IDENTIFYING HOT LEADS FOR X EDUCATION

ENHANCING LEAD CONVERSION EFFICIENCY

OBJECTIVE: IMPROVE LEAD-TO-SALE CONVERSION RATE BY IDENTIFYING 'HOT LEADS'

SUBMITTED BY
CHINNI ANJANEYULU

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, search engines, and even social media sometimes. Once these people land on the website, they might browse the courses, fill out a form for the course, or watch some videos. When these people fill out a form with their email address or phone number, they are classified as leads. 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 into successful sales, while most of the leads do not. The typical lead to successful sale conversion rate at X education is around 30%.**
- **Now, although X Education gets a lot of leads, its lead-to-sale conversion rate is very poor. For example, if they acquire 100 leads in a day, only about 30 of them are converted into successful sales. 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 would go up as the sales team would now be focusing more on communicating with the potential leads rather than making calls to everyone. A typical lead conversion process can be represented using the following funnel.**
- **As you can see, there are a lot of leads generated in the initial stage (the initial pool of leads), but only a few of them come out as paying customers from the bottom (converted leads). In the middle stage (lead nurturing), you need to nurture the potential leads well (i.e., educate the leads about the product, constantly communicate, etc.) in order to get a higher lead conversion.**
- **X Education has appointed you to help them select the most promising leads, i.e., the 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, in particular, has given a ballpark estimate of the target lead conversion rate as being around 80%.**

APPROACH FOR LEAD SCORING CASE STUDY:

- **Reading and Understanding : Firstly I imported all the Libraries which are necessary for my analysis. Imported the Lead Converted Data from the csv file into Jupiter Notebook, Reading and understanding the Data.**
- **Data Handling and Cleaning : Later I cleaned the Data, At firstly Removed the features which are having null values more than 30 percent, Later I gone through each and every feature understand the feature is it important or not for analysis and get rid of Outliers from the data which will affect the analysis. Finally the Data is cleaned and Ready for analysis and Model Building.**
- **Exploratory Data Analysis : The data analysis step, particularly univariate and bivariate analysis, is critical in understanding the underlying patterns and relationships in the data. This step provides insights that guide the model-building process, ensuring that the resulting model is well-informed and robust.**
  - **Univariate analysis involves examining each variable in the dataset individually. It helps in understanding the distribution, central tendency, and variability of the data.**
  - **Bivariate analysis examines the relationship between two variables. It is essential for understanding interactions and dependencies in the data.**
- **Preparing the Data for Modelling : Data preparation is a crucial step in building any predictive model. It ensures that the data is clean, consistent, and ready for analysis, which ultimately leads to more accurate and reliable models.**
  - **Most machine learning algorithms require numerical input. Categorical data needs to be converted into a numerical format.**
  - **Creating dummy variables (one-hot encoding), label encoding, or using techniques like target encoding.**
  - **Dummy variables represent categorical data as binary vectors, making it possible for algorithms to interpret categorical data.**
  - **For a categorical variable with n categories, create n-1 binary columns, each representing the presence (1) or absence (0) of a category**

❖Building Model :

- Normalization/Standardization: Scaling numerical features for equal contribution.
- Train-Test Split: Dividing data into training (70%) and testing sets (30%) to evaluate model performance
- Purpose: Ensure the model generalizes well to unseen data and prevents overfitting.
- Instantiate the Model: Creating a logistic regression model instance.
- Instantiate the Model: Creating a logistic regression model instance.

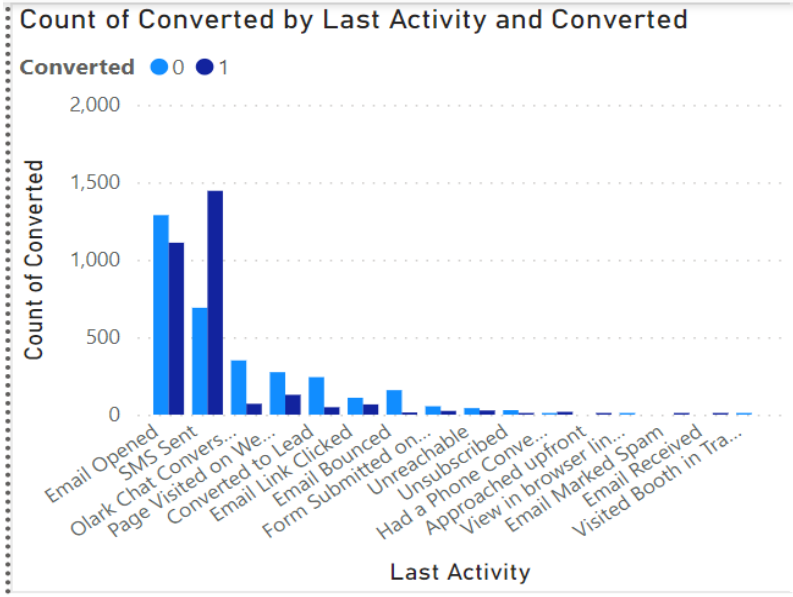
❖ Model Evaluation : Evaluating the model using accuracy score, precision score, recall score, specificity , sensitivity .

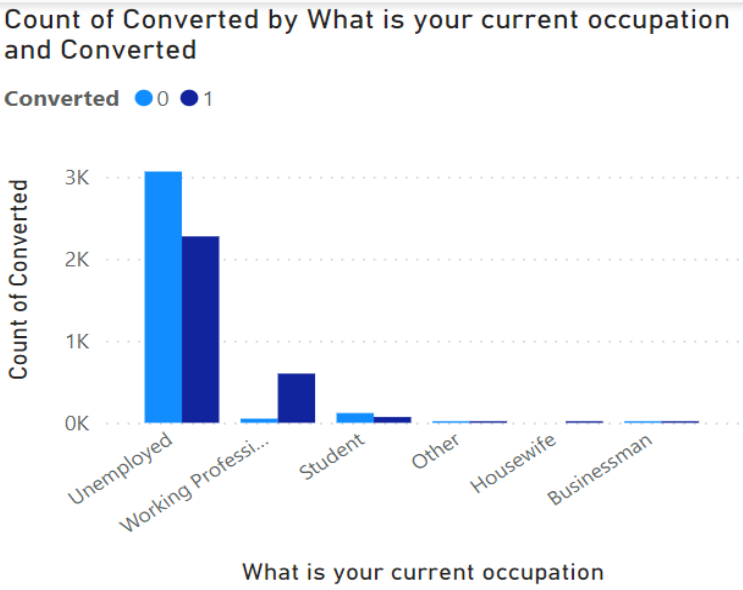
- Accuracy is the ratio of correctly predicted instances to the total instances.
- Precision (also called Positive Predictive Value) is the ratio of correctly predicted positive observations to the total predicted positives.
- Sensitivity (Recall) is the ratio of correctly predicted positive observations to the all observations in actual class.
- Specificity is the ratio of correctly predicted negative observations to the all observations in actual negative class.
- The ROC curve is a graphical representation of a classifier's performance across different thresholds, plotting the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity).
- The threshold is the probability cut-off point that determines the classification of a predicted probability.

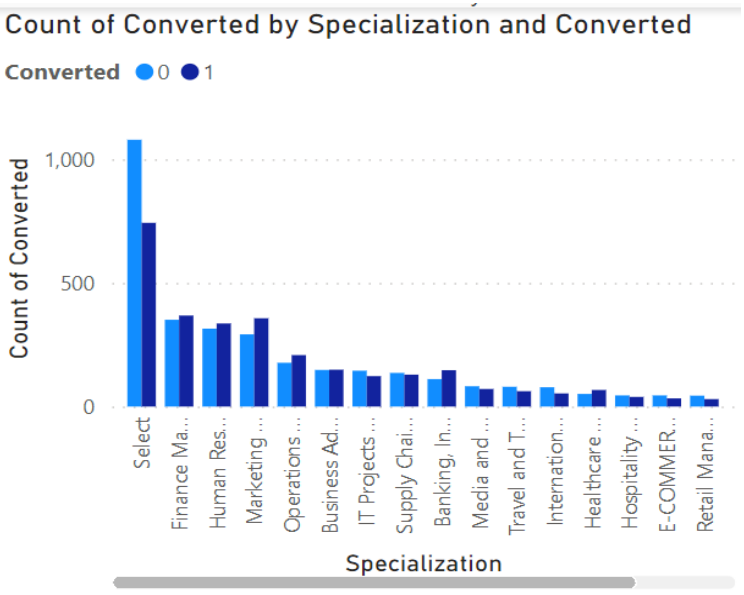
❖ Predicting on Test Data:

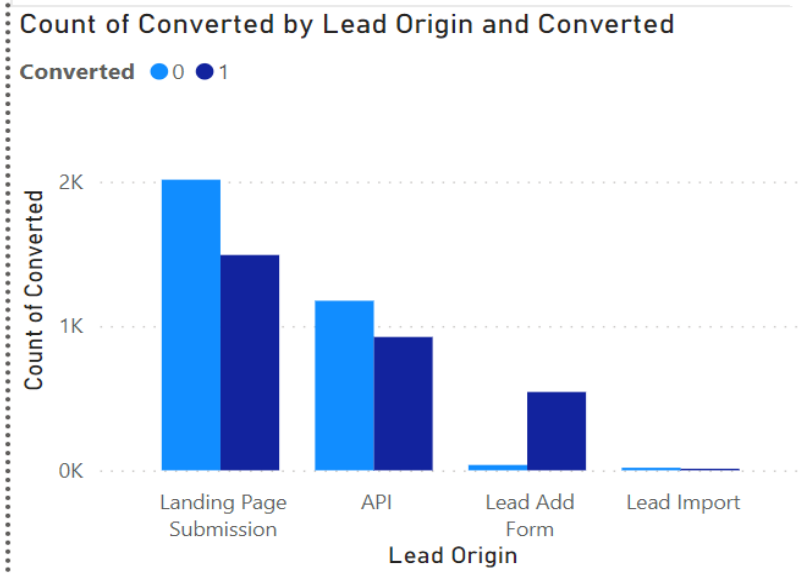
- Test data evaluation involves assessing the performance of a machine learning model using a separate dataset that was not used during training. This helps determine how well the model generalizes to new, unseen data.
- To ensure that the model is not overfitting to the training data and can accurately predict outcomes for new inputs.
- Test the model on which we have trained above.
- Ensure all preprocessing steps applied to the training data (e.g., scaling, encoding) are also applied to the test data to maintain consistency.
- Use the trained model to predict outcomes on the test data.
- Obtain predicted probabilities and class labels.
- Compare the predicted outcomes to the actual outcomes in the test set.
- Calculate performance metrics such as accuracy, precision, recall, specificity, and ROC-AUC score.

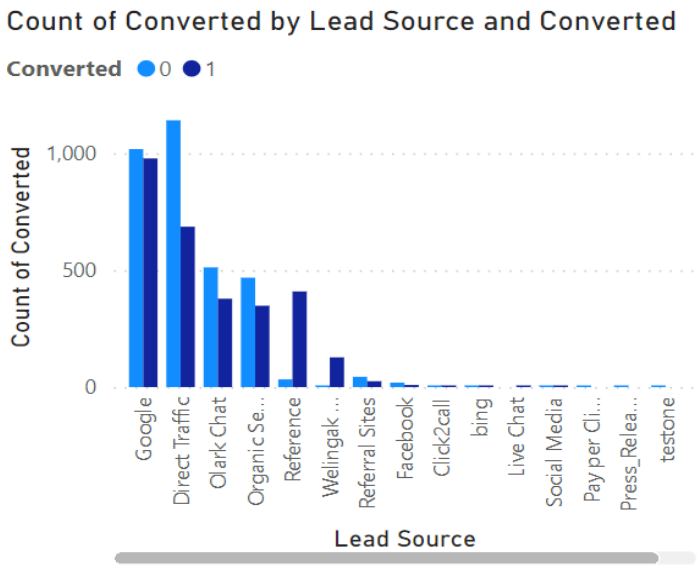
VISUALISATIONS : HOW EACH CATEGORICAL VARIABLE VARY WITH TARGET VARIABLE

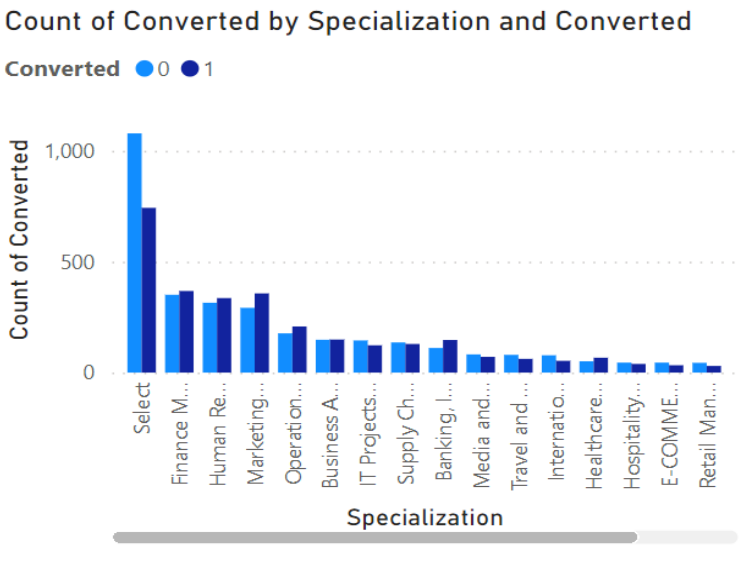




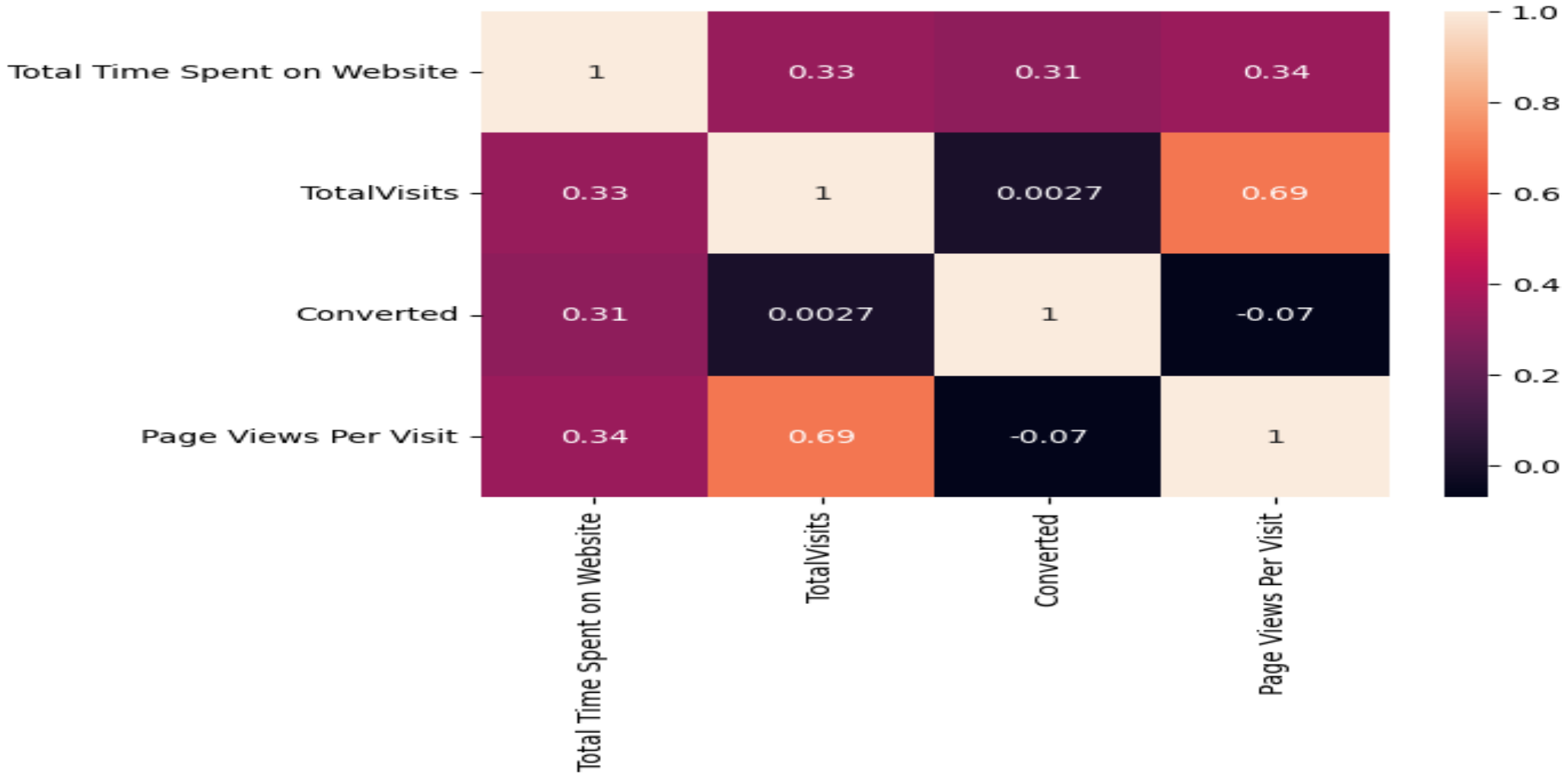




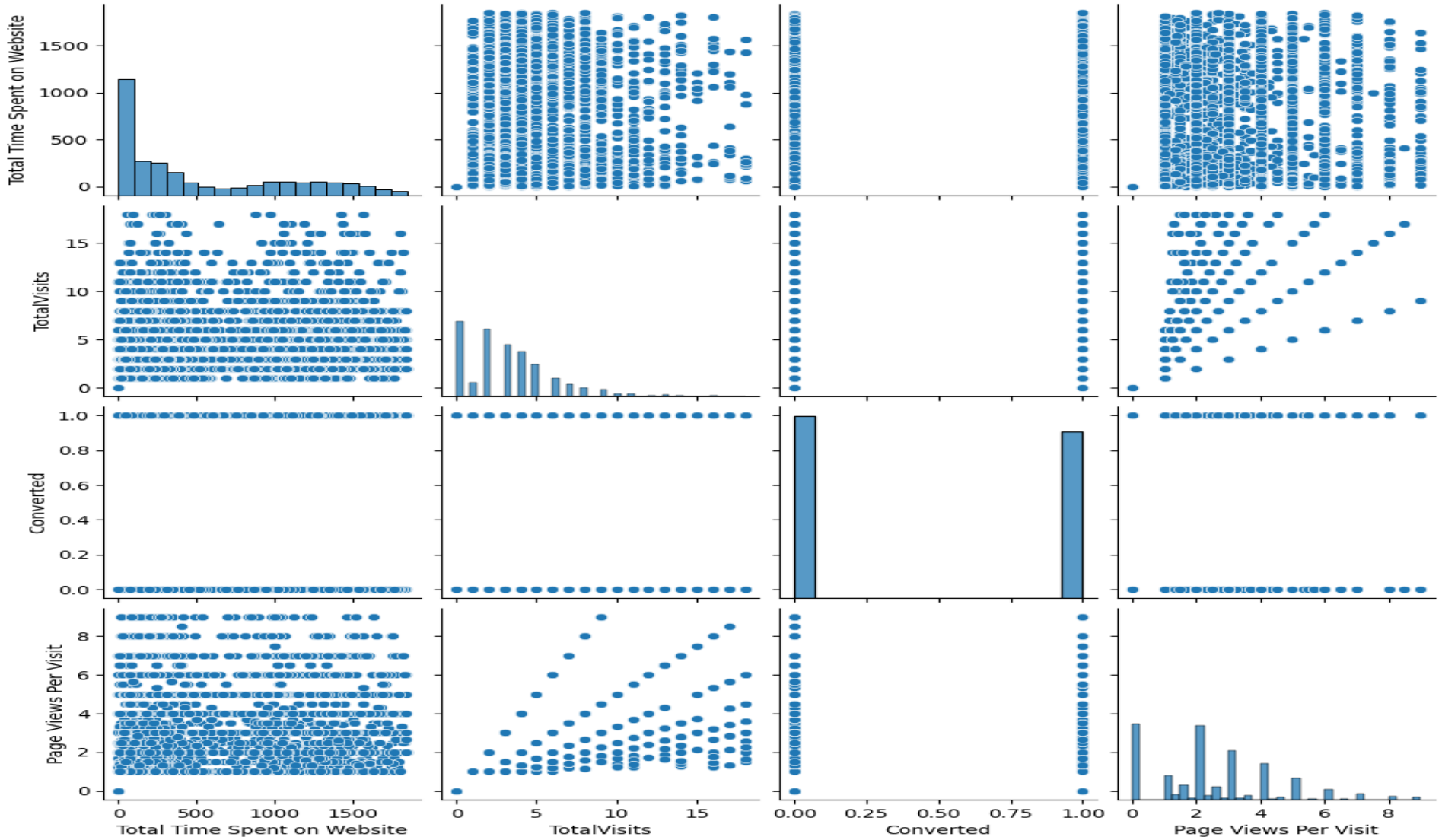




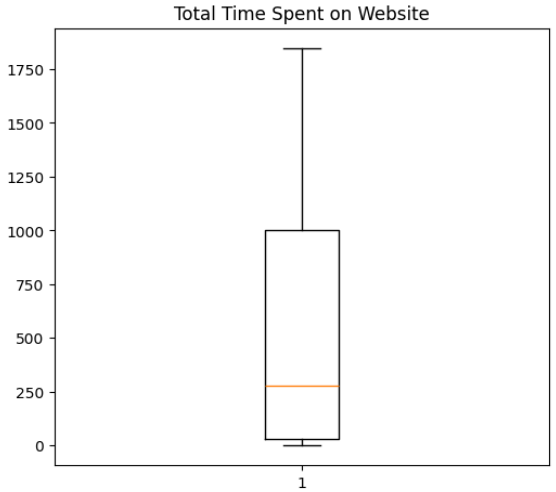
CORRELATION AMONG NUMERICAL VARIABLES:

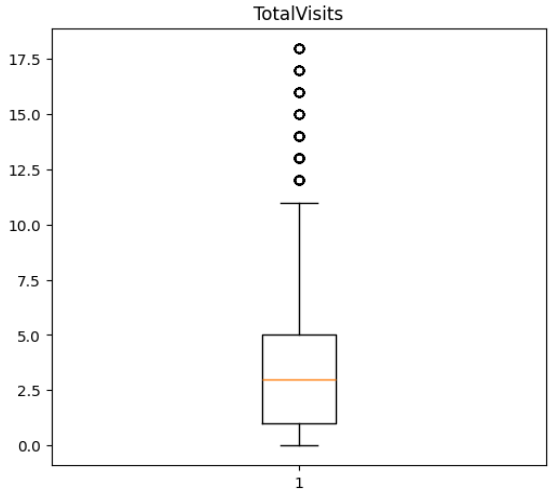


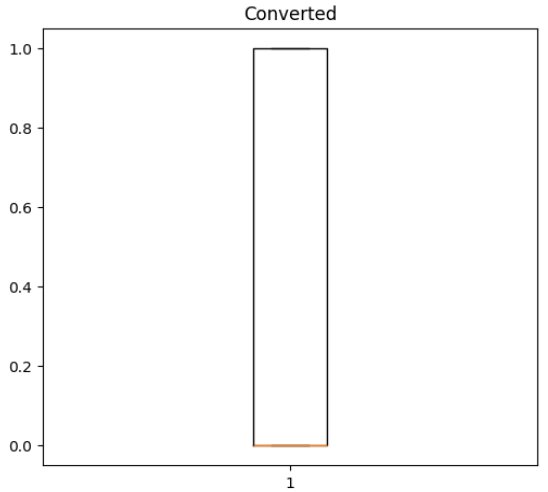
SCATTER PLOT WITH NUMERIC VARIABLES :
RELATIONSHIP BETWEEN TWO CONTINUOUS VARIABLES

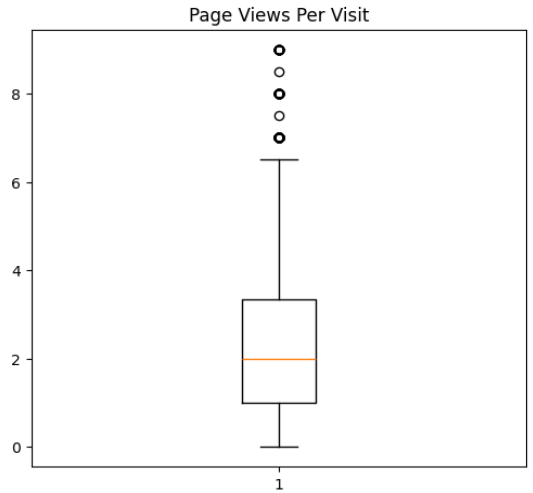


BOX PLOT : SUMMARIZES THE NUMERICAL DISTRIBUTION









FEATURES PREDICTING THE MODEL :

- 'TotalVisits’
- 'Total Time Spent on Website’
- 'Do Not Email_yes'
- 'Leadsource(Olark Chat)’
- 'Leadorigin(Lead Add Form)'
- 'Leadactivity(Converted to Lead)'
- 'Leadactivity(Had a Phone Conversation)'
- 'Leadactivity(Olark Chat Conversation)’
- 'Leadactivity(SMS Sent)',
- 'Cur_Occu(Working Professional)'

RESULTS IN BUSINESS TERMS :

- **Focus on leads with high engagement indicators. Use the variables such as 'Total Visits', 'Total Time Spent on Website', and 'Lead activity(SMS Sent)' to identify high-potential leads. These leads are more likely to convert due to their higher interaction with the website and received communications.**
- **Use a combination of emails, SMS, and phone calls to reach out to potential leads. For leads marked 'Do Not Email yes', prioritize phone calls and SMS.**
- **Analyse competitor strategies and market trends to refine future sales approaches**
- **Conduct market research to identify new potential market segments and opportunities.**
- **Efficient Use of Resources: Effective allocation of leads and follow-ups should result in a more efficient use of the intern’s time and resources.**
- **Feedback and Improvement: Collecting feedback from leads and interns will provide insights into what worked and what needs improvement, leading to refined strategies for future campaigns.**
- **Improved Lead Engagement: Leads should show higher engagement levels if the calling strategy is effective.**

SUMMARY OR INSIGHTS :

- **Based on the analysis and the logistic regression model developed, the following recommendations are made:**

  1. Prioritize Hot leads from the “Total” Visits” and “Total Time spend on web website” these two features identifies Hot Leads.

  2.Prioritize leads from the lead sources "Welingak Websites" and "Reference" for targeted communication.
  3.Focus on working professionals as they are more likely to convert.
  4.Pay attention to leads whose last activity was "SMS Sent."
  5.Engage with leads who have spent more time on the website.
  6.Give special attention to leads from the lead source "Olark Chat.“

  7. Leads with last activity “Converted to Lead” are most likely to convert

- **Conversely, avoid contacting leads with the following attributes:**

  1.Leads whose last activity was "Olark Chat Conversation.“ and “Had a phone conversation”
  2.Leads with the Lead Origin was “Lead Add Form”
  3.Leads who have opted for "Do not Email" as "yes.“

- **In addition to the above recommendations, the company should also consider lead scores, which are probabilities obtained through the algorithm. Leads with scores exceeding 80% are highly likely to convert and should be prioritized to expedite the conversion rate.**

- **By implementing these recommendations and the lead scoring model, X Education can significantly improve its lead conversion rate and make its marketing efforts more efficient and effective.**

INSIGHTS :

- **Leads who spend more time on platform are more likely to convert.**
- **Leads who spend more time on platform are more likely to convert.**
- **Leads with a higher number of page views per visit are more likely to convert.**
- **Leads from India show a high likelihood of converting compared to other countries.**
- **Leads who have received an SMS as their last activity are more likely to convert.**
- **Leads who have opened an email as their last activity show a higher likelihood of converting.**
- **Leads originating from landing page submissions and API show a higher likelihood of converting.**
- **Leads coming from Google (organic search or paid ads) show a high likelihood of converting**
- **Leads from direct traffic have a higher conversion rate.**
- **Leads who interact through Olark Chat show a significant likelihood of converting.**

INSIGHTS :

- **Leads interested in Human Resource Management show a strong likelihood of converting.**
- **Leads focusing on Finance also exhibit a high conversion rate.**
- **Leads interested in Marketing have a significant likelihood of converting.**
- **Working professionals also exhibit high conversion rates.**
- **Unemployed leads show a strong interest in educational programs, particularly those that can enhance their skills and improve job prospects.**
- **By implementing these recommendations and the lead scoring model, X Education can significantly improve its lead conversion rate and make its marketing efforts more efficient and effective.**
- **In addition to the above recommendations, the company should also consider lead scores, which are probabilities obtained through the algorithm. Leads with scores exceeding 80% are highly likely to convert and should be prioritized to expedite the conversion rate.**

RECOMMENDATIONS:

- **Ensure your website is easy to navigate and provides a seamless experience. Reduce load times and make key information readily accessible to encourage longer and more frequent visits.**
- **Offer high-quality, engaging, and relevant content that encourages visitors to spend more time on your site. This could include detailed articles, videos, infographics, and interactive tools**
- **Use SMS for time-sensitive information, reminders, and follow-ups. Ensure messages are concise and have a clear call to action.**
- **Create compelling subject lines and engaging content to increase email open rates. Ensure emails are visually appealing and provide valuable information.**
- **Continuously test and improve landing page design, copy, and call to action. Use A/B testing to determine the most effective elements.**
- **Continue to invest in SEO and Google Ads to capture high-intent traffic. Focus on relevant keywords, high-quality content, and optimized landing pages.**
- **Strengthen brand awareness and loyalty through consistent content marketing, email campaigns, and social media engagement. Encourage repeat visits and direct traffic.**
- **Optimize the live chat experience on your website. Train chat agents to be responsive and knowledgeable, and use chatbots to handle common queries.**
- **Develop targeted marketing campaigns and content for each specialization. Highlight the career benefits, industry demand, and success stories for HRM, Finance, and Marketing.**
- **Create targeted messaging for unemployed individuals and working professionals. For the unemployed, focus on how the program can improve job prospects and offer career support. For working professionals, emphasize career advancement, skill enhancement, and flexibility.**
- **For unemployed individuals, emphasize career counselling, job placement services, and internships. For working professionals, highlight networking opportunities, advanced career support, and industry connections.**