



Lead Scoring Case Study

By- Ajinkya Shakkarwar

Anjali Singh

Problem Statement



X Education , An education company named sells online courses to industry professionals



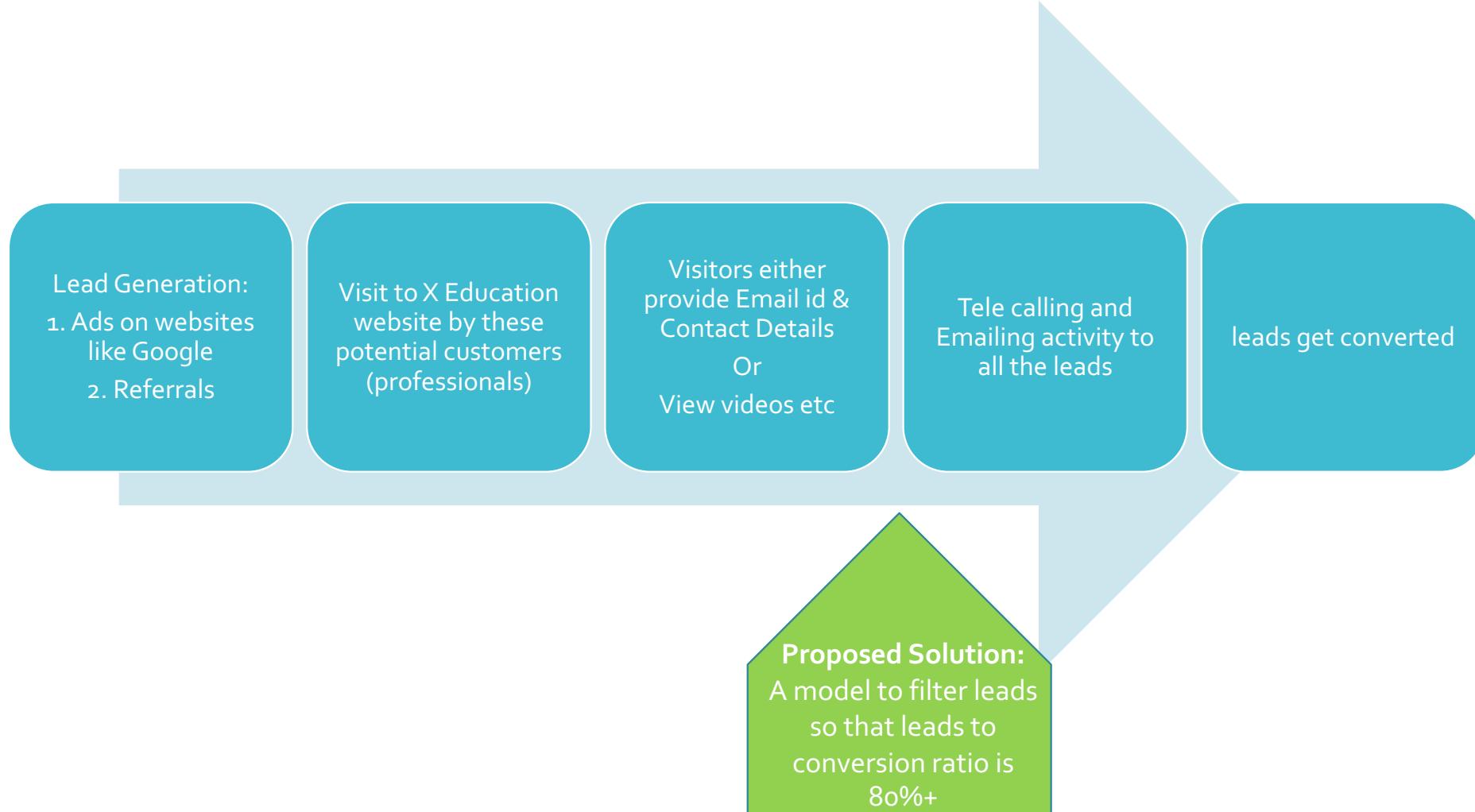
Many interested professionals land on their website



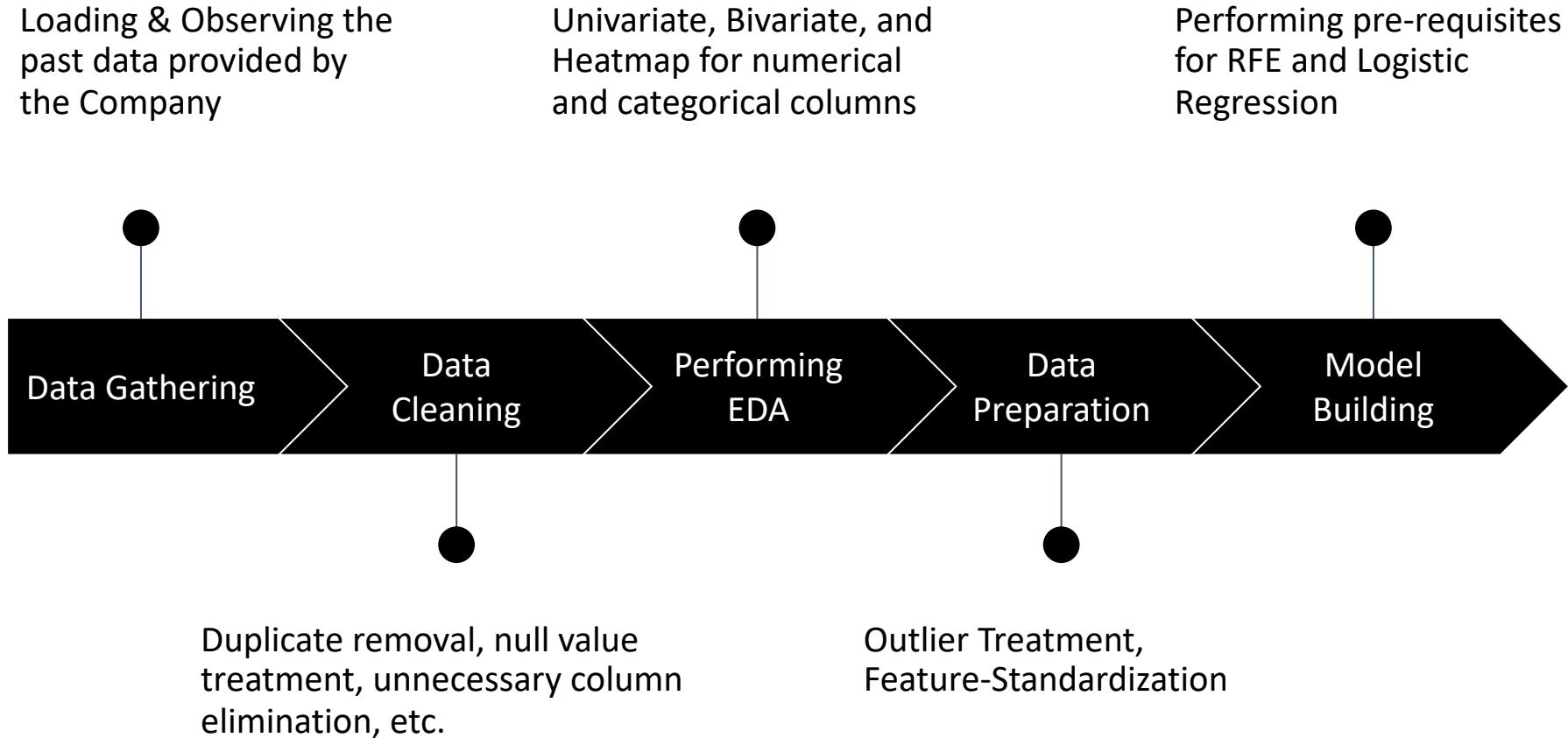
The company markets its courses on several websites like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos

- 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%
- X Education gets a lot of leads but 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 as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Process

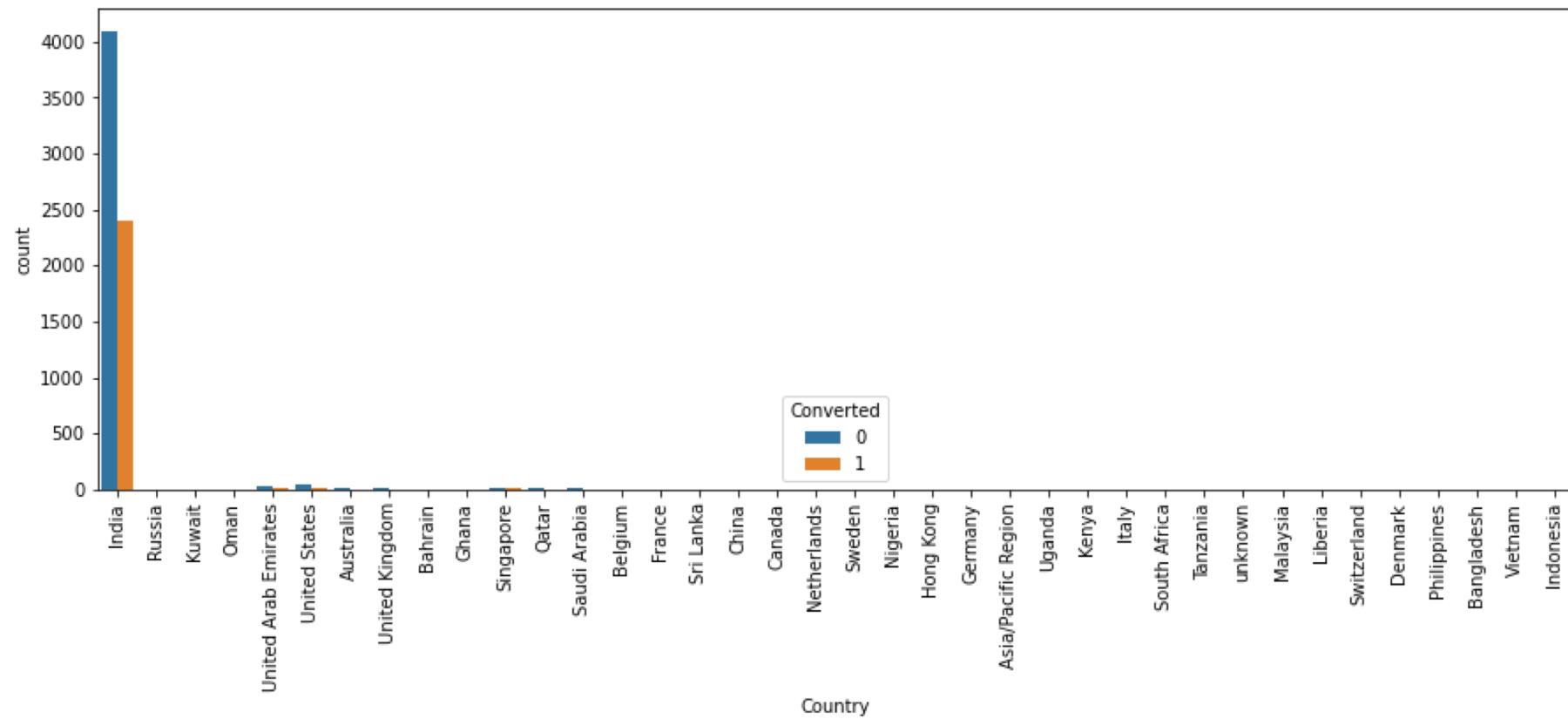


Implementation

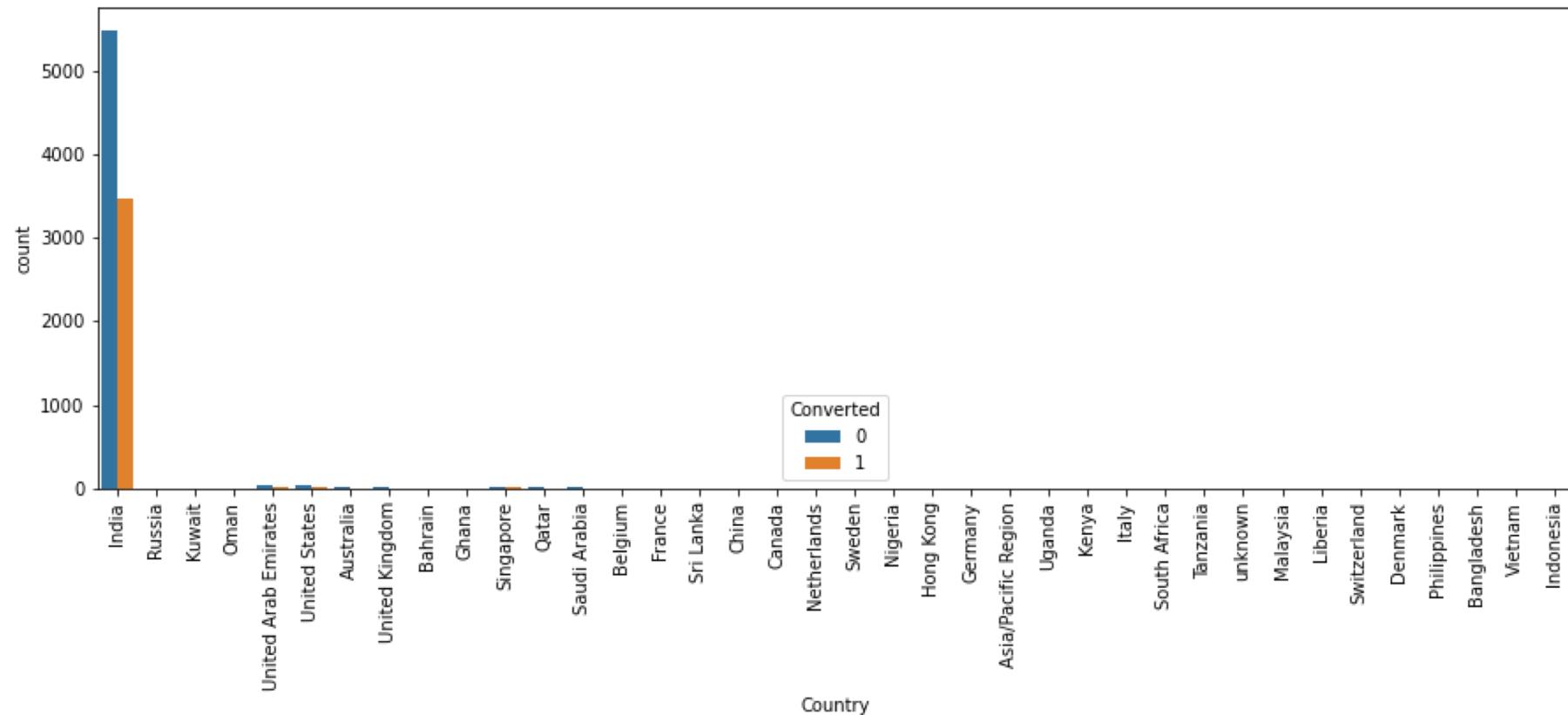


Plots (Visualization)

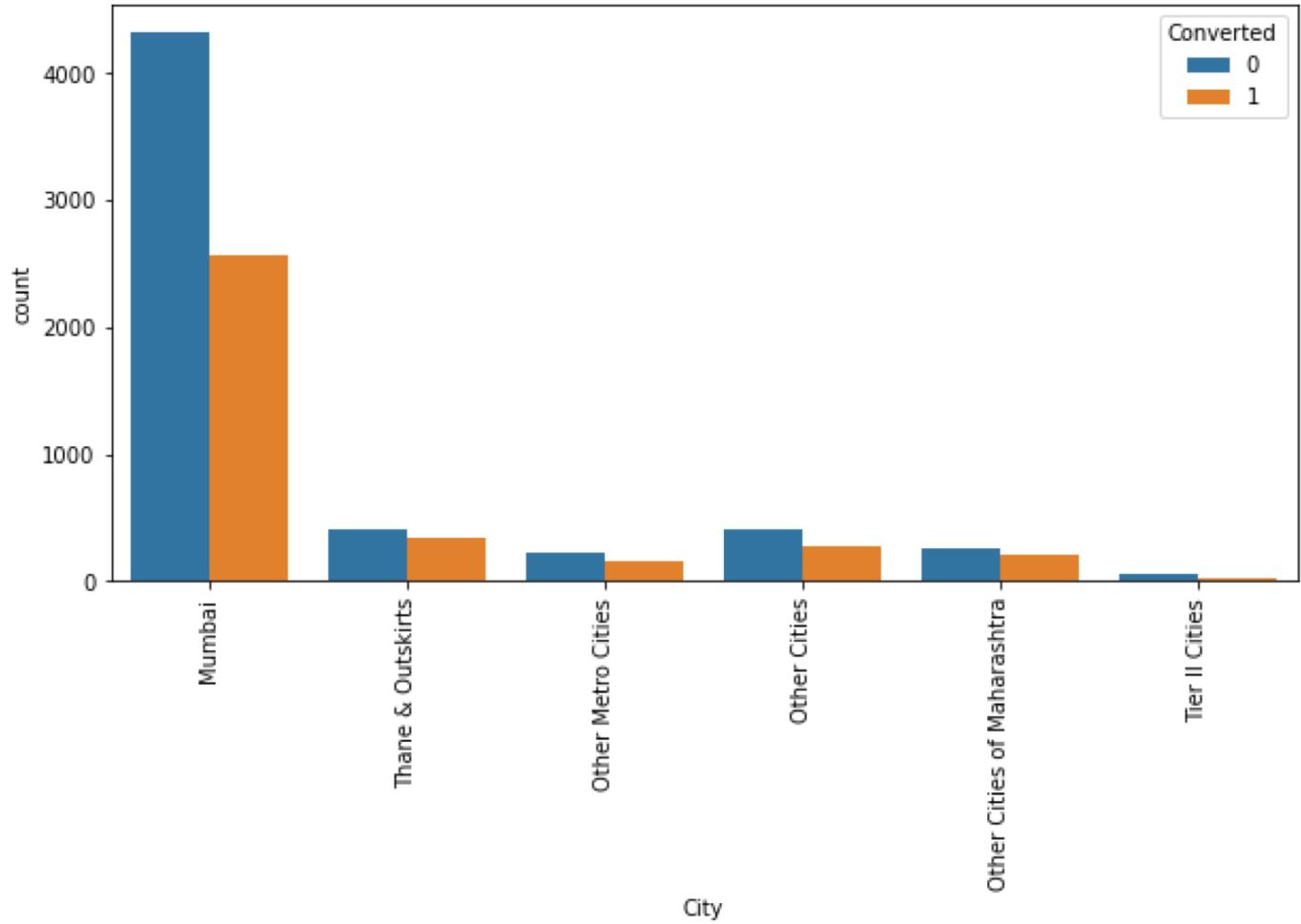
Plotting spread of Country column



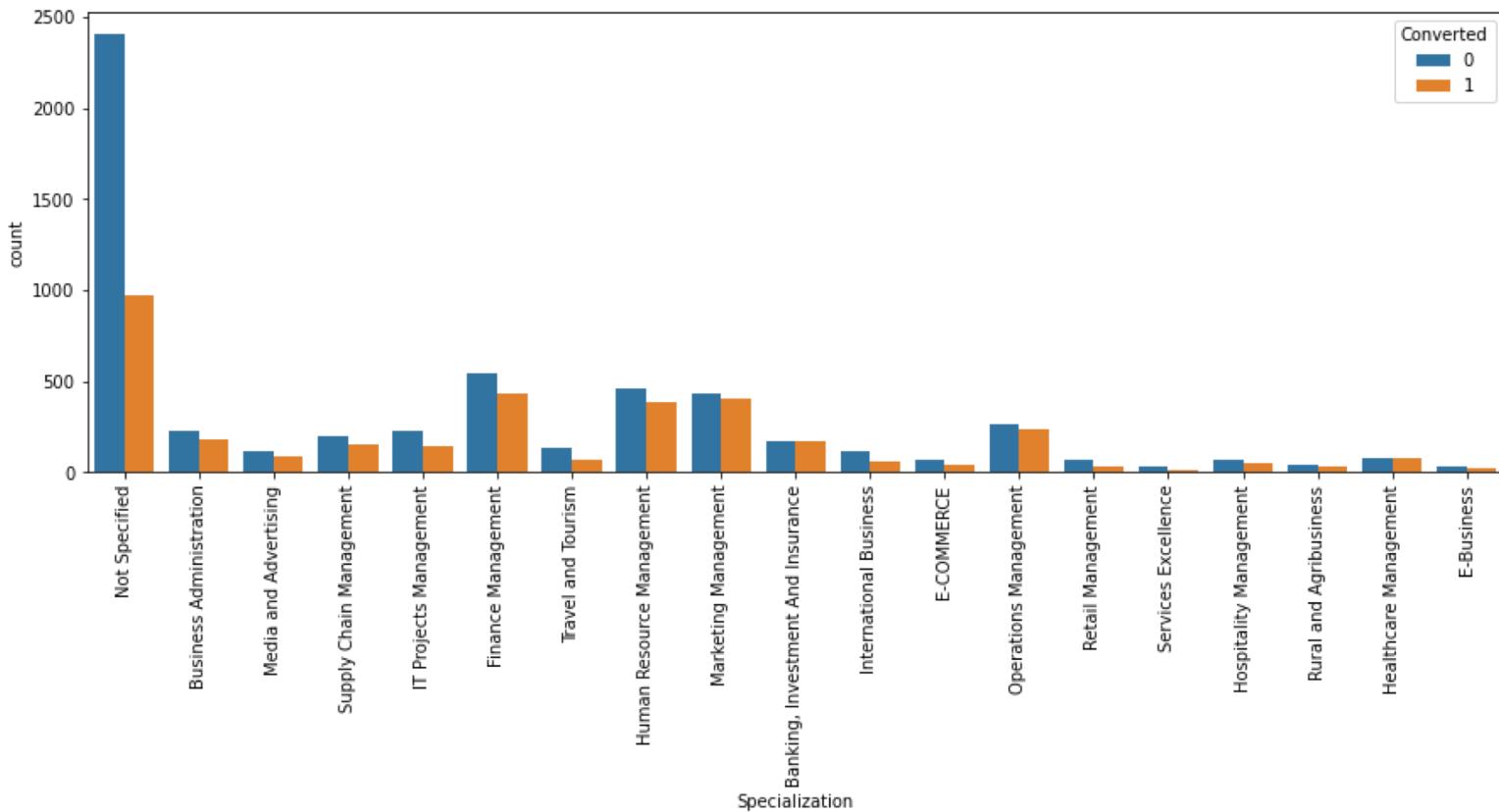
plotting spread of Country columnn after replacing NaN values

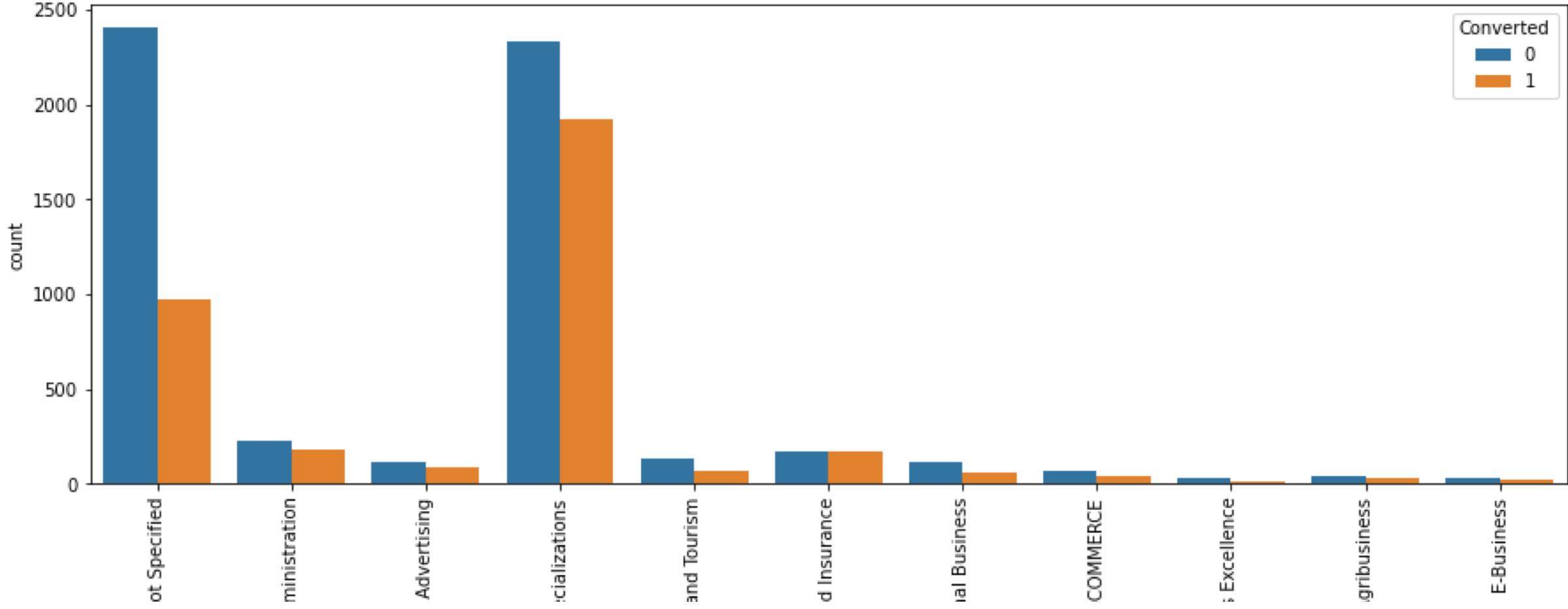


plotting
spread of
City column
after
replacing
NaN values



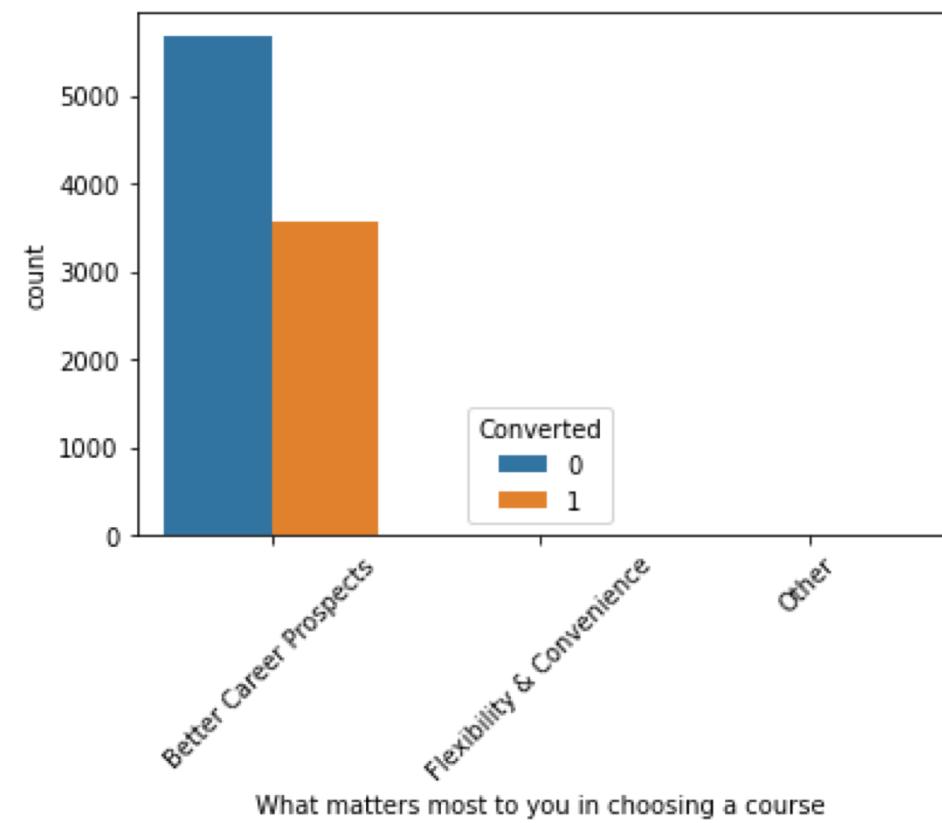
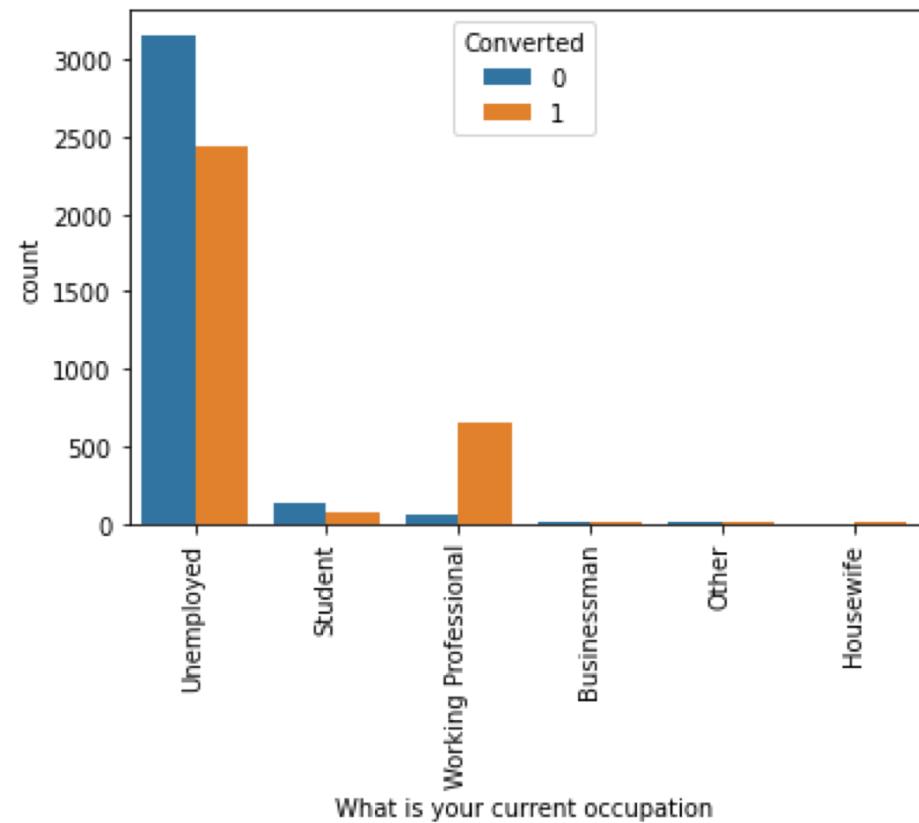
plotting spread of Specialization columnn

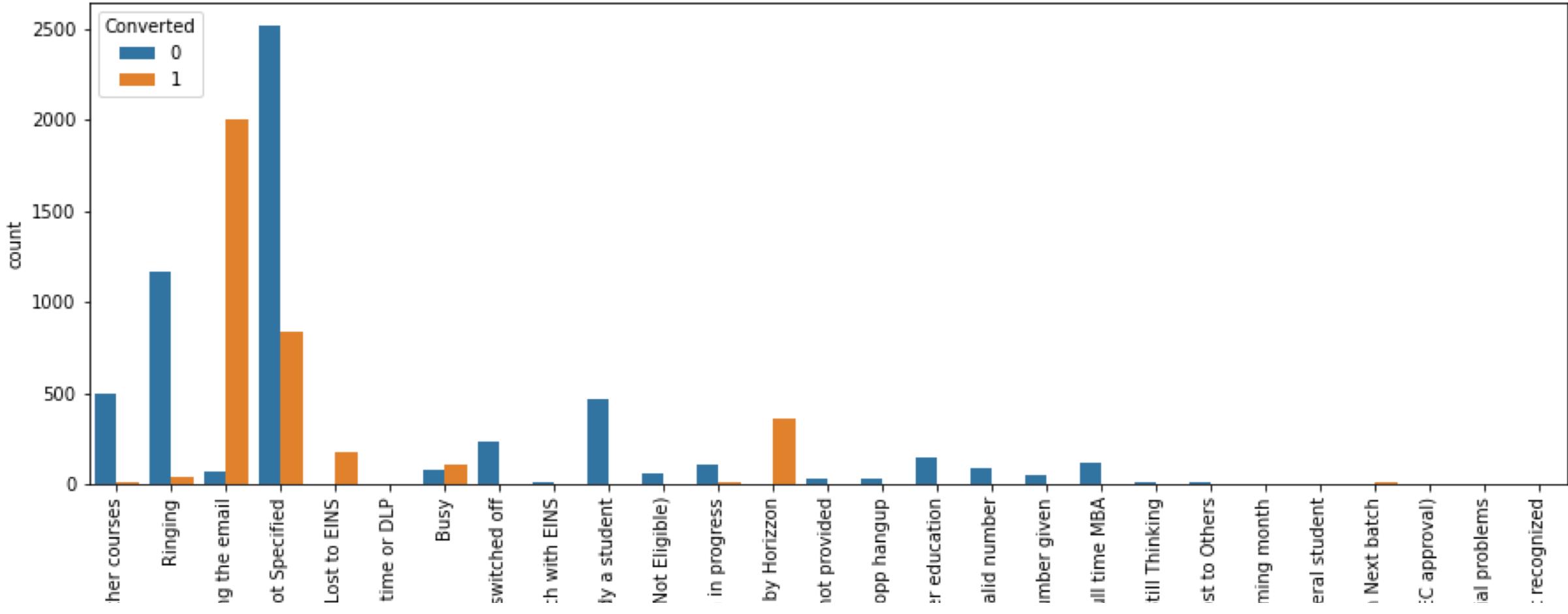




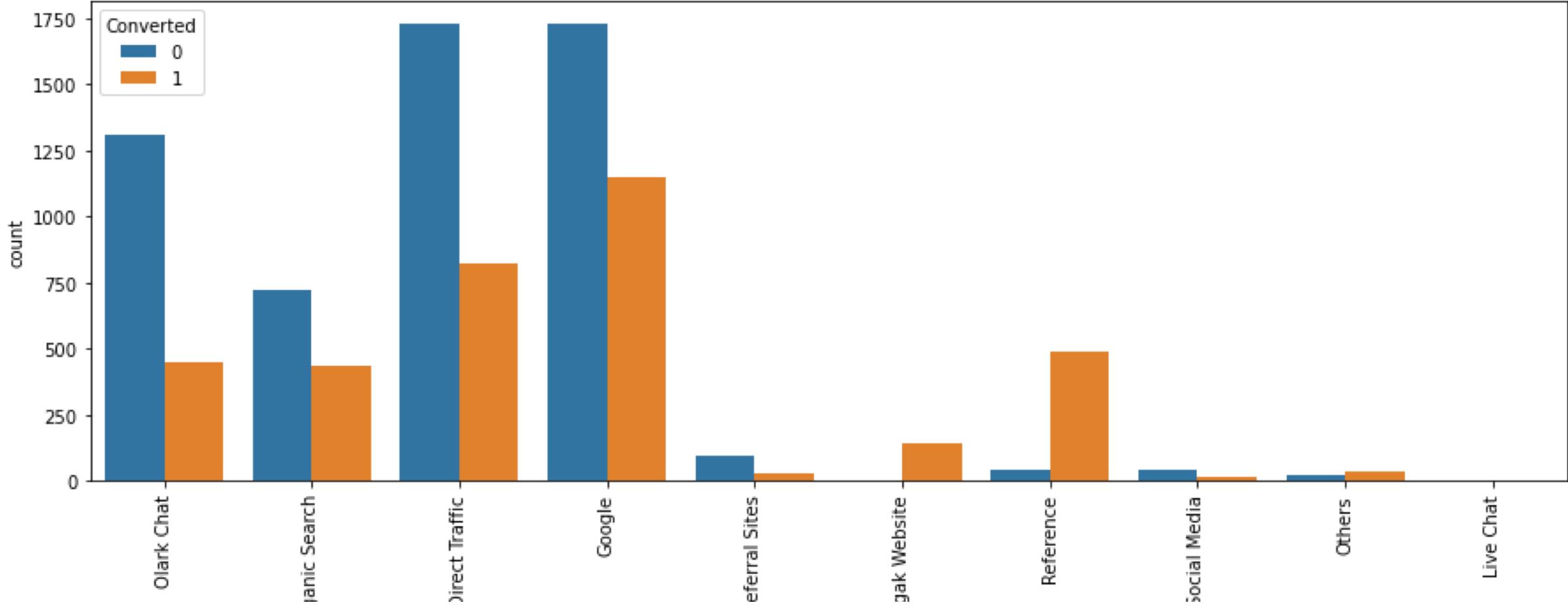
visualizing count of Variable
based on Converted value

visualizing count of Variable based on Converted value

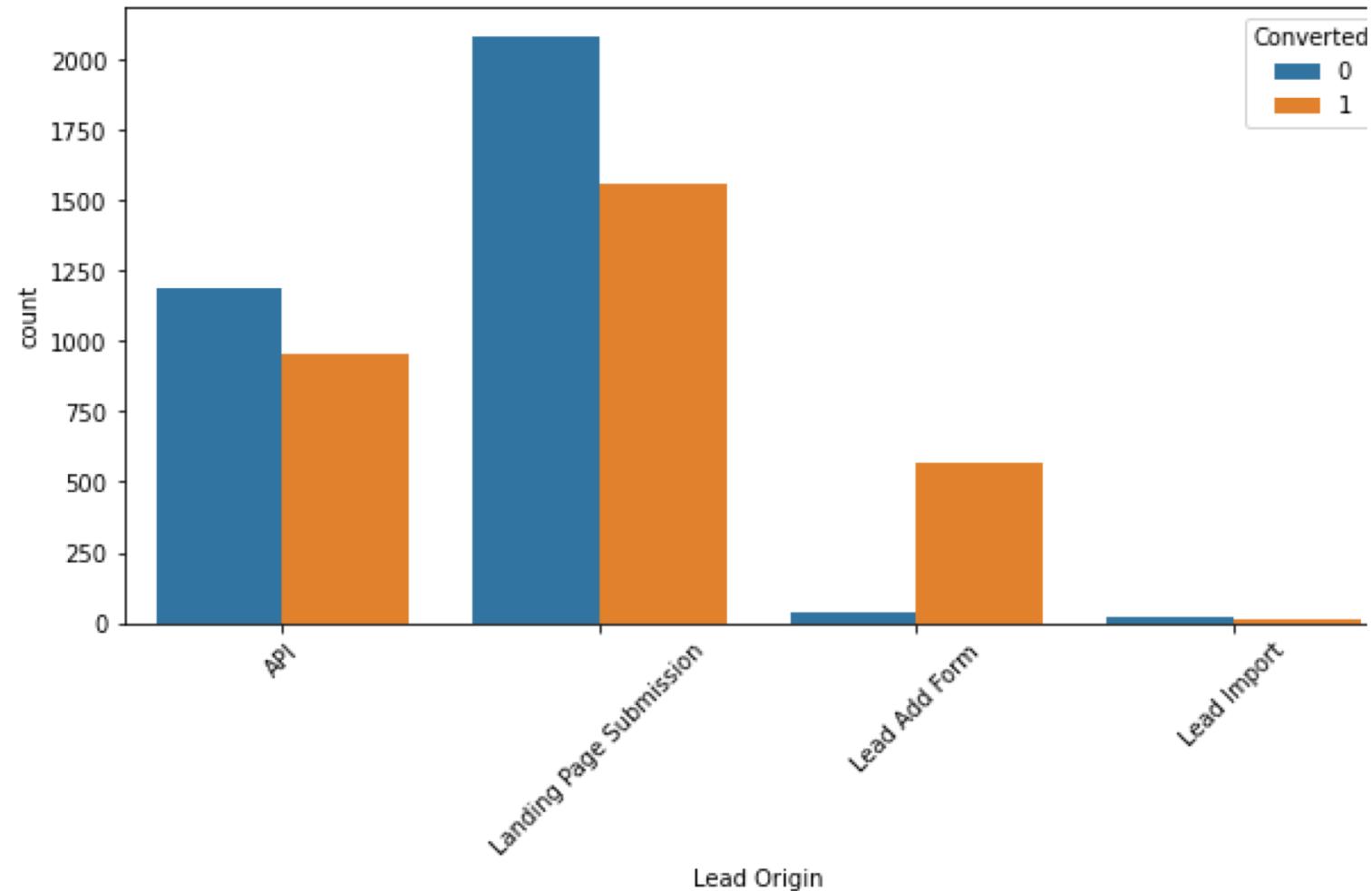




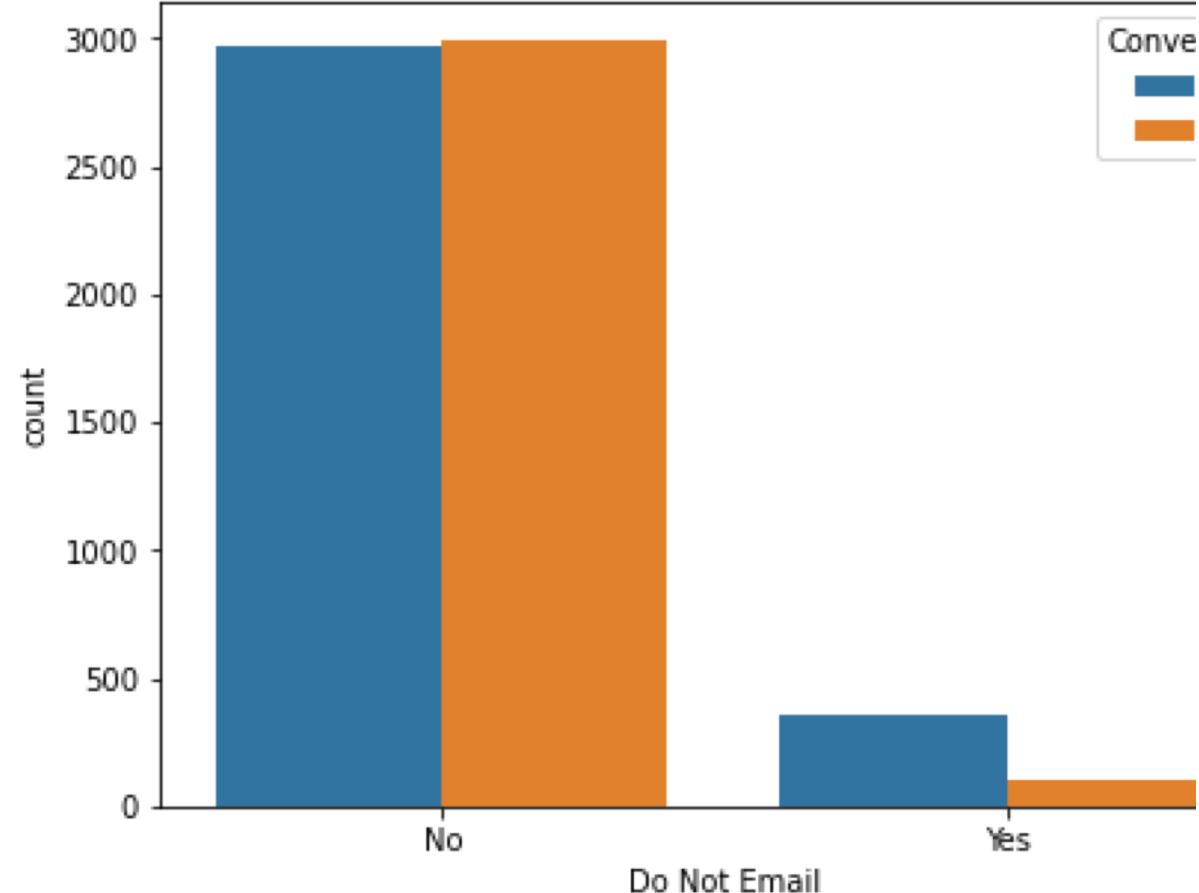
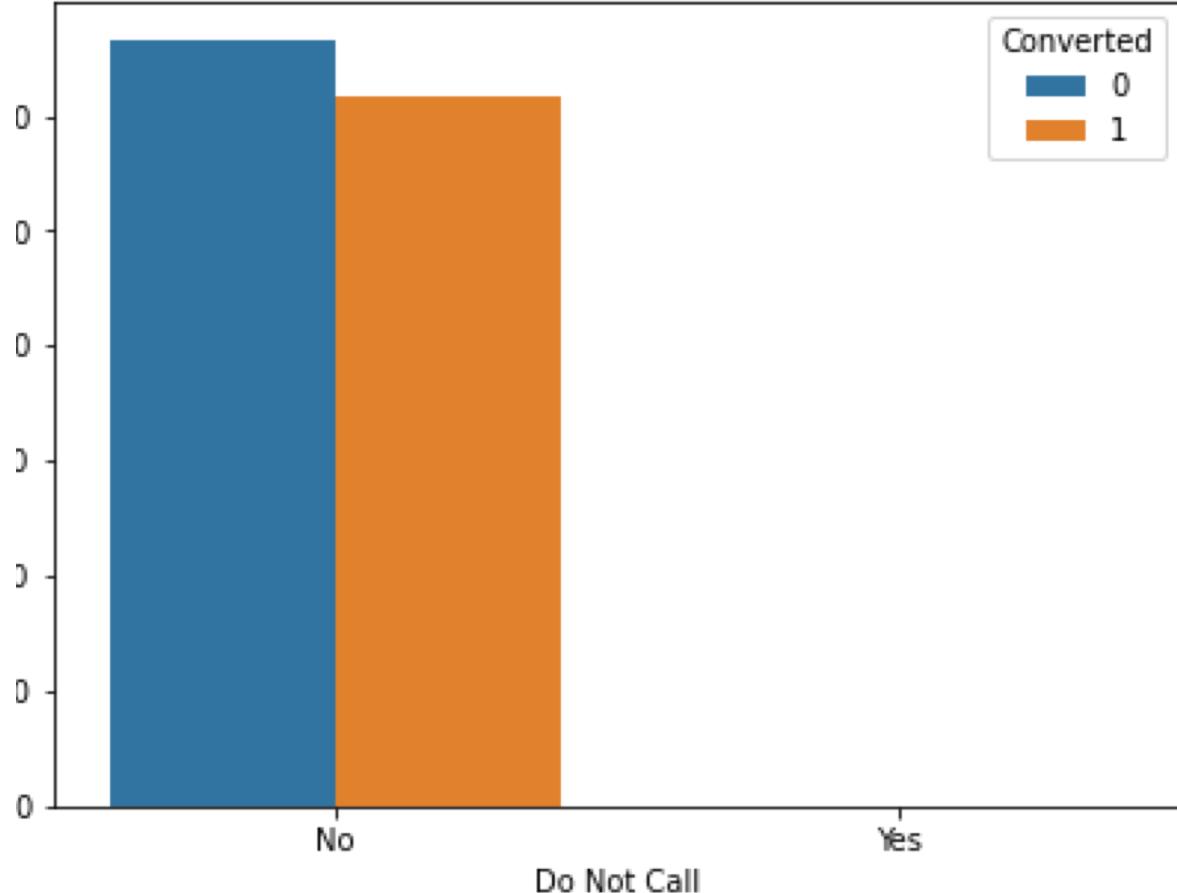
visualizing count of Variable based
on Converted value



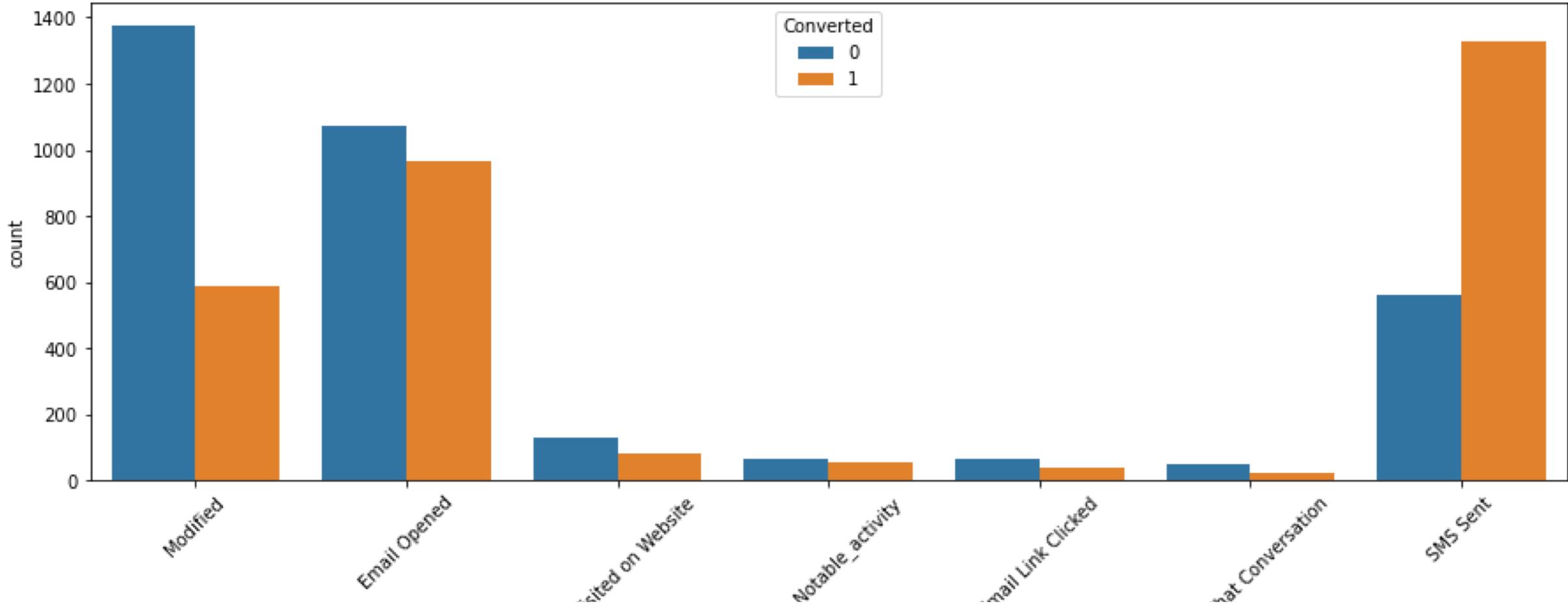
visualizing count of Variable based
on Converted value |



visualizing
count of
Variable
based on
Converted
value

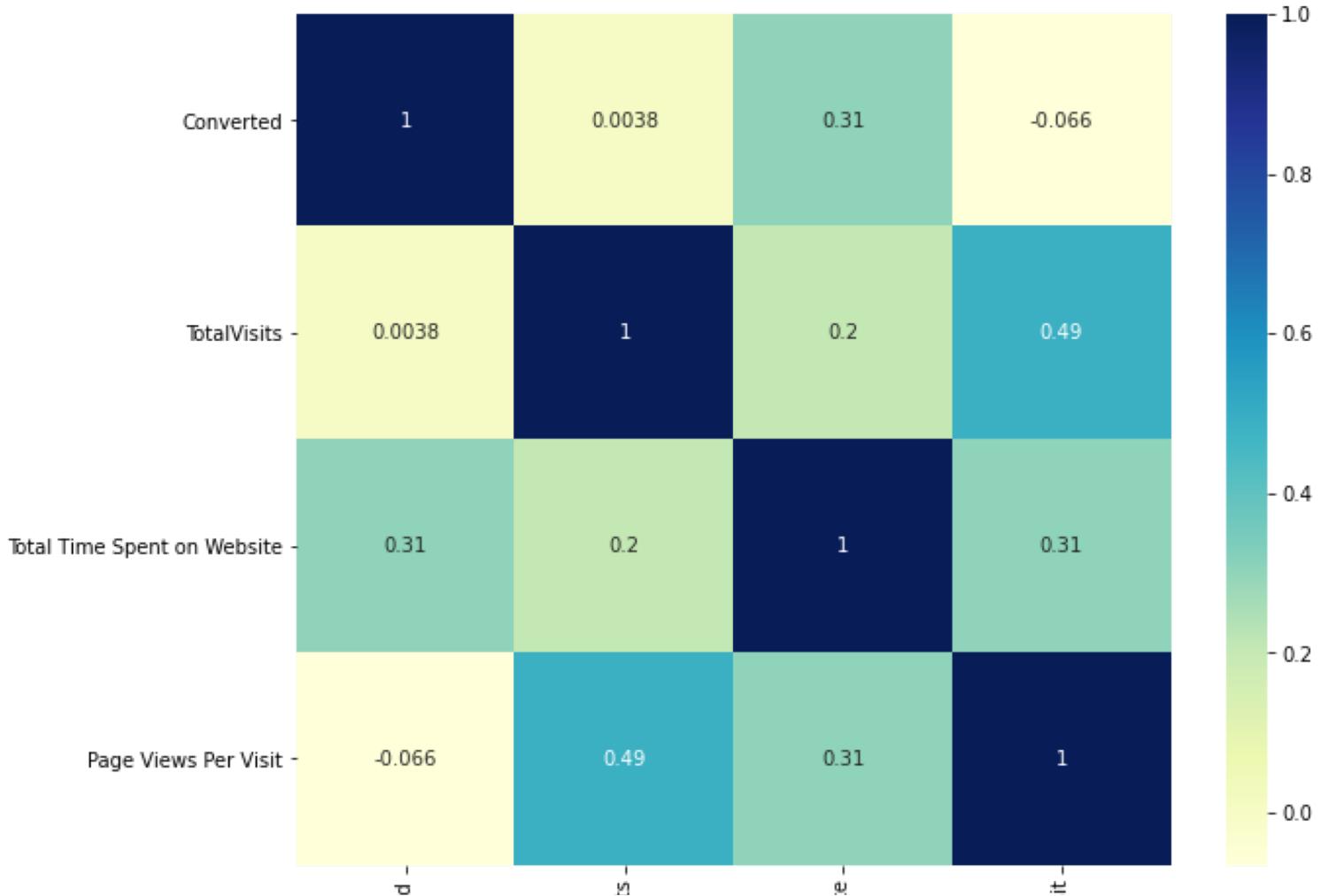


visualizing count of Variable based
on Converted value |

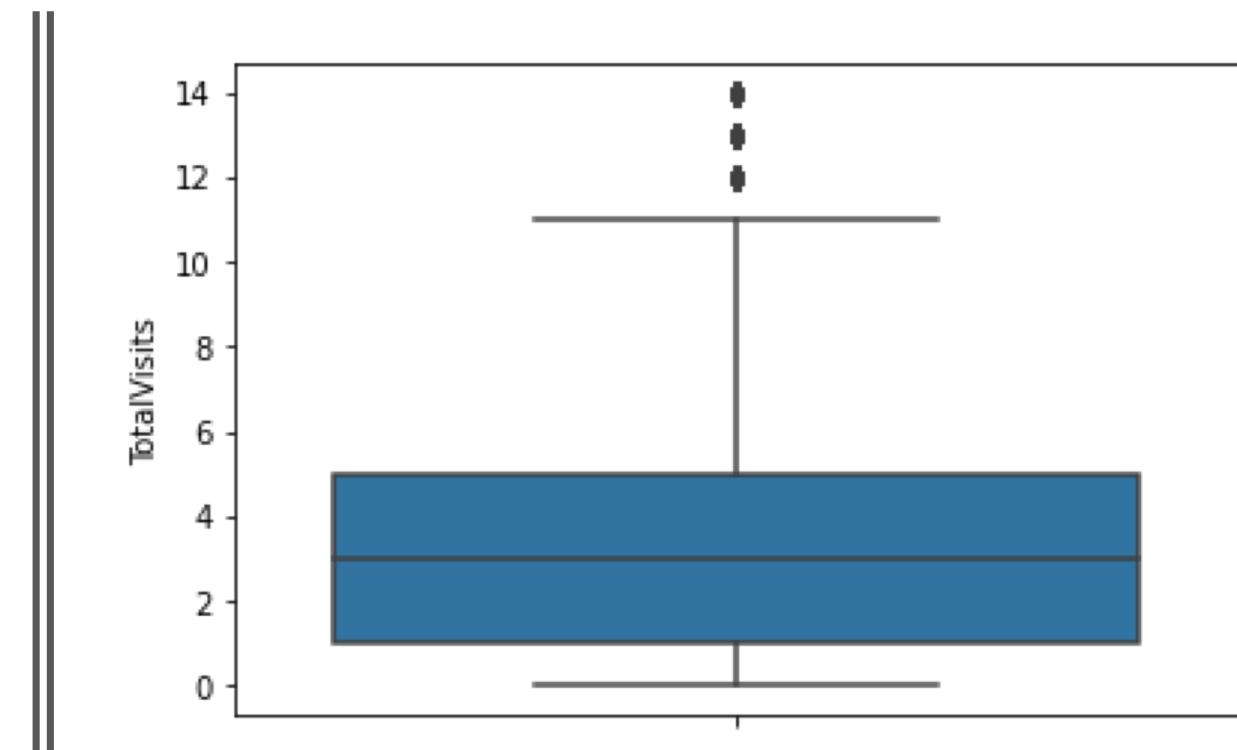
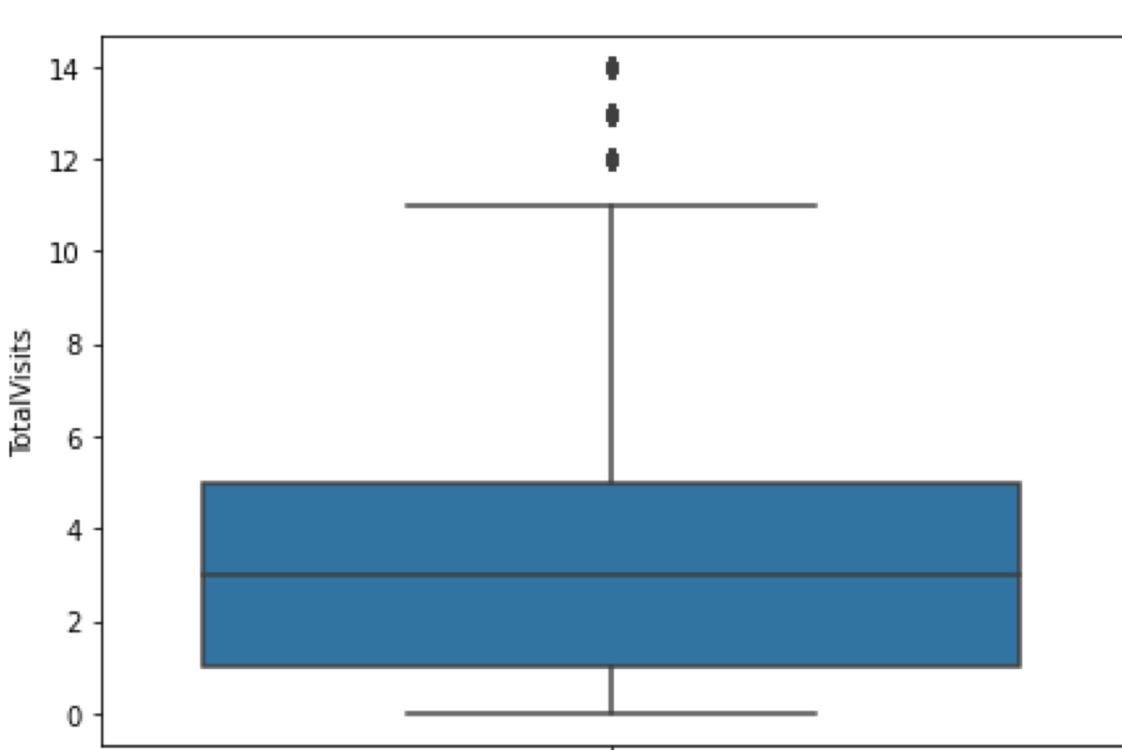


visualizing count of Variable based
on Converted value |

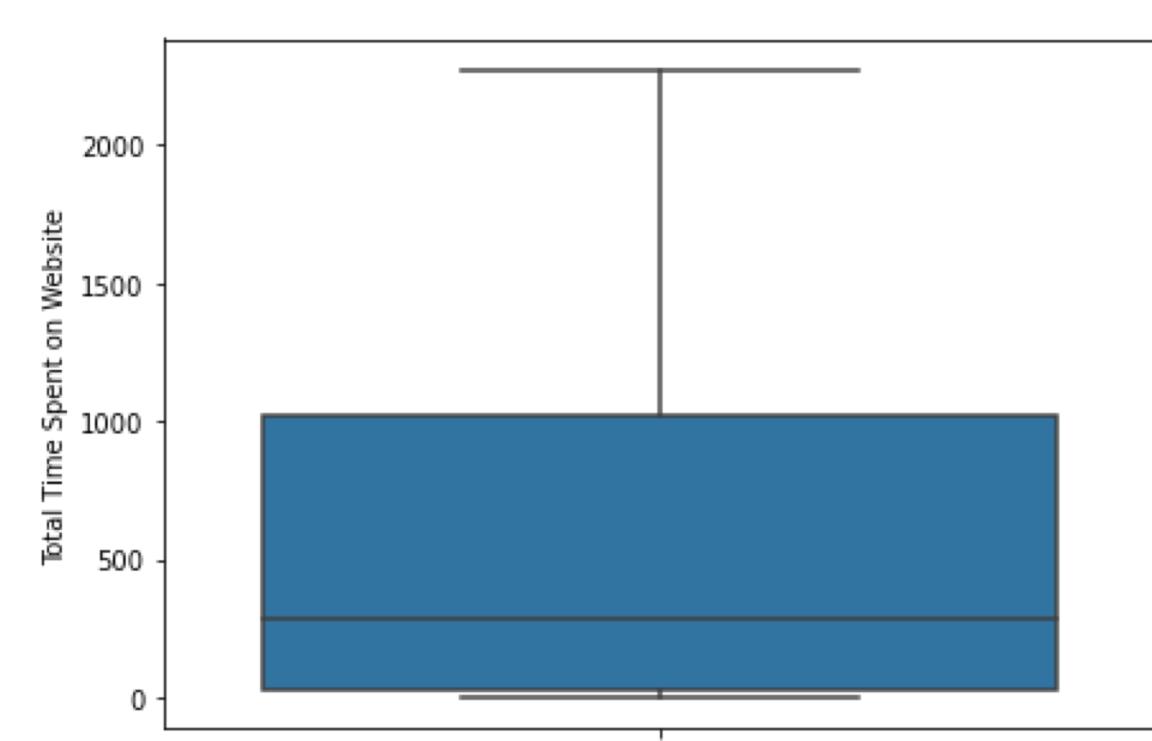
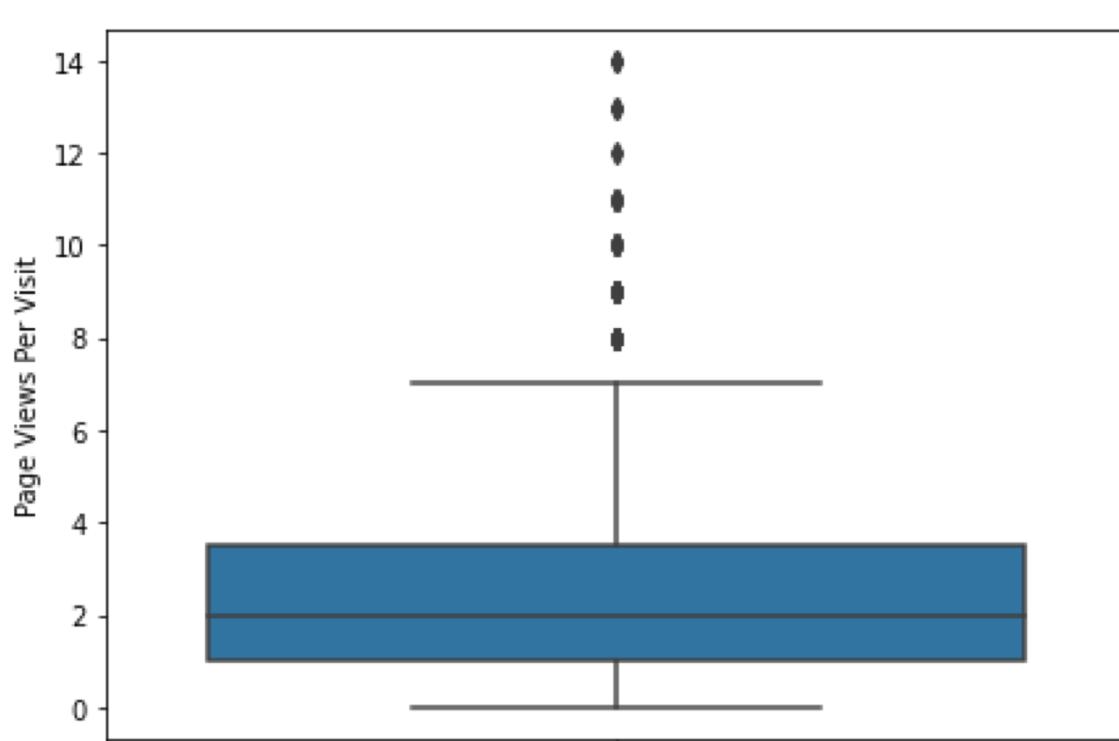
Checking
correlations
of numeric
values



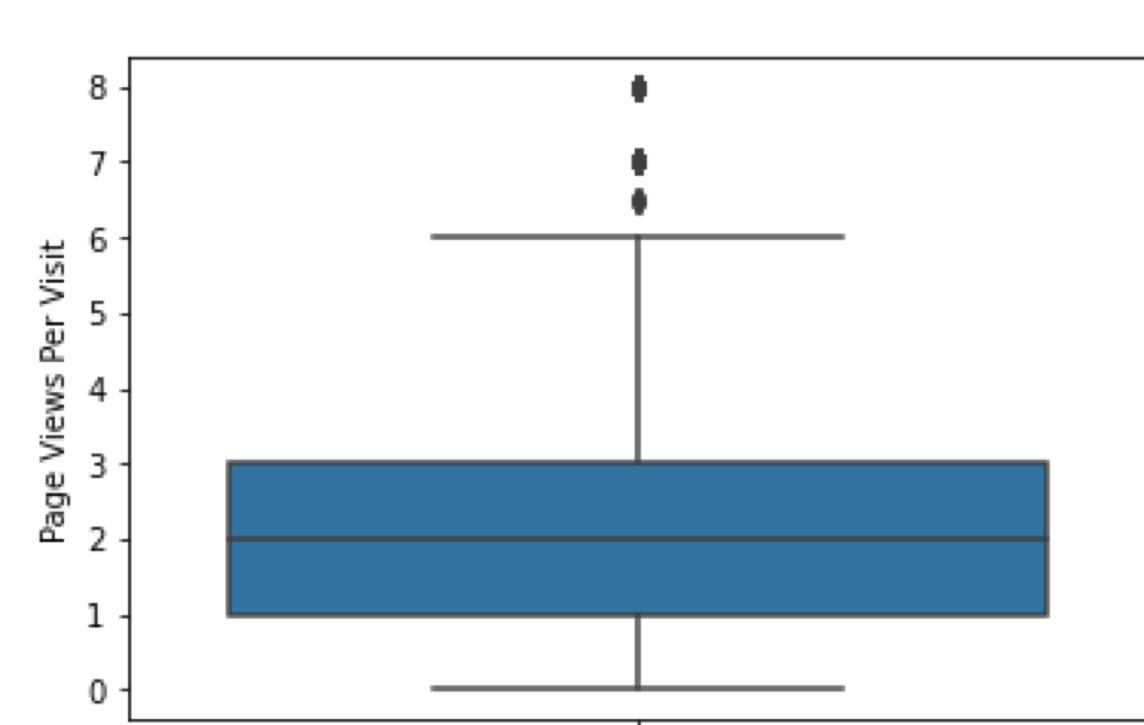
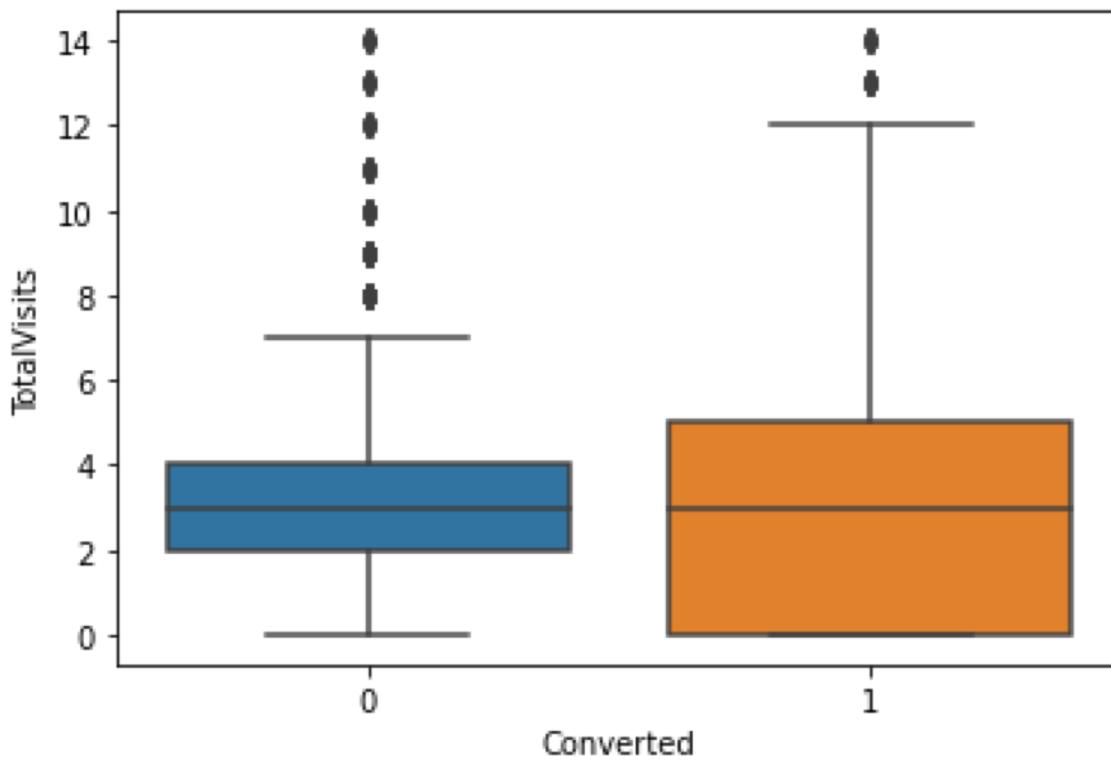
visualizing spread of variable, Remove top & bottom 1%



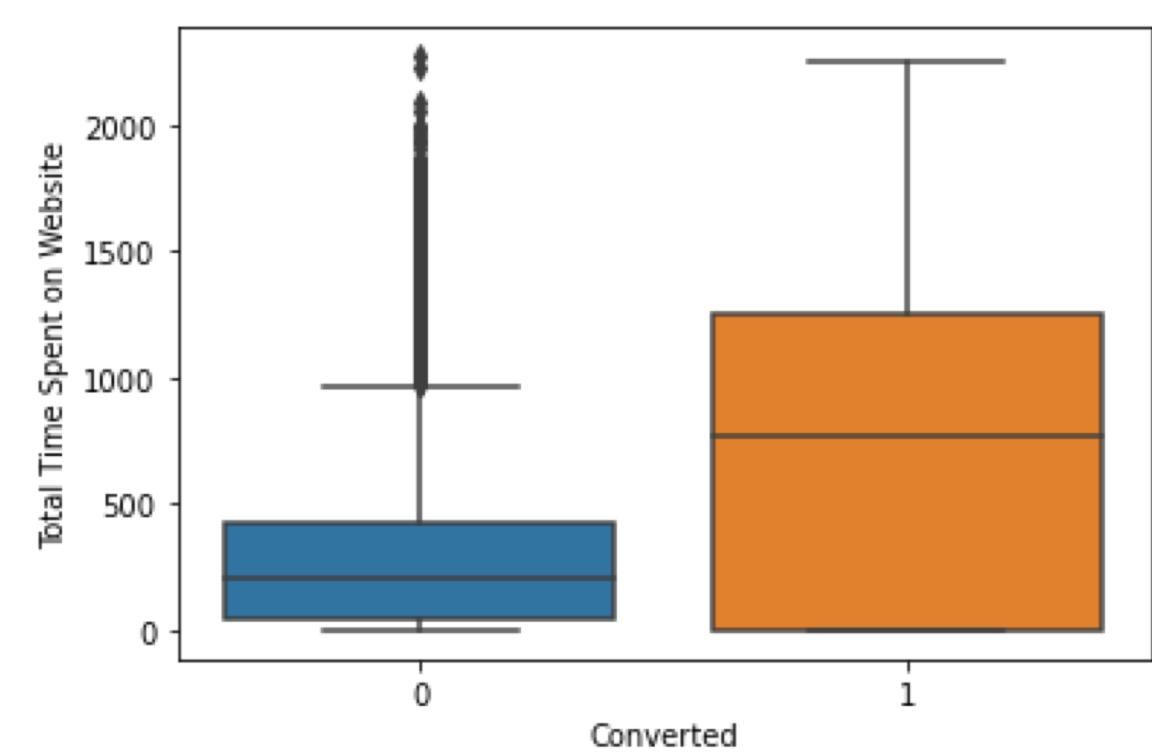
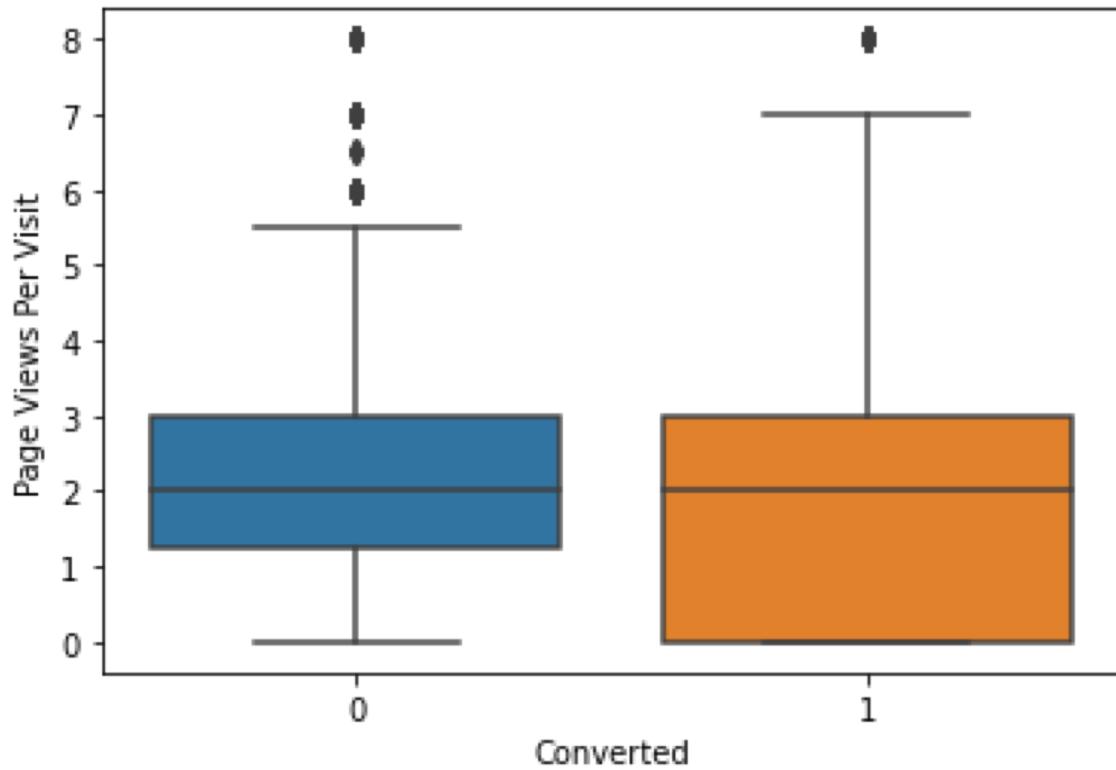
visualizing spread of numeric variable

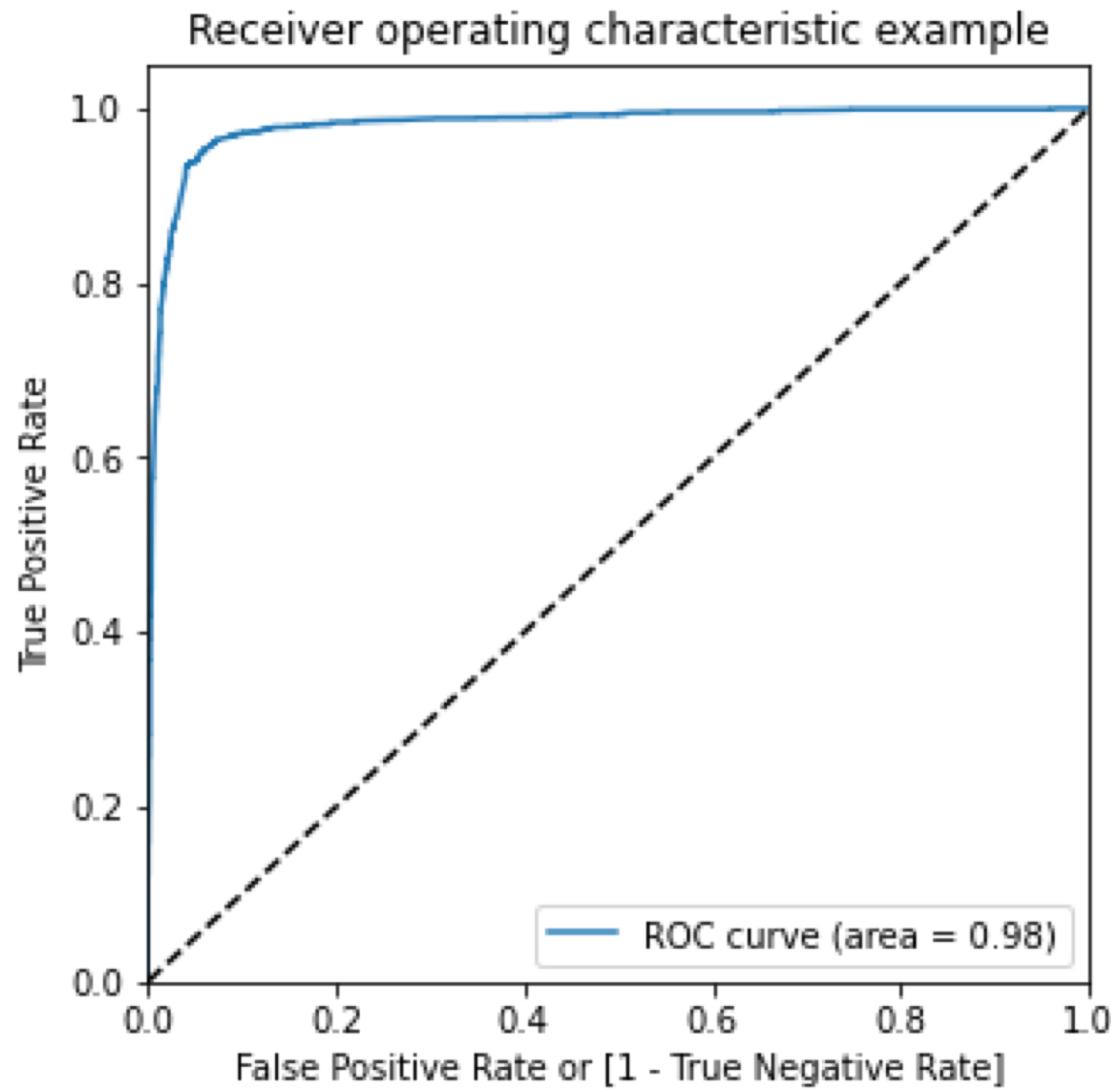
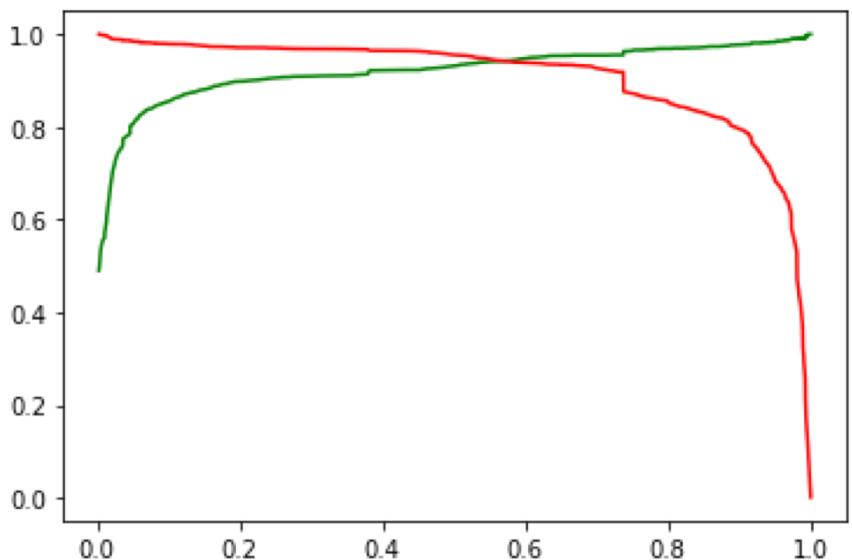
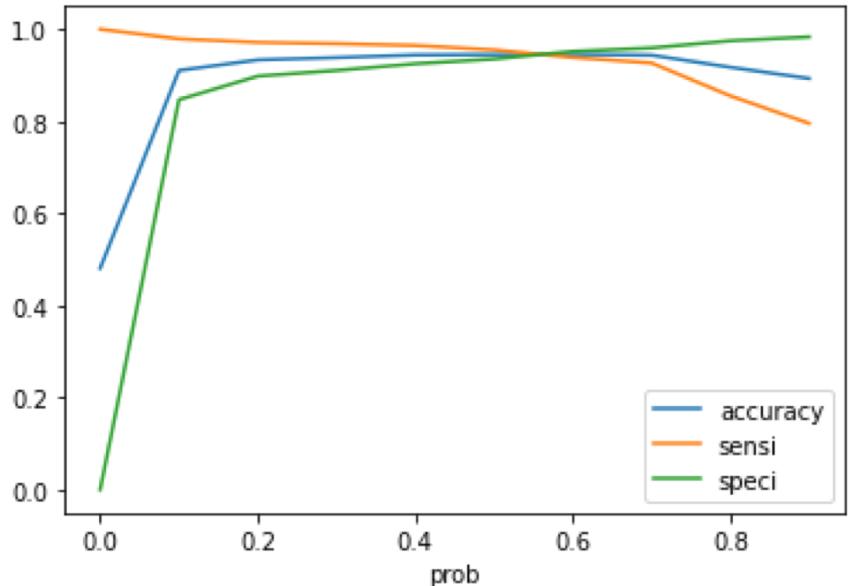


Remove top & bottom 1% , check Spread of "Total Visits" vs Converted variable



Spread of "Total Time Spent on Website" vs Converted variable





Observation:

After running the model on the Test Data these are the figures we obtain:

Accuracy : 95.53%

Sensitivity : 96.67%

Specificity : 90.60%

