

LEAD SCORING

CASE STUDY

X Education implemented a strategy to identify hot leads, prioritizing them for increased attention and resources in order to improve the conversion ratio.



BY:-

Samiksha Bansod

TABLE OF CONTENT

- About Company
- Problem Statement
- Goals of the Case Study
- Data Understanding
- Proposed Solution
- Implementation
- Visualization
- Inference
- Conclusion
- Recommendations

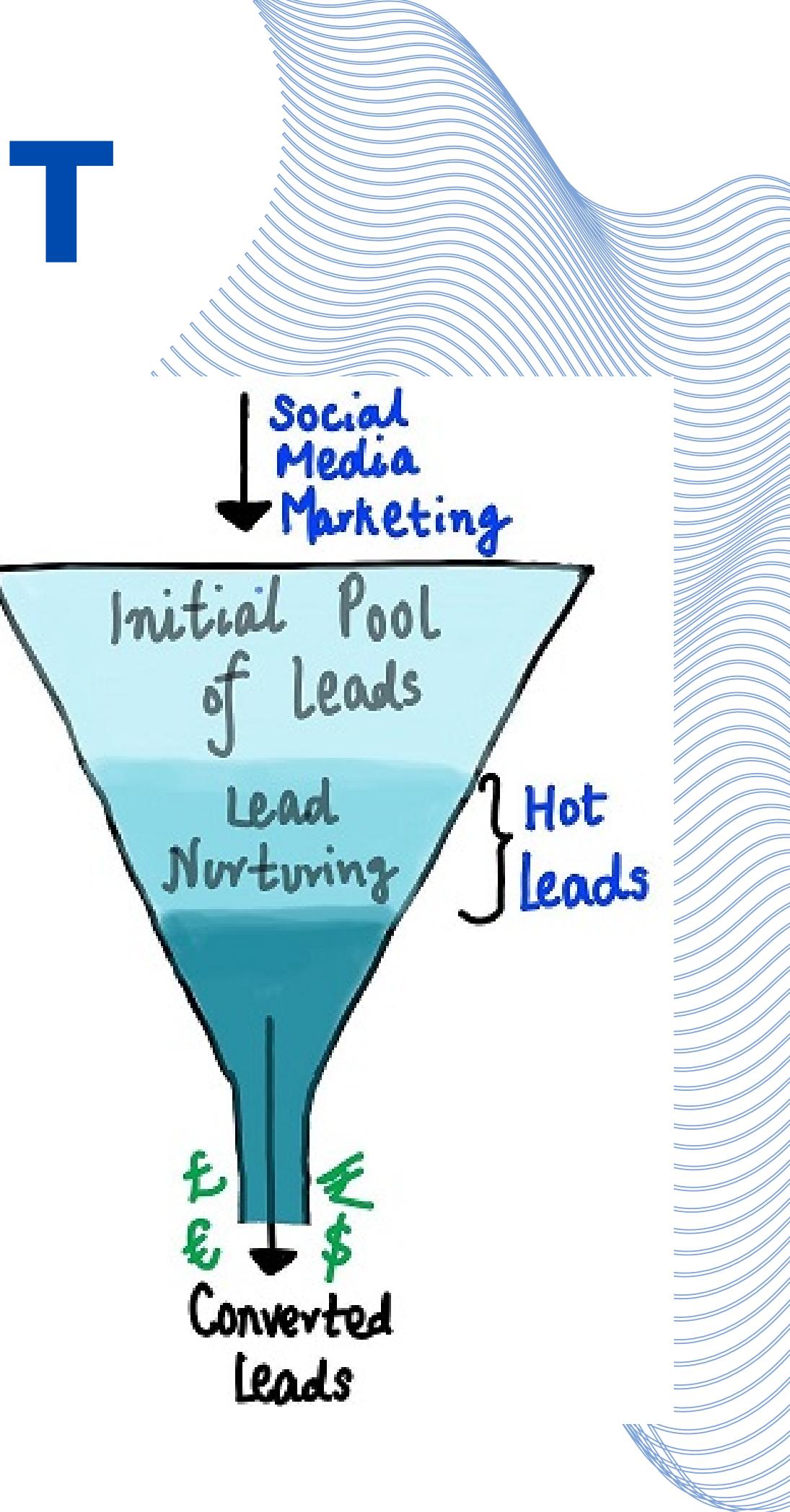


ABOUT COMPANY

- An education company named X Education sells online courses to industry professionals.
- Many professionals who are interested in the courses land on their websites and browse for courses.
- The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- When these people fill up a form providing their email address or phone number, they are classified to be a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.
- Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around 30%.

PROBLEM STATEMENT

- Although X Education gets a lot of leads, its lead conversion rate is very poor.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If they successfully identify this set of leads, the lead conversion rate should go up.
- 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.



GOALS OF THE CASE STUDY

- To 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.
- To adjust to if the company's requirement changes in the future so you will need to handle these as well.



DATA UNDERSTANDING

- We have been provided with a leads dataset from the past with around 9000 data points. (9240, 37)
- This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc.
- The target variable, in this case, is the column 'Converted' which tells whether a past lead was converted or not.

	Prospect ID	Lead Number	Lead Origin	Lead Source	Do Not Email	Do Not Call	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	...	Get updates on DM Content	Lead Profile	City	Asymmetrique Activity Index
0	7927b2df-8bba-4d29-b9a2-b6e0beafe620	660737	API	Olark Chat	No	No	0	0.0	0	0.0	...	No	Select	Select	02.Medium
1	2a272436-5132-4136-86fa-dcc88c88f482	660728	API	Organic Search	No	No	0	5.0	674	2.5	...	No	Select	Select	02.Medium
2	8cc8c611-a219-4f35-ad23-fdfd2656bd8a	660727	Landing Page Submission	Direct Traffic	No	No	1	2.0	1532	2.0	...	No	Potential Lead	Mumbai	02.Medium
3	0cc2df48-7cf4-4e39-9de9-19797f9b38cc	660719	Landing Page Submission	Direct Traffic	No	No	0	1.0	305	1.0	...	No	Select	Mumbai	02.Medium
4	3256f628-e534-4826-9d63-4a8b88782852	660681	Landing Page Submission	Google	No	No	1	2.0	1428	1.0	...	No	Select	Mumbai	02.Medium

5 rows x 37 columns

PROPOSED SOLUTION

1

SELECTION OF HOT LEADS

We cluster the leads into certain categories based on their tendency or probability to convert, thus, getting a smaller section of hot leads to focus more on.

2

COMMUNICATING WITH HOT LEADS

Since we would have a smaller set of leads to have communication with, we might make more impact with effective communication.

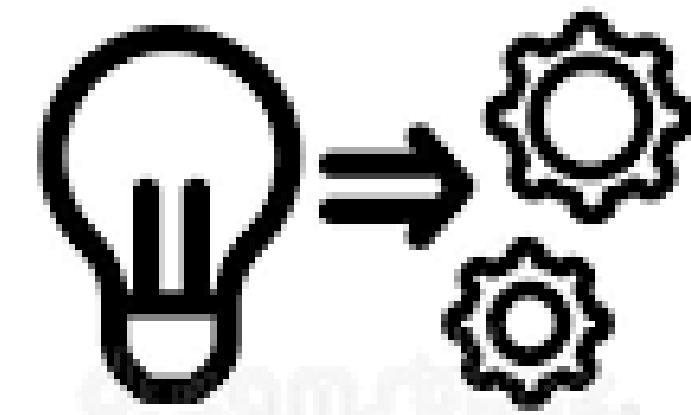
3

CONVERSION OF HOT LEADS

Since we focussed on hot leads, which were more probable to convert, we would have a better conversion rate, and hence we can achieve the 80% target

IMPLEMENTATION

Implementing the proposed solution in a proper format to build a Linear Regression Model.



IMPLEMENTATION

STEPS FOR MODEL BUILDING

Loading & Observing the data provided by the Company

Univariate, Bivariate, and Heatmap for numerical and categorical columns

Performing pre-requisites for RFE and Logistic Regression



Data Gathering

Data Cleaning

Performing EDA

Data Preparation

Model Building

Duplicate removal, null value treatment, unnecessary column elimination, etc.

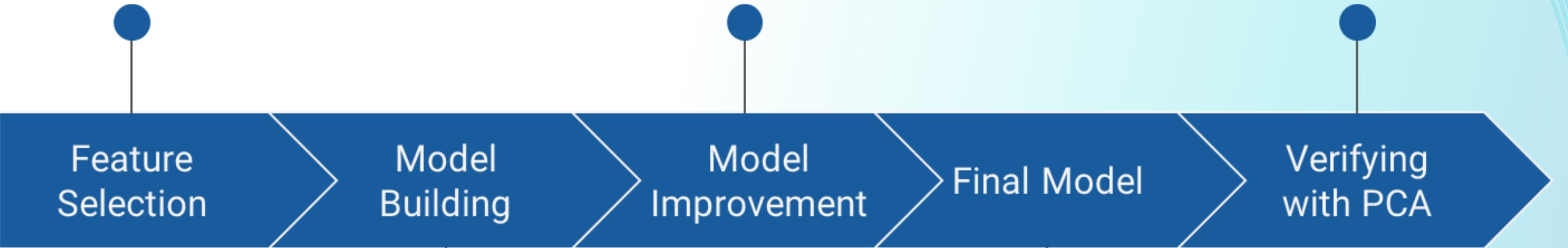
Outlier Treatment, Feature-Selection, Standardization, etc.

STEPS FOR MODEL BUILDING

Selection of top features using RFE and VIF

Reduction of columns and Model re-building

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

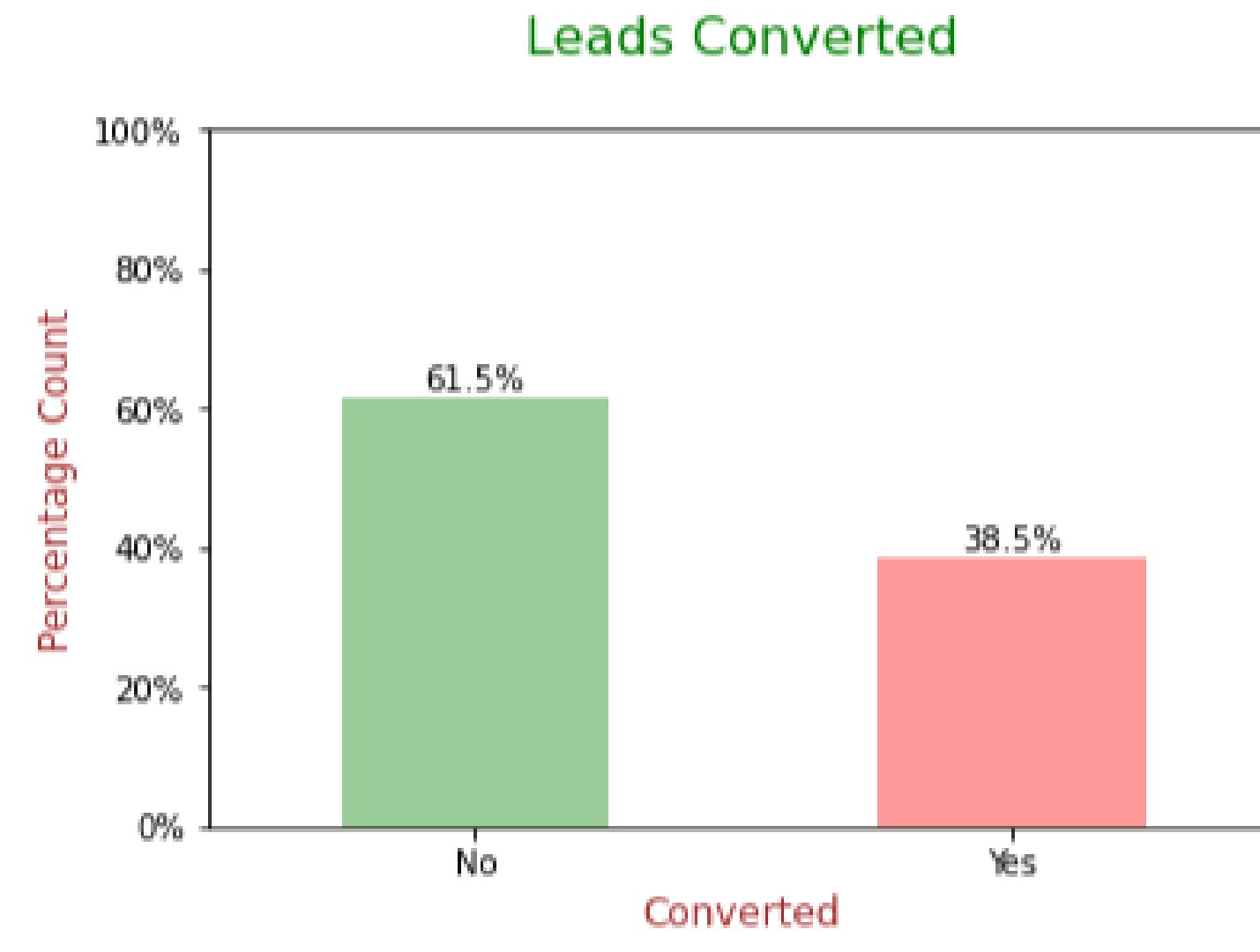


Model building using RFE for selected columns

Final Model Analysis and performance on Test Data

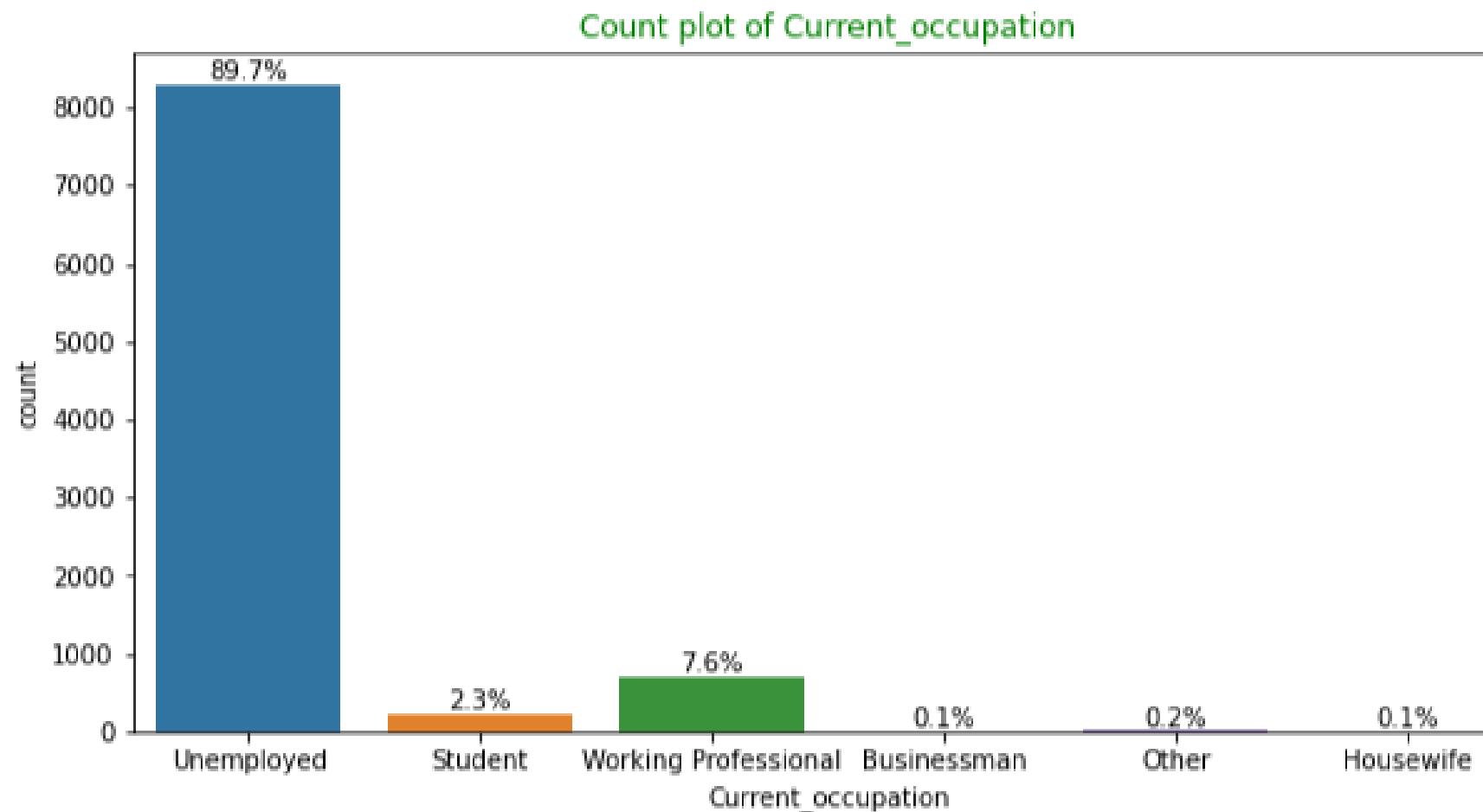
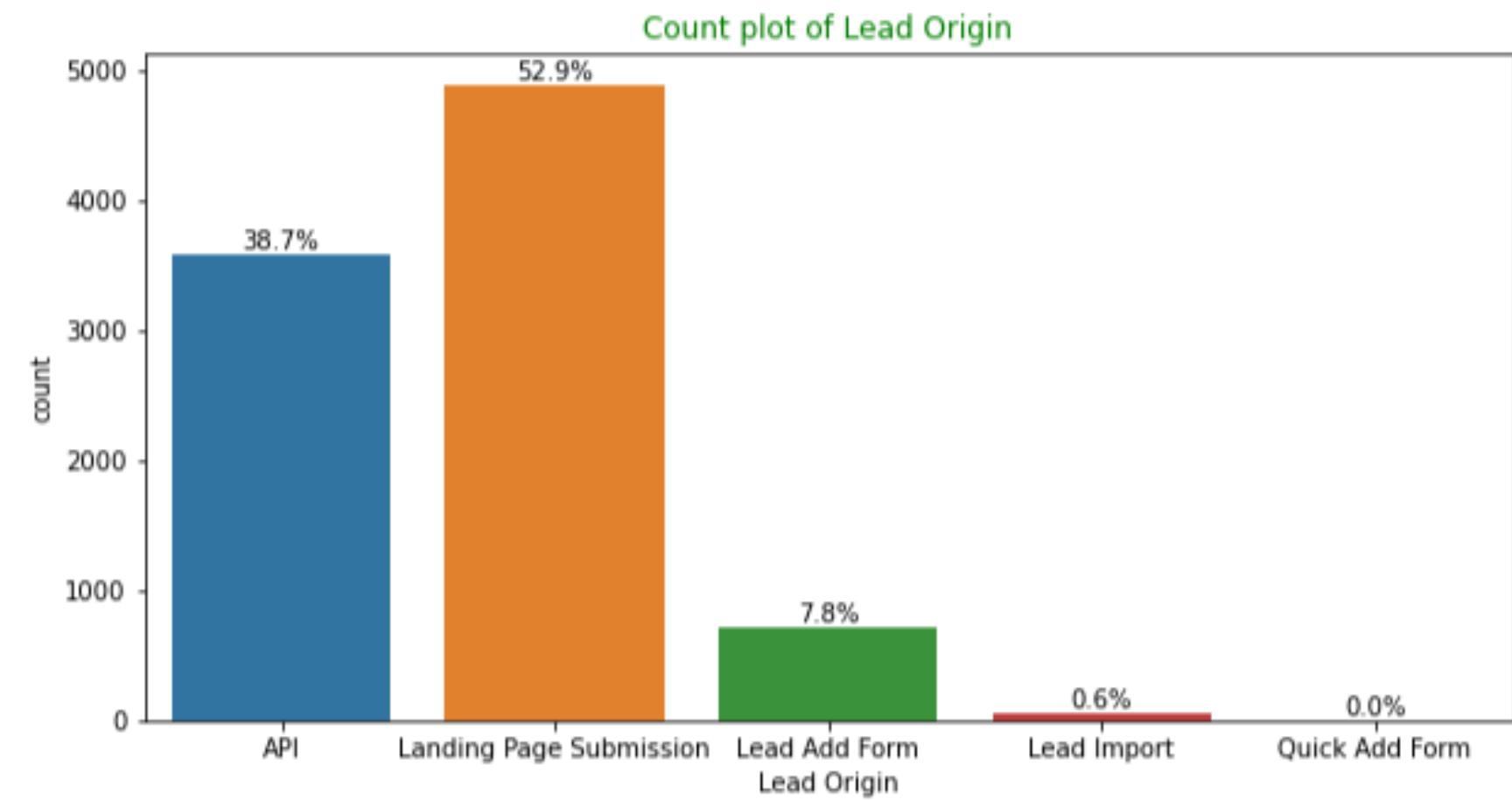
VISUALIZATION



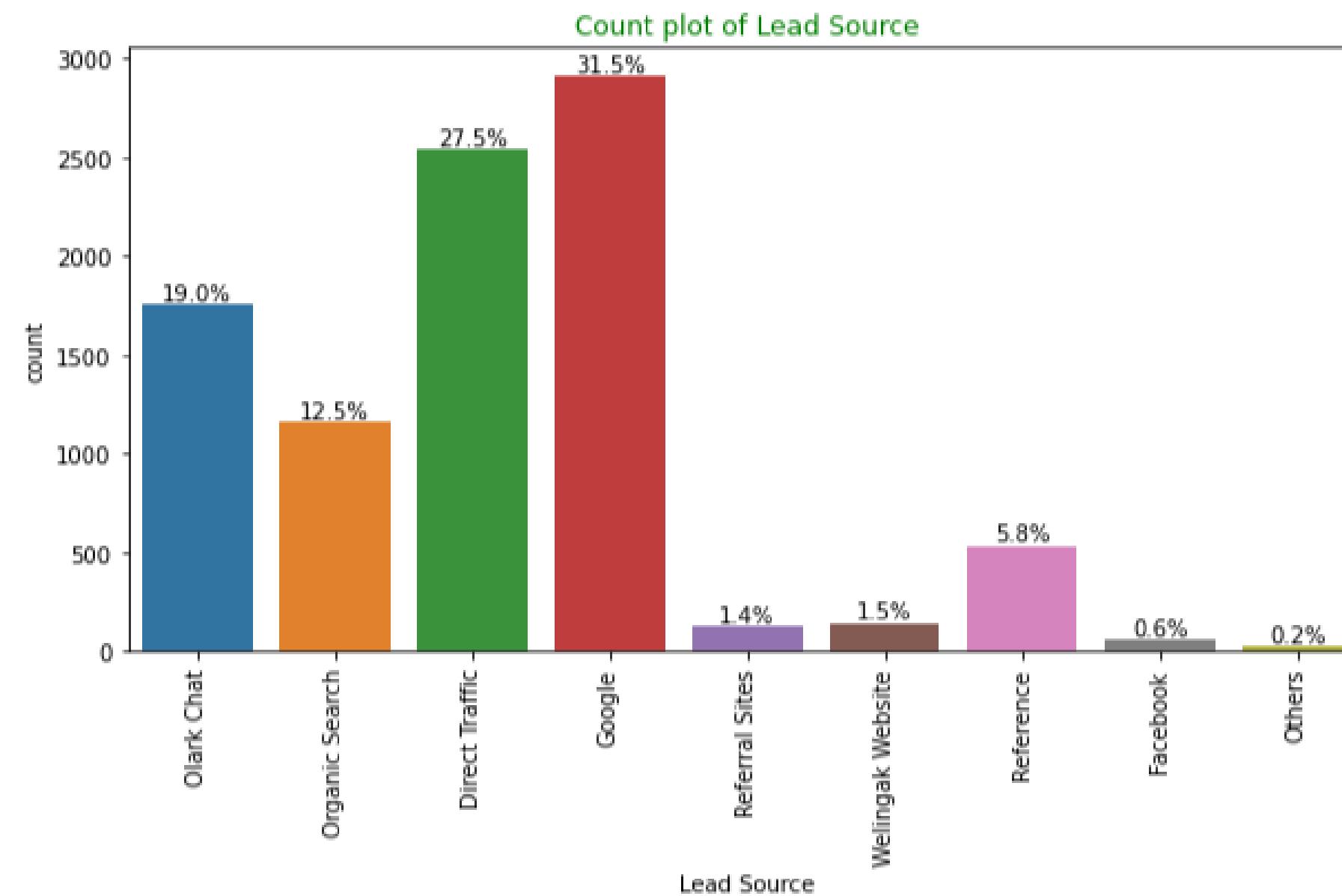


**Leads that are converted have lower rates than non-converted leads.
Almost 23% difference is seen in converted and non converted leads.**

The highest count percent of Leads Origin are from Landing page submission i.e. the website where candidates fill the forms followed by API.

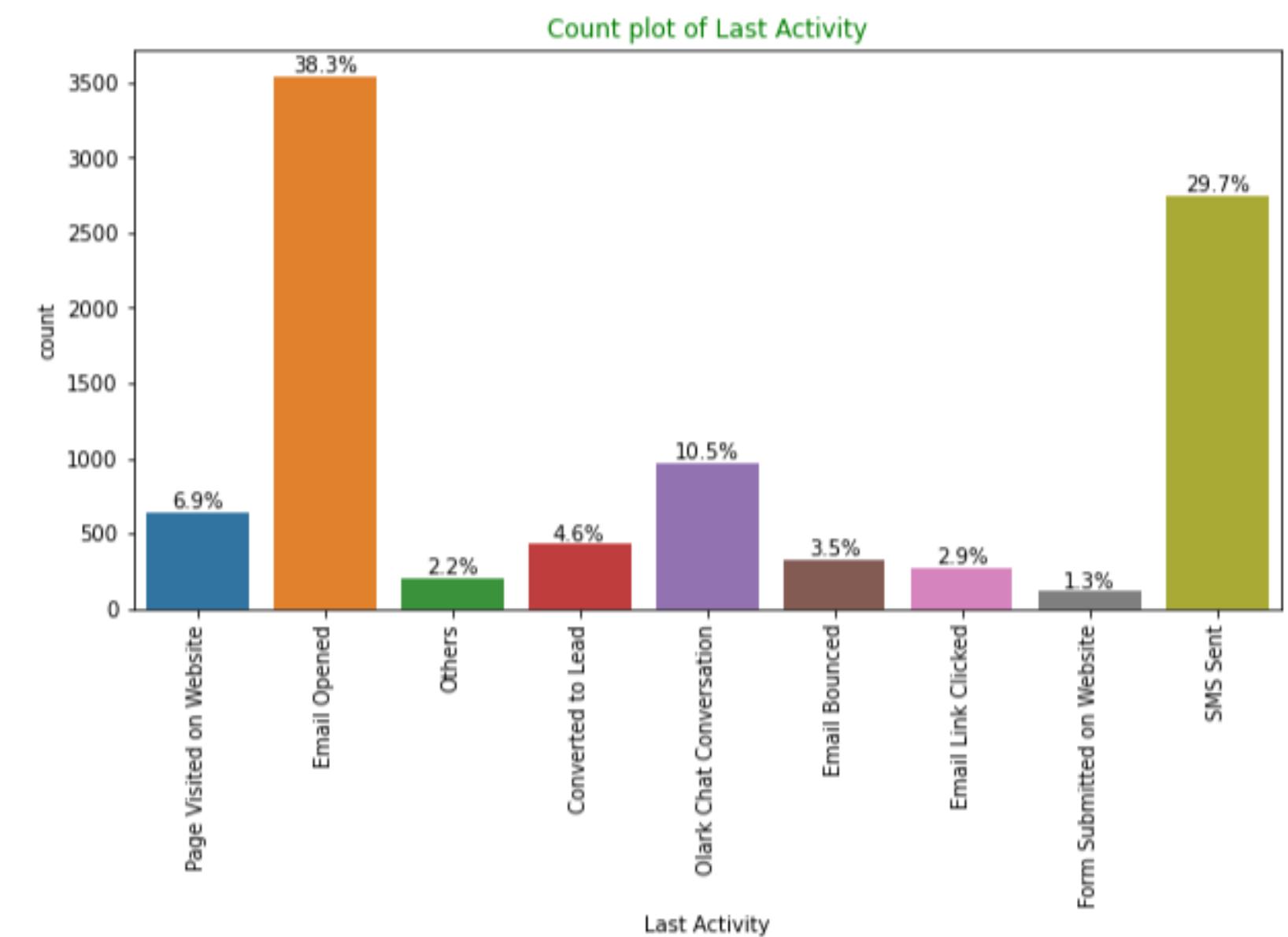


Maximum leads as per the data are unemployed as compared to all other occupations.

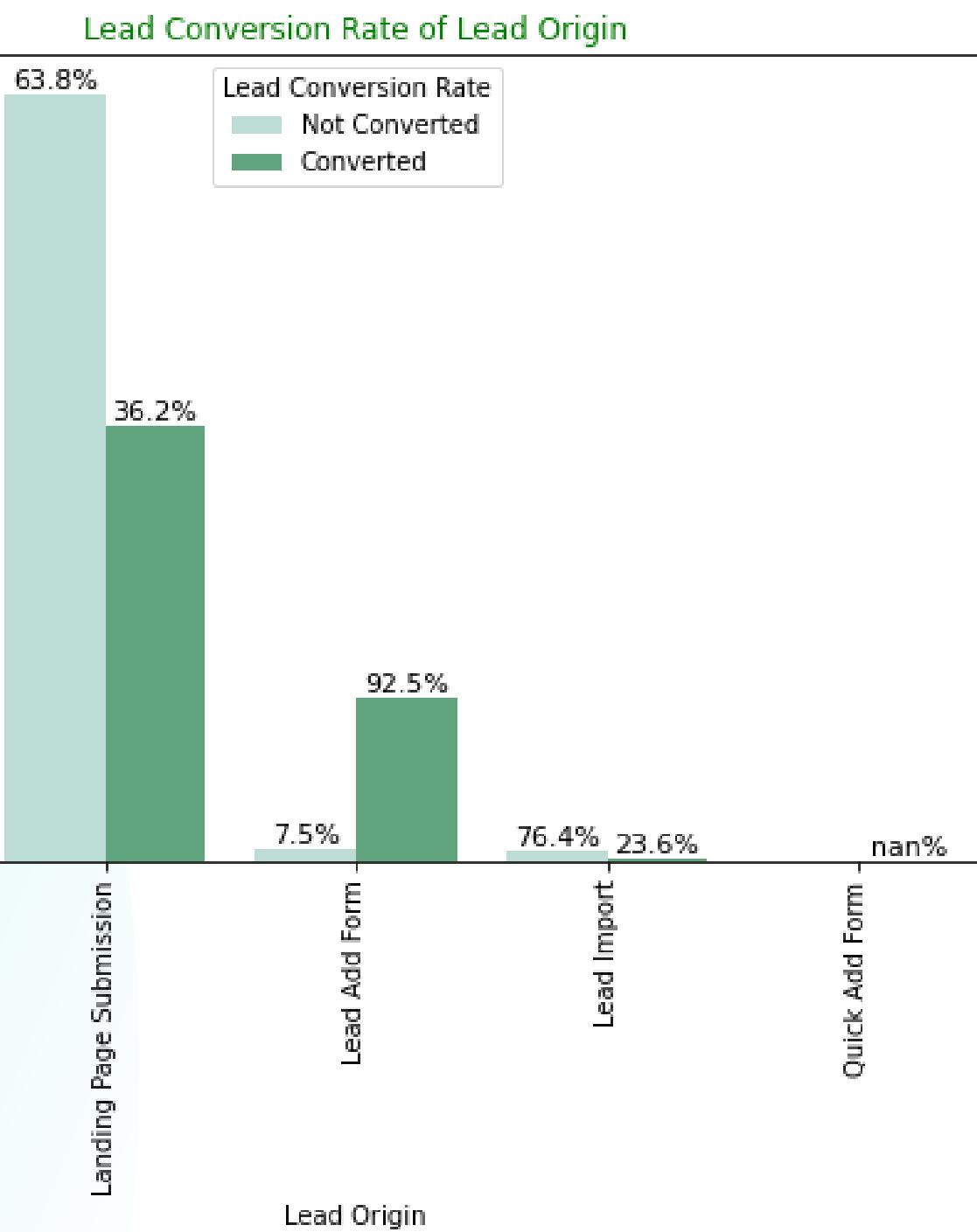
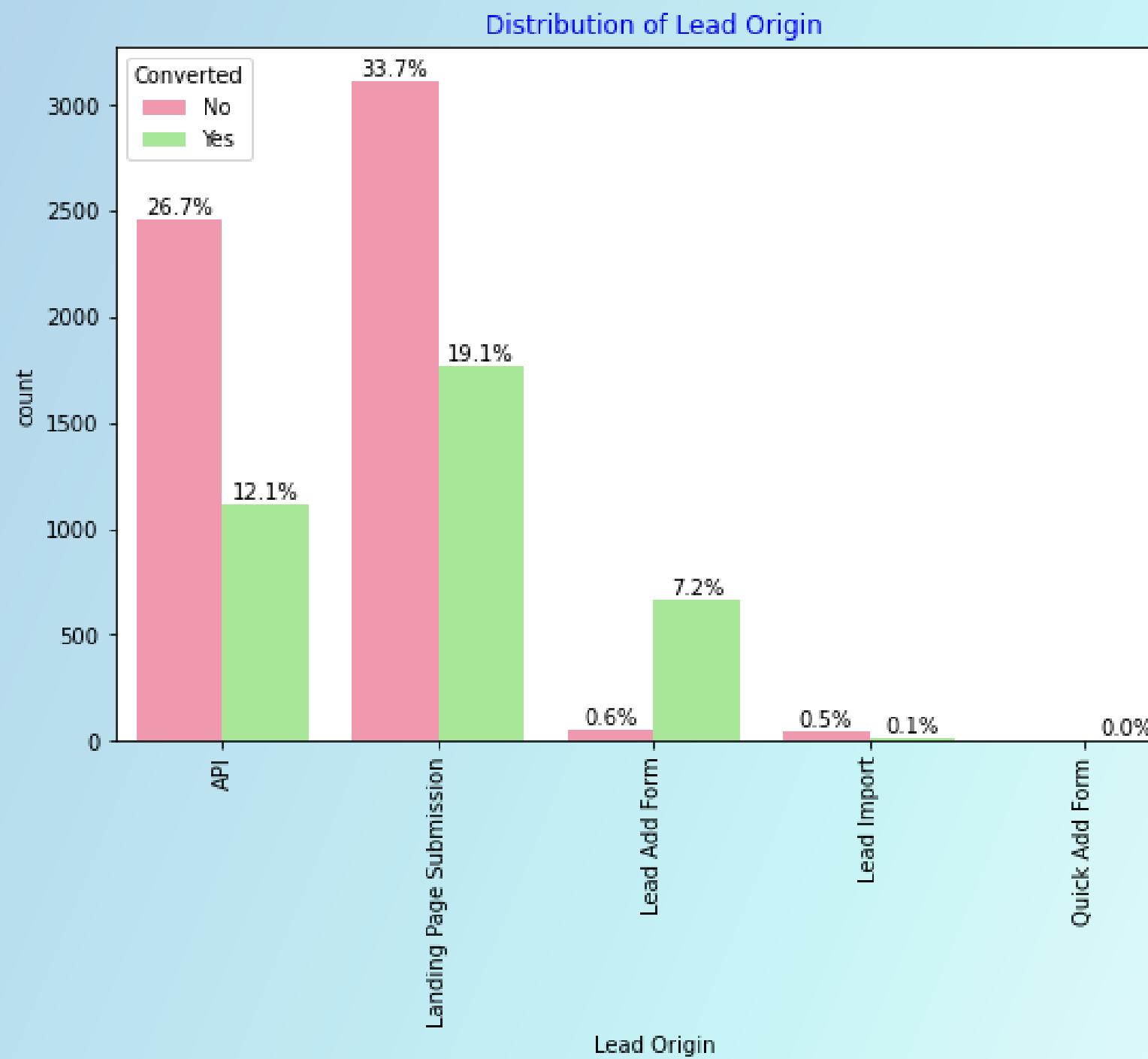


Customer's the last Activities were either Emails opened or SMS sent to the company.

Sources from where the leads are visiting or knowing about the company are from Google and Direct traffic.



Lead Origin Countplot vs Lead Conversion Rates

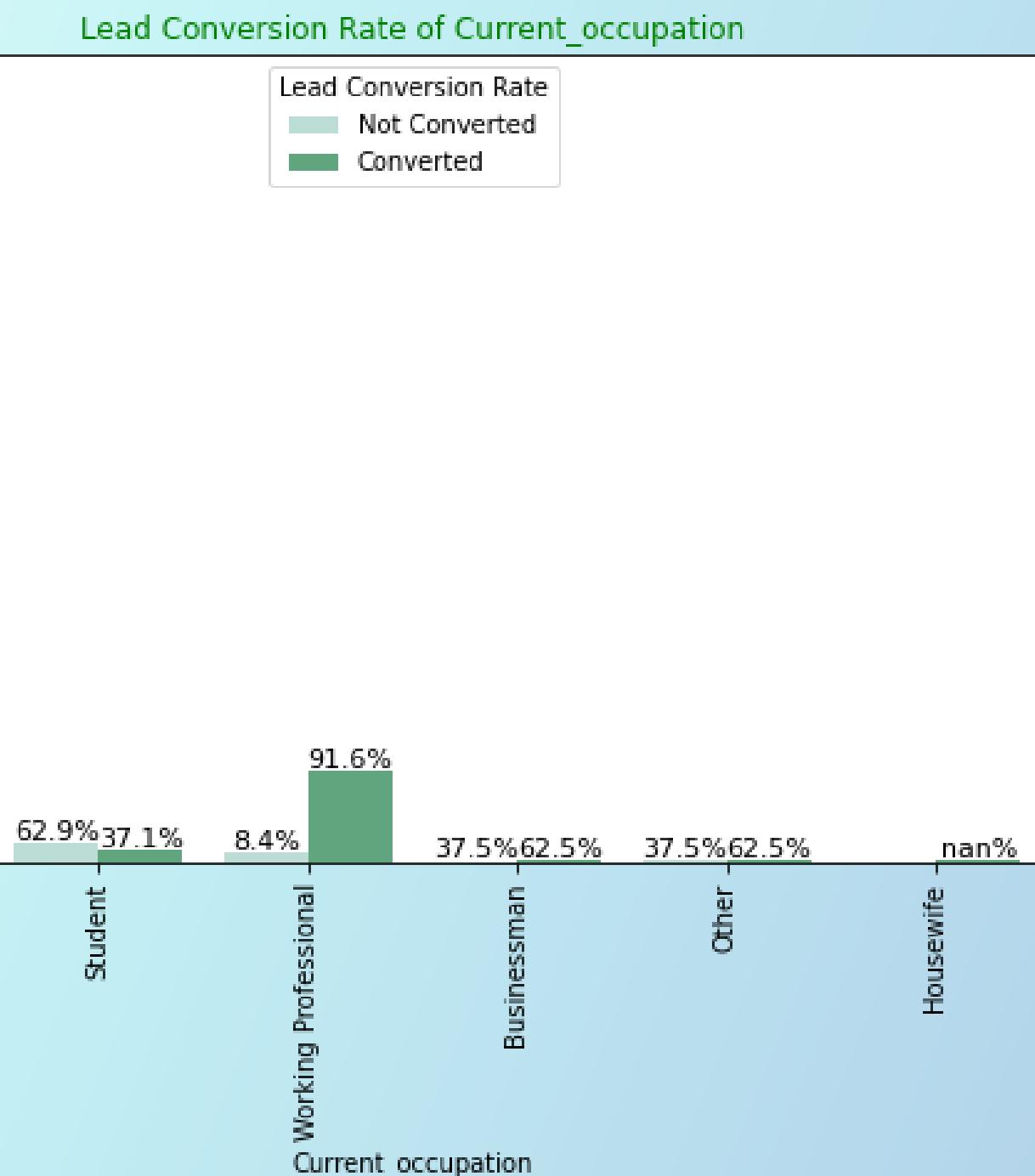
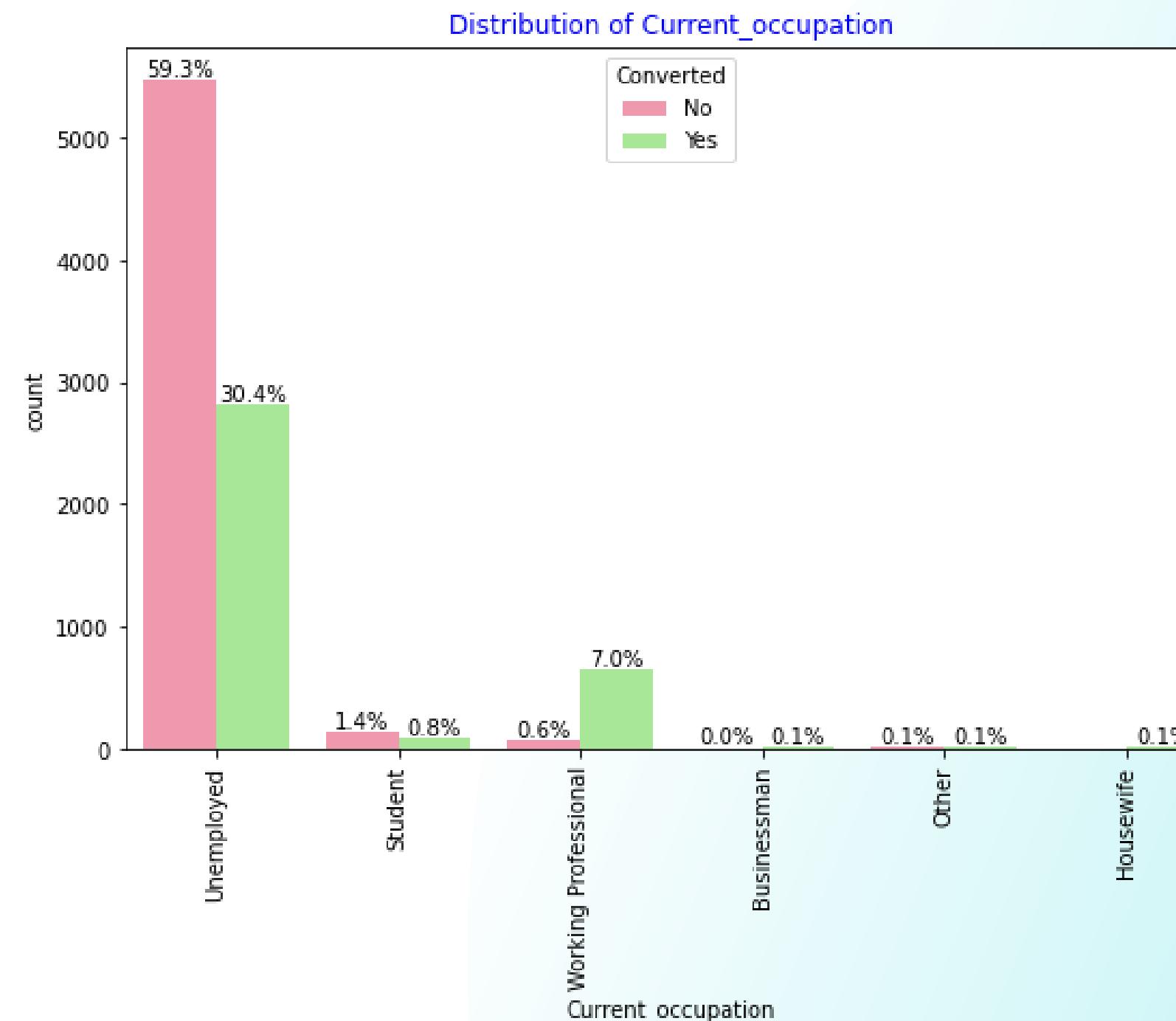


As compared to other lead origins landing page submission has the highest lead conversion rate followed by API.

Even though Landing page submission has the highest conversion rate but non converted leads are even higher for it as well as for API.

Lead Add Form is the only Lead origin where the converted rate is high than nonconverted rate.

Current_occupation Countplot vs Lead Conversion Rates



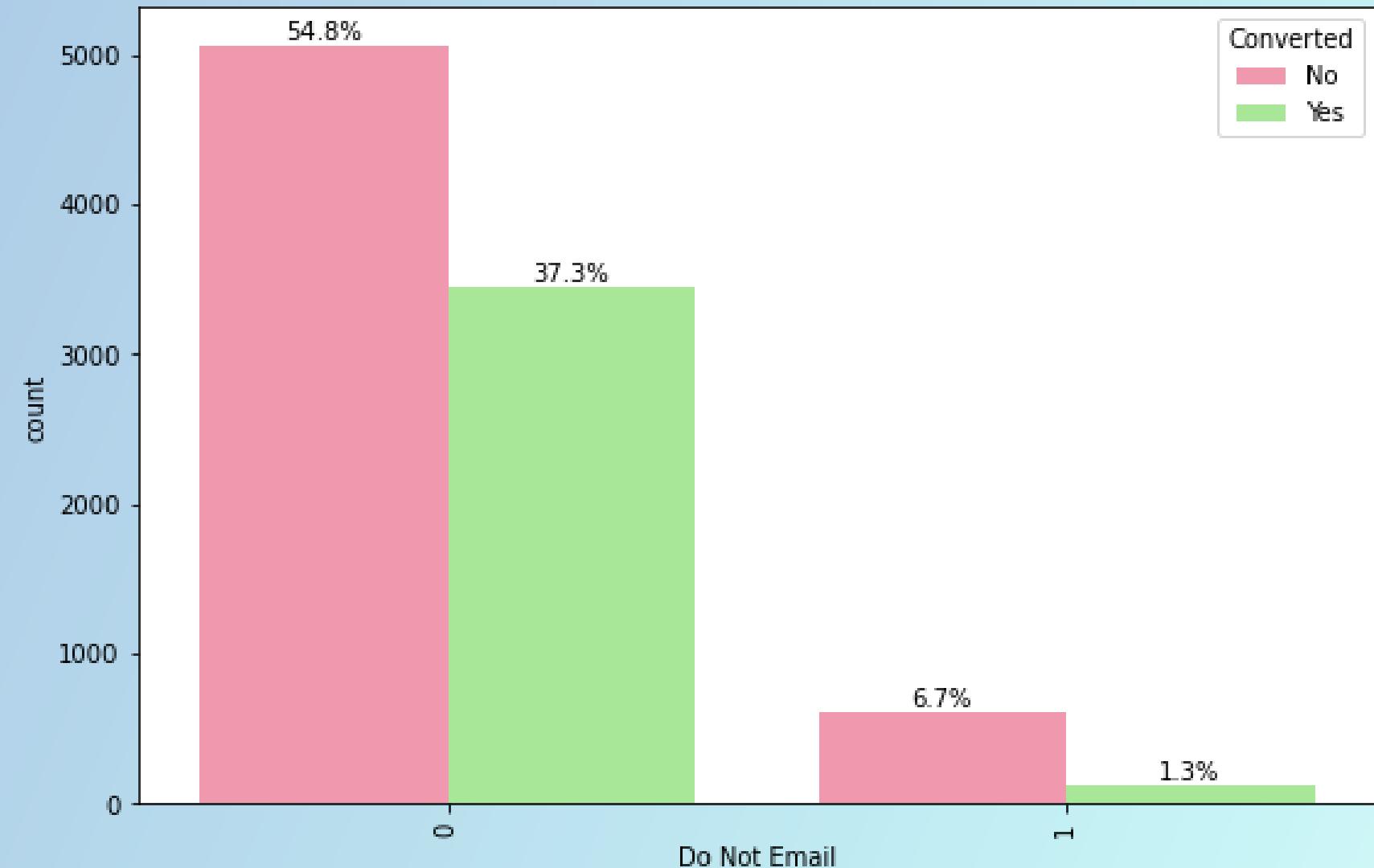
Customers having Current Occupation as Unemployed has highest Conversion rate as well as non-Conversion rate than other Occupations.

But the difference between Non Converted customers is more than Converted customers, except for Working Professional.

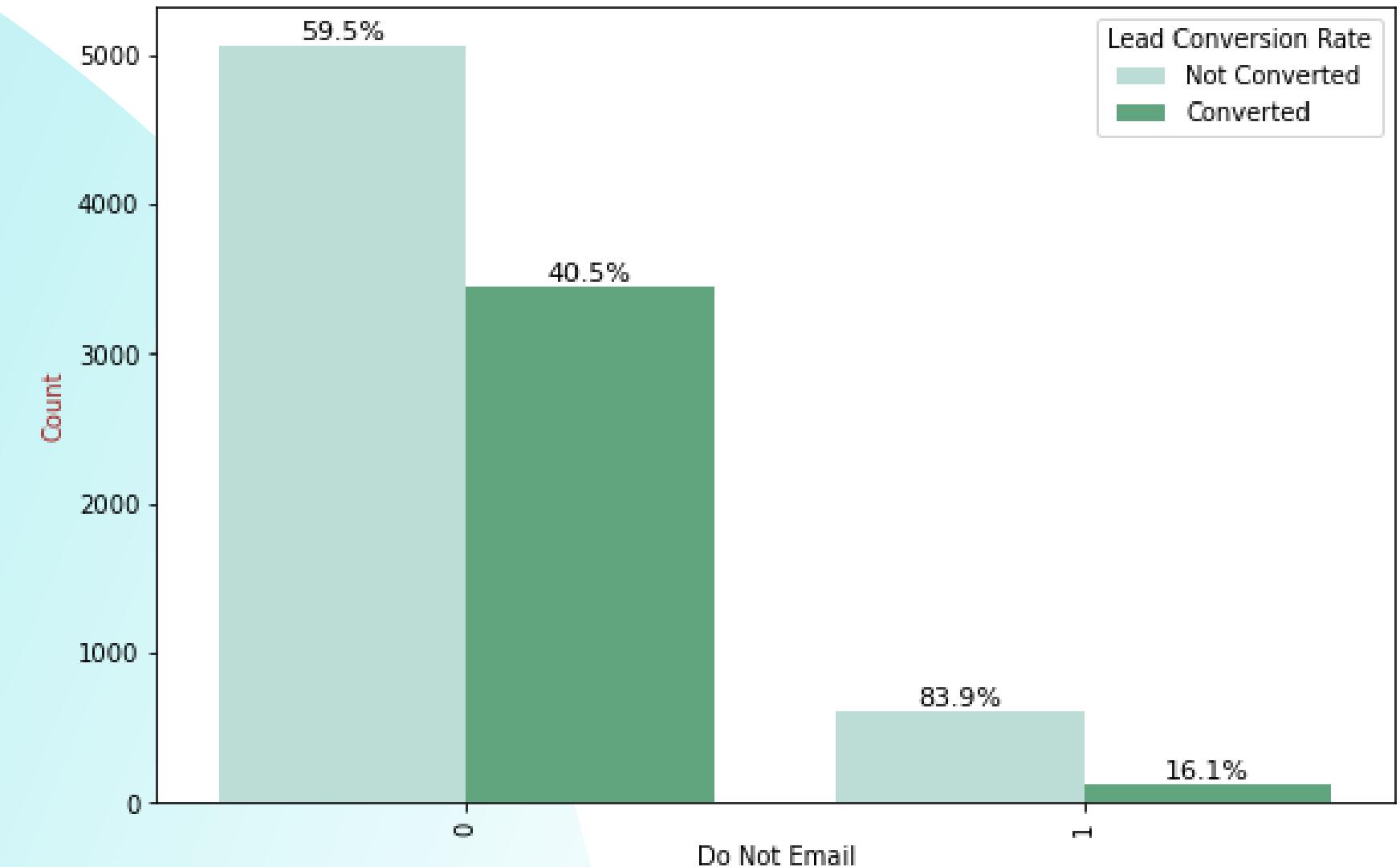
Though Working Professional total is not more than 1000 but converted leads percent is higher .

Do Not Email Countplot vs Lead Conversion Rates

Distribution of Do Not Email

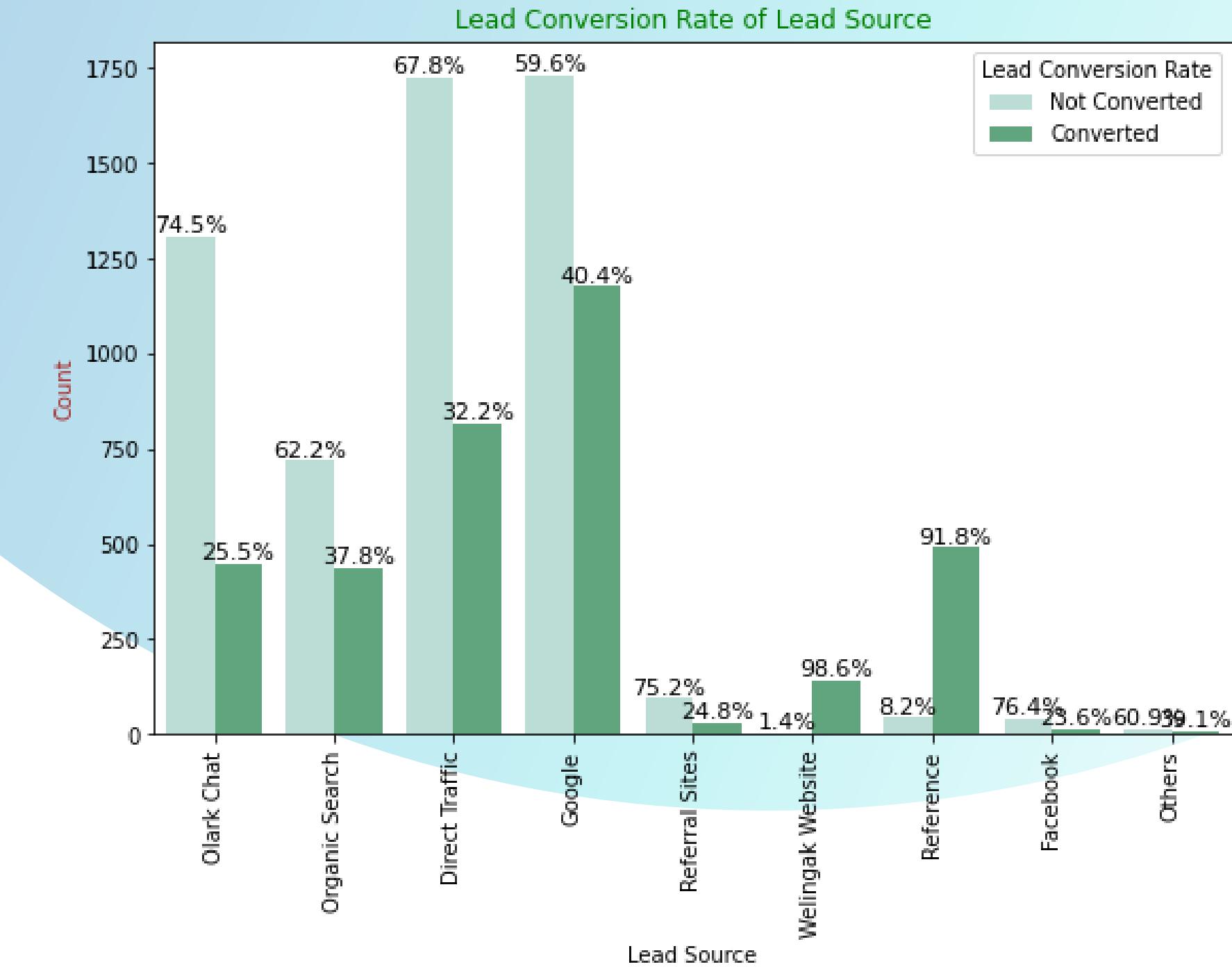
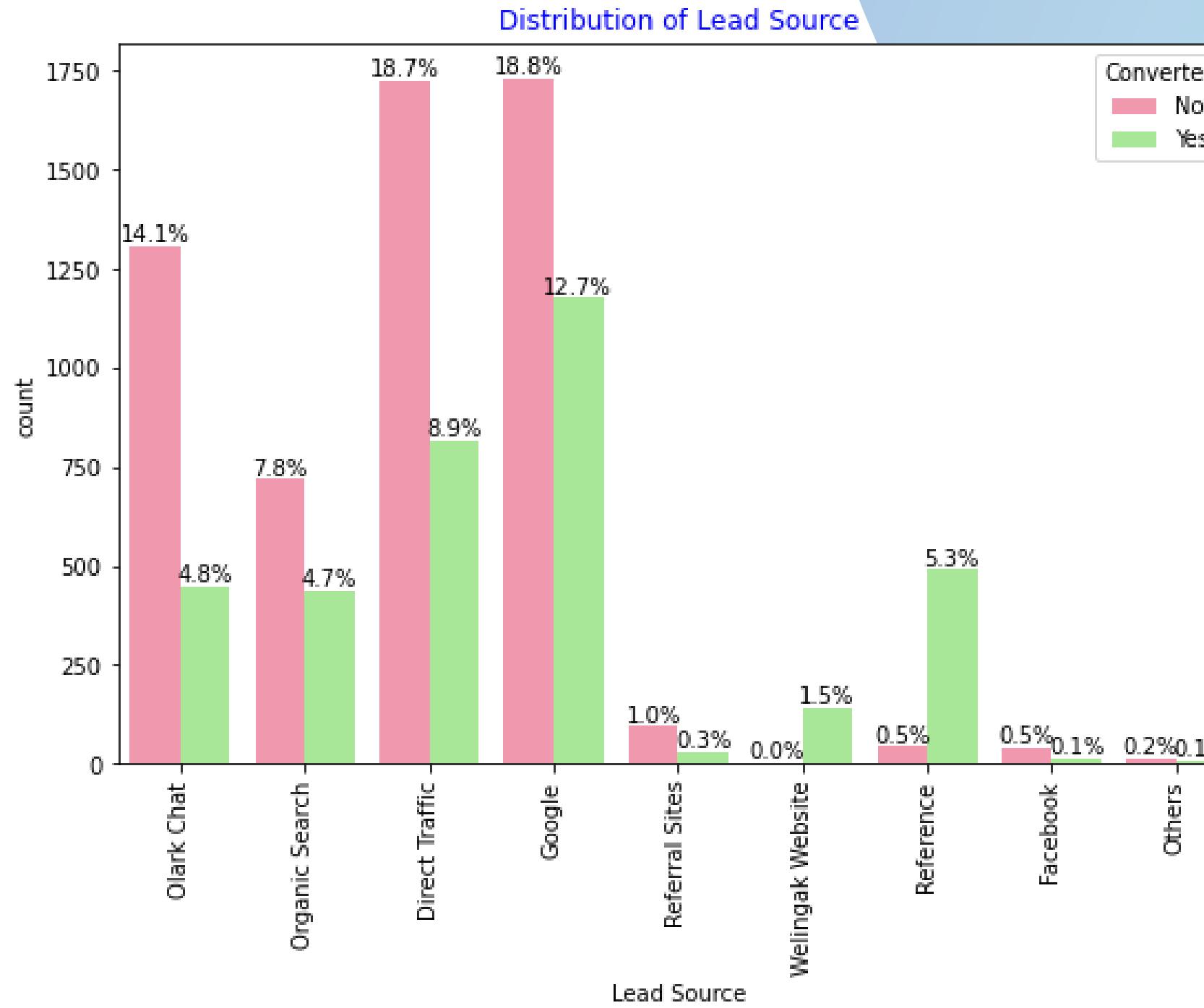


Lead Conversion Rate of Do Not Email



Customers who do not want to mail them are much less in number as compared to the one who do not want company to email them.

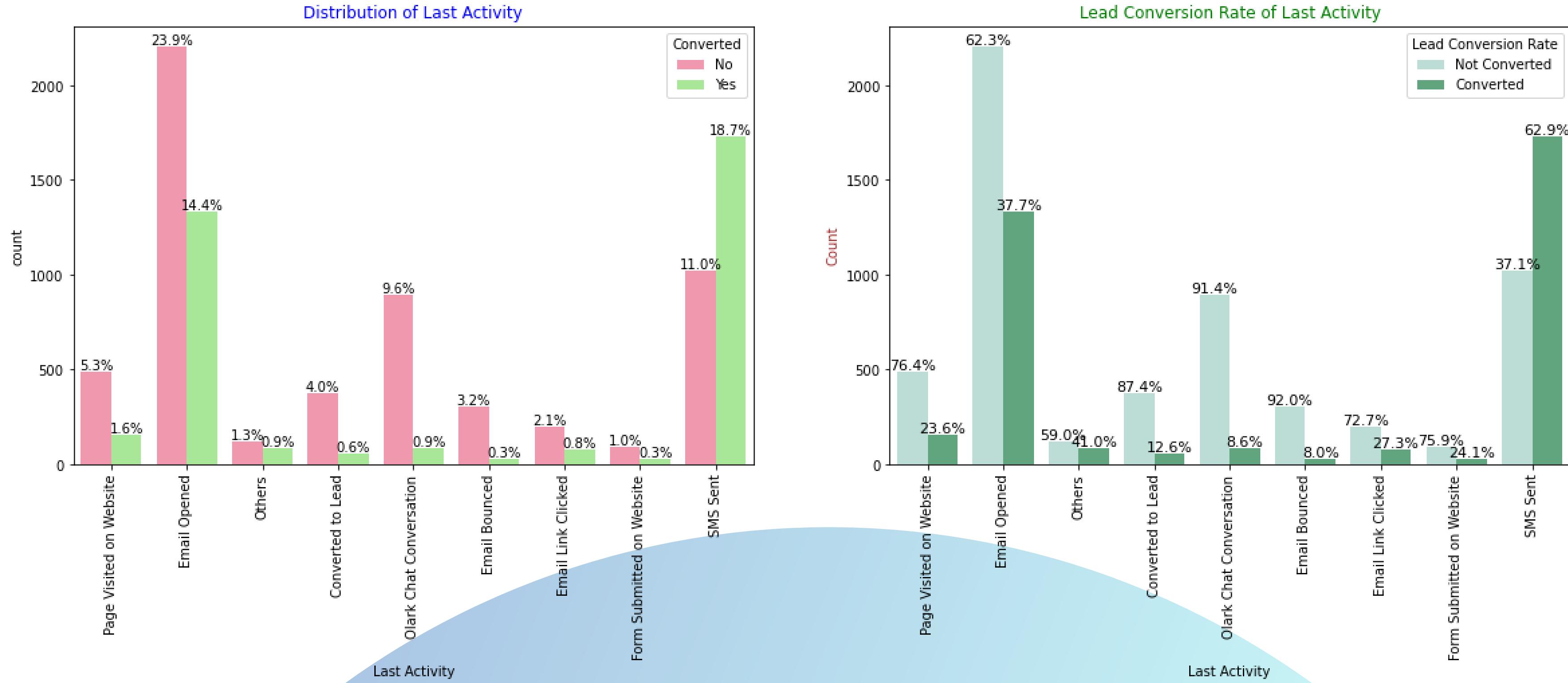
Lead Source Countplot vs Lead Conversion Rates



Lead Source which are Referred have highest lead conversion rates as compared to others which is then followed by Welingak website.

Though Google and Other sources have high number of leads generated, their conversion rate is very less.

Last Activity Countplot vs Lead Conversion Rates

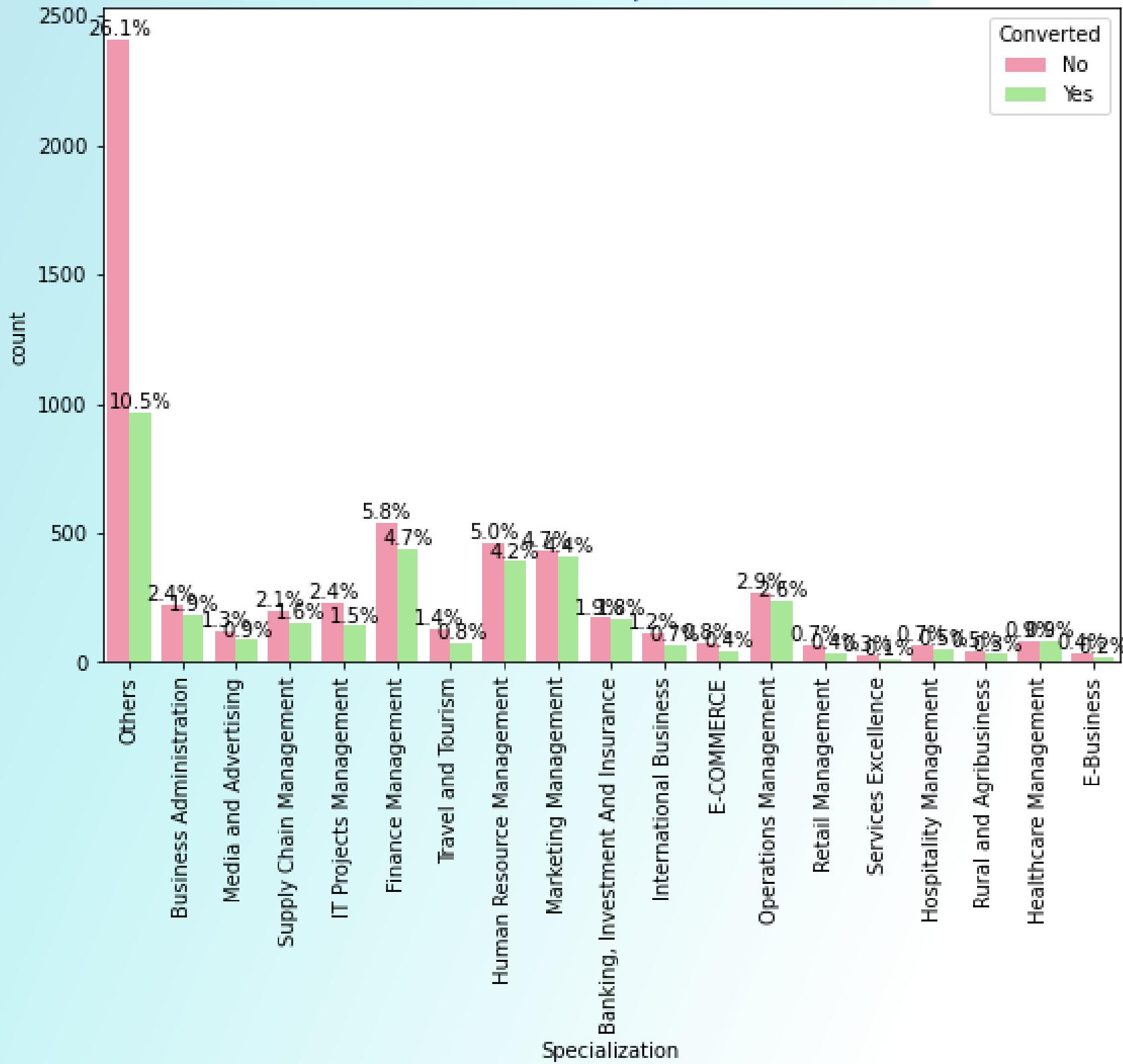


The last Activity which is converted to hot leads is only for customers whose SMSs are Sent by the company.

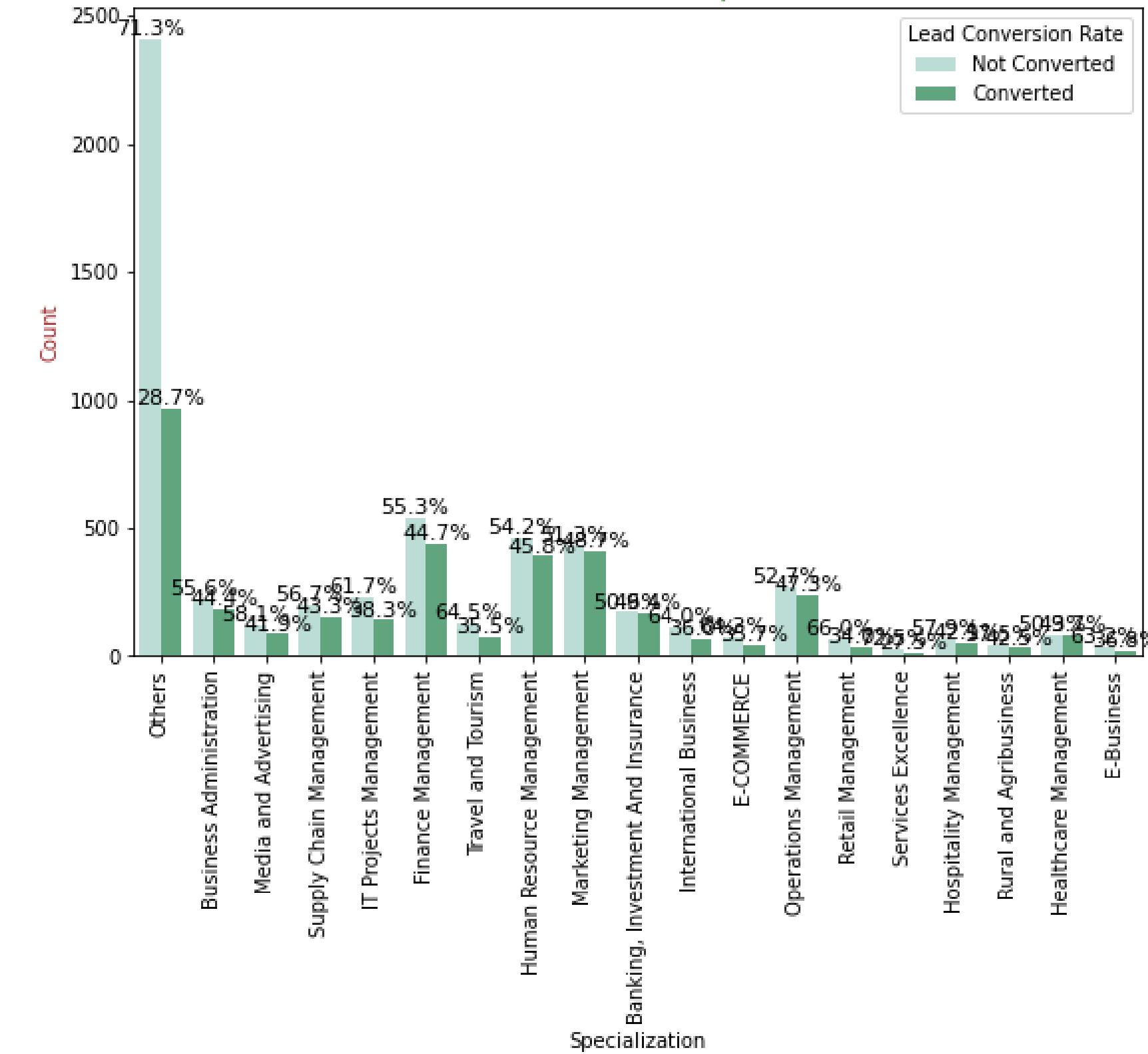
Email opened have high number of leads generated overall but conversion rate is quite low.

Specialization Countplot vs Lead Conversion Rates

Distribution of Specialization

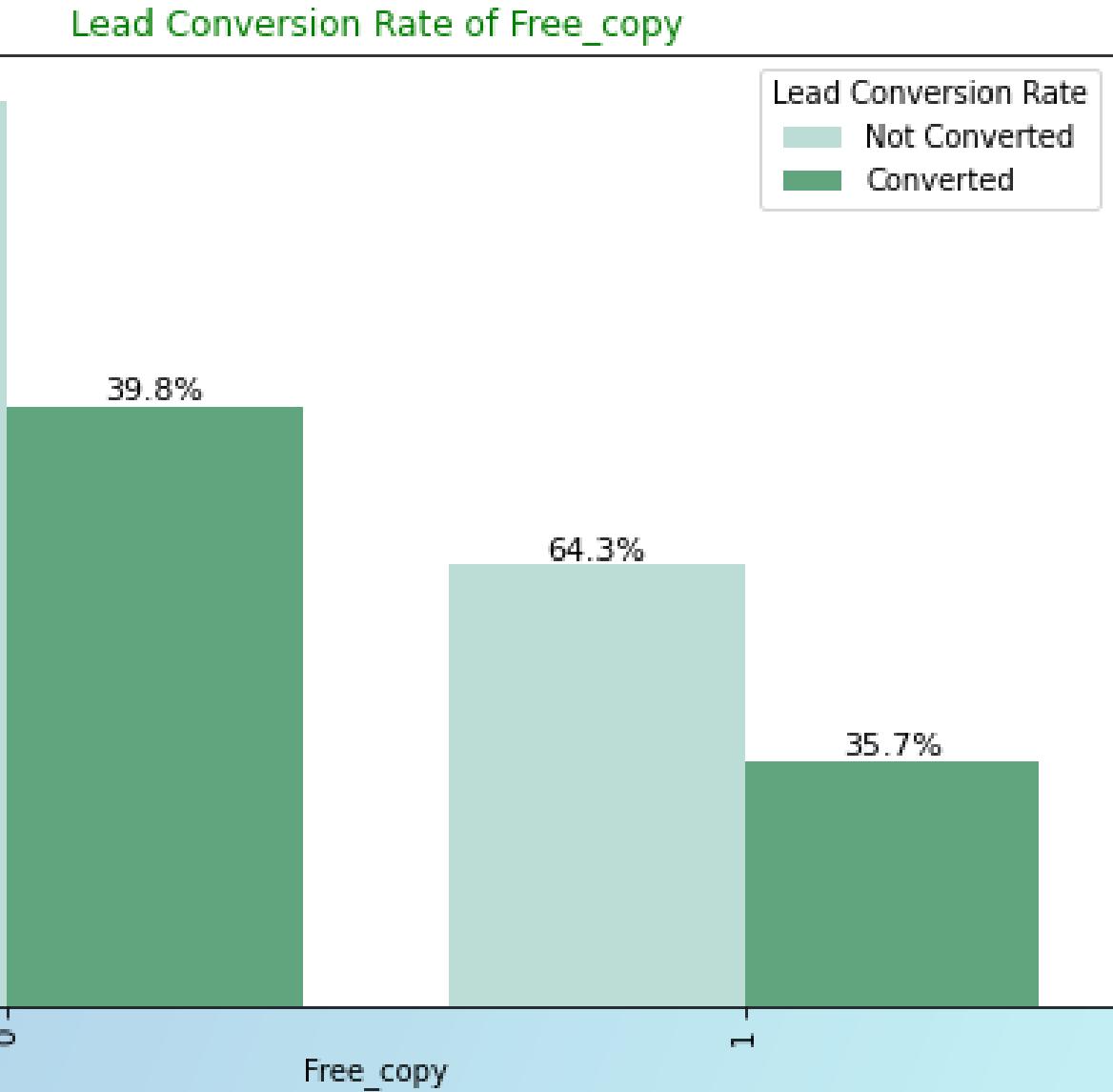
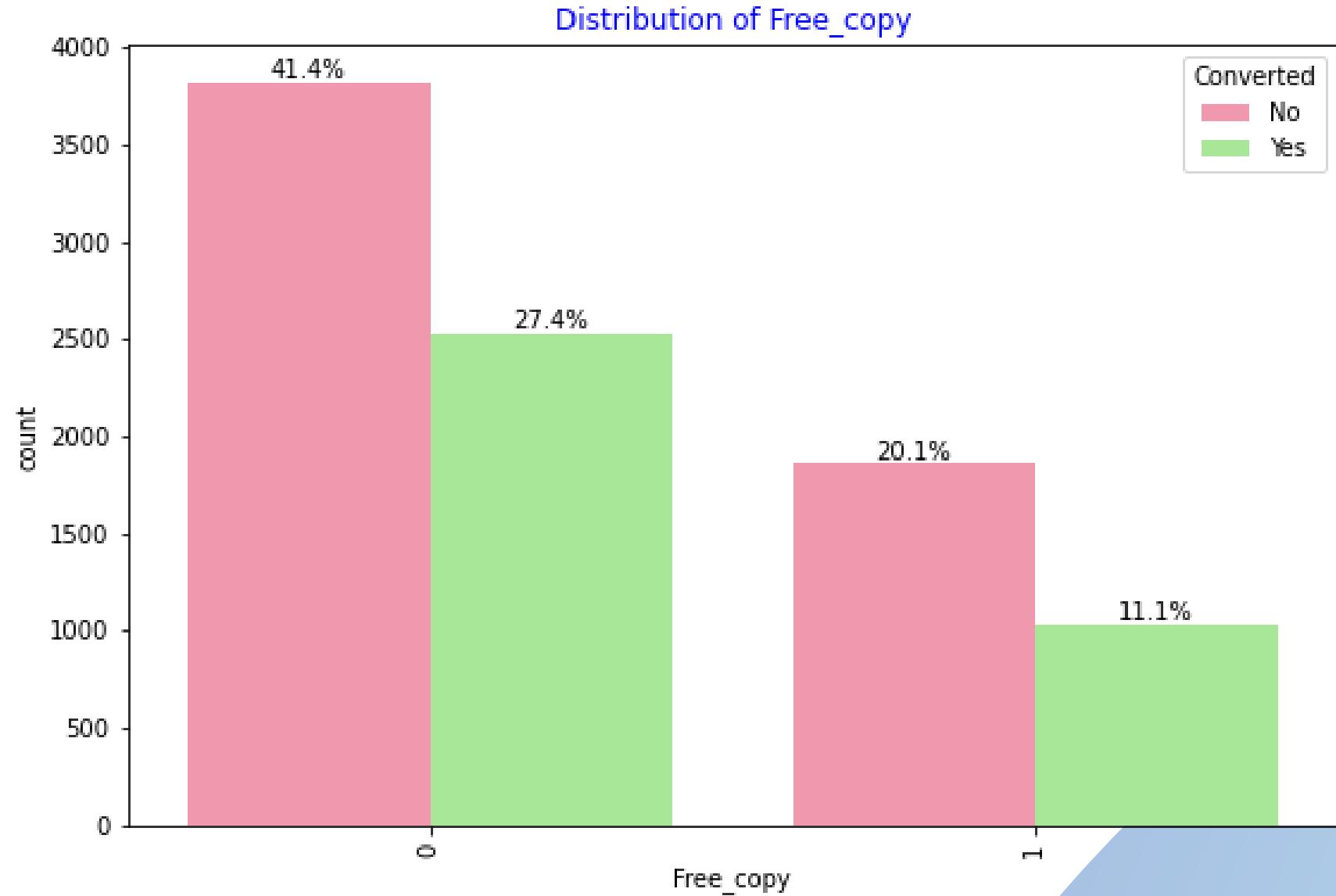


Lead Conversion Rate of Specialization

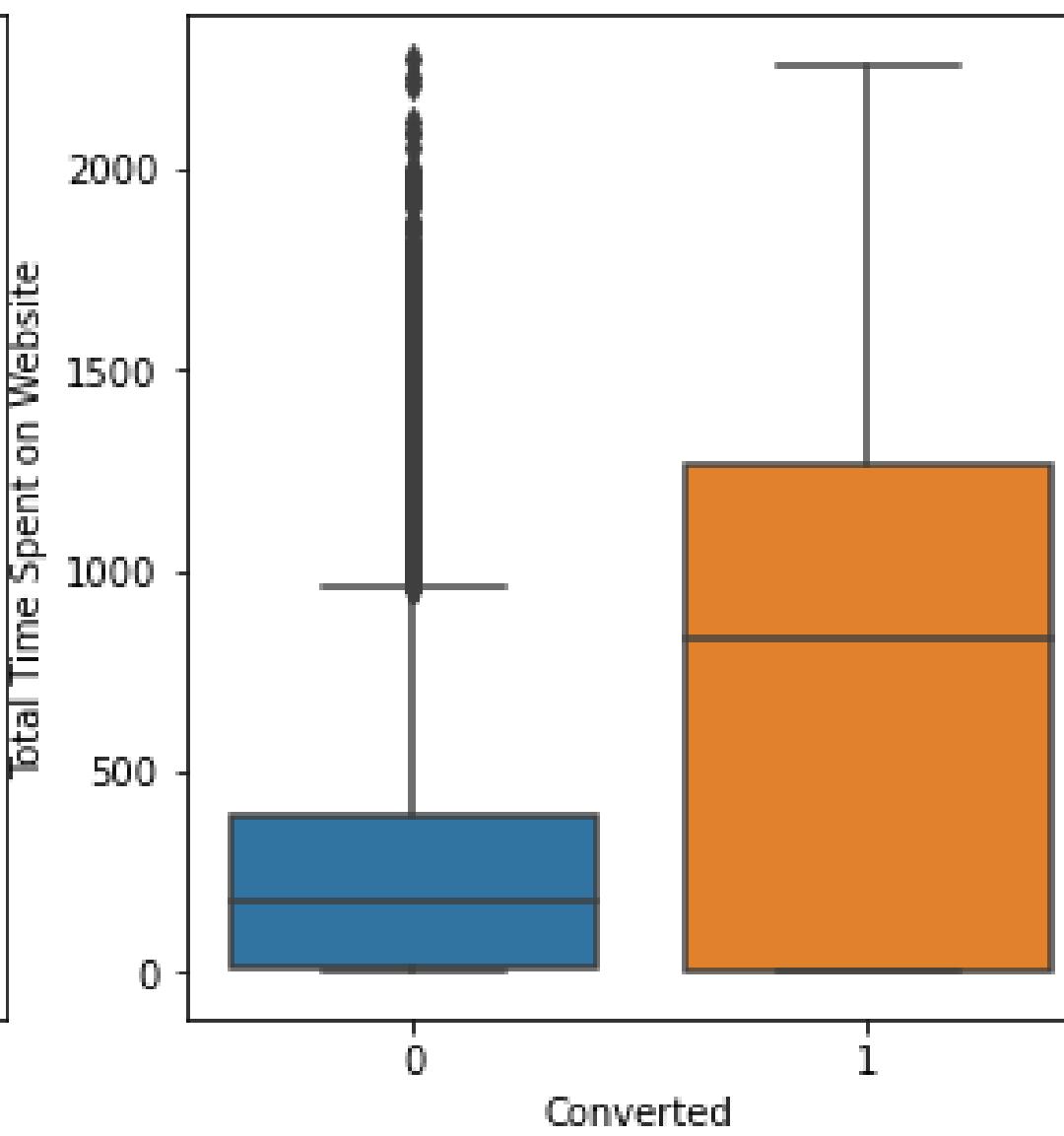
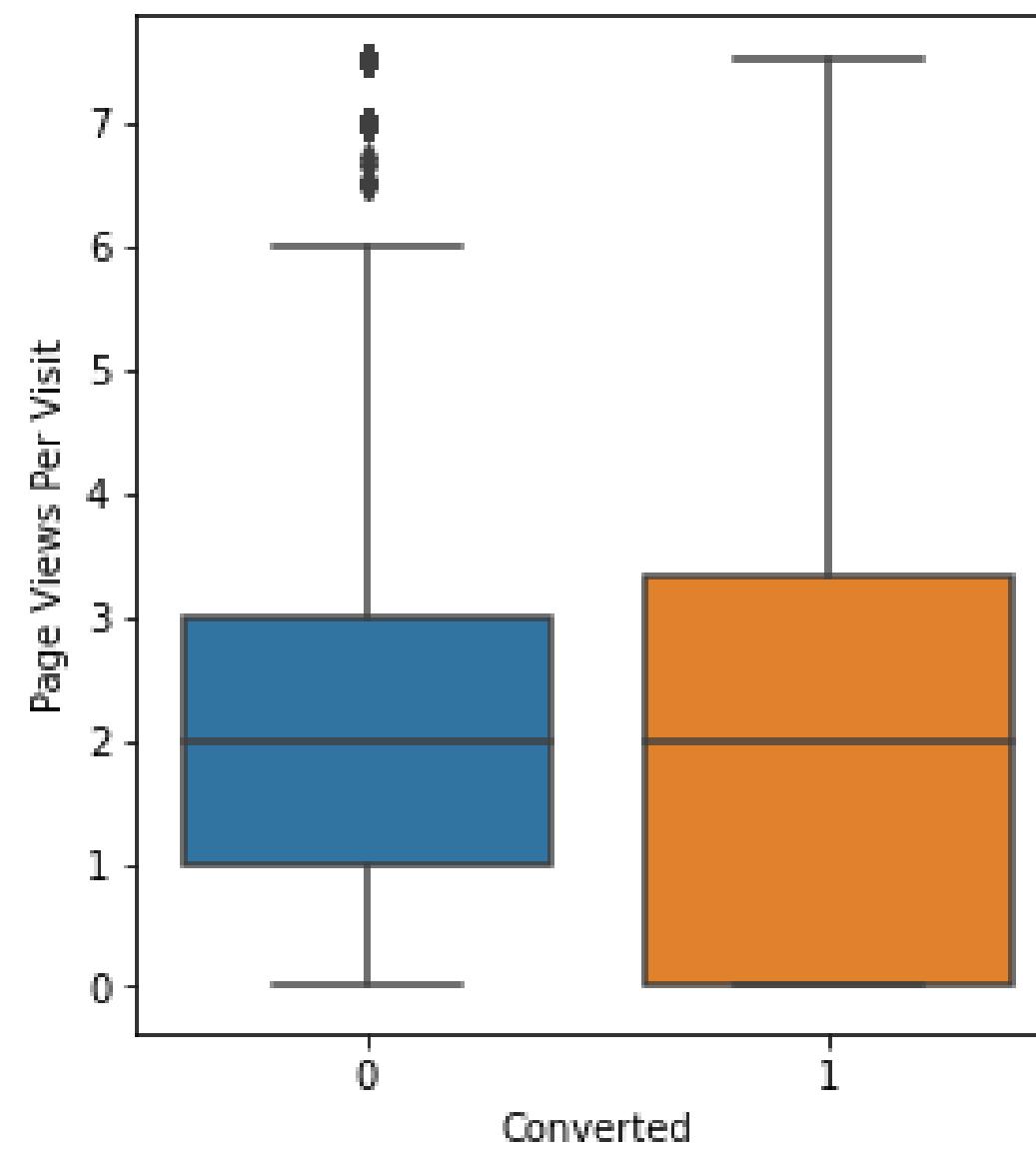
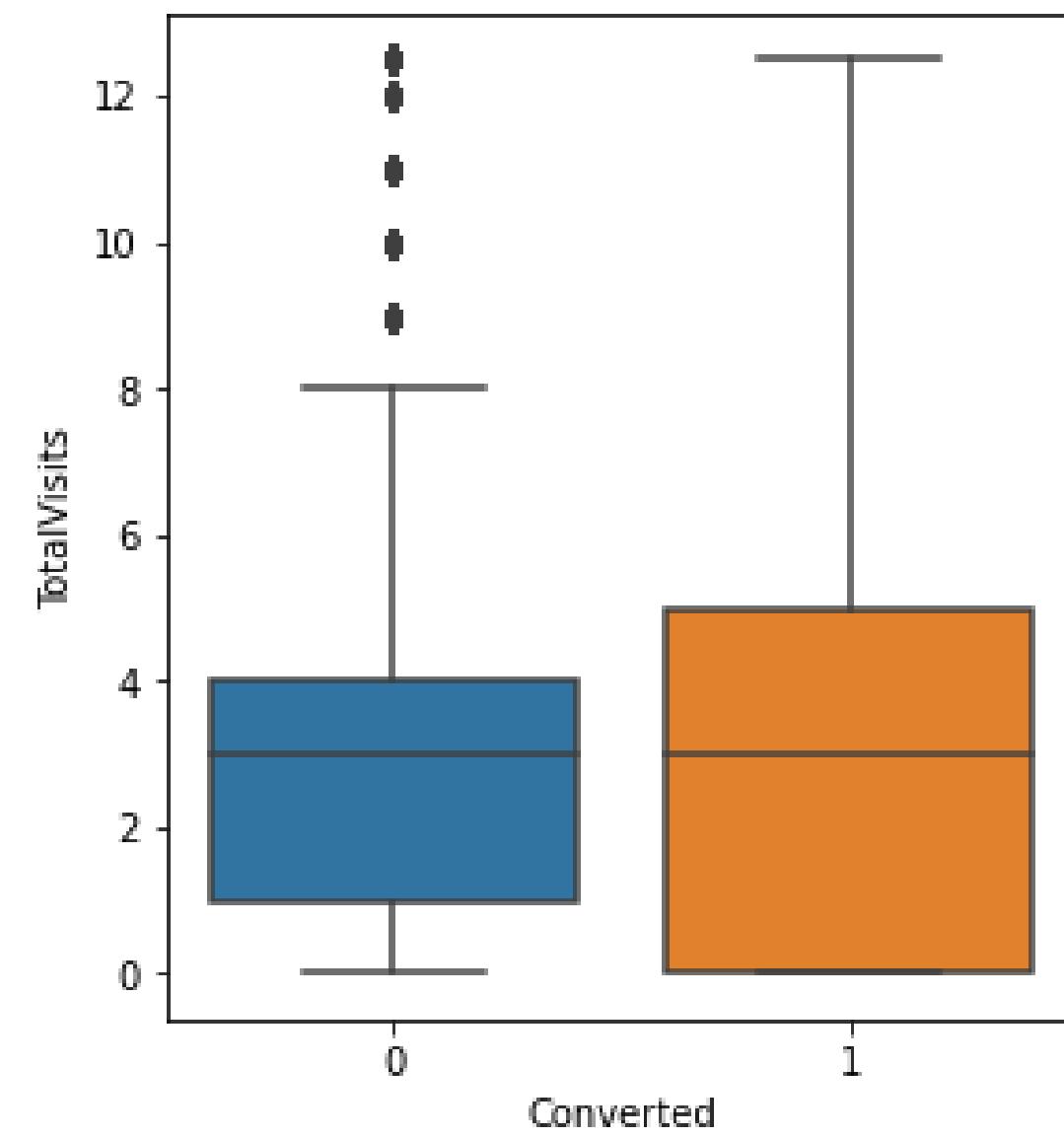


Among all the Specializations the leads mostly converted into hot leads is Healthcare Management, but the overall count is very less.

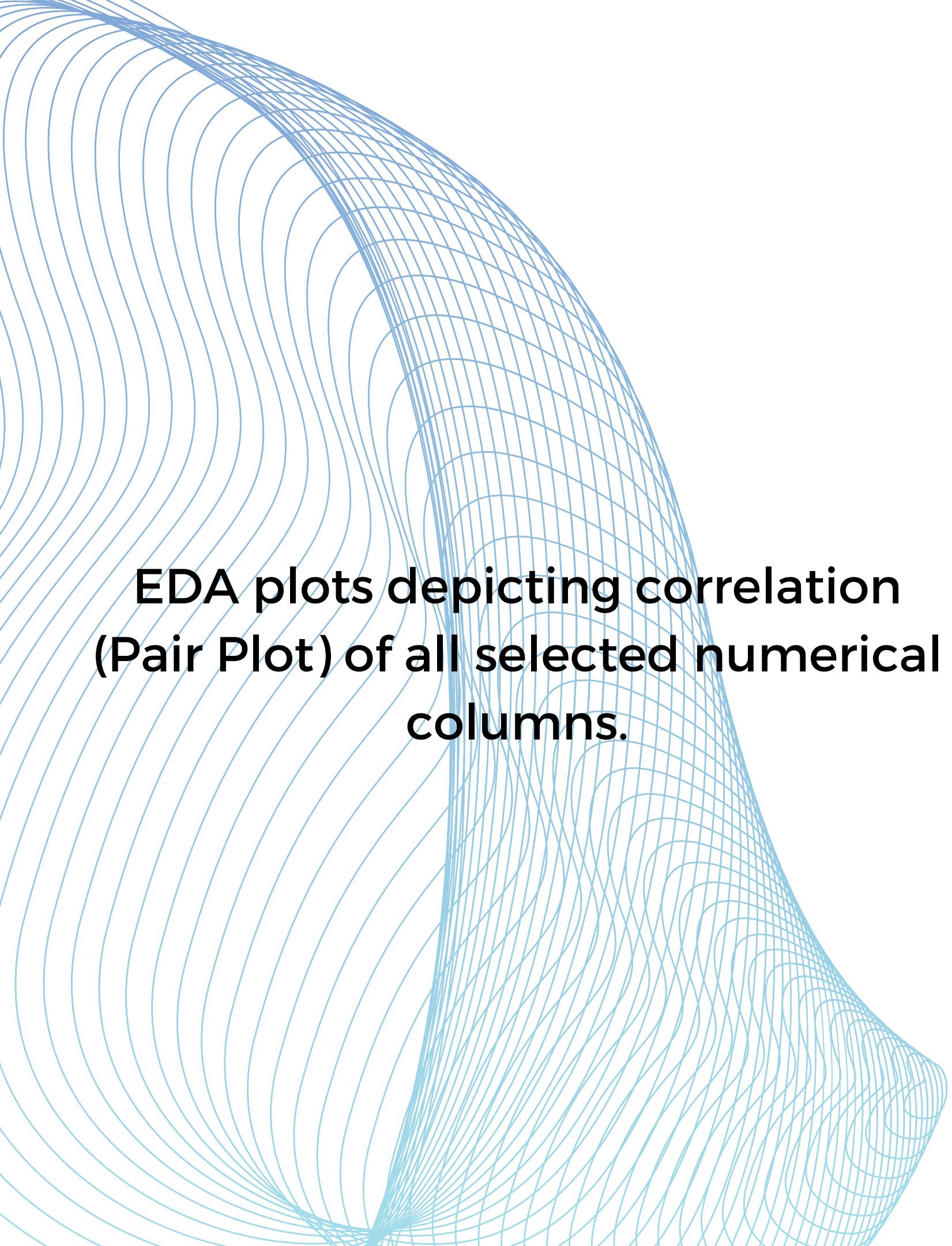
Free_copy Countplot vs Lead Conversion Rates



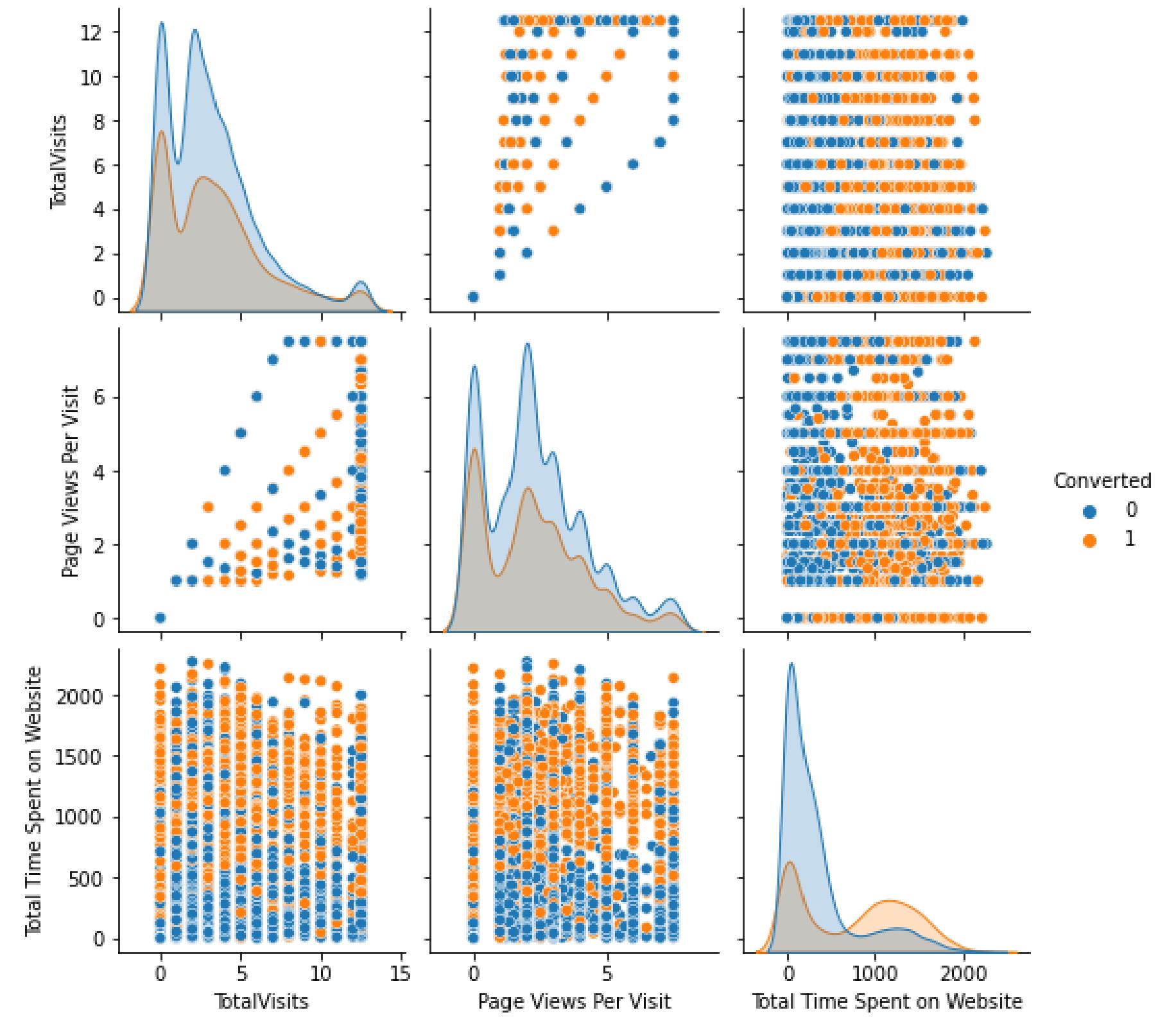
The Customers who want free copies of 'Mastering the Interview' are less than who don't.
But still the lead conversion rate is not much in either of the cases.
As compared to Distribution rate and Lead Conversion Rate the one who don not want Free Copy is higher.



EDA plots depicting variation in numerical columns for those who Converted and those who didn't.



EDA plots depicting correlation (Pair Plot) of all selected numerical columns.

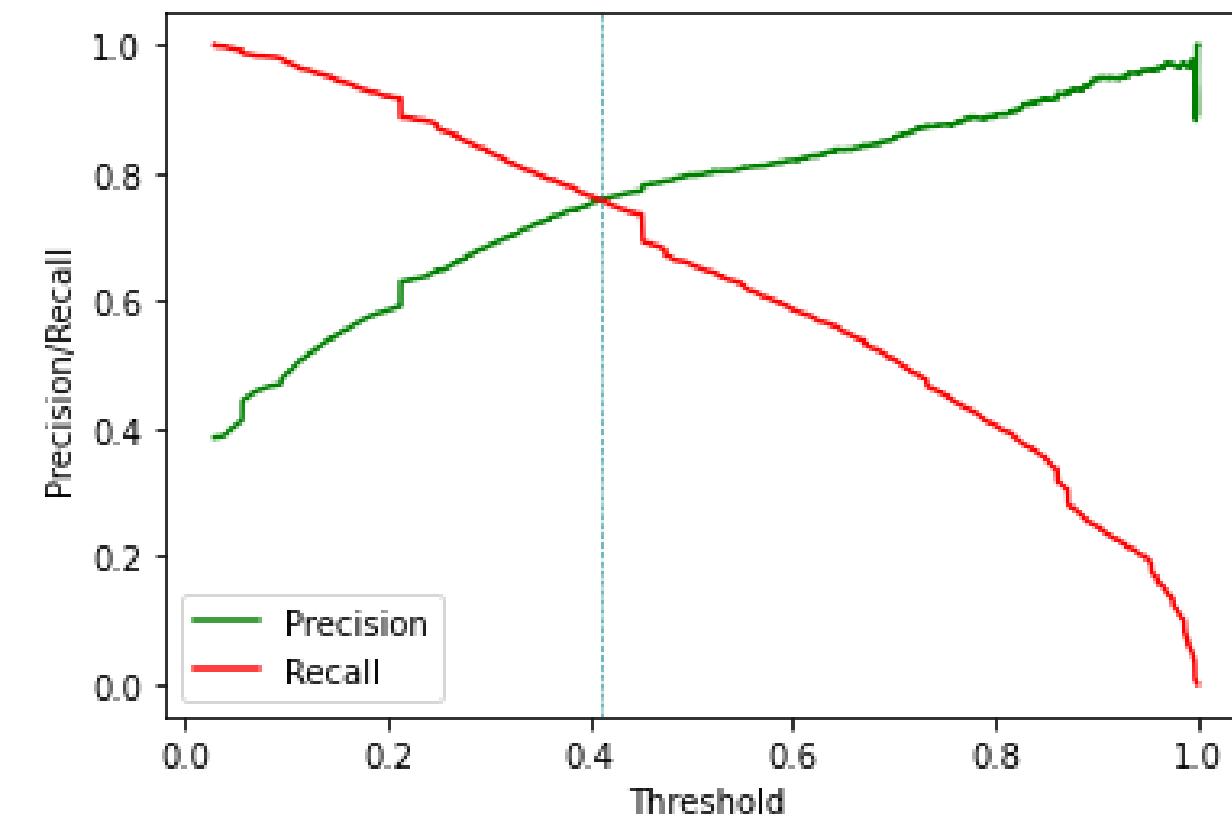
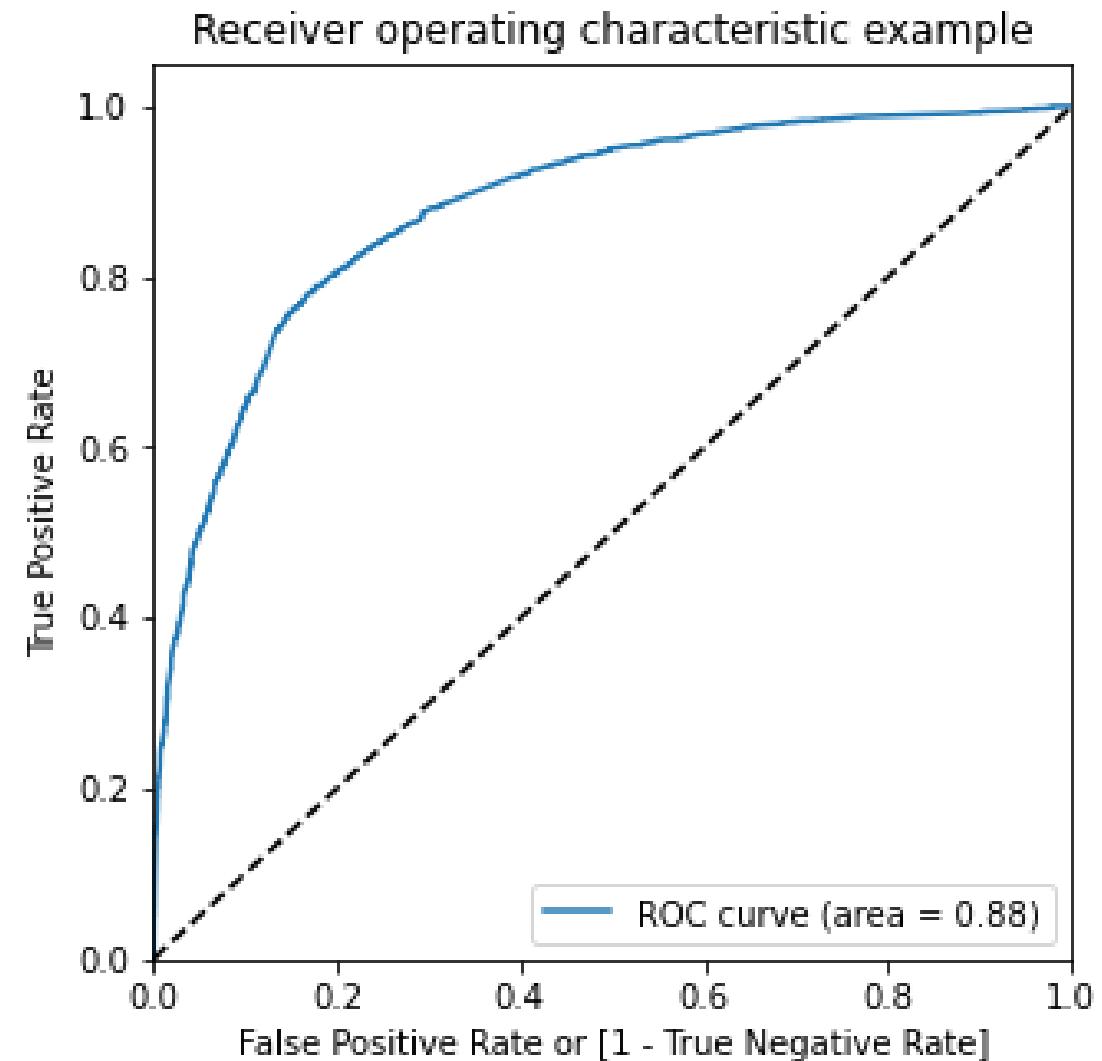
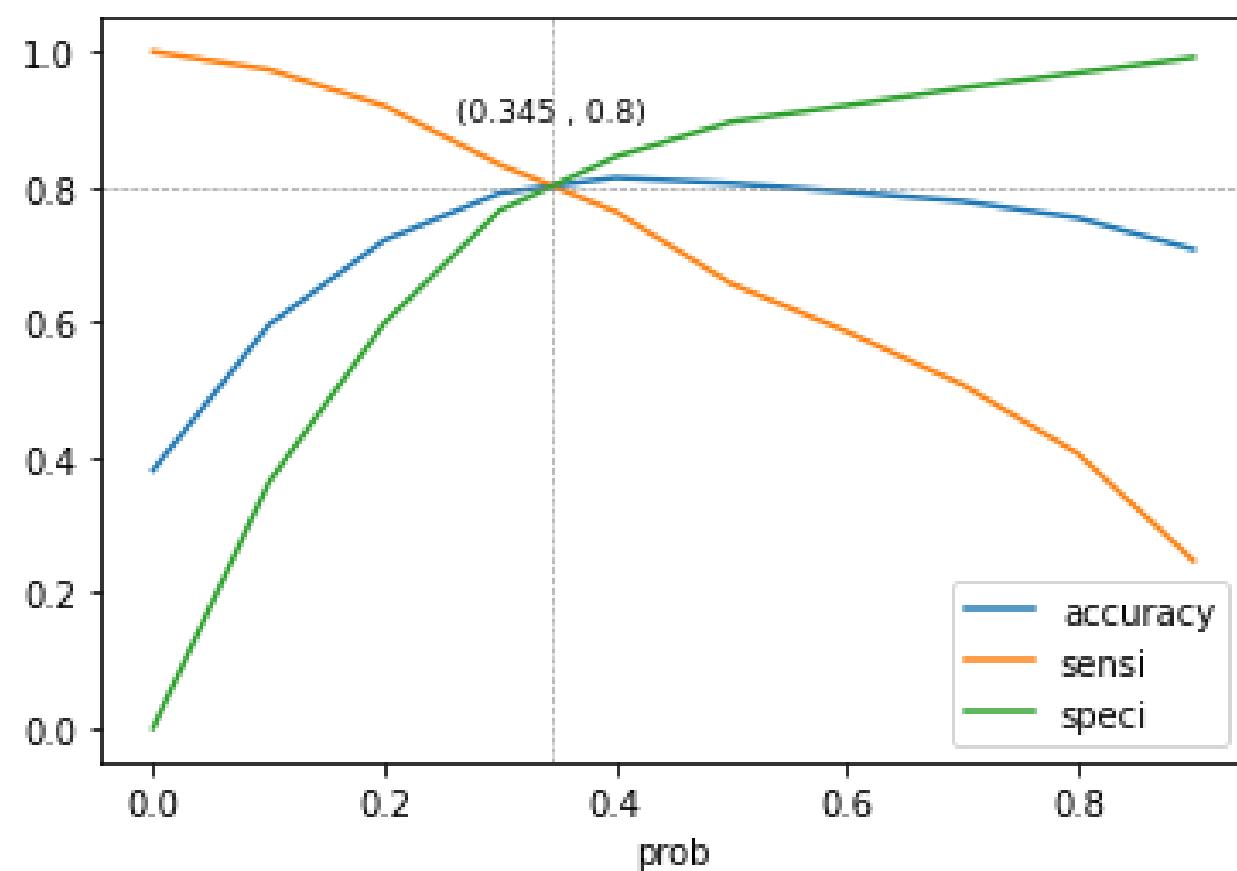


Linear Regression Final Model Parameters

Area under ROC = 0.88

Intermediate cut-off = 0.345

Final cut-off = 0.42



INFERENCES

Model Analysis

Performance of our Final Model
after model building and
selecting features.

Overall accuracy on Test
set: 0.805

Sensitivity of our logistic
regression model: 0.656

Specificity of our logistic
regression model: 0.896

INFERENCES FROM MODEL

Business Insights from
the Model

- The factors include '**the number of total visits to the website**', '**the total time spent on the website**', and '**the page views per visit**' have the highest impact on the probability of a lead getting converted.
- Lead Source_Welingak Website, Lead Source_Reference, Current_occupation_Working Professional are the top 3 features that contribute positively in getting hot leads.

INFERENCES FROM MODEL

Business Insights from
the Model

The top three variables in my model, that should be focused on are:

1. Lead Origin_Landing Page Submission
2. Specialization_Others
3. Specialization_Hospitality Management

Some specializations have negative impact on leads conversion, even chat conversation impacts negatively.

RECOMMENDATIONS

- Focus on High-Probability Leads: Prioritize leads that have been predicted as '1' by the model, indicating a higher likelihood of conversion. Allocate resources and prioritize phone calls specifically to these leads.
- Increase Phone Call Volume: Increase the number of phone calls made by the sales team to reach out to as many potential leads as possible. This will maximize the chances of conversion during this intensive phase.
- Personalized Communication: Ensure that phone calls are personalized and tailored to the specific needs and interests of each lead. Engage in meaningful conversations, address any concerns, and highlight the value proposition of X Education's offerings.
- Prompt Follow-up: After the initial phone call, promptly follow up with potential leads through email or additional phone calls. Maintain regular communication to keep the lead engaged and address any further questions or concerns they may have.
- Offer Incentives or Limited-Time Promotions: Provide attractive incentives or limited-time promotions to create a sense of urgency and encourage potential leads to convert. Special discounts, exclusive offers, or additional benefits can motivate leads to take action.
- Track and Analyze Results: Continuously track and analyze the outcomes of these aggressive conversion efforts. Monitor the conversion rates, assess the effectiveness of different strategies, and make necessary adjustments to optimize the process.

RECOMMENDATIONS

- To improve the lead conversion rate, it is crucial to effectively nurture the potential leads during the middle stage. This involves educating them about the product or service, maintaining constant communication, and building a relationship.
- Maintaining a list of leads is essential for keeping them informed about new courses, services, job opportunities, and future higher education options. It is important to monitor each lead closely to understand their specific interests and preferences. By tailoring the information sent to them, such as job offerings, course details, or relevant resources, the chances of capturing their interest and converting them into prospects increase.
- The key is to develop a comprehensive plan to understand the unique needs of each lead. This involves engaging in question-answer sessions with the leads to gather relevant information about their background, preferences, and goals. Based on these insights, further inquiries and appointments can be made to assess the lead's intention and mindset regarding joining online courses.
- By focusing on the leads that have already converted into customers, the sales team can gain valuable insights and feedback. This information can be used to refine the lead nurturing process and better align the offerings with the needs and expectations of potential leads.

**THANK
YOU!**

