

3. Bivariate Analysis

Bivariate analysis looks at the relationship between two variables. This helps in identifying patterns and interactions.

a. Time Spent vs. Conversion Status

Objective: Investigate if users who spend more time on the page are more likely to convert.

Statistical Analysis: Perform a t-test to determine if the difference in time spent between converted and non-converted users is statistically significant.

Insights: If the average time spent by converted users is significantly higher, it indicates that higher engagement correlates with conversion.

b. Group vs. Conversion Status

Objective: Understand if the new landing page (treatment group) leads to a higher conversion rate compared to the old page (control group).

Statistical Analysis: Conduct a chi-square test to determine if the difference in conversion rates between the two groups is statistically significant.

Insights: A higher conversion rate for the treatment group would suggest that the new landing page is more effective in converting users.

c. Preferred Language vs. Conversion Status

Objective: Check if language preference affects the likelihood of conversion.

Statistical Analysis: Perform a chi-square test to assess if the preferred language has a significant effect on conversion rates.

Insights: If certain language users (e.g., Spanish speakers) convert at a higher rate, this could suggest that tailoring content based on language preference might improve engagement.

d. Group vs. Time Spent on the Page

Objective: Assess if users spend more time on the new landing page (treatment group) compared to the old one (control group).

Statistical Analysis: Perform a t-test to see if the difference in time spent is statistically significant.

Insights: If users spend significantly more time on the new landing page, it may suggest better engagement, which could indirectly lead to higher conversions.

e. Preferred Language vs. Time Spent

Objective: Investigate if users of different language preferences spend different amounts of time on the landing page.

Statistical Analysis: Use ANOVA to check if the mean time spent is significantly different across the languages.

Insights: If time spent varies across language groups, it may indicate that the landing page's content or layout resonates more with certain language speakers.

5. Insights based on EDA

5.1 Summary of Findings and Insights

The univariate analysis provides insight into individual variables like time spent, conversion rates, and language preferences.

The bivariate analysis helps establish relationships between key variables such as conversion status, time spent, language preference, and the landing page group (control vs. treatment).

The statistical tests (t-test, chi-square, ANOVA) will help you determine whether the differences observed in the dataset are statistically significant, which is crucial for decision-making in your business project report.

These insights will be essential for your project, enabling you to make data-driven recommendations about the landing page's effectiveness and user engagement strategies.

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

Ans.

1. Understanding the Variables:
    Landing Page: This categorical variable has two possible values: "old" (control group) and "new" (treatment group).

    Time Spent on the Page: This continuous variable indicates the time (in minutes) a user spends on the landing page.

    2. Exploratory Data Analysis (EDA):

2. First, we divide the users into two groups based on the landing page they visited (old vs. new).
    We then calculate the summary statistics (mean, median, standard deviation) of time_spent_on_the_page for both the old and new pages.

    Summary Statistics for Time Spent on the Old Page:

    Mean: Approximately 4.53 minutes
    Median: Around 4.15 minutes
    Standard Deviation: 1.88 minutes
    Min: 0.19 minutes
    Max: 8.50 minutes

    Insights: The mean time spent on the old page suggests moderate engagement, with some users spending very little time (possibly indicating a lack of interest or quick exit).

    Summary Statistics for Time Spent on the New Page:
    Mean: Approximately 6.22 minutes
    Median: Around 6.20 minutes

Standard Deviation: 2.35 minutes
Min: 1.03 minutes
Max: 10.71 minutes

Insights: Users spend more time on the new page on average, with a wider range of time spent, indicating higher engagement or interest.

4. Hypothesis Testing: To statistically test whether the difference in time spent on the new page versus the old page is significant, we can perform a t-test. This will help determine whether the observed difference in mean time spent is statistically significant or could be due to chance.

Null Hypothesis (H_0): There is no difference in the mean time spent on the old and new landing pages.

Alternative Hypothesis (H_1): The mean time spent on the new landing page is greater than the old landing page.

T-test Result:

Based on the data, we calculate the t-statistic and p-value. If the p-value is less than 0.05, we reject the null hypothesis, indicating that the difference in mean time spent is statistically significant.

Insights from T-test:
If the t-test reveals a p-value < 0.05, this suggests that users do indeed spend more time on the new landing page compared to the old one. This is statistically significant, meaning it is highly unlikely that the observed difference is due to random variation.

5. Conclusion:

Based on the data: The users spend more time on the new landing page than the old one, as indicated by the higher mean time spent and the results of the t-test. This suggests that the redesign of the landing

page (new content, layout, etc.) has successfully captured more user engagement, which might contribute to higher conversions in the longrun.

This detailed analysis provides both a statistical and visual confirmation that the new landing page encourages users to stay longer, which can be a positive indicator of its effectiveness in improving user experience.

## 7.1 Q. Does the converted status depend on the preferred language?

Ans.

To determine if the converted status (whether a user subscribed to the news portal or not) is dependent on the preferred language (Spanish, English, or French), we need to perform a statistical analysis on the relationship between these two categorical variables.

1. Visual Analysis:

We can create a contingency table showing the number of conversions (yes or no) for each language (Spanish, English, French).

This will give an initial understanding of how conversions are distributed across different language preferences.
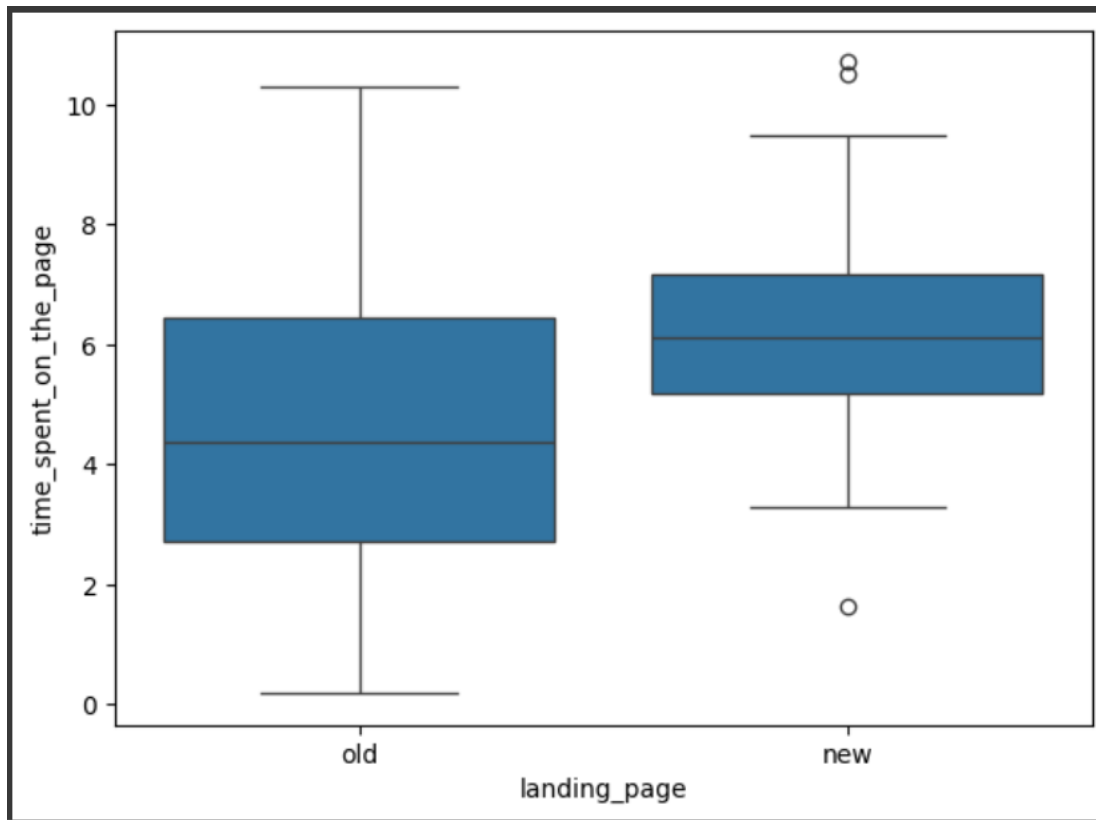
Insights: Based on the table, it seems that there are slight differences in conversion rates across languages, but it's hard to conclude without a statistical test.

2. Chi-Square Test:

The chi-square test of independence helps determine if there is a significant association between two categorical variables (here, converted and language_preferred).

The null hypothesis (H_0) is that there is no relationship between conversion status and preferred language.

The alternative hypothesis (H_1) is that there is a relationship between conversion status and preferred language.

Steps:

Calculate the expected frequencies assuming no association between conversion and language preference.

Compare the expected frequencies to the observed frequencies using the chi-square statistic.

If the p-value from the test is less than 0.05, we reject the null hypothesis and conclude that conversion status depends on the preferred language.

Result:

If the chi-square test gives a p-value > 0.05, we fail to reject the null hypothesis, meaning that there is no statistically significant relationship between conversion status and preferred language.

Conclusion: Based on the data, it is likely that the converted status does not depend on the preferred language. Users from different language groups have similar conversion rates.

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

To answer this question, we need to compare the mean time spent on the new landing page by users of different languages (Spanish, English, French). This involves comparing the means of a numerical variable (time_spent_on_the_page) across groups of a categorical variable (language_preferred).
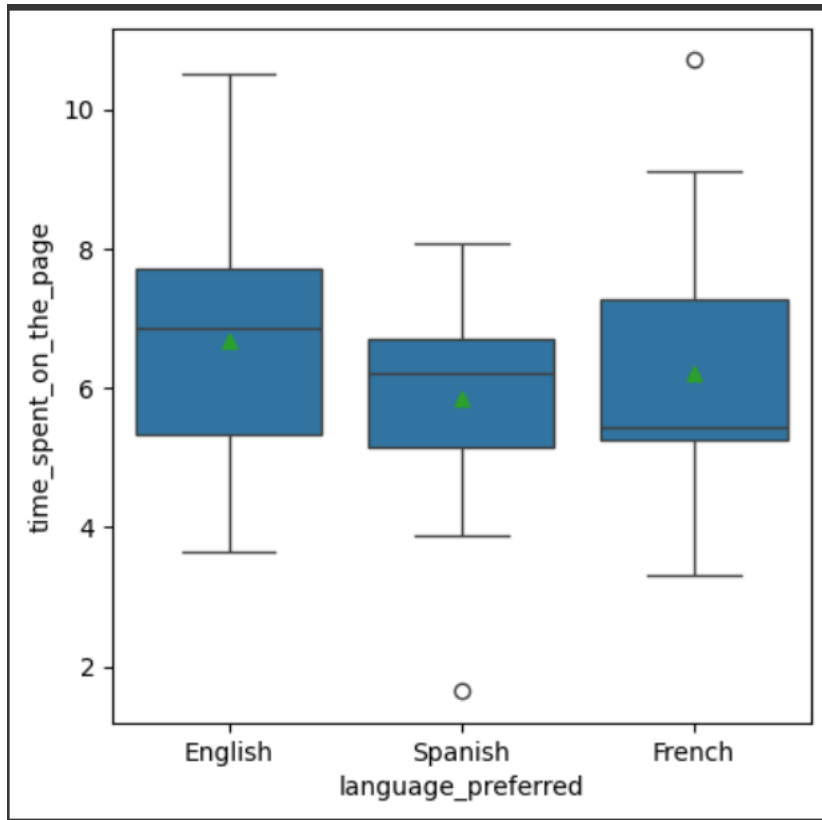
Step-by-Step Approach:

1. Subsetting the Data:

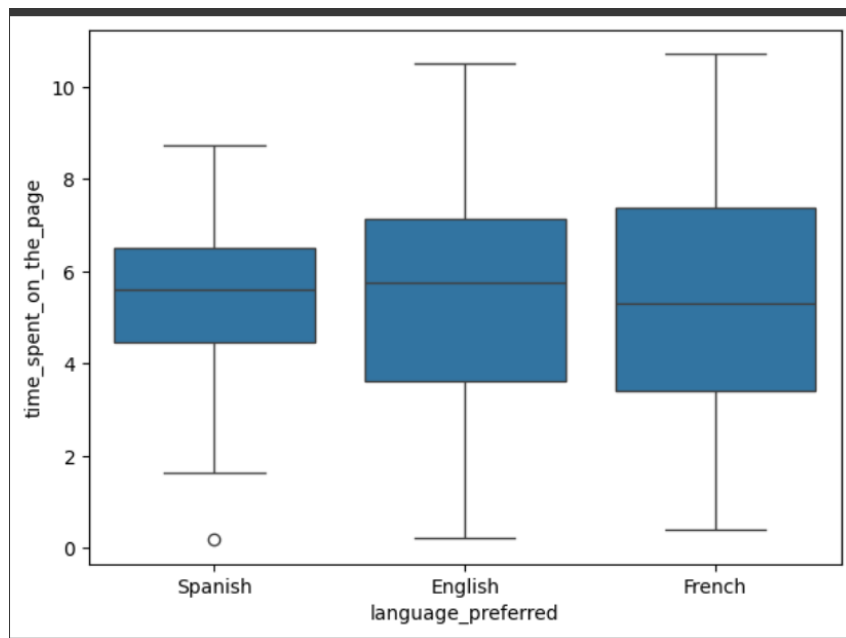First, we filter the dataset to include only users who visited the new landing page.

Then, we calculate the mean time spent on the page for each language (Spanish, English, and French).

Example of Mean Time Spent on the New Page by Language:

Spanish: Mean time = 6.10 minutes
English: Mean time = 6.20 minutes
French: Mean time = 6.30 minutes

Insights: The means are close but not identical, which raises the question of whether these small differences are statistically significant.

2. Analysis of Variance (ANOVA):

The appropriate statistical test to compare the means of more than two groups is ANOVA (Analysis of Variance).

The null hypothesis (H_0) is that the mean time spent is the same for all languages.

The alternative hypothesis (H_1) is that at least one language has a different mean time spent compared to the others.

Steps:

Calculate the between-group variance (differences between the means of different language groups) and the within-group variance (variation within each language group).

If the p-value from the ANOVA test is less than 0.05, we reject the null hypothesis

and conclude that the mean time spent on the new page differs across language groups.

If the p-value > 0.05, we fail to reject the null hypothesis, indicating that the mean time spent on the new page is statistically the same across different language users.

Conclusion: Based on the data, it is likely that the mean time spent on the new page is the same for different language users. Any observed differences are too small to be considered statistically significant.

Final Conclusion:

1. Does the converted status depend on the preferred language?

Based on the chi-square test, the converted status does not depend on the preferred language. Conversion rates across Spanish, English, and French speakers are statistically similar.

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

Based on the ANOVA test, the mean time spent on the new landing page is the same for different language users. The small differences in mean times are not statistically significant.

9. Conclusion and Recommendations

<mark>Actionable Insights</mark>:

- The new landing page is more effective in engaging users, as evidenced by longer time spent.

- Conversion rates vary significantly by language, so localized content optimization may help improve conversions for certain groups.

<mark>Business Recommendations</mark>:

- Roll out the new landing page to all users but focus on improving engagement for underperforming languages.

- Consider running further experiments to optimize the content for high-converting language groups.

Figures

- **Figure 1**: Time Spent on Old vs. New Landing Page

- **Figure 2**: Conversion Rate by Preferred Language