



# Lead Scoring Case Study

***Submitted By***

*Abhishek Rathore*

*Abhisek Kundu*

*Abhradeep Chandra Paul*

# Problem Statement

---

- **Company:** X Education sells online courses to industry professionals.
- **Process:** Interested professionals fill out a form on the website and become leads.
- **Current Conversion Rate:** 30% (only 30 out of 100 leads convert).
- **Goal:** Identify "Hot Leads" to improve efficiency.
- **Objective:**
  - Build a model to score each lead between 0-100.
  - Focus sales efforts on high-potential leads.
  - Increase conversion rate to 80%.
- **Requirements:**
  - Handle future constraints.
  - Optimize peak time actions.
  - Utilize full manpower efficiently.
  - Adapt post-target achievement strategies.
- Making targeted efforts should boost conversion and streamline the sales process!

# Approach

---

- 1. Data Preparation**
- 2. Exploratory Data Analysis (EDA)**
- 3. Dummy Variable Creation**
- 4. Train-Test Split**
- 5. Feature Scaling**
- 6. Correlation Analysis**
- 7. Model Building**
  1. Recursive Feature Elimination (RFE)
  2. R-squared, VIF, and p-values
- 8. Model Evaluation**
- 9. Threshold Tuning**
- 10. Precision and Recall Evaluation**
- 11. Making Predictions on Test Set**

# Data Preparation

- **Data Insights**
- **Number of Rows and Columns:** 9,240 rows, 37 columns
- **Duplicates:** No duplicate rows found
- **Data Types:** 7 numeric columns, 30 non-numeric columns
- **Null Handling:** 'Select' values were converted to null as per the problem statement

## Cleaning of categorical data

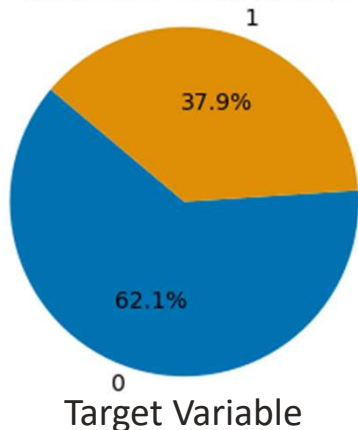
- **Columns Deleted ( > 40% Null Data)**
  - Lead Quality
  - Lead Profile
  - Asymmetrique Activity Index
  - Asymmetrique Profile Index
  - How did you hear about X Education
- **Columns Dropped (Insignificant Data):**
  - What matters most to you in choosing a course
  - Country
- **Null Data is filled with “Other”**
  - What is your current occupation
  - Specialization
- **Null data filled with Major category**
  - City
  - Tags

## Cleaning of Numerical Data

- **Columns Deleted ( > 40% Null Data):**
  - Asymmetrique Activity Score
  - Asymmetrique Profile Score
- **Handling Null Rows:**
  - Null rows were deleted where data is relatively small for:
    - TotalVisits
    - Page Views Per Visit

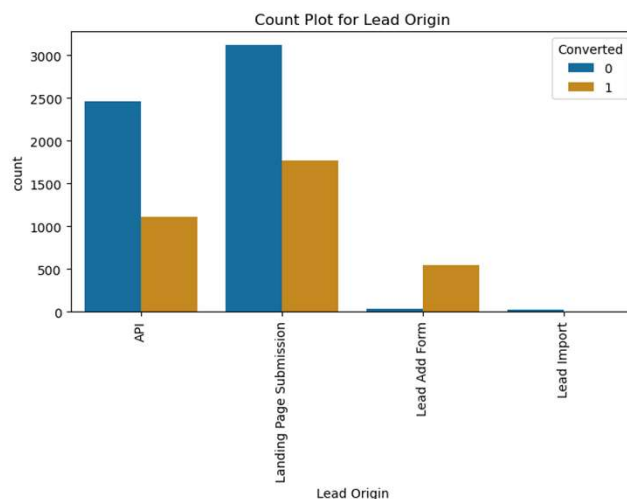
# EDA

Converted Value Counts

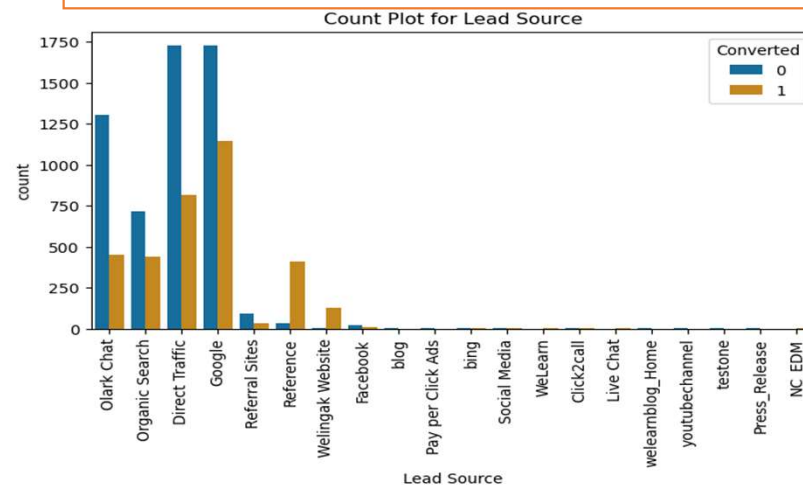


## Non Numeric Data Bivariate Analysis: Major Highlights

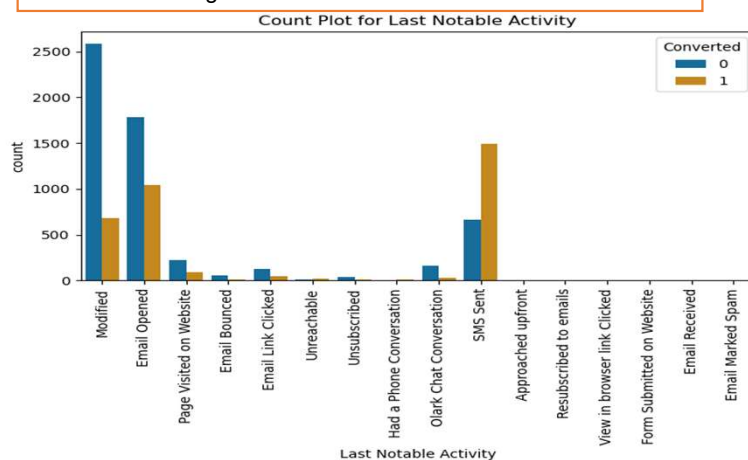
API and Landing Page Submissions have 30-35% conversion rates with high lead counts



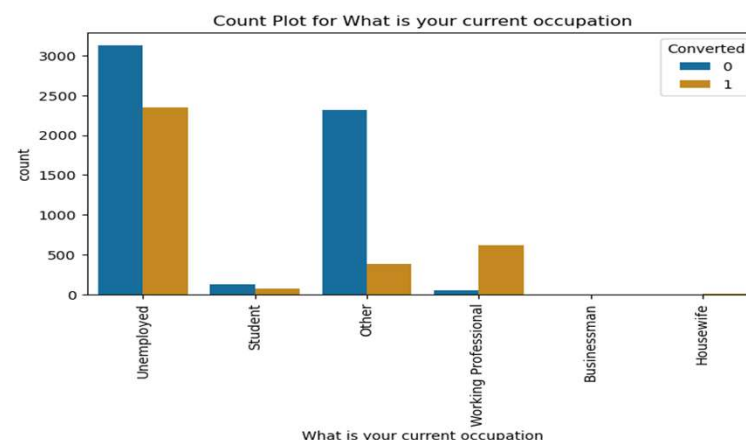
Improve conversions for Olark Chat, Organic Search, Direct Traffic, and Google leads. Generate more leads from Reference and Welingak Website.



SMS Sent has high conversion



Working Professionals going for the course have high chances of joining it

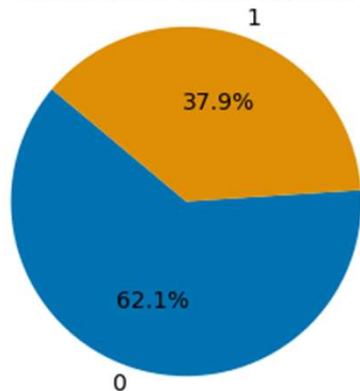


Columns Removed since no inference can be drawn

- Tags
- 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

# EDA

Converted Value Counts



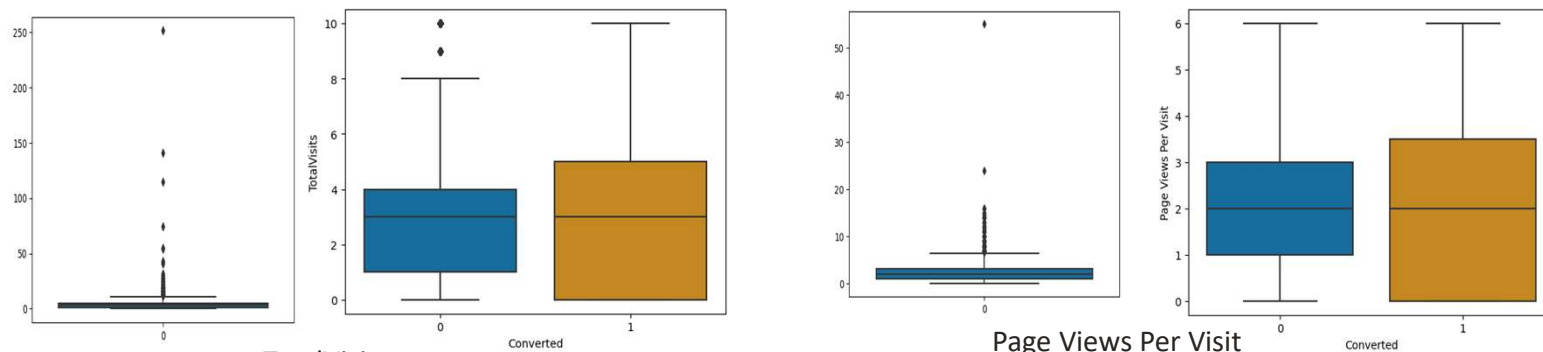
Target Variable

**No conclusion on Data**

Medians values were same

- TotalVisits
- Lead Number
- Page Views Per Visit

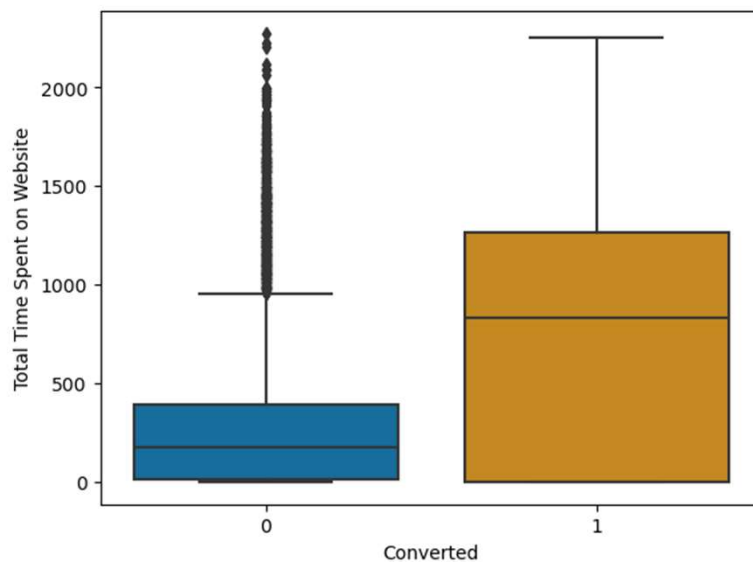
## Numeric Data Bivariate Analysis



Removing Outliers

Leads converted more when spending more time on website. Make site engaging to boost conversions

Numeric columns are correlated. Total visit has high correlation with page views per visit.



# Model Evaluation

## Final Model after RFE, Pvalue and VIF analysis

P=0

	coef	std err	z	P> z
const	-0.9327	0.137	-6.822	0.000
Do Not Email	-1.5291	0.178	-8.599	0.000
Total Time Spent on Website	4.5421	0.168	27.018	0.000
Lead Origin_Landing Page Submission	-1.0650	0.129	-8.259	0.000
Lead Source_Olark Chat	1.2221	0.125	9.739	0.000
Lead Source_Reference	3.1217	0.241	12.946	0.000
Lead Source_Welingak Website	5.4986	0.728	7.553	0.000
Last Activity_Had a Phone Conversation	2.7069	0.761	3.558	0.000
Last Activity_Olark Chat Conversation	-0.9284	0.174	-5.333	0.000
Last Activity_SMS Sent	1.2612	0.076	16.511	0.000
Specialization_Others	-0.9829	0.127	-7.734	0.000
What is your current occupation_Other	-1.1395	0.089	-12.750	0.000
What is your current occupation_Working Professional	2.3689	0.193	12.243	0.000
Last Notable Activity_Modified	-0.7891	0.083	-9.533	0.000
Last Notable Activity_Unreachable	1.7622	0.497	3.546	0.000

### Confusion Matrix:

[[3467 438]

[ 724 1722]]

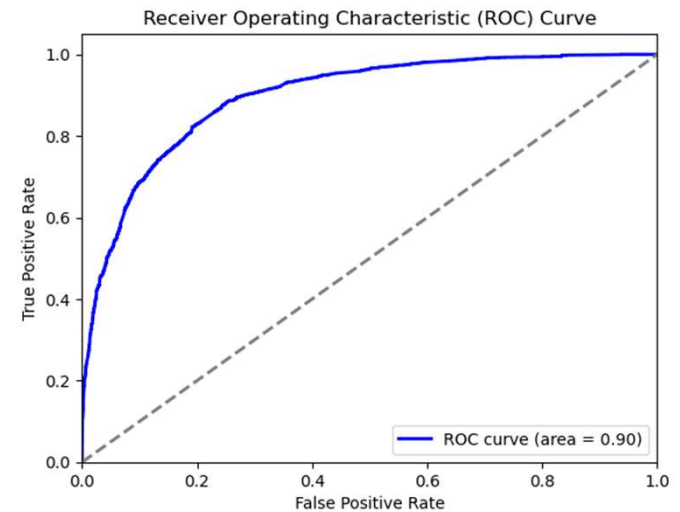
**Sensitivity:** 0.7040065412919051

**Specificity:** 0.8878361075544174

**False Positive Rate:** 0.11216389244558259

**Positive Predictive Value:** 0.7972222222222223

**Accuracy:** 0.8170366871358841



Given our model's high ROC AUC score of 0.90, it demonstrates excellent performance and reliability.



# Model Evaluation

**Threshold Tuning:** Experiment with different probability thresholds to find the best balance between sensitivity and specificity

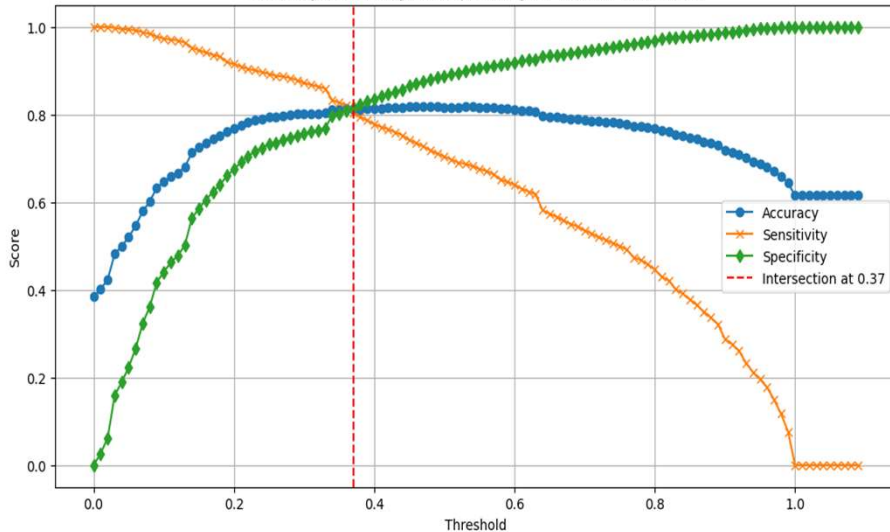
Intersection Threshold: 0.37

Accuracy at Intersection: 0.8110533774208786

Sensitivity at Intersection: 0.8045789043336059

Specificity at Intersection: 0.8151088348271447

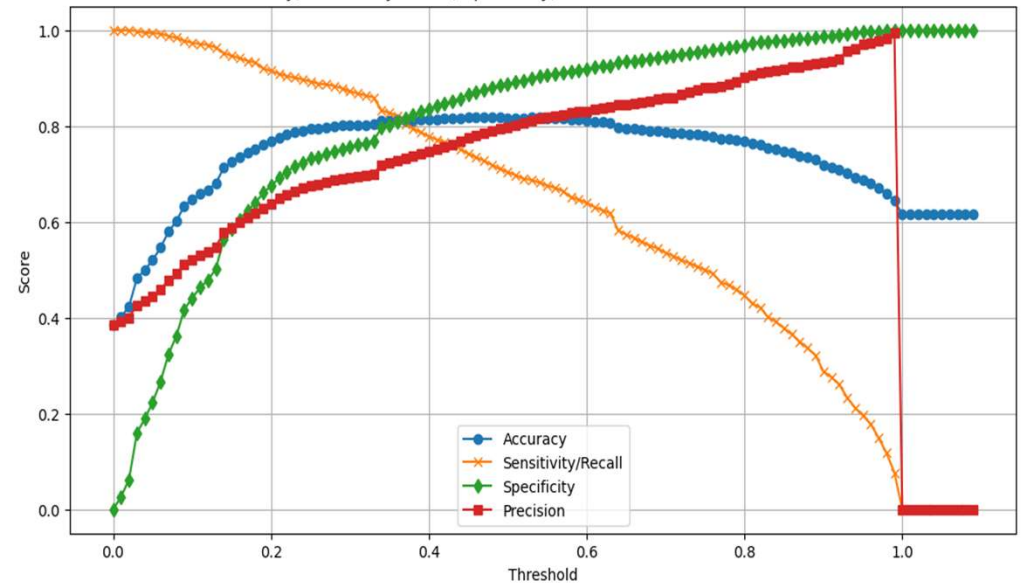
Accuracy, Sensitivity, and Specificity for Various Thresholds



**Precision and recall tradeoff :**

The Precision-Recall trade-off reduced True Positives and increased True Negatives. For better recall and increased True Positives, we chose 0.3 as the optimal cutoff point.

Accuracy, Sensitivity/Recall, Specificity, and Precision for Various Thresholds





# Conclusion

## Train Data:

*Accuracy : 81.1%*

*Sensitivity: 80.5%*

*Specificity: 81.5%*

## Test Data:

*Accuracy : 80.9%*

*Sensitivity: 79.1%*

*Specificity : 81.9%*

## Recommendations based on Prediction

### Call Priority Leads:

- Leads from "Welingak Websites" and "Reference" sources are highly likely to convert.
- Working professionals show a higher conversion likelihood.
- Leads who spend more time on the website are promising candidates.
- Leads from "Olark Chat" are more likely to convert.
- Leads whose last activity was SMS Sent are promising.
- Leads having Phone conversion are promising candidates

### Avoid Calling:

- Leads whose last activity was "Olark Chat Conversation" are less likely to convert.
- Leads from "Landing Page Submission" origin are not likely candidates.
- Leads with "Specialization: Others" show lower conversion rates.
- Leads who selected "Do not Email" as "yes" are less likely to convert.

