summry report

The analysis is done for x eduction company and to find ways to get more industry professional to join their resources.the basic data provided me which gives a lot of information about the potential customer visit the site, the time they spend there,how they reached the site and the concersion rate.

We have used the following step to analysis this case study to get potential students eho actually interested.

1.Read and understand the- data-using pandas librery

2.Data cleaning-the data partially clean except for a few null value and option select hed to be replaced with a null value since it did not give us mucg=h information.Few of the null values were changed to "not provided ' so as tonot lose much data .although they were many from India and few from outside .the elements were changd ti india ,outside india not provided

3.EDA-A quick EDA was done to check the condition of our data . it was found that a lot of elements in the categorical were irrelevant.the numeric values seem good and no outliers are found.

4.Data preparation- *We have two binary columns: do_not_email, mastering_interview*

5.Train and split thye data - *The conversion rate is 38.5%*

6.Model building

- *7.Conclusion-*
- *The lead_score column can be used to identify the potential leads to focus first.*
- *Higher the score, higher are the chances for the lead to convert.*

- *In case, there are limited sales representatives, then the score cut-off should be higher to ensure a higher conversion probability people are contacted further to turn them into a potential customer.*
- *It is the same as increasing the precision value of the model by adjusting the cut-off point to a higher value.*
- *In case there are more resources available in the sales team (i.e., interns, etc.), then the score cut-off can be lowered.*

As there are more human resources, the company can afford a higher rate of False positives as it will increase the customer outreach and, in turn, increase the potential customer who will take the online courses