



upGrad

# Lead Scoring Case Study

TEAM :

Himanshu (himuchaudhari@gmail.com)

Himank (himank.Khandelwal@gmail.com)

Hemang (hhra2323@gmail.com)

# Agenda

---

- Introduction
- Business Objective
- Problem Statement
- Lead – Conversion Process
- Proposed solution
- Solutions
- Implementation
- EDA
- Logistic Regression Final Model
- Conclusion
- Recommendation

# Business Objective

---

- X Education, an education company, sells online courses to industry professionals.
- Many interested professionals visit their website.
- X Education wants to identify the most promising leads.
- To achieve this, they aim to build a model that identifies hot leads and deploy this model for future use

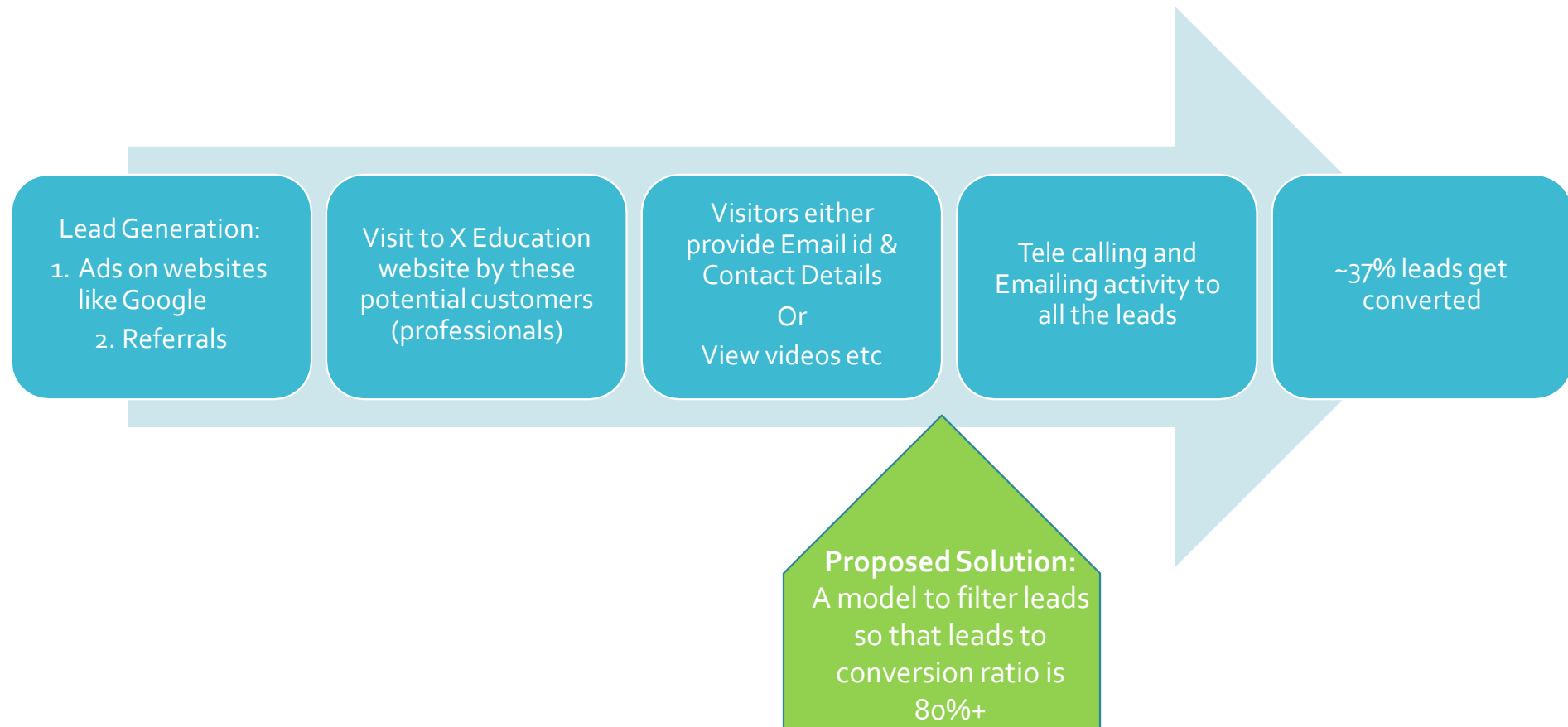
# Problem Statement

---

- X Education receives many leads, but its lead conversion rate is very poor.
- To improve this process, the company aims to identify the most promising leads, also known as 'Hot Leads'.
- By successfully identifying these leads, the lead conversion rate should increase as the sales team will focus on communicating with potential leads rather than calling everyone.
- We will assist them in selecting the most promising leads, i.e., those most likely to convert into paying customers.
- We need to build a model that assigns a lead score to each lead, indicating the likelihood of conversion, with higher scores representing a higher conversion chance.
- The CEO has set a target lead conversion rate of 80%

# Lead - Conversion Process

---



# Proposed Solution

---

## ➤ Selection of Hot Leads

We cluster the leads into categories based on their likelihood of conversion, resulting in a smaller group of hot leads to focus on.

## ➤ Communicating with Hot Leads

With a smaller set of leads to communicate with, we can achieve a greater impact through effective communication.

## ➤ Conversion of Hot Leads

By focusing on leads that are more likely to convert, we can improve our conversion rate and achieve the 80% target.

# Solution

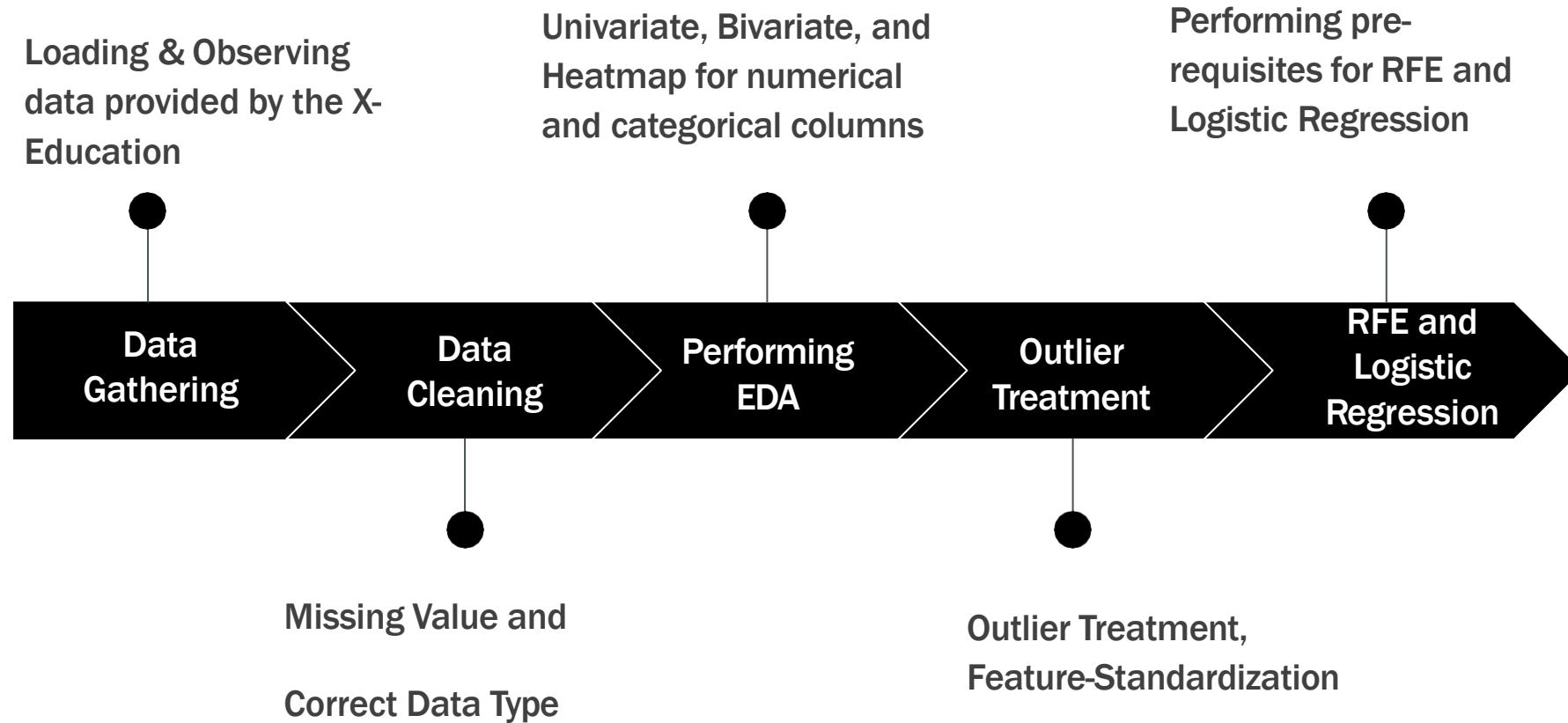
---

## ➤ Selection of Hot Leads

For our solution, the crucial part is accurately identifying hot leads. The more accurately we identify hot leads, the higher our conversion ratio will be. Given our target of an 80% conversion rate, achieving high accuracy in identifying hot leads is essential.

# Implementation

---





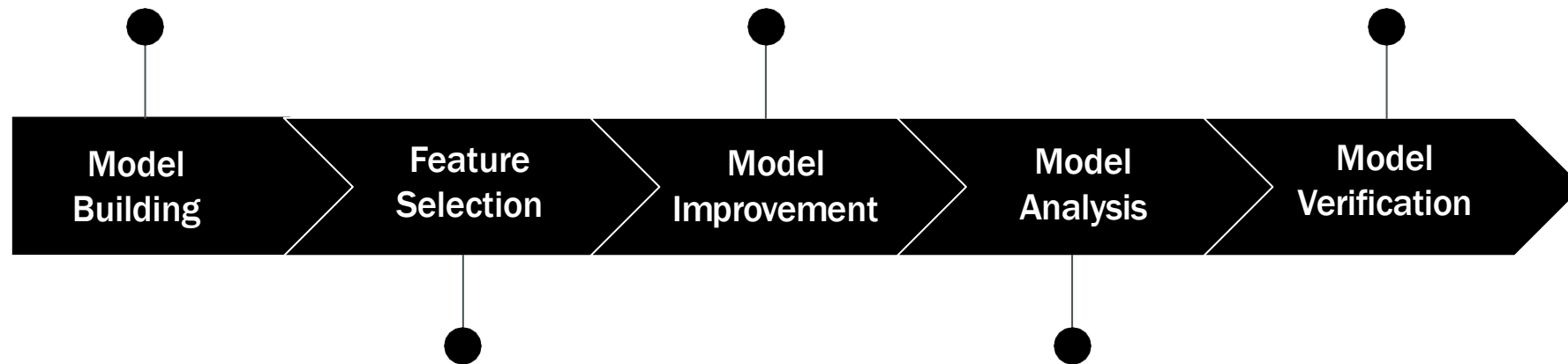
# Implementation

---

Selection of top 15  
features using RFE

Reduction of columns  
and Model re-building

Verifying our Final Model  
Accuracy etc. with model  
built with PCA



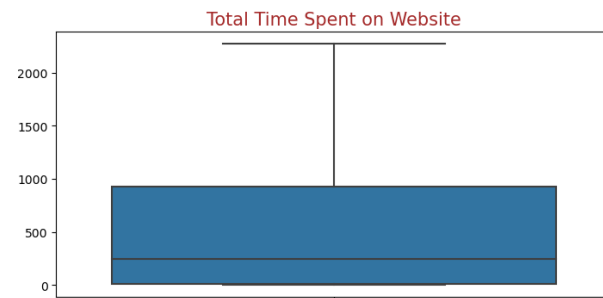
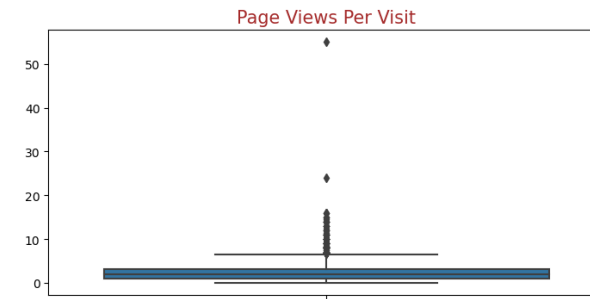
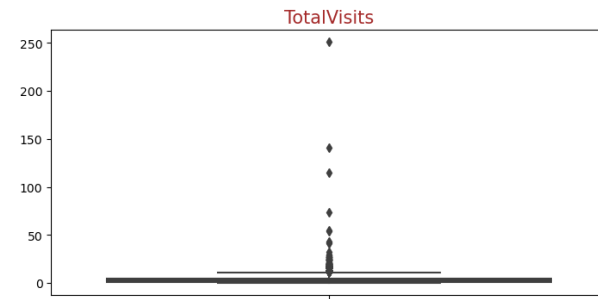
Model building using RFE for  
selected columns

Final Model Analysis and  
performance on Test Data

# EDA - Box Plot - Outliers

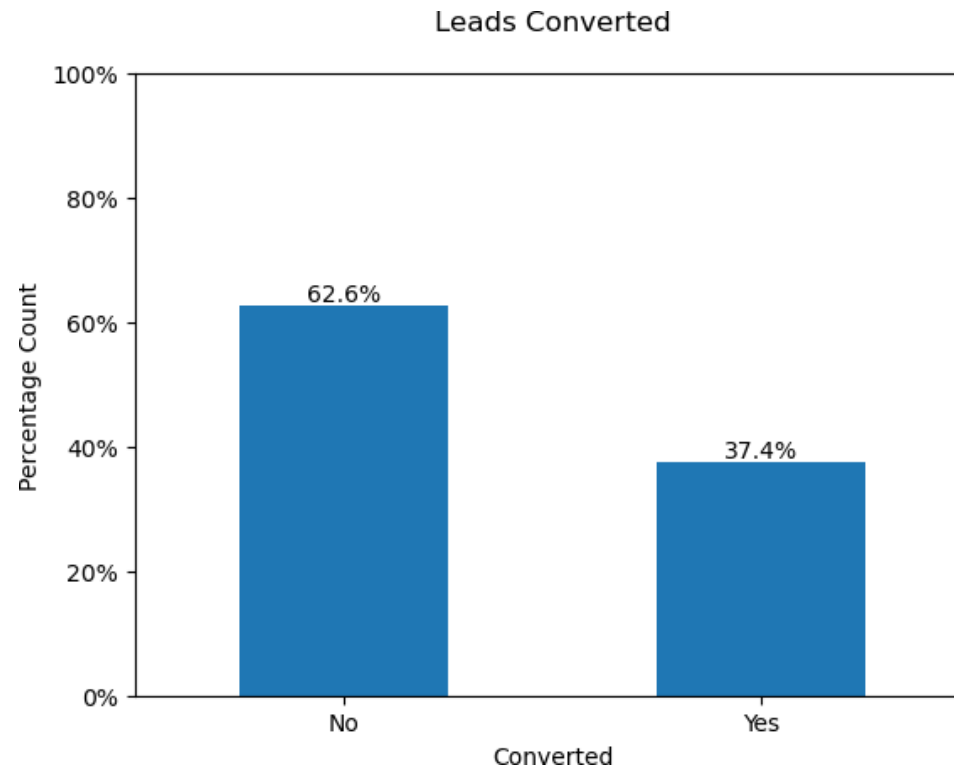
---

Checking Outliers using Boxplot



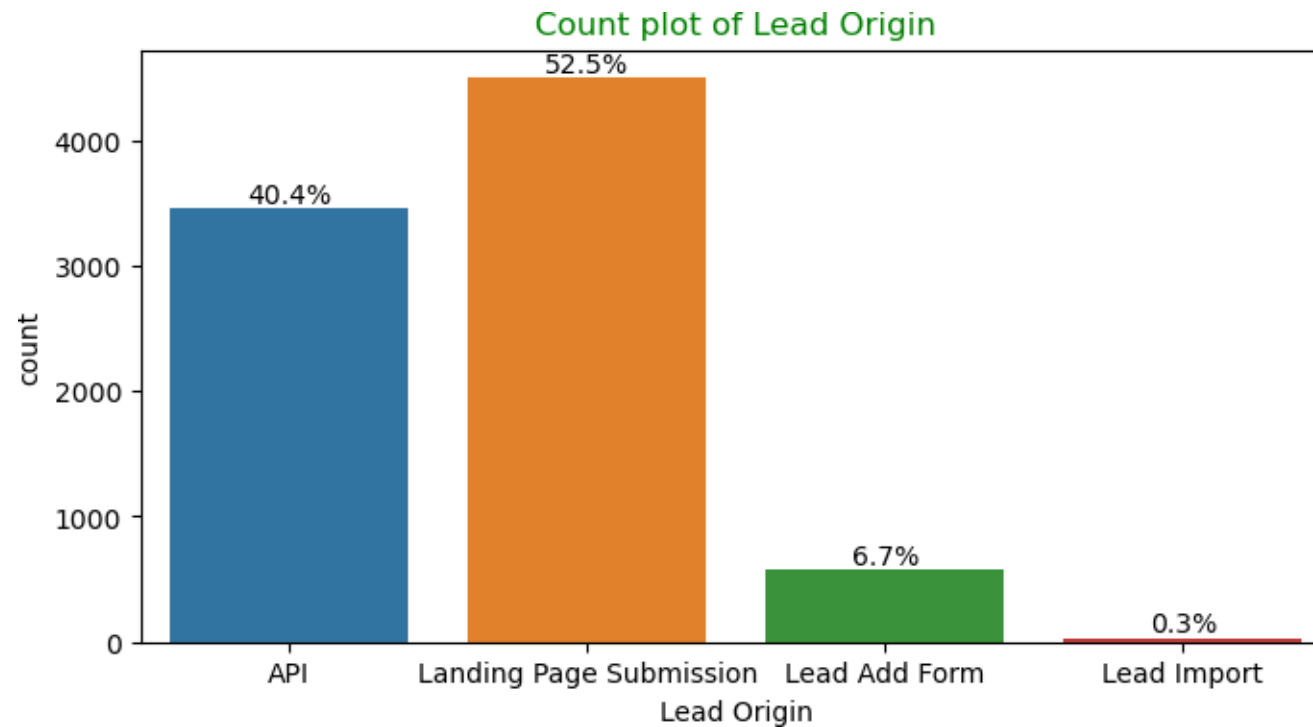
# EDA - Univariate Analysis

---



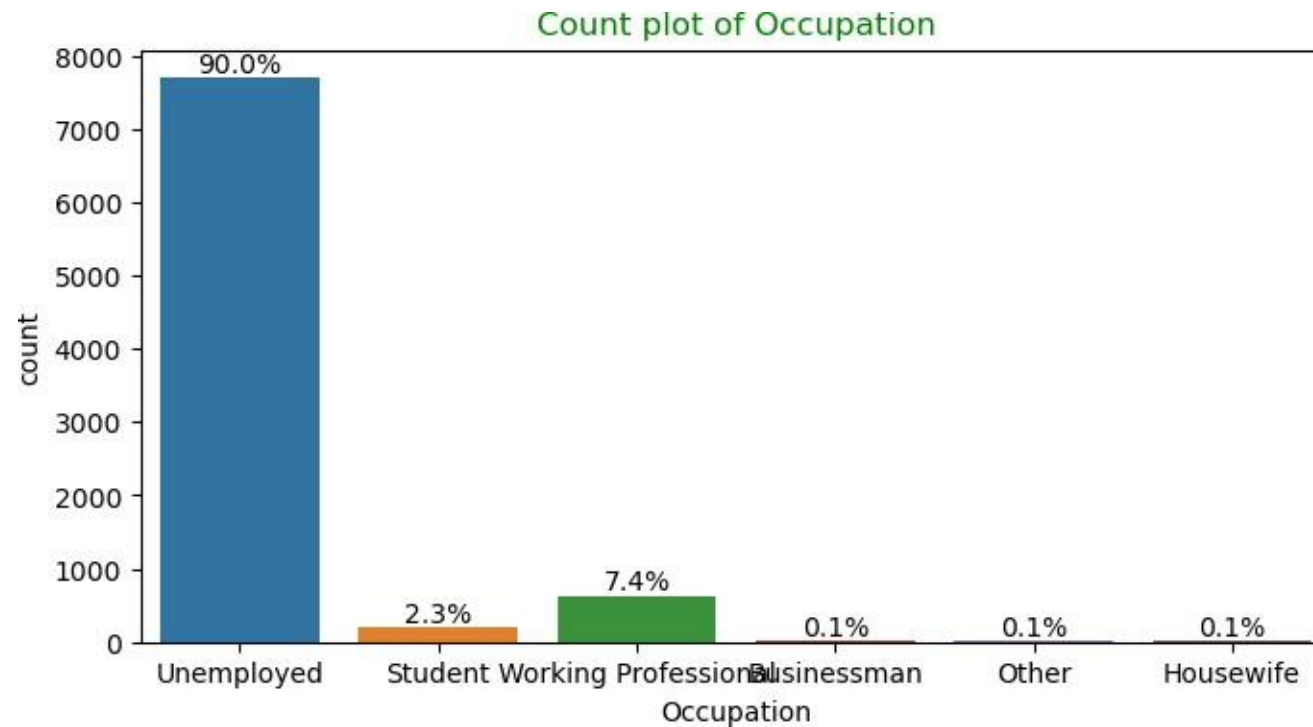
# EDA - Univariate Analysis

---



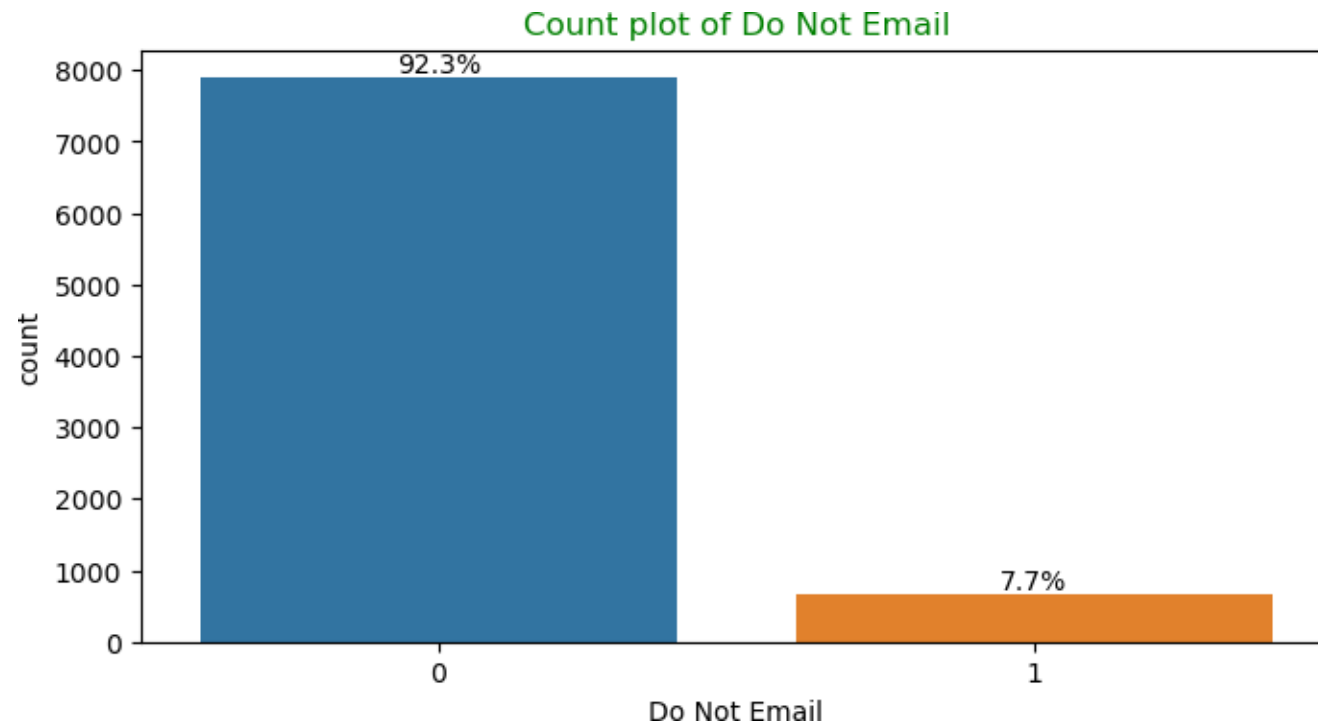
# EDA - Univariate Analysis

---



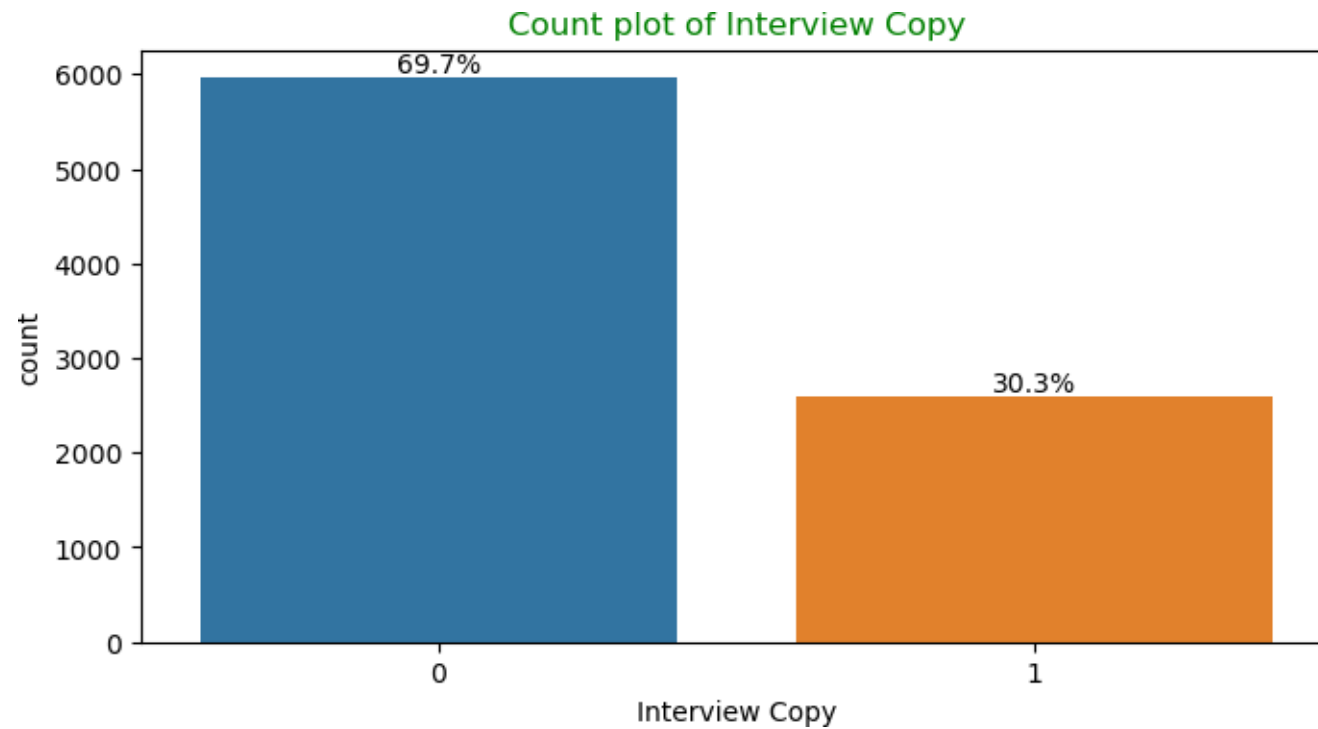
# EDA - Univariate Analysis

---



# EDA - Univariate Analysis

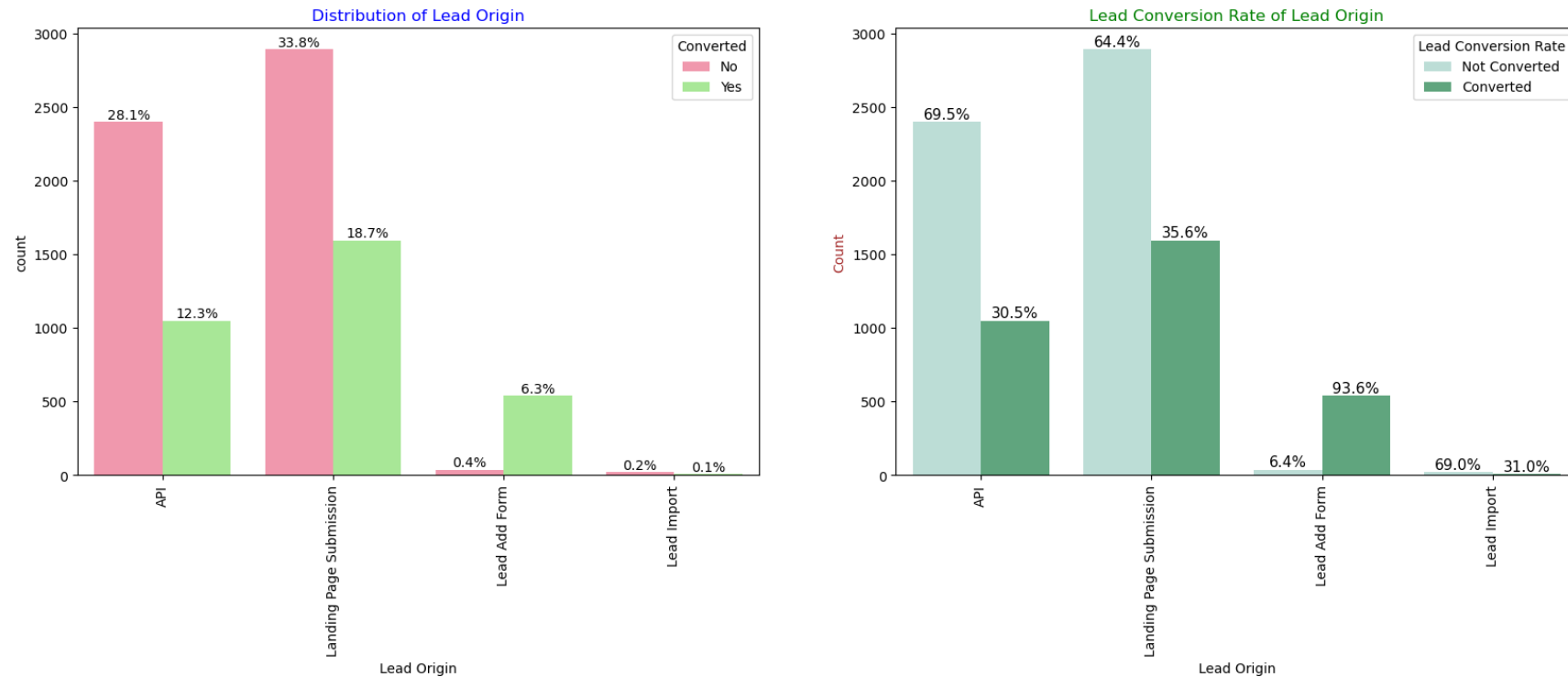
---



# EDA - Bivariate Analysis

---

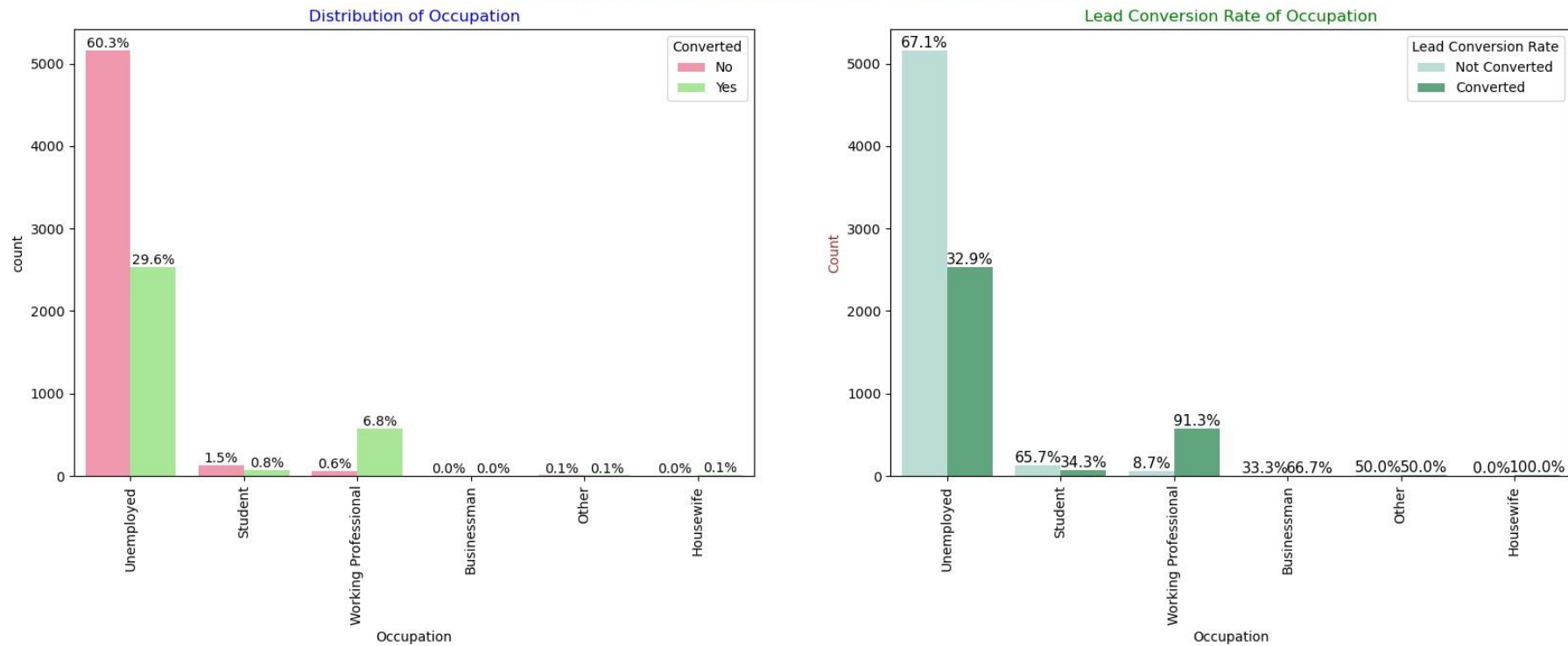
Lead Origin Countplot vs Lead Conversion Rates





# EDA - Bivariate Analysis

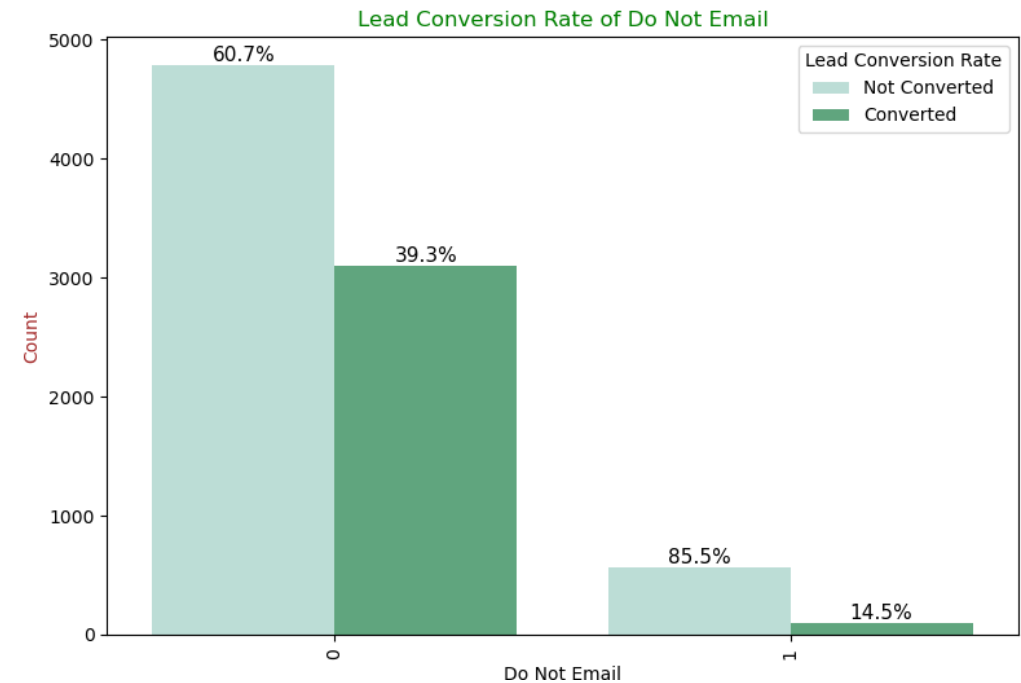
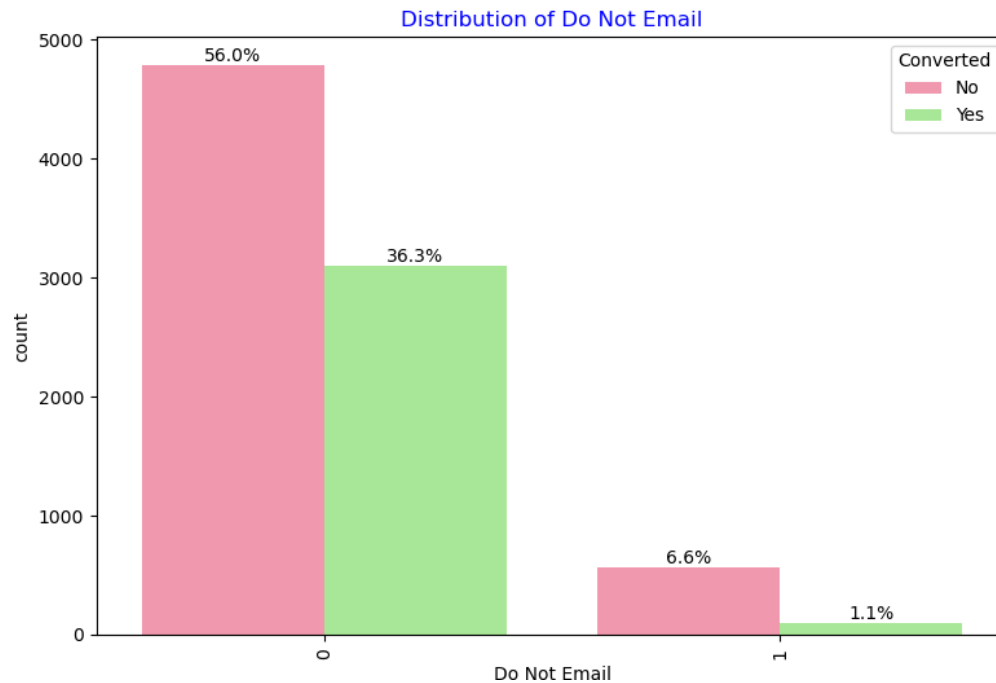
Occupation Countplot vs Lead Conversion Rates



# EDA - Bivariate Analysis

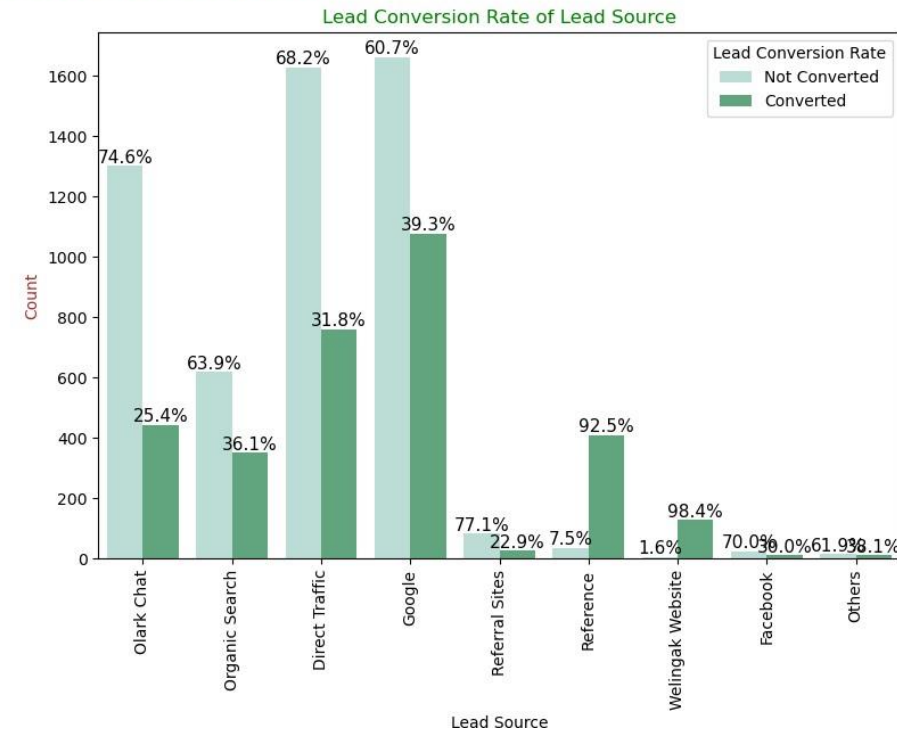
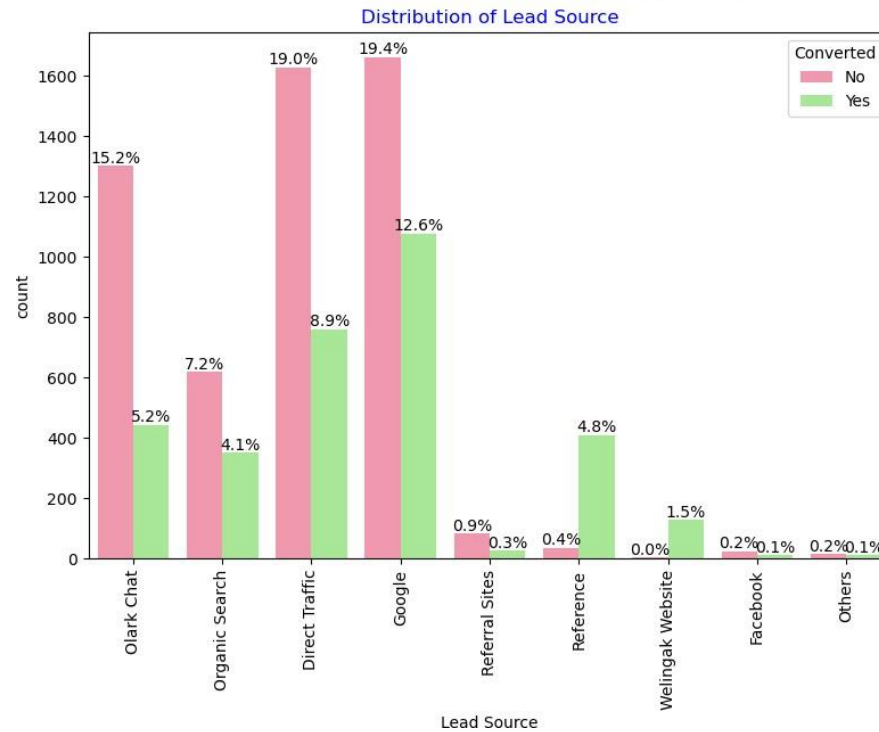
---

Do Not Email Countplot vs Lead Conversion Rates



# EDA - Bivariate Analysis

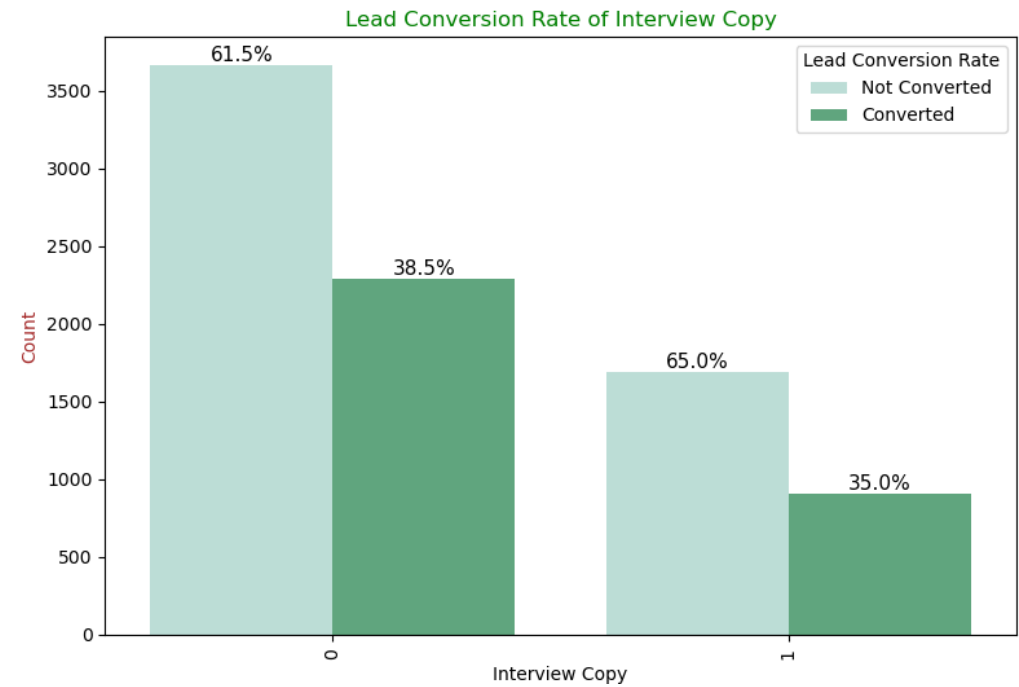
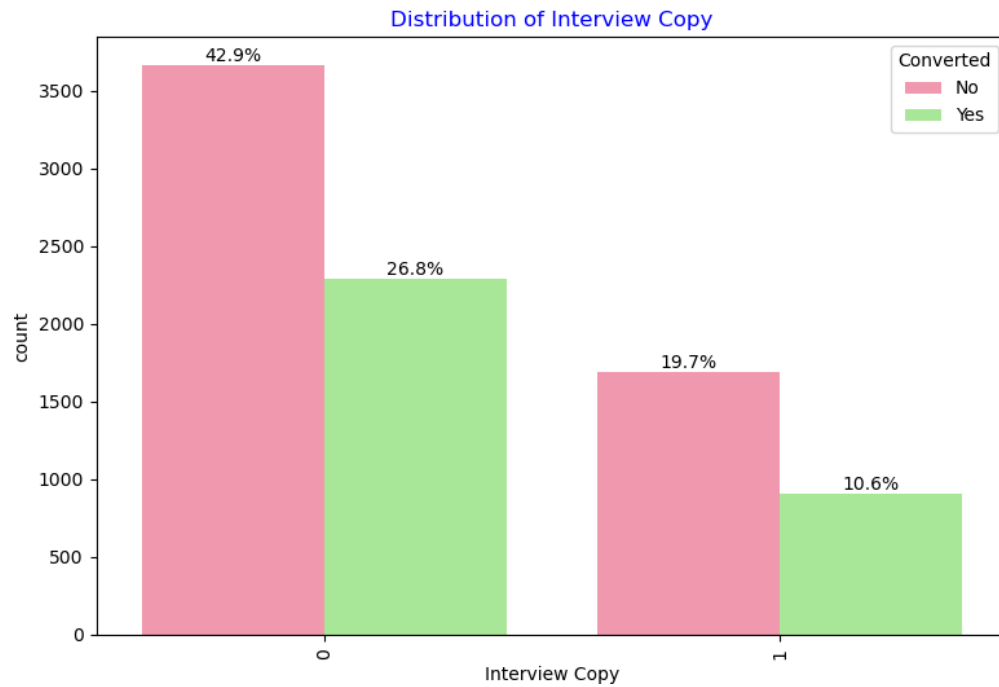
Lead Source Countplot vs Lead Conversion Rates



# EDA - Bivariate Analysis

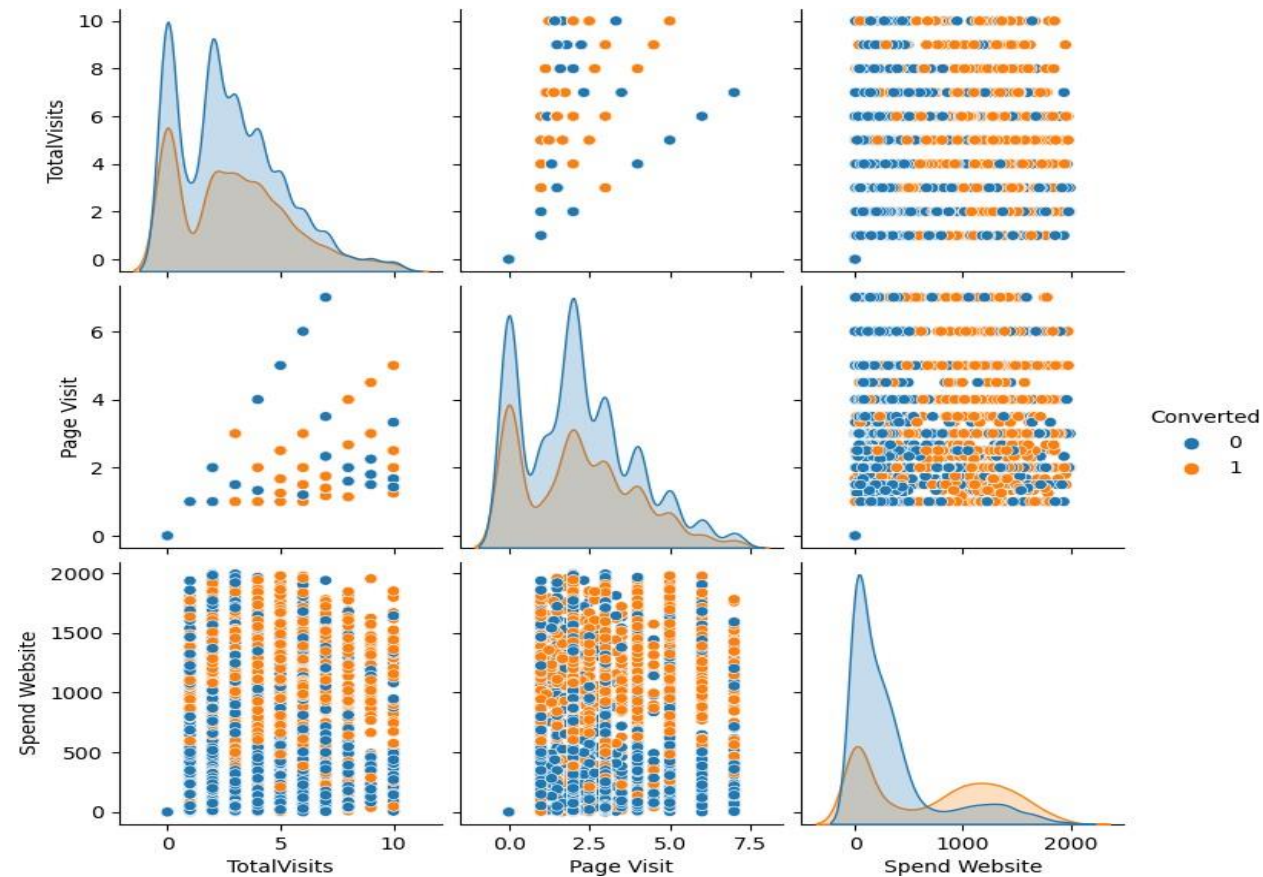
---

Interview Copy Countplot vs Lead Conversion Rates



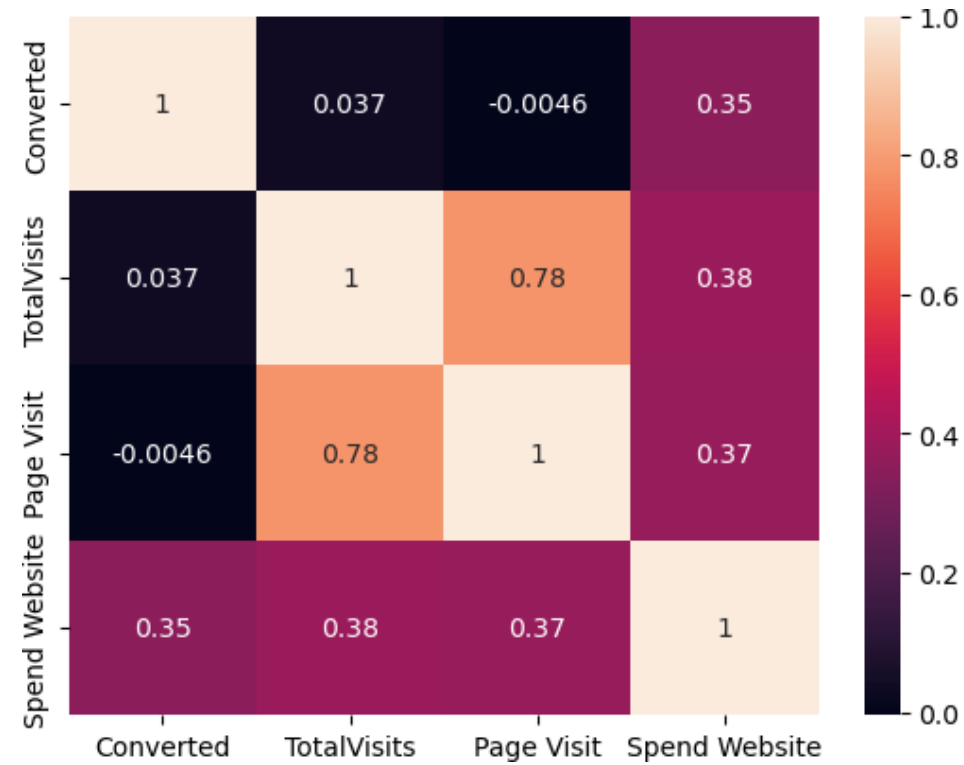
# EDA - Bivariate Analysis

---

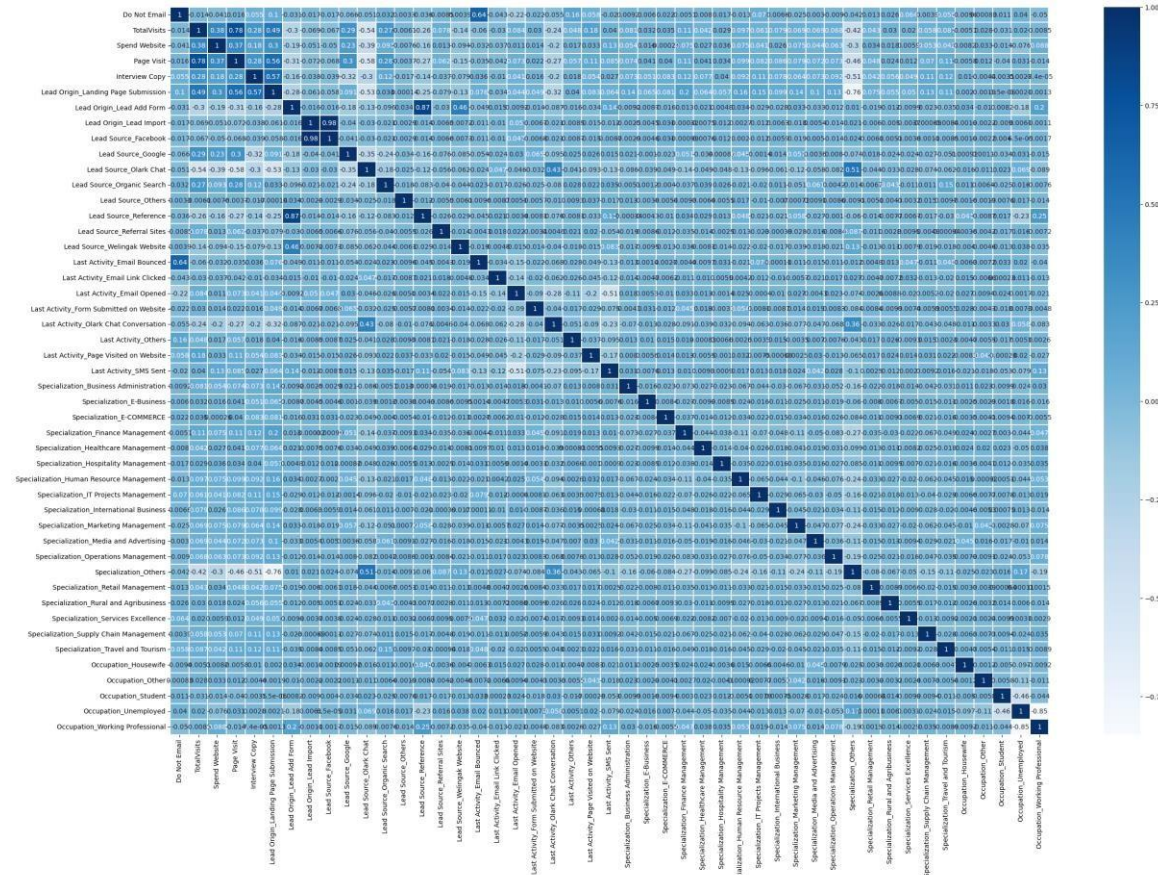


# EDA - Heatmap (All Numerical Columns)

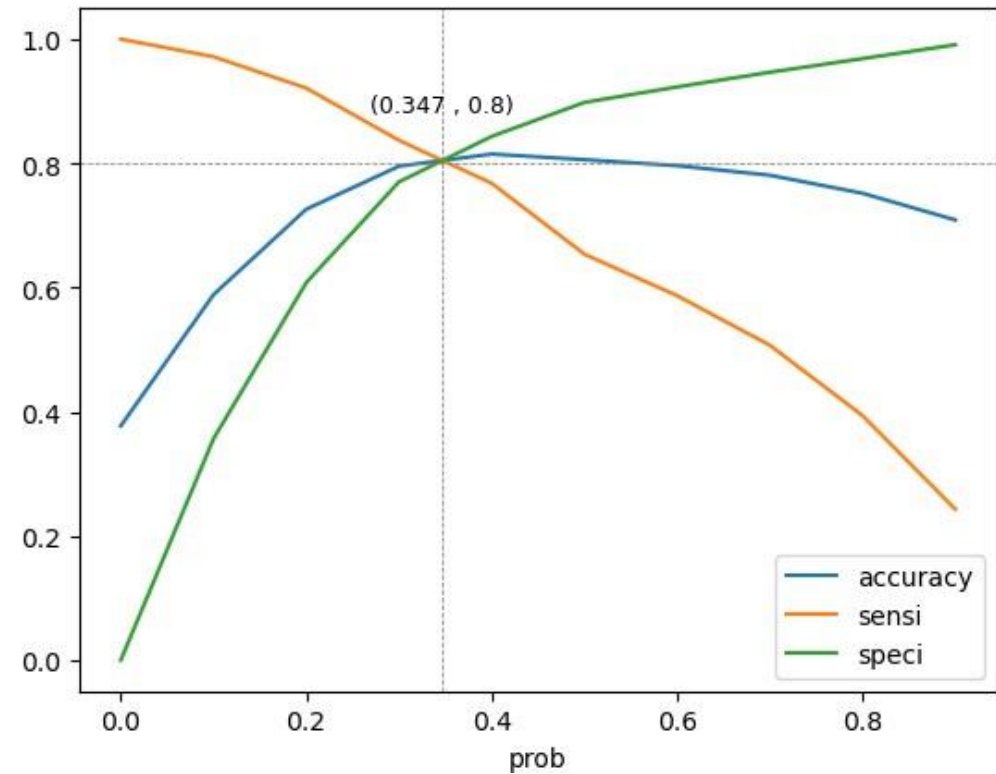
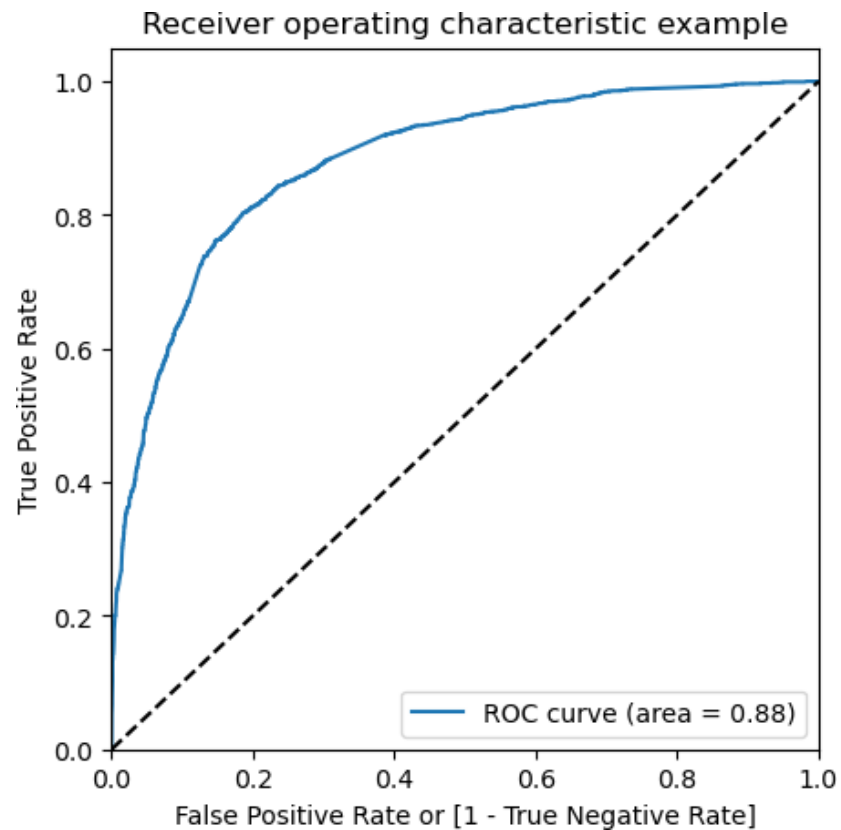
---



# EDA - Heatmap (Numerical & Dummy Columns)



# Linear Regression Final Model





# Conclusion

---

## Performance on Test Data

- Accuracy : 80.63%
- Sensitivity : 79.38%
- Specificity : 81.35%

# Conclusion

---

- The top three variables in the final model that most significantly contribute to the probability of a lead converting are
- Lead Source\_Welingak Website: 5.48
- Lead Source\_Reference: 3.23
- Last Activity\_SMS Sent: 2.80

# Recommendation

---

- **Focus on High-Potential Leads:** Leads from the following sources have a higher likelihood of conversion based on the coefficients: o Welingak Website: 5.48 o Reference: 3.23 o Working Professional: 2.64 Therefore, the sales team should prioritize calling leads from these sources during the intern-hiring period.
- **Leverage Effective Communication Channels:** Leads who have been sent SMS messages and have opened emails are more likely to convert. The coefficients for Last\_Activity\_Others and Last Activity\_Email Opened are 1.91 and 1.65, respectively. Thus, the sales team should prioritize calling leads who have been sent SMS messages or have opened emails from X Education.
- **Maximize Website Engagement:** Total Time Spent on the Website and Lead Source\_Facebook are the good indicator of a lead's interest in X Education's services, with a coefficient of 1.047 and 1.14. Hence, the sales team should also prioritize calling leads who have spent significant time on the website.
- **Maintain a Multi-Channel Approach:** Finally, the sales team should follow up with leads who have interacted with X Education through multiple channels. For example, leads who have used the Olark Chat feature on the website may not have spent much time on the website but may still be interested in X Education's services. Therefore, the sales team should ensure they follow up with leads who have used multiple channels to interact with X Education.

---

**Thank you**

