

Lead Scoring Case Study

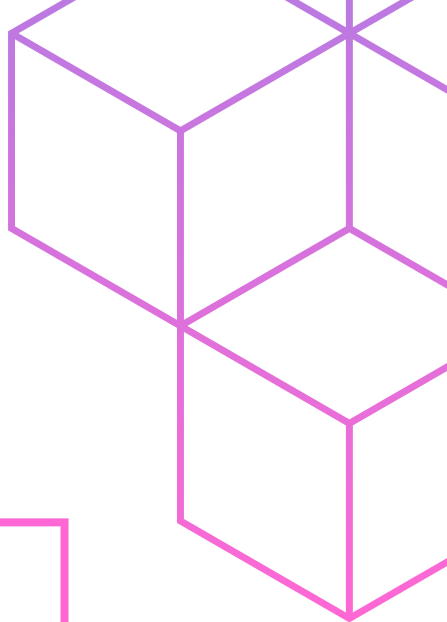
Problem Statement

- Company Overview: X Education sells online courses to industry professionals.
- Lead Generation: Interested professionals visit the website and fill out a form, becoming leads.
- Follow-Up Process: The sales team contacts leads via phone calls and emails.
- Current Conversion Rate: Approximately 30% of leads convert into customers (e.g., 30 out of 100 leads).
- Objective: Improve lead conversion by identifying high-potential leads, termed Hot Leads.
- Strategy: Focus the sales team's efforts on Hot Leads to increase the conversion rate and reduce outreach to less promising prospects.

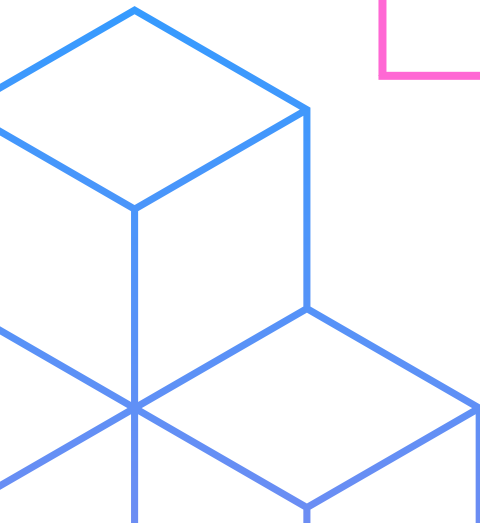
Objectives

- Develop a model to assign a lead score between 0 and 100 for each lead.
- Identify Hot Leads to increase the conversion rate.
- The CEO aims to achieve an 80% lead conversion rate.
- Handle peak-time actions efficiently.
- Optimize the use of manpower.
- Determine appropriate strategies after achieving target goals.

Problem Approach



Data Preparation	EDA	Dummy Variable Creation	Model Building
<ul style="list-style-type: none">Importing the data	<ul style="list-style-type: none">Cleaning the data by checking missing values and other infoPlotting graphs for better analysis	<ul style="list-style-type: none">Creating dummy variables as the data has categorical variables and we follow rfe method for model building	<ul style="list-style-type: none">Includes Test Train SplitScaling if required but suggestedRFE, VIF valuesSensitivity, SpecificityModel Evaluation for both test and train set



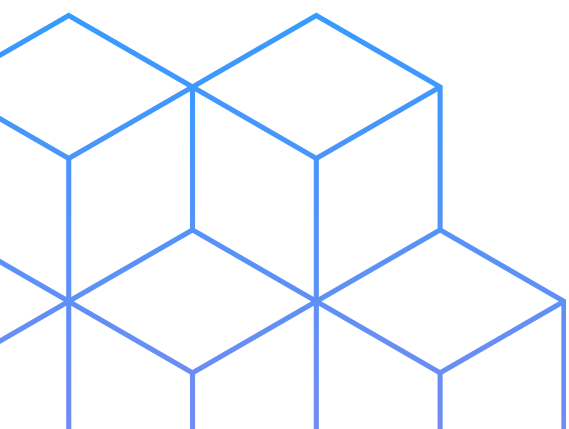
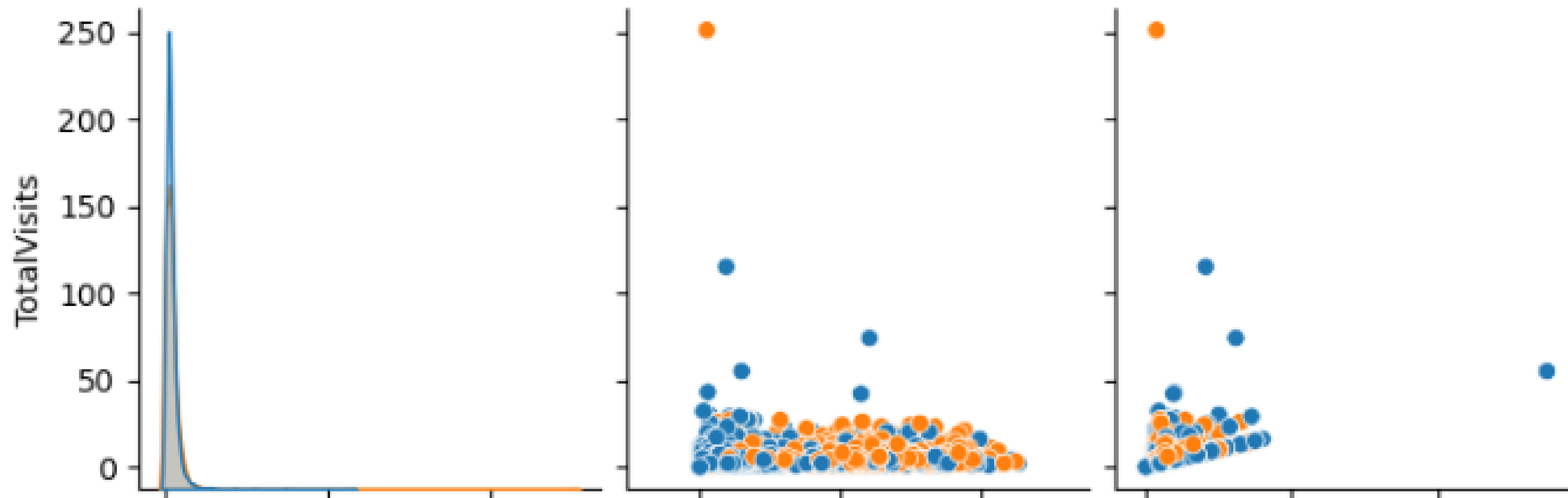
EDA Outcomes

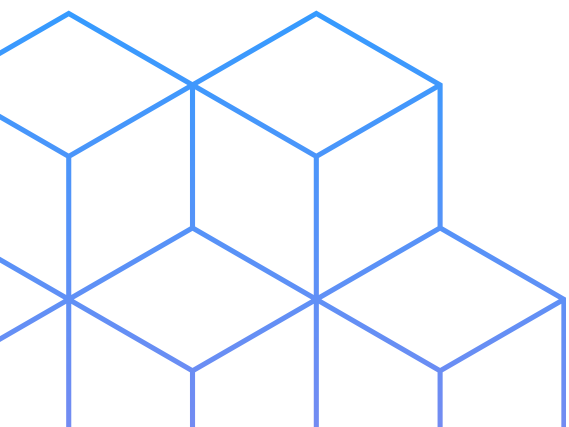
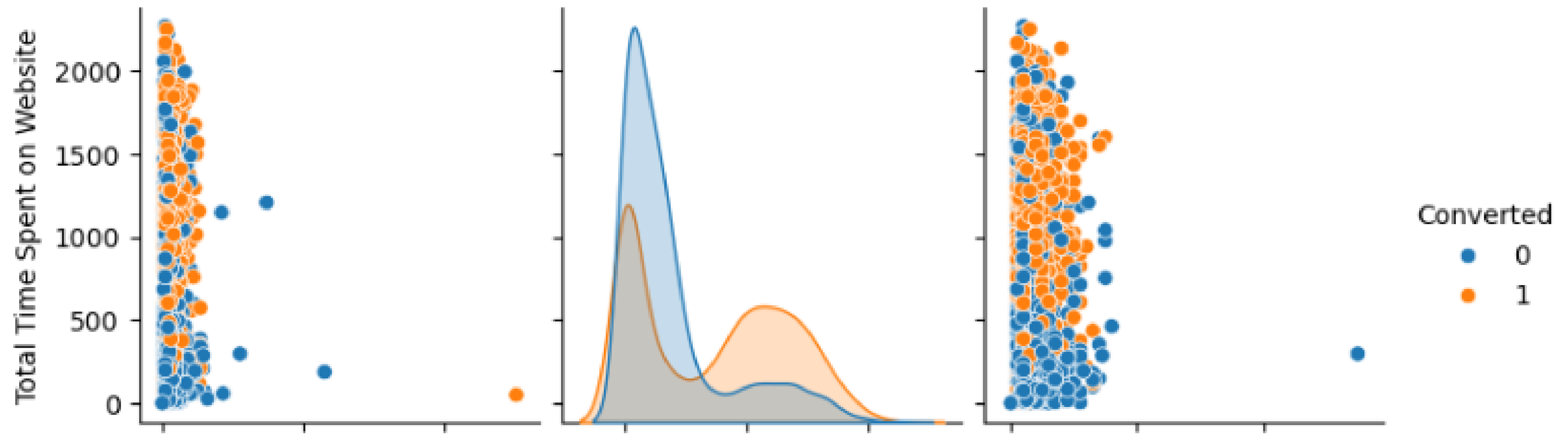
- Data Contained more than 9000 values in which 3000 are missing
- Removed City and Country columns as it was not get interpretation from these columns
- Few Columns Had 'Select' Values which are like default values don't have any weightage in analysis
- Before Removing Select Value columns we have checked the count and then removed

By the end of EDA We retrived 60% of data for further analysis which consistes model building and evaluation

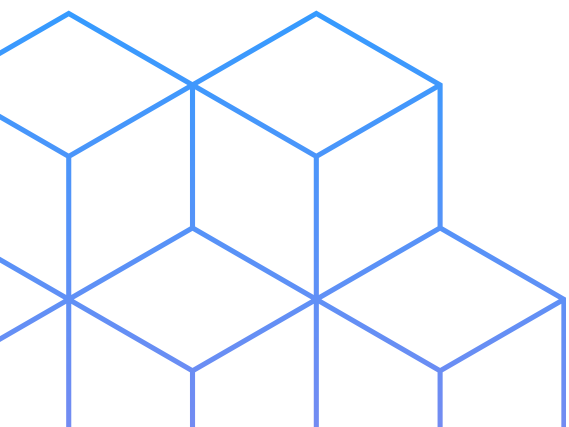
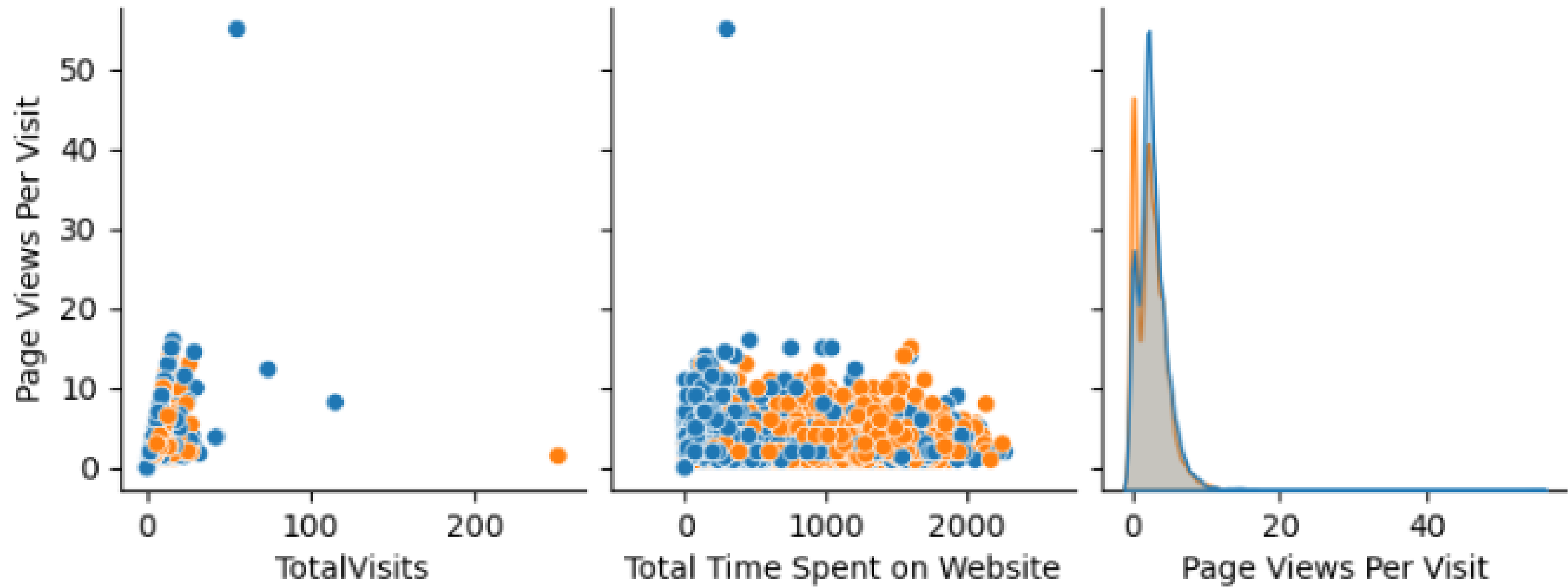
The pairplot visualization shows relationships between three variables: TotalVisits, Total Time Spent on Website, and Page Views Per Visit, categorized by the Converted status (0 for not converted, 1 for converted)

01





03



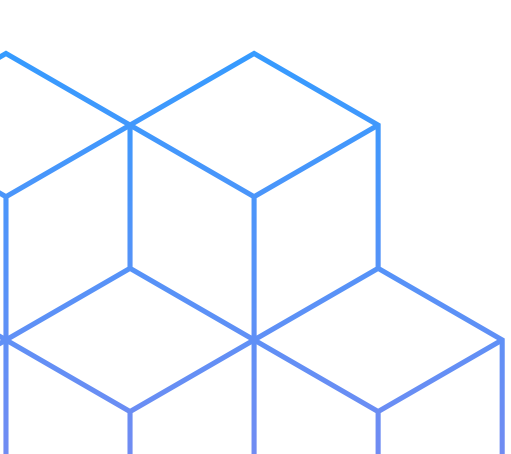
Plot Insights

Total Time Spent on Website vs. Conversion:

- There seems to be a strong positive association between Total Time Spent on Website and conversions. Users who spent more time on the website are more likely to convert (represented by more orange points in higher time regions).
- The kernel density estimates (KDE) show that for both converted and not converted, most users spend a relatively short amount of time on the website, but converted users show a secondary peak at higher time spent values.

TotalVisits vs. Conversion:

- Most users, whether converted or not, have a low number of total visits. However, there is a wider spread of visits among non-converted users.
- A few users with a high number of visits do not convert, suggesting that higher total visits alone don't guarantee conversion.



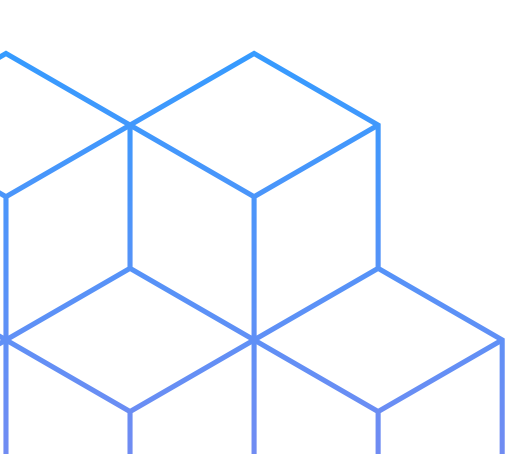
Plot Insights

Page Views Per Visit vs. Conversion:

- The majority of both converted and non-converted users have a low number of page views per visit, indicating that high page views per visit are not a common trait among users.
- There is a slight indication that users with very low page views per visit tend not to convert, but this trend is not strongly marked.

Interactions Between Variables:

- Total Time Spent on Website and Page Views Per Visit seem to show some interaction; users who spend more time on the website also tend to have more page views per visit, especially among converted users.
- There seems to be no strong pattern between TotalVisits and the other two variables when segmented by conversion status.



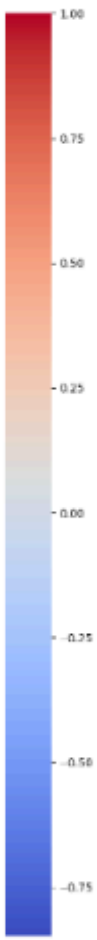
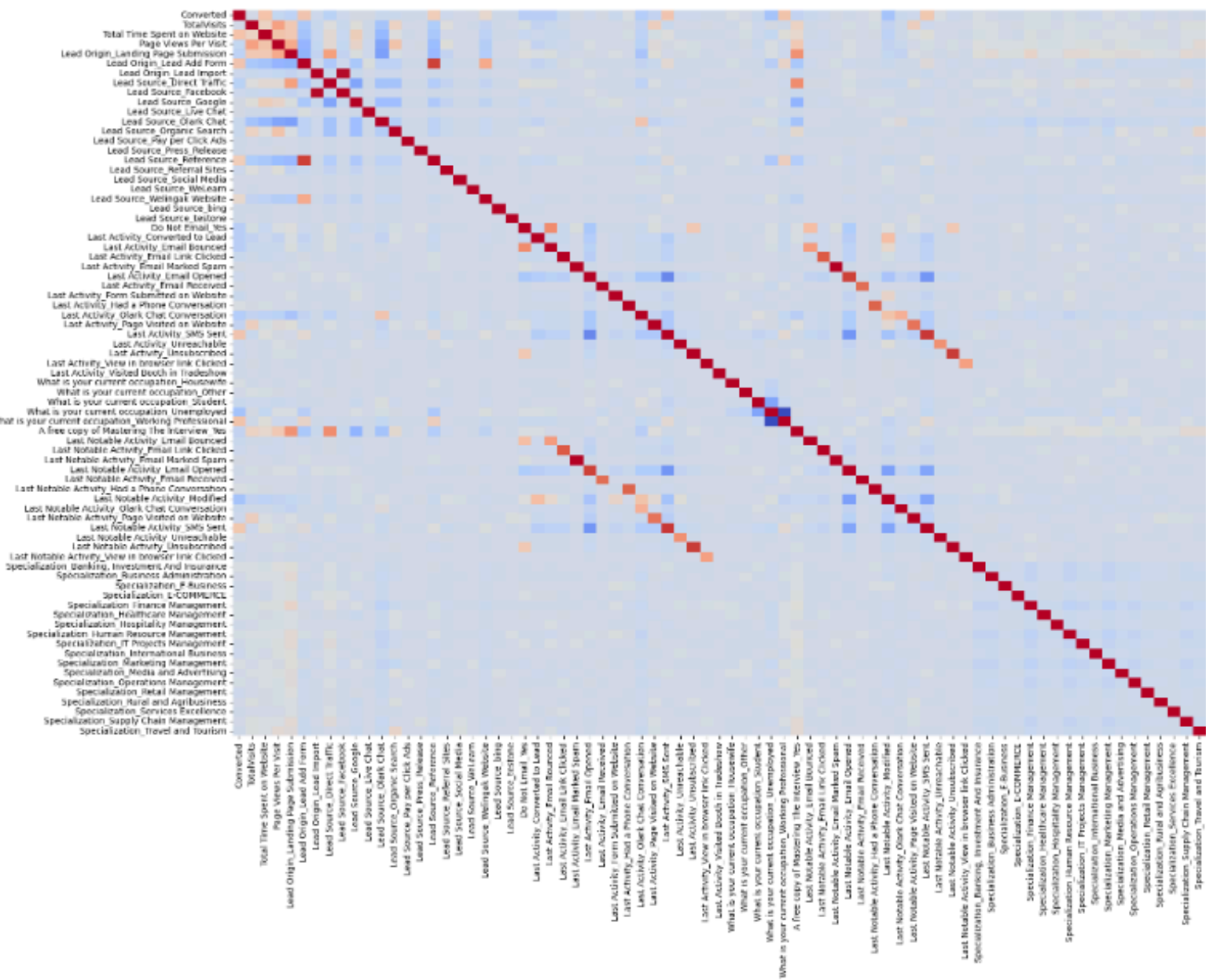
Plot Summary

In summary, the most significant insight is the positive relationship between the time spent on the website and the likelihood of conversion. Efforts to increase engagement time could potentially lead to higher conversion rates.

Dummy Variable Creation

To deal with the categorical data we need to create the dummy variable.

Correlation



- Strong positive correlations are shown by dark red squares off the diagonal. These indicate that the two variables increase together.
- Strong negative correlations are shown by dark blue squares. These indicate that as one variable increases, the other decreases.
- Light colors, particularly those close to white, indicate little to no correlation between variables
- If there are many dark red or dark blue squares clustered together, this could indicate multi-collinearity, where several variables are highly correlated with each other.

Model Evaluation

Accuracy

72.2%

Sensitivity

64%

Specificity

79%

Overall Model looks good

Summary

- Initial Stage: Generate a large number of leads.
- Middle Stage: Nurture leads by educating them and maintaining regular communication to enhance conversion rates.
- Prioritize Key Metrics:
 - Total Visits
 - Total Time Spent on Website
 - Page Views Per Visit
 - Use these to identify high-potential prospects.
- Manage Lead List:
 - Keep a detailed list of leads for targeted updates on new courses, services, job offers, and future study opportunities.
 - Tailor communications based on individual lead interests and needs.
- Engage with Converted Leads:
 - Conduct Q&A sessions to gather insights.
 - Schedule follow-ups to assess their interest and suitability for online courses.

Thank You