

# Lead Score Case Study Presentation



**Submitted by:**  
**Jay Joshi**

---

# Problem and Objectives:



---

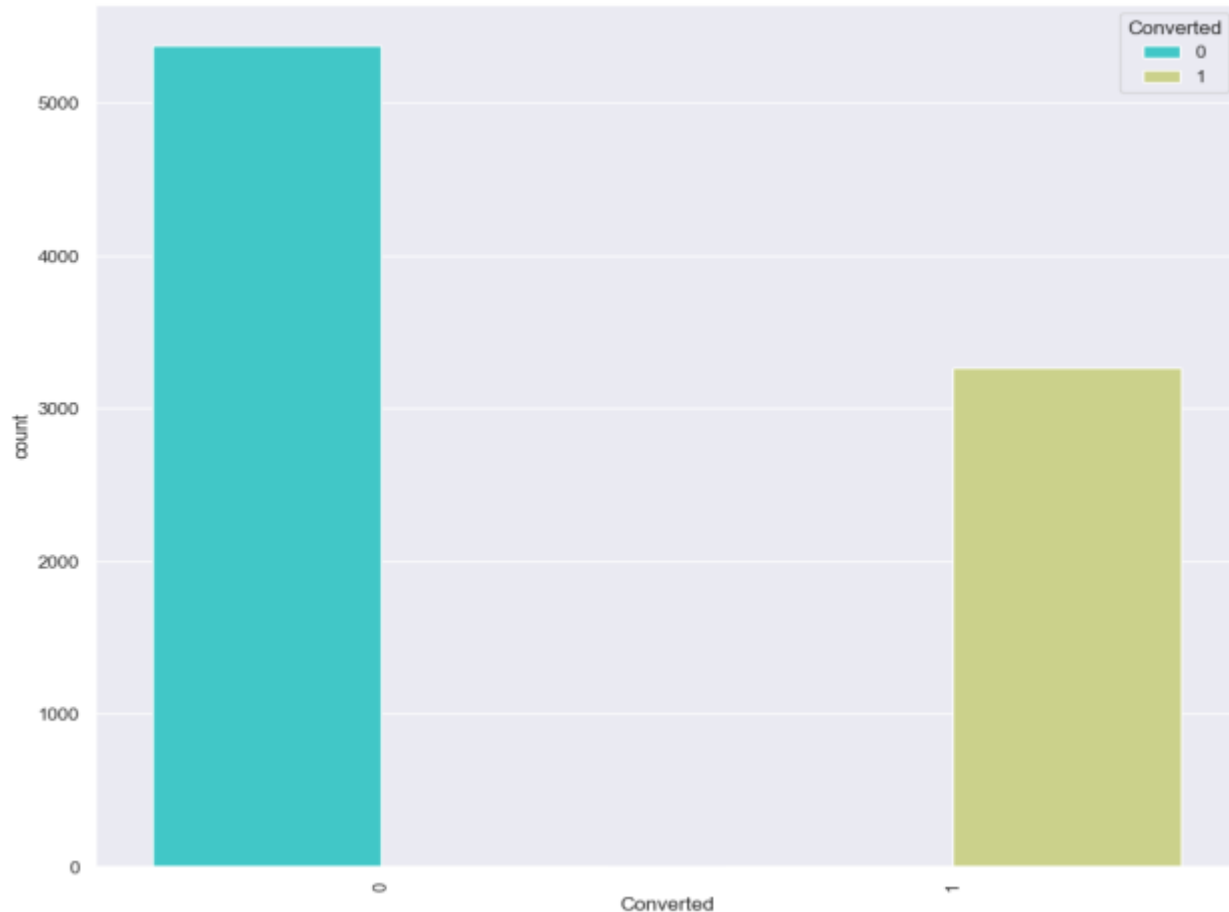
## Problem statement for the business side:

X Education sells online courses and needs help in identifying promising leads with a high likelihood of converting into paying customers. They require a model to assign lead scores based on conversion likelihood, with a target conversion rate of 80%.

## Objective approached by the business side:

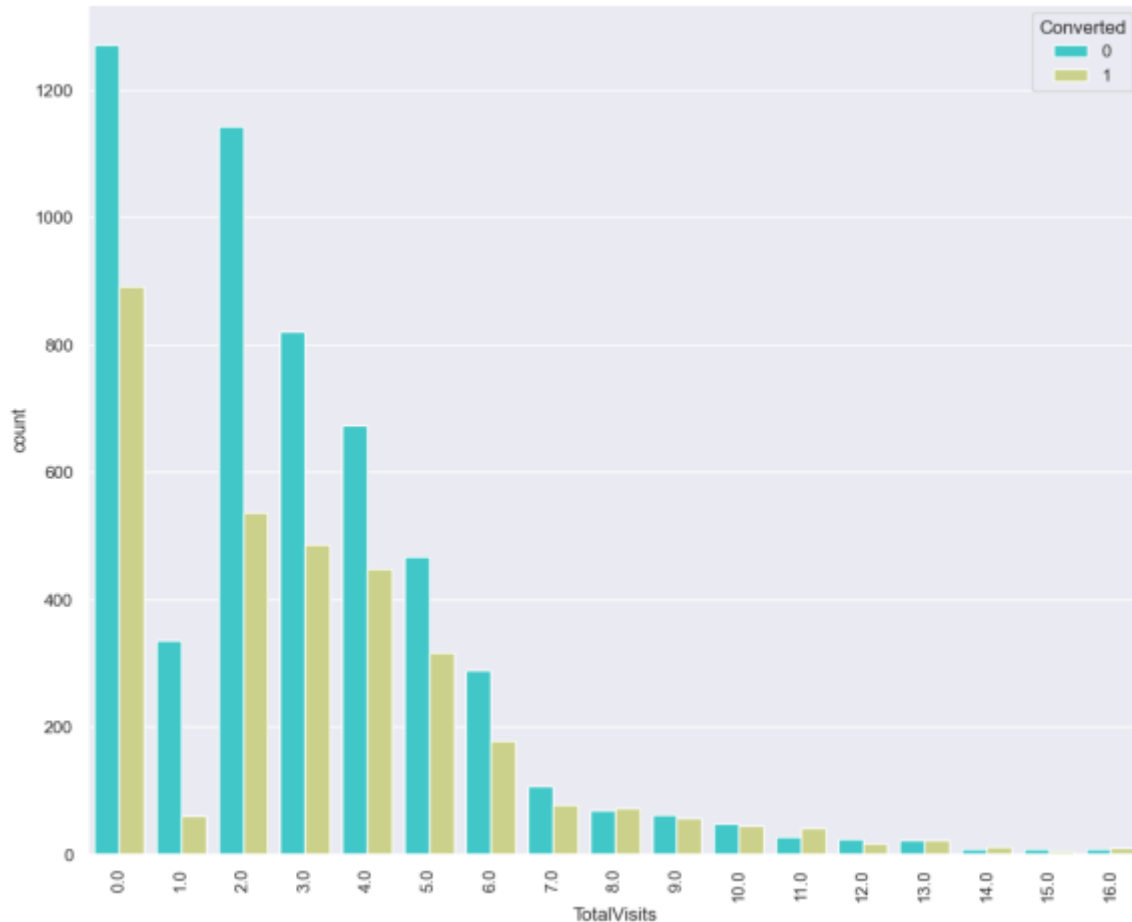
- Build the model for hot leads with greater than 80 % converted rate.
- Deploy model for further uses.
- Finding out the necessary leads which make conversion success.

# DATA ANALYSIS FOR UNIVARIANT ANALYSIS:



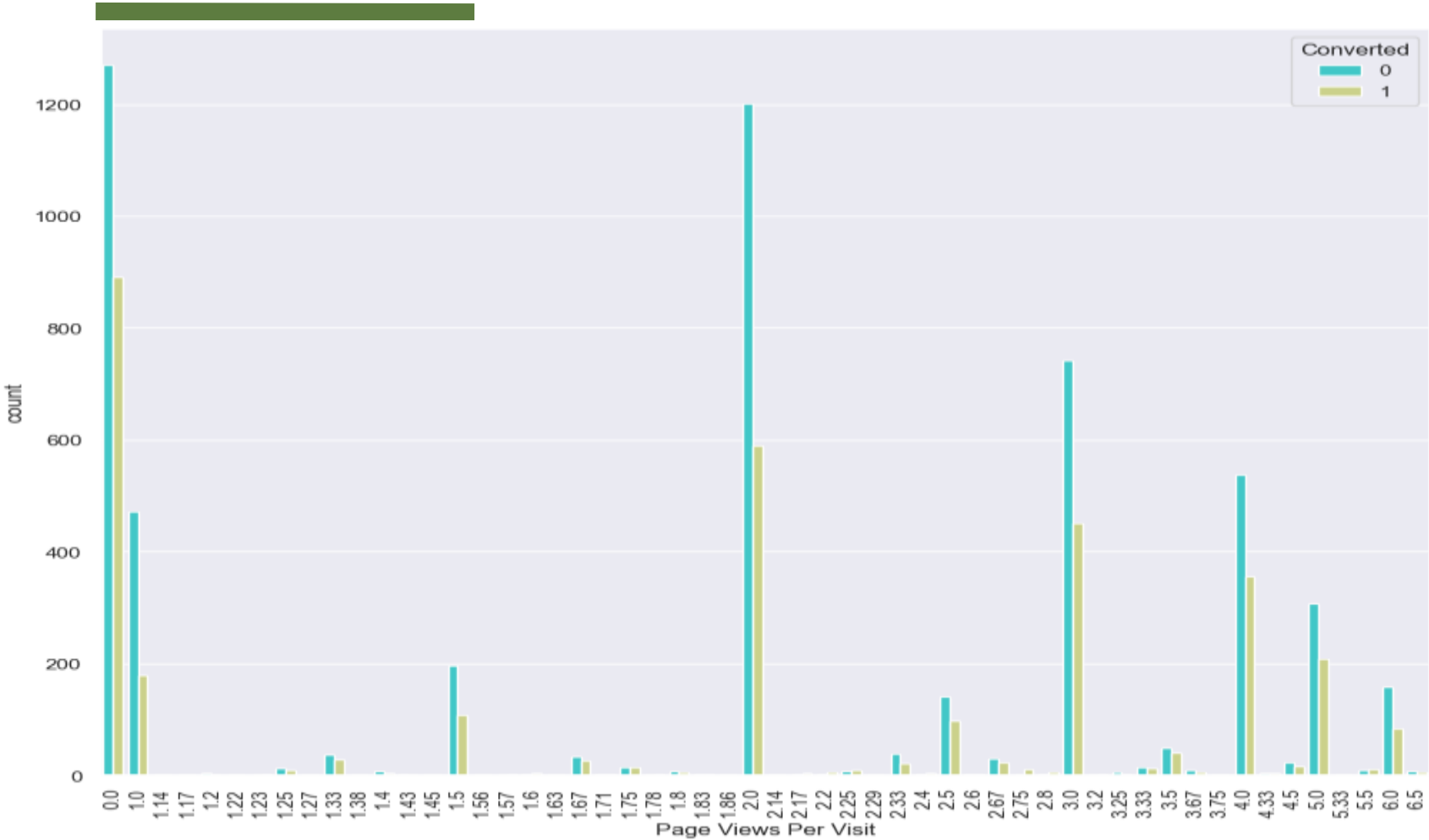
- In this plot, more than 70% of the individuals are not converted to pursue the course, as indicated by the data from the X education business side.

# UNIVARIATE ANALYSIS FOR TOTAL VISIT



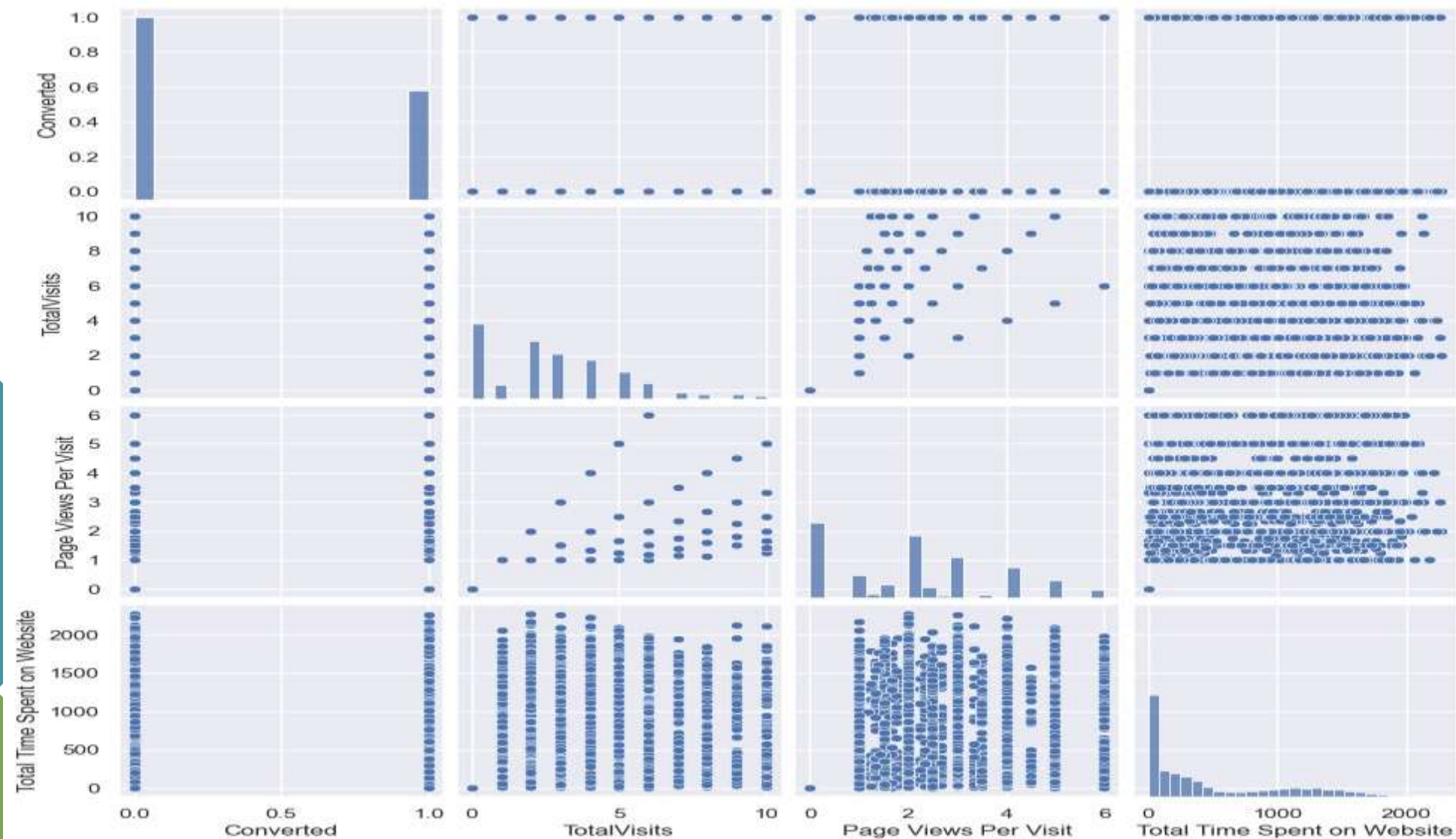
- Check out the graph depicting the total time people visited and whether they were converted.
- In general, the majority of people who have not visited and those who have repeated the course between 2 to 6 times are the most likely to convert.

# UNIVARIATE ANALYSIS FOR PAGE VIEW PER VISIT

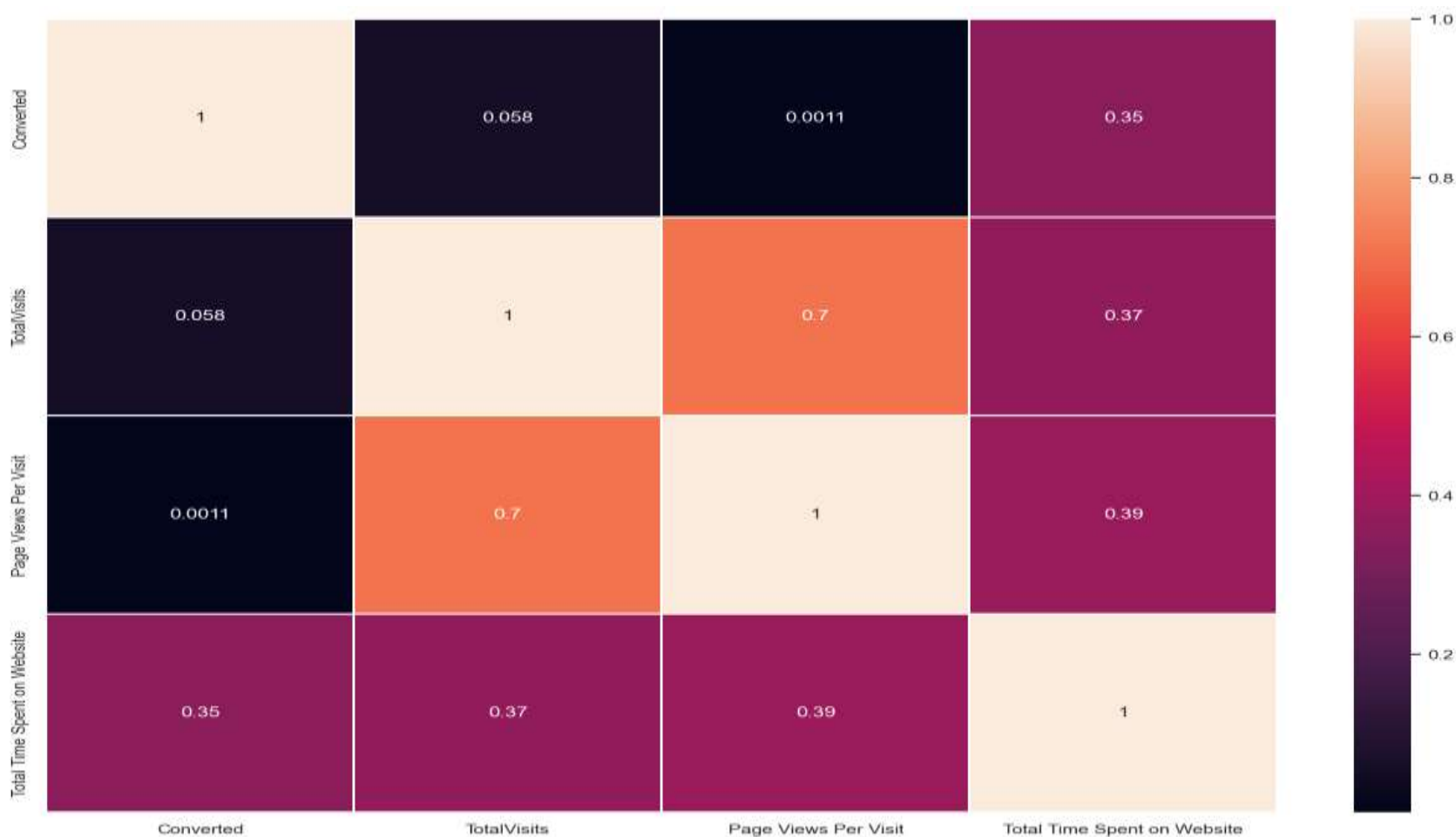


The graph depicts the conversion rate of no visit and 2.0 per page view, which is higher than the other segment.

# BIVARIANT ANALYSIS FOR TARGETED COLUMNS

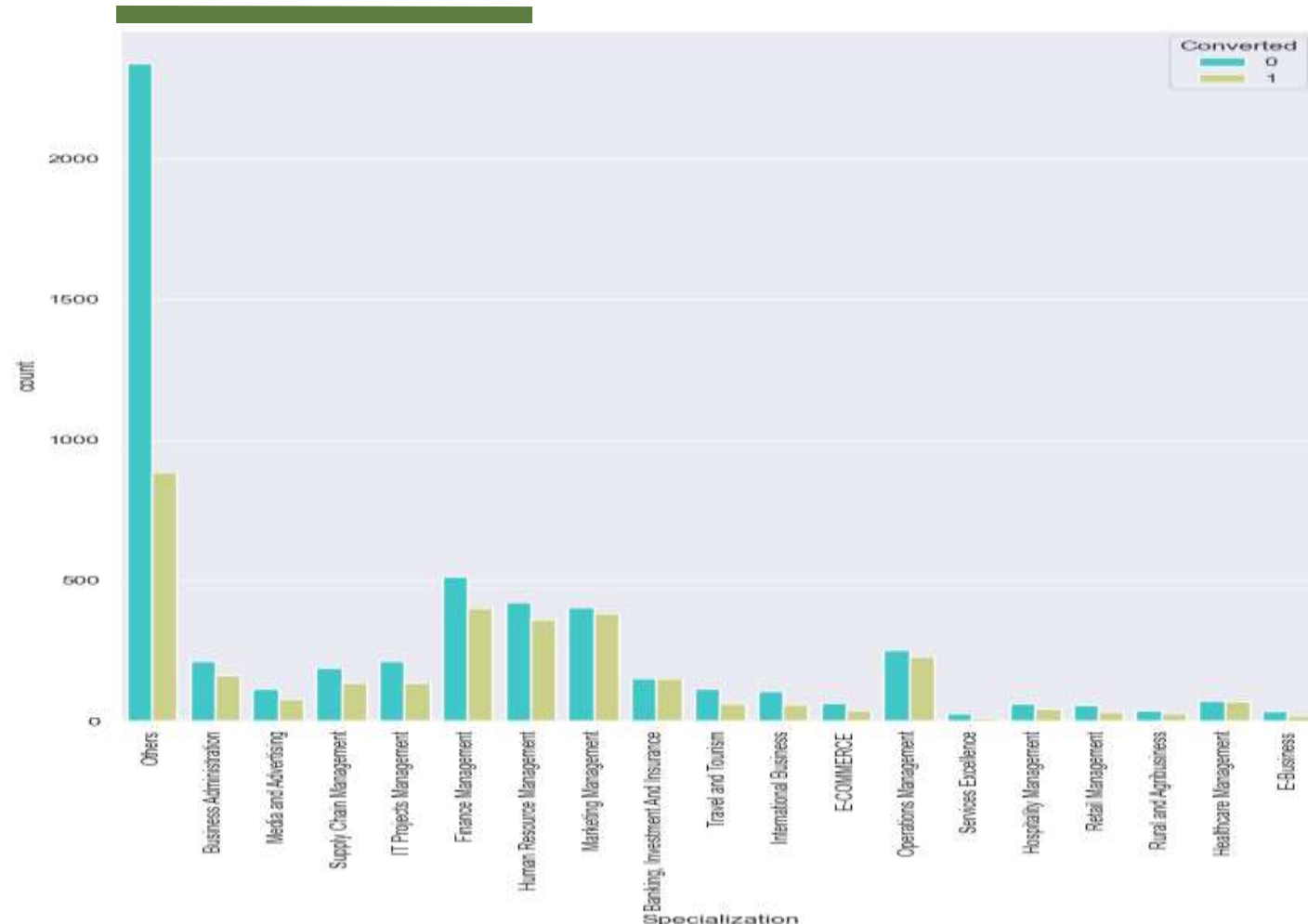


# CORRELATION BETWEEN THESE DATA



According to the image, the columns named "Total Visit" and "Page View Per Visit" have a high correlation value of 0.7.

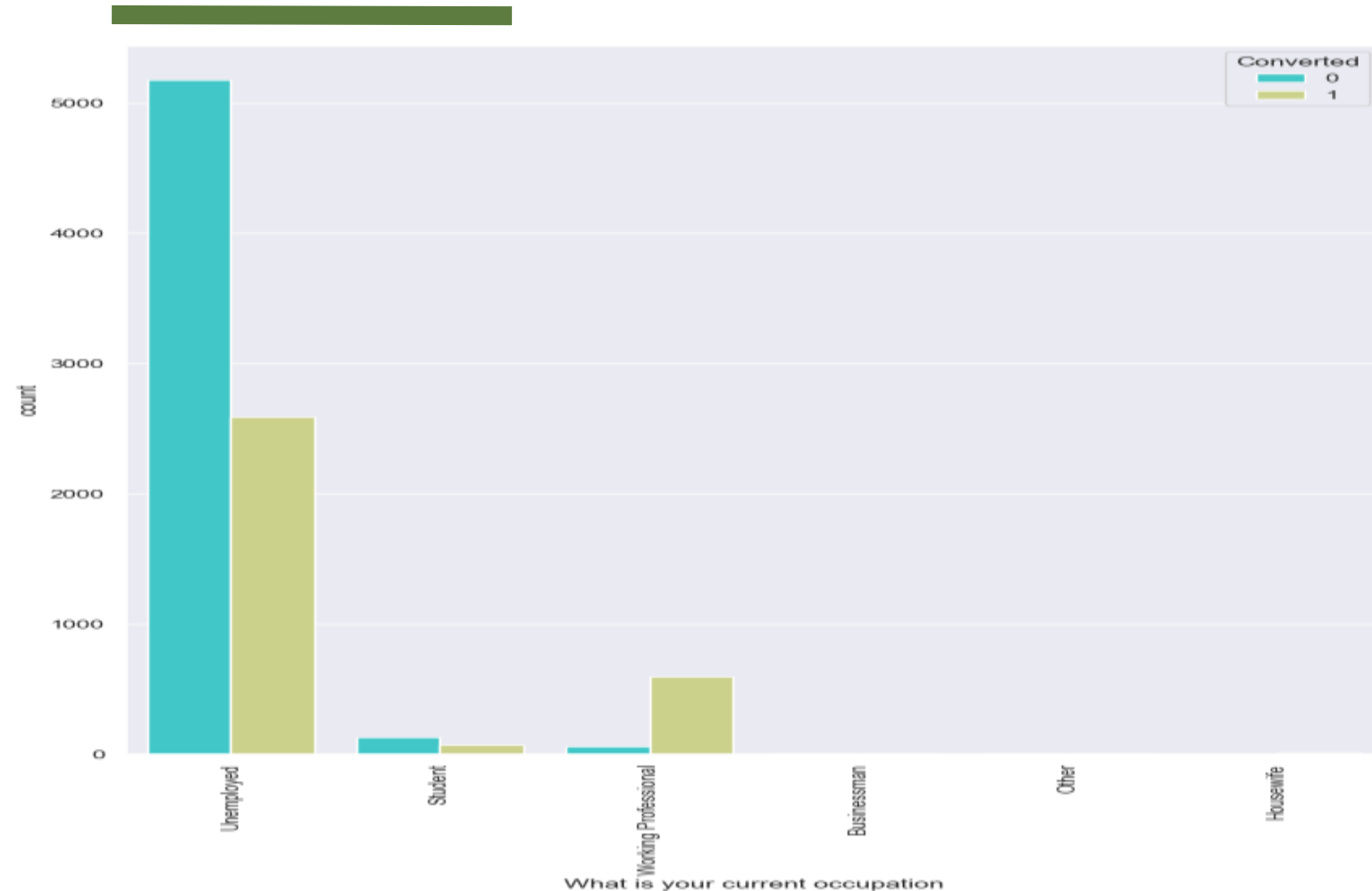
# ANALYSIS FOR SOME COLUMNS



- The data in the plot clearly indicates that there is a significant interest in finance management, HR management, banking, and investment among the majority of people.
- In this graph, "other" is mentioned as an additional variable, but it is unknown.

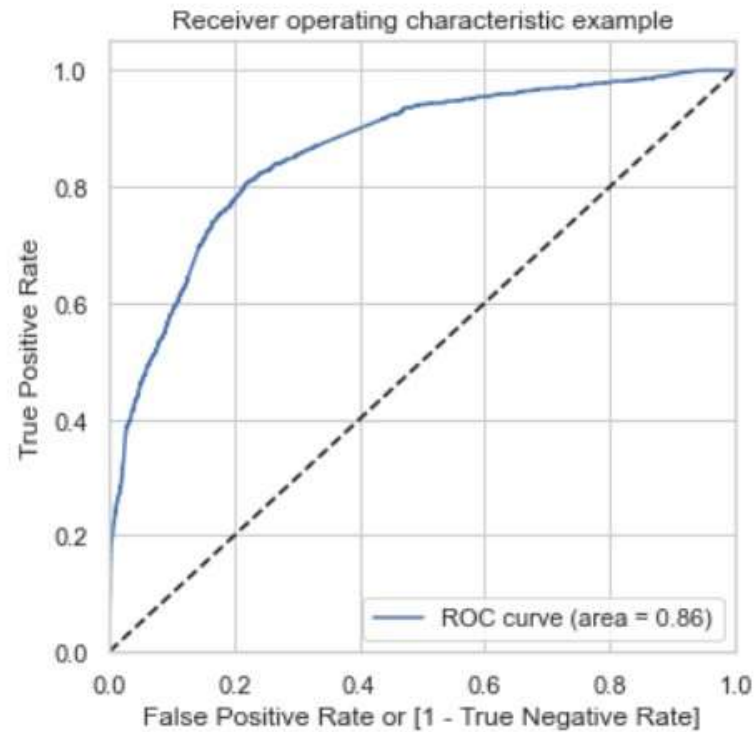


# ANALYSIS BASED ON OCCUPATION



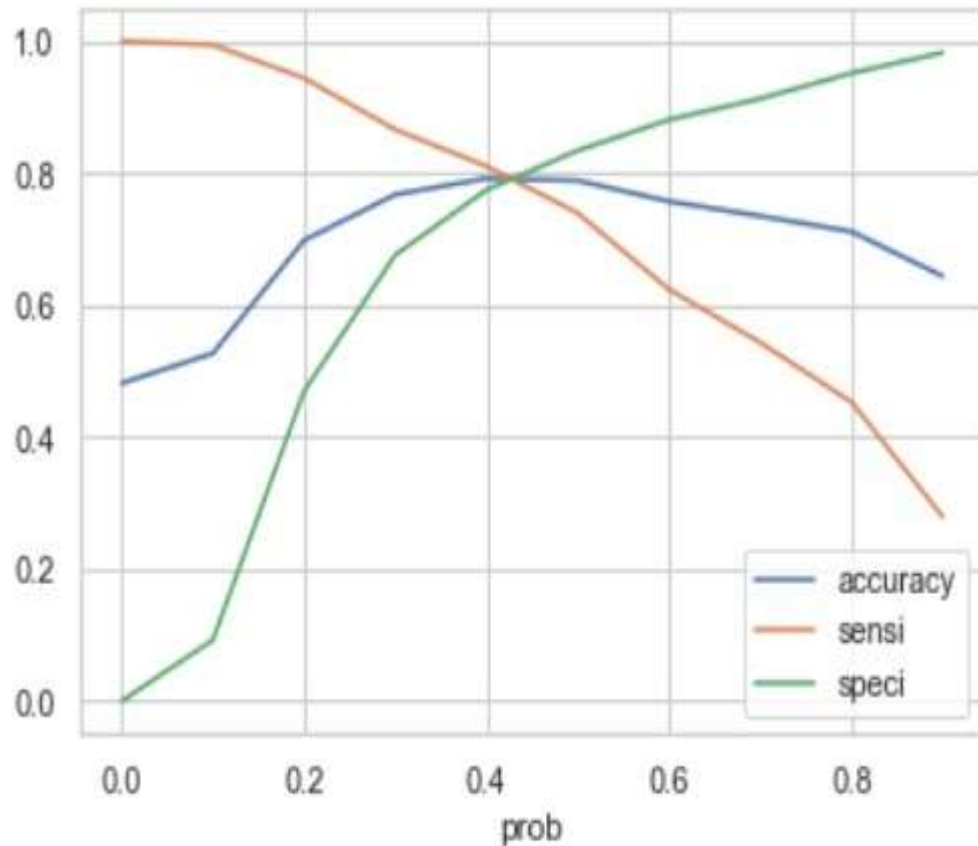
- This graph depicts the transition of unemployed individuals and working professionals to taking courses.

# ROC CURVE



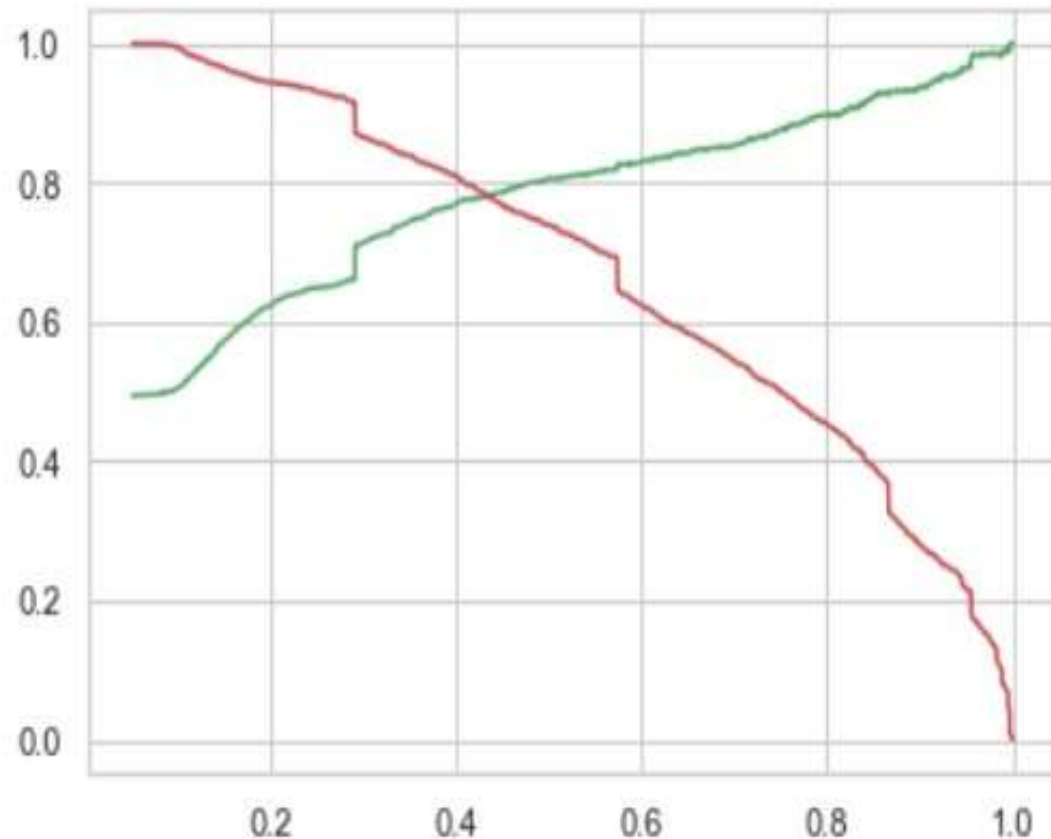
The ROC curve exhibits an impressive area of 0.86, indicating excellent performance.

# OPTIMAL VALUE FOR METRICS



As you can see, the optimal values of the three metrics occur around 0.42. Therefore, let's choose 0.42 as our cutoff now.

# PRECISION AND RECALL



- 0.42 represents the tradeoff between Precision and Recall. Therefore, we can confidently classify any Prospect Lead with a Conversion Probability exceeding 42% as a hot lead.

# SUMMARY



The most influential variables for potential buyers are:

- Total time spent on the website
- Total number of visits
- Lead source: Google, Direct traffic, Organic search website
- Last activity: SMS, Olark chat conversation
- Lead Origin: Lead add format
- Current occupation: working professional