

Lead Scoring Assignment

Khanh – Ilan - Anubha

18th March, 2025

Business Objectives

- Develop an accurate lead scoring model
- Prioritize potential leads to optimize sales resources
- Maximize conversion rates through targeted marketing strategies

Overview of Approach

- **Data Preparation:** Handled missing values, outliers, categorical encoding
- **Model Building:** Logistic Regression with Recursive Feature Elimination (RFE)
- **Model Evaluation:** Accuracy, Precision, Recall, Sensitivity, Specificity, ROC Curve
- **Actionable Insights:** Identified key variables impacting conversions

Data Understanding

- We analyze based on the dataset:
 - **Lead.csv**: *all leads data*
 - **Leads Data Dictionary.csv**: data points explanation

Data Understanding

- Binary vars

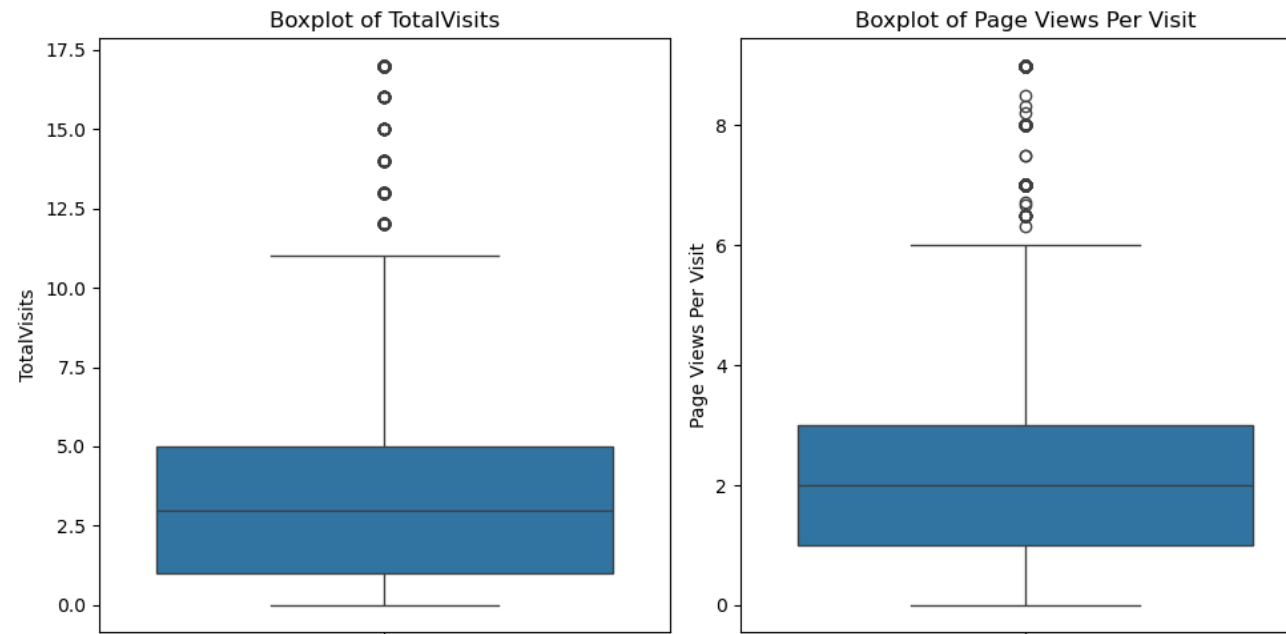
```
# List of binary variables to convert
binary_vars = ['Do Not Email', 'Do Not Call', 'Search', 'Magazine', 'Newspaper Article',
               'X Education Forums', 'Newspaper', 'Digital Advertisement',
               'Through Recommendations', 'Receive More Updates About Our Courses',
               'Update me on Supply Chain Content', 'Get updates on DM Content',
               'I agree to pay the amount through cheque', 'A free copy of Mastering The Interview']
```

- Multi-level vars

```
# List of categorical variables for one-hot encoding with multiple levels
multi_level_vars = ['Lead Origin', 'Lead Source', 'Last Activity', 'Country', 'Specialization',
                   'How did you hear about X Education', 'What is your current occupation',
                   'What matters most to you in choosing a course', 'Tags', 'Lead Quality',
                   'Lead Profile', 'City', 'Asymmetrique Activity Index',
                   'Asymmetrique Profile Index', 'Last Notable Activity']
```

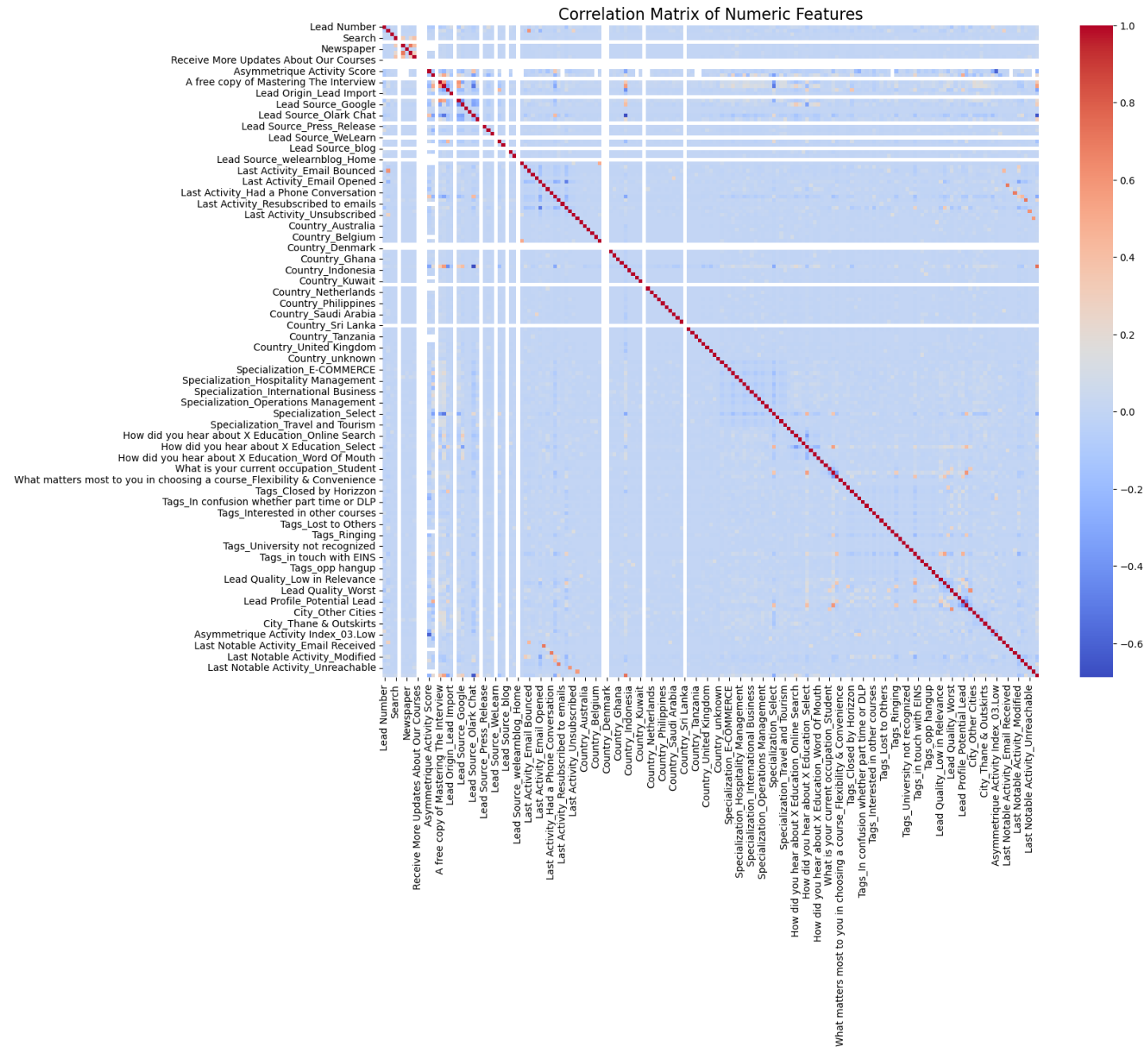
Data Understanding

- Visualize the post-Outliers handling for TotalVisits and Page Views Per Visit



Data Understanding

- Correlation Matrix

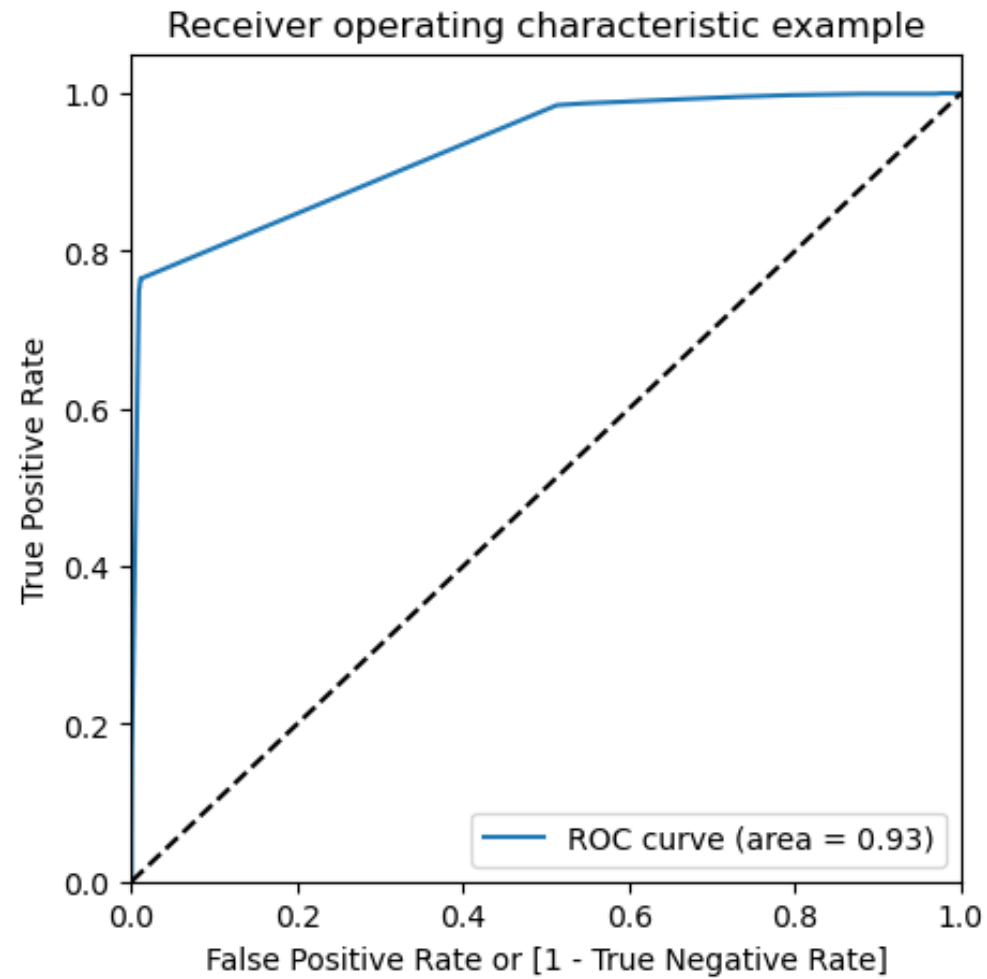


Features Selection

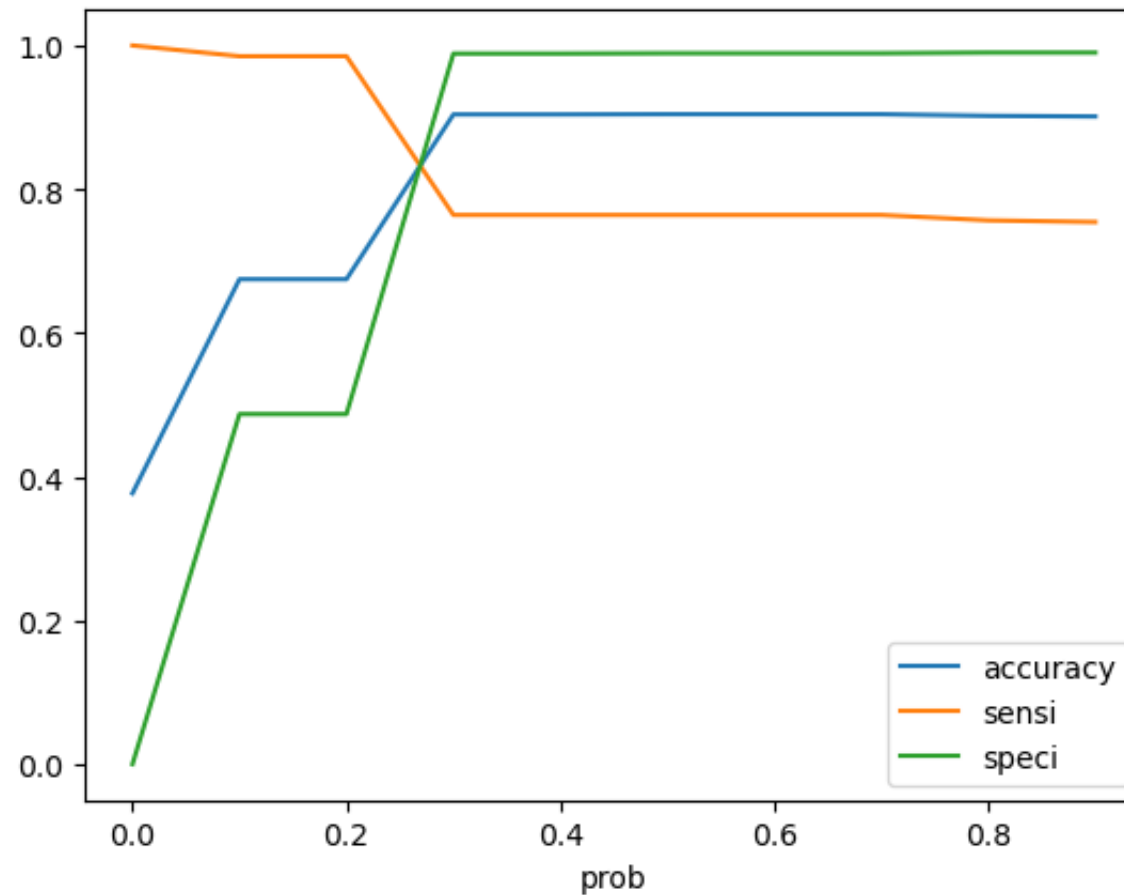
- 10 selected features post-RFE

```
Selected features: Index(['Lead Source_Welingak Website', 'Tags_Closed by Horizzon',  
                        'Tags_Interested in other courses', 'Tags_Lost to EINS', 'Tags_Ringing',  
                        'Tags_Will revert after reading the email', 'Tags_invalid number',  
                        'Tags_switched off', 'Lead Quality_Worst',  
                        'Asymetrique Activity Index_03.Low'],  
                        dtype='object')
```


ROC curve



Plotting accuracy sensitivity and specificity



Key Findings – Technical Insights

- **Model Accuracy:** 88.6% on test data
- **Area Under ROC Curve (AUC):** 0.93, indicating high predictive accuracy
- **Top Predictors for Conversion:**
 - Tags_Lost to EINS
 - Tags_Closed by Horizon
 - Tags_Will revert after reading the email

Business Recommendations (Aggressive Conversion Period)

- **Immediate Contact Strategy:**

- Prioritize leads with conversion probability $> 70\%$
- Focus specifically on high-impact categories (Lost to EINS, Closed by Horizon, Will revert after email)

- **Intern Utilization:**

- Assign clear targets and specific scripts to interns
- Track daily progress and dynamically adjust workloads

Strategic Benefits to X Education

- **Optimized Resource Allocation:** Intern and sales team efficiency
- **Increased Conversion Rates:** Through targeted and prioritized leads
- **Scalable Model:** Adaptable strategies for aggressive sales periods and quieter quarters

Next Steps and Recommendations

- Continuously refine predictive model with new data
- Implement real-time monitoring dashboards for sales performance
- Regularly train sales and marketing teams on insights from predictive analytics

Thank you