Description

Business Context

The advent of e-news, or electronic news, portals has offered us a great opportunity to quickly get updates on the day-to-day events occurring globally. The information on these portals is retrieved electronically from online databases, processed using a variety of software, and then transmitted to the users. There are multiple advantages of transmitting new electronically, like faster access to the content and the ability to utilize different technologies such as audio, graphics, video, and other interactive elements that are either not being used or aren't common yet in traditional newspapers.

E-news Express, an online news portal, aims to expand its business by acquiring new subscribers. With every visitor to the website taking certain actions based on their interest, the company plans to analyze these actions to understand user interests and determine how to drive better engagement. The executives at E-news Express are of the opinion that there has been a decline in new monthly subscribers compared to the past year because the current webpage is not designed well enough in terms of the outline & recommended content to keep customers engaged long enough to make a decision to subscribe.

[Companies often analyze user responses to two variants of a product to decide which of the two variants is more effective. This experimental technique, known as A/B testing, is used to determine whether a new feature attracts users based on a chosen metric.]

Objective

The design team of the company has researched and created a new landing page that has a new outline & more relevant content shown compared to the old page. In order to test the effectiveness of the new landing page in gathering new subscribers, the Data Science team conducted an experiment by randomly selecting 100 users and dividing them equally into two groups. The existing landing page was served to the first group (control group) and the new landing page to the second group (treatment group). Data regarding the interaction of users in both groups with the two versions of the landing page was collected. Being a data scientist in E-news Express, you have been asked to explore the data and perform a statistical analysis (at a significance level of 5%) to determine the effectiveness of the new landing page in gathering new subscribers for the news portal by answering the following questions:

1. Do the users spend more time on the new landing page than on the existing landing page?
2. 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?
3. Does the converted status depend on the preferred language?

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

Data Dictionary

The data contains information regarding the interaction of users in both groups with the two versions of the landing page.

1. user_id - Unique user ID of the person visiting the website
2. group - Whether the user belongs to the first group (control) or the second group (treatment)
3. landing_page - Whether the landing page is new or old
4. time_spent_on_the_page - Time (in minutes) spent by the user on the landing page
5. converted - Whether the user gets converted to a subscriber of the news portal or not
6. language_preferred - Language chosen by the user to view the landing page

Submission Guidelines

1. There are two ways to work on this project:

i. Full-code way: The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

ii. Low-code way. The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

S u b m i s s i o n t y	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	S u b m i s s i o n F o r m

p e					a t
Full-code	Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch	Perform exploratory data analysis to identify insights and recommendations for the problem	Focus on code writing: 10-20% grading on the quality of the final code submitted	Solution notebook from the full-code template submitted in .html format	.html
Low-code	Learners who aspire to be in managerial roles in the future-focussed on solution review, interpretation, recommendations, and communicating with business		Focus on business presentation: 10-20% grading on the quality of the final business presentation submitted	Business presentation in .pdf format with problem definition, insights, and recommendations	.pdf

Please follow the below steps to complete the assessment. Kindly note that if you submit a presentation, ONLY the presentation will be evaluated. Please make sure that all the sections mentioned in the rubric have been covered in your submission.

i. Full-code version

- Download the full-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Clearly write down insights and recommendations for the business problems in the comments.
- Submit only the solution notebook prepared from the learner notebook [format: .html]

ii. Low-code version

- Download the low-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Prepare a business presentation with insights and recommendations to the business problem.
- Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

- it is submitted post-deadline, or,
- more than 1 file is submitted.

Best Practices for Full-code submissions

- The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission.
- It is important to remove all warnings and errors before submission.
- The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
- Please refer to the FAQ page for common project-related queries.

Best Practices for Low-code submissions

- The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
- The key points in the presentation should be the following:
  - Business Overview of the problem and solution approach
  - Key findings and insights which can drive business decisions
  - Business recommendations
  - Focus on explaining the key takeaways in an easy-to-understand manner.
  - The inclusion of the potential benefits of implementing the solution will give you the edge.
- Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
- The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
- Please refer to the FAQ page for common project-related queries.

Happy Learning!

Scoring guide (Rubric) - E-news Express

Criteria	Points
Define the problem and perform an Exploratory Data Analysis - Problem definition, questions to be answered - Data background and contents - Univariate analysis - Bivariate analysis	8
Illustrate the insights based on EDA Key meaningful observations on individual variables and the relationship between variables	6
Do the users spend more time on the new landing page than the old landing page? - Perform visual analysis - Formulate null and alternative hypotheses - Select the appropriate test - Calculate the p-value - Write the inference based on the p-value	10
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? - Perform visual analysis - Formulate null and alternative hypotheses - Select the appropriate test - Calculate the p-value - Write the inference based on the p-value	10
Does the converted status depend on the preferred language? - Perform visual analysis - Formulate null and alternative hypotheses - Select the appropriate test - Calculate the p-value - Write the inference based on the p-value	10
Is the mean time spent on the new page same for the different language users? - Perform visual analysis - Formulate null and alternative hypotheses - Select the appropriate test - Calculate the p-value - Write the inference based on the p-value	10

Criteria	Points
Presentation/Notebook - Overall Quality	
- Structure and flow - Crispness - Visual appeal - Conclusion and Business Recommendation OR - Structure and flow - Well commented code - Conclusion and Business Recommendation	6
Points	60