SUMMARY

1. In this lead scoring case study Logistic regression model has been used to make predictions whether the leads convert or not.
2. Insights from EDA is that most of the leads are from INDIA and in India most are from Mumbai.
3. The lead conversion rate is very high in lead source Welingak Website and References and google has highest leads, but conversion rate is not that high.
4. Total time spent on website may increase the chances of lead getting converted.
5. Talking to last notable Activity, making improvement in customer engagement through email and calls helps in lead conversion. As the leads which are opening email have high chances to convert also sending SMS will also benefit.
6. Most of the leads current occupation is Unemployed, which means gave more focus on unemployed leads.
7. The final model built has following details:
8. Train data:

   Sensitivity :79.4%, Specificity :79.4%, Accuracy :80.0%

   Test data:

   Sensitivity :81.8%, Specificity :79.7%, Accuracy : 80.5%

9. Company should not make calls to leads coming from "Lead Source_Olark Chat","Last Activity_Email Opened","Last Activity_Olark Chat Conversation","Lead Origin_Landing Page Submission","Do Not Email" and "Specialization_Others ".
10. Company should make calls to leads coming from "Lead Source_Welingak Website ","Lead Source_Reference","Last Activity_Other_Activity","Last Notable Activity_Unreachable ","Last Notable Activity_SMS Sent ","Last Activity_Unsubscribed","Last Notable Activity_Email Bounced" and "Total Time Spent on Website".