LEAD SCORING CASE STUDY

FOCUSED BUSINESS APPROACH USING LOGISTIC REGRESSION TECHNIQUE

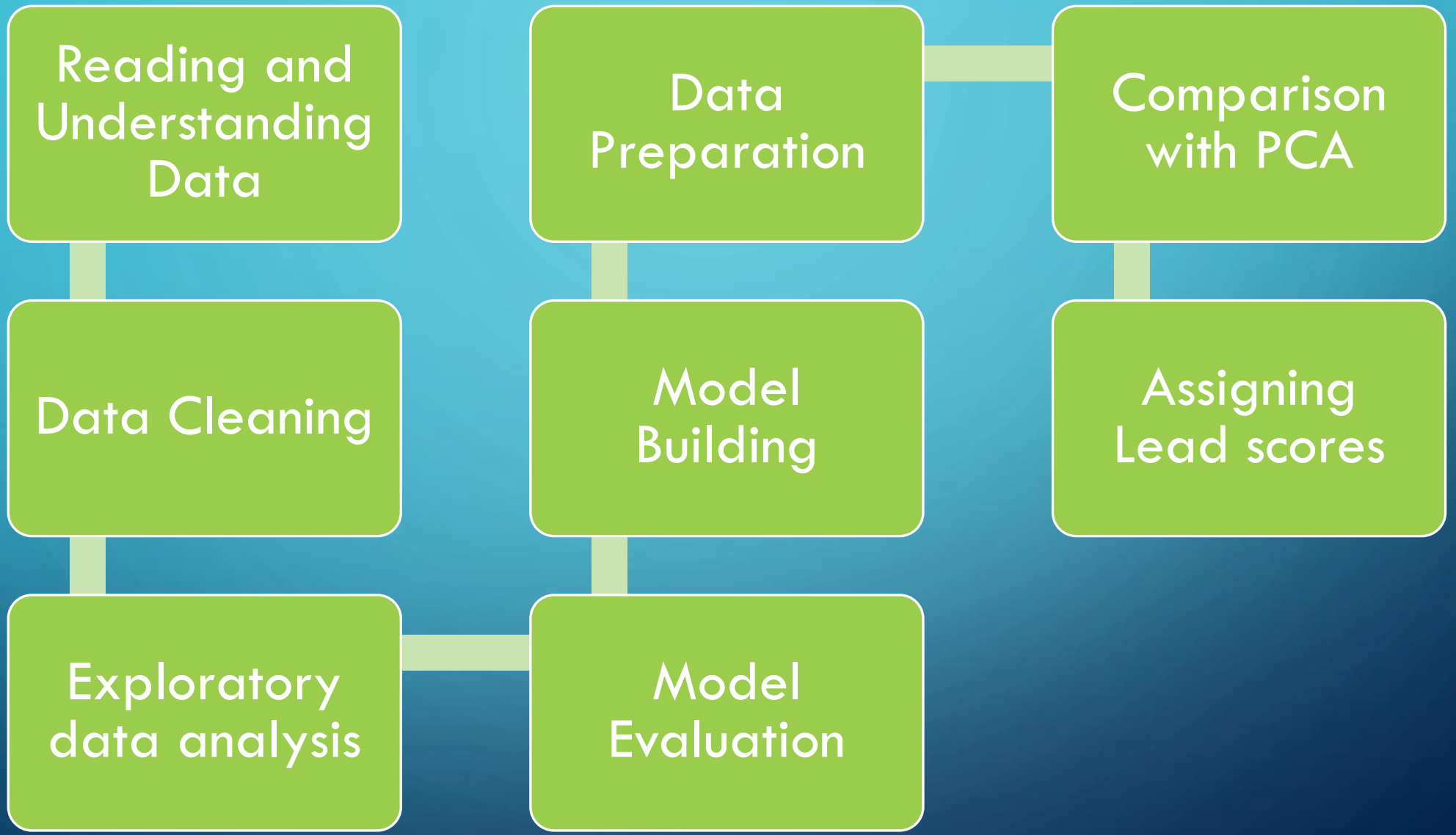
RAINA SHARMA

BUSINESS OBJECTIVE

- To help X Education select most promising leads that are most likely to convert into paying customers.

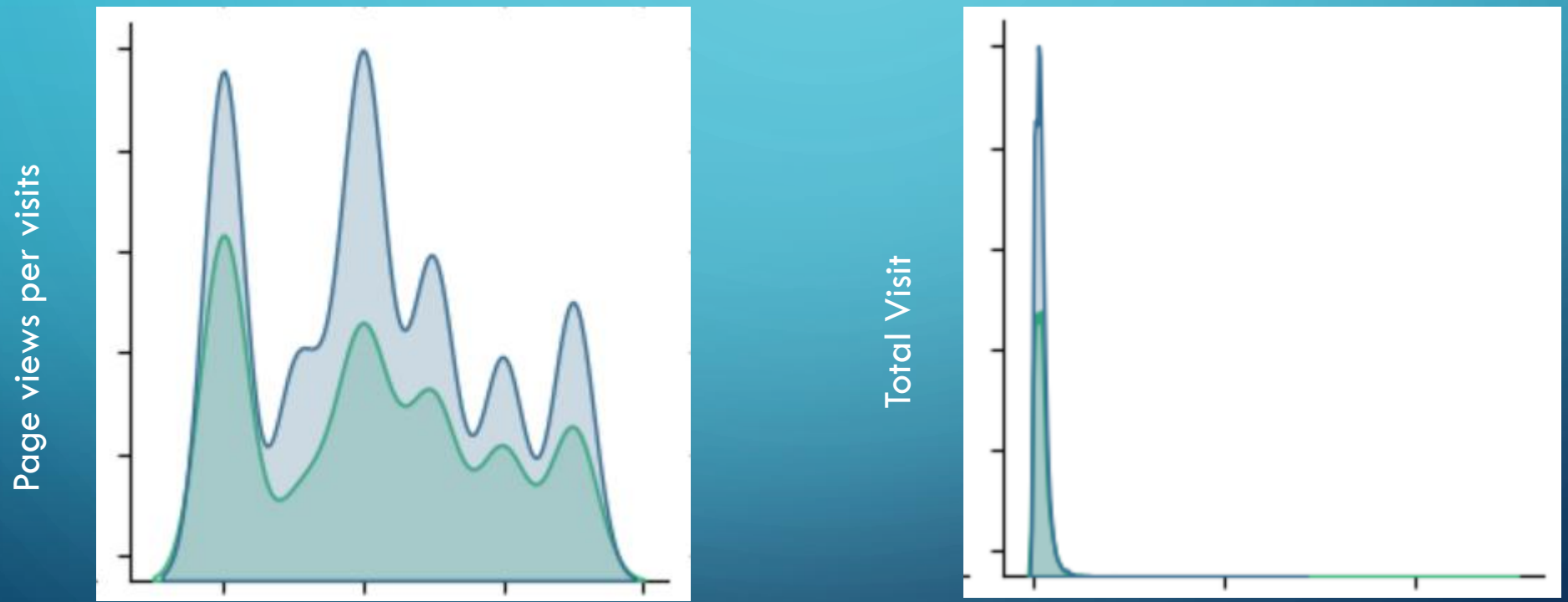
METHODOLOGY

To build a Logistic Regression Model that assigns lead scores to all leads such that the customers with higher lead score will have a higher conversion chance.

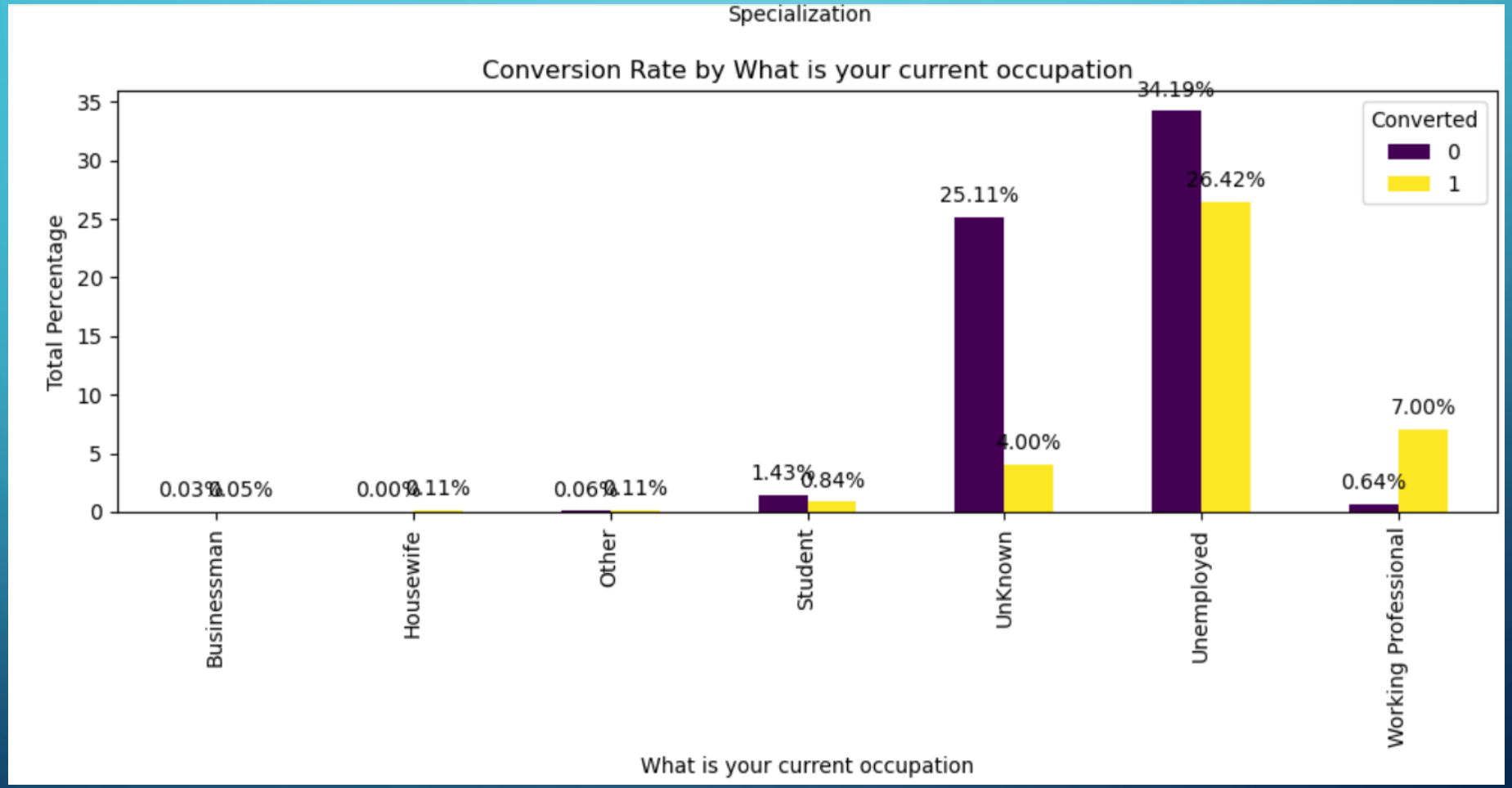

Reading and Understanding Data
Data Cleaning
Exploratory data analysis
Data Preparation
Model Building
Model Evaluation
Comparison with PCA
Assigning Lead scores

DATA VISUALIZATION

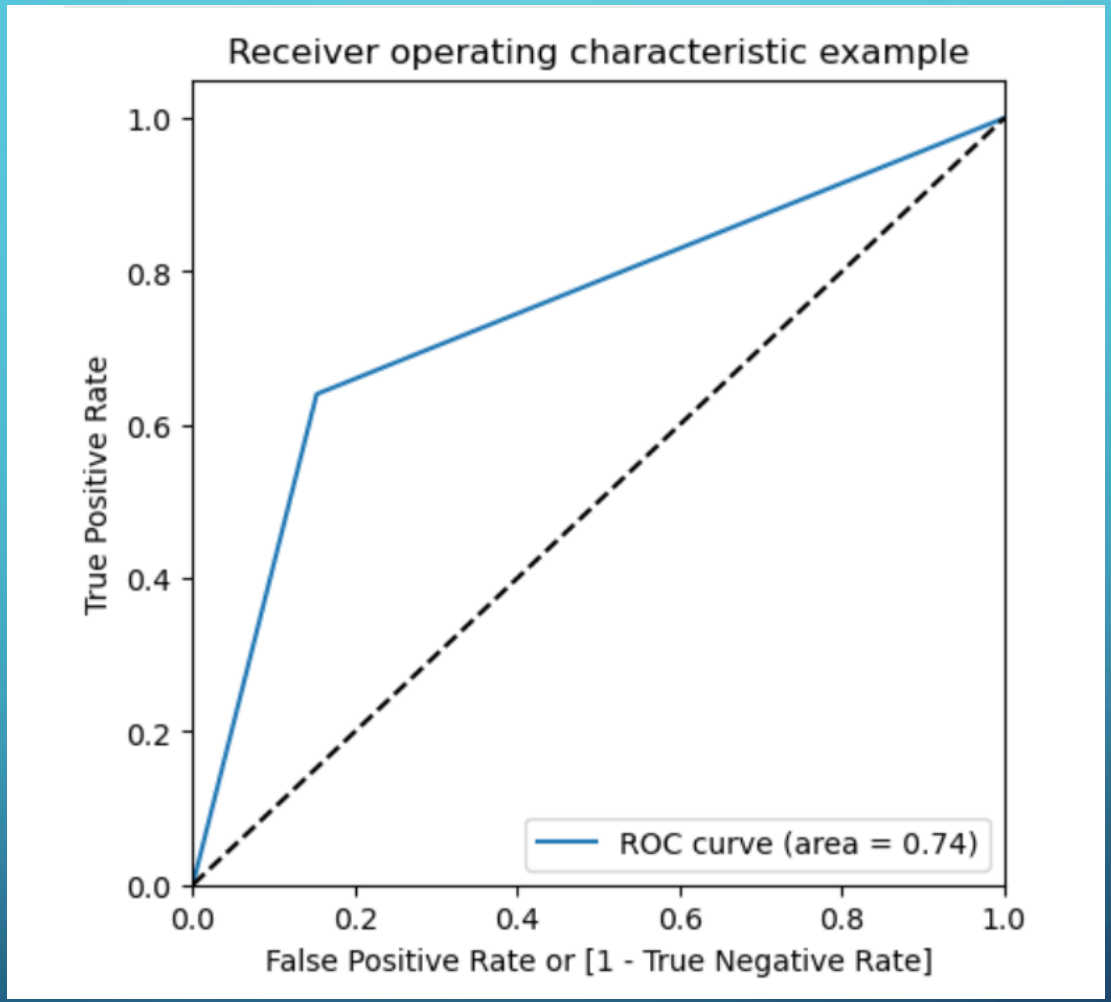
The Page views per visit is highly correlated to Total Visit More the Total Time Spent on Website highest is the rate of conversion.



What is your current occupation_Unemployed has a strong positive impact on lead conversion.

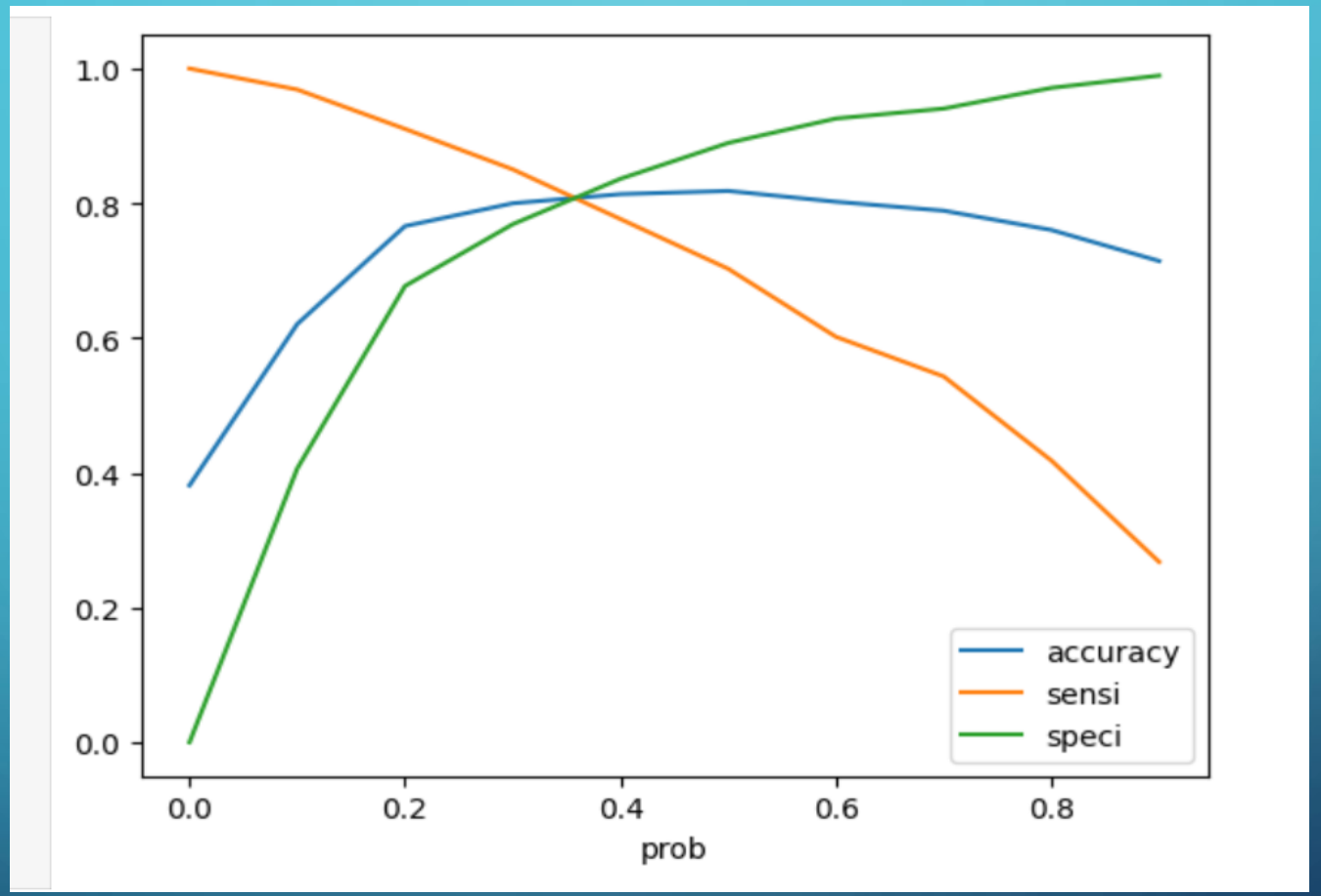


ROC CURVE



The ROC curve's area is 0.74, indicating a strong model performance.

FINDING OPTIMAL THRESHOLD



From the curve above, 0.35 is the optimum point to take it as a cutoff probability.

CONCLUSION

- While we have checked both sensitivity-specificity as well as Precision & Recall metrics, we have considered the optimal cut off based on sensitivity & specificity for calculating the final prediction.
- Accuracy, Sensitivity & Specificity values of test set are around 76.8, 63.9%, 84.7% which are approximately closer to values using trained Data set.
- Overall, this model proves to be accurate.

THANK YOU