Executive Summary

X Education is an online education company that offers courses to industry professionals. The company markets its courses on several websites and search engines. Once people land on the website, they may browse the courses, fill out a form for a course. When people provide their email address or phone number, they are classified as a lead. The leads are then contacted by the sales team to convert them into customers. The typical lead conversion rate at X Education is around 39%. The dataset was pre-processed and cleaned, which included handling missing values, dropping unnecessary columns, imputing data and treating outliers. To reduce the dimensionality of the dataset and select only the important features, Recursive Feature Elimination (RFE) was performed, and only 20 columns were selected out of the original 63.

Logistic regression models were built, and manual feature reduction processes were used to build models by dropping variables with a p-value greater than 0.05 and VIF value greater than 5. Logistic Regression Model_6 was chosen as the final model as it showed significant p-values within the threshold, and there was no sign of multicollinearity with VIFs less than 5.

The top three variables that influenced the lead conversion rate were 'total_time_spent_on_website', 'last_notable_activity_had_a_phone_conversation', 'lead_source_welingak_website'. Customers who spent more time on the website were more likely to convert, and leads with which we had a phone call and leads generated through welingak websites were more likely to convert. The sales team can use this information to prioritize their efforts and focus on leads that have a higher chance of conversion.

Based on these observations from EDA, several recommendations were made, like:
     1. Features such as 'total_time_spent_on_website', last_notable_activity_had_a_phone_conversation', and 'lead_source_welingak_website' have a high conversion rate and should be utilized more in lead generation efforts.
     2. Working professionals should be aggressively targeted as they have a higher probability of converting and are likely to have better financial situations to pay for services.
     3. Increasing the frequency of media usage such as Google ads or on site can save time and increase the conversion rate as we have conversions more from google and direct traffic to site.
     4. Leads whose 'Last Activity' is 'SMS Sent' or 'Email Opened' tend to have a higher conversion rate and should be targeted more frequently along with phone conversation.
     5. Analyzing the behaviour of customers who spend more time on the website can help improve the user experience and increase conversion rates, as this is an important factor , means people spending more time have good chances of being converted.
     6. Making calls and emails to leads also impact its conversion to high extent , so if can focus more on tailored mail and personalized calls.