

# LEAD SCORING CASE STUDY FOR X - EDUCATION COMPANY

Anand Singh Bhandari

Anish Gautam

Kushagra Rao

# CONTENTS

- Introduction
- Problem Statement
- Assumptions
- Areas Of Focus – Graphs and Insights
- Recommendations

# INTRODUCTION

---

- An education company named X Education sells online courses to industry professionals. The company markets its courses on their website, several other websites and search engines like Google.
- When 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.
- Although X Education gets a lot of leads, its lead conversion rate is very poor around 30%.

# PROBLEM STATEMENT

---

- To make conversion rate more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If the company 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.
- Need to find out most promising 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 a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%.

# ASSUMPTIONS

---

## Assumptions based on Linear regression model:

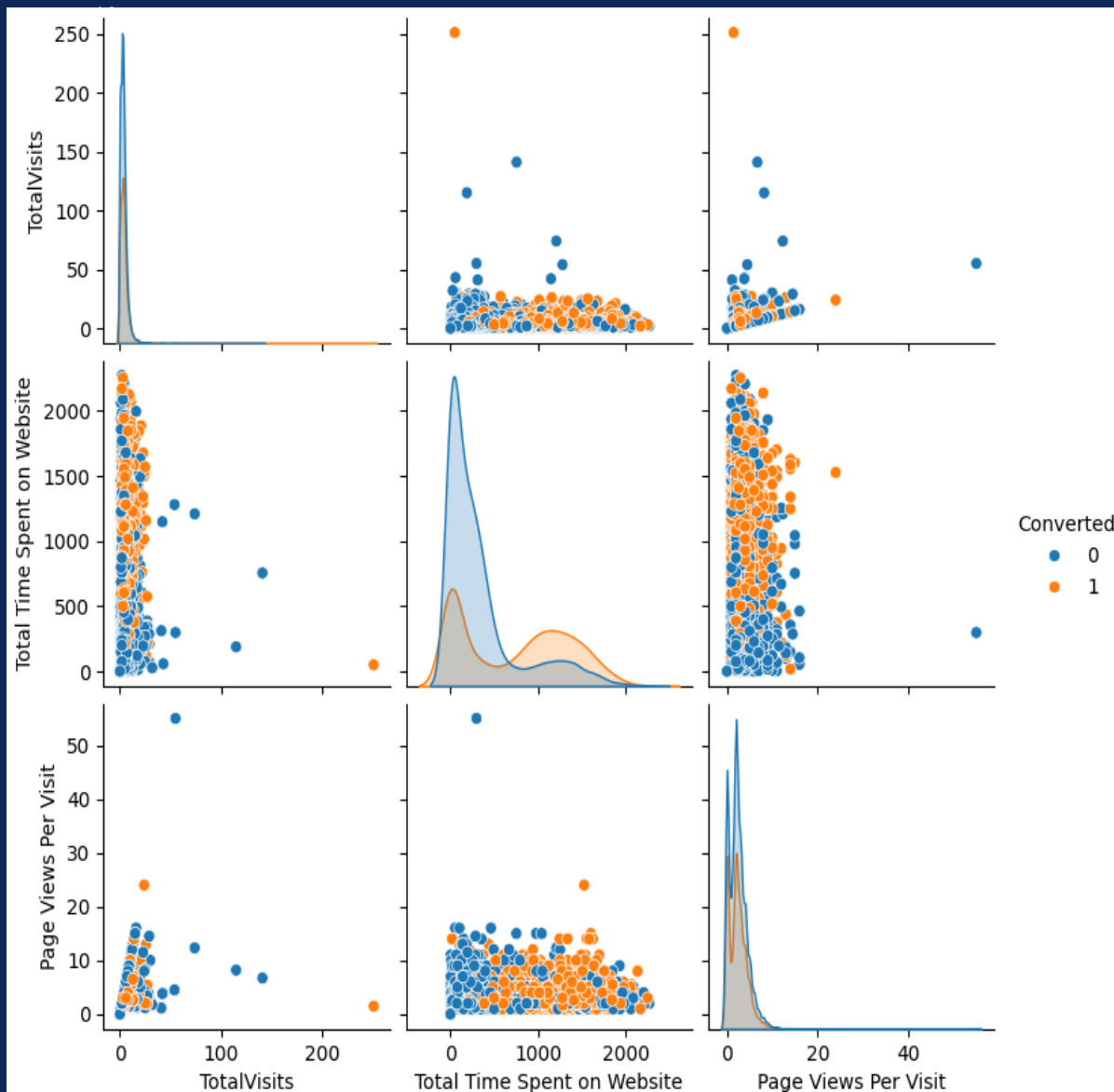
- **Linearity:** We assume that there is a linear relationship between the independent variables (such as TotalVisits, Total Time Spent on Website, Page Views Per Visit) and the dependent variable (Converted). This can be examined by plotting scatterplots of each independent variable against the dependent variable.
- **Homoscedasticity:** We assume that the variance of the errors is constant across all levels of the independent variables. This means that the spread of the residuals should be roughly the same for different values of the independent variables.
- **Normality of errors:** We assume that the errors follow a normal distribution. This assumption is important for hypothesis testing and confidence interval estimation. We can check this assumption by examining a histogram or Q-Q plot of the residuals.
- **No multicollinearity:** We assume that there is no high correlation between independent variables. Multicollinearity can inflate the standard errors of the coefficients and make the estimates unstable. We can assess multicollinearity by calculating the correlation matrix or variance inflation factors (VIF) for the independent variables.
- **Correct specification of the model:** We assume that the model includes all relevant independent variables and excludes irrelevant ones. This assumption ensures that the model adequately captures the relationship between the independent and dependent variables.



# GRAPHS & INSIGHTS

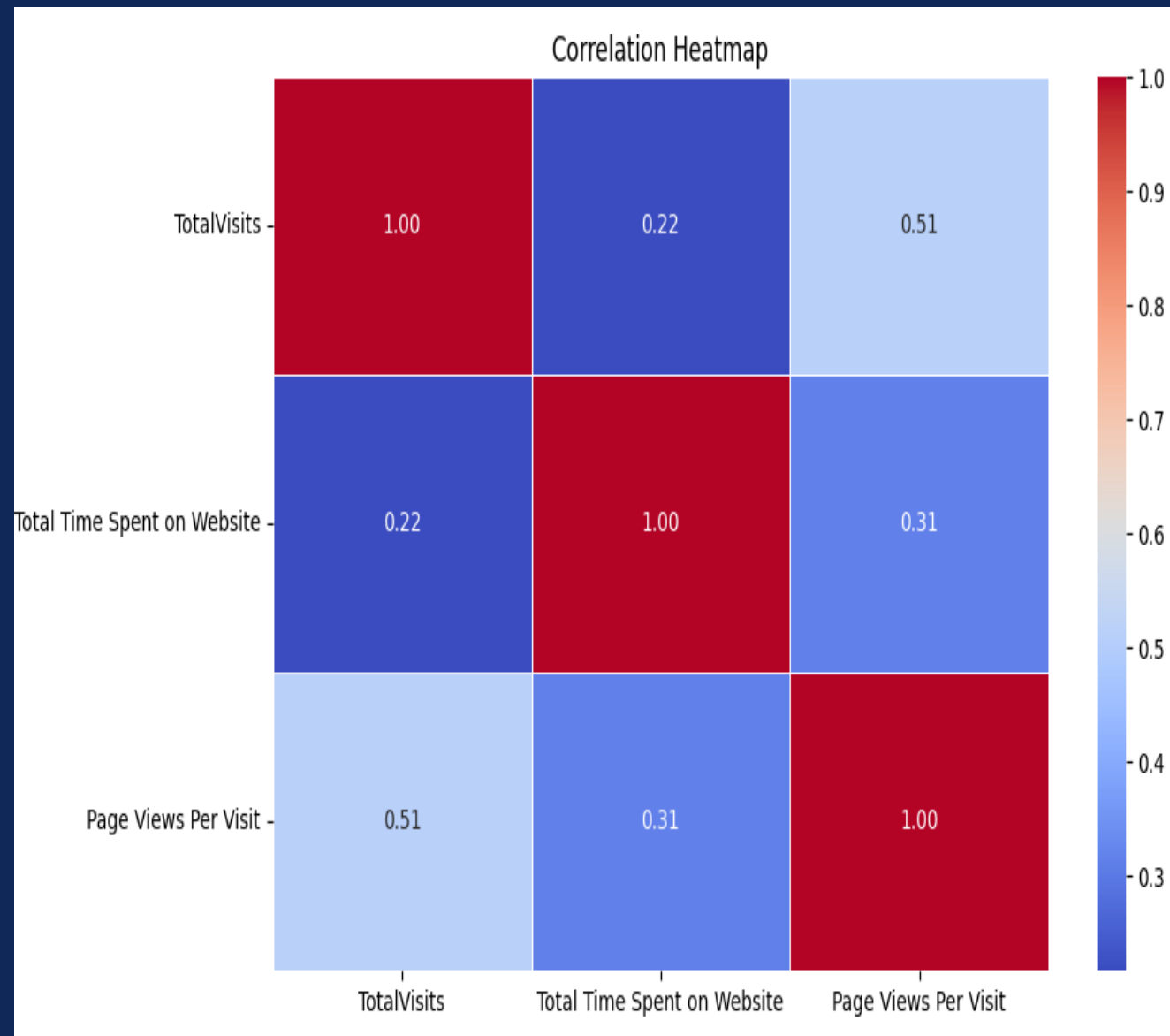
---

# NUMERICAL - PAIR PLOT & CORRELATION



## Observations:

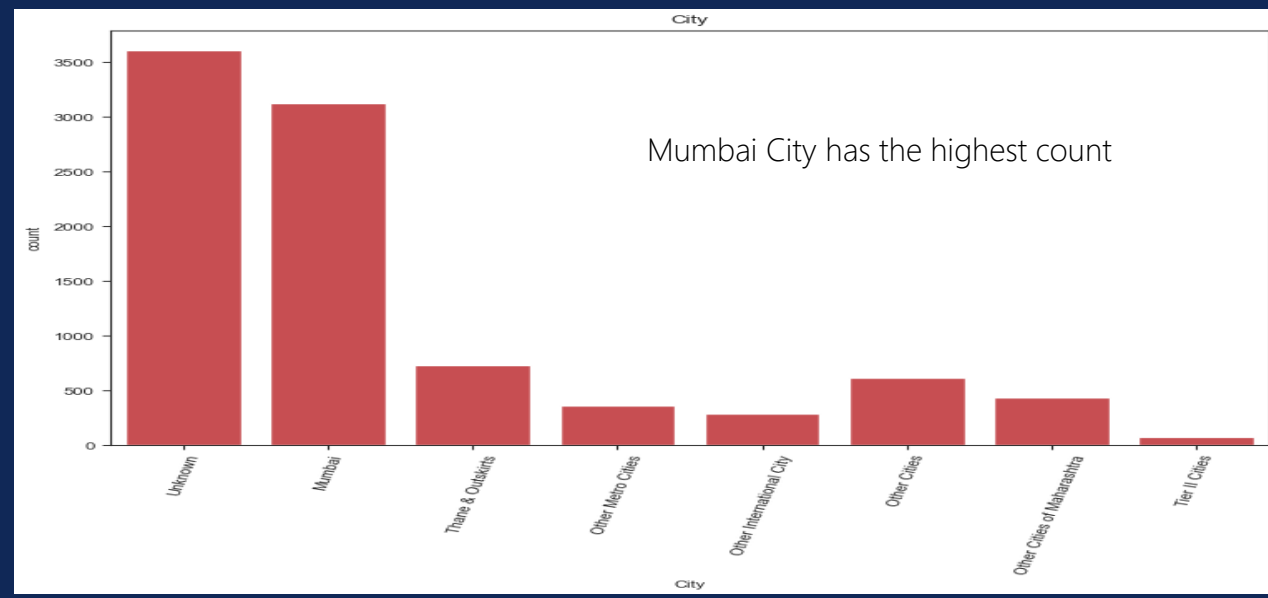
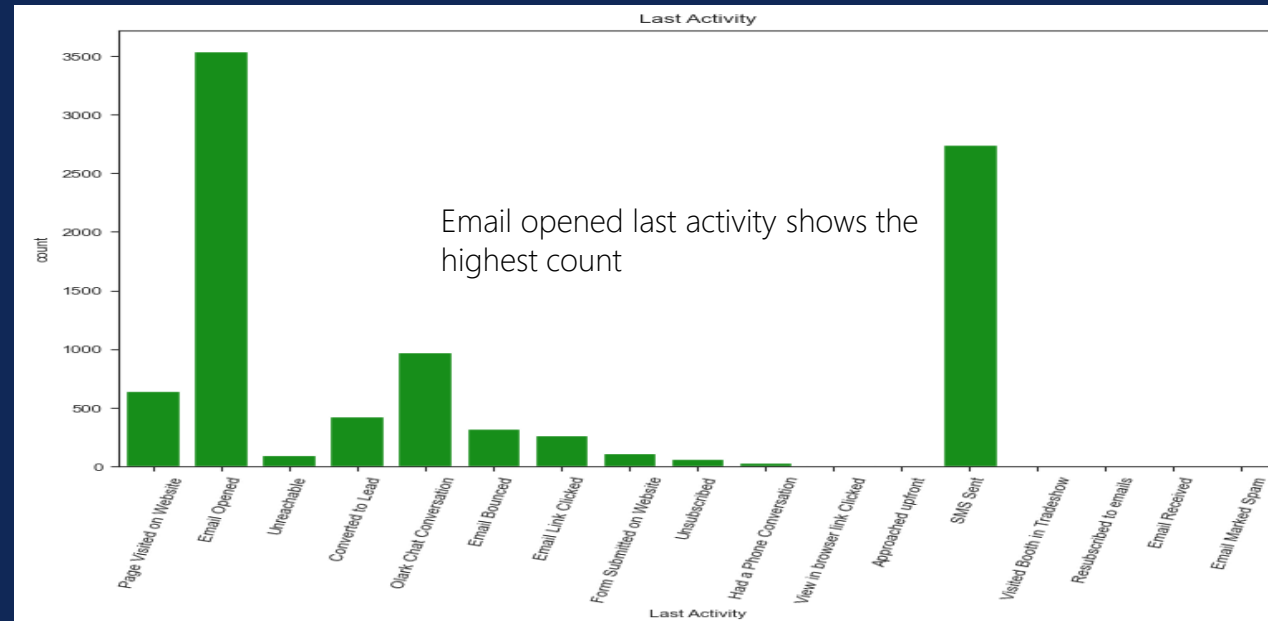
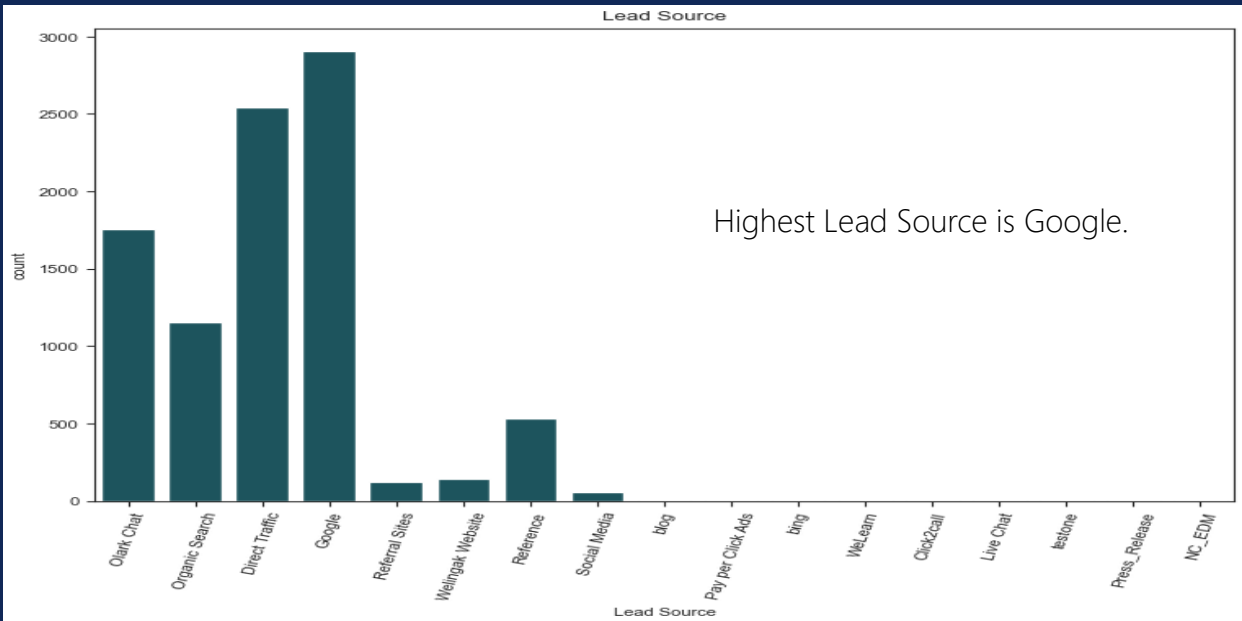
- Converted leads tend to visit less often but spend more time per visit
- Converted leads tend to visit less often but view more pages per visit



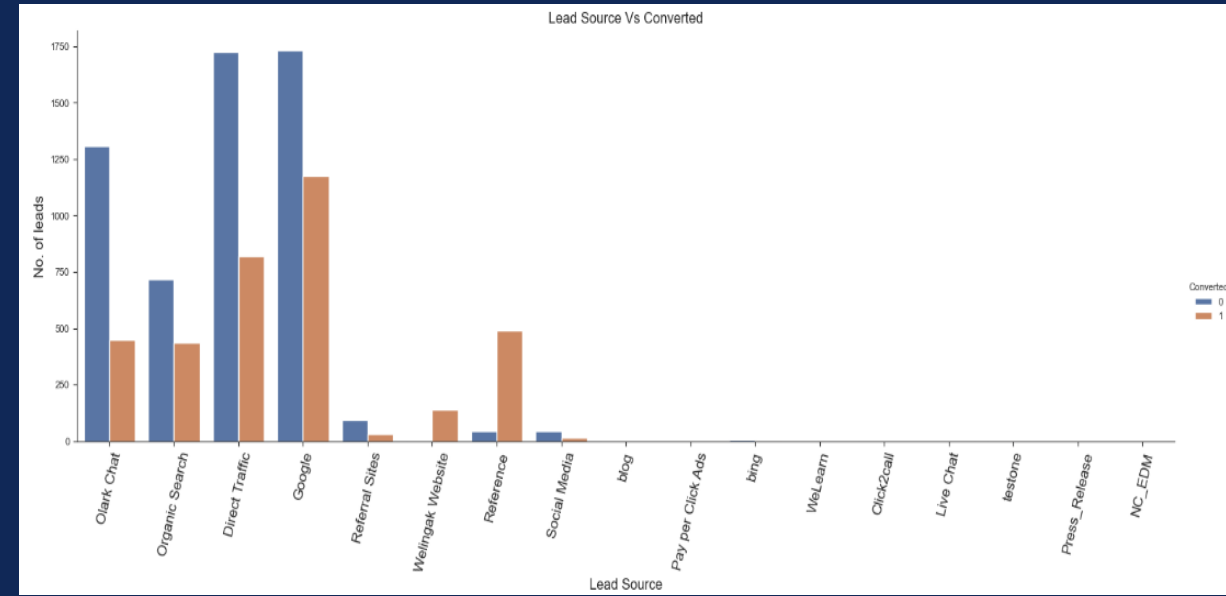
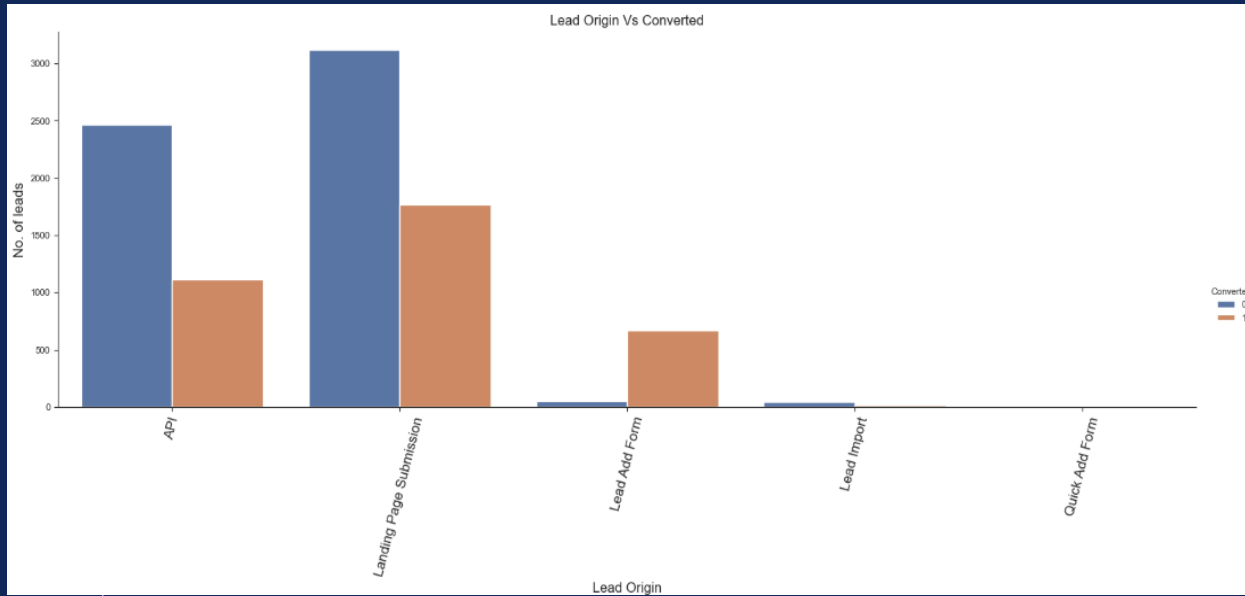
## Observations:

- There is not much correlation between numerical variables

# UNIVARIATE- CATEGORICAL

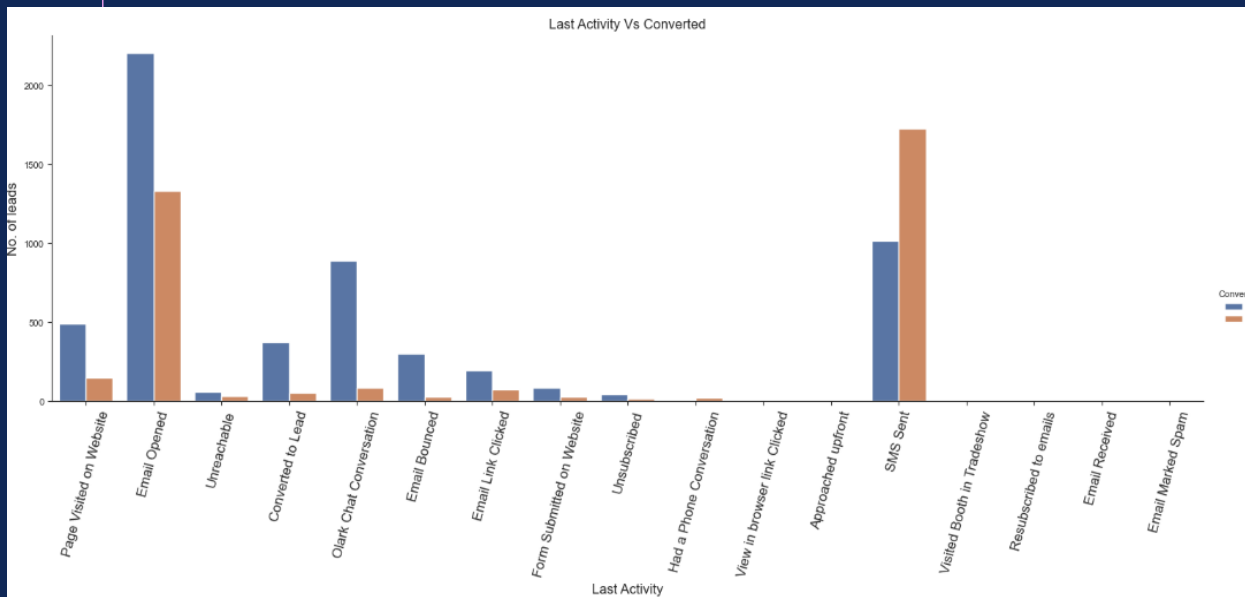


# CATEGORY WISE CONVERSATION RATE

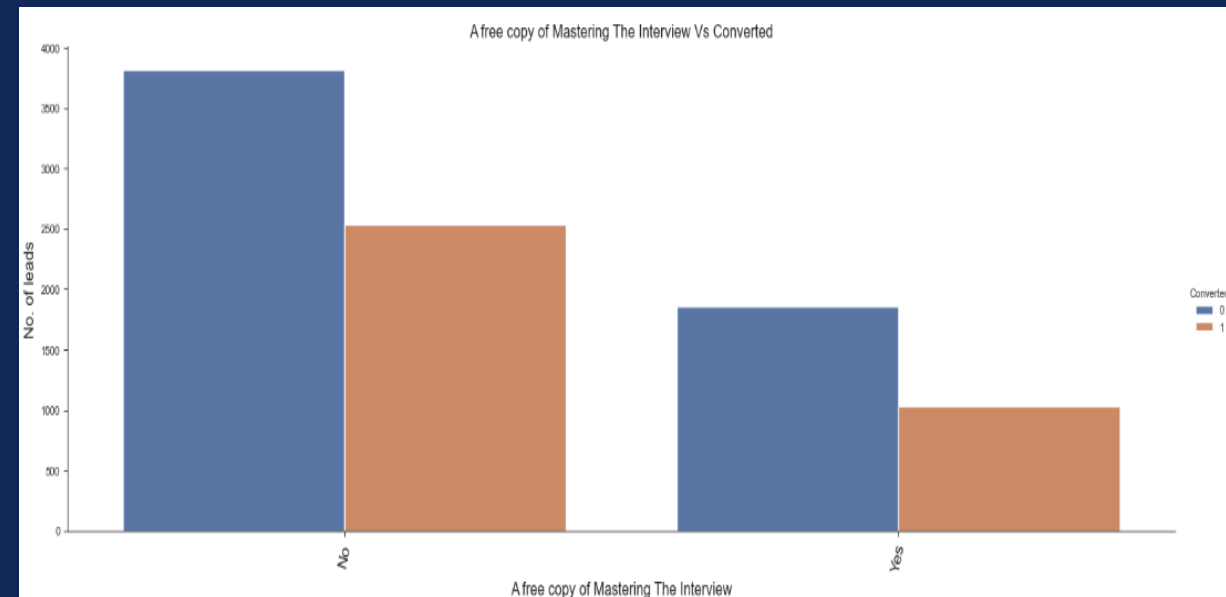


Landing Page Submission contributes most to both the categories.

Most converted leads come from Google

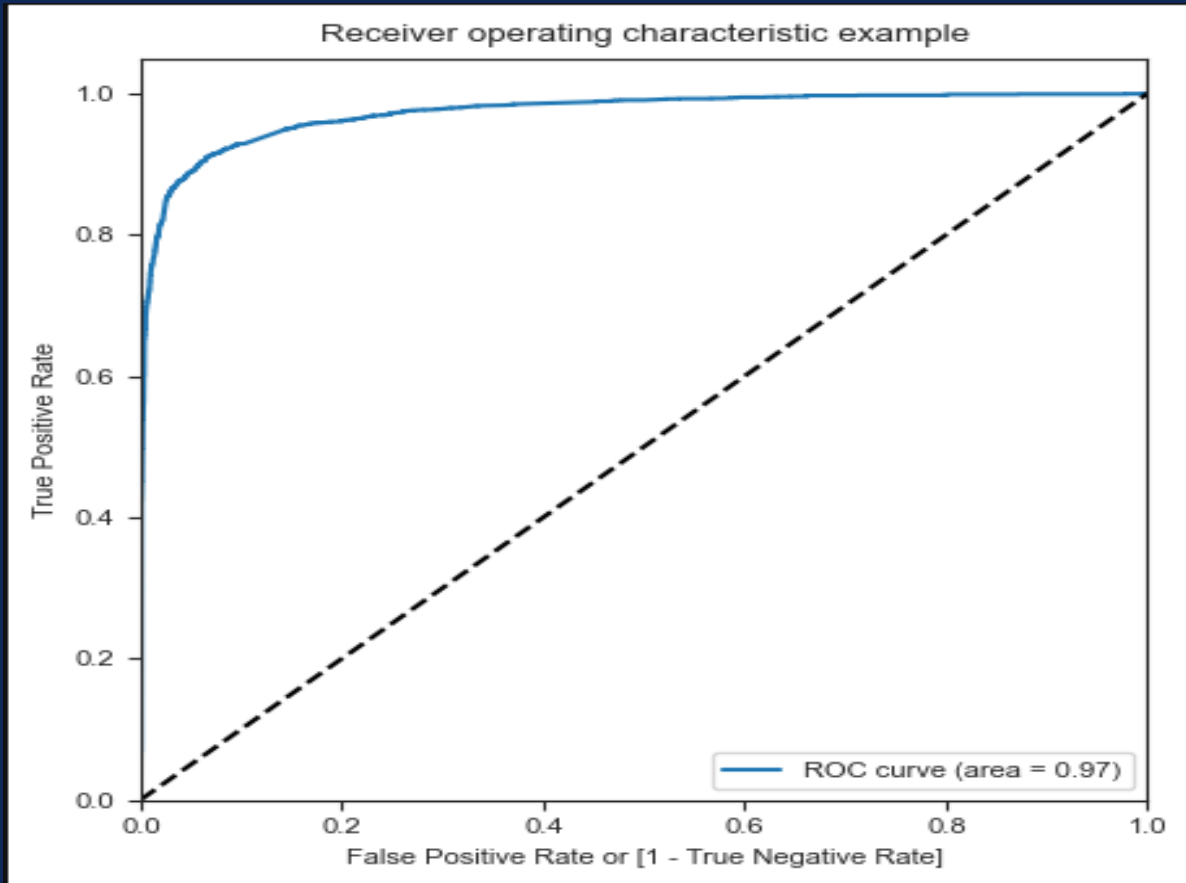


Most converted leads in Last Activity was 'SMS Sent'



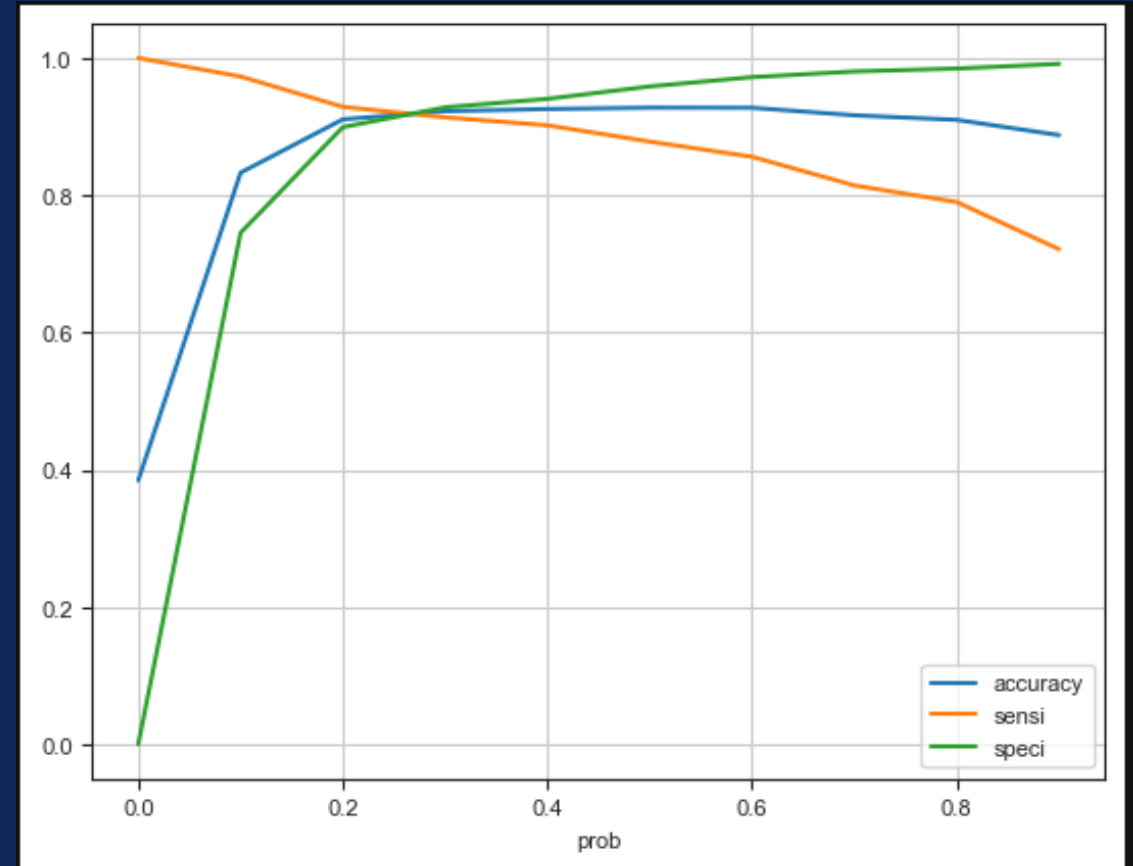
People not taking the free copy of mastering the interview are more likely to convert than those who've taken the free book.

# ROC CURVE



ROC curve has area of 0.97. Curve seems to be accurate showing trade off between sensitivity and specificity.

# ACCURACY, SENSITIVITY, SPECIFICITY



From the curve above intersection point is approximately at 0.285 cut off probability

# RECOMMENDATIONS

---

To improve the potential lead conversion rate X-Education will have to mainly focus important features responsible for good conversion rate are :-

- ❖ Focus on High-Value Tags: Prioritize leads with tags such as "Closed by Horizon" and "Lost to EINS," as they have a strong positive impact on conversion.
- ❖ Engage with Active Leads: Actively engage with leads who exhibit positive engagement signals, such as "Will revert after reading the email" and "Had a Phone Conversation."
- ❖ Leverage Effective Channels: Utilize channels like SMS communication and the Welingak website to effectively engage leads and increase conversion rates.
- ❖ Address Negative Signals: Identify and address negative signals such as "Busy" or "Unsubscribed" to mitigate their impact on conversion.
- ❖ Optimize Lead Origin and Specialization: Further investigate the impact of lead origin and specialization on conversion and consider optimizing strategies accordingly.
- ❖ Continuous Monitoring and Adaptation: Regularly monitor the performance of predictor variables and adapt strategies to maximize conversion rates over time.



**THANK YOU**

---