Great Learning

Context

The EdTech industry has been surging in the past decade immensely, and according to a forecast, the Online Education market, would be worth $286.62bn by 2023, with a compound annual growth rate (CAGR) of 10.26% from 2018 to 2023. The modern era of online education has enforced a lot in its growth and expansion beyond any limit. Due to having many dominant features like ease of information sharing, personalized learning experience, transparency of assessment, etc., it is now preferable to traditional education.

The online education sector has witnessed rapid growth and is attracting a lot of new customers. Due to this rapid growth, many new companies have emerged in this industry. With the availability and ease of use of digital marketing resources, companies can reach out to a wider audience with their offerings. The customers who show interest in these offerings are termed as**leads**. There are various sources of obtaining leads for Edtech companies, like:

- The customer interacts with the marketing front on social media or other online platforms.
- The customer browses the website/app and downloads the brochure.
- The customer connects through emails for more information.

The company then nurtures these leads and tries to convert them to paid customers. For this, the representative from the organization connects with the lead on call or through email to share further details.

Objective

ExtraaLearn is an initial stage startup that offers programs on cutting-edge technologies to students and professionals to help them upskill/reskill. With a large number of leads being generated on a regular basis, one of the issues faced by ExtraaLearn is to identify which of the leads are more likely to convert so that they can allocate the resources accordingly. You, as a data scientist at ExtraaLearn, have been provided the leads data to:

- Analyze and build an ML model to help identify which leads are more likely to convert to paid customers.
- Find the factors driving the lead conversion process.
- Create a profile of the leads which are likely to convert.

Data Description

The data contains the different attributes of leads and their interaction details with ExtraaLearn. The detailed data dictionary is given below.

- **ID:**ID of the lead

- **age:**Age of the lead

- **current_occupation:**Current occupation of the lead. Values include 'Professional', 'Unemployed', and 'Student'

- **first_interaction:**How did the lead first interact with ExtraaLearn? Values include 'Website' and 'Mobile App'

- **profile_completed:**What percentage of the profile has been filled by the lead on the website/mobile app? Values include Low - (0-50%), Medium - (50-75%), High (75-100%)

- **website_visits:**The number of times a lead has visited the website

- **time_spent_on_website:**Total time (seconds) spent on the website.

- **page_views_per_visit:**Average number of pages on the website viewed during the visits

- **last_activity:**Last interaction between the lead and ExtraaLearn

    - **Email Activity:**Seeking details about the program through email, Representative shared information with a lead like a brochure of the program, etc.
    - **Phone Activity:**Had a phone conversation with a representative, had a conversation over SMS with a representative, etc.
    - **Website Activity:**Interacted on live chat with a representative, updated profile on the website, etc.
- **print_media_type1:**Flag indicating whether the lead had seen the ad of ExtraaLearn in the Newspaper

- **print_media_type2:**Flag indicating whether the lead had seen the ad of ExtraaLearn in the Magazine

- **digital_media:**Flag indicating whether the lead had seen the ad of ExtraaLearn on the digital platforms

- **educational_channels:**Flag indicating whether the lead had heard about ExtraaLearn in the education channels like online forums, discussion threads, educational websites, etc.

- **referral:**Flag indicating whether the lead had heard about ExtraaLearn through reference.

- **status:**Flag indicating whether the lead was converted to a paid customer or not. The class 1 represents the paid customer and class 0 represents the unpaid customer.

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 report 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.

Submission type	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	Submission Format
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 report: 10 - 20% grading on the quality of the final businessreportsubmitted	Businessreportin .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 report, ONLY thereportwill 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 thesolution 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 businessreportwith insights and recommendations to the business problem.
- Submit only thereport [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

- Thereportshould be made keeping in mind that the audience will be the Data Science lead of a company.
- The key points in thereportshould 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 report.
- Thereportshould be submitted as a PDF file (.pdf) and NOT as a .doc file.
- Please refer to the FAQ page for common project-related queries.

Happy Learning!