LEAD-SCORING

CASE STUDY

By: Payal Sharma

Problem Statement

➢ X Education sells online courses to industry professionals.

➢ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

➢ 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 as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective:

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Solution Methodology

1) Data cleaning and data manipulation.

- ➢ Check and handle duplicate data.
- ➢ Check and handle NA values and missing values.
- ➢ Drop columns, if it contains large amount of missing values and not useful for the analysis.
- ➢ Imputation of the values, if necessary.
- ➢ Check and handle outliers in data.

2) EDA

- ➢Uni-variate data analysis: value count, distribution of variable etc.
- ➢ Bi-variate data analysis: correlation coefficients and pattern between the variables etc.

Data Manipulation

➢ Total Number of Columns=37, Total Number of Rows =9240.

➢ Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply”

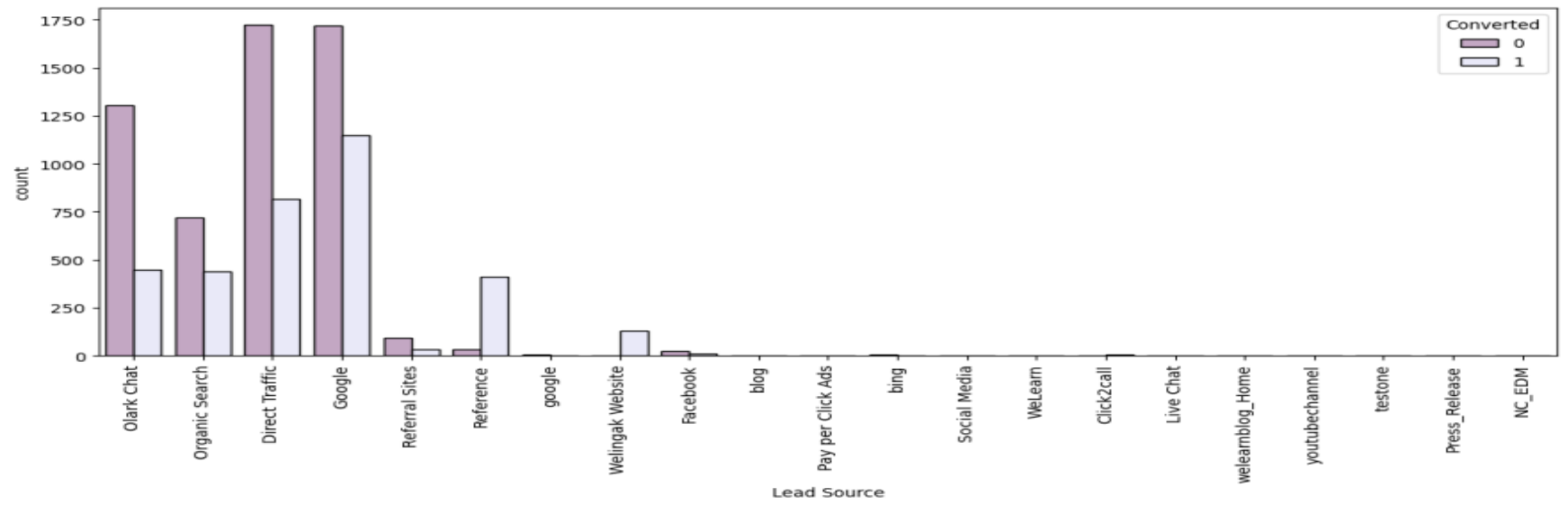
➢ Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque” etc. have been dropped.

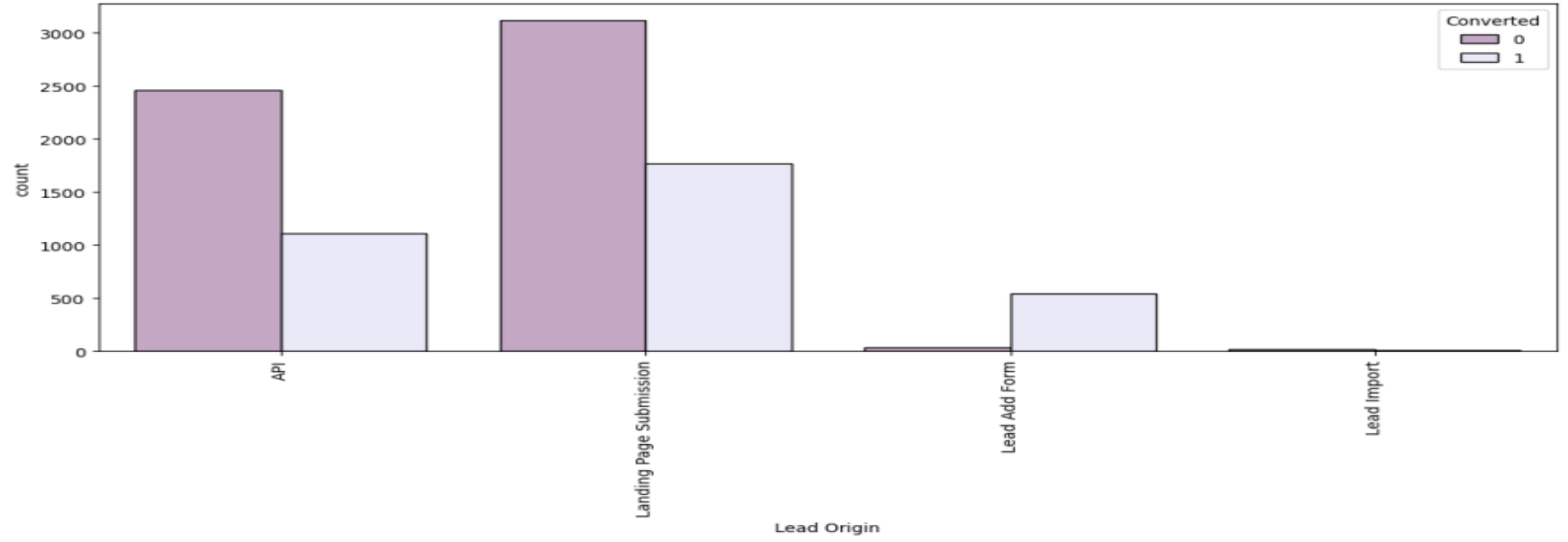
➢ Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.

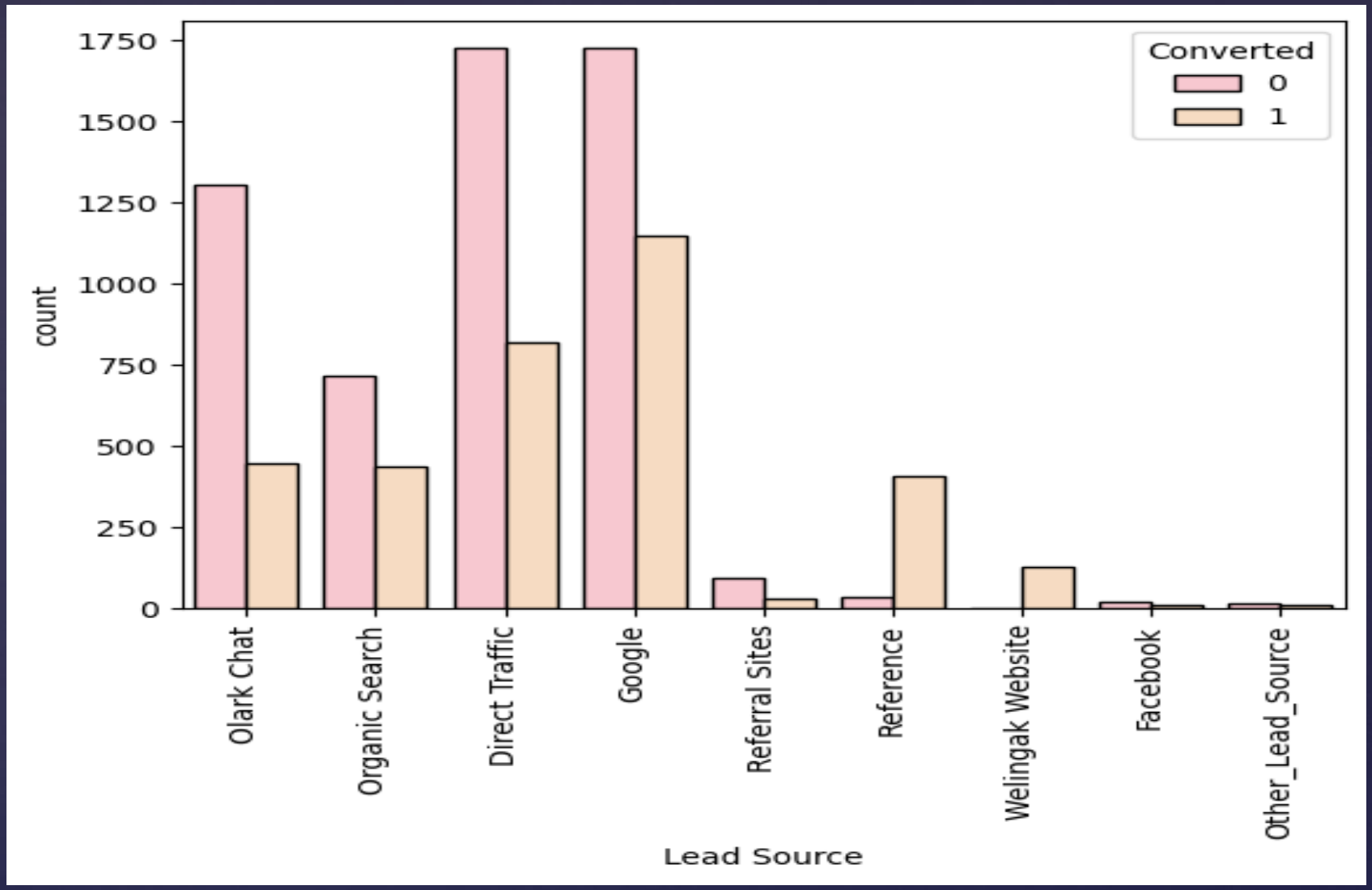
➢ After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.

➢ Dropping the columns having more than 35% as missing value such as ‘How did you hear about X Education’ and ‘Lead Profile’.

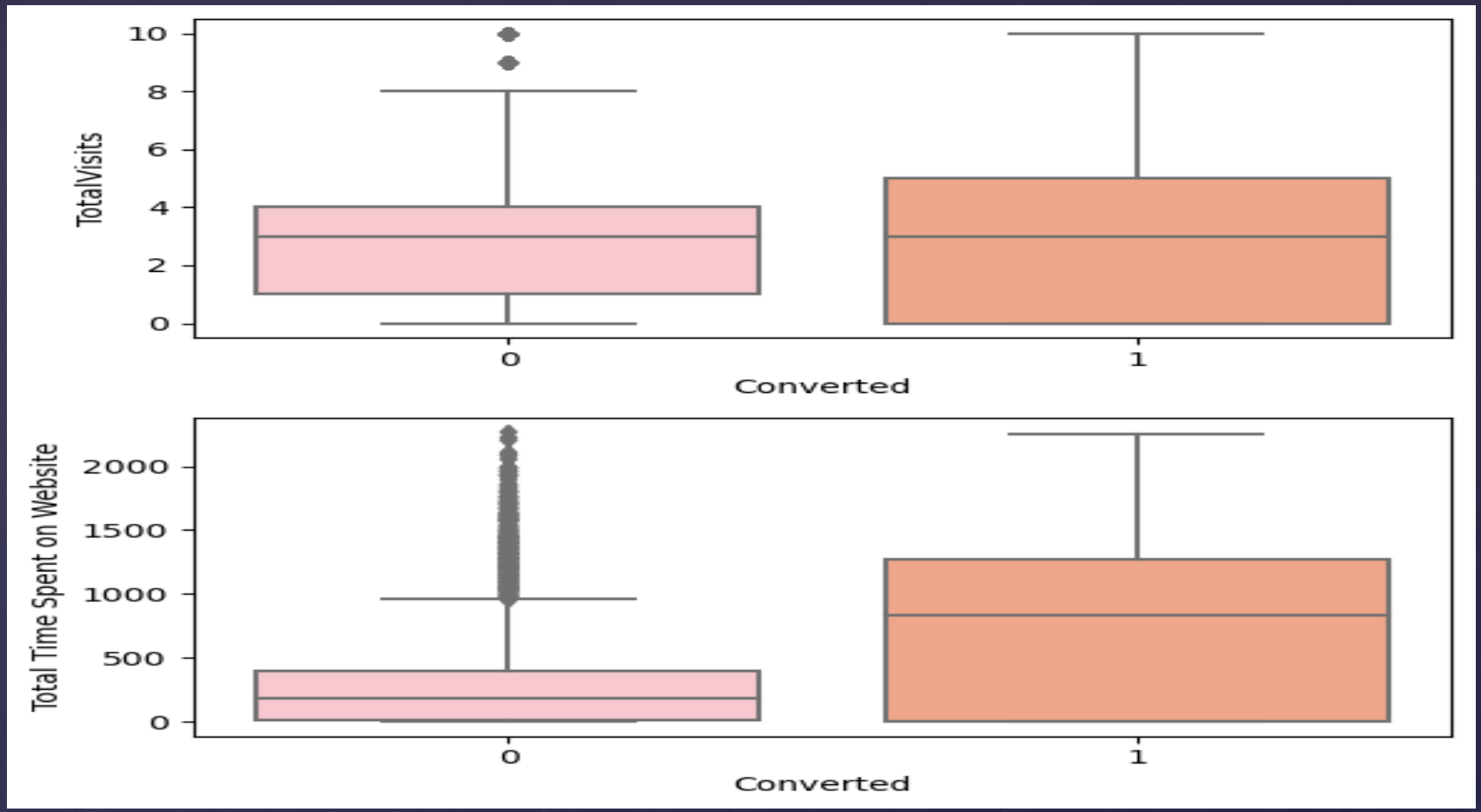
EDA


Converted
0
1
count
1750
1500
1250
1000
750
500
250
0
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Reference
google
Welingak Website
Facebook
blog
Pay per Click Ads
bing
Social Media
WeLearn
Click2call
Live Chat
welearnblog_Home
youtubechannel
testone
Press_Release
NC_EDM
Lead Source


Converted
0
1
count
3000
2500
2000
1500
1000
500
0
API
Landing Page Submission
Lead Add Form
Lead Import
Lead Origin

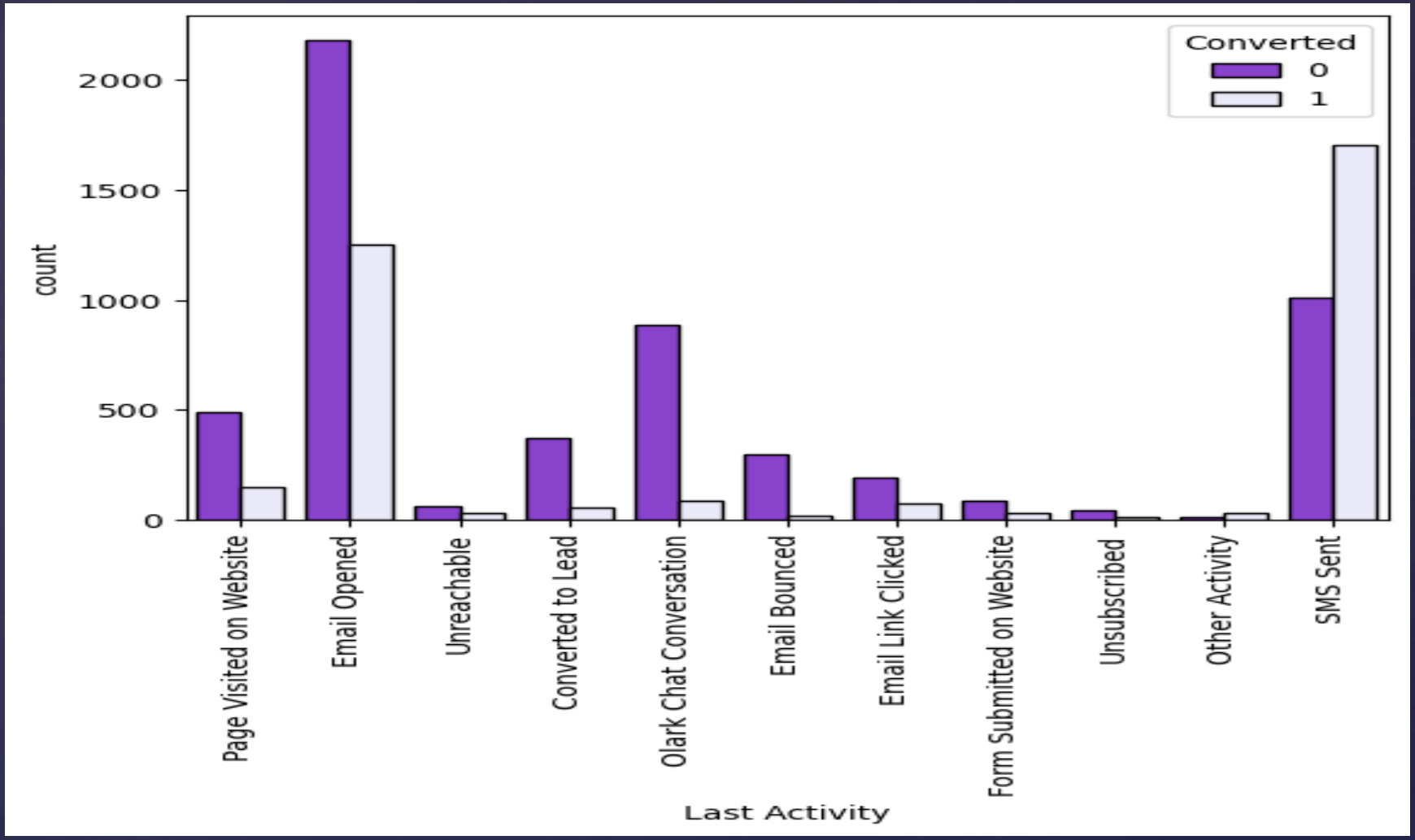


The count of leads from the Google and Direct Traffic is maximum

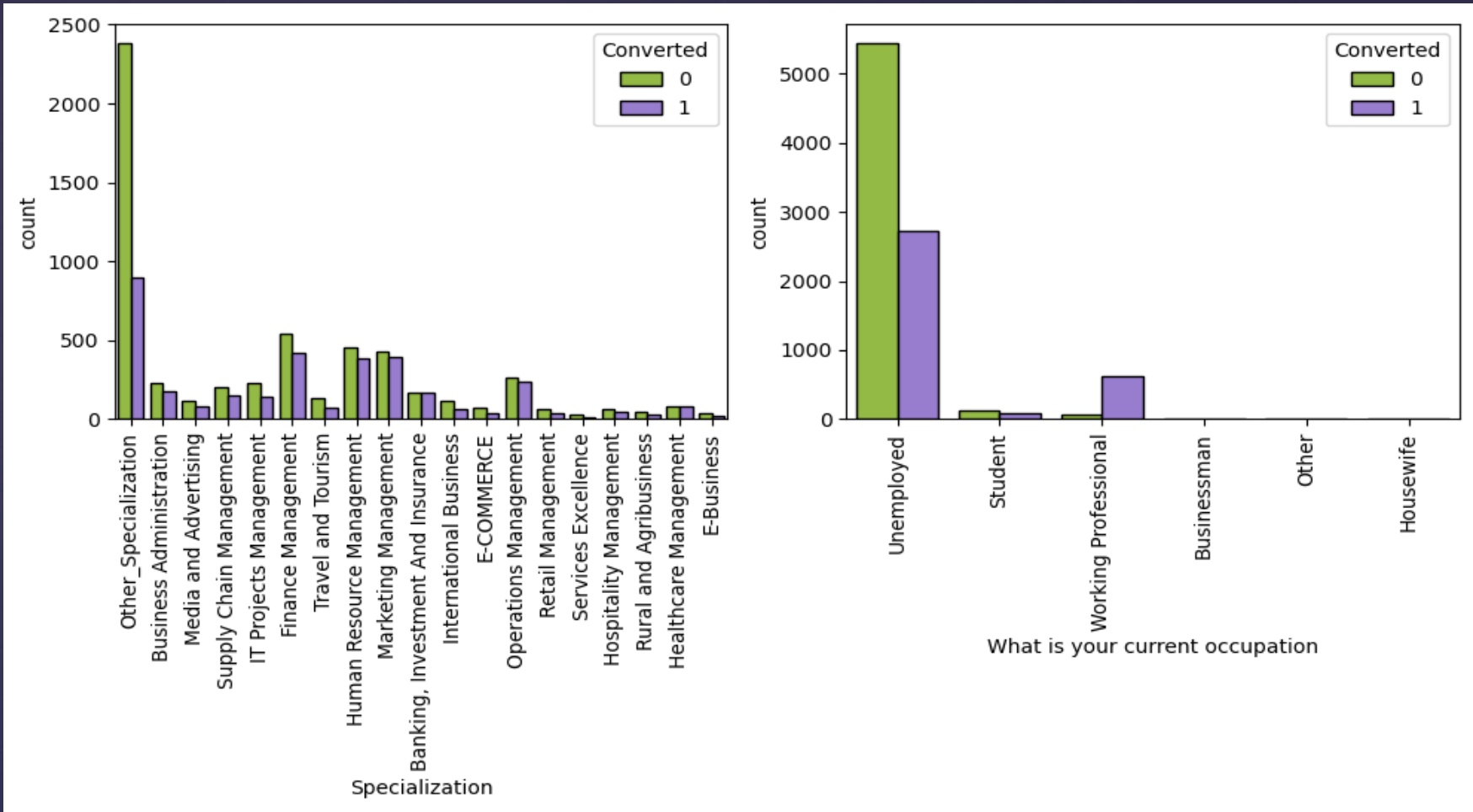


a)The median of both the conversion and non-conversion are same and hence nothing conclusive can be said using this information

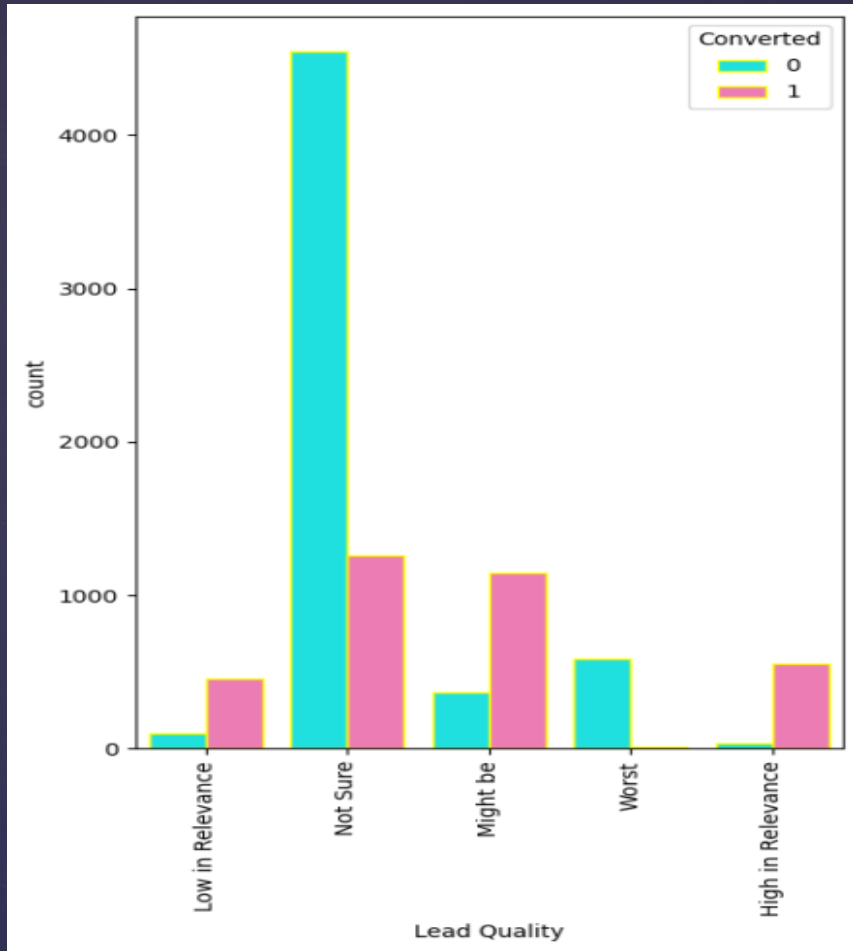
b)Users spending more time on the website are more likely to get converted

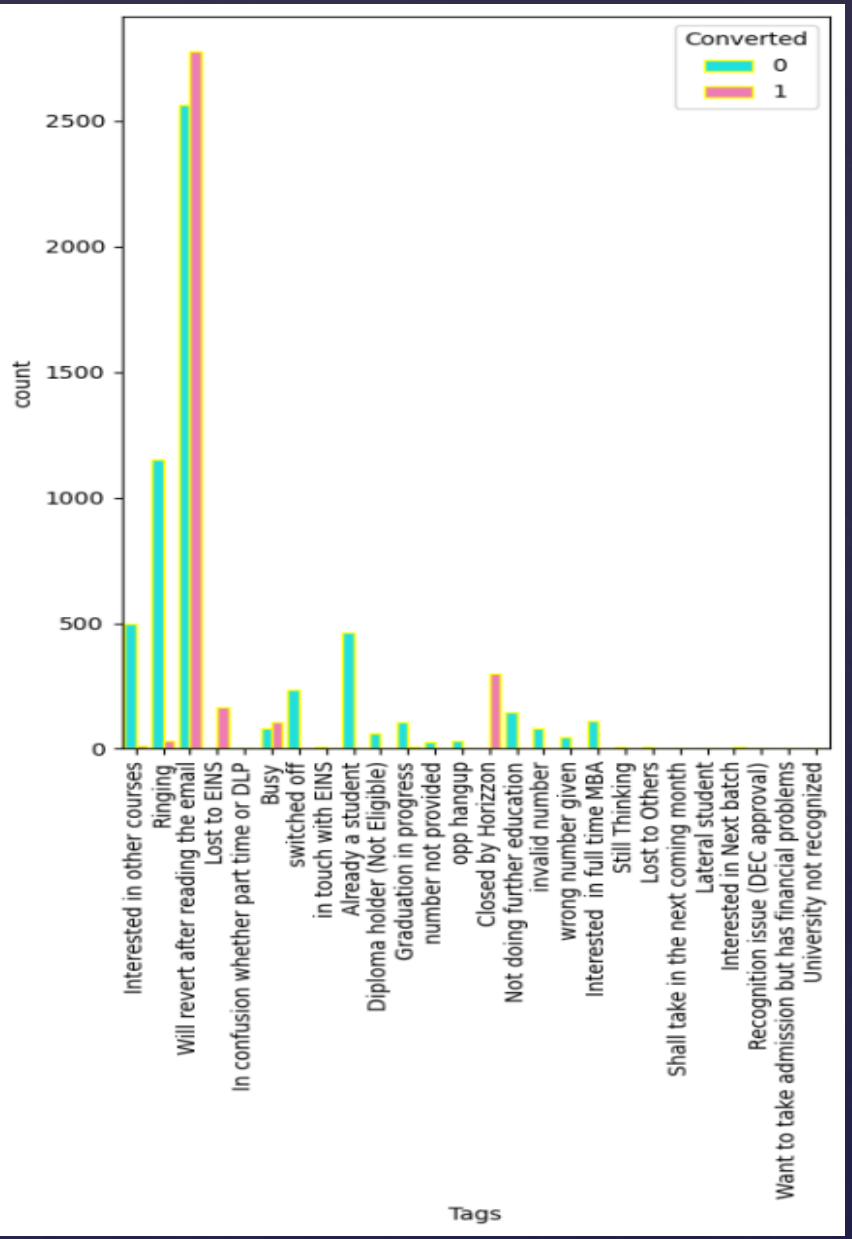


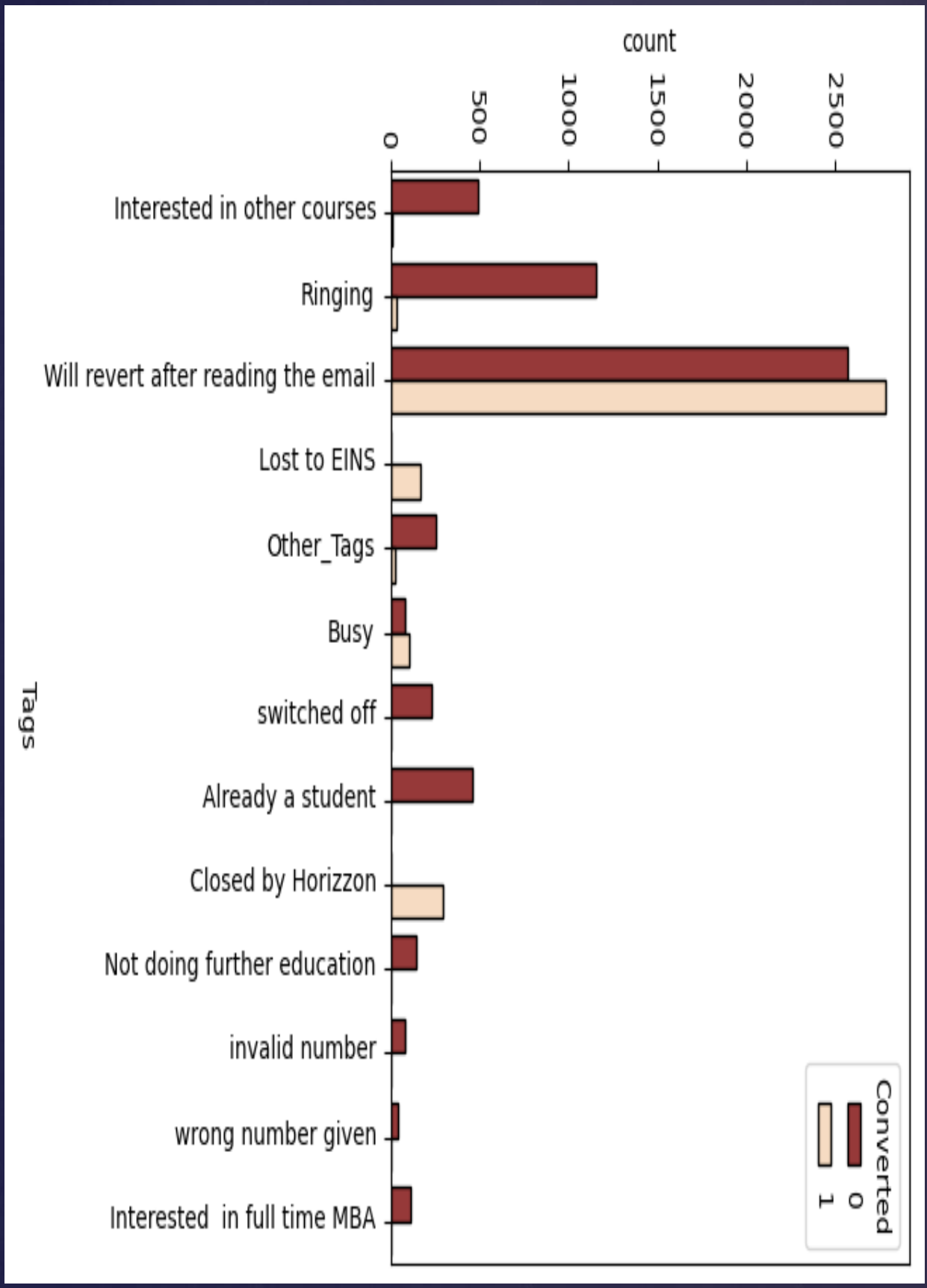
a)The count of lst activity as "Email Opened" is max
b)The conversion rate of SMS sent as last activity is maximum



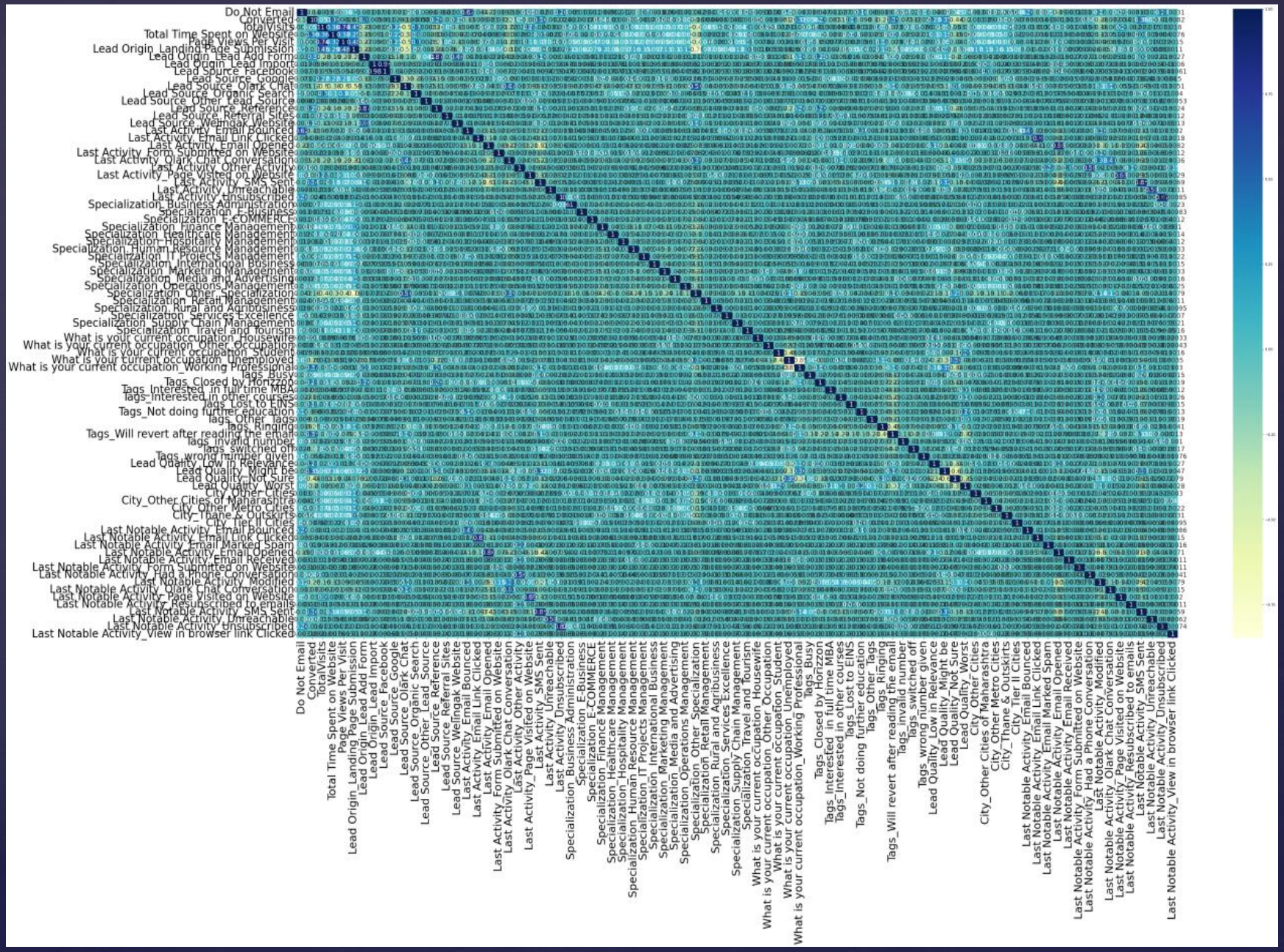
a)Looking at above plot, no particular inference can be made for Specialization
b)Looking at above plot, we can say that working professionals have high conversion rate
c)Number of Unemployed leads are more than any other category


Converted
0
1
count
4000
3000
2000
1000
0
Low in Relevance
Not Sure
Might be
Worst
High in Relevance
Lead Quality


Converted
0
1
count
2500
2000
1500
1000
500
0
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
In confusion whether part time or DLP
Busy
switched off
in touch with EINS
Already a student
Diploma holder (Not Eligible)
Graduation in progress
number not provided
opp hangup
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
Shall take in the next coming month
Lateral student
Interested in Next batch
Recognition issue (DEC approval)
Want to take admission but has financial problems
University not recognized
Tags


count
0
500
1000
1500
2000
2500
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
Other_Tags
Busy
switched off
Already a student
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Tags
Converted
0
1

a) To improve the overall lead conversion rate, we need to focus on increasing the conversion rate of 'API' and 'Landing Page Submission' Lead Origins and also increasing the number of leads from 'Lead Add Form'
b) To improve the overall lead conversion rate, we need to focus on increasing the conversion rate of 'Google', 'Olark Chat', 'Organic Search', 'Direct Traffic' and also increasing the number of leads from 'Reference' and 'Welingak Website'
c) Websites can be made more appealing so as to increase the time of the Users on websites
d) We should focus on increasing the conversion rate of those having last activity as Email Opened by making a call to those leads and also try to increase the count of the ones having last activity as SMS sent
e) To increase overall conversion rate, we need to increase the number of Working Professional leads by reaching out to them through different social sites such as LinkedIn etc. and also on increasing the conversion rate of Unemployed leads
f) We also observed that there are multiple columns which contains data of a single value only. As these columns do not contribute towards any inference, we can remove them from further analysis


Do Not Email
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Lead Origin_Landing Page Submission
Lead Origin_Lead Add Form
Lead Origin_Lead Import
Lead Source_Facebook
Lead Source_Google
Lead Source_Olark Chat
Lead Source_Organic Search
Lead Source_Other_Lead Source
Lead Source_Reference
Lead Source_Referral Sites
Lead Source_Welingak Website
Last Activity_Email Bounced
Last Activity_Email Link Clicked
Last Activity_Email Opened
Last Activity_Form Submitted on Website
Last Activity_Olark Chat Conversation
Last Activity_Other Activity
Last Activity_Page Visited on Website
Last Activity_SMS Sent
Last Activity_Unreachable
Last Activity_Unsubscribed
Specialization_Business Administration
Specialization_E-Business
Specialization_E-COMMERCE
Specialization_Finance Management
Specialization_Healthcare Management
Specialization_Hospitality Management
Specialization_Human Resource Management
Specialization_IT Projects Management
Specialization_International Business
Specialization_Marketing Management
Specialization_Media and Advertising
Specialization_Operations Management
Specialization_Other_Specialization
Specialization_Retail Management
Specialization_Rural and Agribusiness
Specialization_Services Excellence
Specialization_Supply Chain Management
Specialization_Travel and Tourism
What is your current occupation_Housewife
What is your current occupation_Other_Occupation
What is your current occupation_Student
What is your current occupation_Unemployed
What is your current occupation_Working Professional
Tags_Busy
Tags_Closed by Horizzon
Tags_Interested_in full time MBA
Tags_Interested_in other courses
Tags_Lost to EINS
Tags_Not doing further education
Tags_Other_Tags
Tags_Ringing
Tags_Will revert after reading the email
Tags_invalid number
Tags_switched off
Tags_wrong number given
Lead Quality_Low in Relevance
Lead Quality_Might be
Lead Quality_Not Sure
Lead Quality_Worst
City_Other Cities
City_Other Cities of Maharashtra
City_Other Metro Cities
City_Thane & Outskirts
City_Tier II Cities
Last Notable Activity_Email Bounced
Last Notable Activity_Email Link Clicked
Last Notable Activity_Email Marked Spam
Last Notable Activity_Email Opened
Last Notable Activity_Email Received
Last Notable Activity_Form Submitted on Website
Last Notable Activity_Had a Phone Conversation
Last Notable Activity_Modified
Last Notable Activity_Olark Chat Conversation
Last Notable Activity_Page Visited on Website
Last Notable Activity_Resubscribed to emails
Last Notable Activity_SMS Sent
Last Notable Activity_Unreachable
Last Notable Activity_Unsubscribed
Last Notable Activity_View in browser link Clicked

Data Conversion Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 79.43% on test data, and 79.17% on train data set

Finding Optimal Cut off Point

- 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.
- From Precision and recall chart suitable threshold is 0.38 or 0.39
- Although the values are same at both cutoff 0.35 and 0.38, so kept 0.35

CONCLUSION

It was found that the variables that mattered the most in the lead converted are

- Lead Origin
- Lead Add Form
- Do Not Email
- Yes
- Last Activity
- Converted to Lead
- Olark Chat Conversation
- What is your current occupation_
- Unemployed
- Working Professional
- Tags
- Busy
- Closed by Horizon
- Lost to EINS
- Will revert after reading the email
- in touch with EINS
- Last Notable Activity
- SMS Sent C