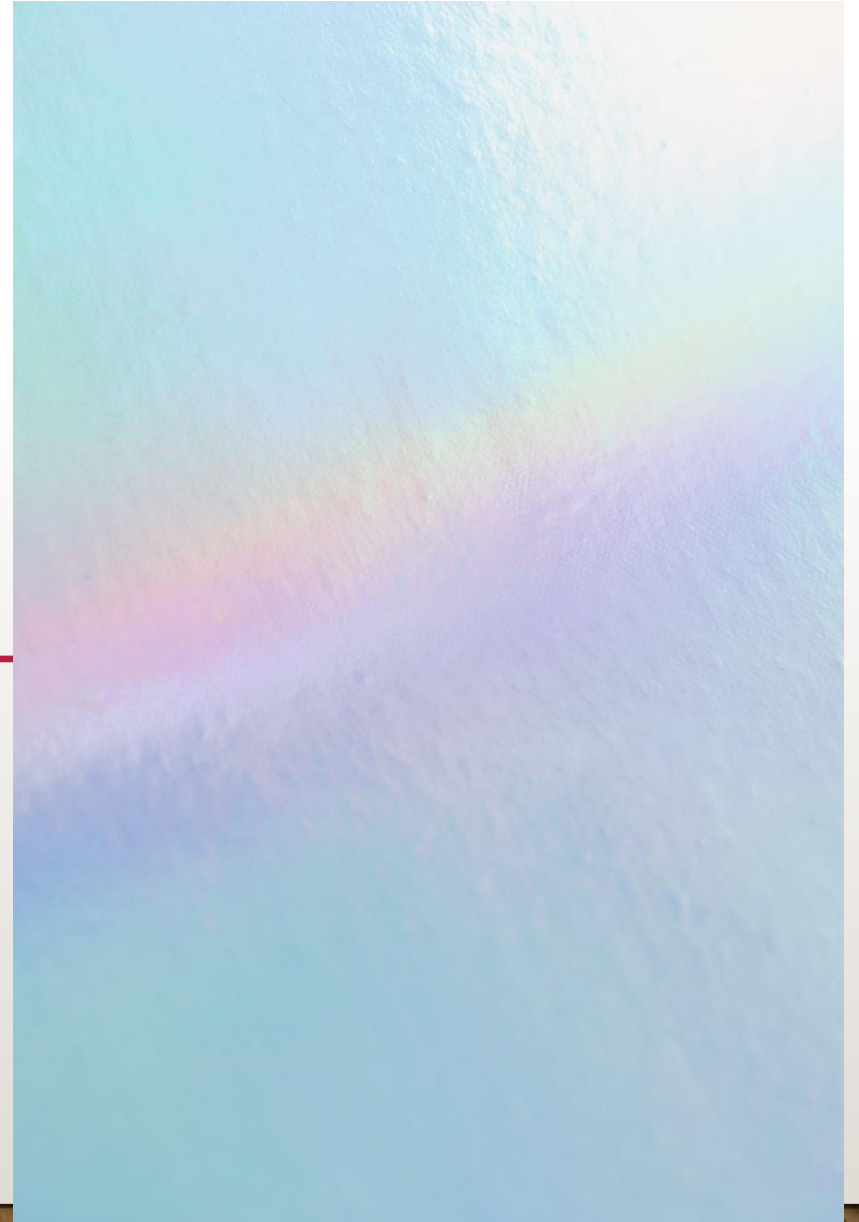


# LEAD SCORING ASSIGNMENT

---

UMAMAGESH R



# PROBLEM STATEMENT

---

X Education, an online education company, faces challenges with low lead conversion rates despite generating numerous leads daily. To improve efficiency and focus on potential customers, they seek to identify "Hot Leads" with higher conversion chances.

The goal is to build a lead scoring model to assign scores to leads, prioritizing higher conversion chances. By targeting leads with higher scores, the company aims to increase the overall lead conversion rate. The CEO's target is to achieve around 80% lead conversion.

To achieve this, we will analyze historical data, including lead source, website interactions, past referrals, and lead characteristics. Through data preprocessing, feature engineering, and machine learning techniques like logistic regression, we will create a lead scoring model.

This model will enable the sales team to concentrate efforts on the most promising leads, optimizing lead conversion rates and maximizing the company's revenue.



## GOALS OF THE CASE STUDY

:Build a logistic regression model to assign a lead score between 0 and 1 to each of the leads which can be used by the company to target potential leads.

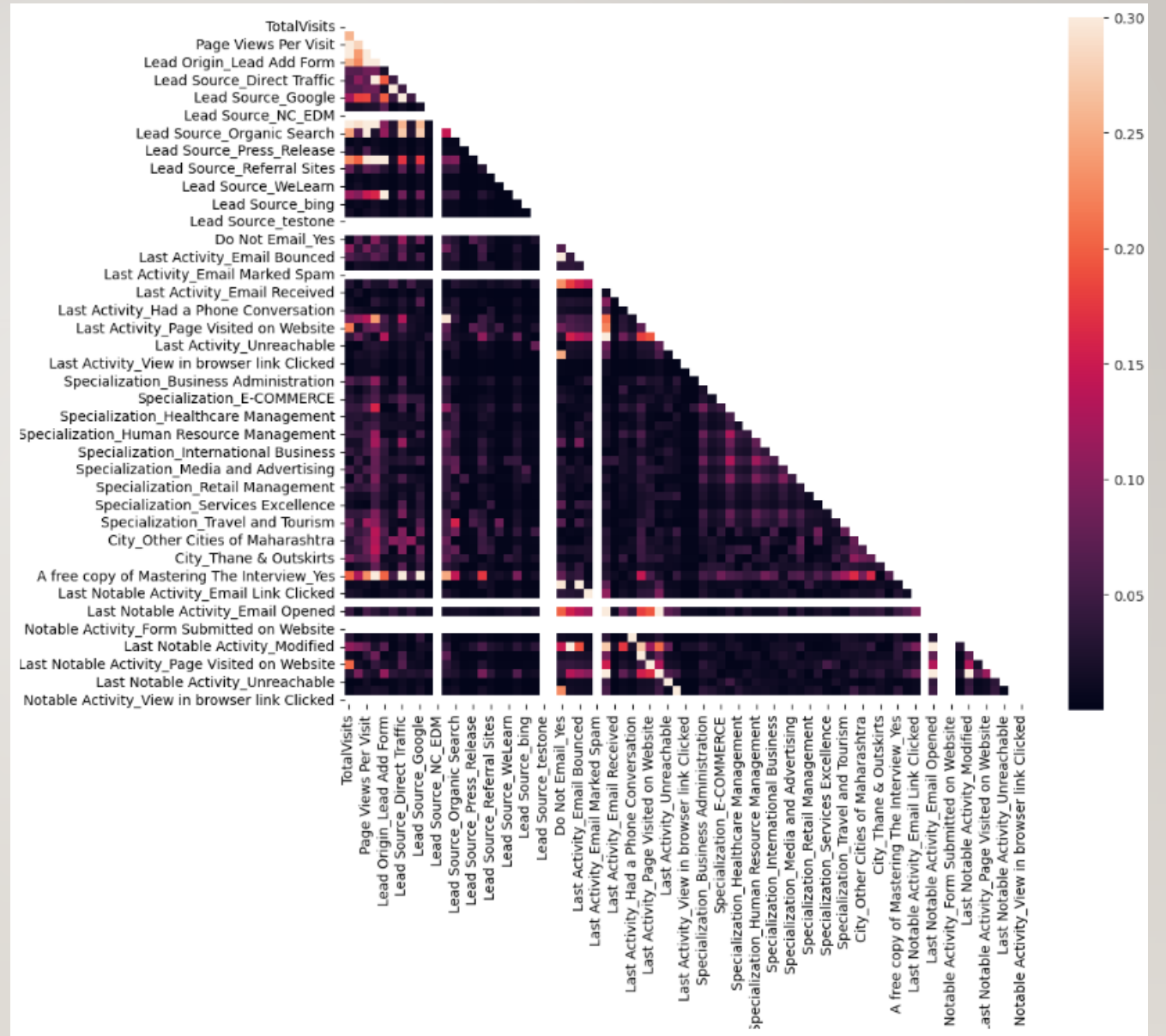
A higher score would mean that the lead is hot, i.e., is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

# DATA PREPARATION

1. Import the raw data. Shape of data frame: 9240-rows x 37-columns.
2. Variables with more than 20 percent null values are dropped. 11 variables were dropped.
3. Columns with unique values are dropped as it does not contribute to the modelling.
4. Categorical columns with does not have much variation in the values(Imbalance data) & its distribution are dropped as it does not help in model building. 12 such columns were dropped.
5. Next, the rows with null values are removed.
6. At last, 17% of data were during data cleaning. Remaining data will be used for Model building.
7. Dummy Variable creation
8. Test Train split

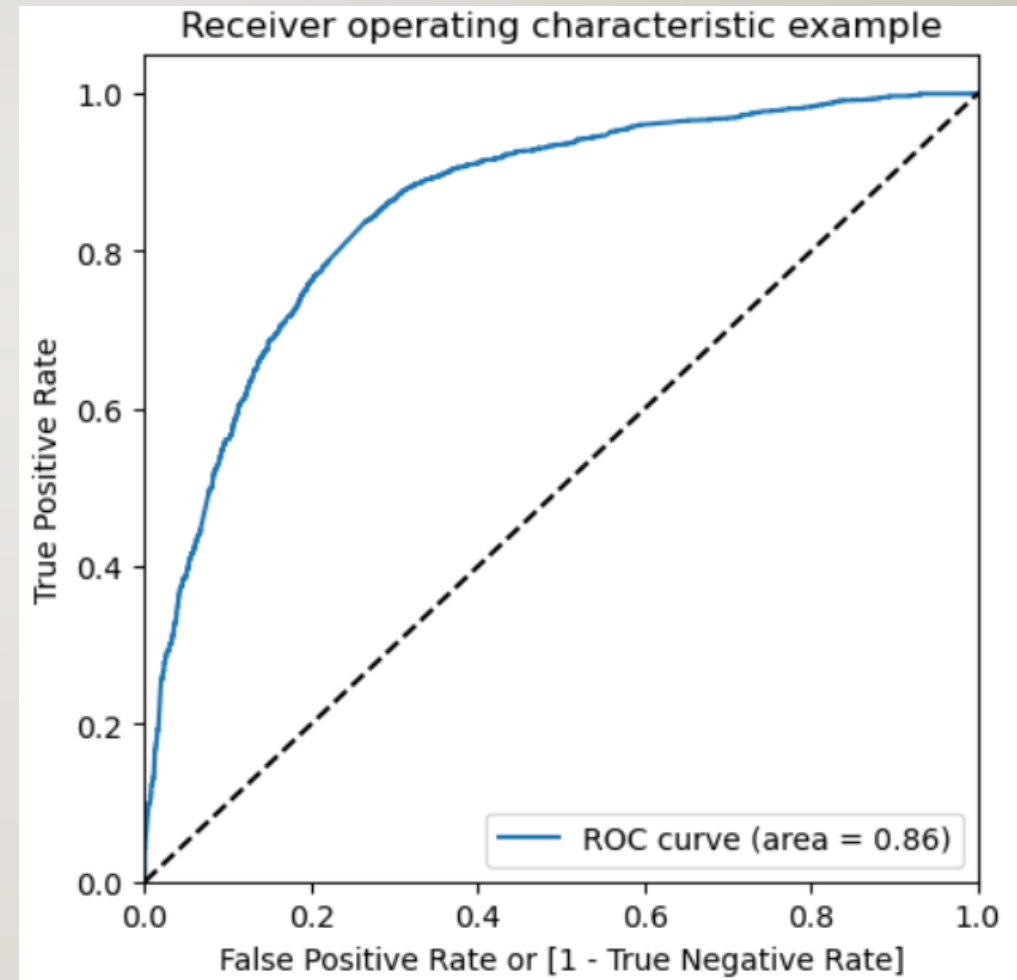


# CORRELATION

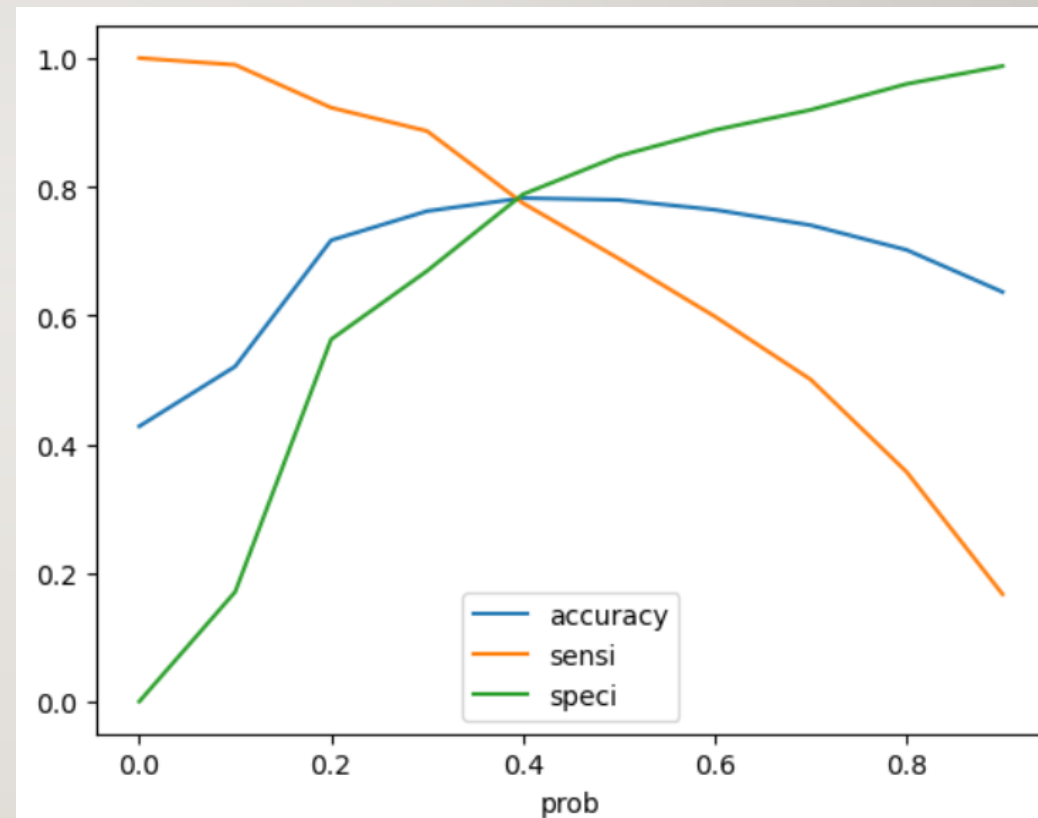


# ROC CURVE

---



# COMPARISON OF ACCURACY, SENSITIVITY & SPECIFICITY BASED ON VARIOUS PROBABILITIES



# MODEL PARAMETERS

---

- Accuracy: 0.76
- Sensitivity : 0.82
- Specificity : 0.72
- Precision : 0.69
- Recall : 0.82



## CONCLUSION

- The lead score calculated in the test set of data shows the conversion rate of 76% on the final predicted model.
- Good value of sensitivity of our model will help to select the most promising leads.
- Features which contribute more towards the probability of a lead getting converted are:
  - Total Time Spent on Website
  - Lead Source\_Welingak Website
  - Lead Source\_Reference