

# Lead Scoring Case Study

Presented by: Devesh Parashar

# Problem Statement

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website 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. Moreover, the company also gets leads through past referrals. 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%.



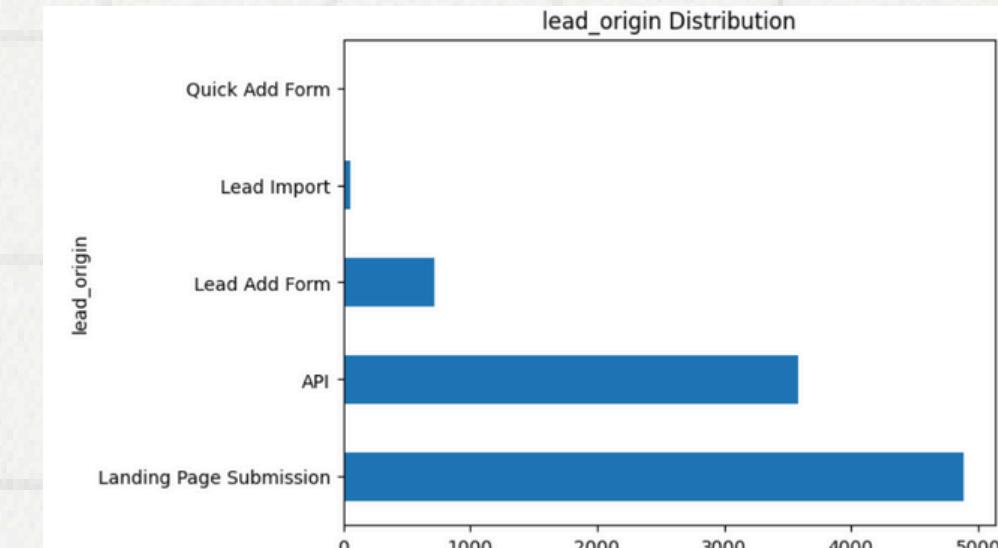
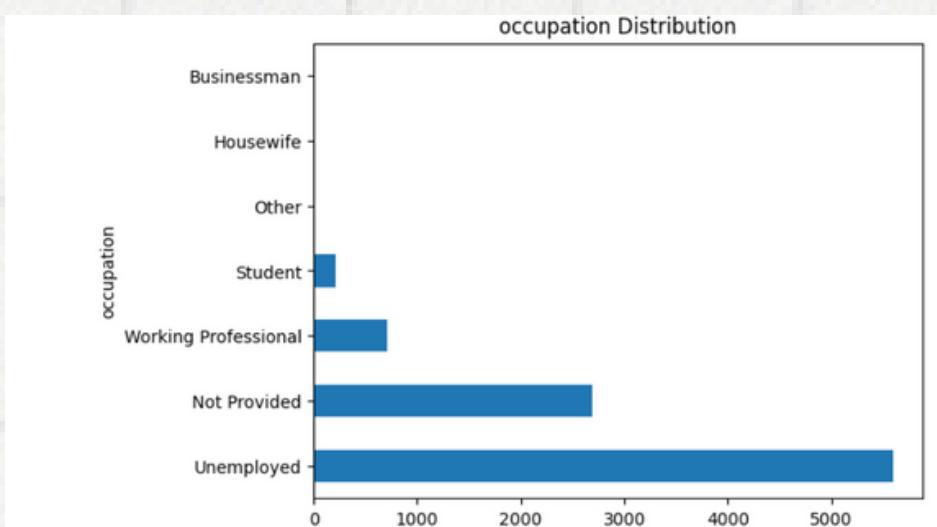
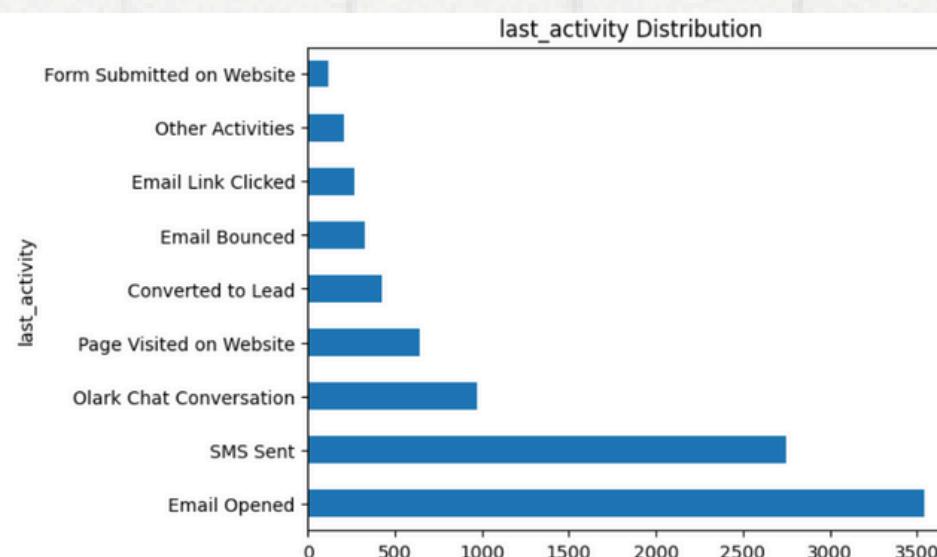
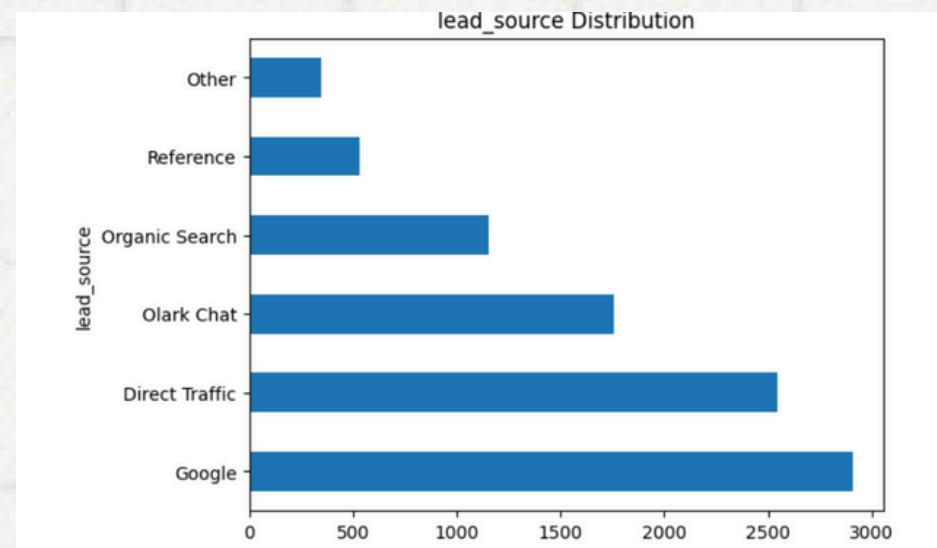
# Business Goal

- Although 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.

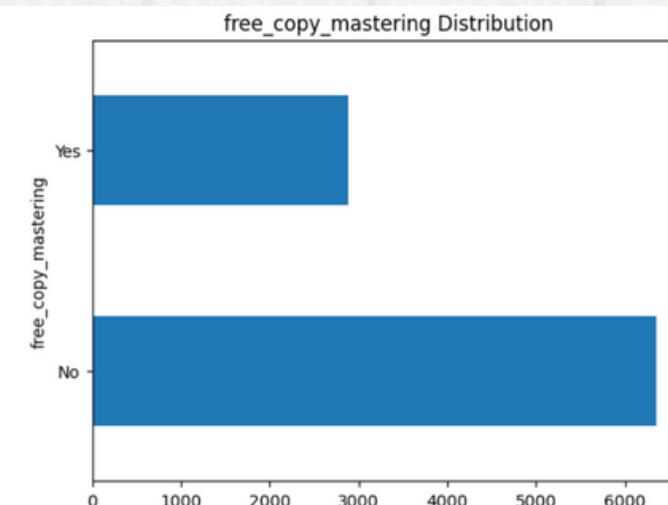


# Exploratory Data Analysis

Univariate Analysis  
(Categorical Columns)



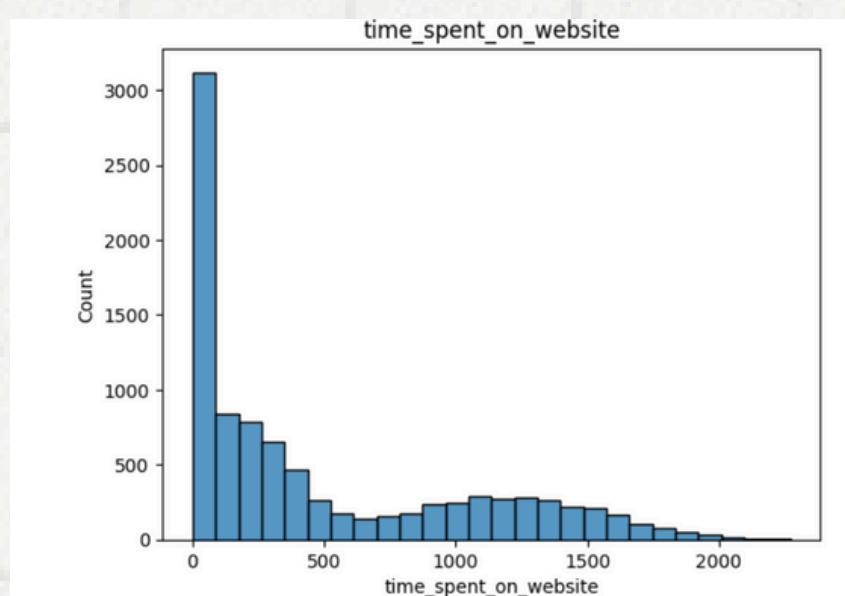
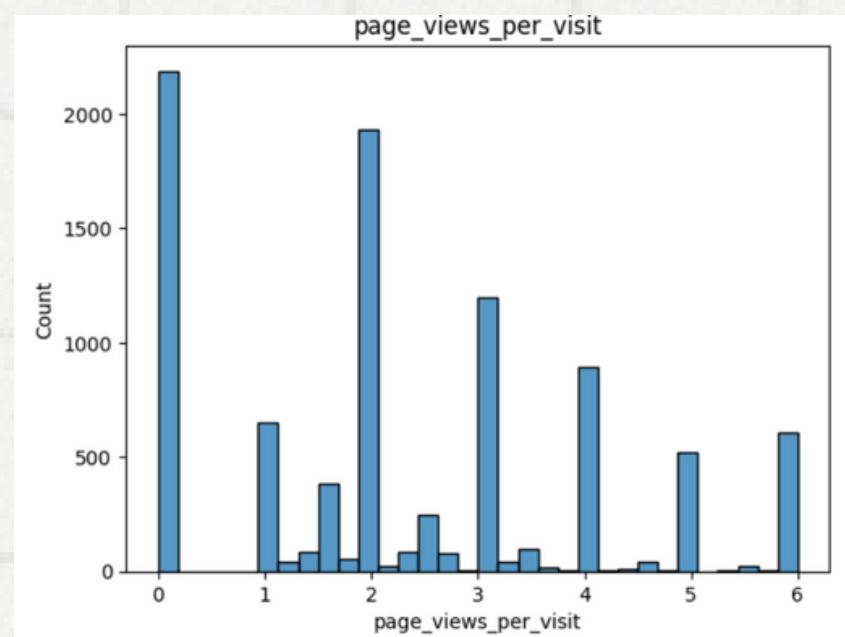
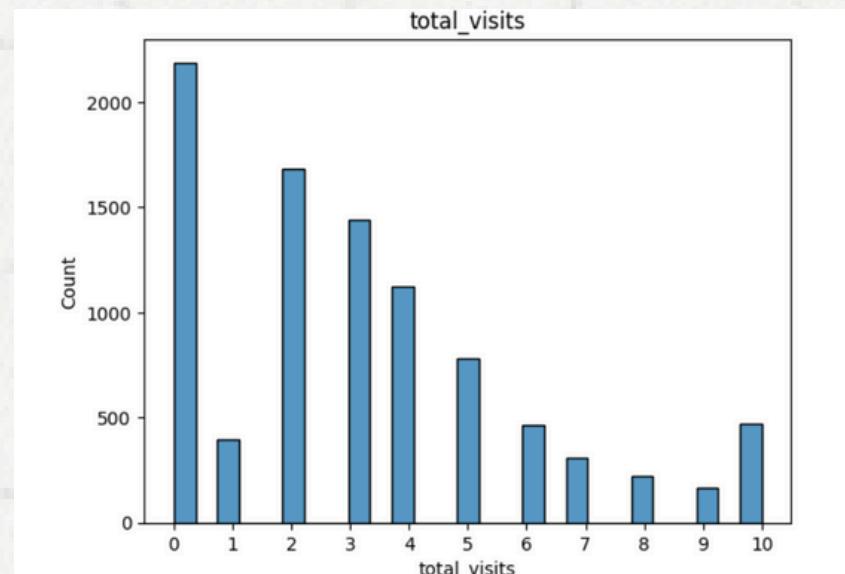
- API & Landing Page Submission are two major contributor of Lead Origin.
- Direct Traffic and Google are the two main source of Leads.
- Email Opened and SMS Sent are the major Last Activity.
- Most of the lead generated by Unemployed.
- Majority don't want a free copy of Mastering The Interview.



# Exploratory Data Analysis

## Univariate Analysis (Continuous Columns)

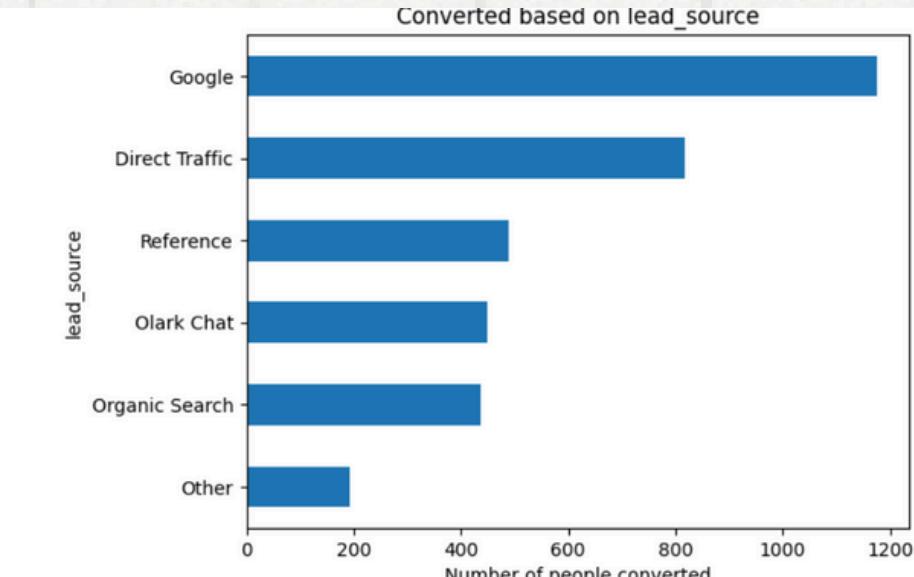
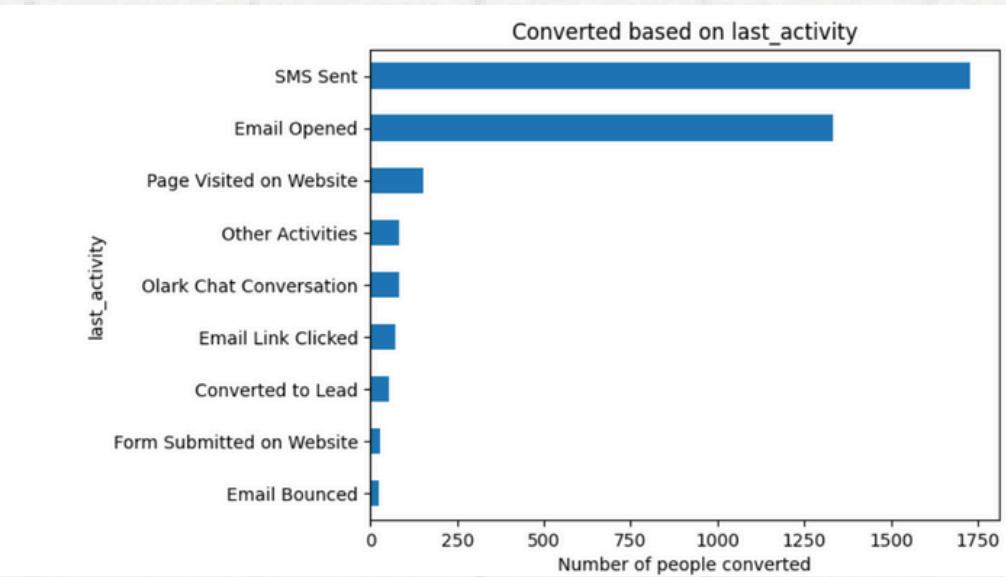
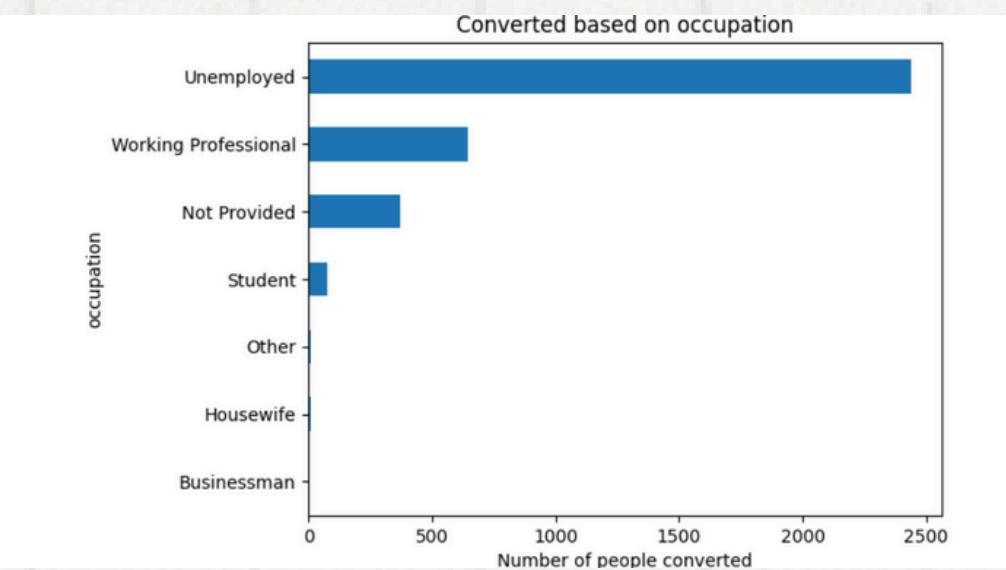
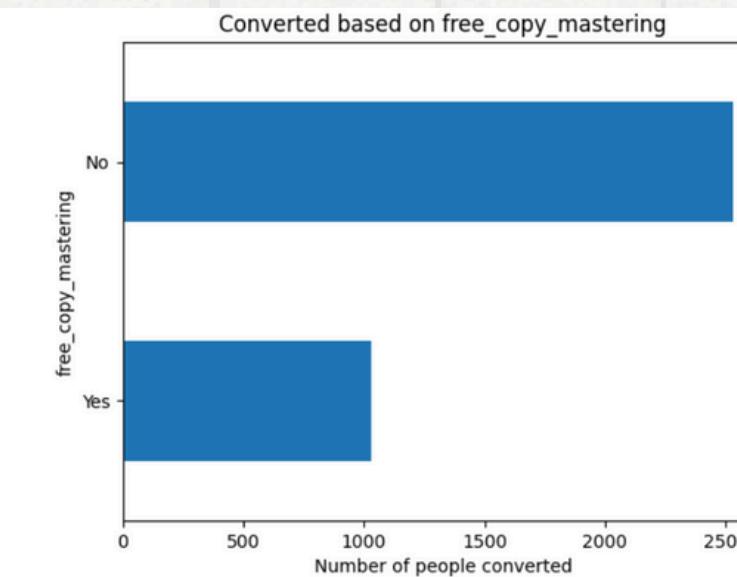
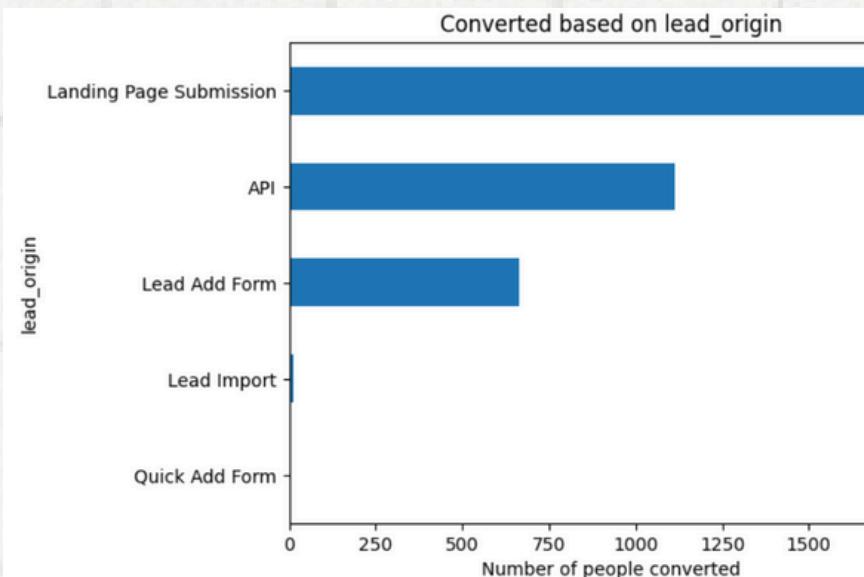
- More than 2000 people have not visited our website, around 1500 have visited twice and thrice, and this number keeps on falling with increase in visit count.
- More than 2000 people have not spent a single minute on our website. We have 600 people spending 1-60 mins on our website and almost same amount of people spending between 1-2 hrs on the website and so on.
- Once again, more than 2000 people have not viewed any page on our website, this is simply because the same number of people did not visit the website, then 1900 people have visited 2 pages, 1200 have visited 3 pages and so on.



# Exploratory Data Analysis

## Bivariate Analysis (Continuous Columns)

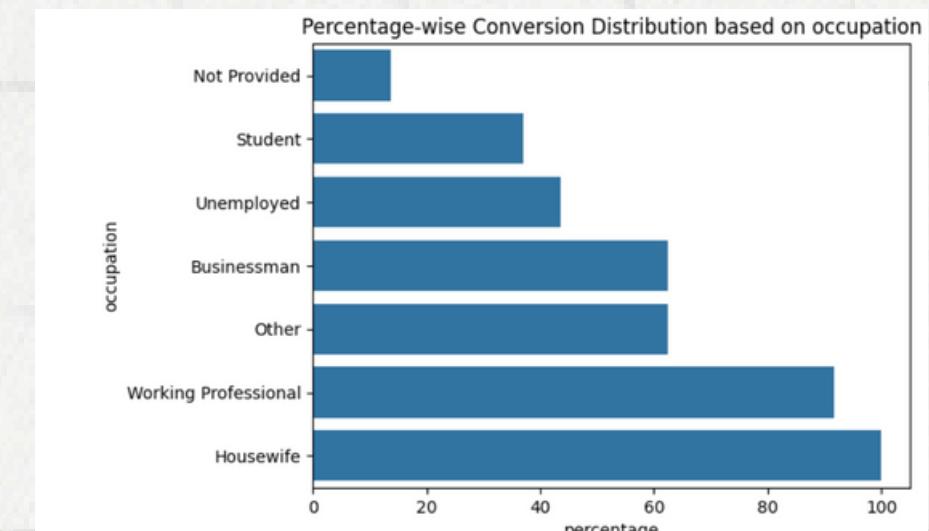
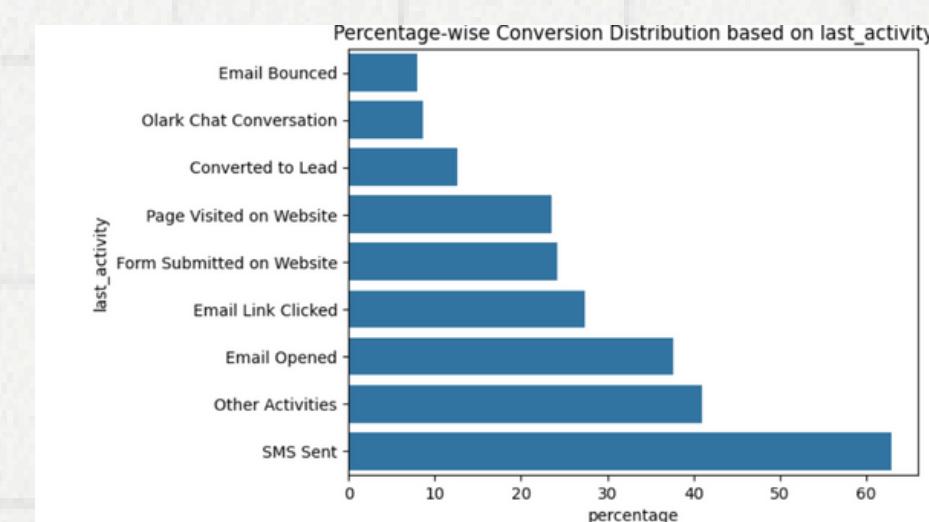
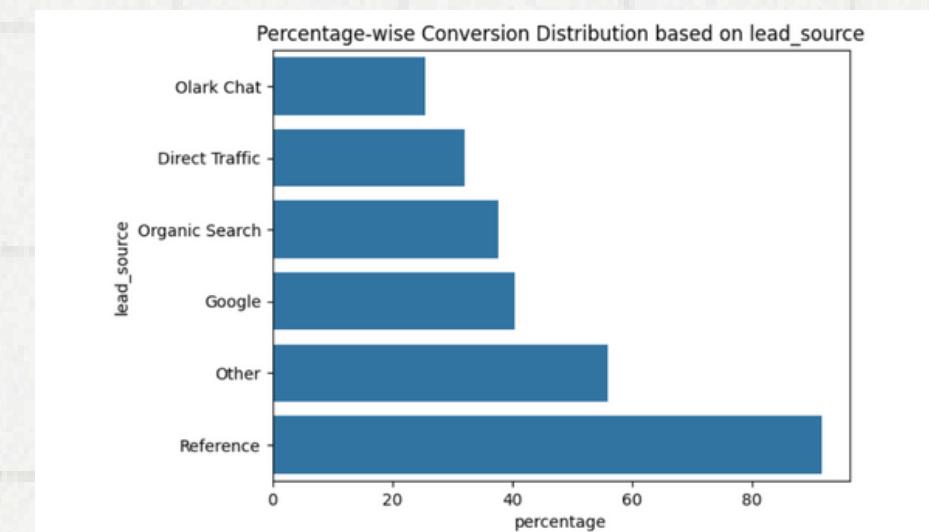
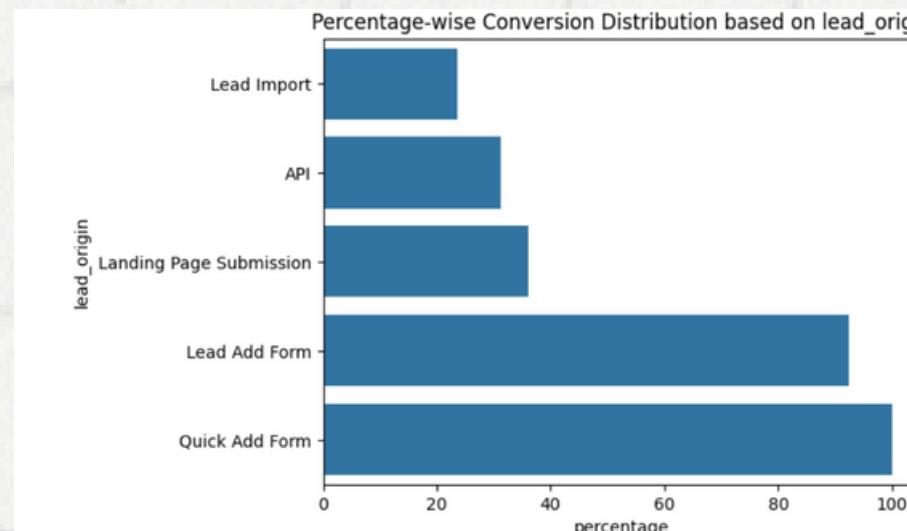
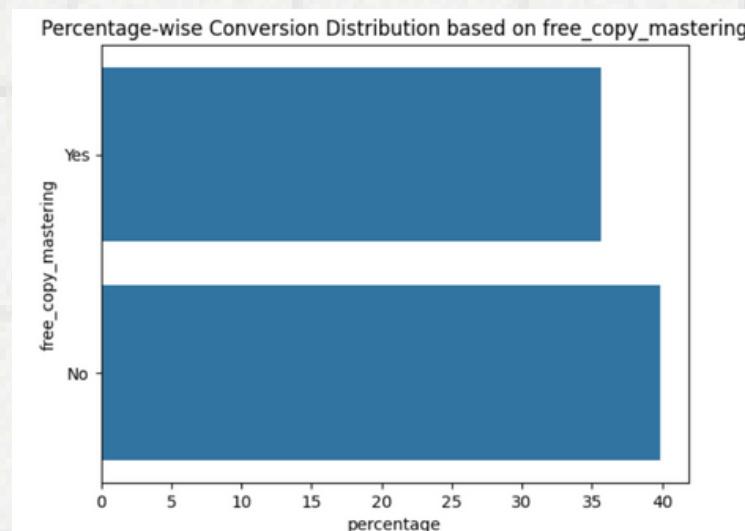
- Based on lead's origin, Landing Page Submission leads to most number of conversion i.e. more than 1750.
- Based on lead's source, people coming from Google gets converted the most i.e. ~1200.
- People with SMS sent as their last activity got most converted i.e. ~1750 followed by Email Opened.
- Most number of Unemployed people got converted i.e. ~2500
- Most converted people (~2500) did not do Free copy mastering.



# Exploratory Data Analysis

## (Percentage wise Calculations)

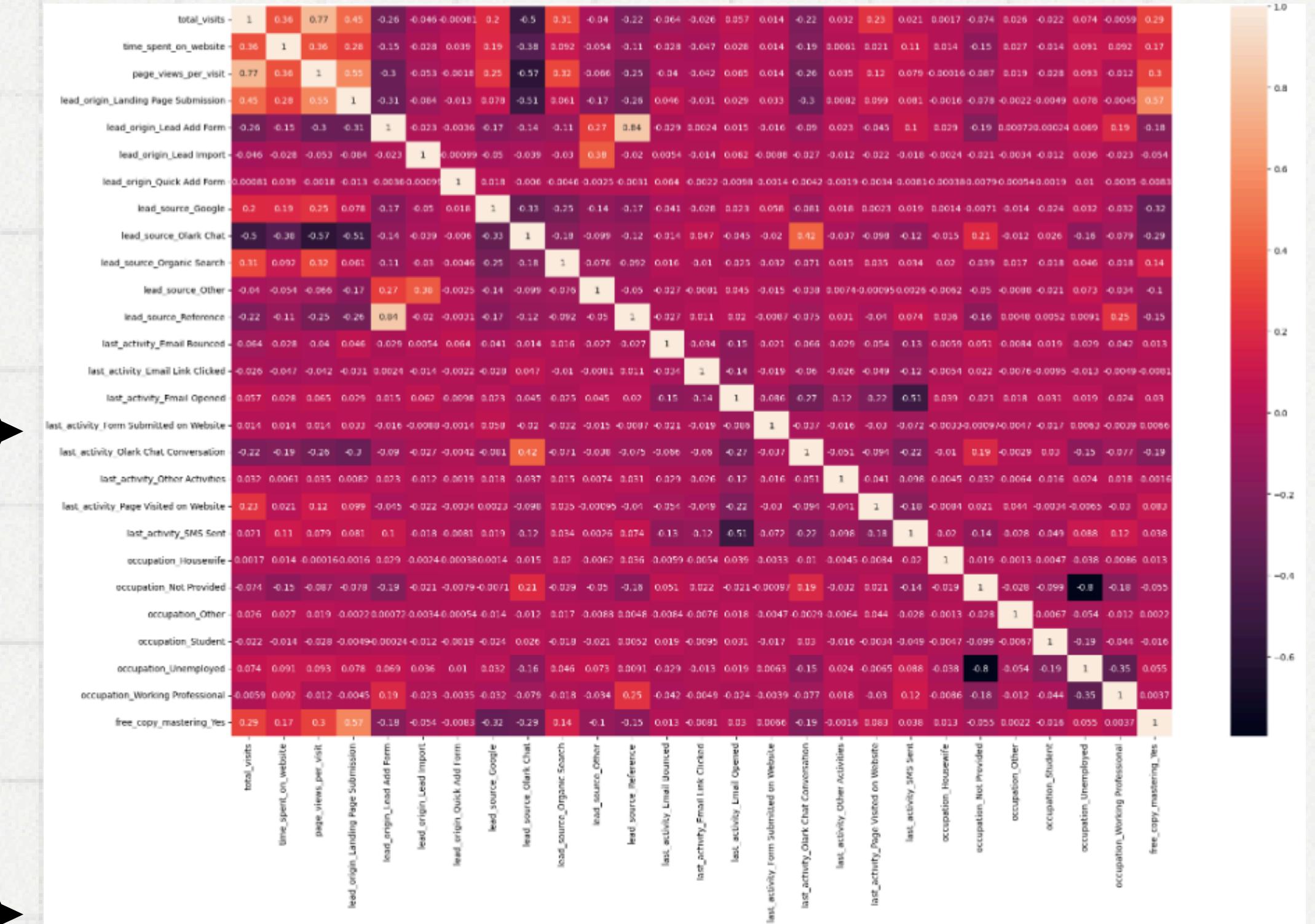
- Based on lead's origin, all leads from Quick add form got converted showing a 100% conversion rate followed by a 92% conversion rate from lead add form.
- Based on lead's source, a 92% conversion rate is seen if the lead is a reference.
- Based on lead's last activity, SMS sent has 62% conversion rate. SMS sent is the last activity showing high conversion based on both absolute and percentage-wise terms.
- All Housewives got converted showing a 100% conversion rate followed by working professionals that also show a very high ~92% conversion rate.
- For people with/without free copy mastering, conversion rate is same only i.e. around 35%.



# Exploratory Data Analysis

## (Correlations)

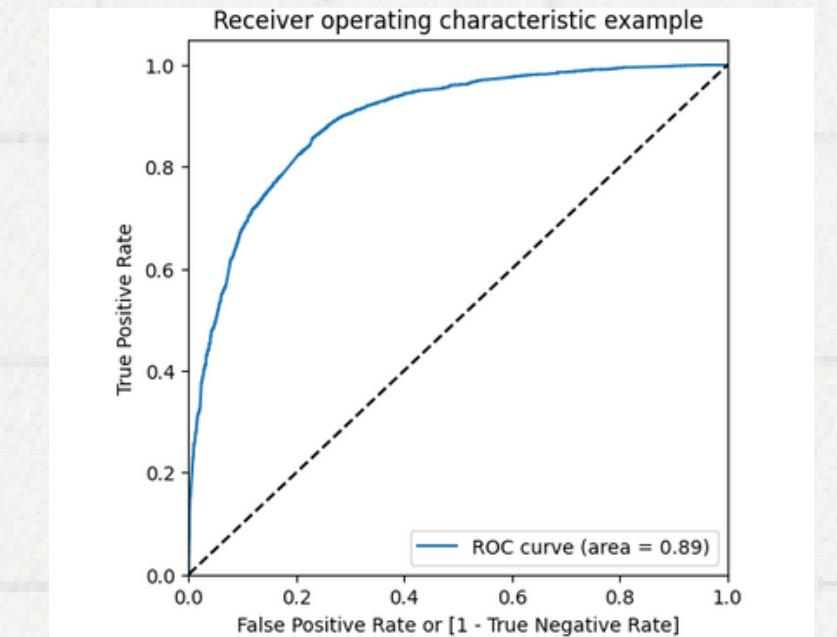
- There is high positive correlation between lead\_source\_Reference and lead\_origin\_Lead Add Forum. Also, between total\_visits and page\_views\_per\_visit which we also found previously.
- There is a strong negative correlation between occupation\_Unemployed and occupation\_Not Provided.



# Model Evaluation

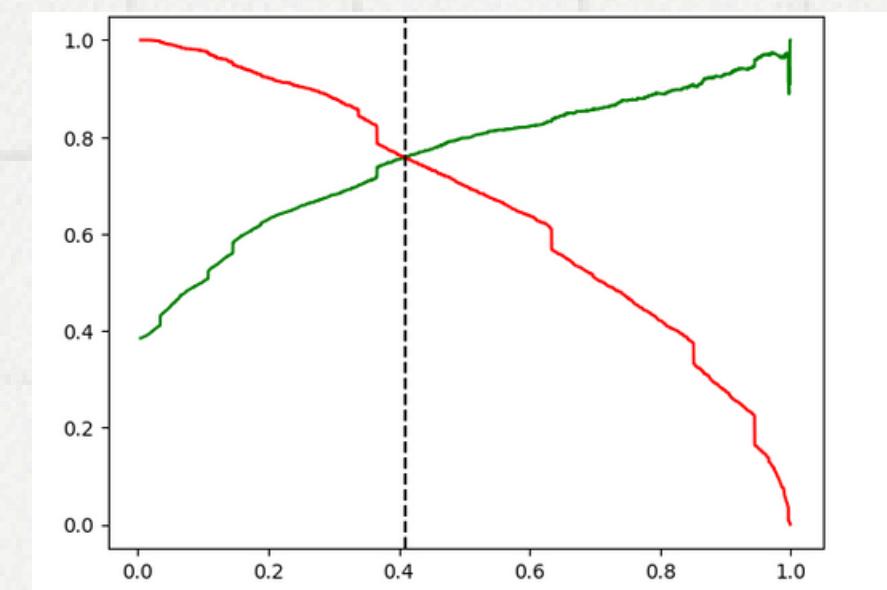
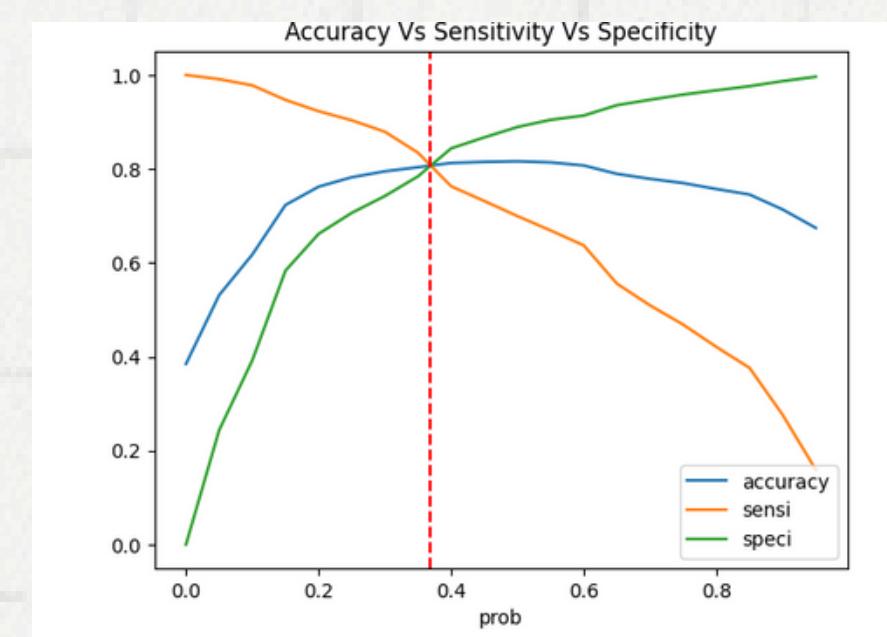
## Plotting ROC Curve

- The area under curve (auc) is 0.89 which is a very good value.
- It means our model is showing good performance and able to distinguish between positive and negative classes.
- Optimal Cut-off point : 0.369



## Precision and Recall Trade-off

- Based on tradoff between Precision and Recall, we are getting 0.41 as our optimal cutoff.
- This is 0.4 higher than previous cutoff i.e. 0.369.
- Final Cut-Off
- We know that recall is nothing but Sensitivity or true-positive rate i.e. percentage of true predicted out of actual true.
- We need to focus on correctly predicting actual true, i.e. high recall/sensitivity.
- Hence, we will take 0.369 as our final cutoff, since its giving 75.9% of sensitivity and 80.9% of overall accuracy.



# Summary

- It was found that the variables that mattered the most in the potential buyers are:
- The total time spend on the Website.
- Total number of visits.
- Page Views per visit
- When the lead Origin was:
  - Lead Add Form
- When the lead source was:
  - Olark Chat
- When the last activity was:
  - Email Bounced
  - Email Opened
  - SMS sent
  - Olark chat conversation
- When their current occupation is a "working professional".
- Keeping these in mind the X Education can flourish as they have a very high chance to convert almost all the potential leads. Leads with score of more than 36 can be targeted by the X Education. Anyone having a score of  $\geq 36$  can be considered a **Hot Lead**.
- **We have:**
  - 87% conversion rate if the score of the lead is **more than equal to 36**.
  - 98% conversion rate if the score of the lead is **more than equal to 40**.
  - 100% conversion rate if the score of the lead is **more than equal to 42**.

**Thank you  
very much!**