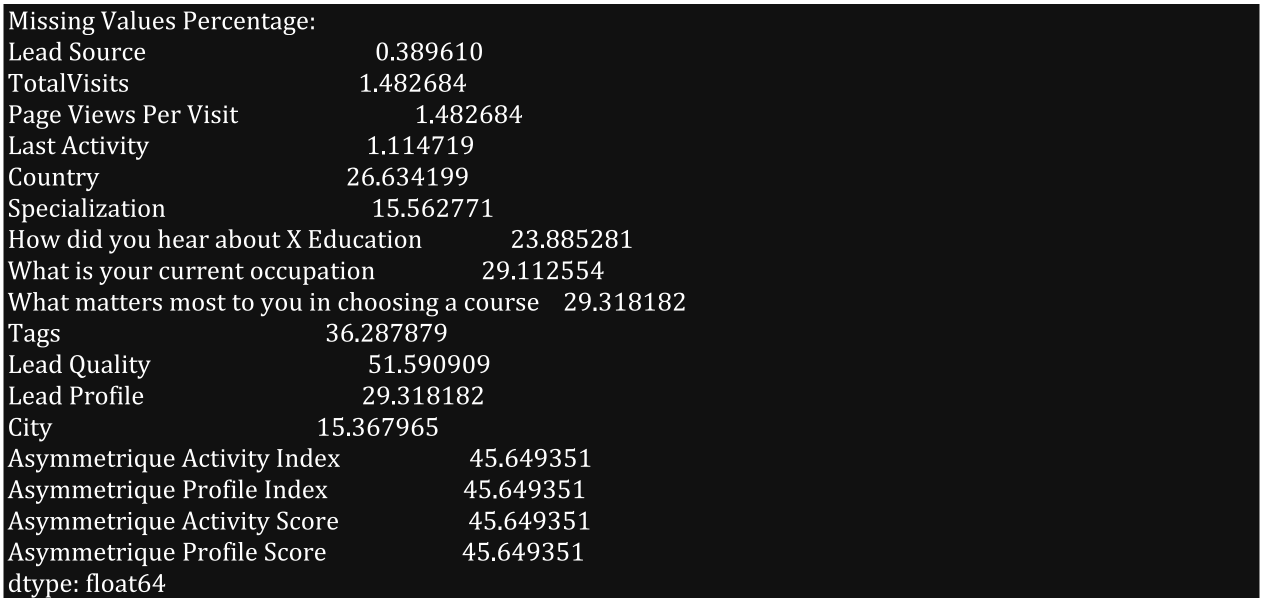
Lead Scoring Assignment - Presentation

1. Introduction

- The purpose of this lead scoring assignment was to optimize lead conversion by identifying high-potential leads.
- We built a logistic regression model to assign a lead score, enabling the sales team to prioritize outreach efforts.

2. Data Overview

- Dataset contained 9,240 leads with 37 features .
- Key variables included `Lead Source`, `Total Time Spent on Website`, `Tags`, and `Last Activity`.
- Missing values were handled using strategic imputations (e.g., median for numerical, 'Unknown' for categorical).



```
Missing Values Percentage:
Lead Source                                  0.389610
TotalVisits                                  1.482684
Page Views Per Visit                         1.482684
Last Activity                                1.114719
Country                                     26.634199
Specialization                              15.562771
How did you hear about X Education          23.885281
What is your current occupation             29.112554
What matters most to you in choosing a course    29.318182
Tags                                        36.287879
Lead Quality                                51.590909
Lead Profile                                29.318182
City                                        15.367965
Asymmetrique Activity Index                 45.649351
Asymmetrique Profile Index                  45.649351
Asymmetrique Activity Score                 45.649351
Asymmetrique Profile Score                  45.649351
dtype: float64
```

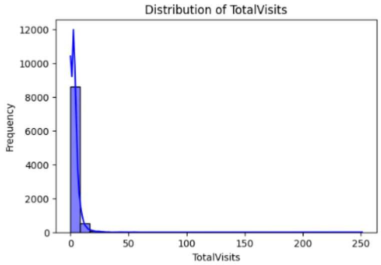
3. Exploratory Data Analysis

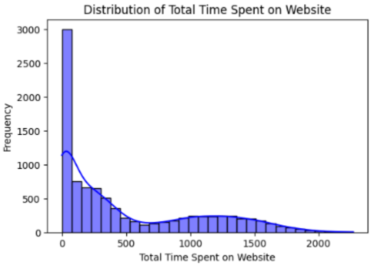
- **Univariate Analysis:**

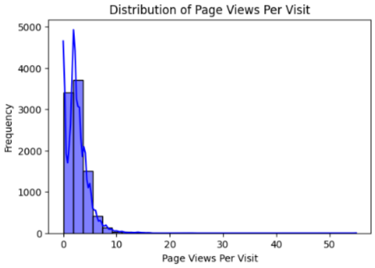
- Numerical Features (Histograms & Box Plots)
  - Total Visits is highly skewed, with a majority of users having very few visits.
  - Total Time Spent on Website shows bimodal behavior, indicating distinct user groups.
  - Page Views Per Visit is right-skewed, with a few users having exceptionally high values.
  - Box plots reveal that all three numerical features have outliers, especially in Total Visits.
- Categorical Features (Bar Charts & Conversion Rate)
  - Lead Source:
    - Google, Direct Traffic, and Olark Chat contribute the highest number of leads.

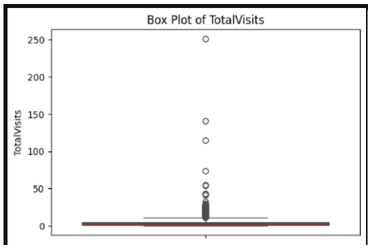
- Reference & Click2Call show the highest conversion rates.

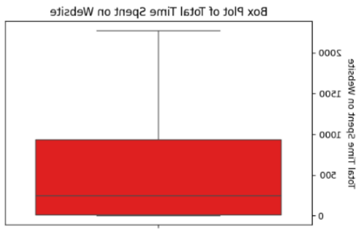
- Last Activity:
  - Email Opened and SMS Sent dominate activity types.
  - High-converting activities include SMS Sent and Direct Conversations.
- Specialization:
  - Finance, HR, and Marketing Management have the highest leads.
  - Most specializations have conversion rates around 40-50%.
- City:
  - Mumbai and “Select” category have the highest number of leads.
  - Conversion rates are fairly even across cities, with Tier II Cities showing a lower rate.

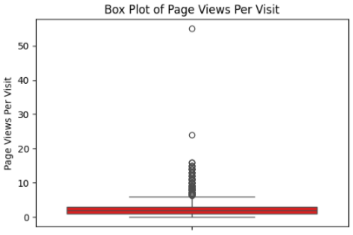


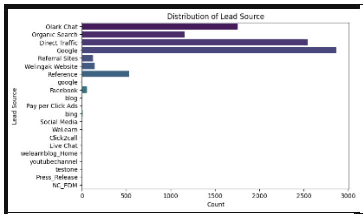


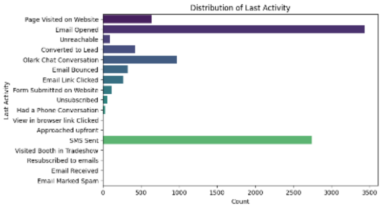


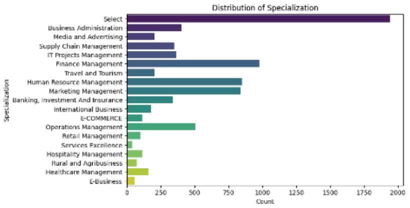


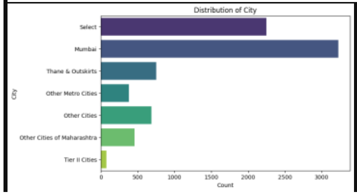


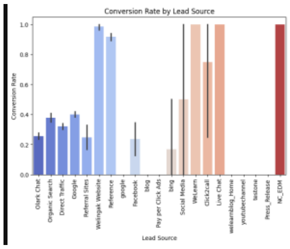


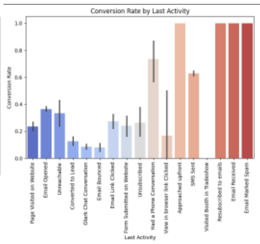


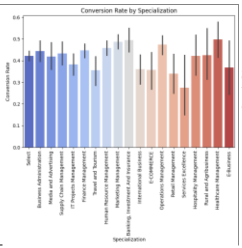


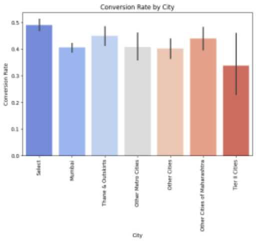




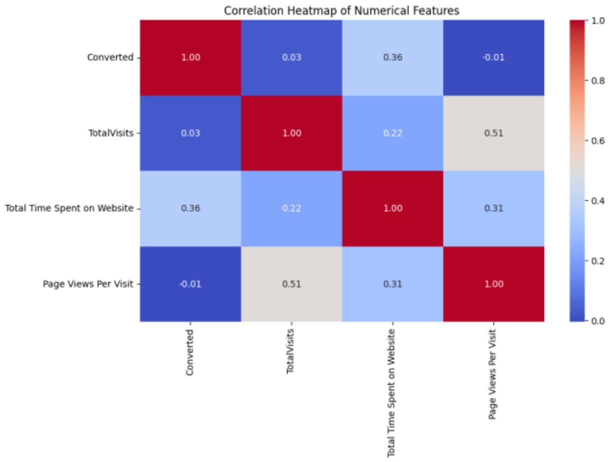








- **Correlation Analysis:**
    - Total Time Spent on Website has a moderate positive correlation (0.36) with conversion.
    - Other numerical features (Total Visits, Page Views Per Visit) have weak correlations with conversion.
    - Lead Number is just an identifier and has no predictive value.



4. **Model Development**

 - The dataset was preprocessed by encoding categorical variables and scaling numerical features.

 - Logistic regression was trained, and Recursive Feature Elimination (RFE) was used to select the top 15 predictive features.

5. **Model Performance**

Metric	Initial Model	Final Model (After Feature Selection)
Accuracy	79.06%	79.06%
Precision	75.04%	75.12%
Recall	68.40%	68.26%
F1 Score	71.57%	71.52%

6. **Key Insights**

 - Top 3 Features Contributing to Lead Conversion:
    1. `Total Time Spent on Website`
    2. `Tags`
    3. `What is your current occupation`
 - Top 3 Categorical Variables to Focus On:
    1. `Tags`
    2. `Lead Source`
    3. `Last Activity`

7. **Business Strategies**

For Aggressive Lead Conversion (Intern Hiring Period):

- Prioritize high-engagement leads first.
- Use automated outreach for lower-priority leads.
- Assign structured lead lists for efficient intern call handling.

For Minimizing Calls When Targets Are Met:

- Focus calls only on leads with >85% conversion probability .
- Use automated emails/SMS for low-probability leads.
- Shift sales team focus to new initiatives like upselling.