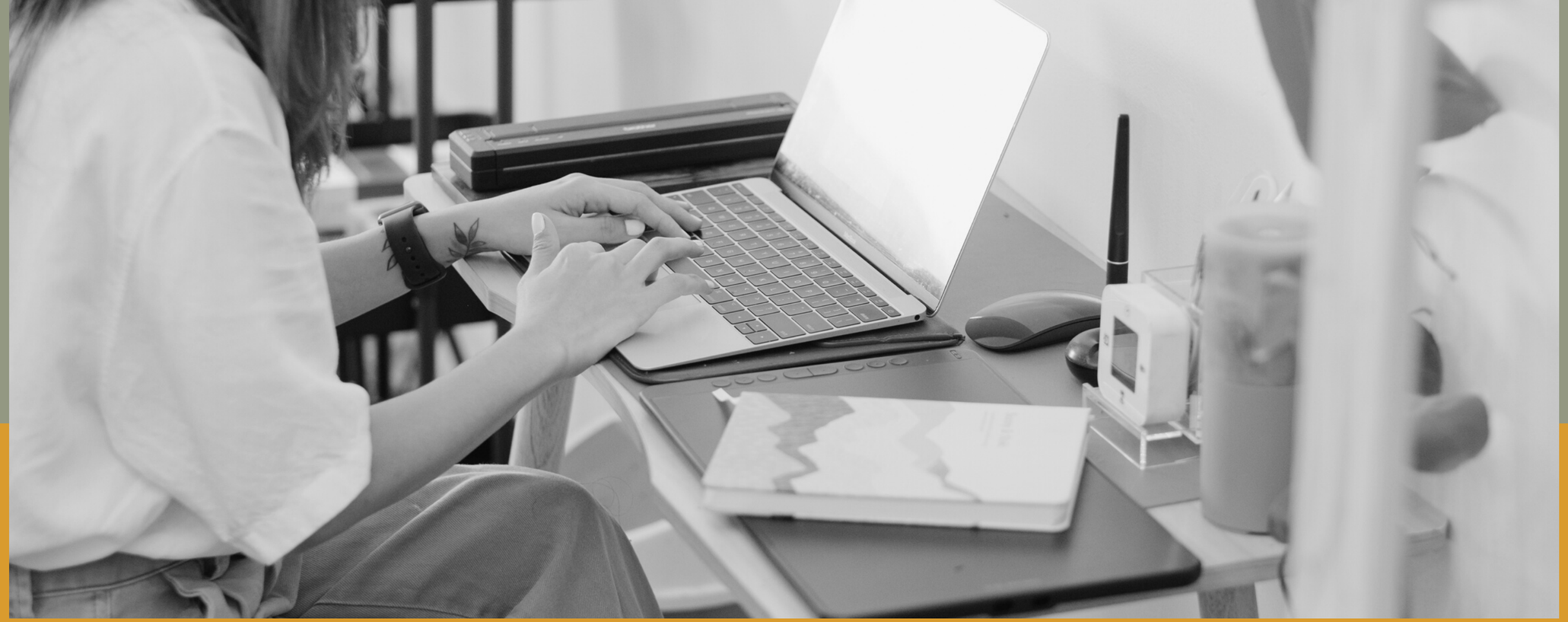


LEAD SCORING CASE STUDY

GOAL

To establish a data backed approach which in turn helps the firm **X Education** identify leads with high potential (aka "Hot Leads") resulting in improvement of lead conversion rate

APPROACH

To tackle this business problem we will be using a logistic Regression model

- Data Clean Up : Perform Null Value & Outlier treatment
- EDA : Check feature correlations and significance
- Data Preparation : Perform One hot Encoding, Feature Scaling & Train Test Split
- Model Building
  - Build a logistic regression model
  - Optimize using RFE and VIF
  - Find optimal cut-off point using 'Sensi-Speci' and 'Precession-Recall' Curve
- Model Evaluation : Use Accuracy score and Confusion metrics to assess the model
- Make Predictions & Share findings

ASSUMPTIONS

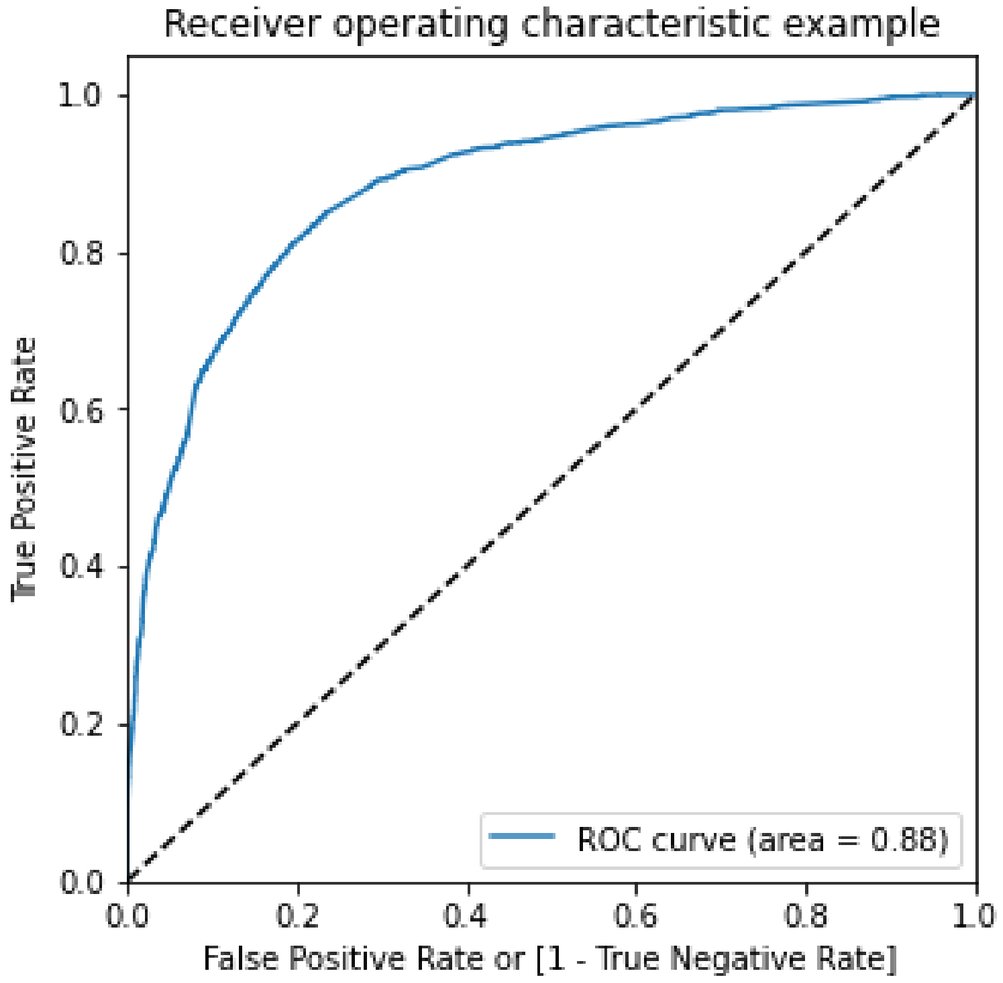
Some of the key assumptions in our projects are :

- Features with 30% or more null values are discarded from Analysis
- Dropping Country column as 96% percent represents a single country and all other countries' composition is insignificant
- Dropping the rows for column where null value <= 1%

THE MODEL

- Logistic Regression Model
- Features narrowed using RFE & VIF
- Optimal cut off : 0.42
- Area under ROC curve : 0.88

Train Data:	Test Data:
• Accuracy : 81.0%	• Accuracy : 81.0%
• Sensitivity : 75.1%	• Sensitivity : 74.0%
• Specificity : 84.7%	• Specificity : 85.0%
• Precision : 75.5%	• Precision : 73.8%
• Recall : 75.1%	• Recall : 74.0%



RECOMMEDATIONS

- **More the time** invested by the customer on our website, higher is the chance of them joining a course
- The leads with 'Lead Origin' value as **Lead Add** Form and **Lead Import**, portray trend of converting
- We noticed that **Working Professional** are more likely to join the course against other class members
- The Customers who had a Phone Conversation with our support team show high probability of enrolling to a course.
- Leads sourced from Welingak Website and Olark Chat have been proven to be valuable in conversion