Summary

This analysis is done for X Education and to find ways to convert more leads to join their courses. Although X Education gets a lot of leads, its lead conversion rate is very poor. There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers. The company needs a model wherein a lead score is assigned to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The data provided has given us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The following are the steps used:

1. **EDA :**

   The data was cleaned by dealing with nulls, dropping customer level columns, dropping those columns which has high percentage of nulls. Some values in columns were categorised to 'others' as they had less than 1% of values. Since there were many from India and few from outside, the elements were changed to 'India', 'Outside India' and 'not provided'. A quick EDA was done to check the condition of our data. Univariate and bivariate analysis was done. There were outliers in 'Page views per visit' and 'Total visits'.

   - Few insights from EDA
     1) Leads from India is high
     2) Working professionals are more likely to join the course
     3) Students are less likely to join the course
     4) People spending more time on website have higher probability of taking the course

2. **Model Evaluation:**

   ROC is 0.89, which indicates that our model is statistically good. Optimal cut off is set to 0.37. The optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity.

3. **Potential Variables leading to conversion of leads:**

   - **The total time spend on the Website** : if customer spent more time on website, he/she is more likely to convert
   - **Total number of visits** : More likely to convert if the customer is visiting site again and again
   - **The lead source** : if customer visited site directly or through google or through referral sites then they are less likely to convert
   - **The last activity** : SMS ,Olark chat conversation
   - **Lead_source_welingak website**: If lead was obtained from this website, conversion chances are high
   - **occupation_working professional**: More likely to convert, than people in other occupation
   - **lead_source_reference**: if lead was obtained through refrence then more like to convert

- **lead_source_direct traffic**: if customer visted site directly or through google or through referral sites then they are less likely to convert

Keeping these factors in mind the X-Education can see potential growth as they have very high chance to get most of the customers to change their mind and buy their courses. They should focus less on students and unemployed people as they are less likely to buy the courses.