



# Lead Score Case Study



- Swati Mhaswadkar
- 

# Problem Statement

---

- ▶ X Education provides online courses tailored for industry professionals.
- ▶ Despite acquiring a large number of leads, the company's lead conversion rate remains low—only about 30 out of every 100 leads convert into customers.
- ▶ To enhance efficiency, X Education aims to identify high-potential leads, referred to as '**Hot Leads**', who have a higher likelihood of conversion.
- ▶ By accurately identifying these leads, the company can improve its conversion rate, allowing the sales team to focus their efforts on the most promising prospects instead of reaching out to every lead indiscriminately.

# Business Objective

---

- ▶ X Education aims to identify the **most promising leads**—those with the highest likelihood of converting into paying customers.
- ▶ To achieve this, they seek to develop a **predictive model** that can analyze lead data and accurately distinguish **high-potential leads** (Hot Leads) from less promising ones. This model will help streamline sales efforts by prioritizing leads that are more likely to convert.
- ▶ Once the model is successfully built and validated, X Education plans to **deploy it into their sales pipeline**, ensuring it is seamlessly integrated for future use. This will enable continuous lead scoring, allowing the sales team to make data-driven decisions and improve overall conversion rates over time.

# Problem Approach

## Data Preprocessing & Model Development

- ✓ Data Cleaning & EDA – Remove duplicates, handle missing values, detect outliers, and analyze correlations for feature selection.
- ✓ Feature Engineering – Standardize numerical data and encode categorical variables for model compatibility.
- ✓ Model Development & Validation – Use Logistic Regression to predict conversions, optimizing accuracy, precision, recall, and ROC Curve Analysis.
- ✓ Deployment & Impact – Enable real-world lead prioritization, improving sales efficiency and conversion rates.

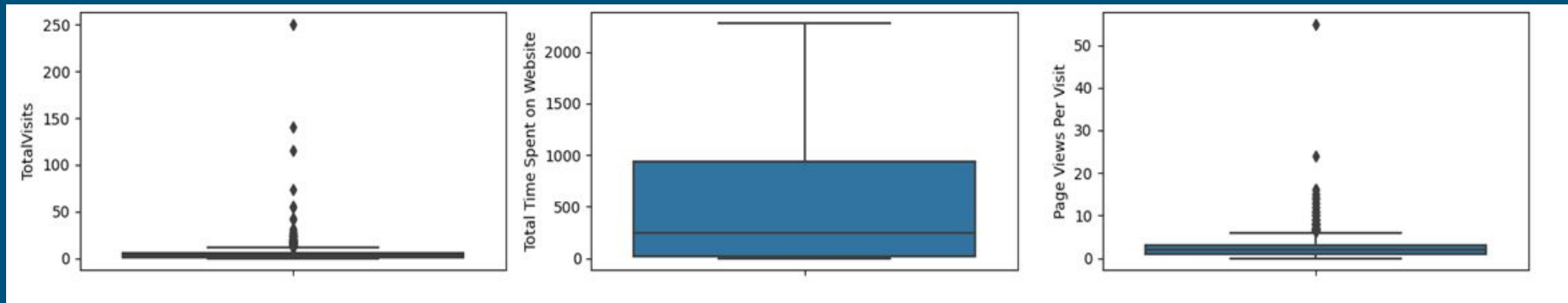
This approach enhances X Education's lead scoring strategy, maximizing sales potential.

# Data Manipulation Summary

- ✓ Dataset Overview: 37 rows, 9,240 columns.
- ✓ Dropped Irrelevant Features: Single-value columns (Magazine, Updates, Content Preferences), unique IDs (Prospect ID, Lead Number), and low-variance features (Do Not Call, Newspaper, Digital Ads).
- ✓ Handled Missing Data: Removed columns with >35% missing values (How did you hear about X Education, Lead Profile).

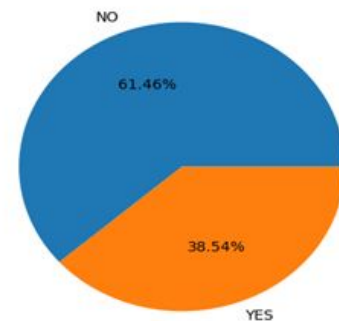
This cleanup enhances data quality, reduces redundancy, and improves model efficiency.

# Performing EDA

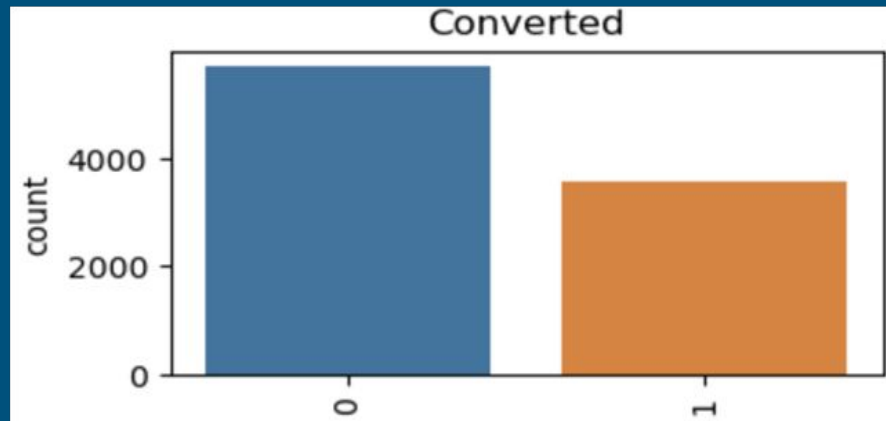
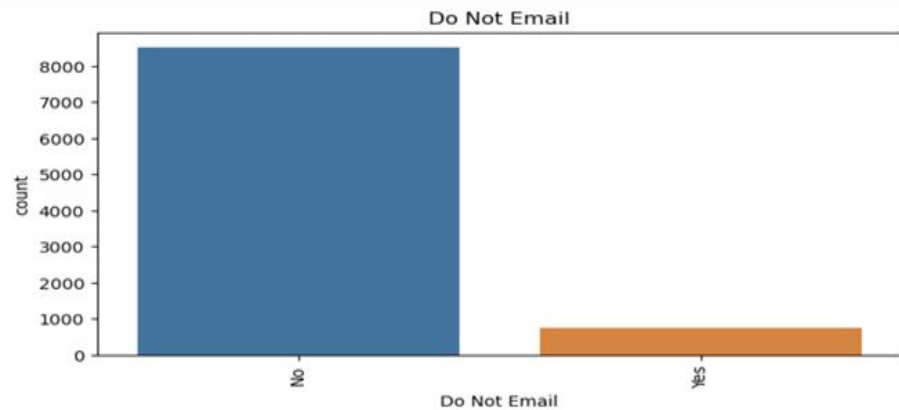
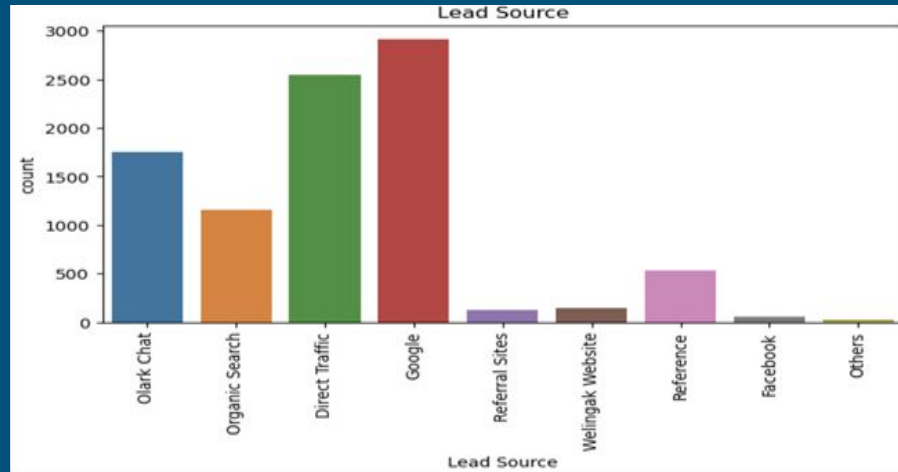
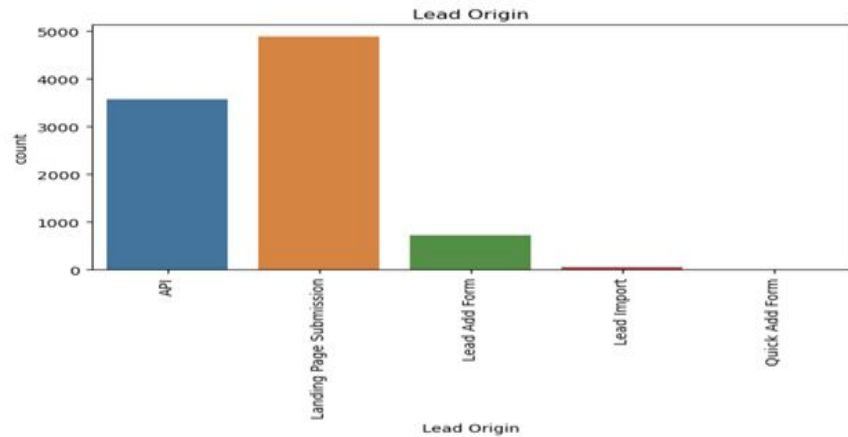


- Assess data distribution for imbalance.
- Perform univariate analysis to understand individual feature distributions.
- Conduct bivariate analysis to examine relationships between variables.
- Analyze multivariate correlations using coefficient metrics.

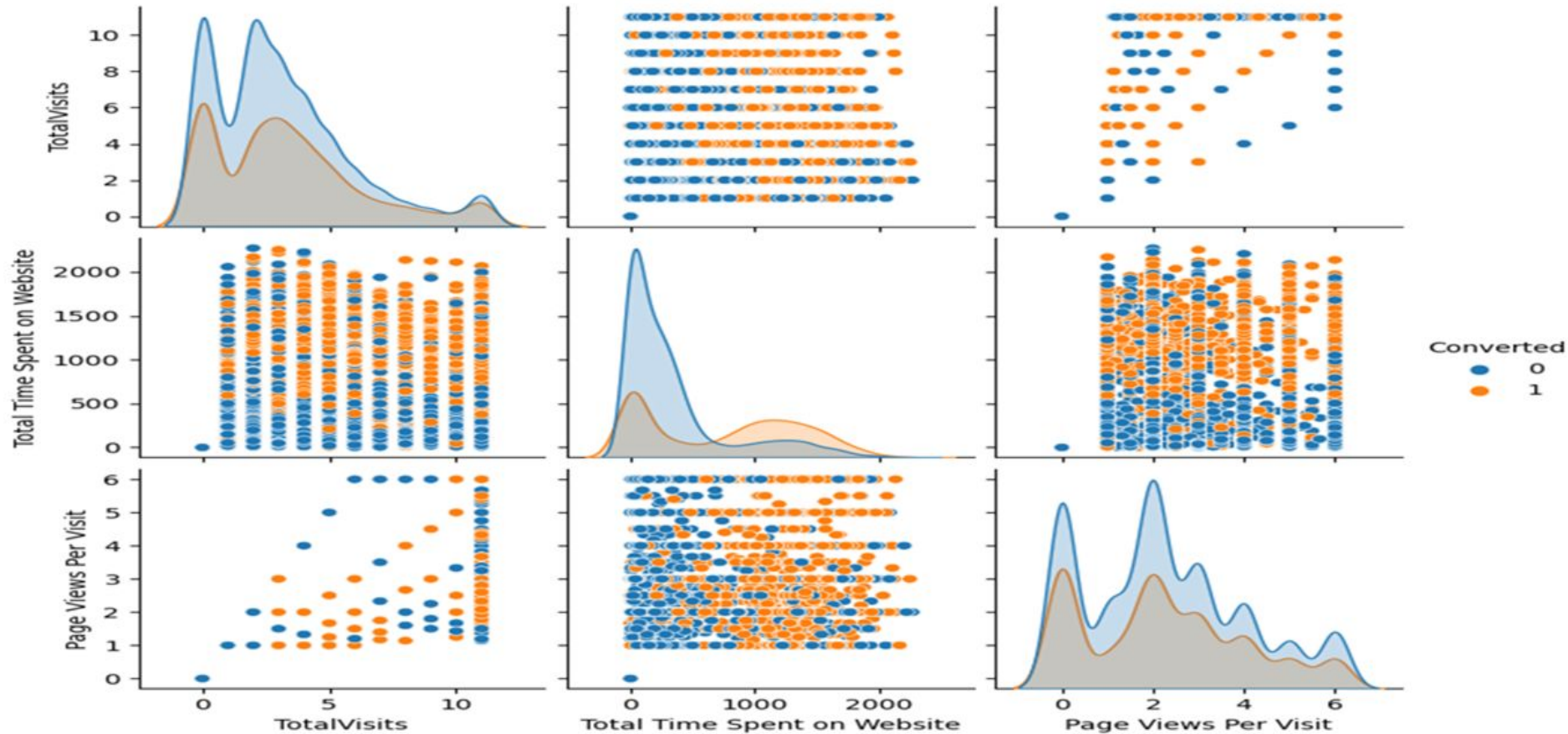
Distribution of Converted Leads



# Univariate Analysis

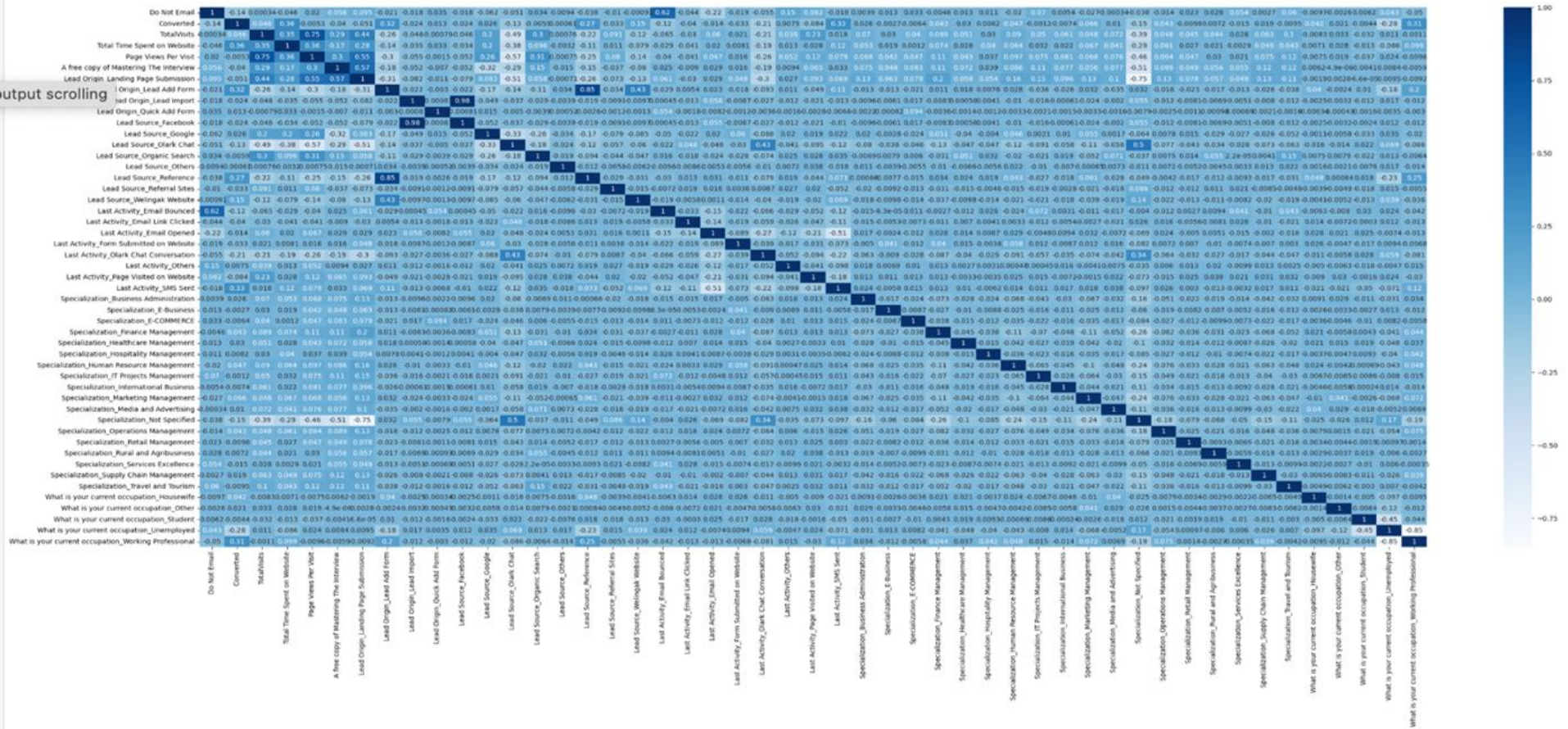


# Bivariate Analysis





output scrolling	Origin, Lead
	Lead Origin, Lead



# Insights & Conclusion Derived:

Key factors influencing potential buyers:

- **Website Engagement:** Higher total time spent on the website.
- **Occupation:** Leads who are currently working professionals.
- **Lead Source:** Higher conversion likelihood when the lead source is:
  - Welingak website
  - Reference
  - Landing Page Submission
- **Recent Activity:** Leads who engaged through:
  - SMS interactions
  - Olark chat conversations

# Recommendations/Suggestions:

- Prioritize features with **positive impact** on conversions.
- Increase advertising spend on the **Welingak Website** to attract more leads.
- Utilize **targeted messaging** to engage working professionals.
- Offer **incentives or discounts** for successful referrals to encourage more lead conversions.
- **Aggressively target working professionals**, as they have a high conversion rate.
- Develop strategies to **attract high-quality leads** from top-performing sources.
- Assess features with **negative impact** to refine the approach.
- **Optimize the landing page submission process** to improve lead capture efficiency.