Summary of the Lead Score Assignment

This analysis was done for X Education in order to increase the conversion rate from regular Leads to paying clients. X Education contributed the data set, which contained a sizable amount of data. The organisation needs to raise its 30% conversion rate to 80%. In this scenario, we must develop a model with a ts Lead score range of 010–100. A higher score will increase the likelihood of conversion, but a lower score will significantly reduce the likelihood of a successful conversion into a paying customer.

The technical procedures employed for the analysis are as follows:

1. **Data cleaning.**
   A tiny number of columns that included a high proportion of null values were removed. spotted the outliers and adjusted the decline.

   After those steps, data was preserved in up to 68% of instances, and this cleaned data was then subjected to analysis.

2. **EDA**
   By using a variety of plot styles and analysing the continuous and categorical variables, EDA was carried out on the cleaned data.

   A univariate analysis was done on the target variable to aid with comprehension.

   These are some of the conclusions:
   1. More time-consuming individuals are promising Leads.
   2. Among the others, the Lead Origin-Landing Page Submission has the highest conversion rate.
   3. The best conversion rate can be found on Google.
   4. The best conversion rate was seen among leads whose most recent activity was an SMS.
   5. Select columns have the highest rate of lead from specialisation who are unknown
   6. Person who are unemployed has the highest conversion rate relatively to working professional.

3. **Scaling, Train-Test separation, and development of dummy variables.**
   For categorical columns, scaling and the introduction of dummy variables were done. All of the features were scaled to be within a comparable range. A split of 30% went to the test and 70% to the train was used.

4. **Model Building and Predictions**
   The features were selected using the RFE technique, and five modelling attempts were done until the VIF and p-values were within allowable bounds.

   Final accuracy, sensitivity, and specificity for the test set were 78%, 79%, and 77% respectively. Recall was 78% and Precision was 77%.

0.43 as chosen as the cutoff, and predictions were produced using that value.

5. Conclusion

The elements that are crucial for prospective Leads include Current employment status is "unemployed,"

"Total visits," "LeadOrigin Lead Add Form," "Latest Activity as SMS delivered," and "Total time spent on the website" are some examples.

Recommendations:

- Regular data collection, model running, and lead updates are all beneficial. It's generally accepted that the optimal time to call prospective leads is shortly after they express interest in your courses.
- Email is just as effective as cold calling, so it's a good idea to mail the leads in addition to making phone calls to remind them.
- You can save a lot of time by limiting the number of call efforts to two to four and boosting the frequency with which you use other media, such as Google ads or emails, to stay in touch with leads.
- By concentrating on hot leads, which have a low contact rate but high conversion rates, we have a better chance of bringing in more value to the company.

Business advice for improving conversion rate:

1. Lead score more than 68. Definitely converted leads.
2. More than 12 hours were spent online overall.
3. Website and references from the lead source Welingak.

Very little chance of leads being converted:

1. Clients chose the "Do not email" option.
2. A lead score under 15.
3. Less than 5 hours were spent on the website overall.
4. Lead sources include Google, organic search, referral sites, and direct traffic.
5. The last activity of the users was either an Olark chat session, a website page visit, a bounced email, a website form submission, or an email link click.