Lead Scoring Case Study

Introduction

- **Company Background:** X Education sells online courses to industry professionals.
- **Marketing Channels:** Google, websites, referrals.
- **Lead Acquisition:** Forms, videos, referrals.
- **Conversion Rate:** 30%

Business Goals

- **Objective:** Identify "Hot Leads.
- **Model Requirement:** Assign lead scores
- indicating conversion probability.
- **CEO's Target:** 80% conversion rate.

Overall Approach

1. Data Cleaning and Imputing Missing Values
2. Exploratory Data Analysis
3. Feature Scaling and Dummy Variable Creation
4. Logistic Regression Model Building
5. Model Evaluation
6. Conclusion and Recommendation

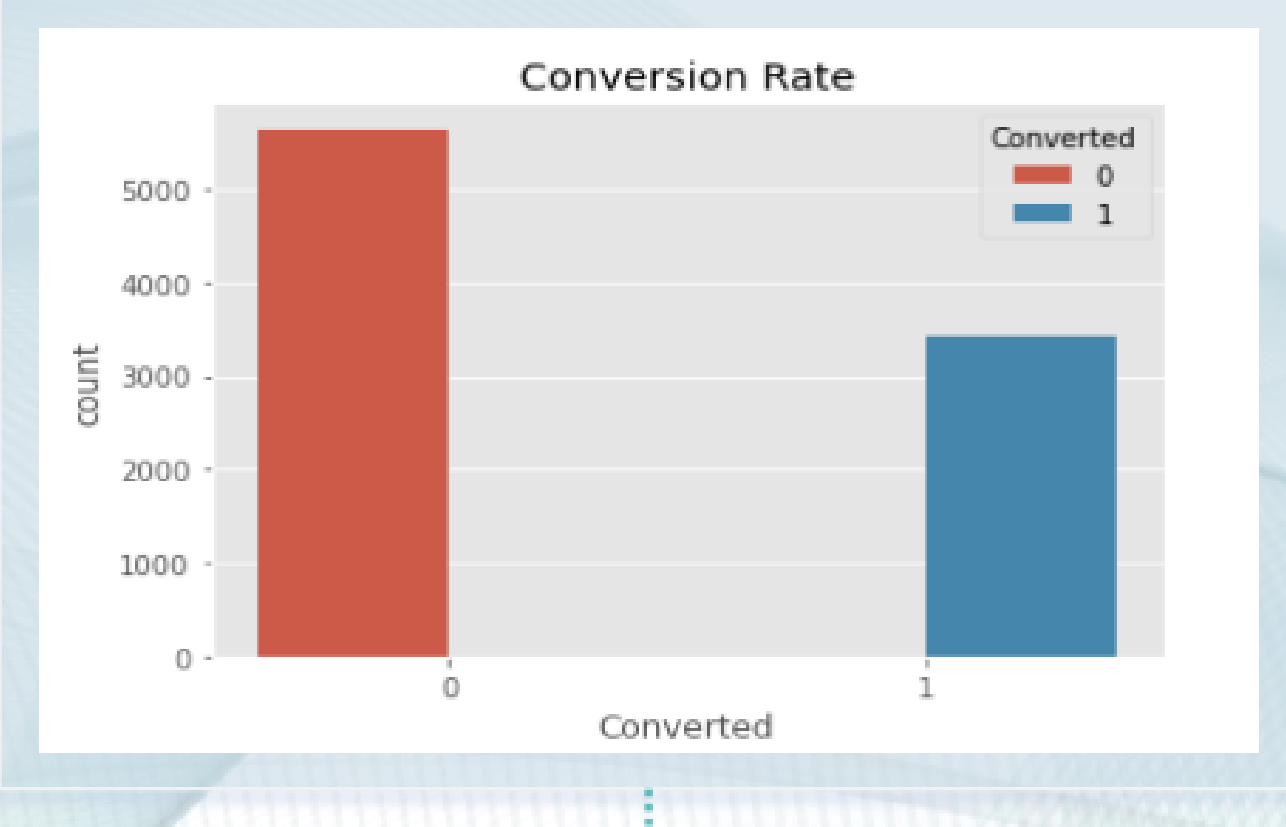
Data Cleaning and Preparation

- ❑ Read data from source.
- ❑ Clean and format data.
- ❑ Remove duplicates.
- ❑ Outlier treatment.
- ❑ Perform EDA.

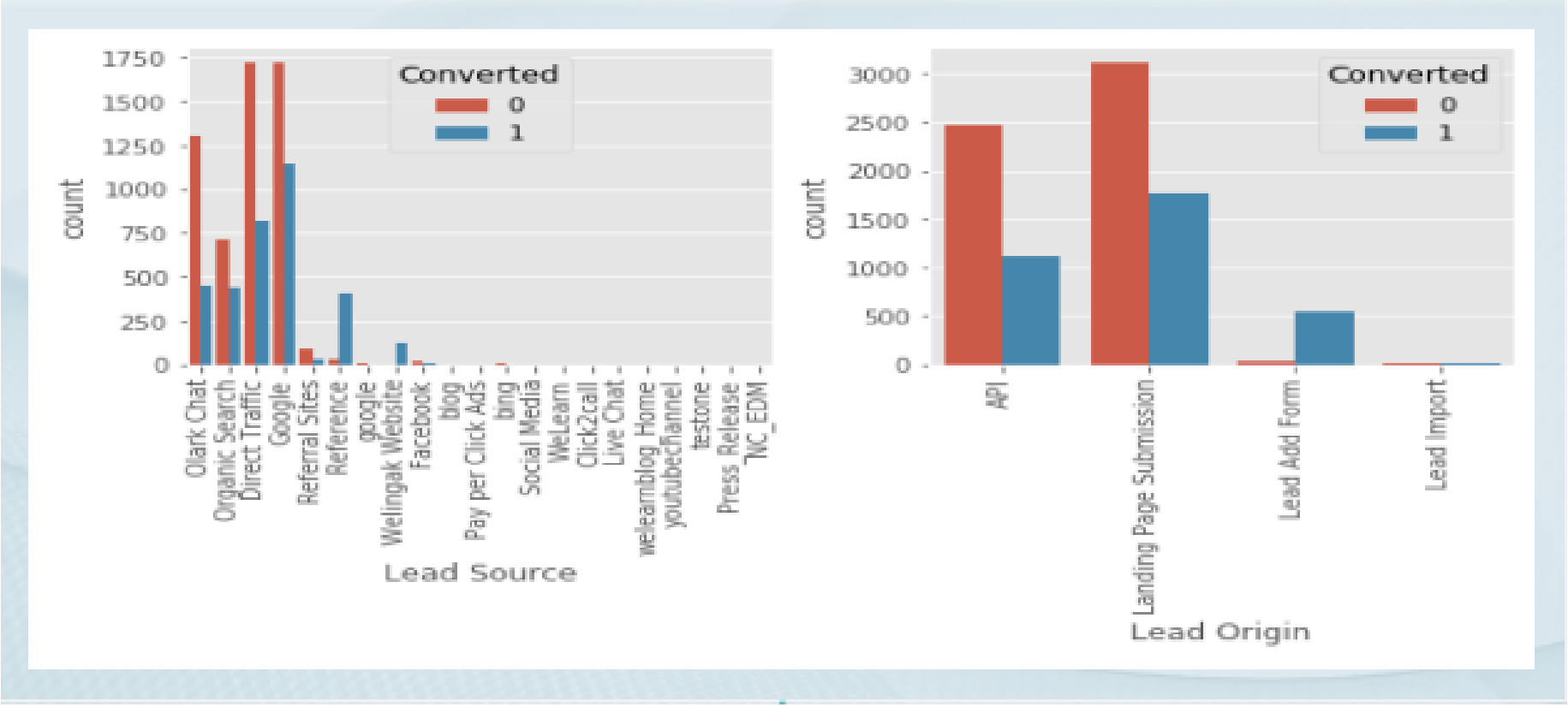
Data Conversion

- ❑ Convert Yes/No values to 1/0.
- ❑ Handle 'Select' values with NaNs.
- ❑ Drop columns with >70% null values.
- ❑ Remove unnecessary columns.
- ❑ Drop rows with <2% null values.

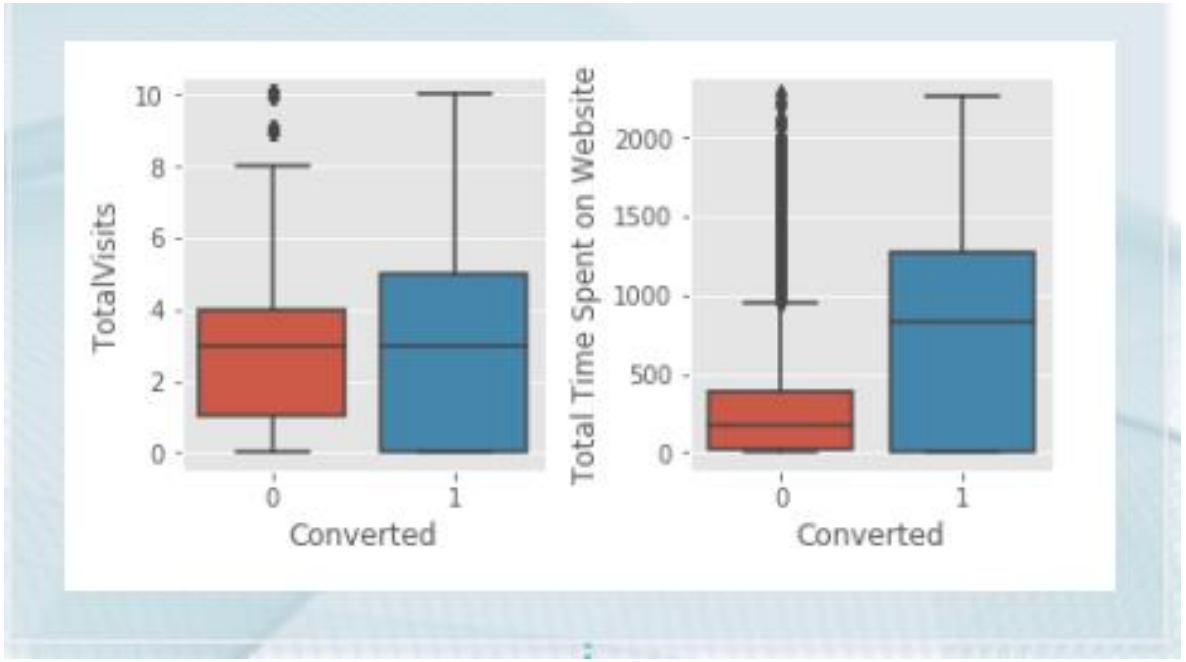
Exploratory Data Analysis (EDA)

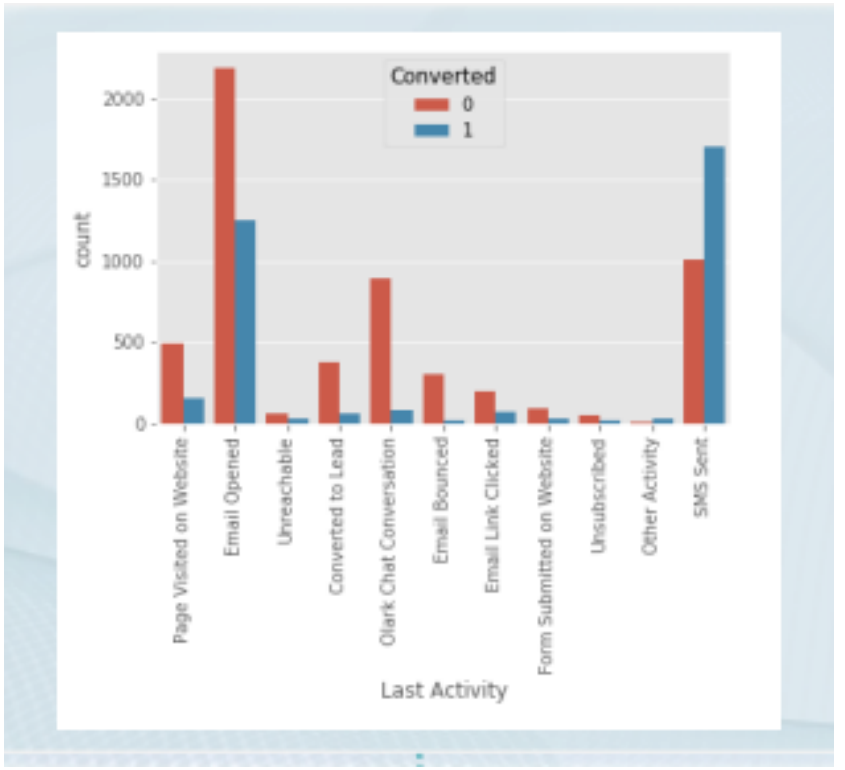


- Conversion rate: 30%.



- ❑ Maximum leads from Google and Direct Traffic.
- ❑ Highest conversion rate from References and Welingak Website.
- ❑ Low count but high conversion rate from Lead Add Form.





- Users spending more time are likely to convert.
- 'Email Opened' has maximum lead activity.
- Highest conversion rate for SMS sent as last activity.
- No conclusive inference for Specialization.

Model Building

- ❑Split data into train (70%) and test (30%) sets.
- ❑Feature selection using RFE, VIF, and p-value.
- ❑Logistic Regression model building.
- ❑Test dataset predictions.
- ❑Achieved 92% accuracy.

Model Evaluation

- **Metrics:**
  - Accuracy, Sensitivity, Specificity for various cutoffs (0.1 to 0.9).
  - Optimal cutoff: 0.27.
- **Train Data Confusion Matrix:**
  - Accuracy: 83.59%
  - Precision: 71.6%
  - Sensitivity: 94.9%
  - Specificity: 76.5%

Conclusion

- **Summary:**
  - Logistic regression model predicts conversion probability.
  - Model performance: 92% conversion rate on test data, 95% on train data.
  - Key contributing variables: Tags and Lead Quality.
  - Model aligns with business requirements and is adaptable for future needs.