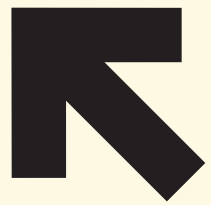
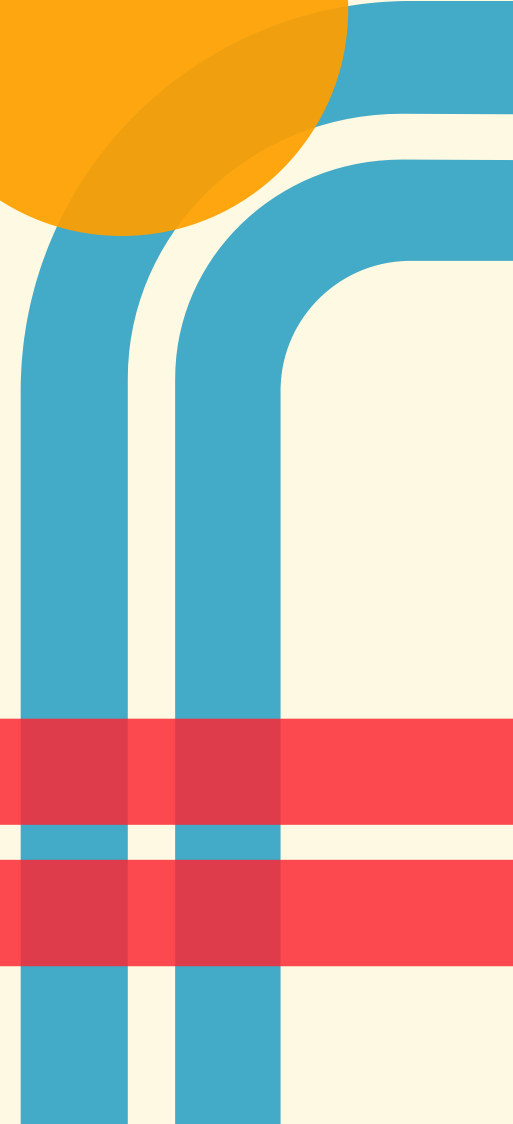
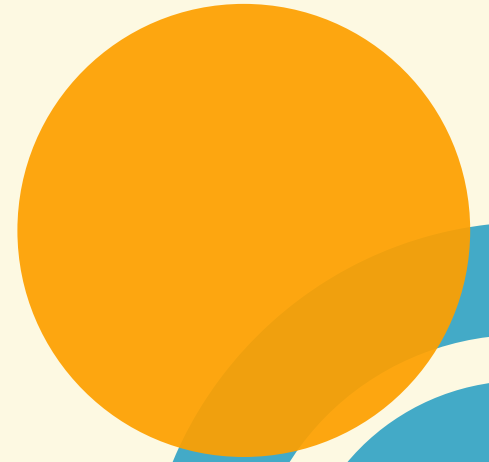




Bharathiseshan Nagarajan & Bhumika Jain



Lead Scoring Case Study



Agenda

Introduction

3

Solution Methodology

4

EDA

5

Analysis Outcomes

17

Conclusion

18



Introduction

Introduction to Problem Statement:

- An education company named X Education sells online courses to industry professionals.
- X Education get's so many leads but its lead conversion rate is not good. 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 i.e. 'Hot Leads'.
- We have to build a model to identify the most potential leads. If the generated leads got converted successfully, it will increase their lead conversion rate.

Business Objective:

- X Education wants to know most potential leads to increase the team efficiency.
- Target lead conversion rate to be around 80%.



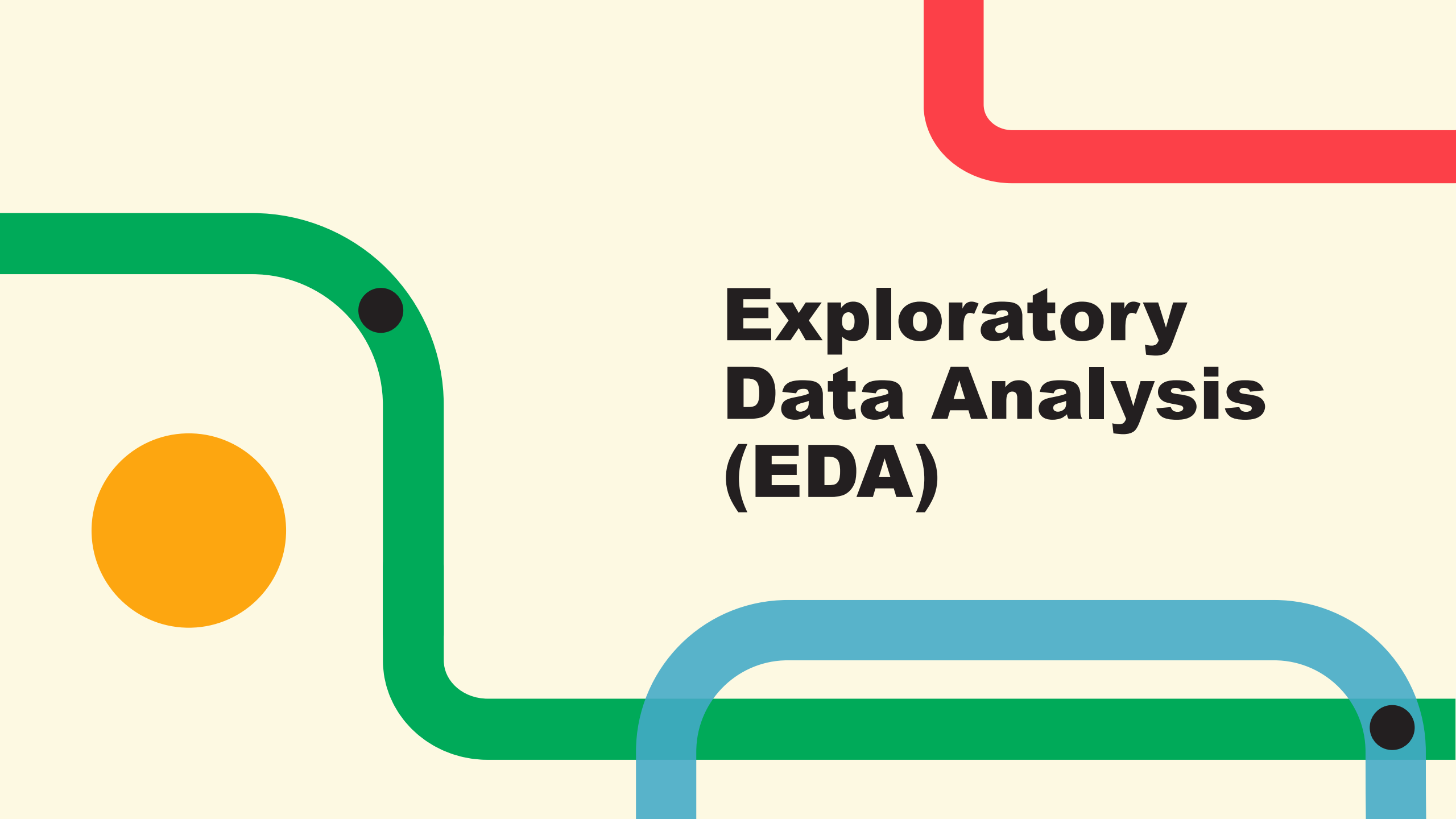


4

Solution Methodology

- Data cleaning and data manipulation
- Exploratory Data Analysis (EDA)
- Feature Scaling and Dummy Creation of the Data
- Classification Technique: Logistic Regression used for this case Study
- Train-Test Split
- Feature Rescaling
- Model Building
- Conclusion and Recommendation



An abstract graphic design on a light cream background. A thick green line enters from the left, curves 90 degrees down, and then continues horizontally to the right. A thick blue line enters from the bottom, curves 90 degrees up, and then continues horizontally to the left, overlapping the green line. A thick red line enters from the top right and curves 90 degrees left, continuing horizontally to the left. A solid orange circle is positioned to the left of the green line's first curve. Two small black dots are located on the green line: one on the upper curve and one on the horizontal segment to the right of the blue line's curve.

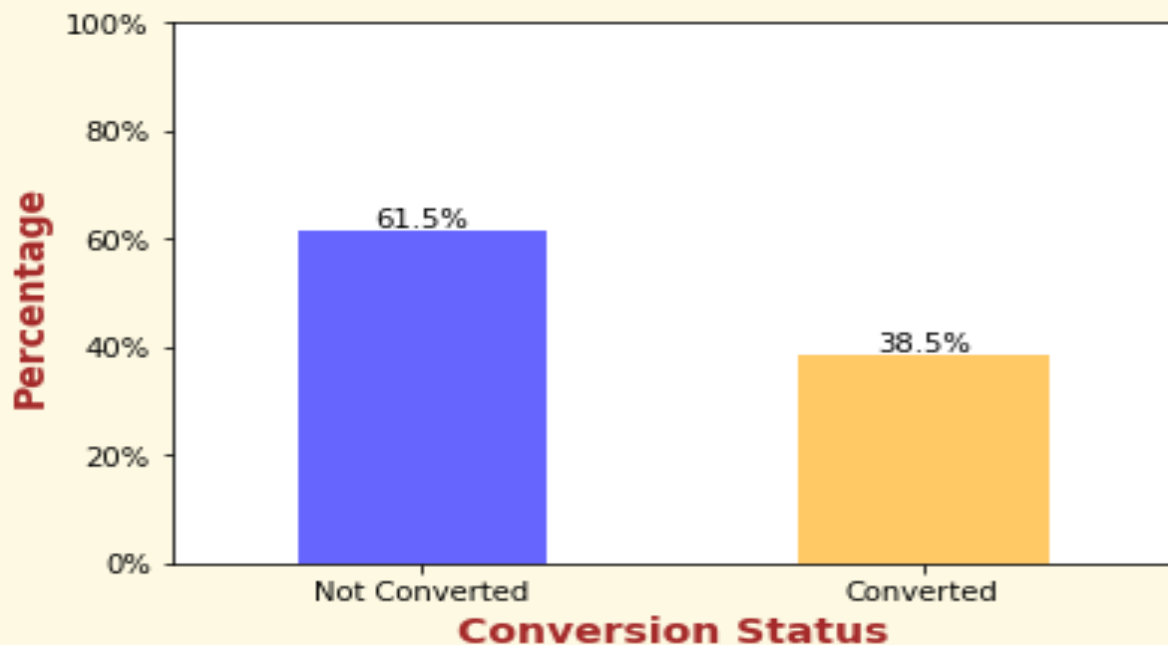
Exploratory Data Analysis (EDA)



6

Lead Conversion Rate

Leads Conversion Rate



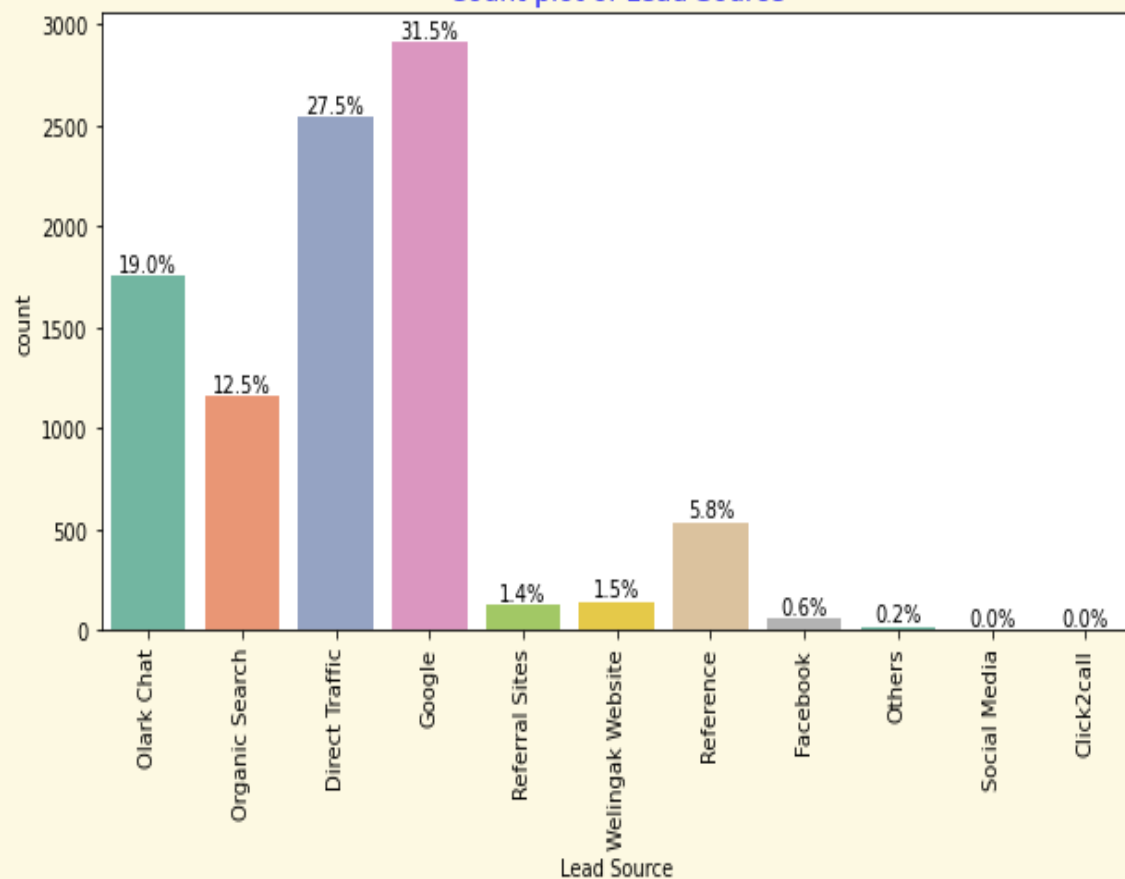
Previously the converted leads were 38.5% which was really low for increasing the lead conversion rate we will build a model which will define most important variables which will help us increasing this lead conversion rate.



7

Categorical Variable Analysis

Count plot of Lead Source

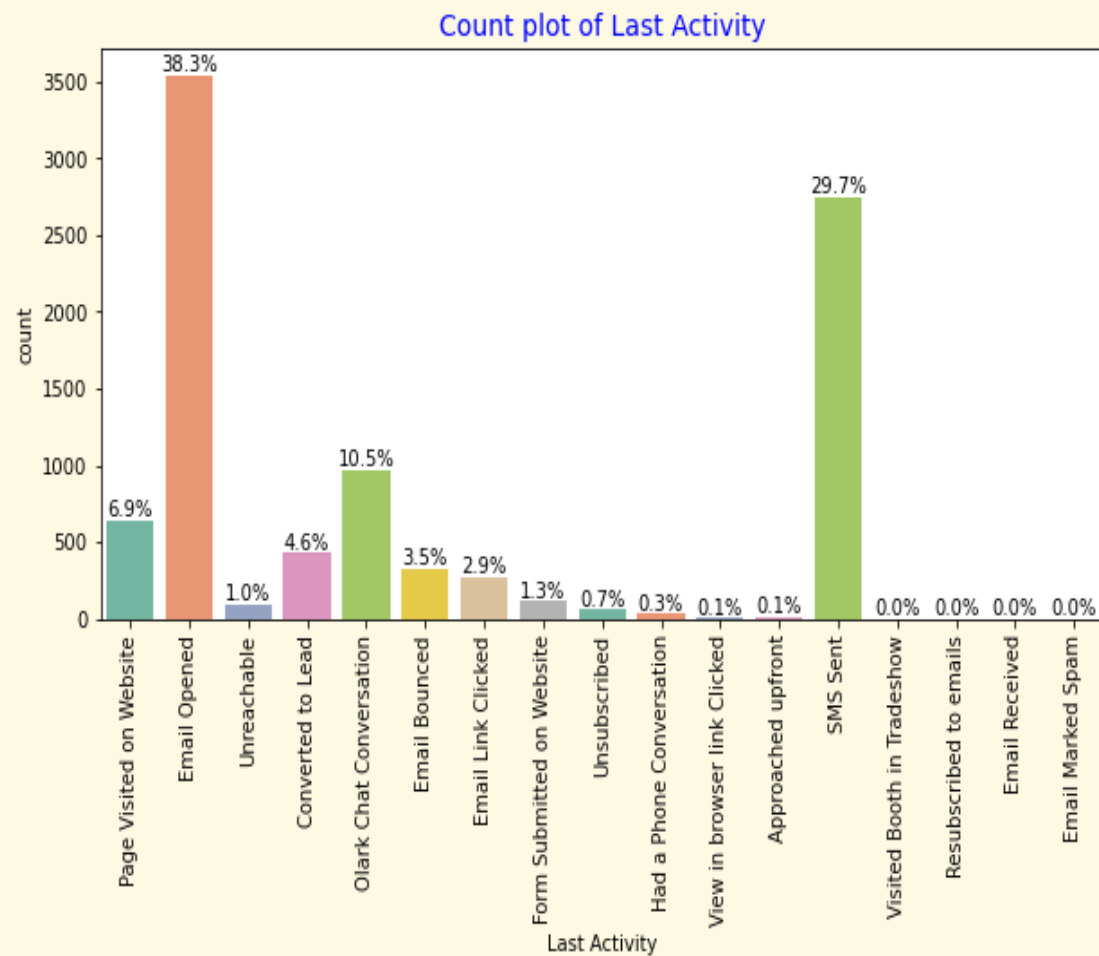


Google, Direct Traffic, and Olark chat conversation are the lead source which contributes most.



8

Categorical Variable Analysis



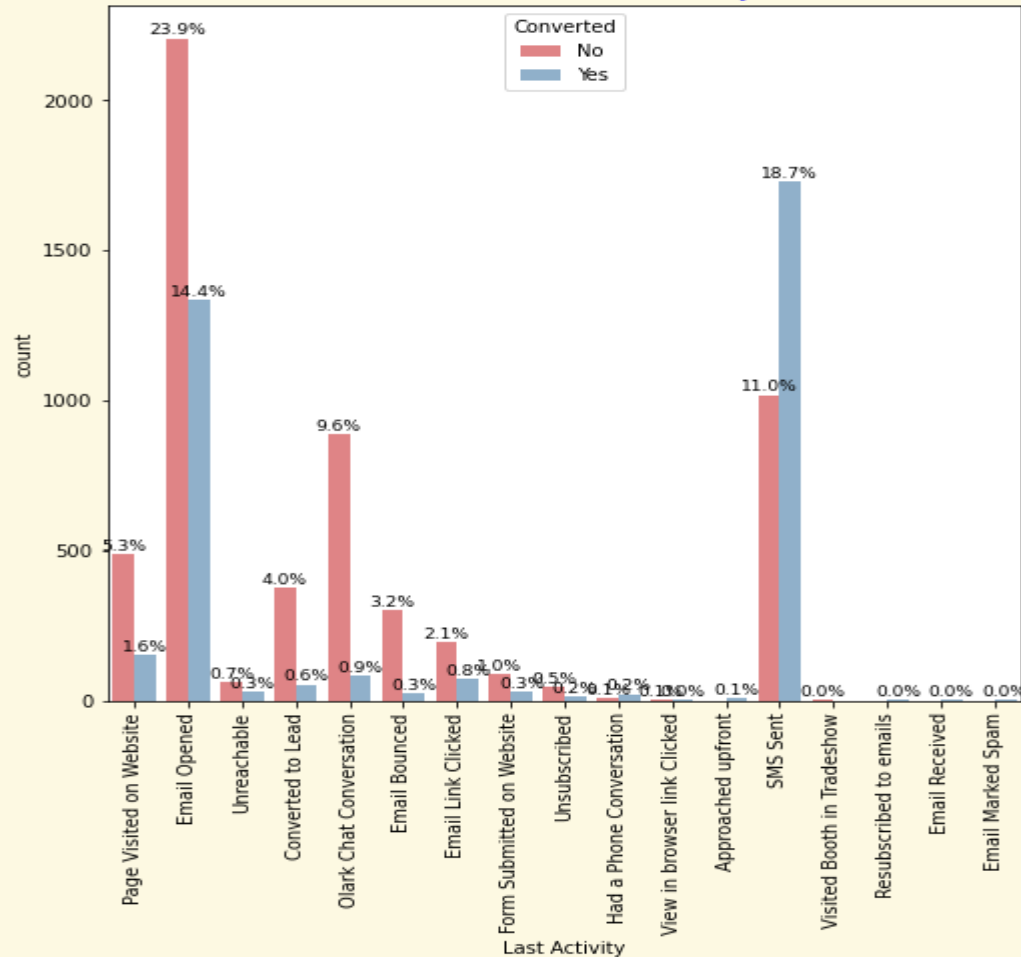
In last activity SMS sent and Email opened are most important variables.



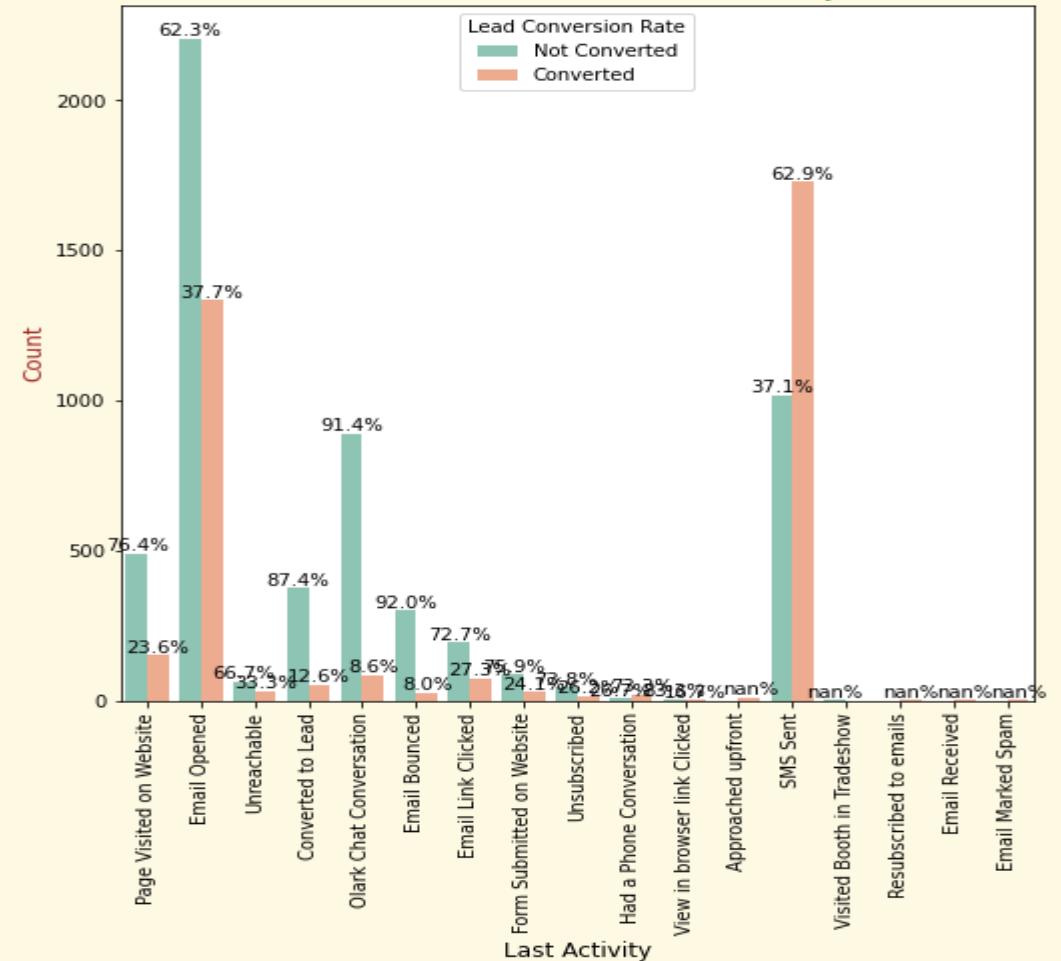
9

Bivariate Analysis

Distribution of Last Activity



Lead Conversion Rate of Last Activity

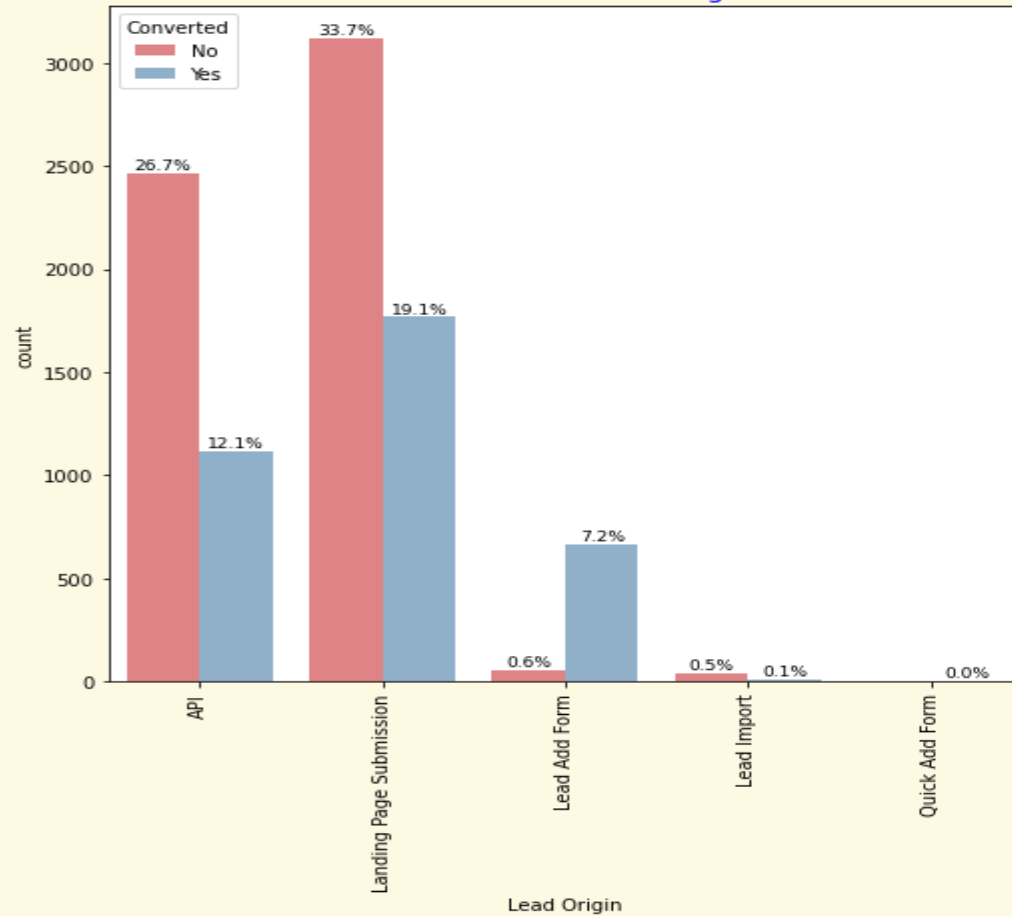




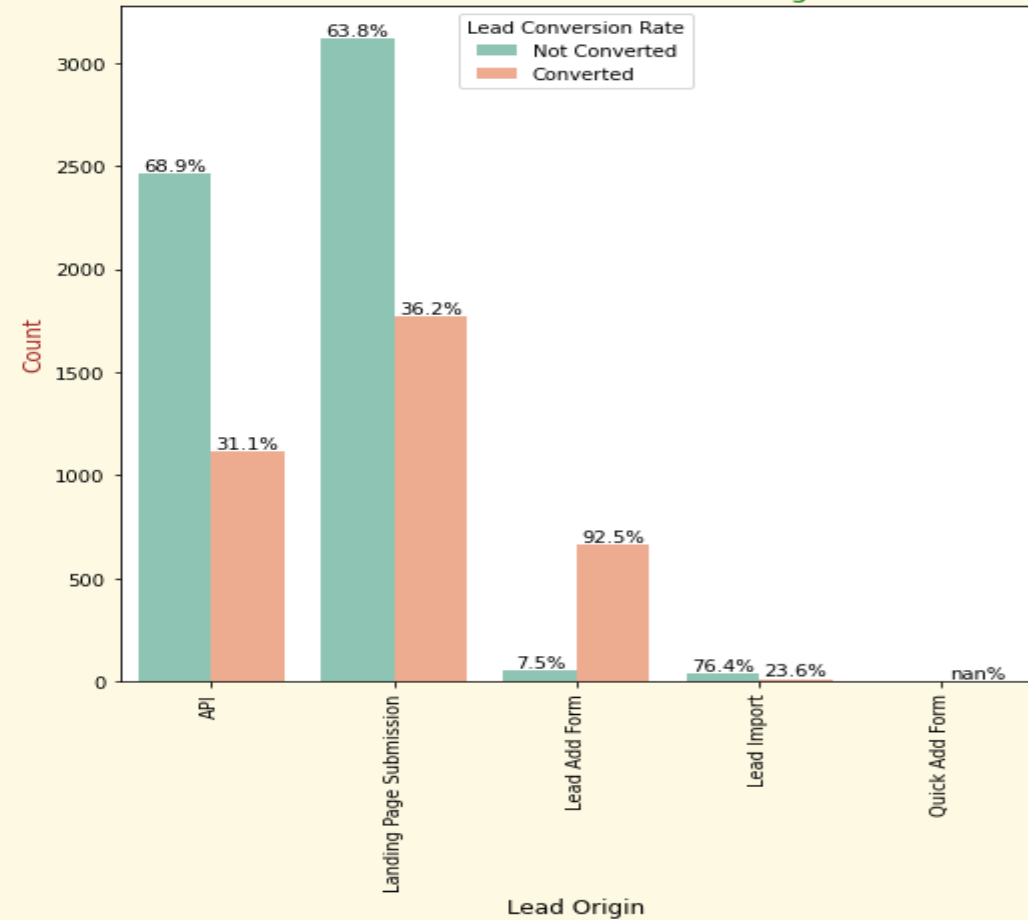
10

Bivariate Analysis

Distribution of Lead Origin



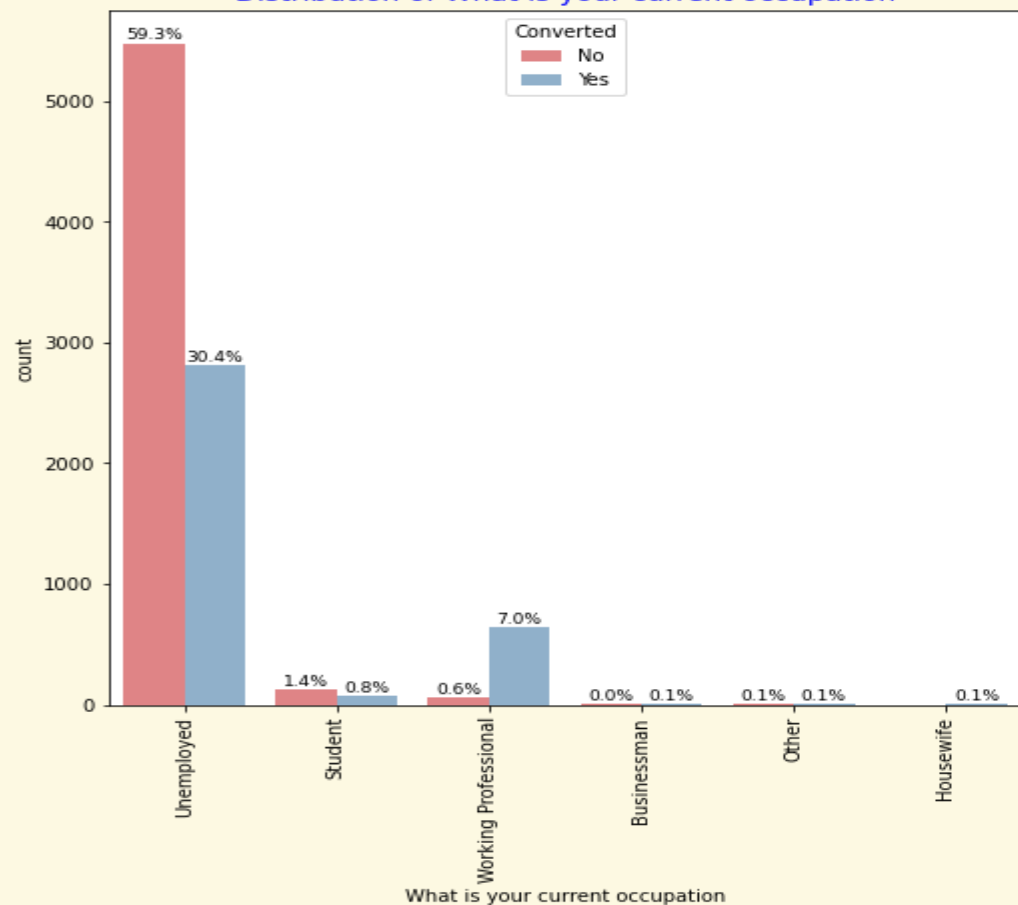
Lead Conversion Rate of Lead Origin



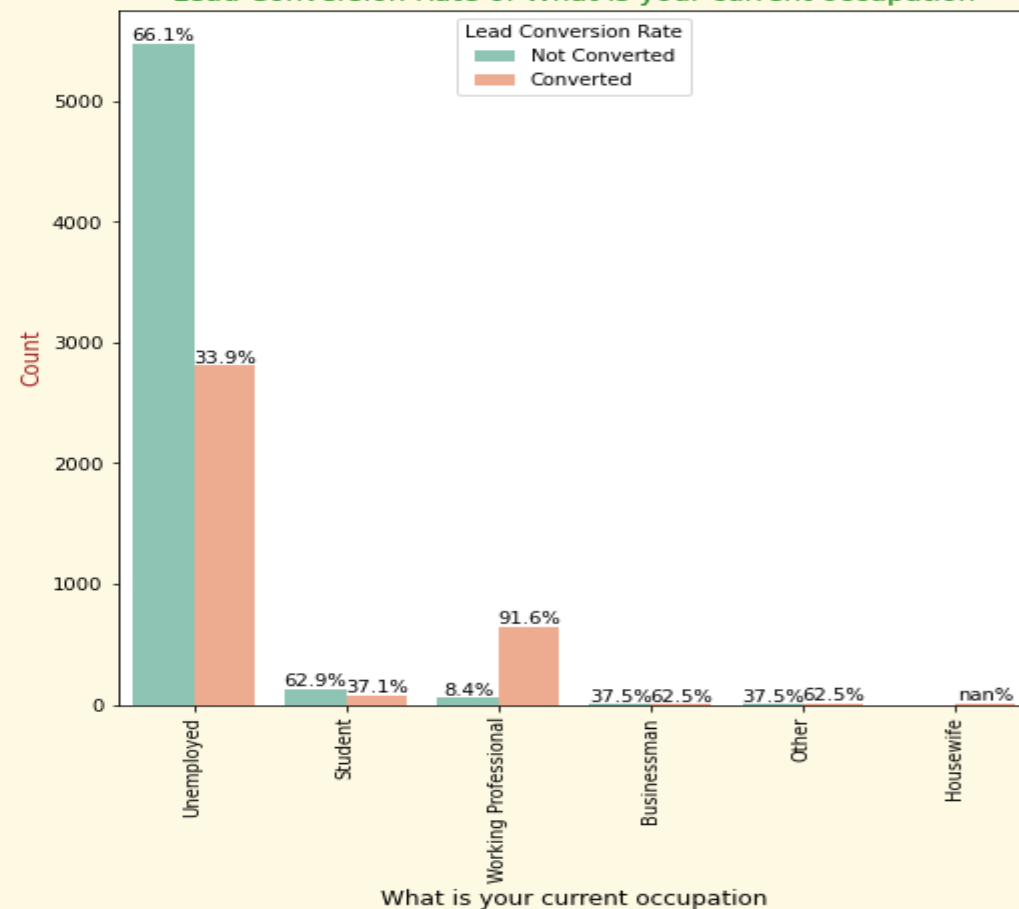
**11**

Bivariate Analysis

Distribution of What is your current occupation



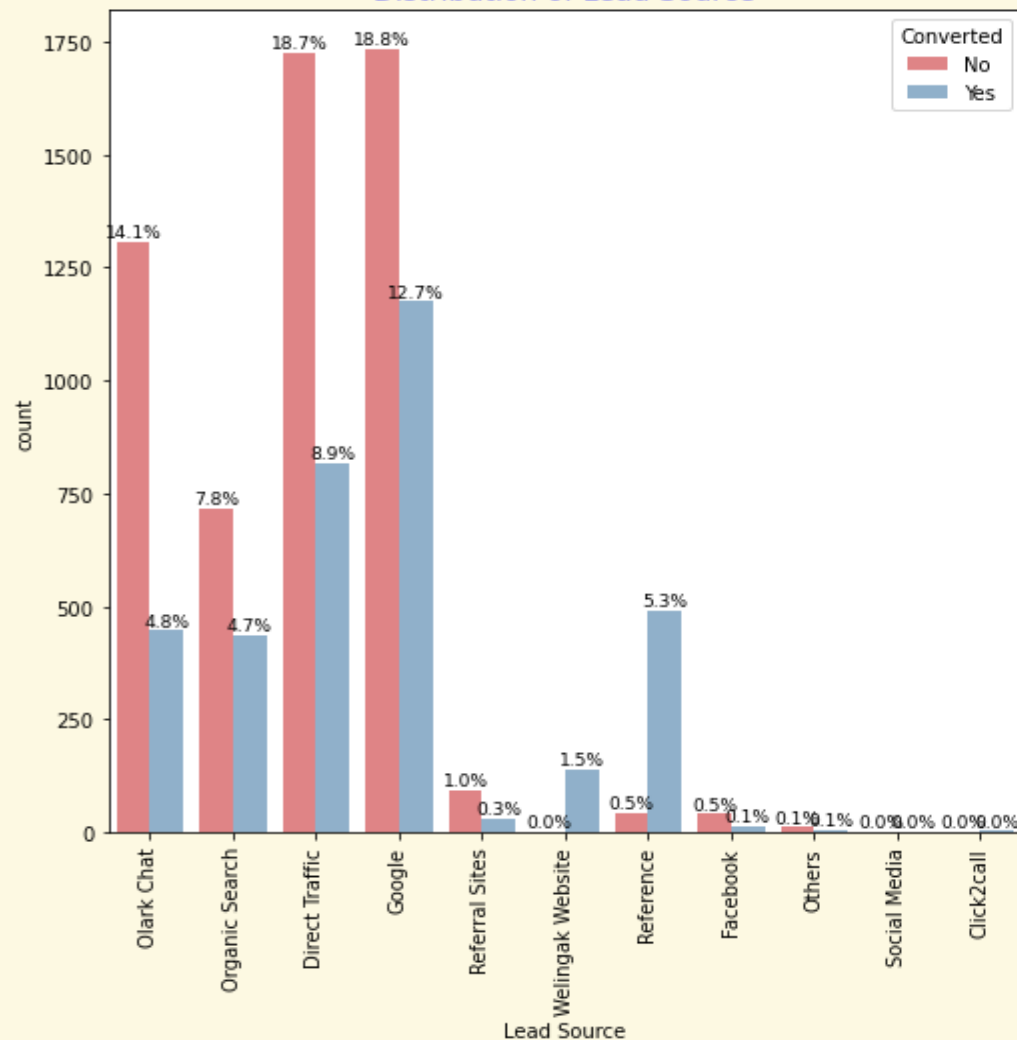
Lead Conversion Rate of What is your current occupation



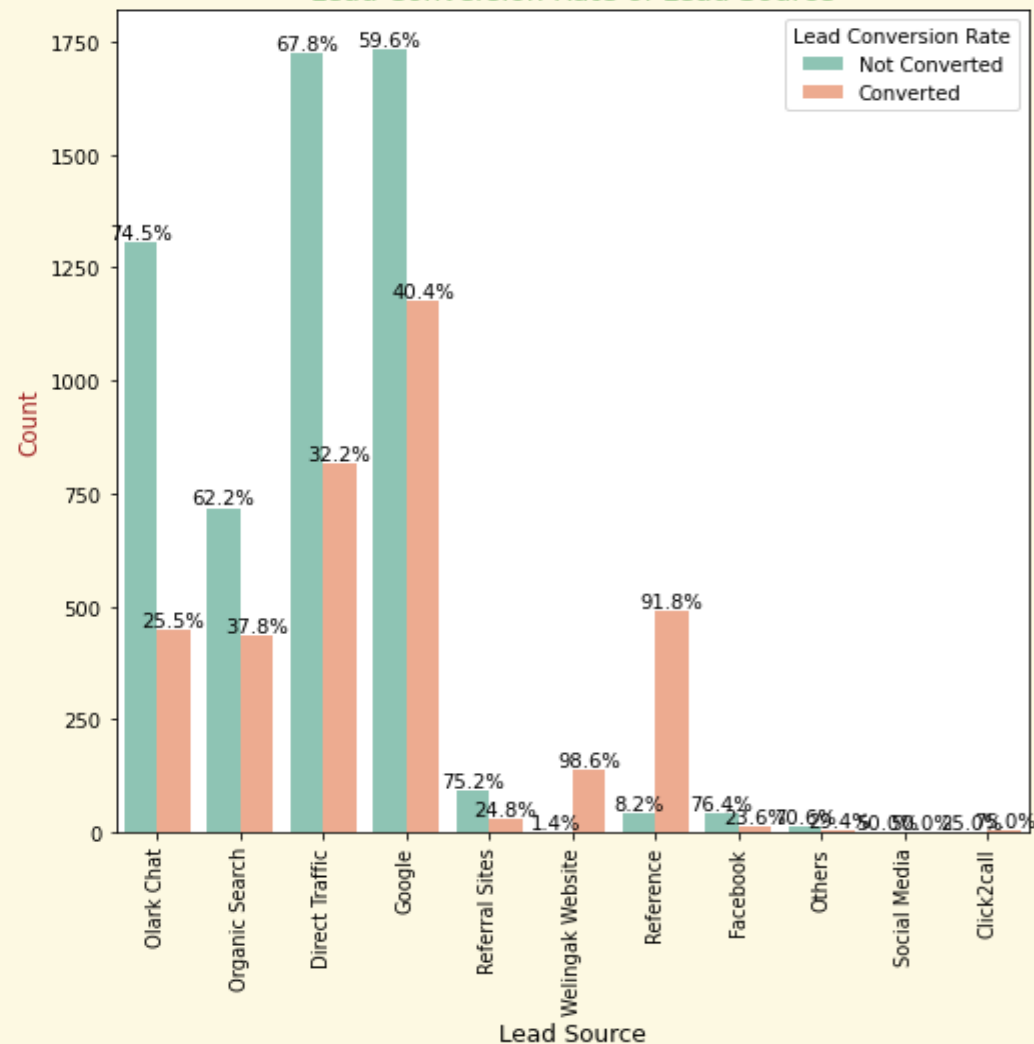
**12**

Bivariate Analysis

Distribution of Lead Source



Lead Conversion Rate of Lead Source

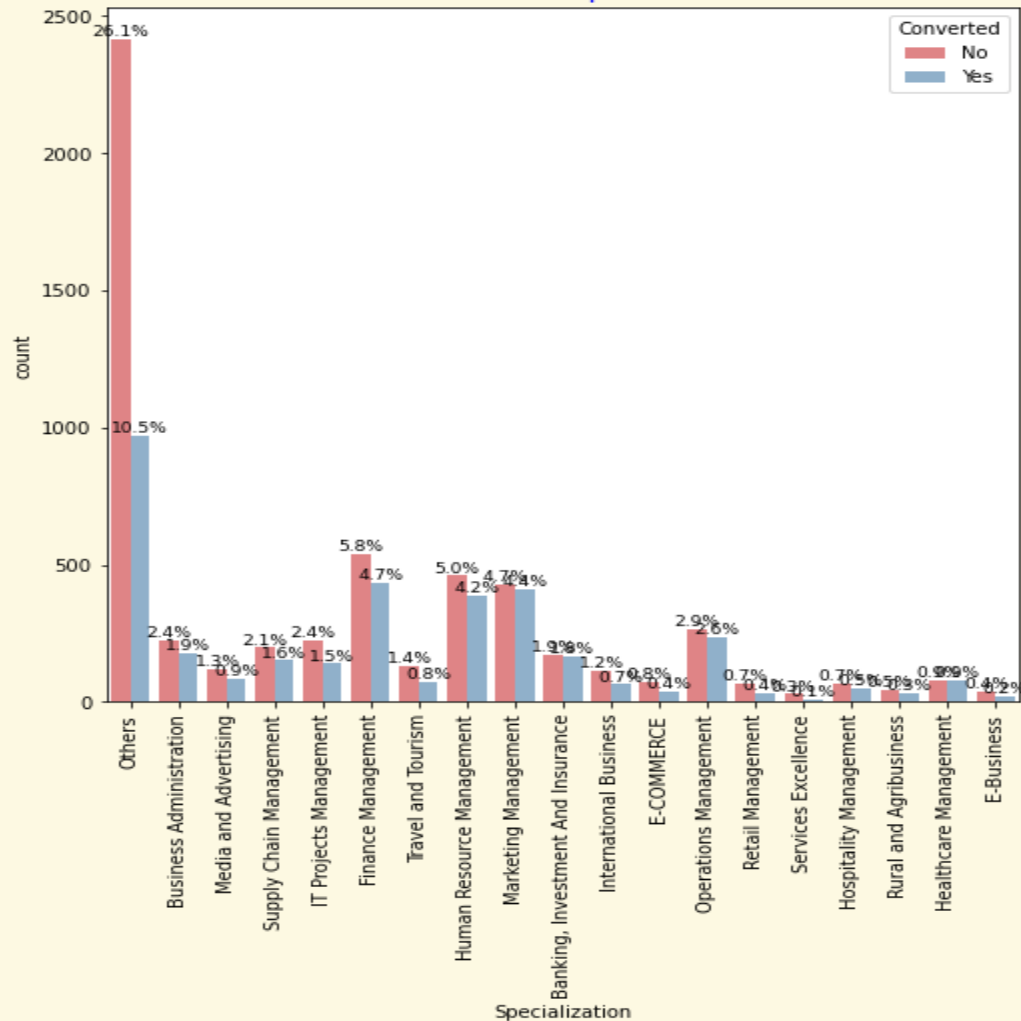




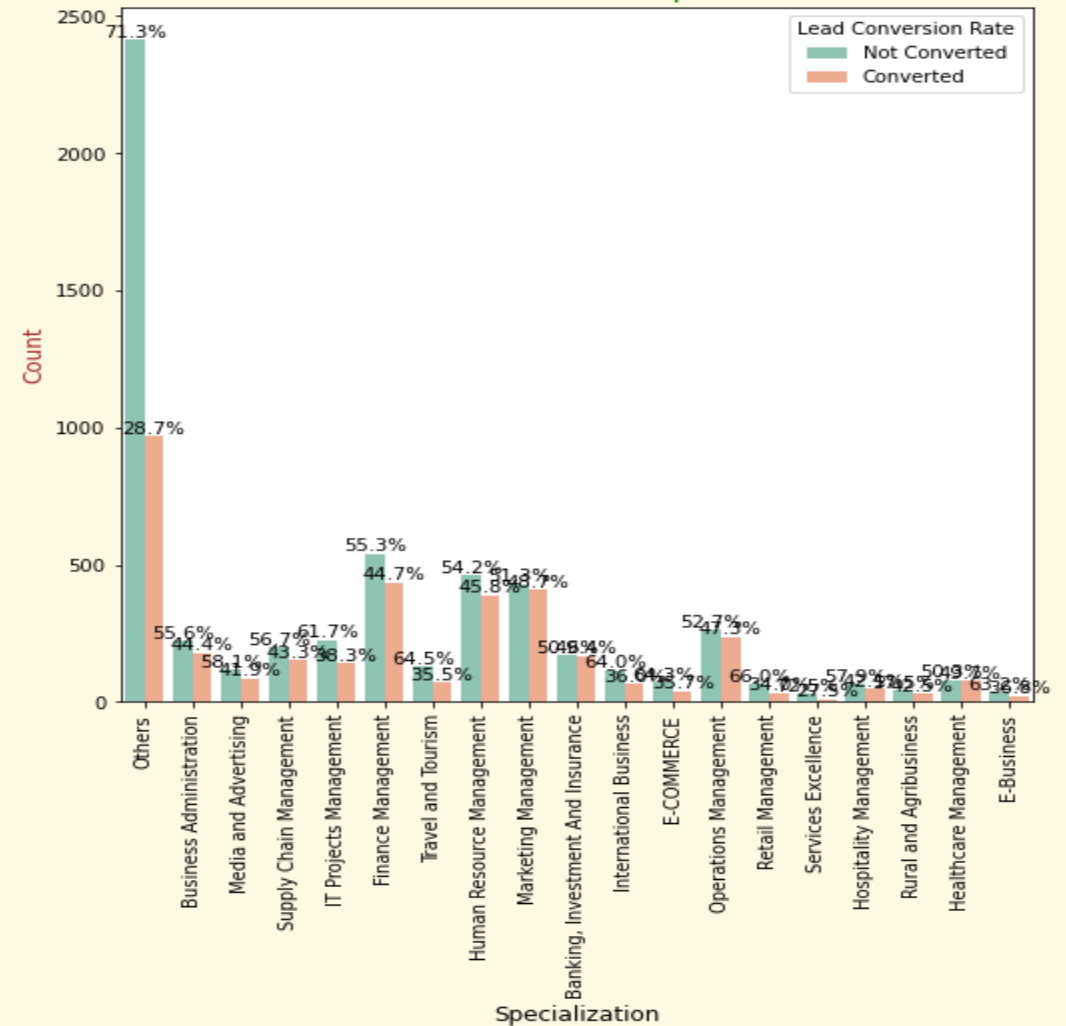
13

Bivariate Analysis

Distribution of Specialization

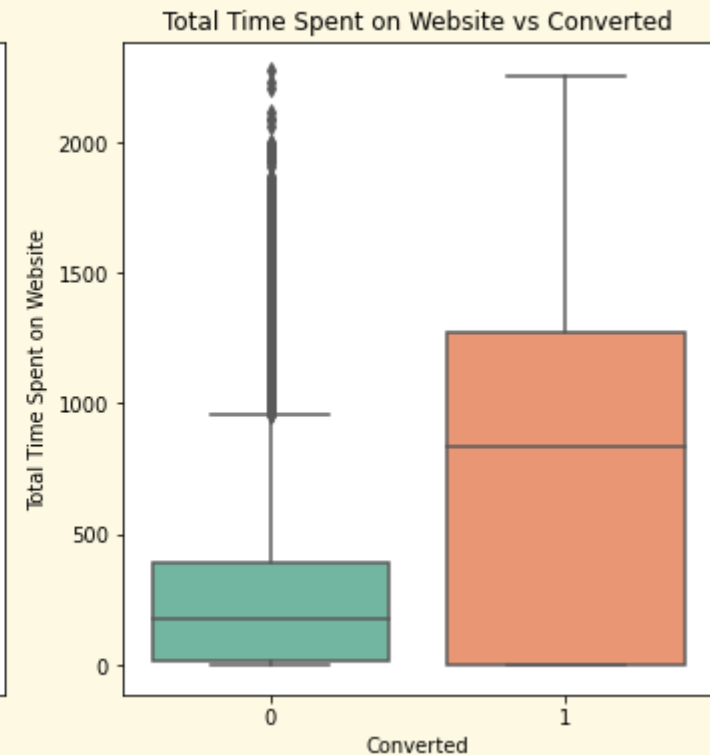
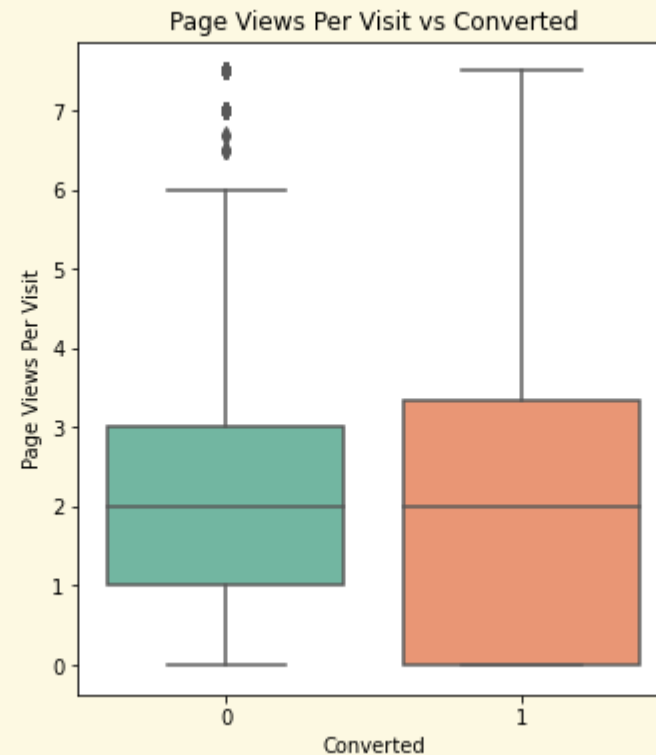
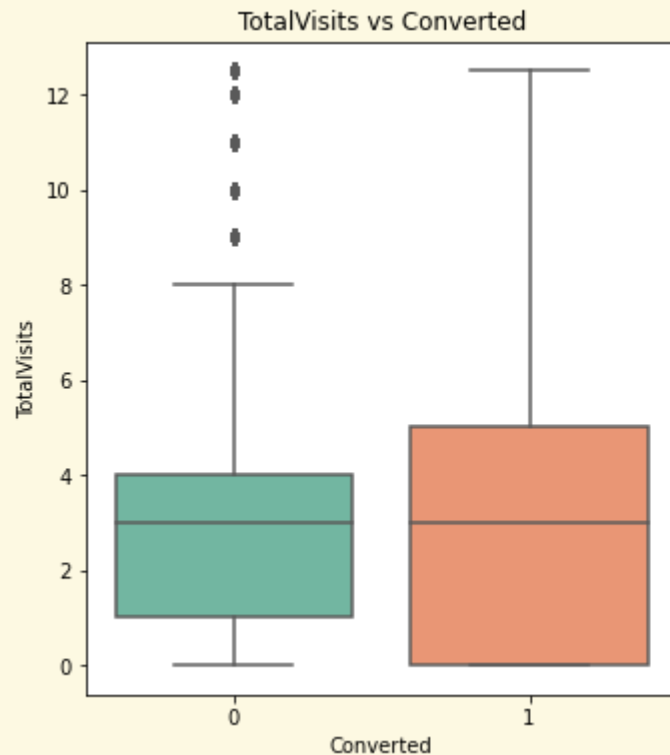


Lead Conversion Rate of Specialization



**14**

Bivariate Analysis

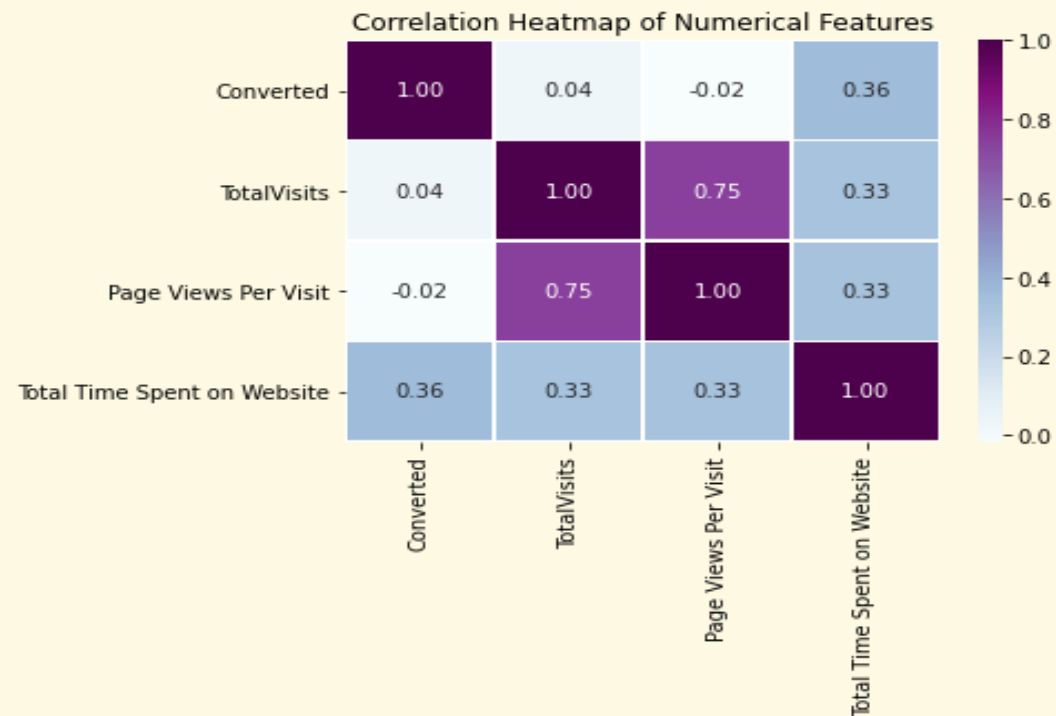
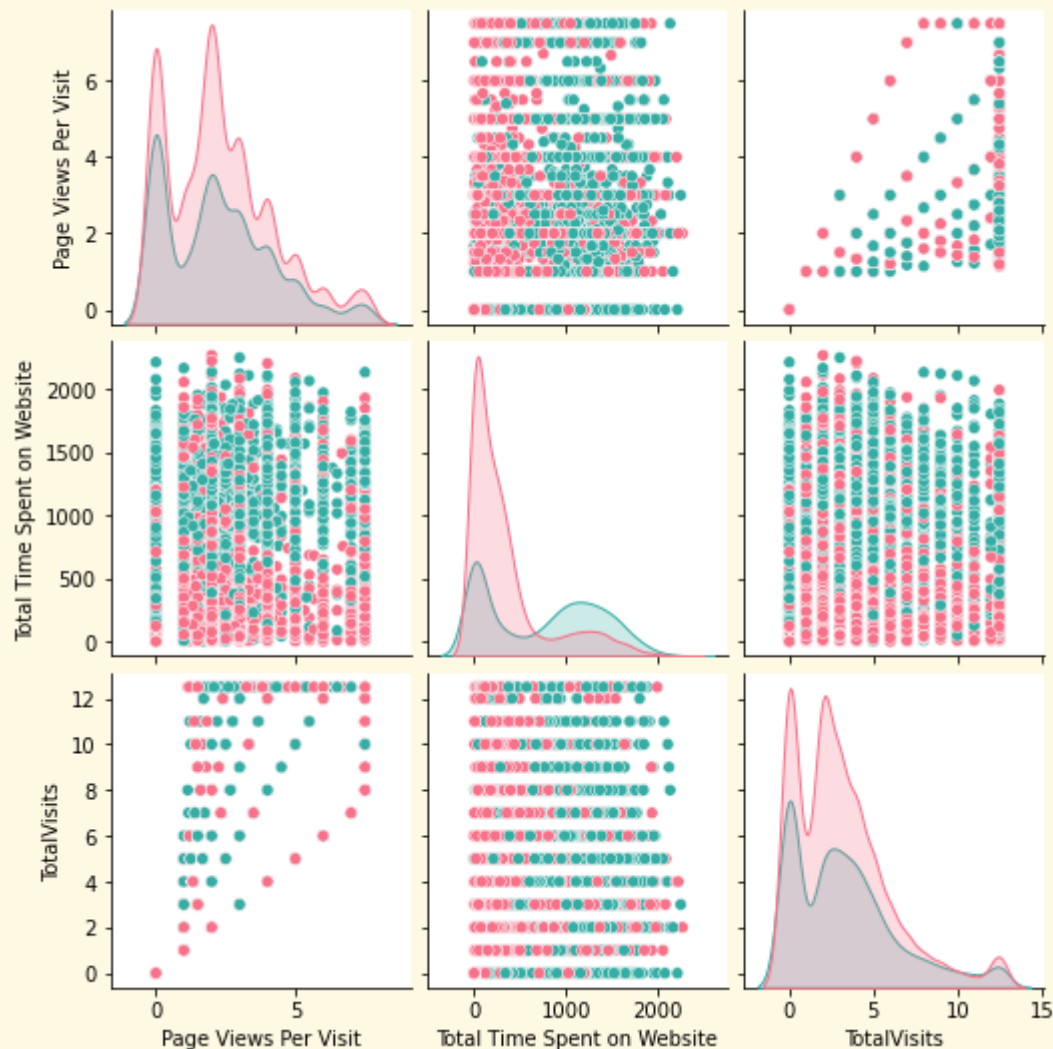


The boxplot reveals that leads who spend more time on the website are more likely to be successfully converted compared to those who spend less time.



15

Multivariate Analysis

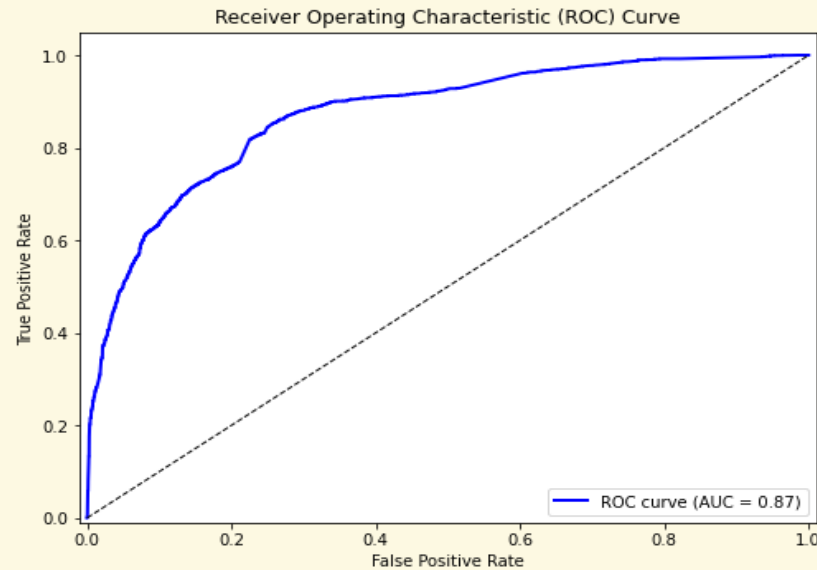
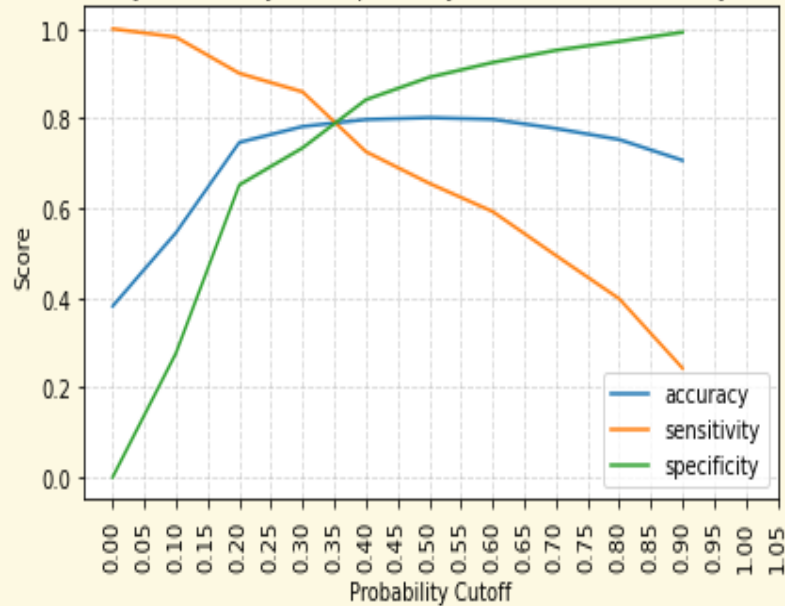


Total Visits and Page views per visit are highly correlated.

**16**

ROC Curve

Accuracy, Sensitivity, and Specificity for Different Probability Cutoffs



Insights

- **The optimal cut off point is where we get balanced sensitivity and specificity.**
- **From first graph Optimal Probability cutoff is 0.35.**
- **ROC curve is giving the area coverage of 87%.**



17

Analysis outcomes

- Approximately 52% of leads originated through "Landing Page Submission," boasting a lead conversion rate (LCR) of 36%. Meanwhile, the "API" detected roughly 39% of customers, yielding a lead conversion rate (LCR) of 31%.
- Approximately 90% of customers are categorized as Unemployed, demonstrating a lead conversion rate (LCR) of 34%. Conversely, Working Professionals comprise only 7.6% of the total customer base but exhibit an almost 92% lead conversion rate (LCR).
- 92% of individuals have indicated their preference not to receive emails regarding the course.
- Google achieves a lead conversion rate (LCR) of 40.4% from 31% of its customers. Direct Traffic, on the other hand, yields a slightly lower LCR of 32% with 27% of customers, which is less than that of Google. Organic Search accounts for 37.8% of LCR, but it is represented by only 12.5% of customers. In comparison, Reference demonstrates an impressive LCR of 91%, albeit with a smaller customer base of around 6%.
- The feature 'SMS Sent' exhibits a notably high lead conversion rate of around 63%.
- Marketing Management, HR Management, Finance Management shows good conversion rates.



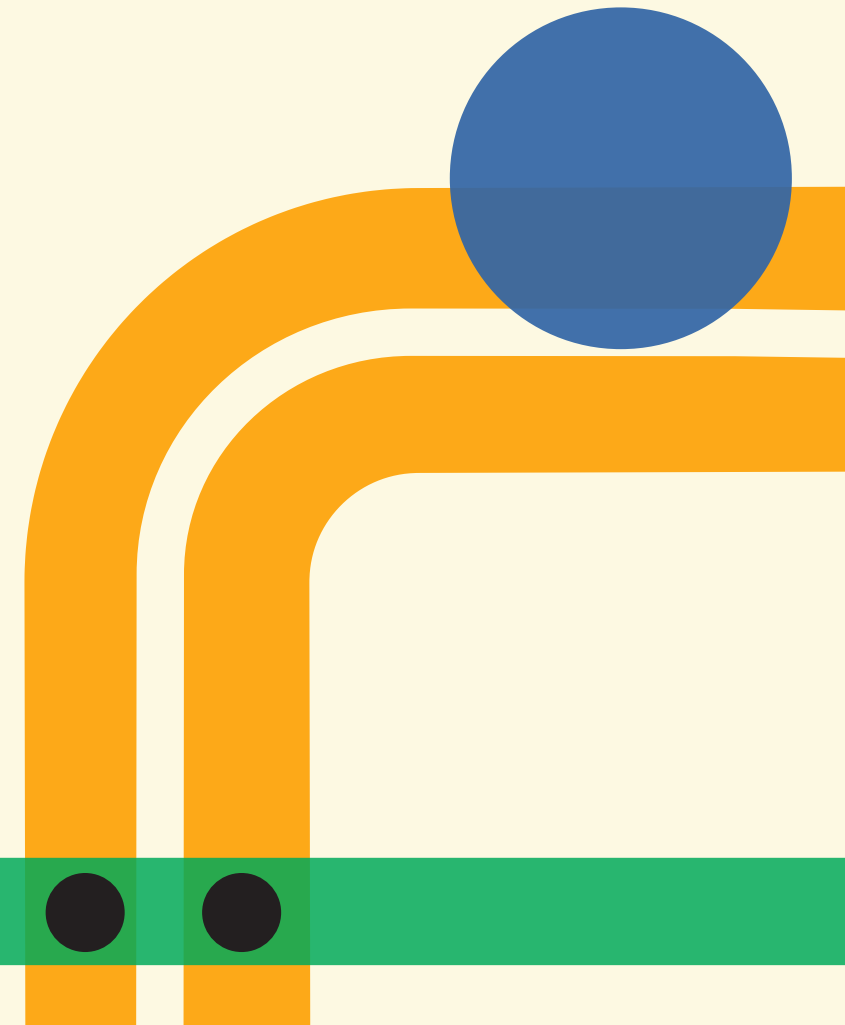
18

Conclusion

It was found that the most important variables that count more in the potential buyers:

- When Lead Source:
 - a. Welingak Website
 - b. Reference
- When current occupation is Working Professional
- The total time spent on the website
- Total number of visits
- When the last activity was:
 - a. SMS sent
 - b. Olark chat conversation
 - c. Converted to lead
 - d. Had a phone conversation
- When lead origin is landing page submission

With all these strong information X Education can increase their lead conversion rate and focus more on potential leads and decrease their business costing.



No matter what you do, Don't pass leads directly to the sales team.

Joan Miller (Marketo)

