# Lead Scoring Assignment

# INDEX

# Introduction

Lead scoring is a process of assigning a value to each lead based on their potential to convert into a paying customer.

Once leads have been scored, company will can use this information to prioritize their sales and marketing efforts. For example, they may choose to focus their attention on leads with a high score, as these leads are more likely to convert.

In this presentation, we will be discussing a case study of a company called X Education , which uses lead scoring to improve its sales and marketing results. We will discuss the challenges that X Education was facing, the solution that they implemented, and the results that they achieved.

## Problem Statement

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

**Business Goal :**

Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

# Steps Taken

❑ Import the Data For Analysis

❑ Data Understanding and inspection

❑ Data Cleaning

> **Treatment for 'Select' values**
>
> **Handling Missing Values**
> **Removing Unwanted Columns**
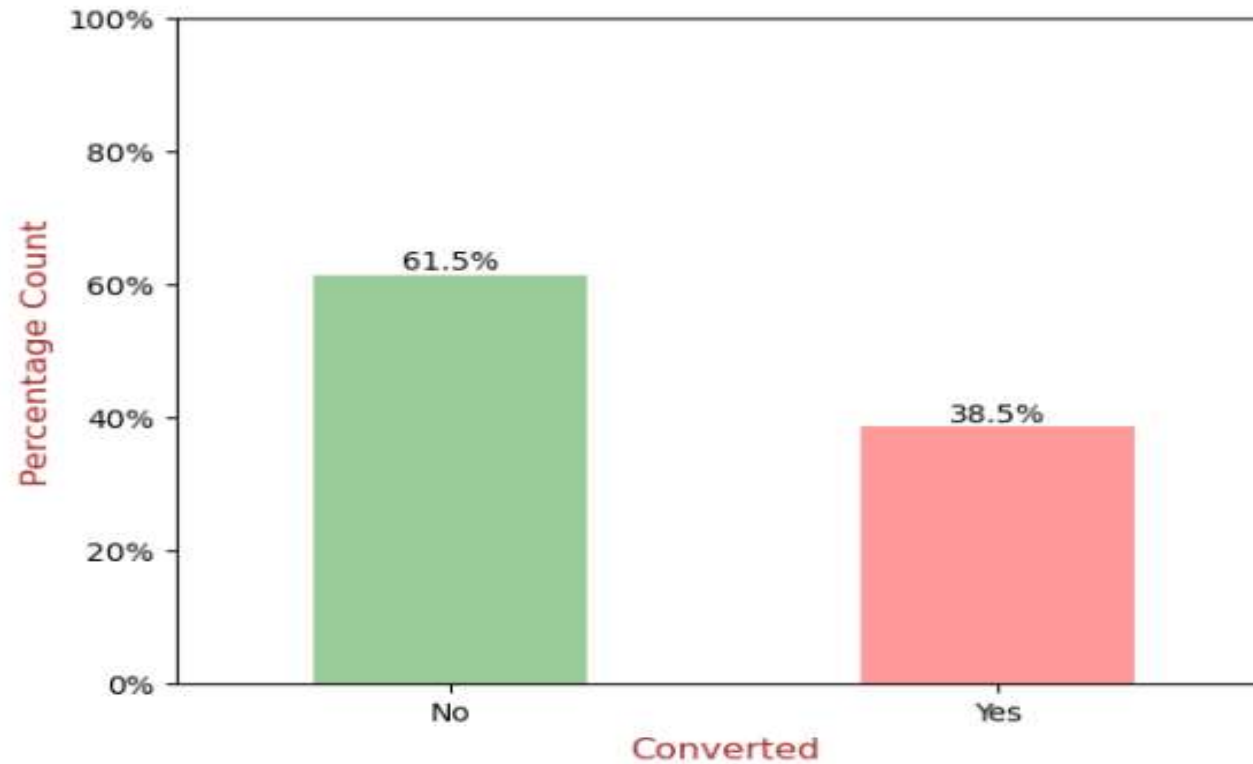> **Checking & Dropping Category Columns that are Skewed**
> **Outlier Analysis**
> **Fixing Invalid values & Standardising Data in columns**

❑ Data Analysis (EDA)

❑ Data Preparation

❑ Feature Scaling

❑ Model Building
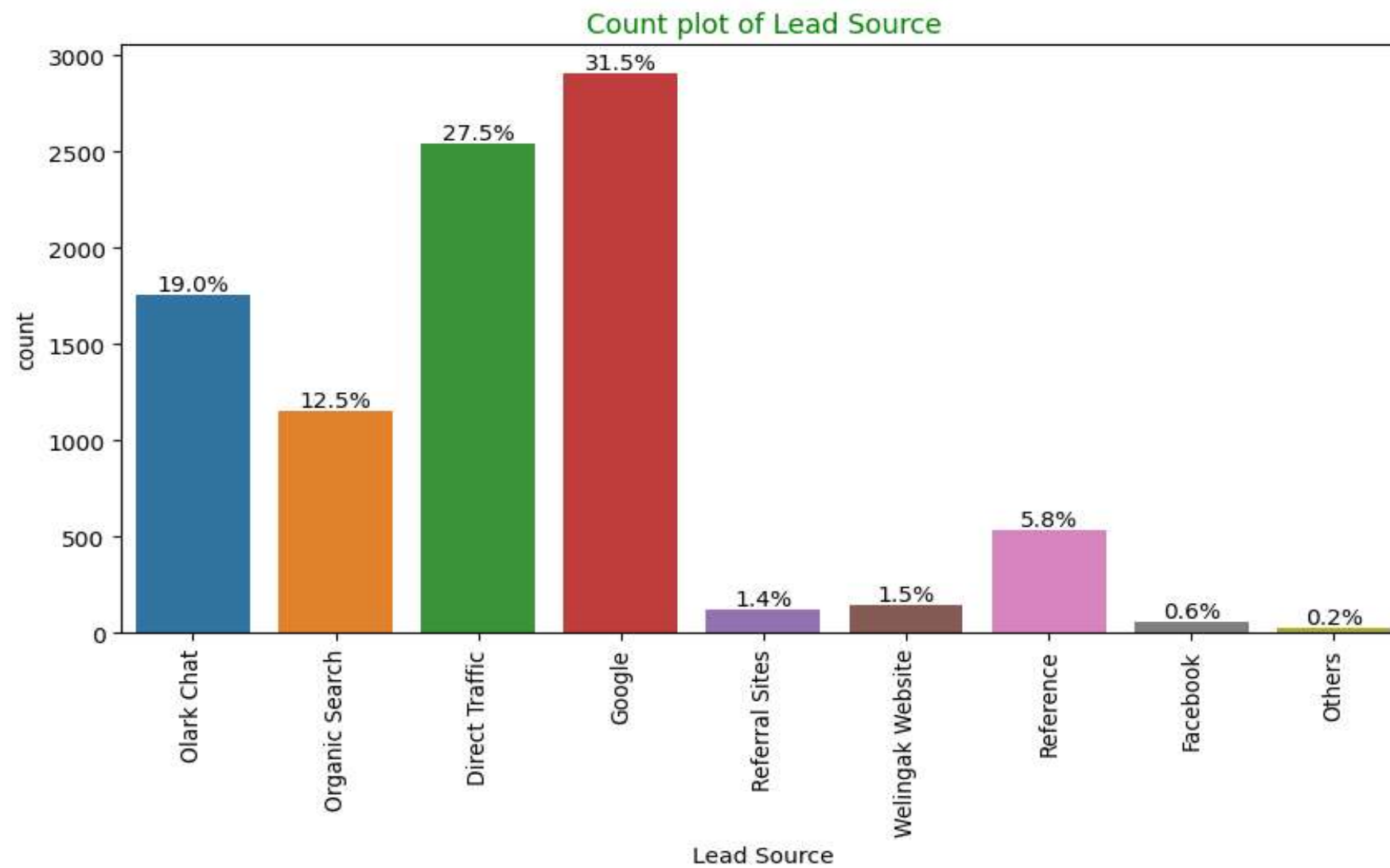
❑ Model Evaluation

❑ Making Predictions on test set

# EDA FINDINGS :

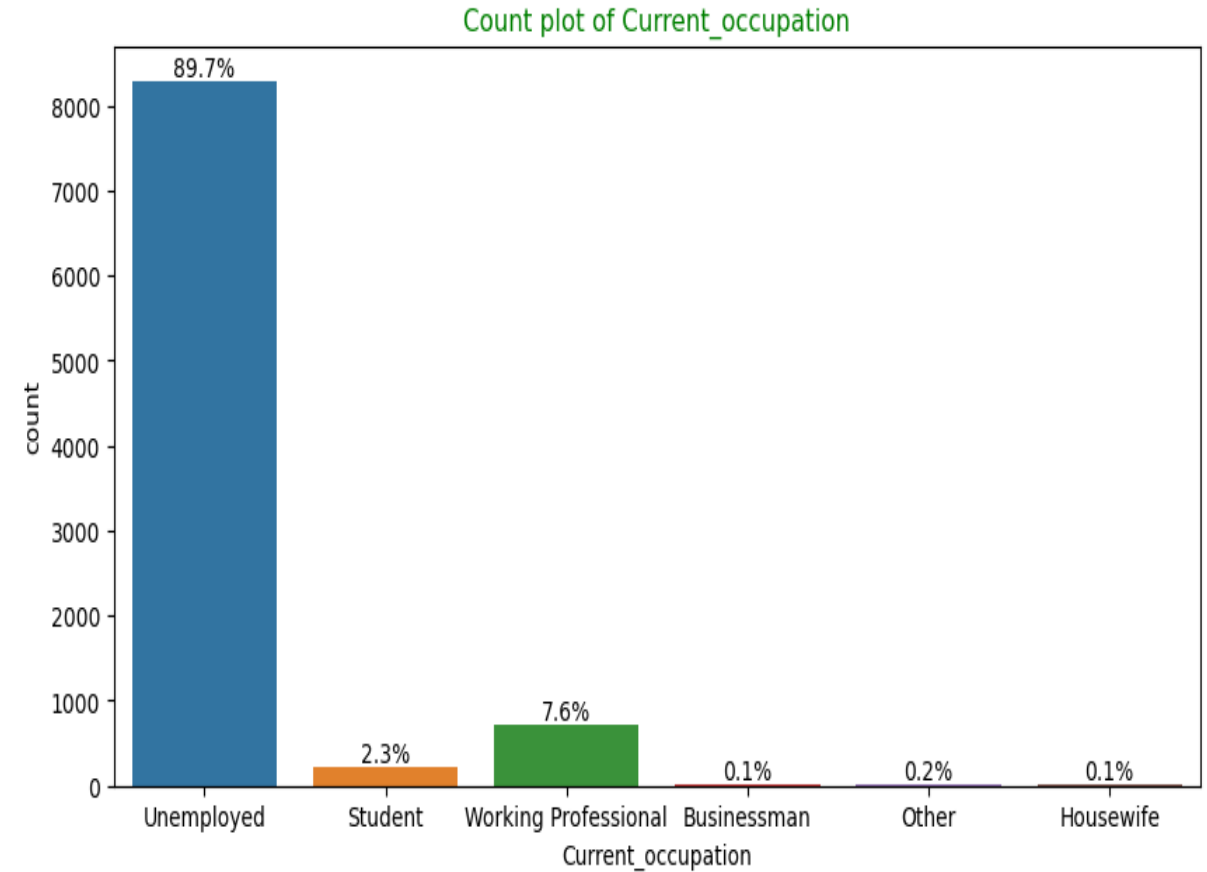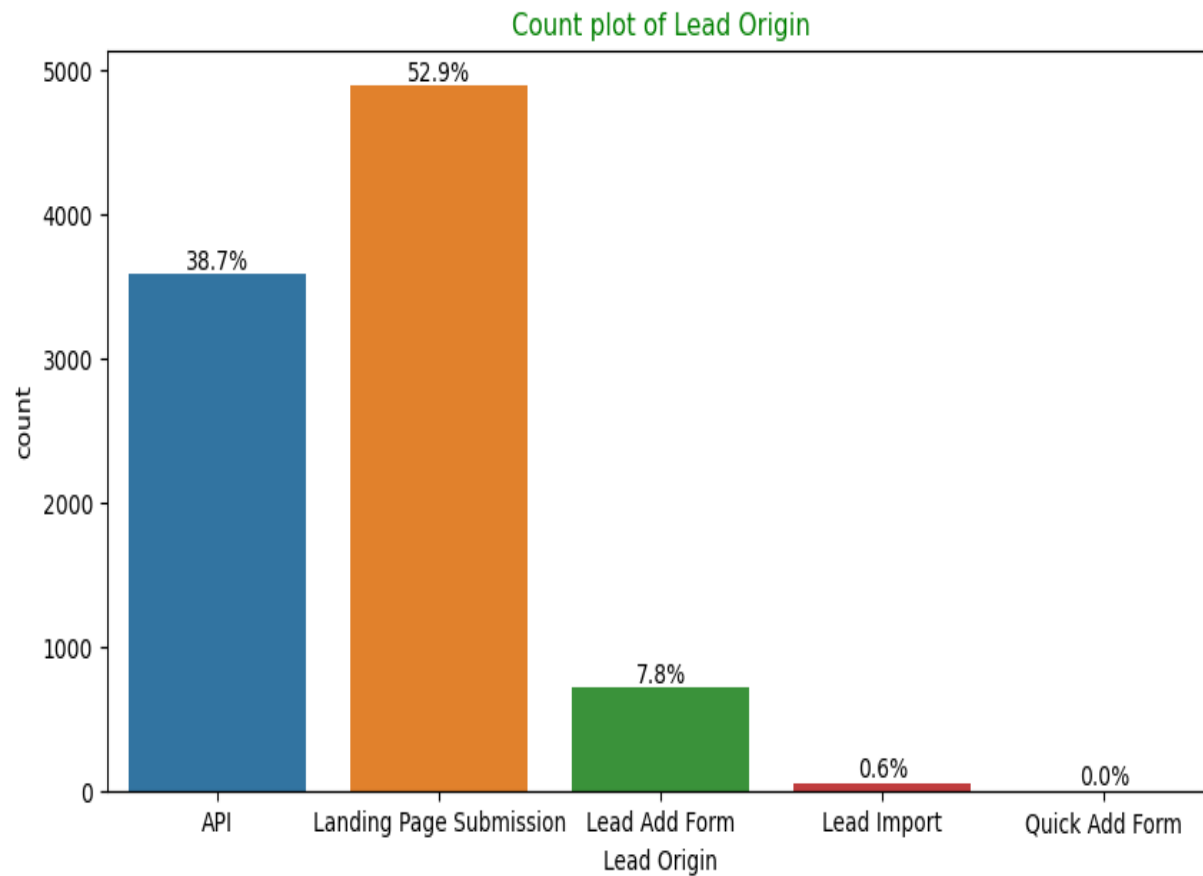## Leads Converted



OBSERVATIONS:

- Conversion rate is of 38.5%, i.e only 38.5% of the people have converted to leads.(Minority)
- While 61.5% of the people didnt convert to leads. (Majority)

Count plot of Lead Source

Count plot of Lead Origin

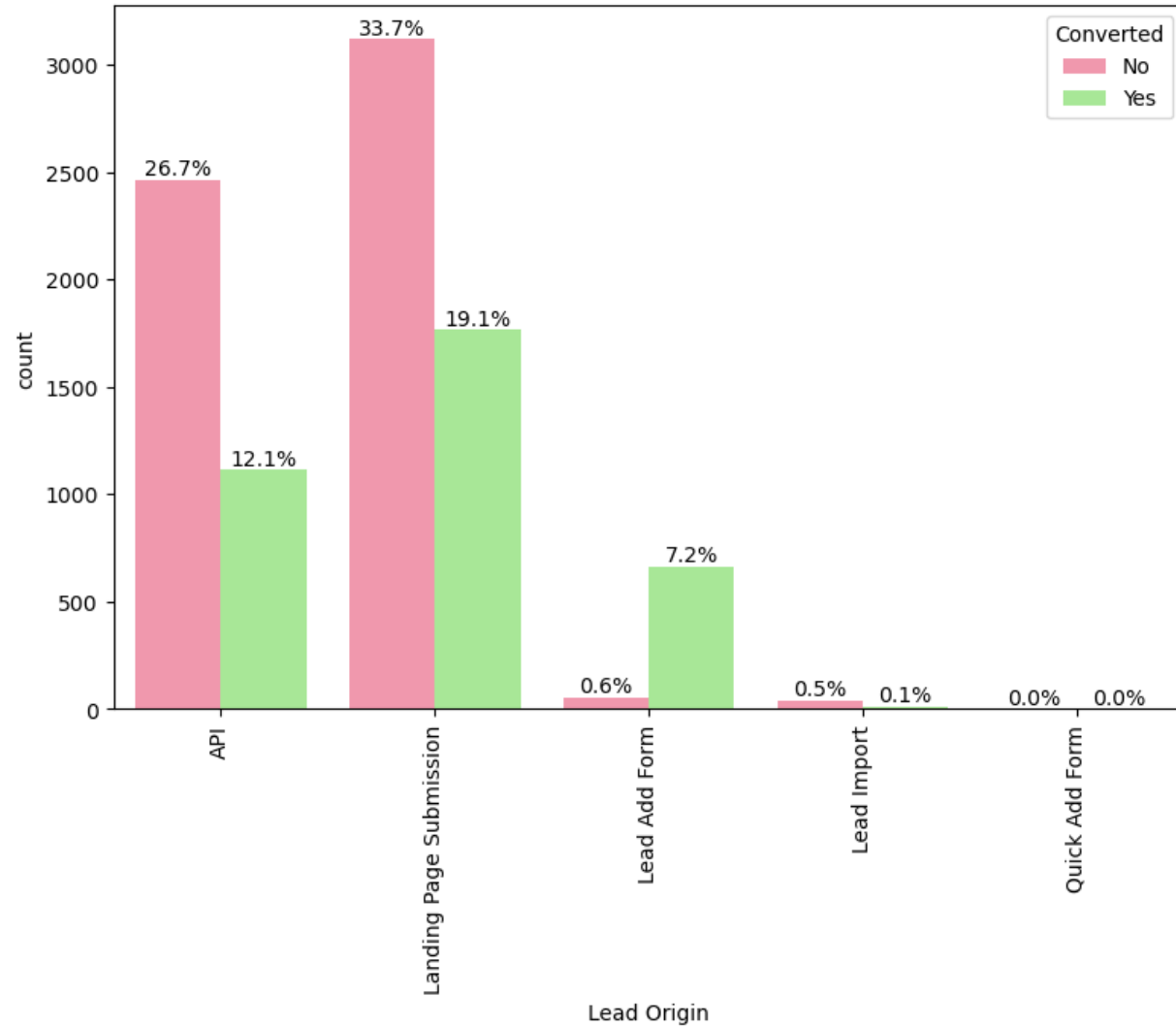Count plot of Current_occupation

**Univariate Observations :**

Here is the list of features from variables which are present in majority (Converted and Not Converted included)

•Lead Origin: "Landing Page Submission" identified 53% customers, "API" identified 39%.

•Current_occupation: It has 90% of the customers as Unemployed

•Do Not Email: 92% of the people has opted that they dont want to be emailed about the course.

•Lead Source: 58% Lead source is from Google & Direct Traffic combined

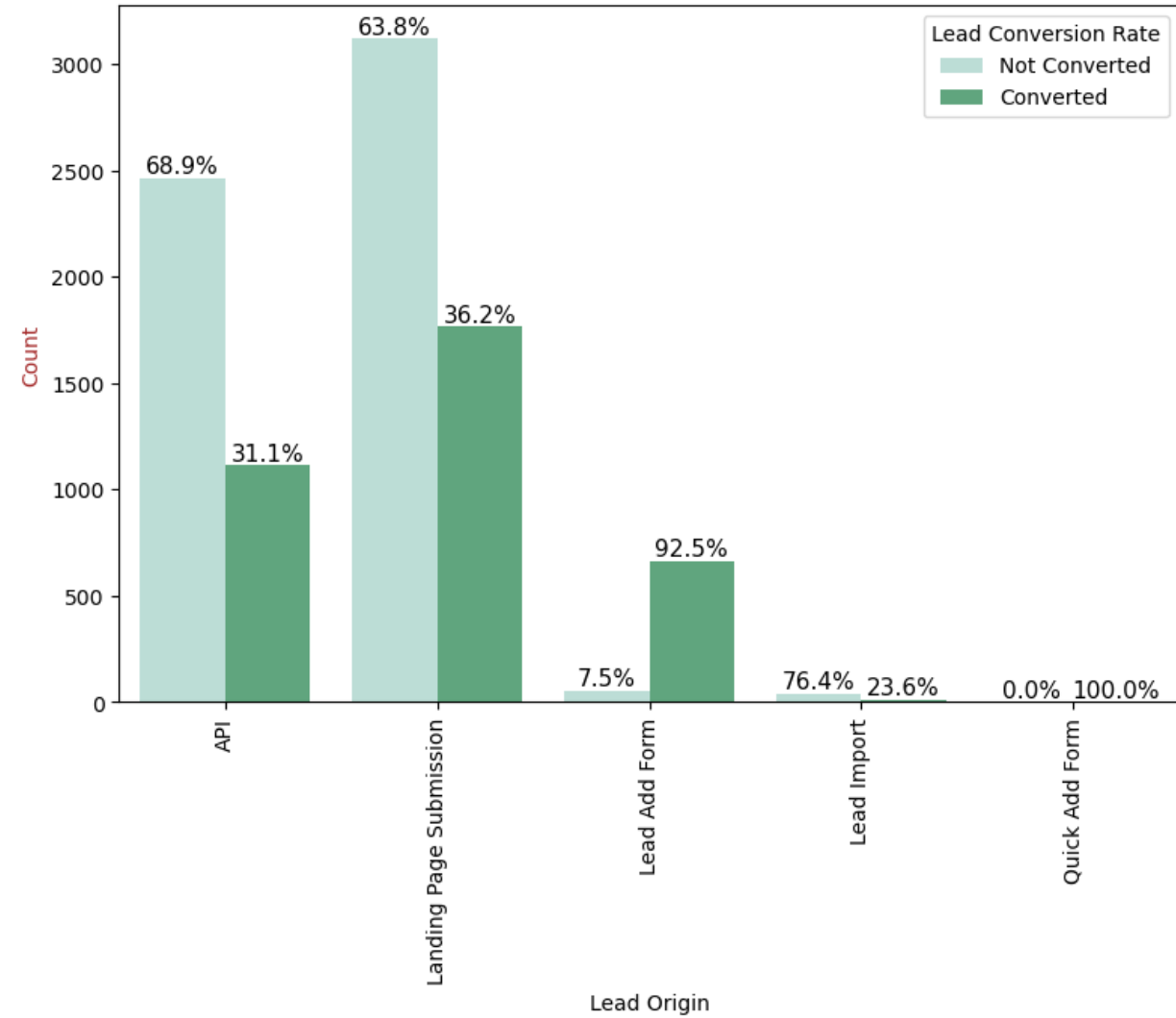•Last Activity: 68% of customers contribution in SMS Sent & Email Opened activities
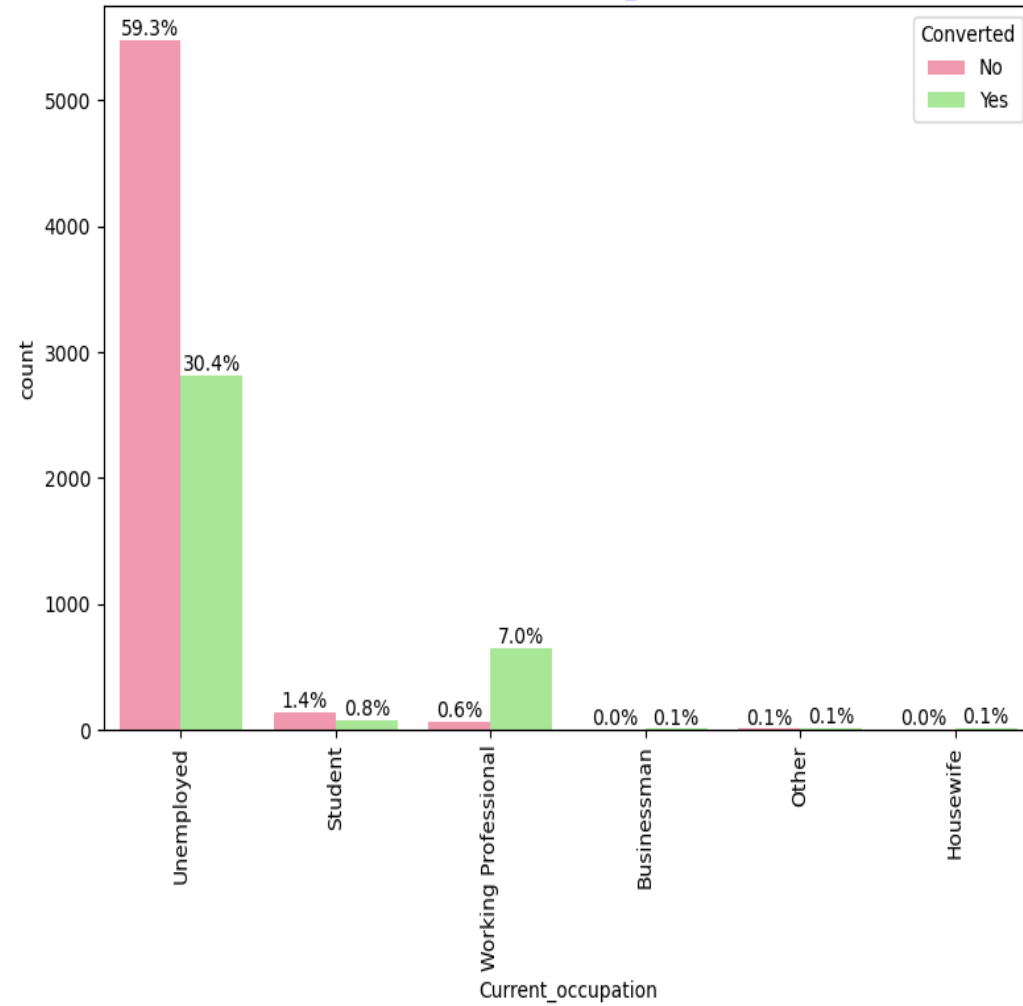
Lead Origin Countplot vs Lead Conversion Rates

Current_occupation Countplot vs Lead Conversion Rates
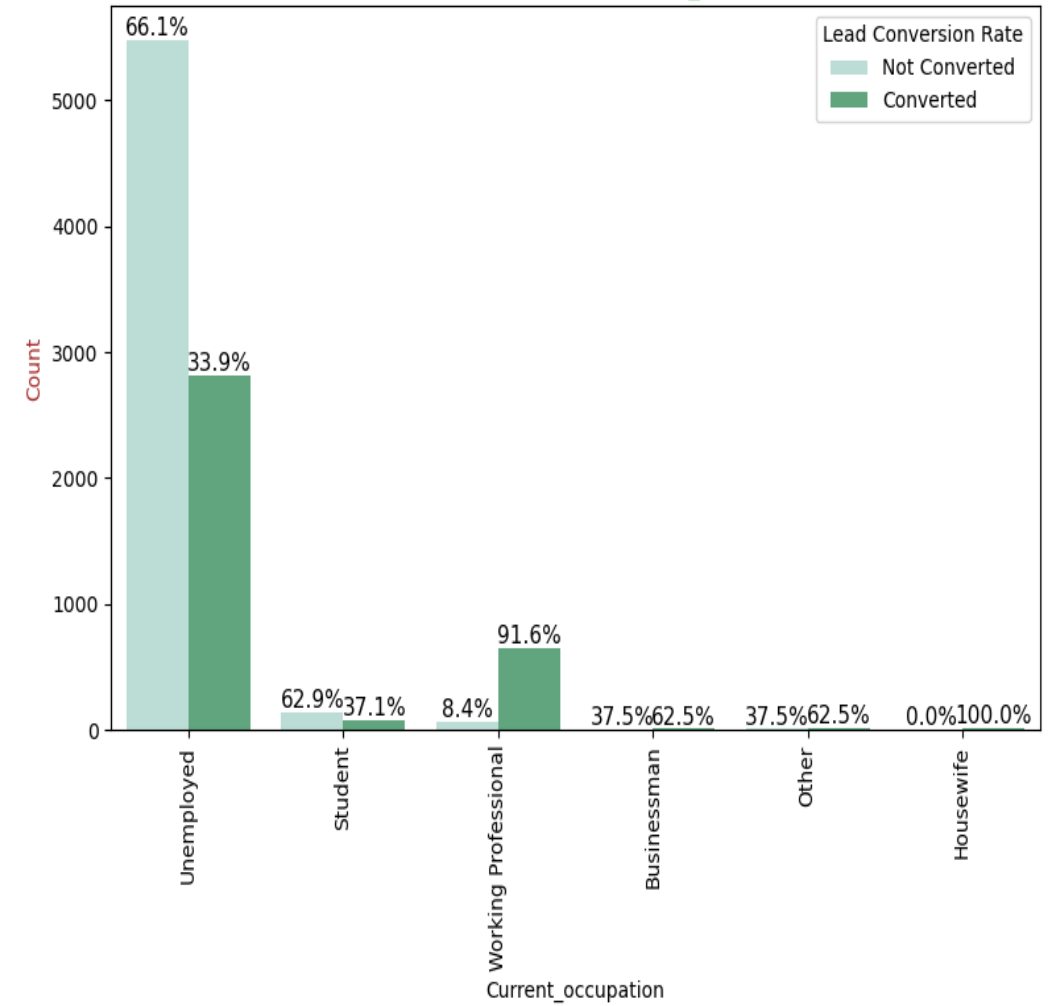
# Lead Source Countplot vs Lead Conversion Rates
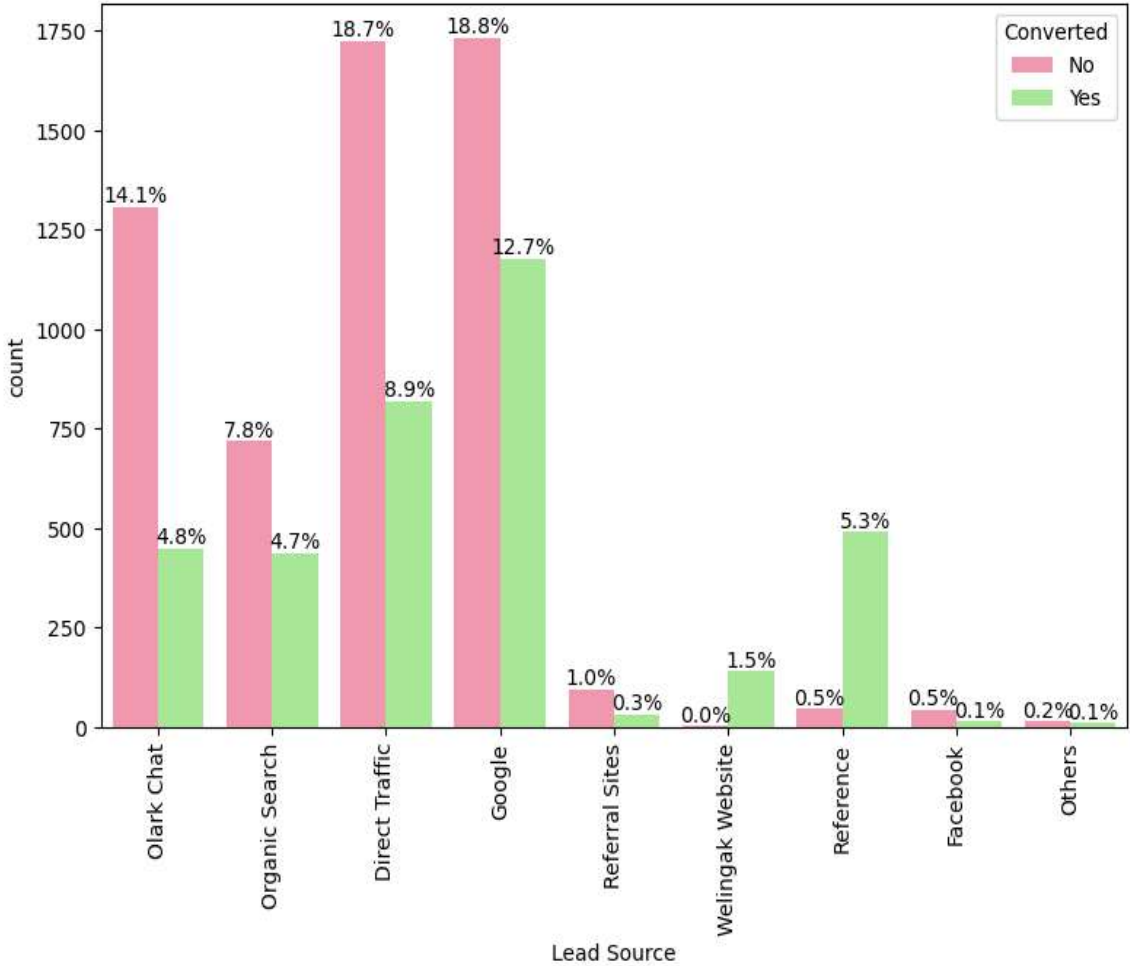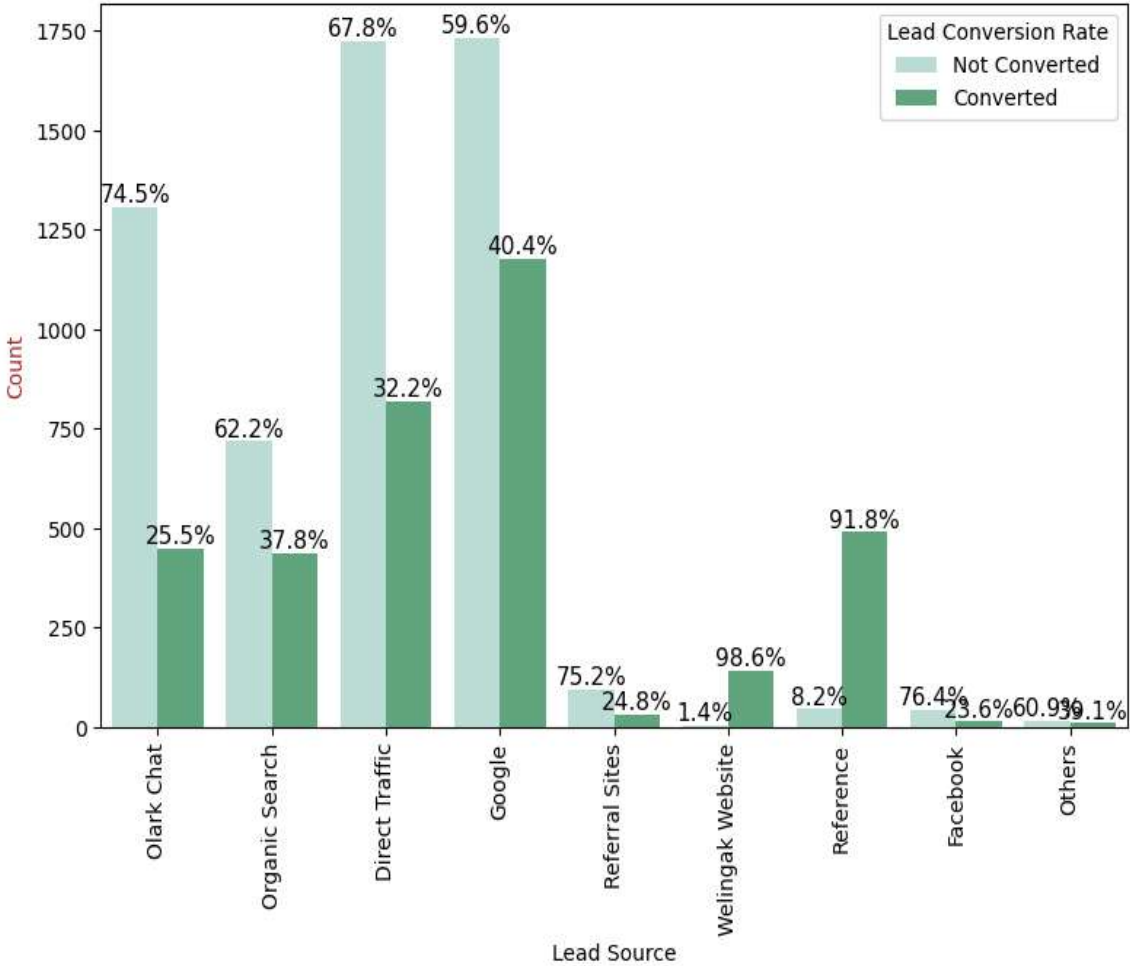
***Observations:***

- Lead Origin: Around 52% of all leads originated from "Landing Page Submission" with a lead conversion rate (LCR) of 36%.The "API" identified approximately 39% of customers with a lead conversion rate (LCR) of 31%.
- Current_occupation: Around 90% of the customers are Unemployed with lead conversion rate (LCR) of 34%. While Working Professional contribute only 7.6% of total customers with almost 92% lead conversion rate (LCR).
- Do Not Email: 92% of the people has opted that they dont want to be emailed about the course.
- Lead Source: Google has LCR of 40% out of 31% customers , Direct Traffic contributes 32% LCR with 27% customers which is lower than Google,Organic Search also gives 37.8% of LCR but the contribution is by only 12.5% of customers ,Reference has LCR of 91% but there are only around 6% of customers through this Lead Source.

***Note:***
- We have assumed LCR as Lead Conversion Rate in short form.
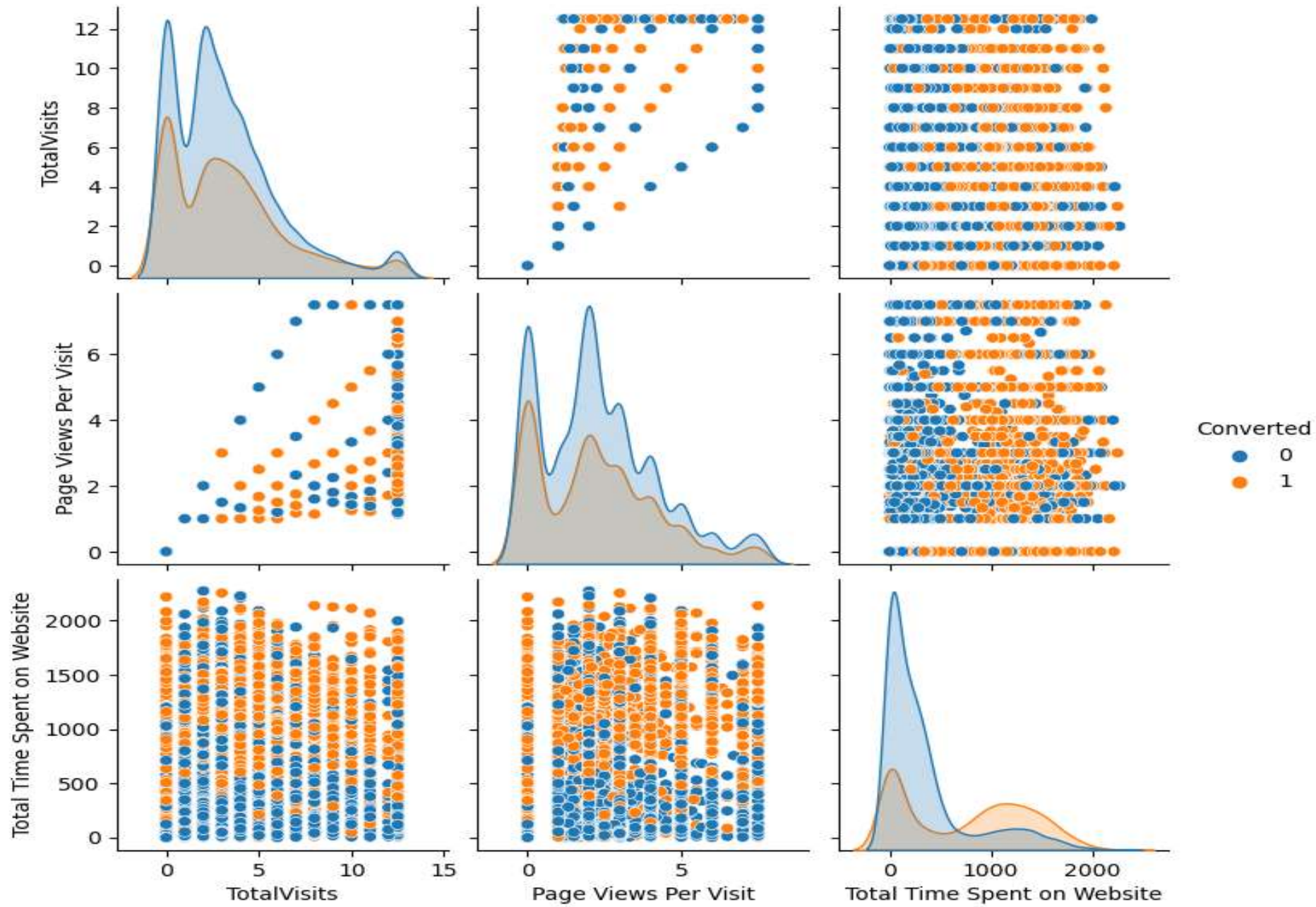
# Model Evaluation
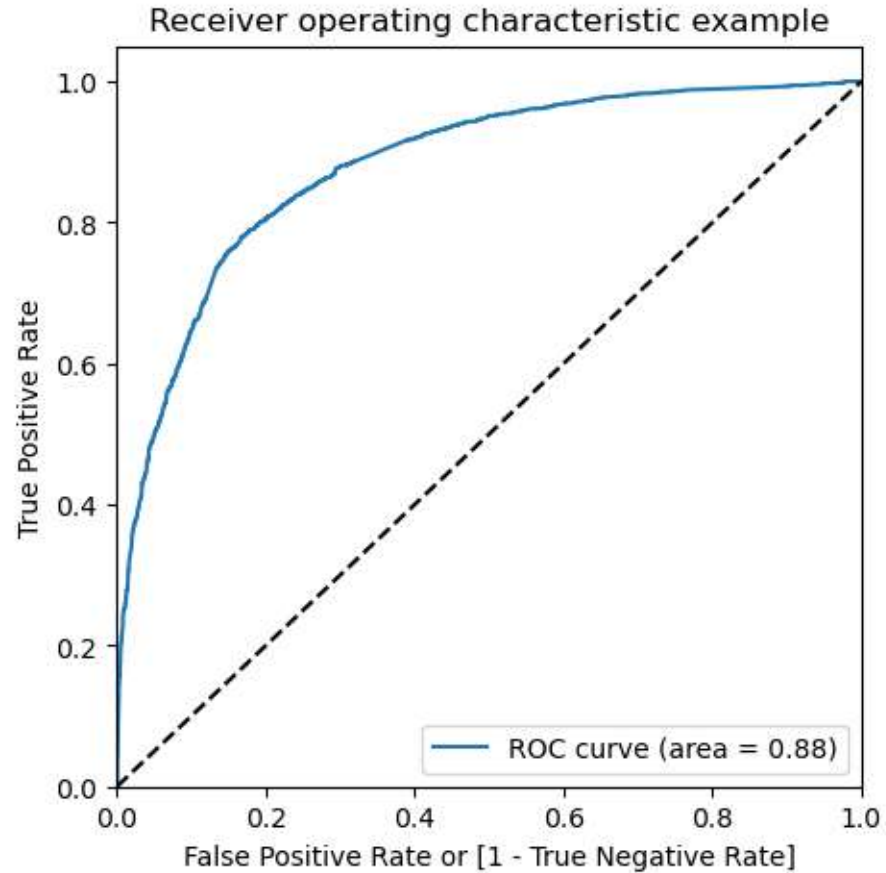


**NOTE:**
•Area under ROC curve is 0.88 out of 1 which indicates a good predictive model
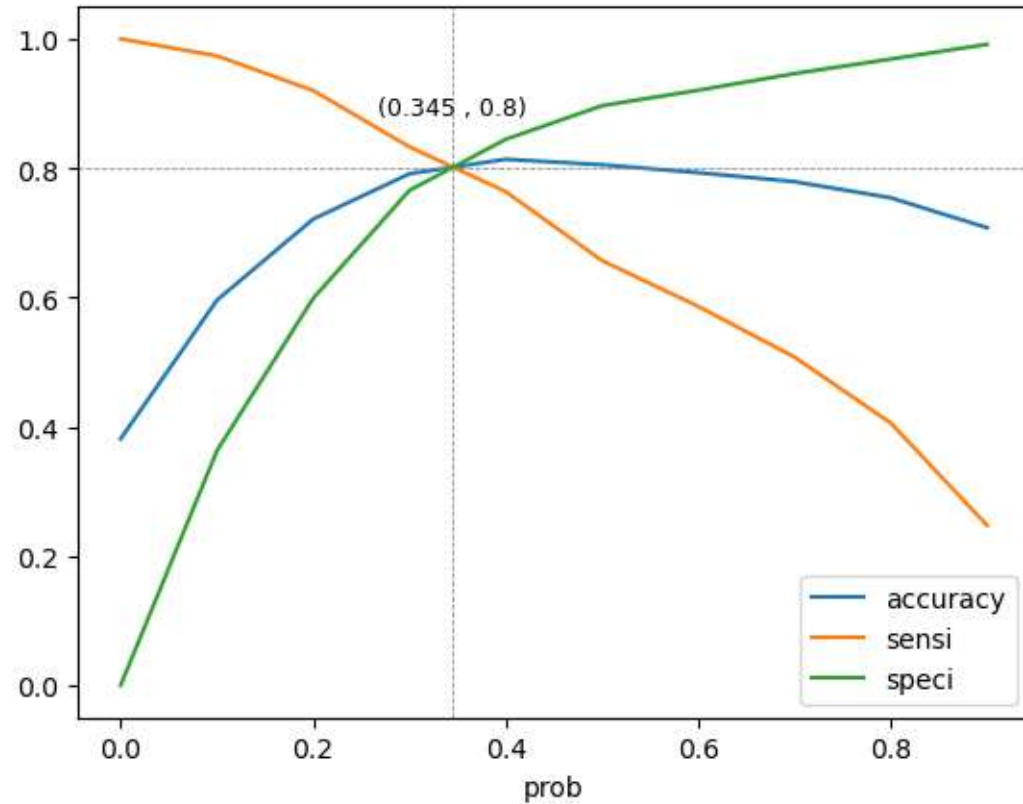
**NOTE:**
•0.345 is the approx. point where all the curves meet, so 0.345 seems to be our Optimal cutoff point for probability threshold .
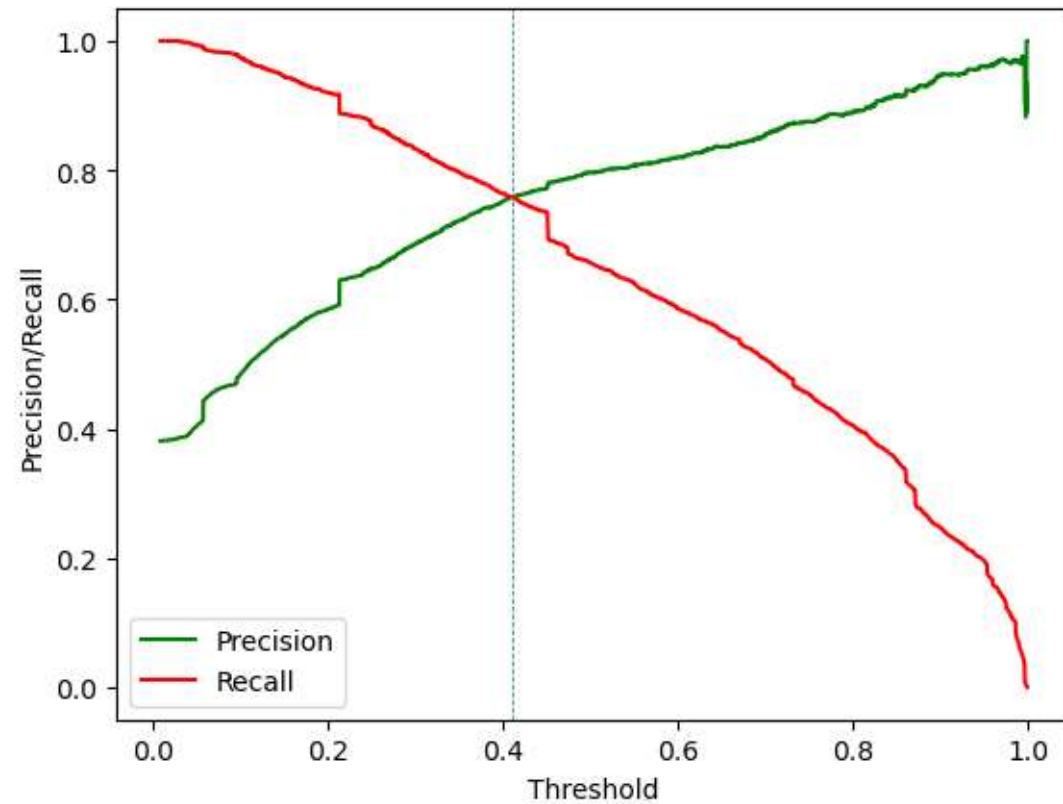
**NOTE:**
•The intersection point of the curve is the threshold value where the model achieves a balance between precision and recall. It can be used to optimise the performance of the model based on business requirement,Here our probability threshold is 0.41 aprrox from above curve.

|   | Prospect ID | Converted | Converted_Prob | final_predicted | Lead_Score |
|---|---|---|---|---|---|
| 0 | 4269 | 1 | 0.697934 | 1 | 70 |
| 1 | 2376 | 1 | 0.860665 | 1 | 86 |
| 2 | 7766 | 1 | 0.889241 | 1 | 89 |
| 3 | 9199 | 0 | 0.057065 | 0 | 6 |
| 4 | 4359 | 1 | 0.871510 | 1 | 87 |

Lead Score: Lead Score is assigned to the customers

- The customers with a higher lead score have a higher conversion chance
- The customers with a lower lead score have a lower conversion chance.

### Train - Test

**Train Data Set:**

Accuracy: 80.46%

Sensitivity: 80.05%

Specificity: 80.71%

**Test Data Set:**

Accuracy: 80.34%

Sensitivity: 79.82% ≈ 80%

Specificity: 80.68%

### NOTE:

The evaluation matrics are pretty close to each other so it indicates that the model is performing consistently across different evaluation metrics in both test and train dataset.

- The model achieved a sensitivity of 80.05% in the train set and 79.82% in the test set, using a cut-off value of 0.345.
- Sensitivity in this case indicates how many leads the model identify correctly out of all potential leads which are converting
- The CEO of X Education had set a target sensitivity of around 80%.
- The model also achieved an accuracy of 80.46%, which is in line with the study's objectives.

# Recommendations :

To increase our Lead Conversion Rates:

➢ Focus on features with positive coefficients for targeted marketing strategies.
➢ Develop strategies to attract high-quality leads from top-performing lead sources.
➢ Engage working professionals with tailored messaging.
➢ Optimize communication channels based on lead engagement impact.
➢ More budget/spend can be done on Welingak Website in terms of advertising, etc.
➢ Incentives/discounts for providing reference that convert to lead, encourage providing more references.
➢ Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.


To identify areas of improvement:

➢ Analyze negative coefficients in specialization offerings.
➢ Review landing page submission process for areas of improvement

Thank you