

# Lead Scoring Case Study

By:-

Sachin

Harikrishnan

Gitika

# Problem Statement and Business Objective



People searching for courses, landing up in X Education Course website and filling up a form using email and phone number for an enquiry about the course details are classified as a lead.



Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. The typical lead conversion rate at X education is around 30%.



To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.



Build a model wherein need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.



Ballpark of the target lead conversion rate to be around 80%.

# #Approach / Steps

Importing Data, Understanding the data

Data Preparation ( Handling Null values)

EDA (Univariate Analysis , Checking Outliers, Data Imbalance

Dummy Variable creation

Train-test split

Feature scaling

Looking at Correlation

First Model building using RFE and later finetuning by P values and VIF

Finding Optimum cut off using ROC and specificity and sensitivity

Final Model

Model evaluation

Conclusion and Insights

# Data cleaning and Imputing

Conversion of categorical variables have a level called 'Select' to null value because it is as good as a null value.

Removed the columns which were having more than 70% of null values.

Filling the null values in columns with 'Unknown' for categorical columns.

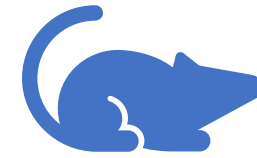
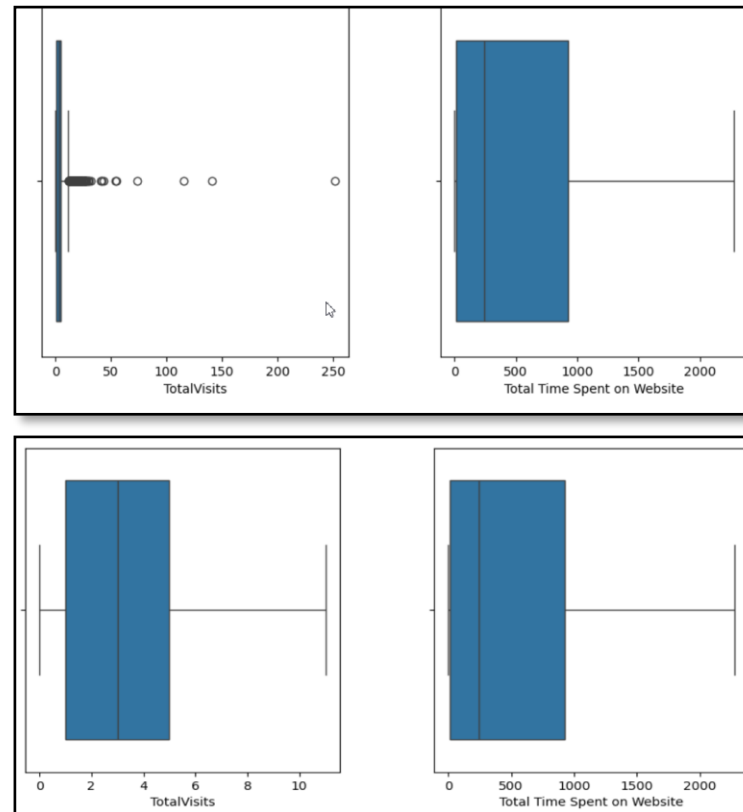
Converting the appropriate data types of columns for better analysis.

Dropping the irrelevant columns from the data set.

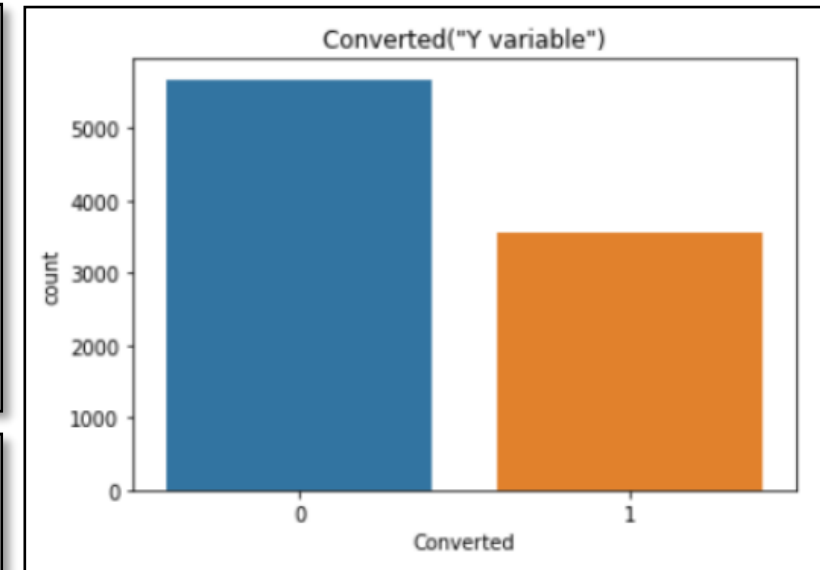
# Handling Outliers, Data Imbalance



TotalVisits and Total Time Spent on Website columns has an outlier, which has been handled.



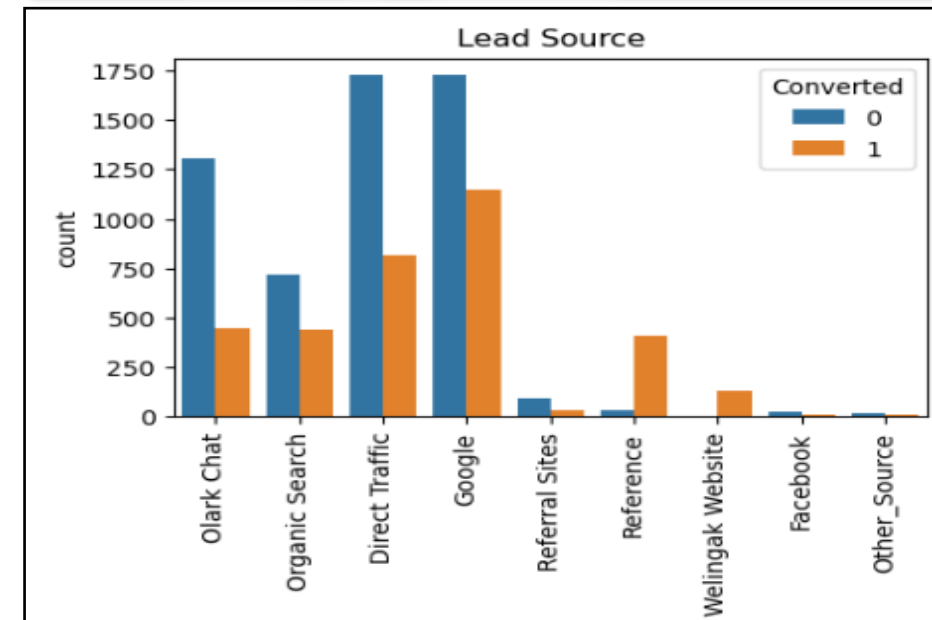
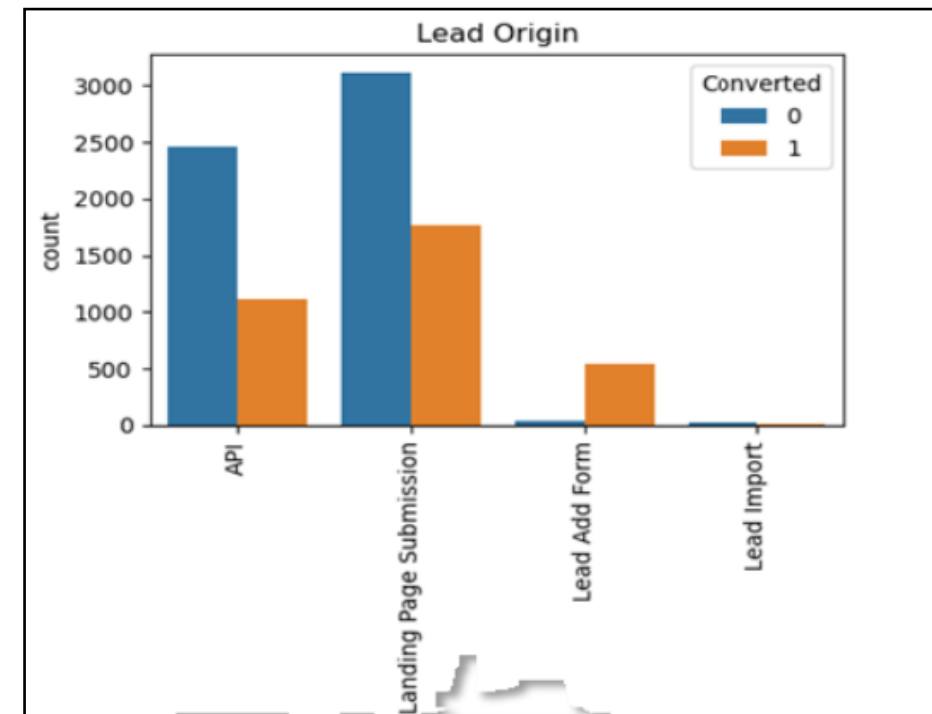
No data imbalance found on target variable.



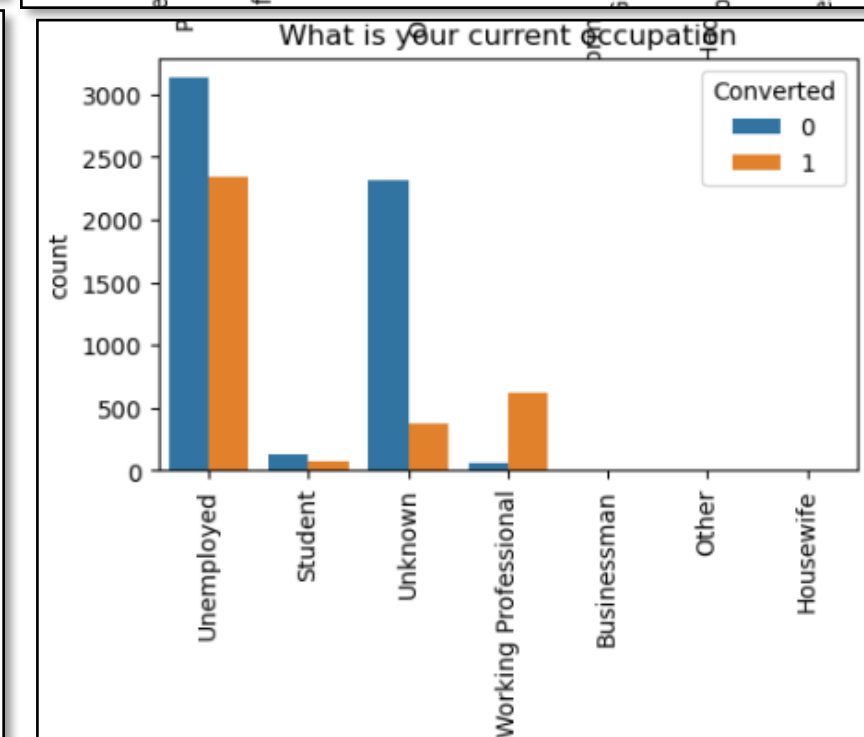
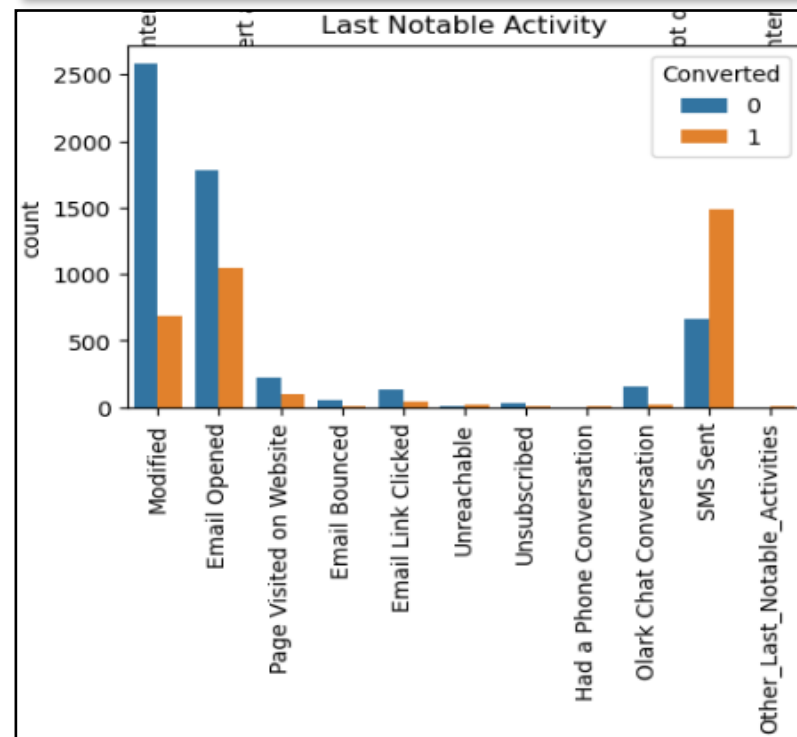
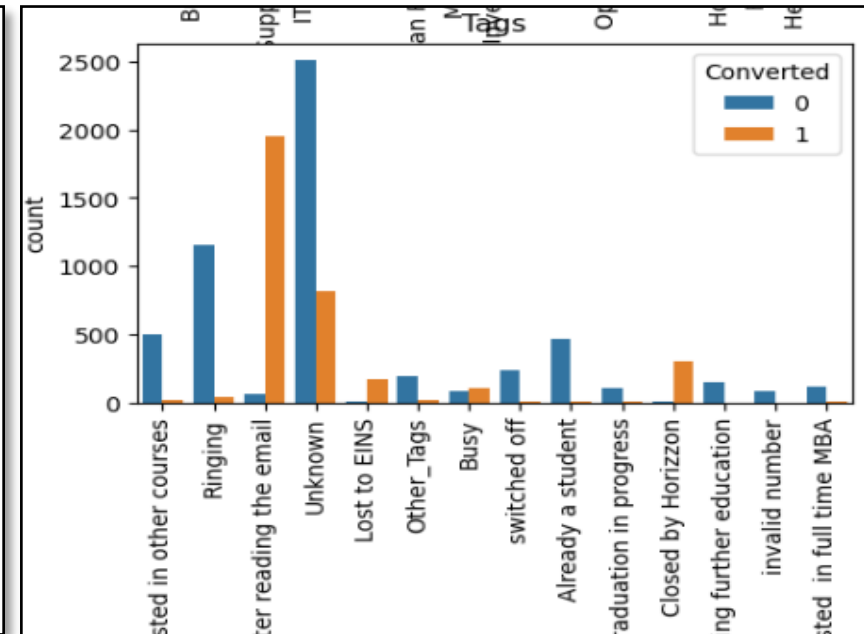
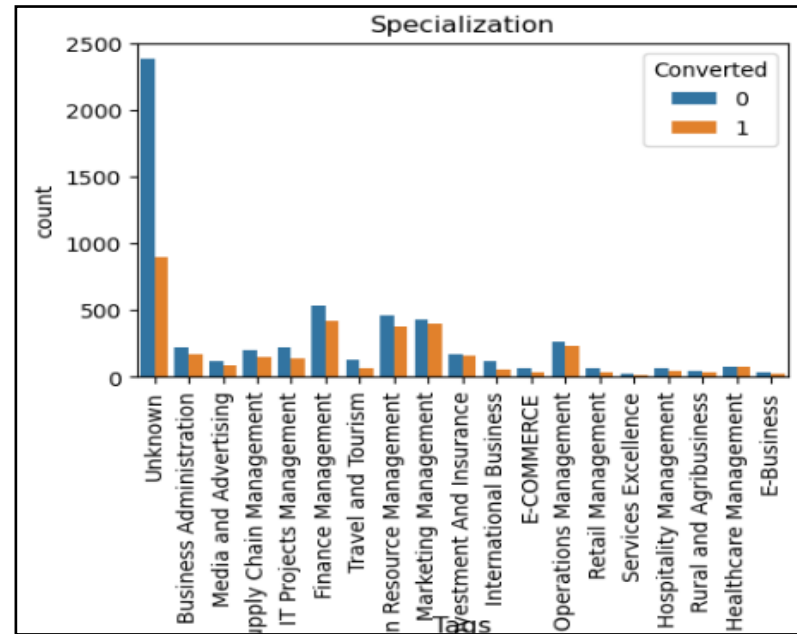
# Insight details

## – Bivariate analysis

- When Lead origin compared with converted, API and landing page submission has lower conversion rate but has considerable conversion counts. Lead Add Form has good conversion rate.
- Lead Source - Reference and welingak website has good conversion rate, Google has the highest conversion.



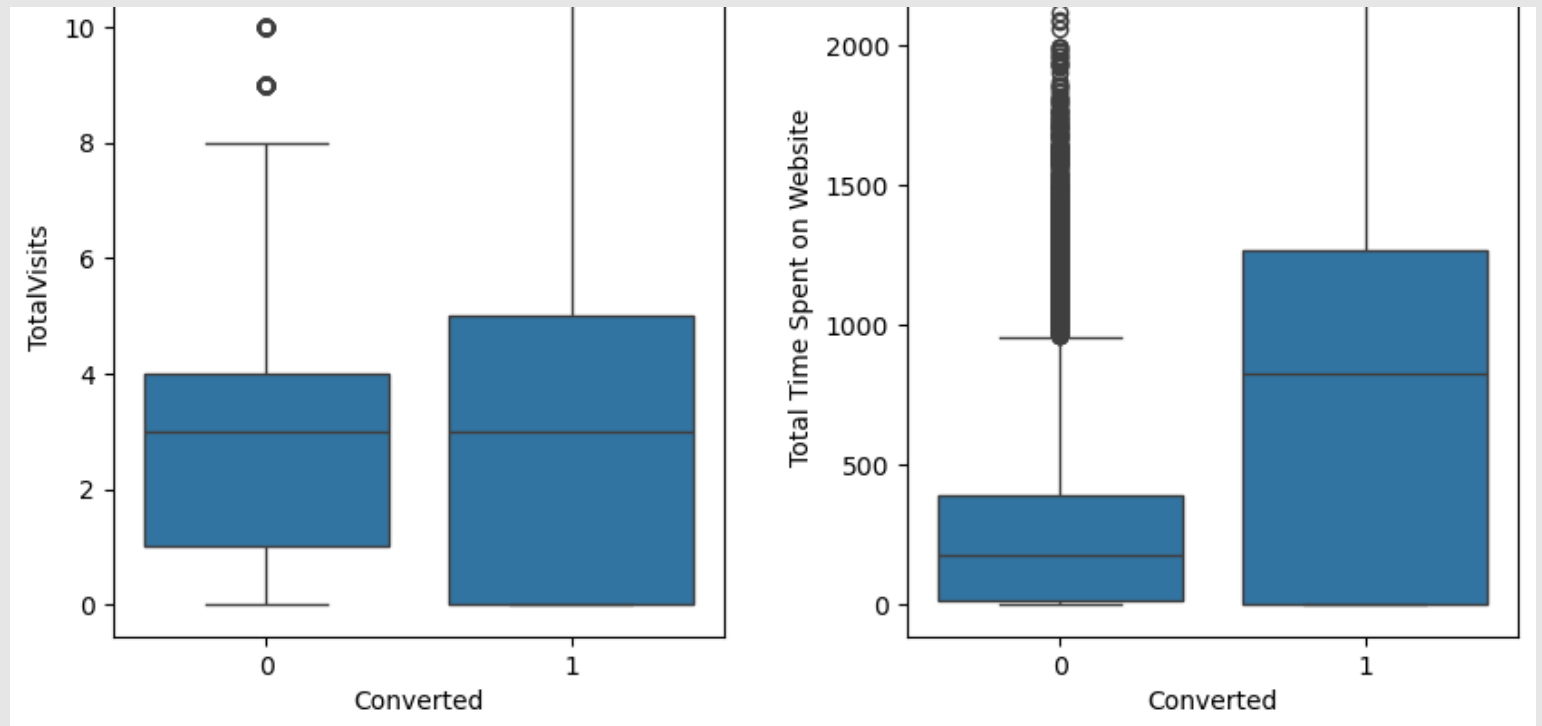
# Bivariate analysis



# Numerical column vs Target column

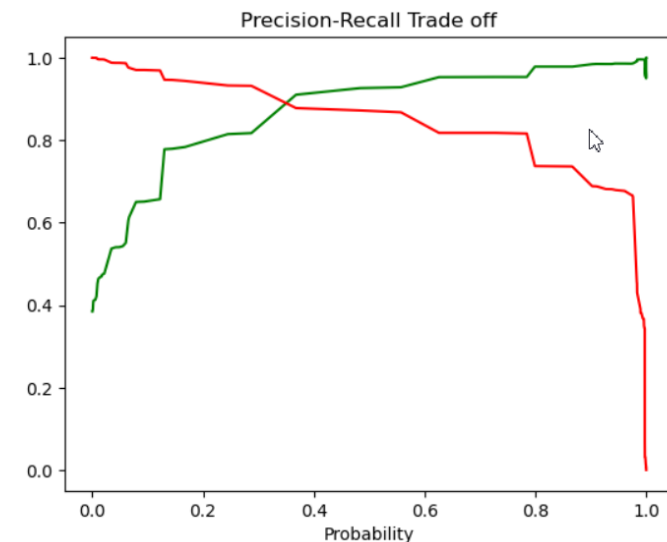
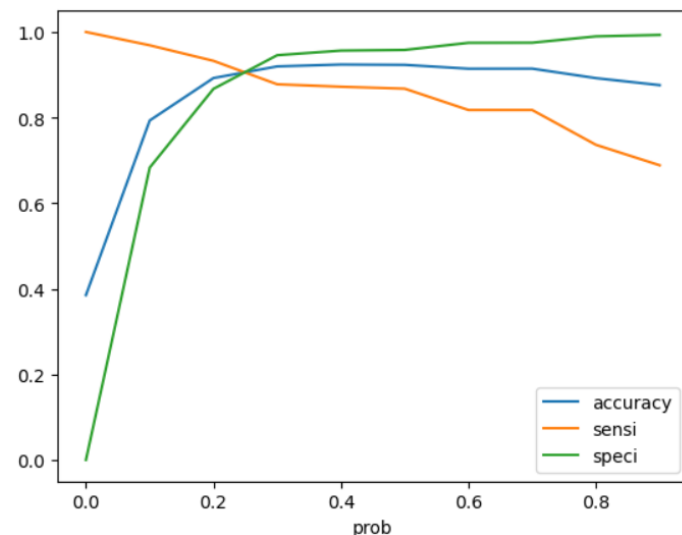
- People those who are spending more time on Website are most likely to join the courses.
- Similarly, leads which visiting most of the time on website has high probability of joining the course.

**Suggestion:** X Education company can make website more interactive to increase the potential lead conversion.





# ROC curve and Cut-off/ Metrics



## Finding Optimal Cut off Point:

- Area under ROC curve is 0.97
- Optimal cut off probability is that probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.35.

## Confusion Matrix :

```
[[1635  99]
 [ 138 851]]
```

Accuracy : 0.9129636430407638

Sensitivity : 0.8604651162790697

Specificity : 0.9429065743944637

Precision : 0.8957894736842106

```
: (0.9129636430407638,
  0.8604651162790697,
  0.9429065743944637,
  0.8957894736842106)
```

```
-----Feature Importance-----
const -0.908699
Do Not Email -1.356992
Lead Source_Welingak Website 4.121405
Last Activity_SMS Sent 2.199831
What is your current occupation_Unknown -1.061079
Tags_Closed by Horizon 6.828270
Tags_Lost to EINS 7.471301
Tags_Ringing -3.933288
Tags_Will revert after reading the email 4.613101
Tags_switched off -3.954568
Lead Quality_Worst -3.746908
Last Notable Activity_Modified -1.832625
Last Notable Activity_Olark Chat Conversation -0.867779
dtype: float64
```

# Conclusion

---



Tags - Lost to EINS, Closed by Horizzon, Ringing, Will revert after reading the email has high potential customers those who are likely to join the course.



Lead Source – Welingak Website is giving more conversion customers.



Customers who using chat assistant in the website - Olark Chat Conversation are most likely join the course.



Our model's precision is 89% and accuracy is 91% which is more than the ballpark.