LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

SUBMITTED BY :

1. SUNILKUMAR MENON

2. MEDINI HEGDE

3. MITALI SINGH RAJPUT

CONTENTS

- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around 30%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

BUSINESS OBJECTIVE

- Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.

PROBLEM APPROACH

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE R-squared VIF and p-values)
- Model Evaluation
- Making predictions on test set

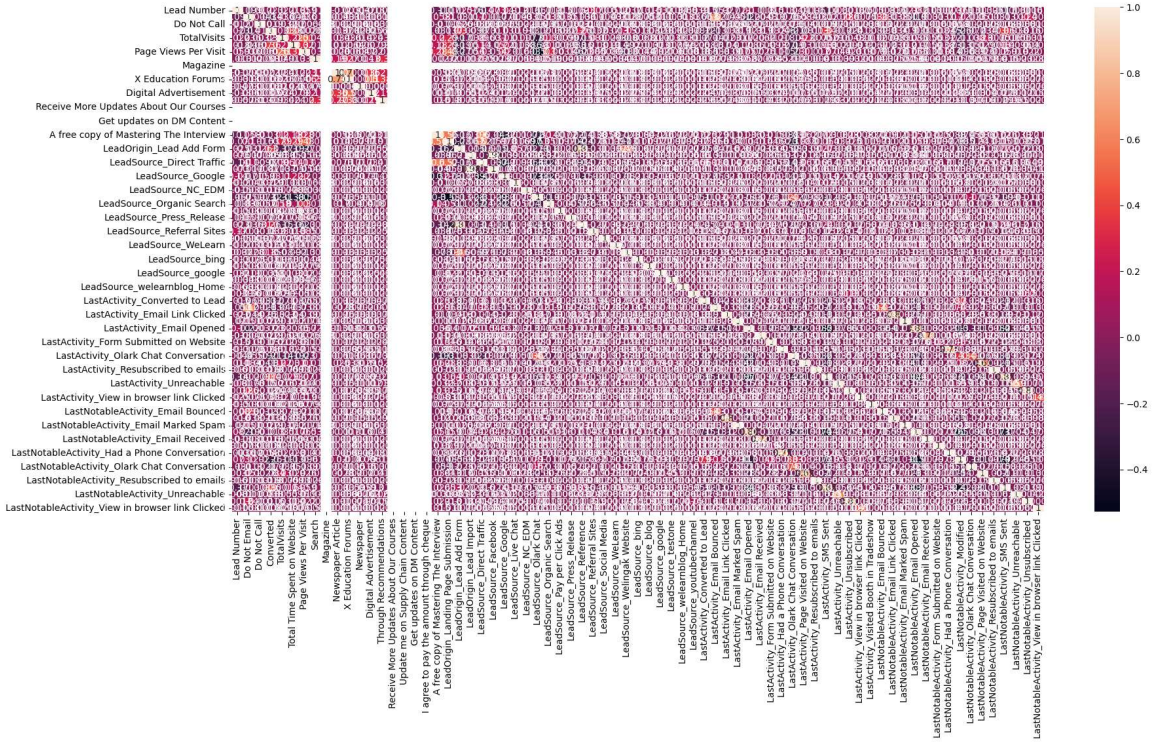
EDA – DATA CLEANING

Key findings from the data:

- Most leads are from India, with Mumbai leading among cities.
- Columns like "Select" need mandatory selection for better insights.
- High engagement metrics (visits and time spent) positively impact conversions.

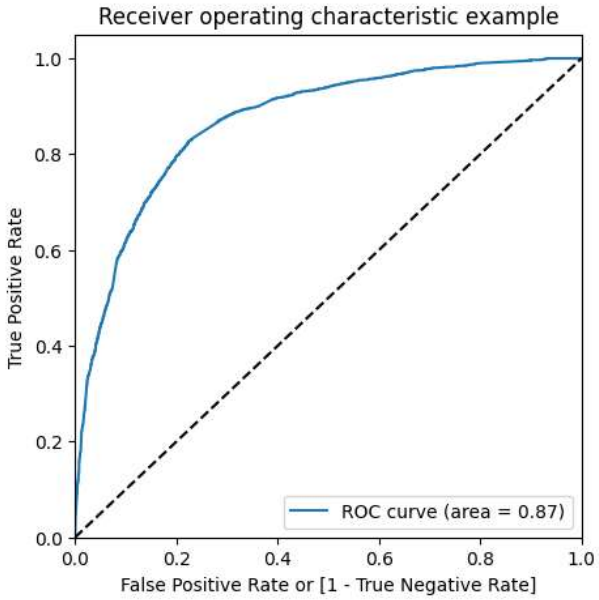
CORRELATIONS

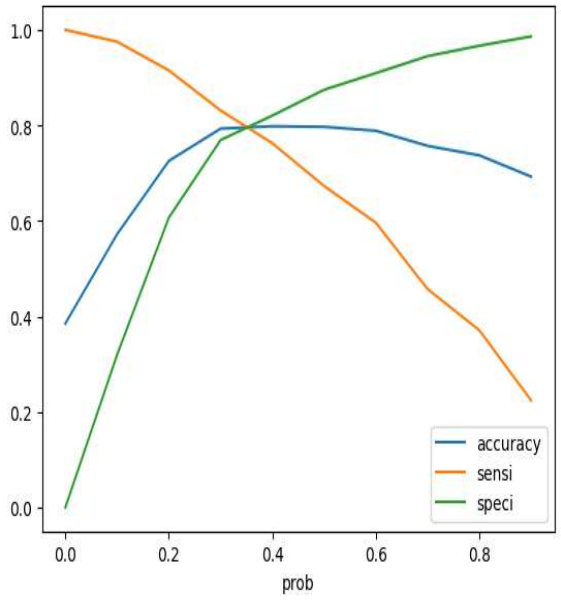
- High correlation observed between:
- **Total visits** and **Conversion Probability**
- **Time spent on platform** and **Conversions**



MODEL EVALUATION

- Logistic regression model performance:
- Accuracy: Highlight metric value
- Sensitivity and specificity: Explain trade-offs
- It is observed that 0.3 can be the optimum cut off probability.





OBSERVATIONS AND INSIGHTS

- Leads motivated by career prospects show higher conversion potential.
- Specializations like Finance, HR, and Marketing exhibit high conversion probabilities.
- Unemployed leads are a promising segment for targeted efforts.

CONCLUSION AND RECOMMENDATIONS

- Make key data fields mandatory to improve quality.
- Focus on leads with higher engagement metrics and unemployed demographics.
- Use personalized emails and SMS for better customer interaction.