

# SUMMARY

Gayatri Bhinge



## **ANALYSIS OBJECTIVE:**

This analysis focuses on X Education's goal of attracting more industry professionals to enroll in their courses. The initial dataset provides valuable insights into the behavior of potential customers, including how they visit the website, the duration of their visits, their referral sources, and the conversion rates.

Our task is to identify key factors that influence these metrics and suggest strategies to improve conversions and increase professional enrollments.

## THE FOLLOWING ARE THE STEPS USED:

- ❑ **Cleaning Data:** The data was mostly clean, with a few null values. The "option select" field was replaced with null as it didn't provide meaningful information. Some null values were changed to "not provided" to preserve as much data as possible. These values were later removed while creating dummy variables. Since the dataset included many entries from India and a few from other countries, the location data was categorized into "India," "Outside India," and "Not Provided."
- ❑ **EDA:** A quick exploratory data analysis (EDA) was performed to assess the dataset. It revealed that many categorical variables had irrelevant elements. The numeric variables appeared in good condition, and no significant outliers were found.
- ❑ **Dummy Variables:** Dummy variables were created, and those containing "not provided" elements were removed. For numeric variables, the MinMaxScaler was applied to normalize the data.
- ❑ **Train-Test Split:** The data was split into training and testing sets, with a 70% training and 30% testing ratio.

## THE FOLLOWING ARE THE STEPS USED:

- ❑ **Model Building:** Recursive Feature Elimination (RFE) was used to identify the top 15 relevant variables. The remaining variables were manually removed based on their Variance Inflation Factor (VIF) and p-values (keeping those with  $VIF < 5$  and  $p\text{-value} < 0.05$ ).
- ❑ **Model Evaluation:** A confusion matrix was generated, followed by the ROC curve to determine the optimum cut-off value. Accuracy, sensitivity, and specificity were calculated, each achieving approximately 80%.
- ❑ **Prediction:** Predictions were made on the test dataset, using an optimum cut-off value of 0.35. The accuracy, sensitivity, and specificity were all around 80%.
- ❑ **Precision-Recall:** To further validate the model, precision-recall analysis was conducted. A cut-off of 0.41 was determined, with precision at approximately 73% and recall around 75% on the test dataset.

## **THE MOST SIGNIFICANT VARIABLES INFLUENCING POTENTIAL BUYERS, IN DESCENDING ORDER, WERE IDENTIFIED AS FOLLOWS:<sup>5</sup>**

- ☐ Total time spent on the website.
- ☐ Total number of visits.
- ☐ Lead source, particularly when coming from:
  - Google
  - Direct traffic
  - Organic search
  - Welingak website
- ☐ Last activity, especially when it involved:
  - SMS
  - Olark chat conversation
- ☐ Lead origin from a Lead Ad format.
- ☐ Current occupation as a working professional.
- ☐ With these key insights, X Education has a strong opportunity to effectively target potential buyers and convert them into course enrollments, thereby maximizing their growth and success.

The background features a large, light pink circle on the right side, partially overlapping a dark blue circular segment at the top and bottom. The left side of the image is a solid light cream color.

**THANK  
YOU!**