LEAD SCORING CASE STUDY

BY: ANEESH PRABHU, ANANTH KUMAR, EVA ANAND

Problem Statement

- X Education, an online course provider for industry professionals, faces a challenge with its low lead conversion rate, which is typically around 30%. The company generates many leads through marketing on various platforms like Google and referrals, but only a small percentage of these leads convert into paying customers. To improve efficiency, X Education wants to identify the "Hot Leads"—those most likely to convert—so the sales team can focus on them, ultimately increasing the lead conversion rate. The company wants to build a model that assigns a lead score to each prospect.

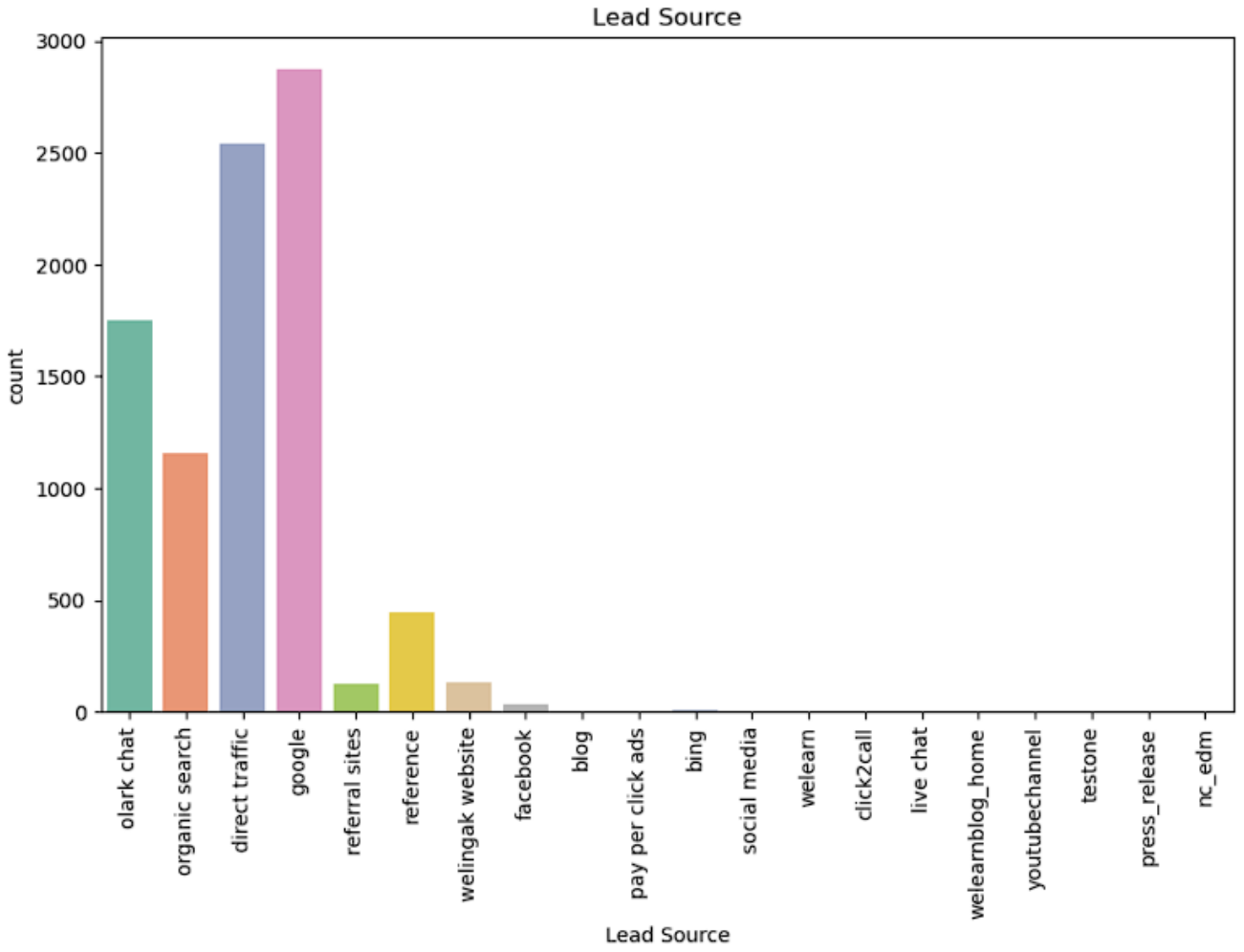
- The leads with higher scores would have a higher chance of converting, while those with lower scores would have a lower chance. The goal is to achieve a lead conversion rate of around 80%, focusing efforts on the most promising leads to improve the sales process and conversion outcomes.

STEPS TO PERFORM ANALYSIS

- Clean and prepare the data by removing duplicates, handling missing values, dropping irrelevant columns, imputing values when needed, and addressing outliers.
- Perform Exploratory Data Analysis (EDA) using both Univariate & Bivariate analysis. The top three variables to focus on are found to be:
    a. Total Visits
    b. Total Time Spent on Website
    c. Page Views Per Visit
- Logistic regression is used for the model making and prediction.
- Model is validated for accuracy.
- Using the model and EDA suitable conclusions can be drawn to help X Education get a higher conversion rate.

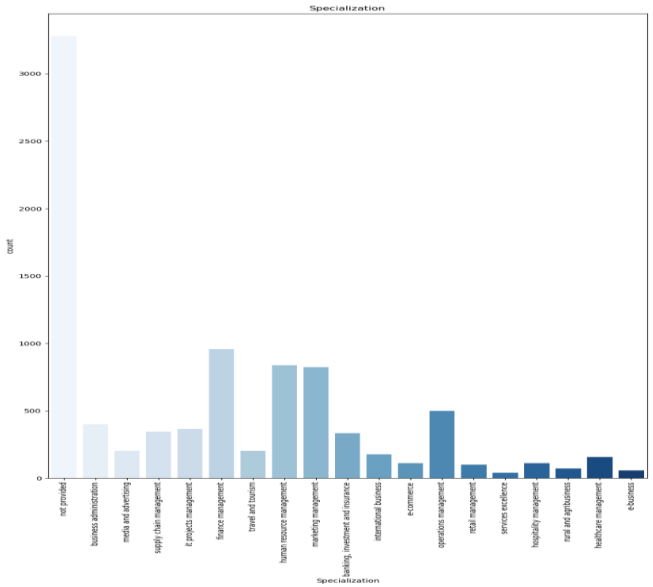
EDA – Univariate Analysis

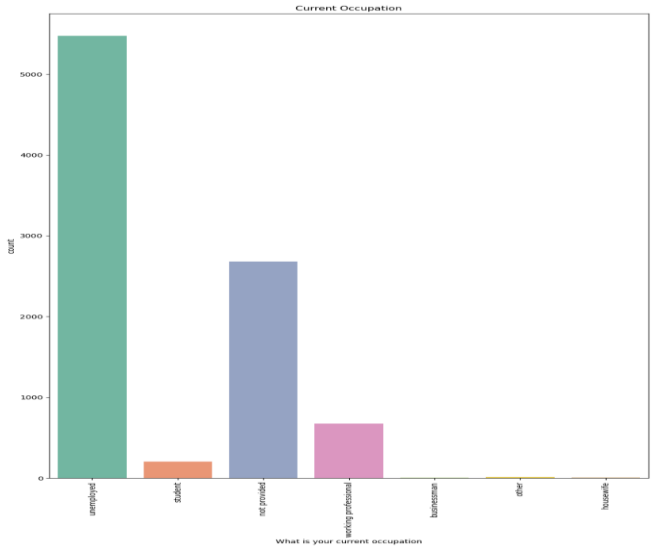
Lead source VS Count:

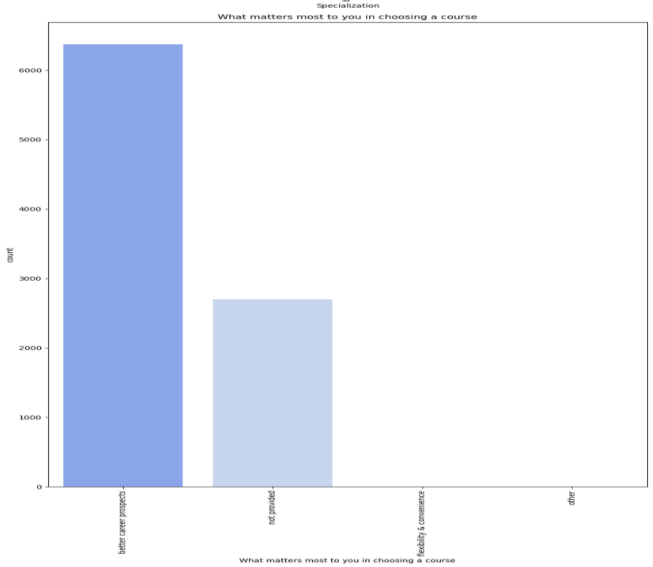


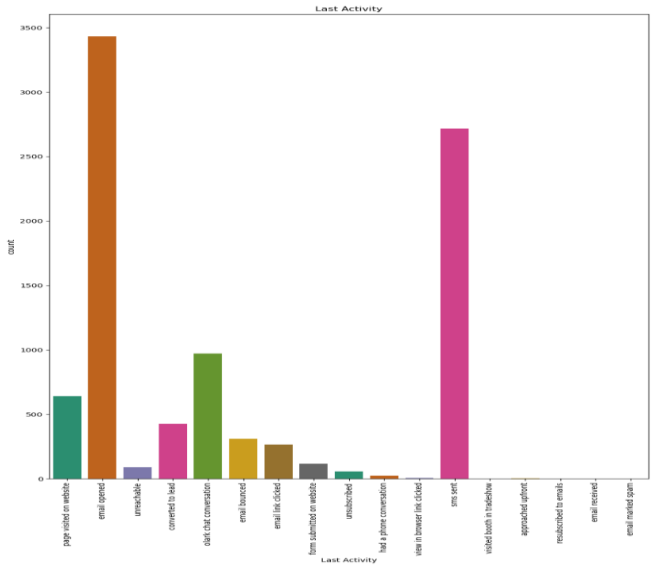
EDA – Univariate Analysis

Potential Lead Sources VS Count:



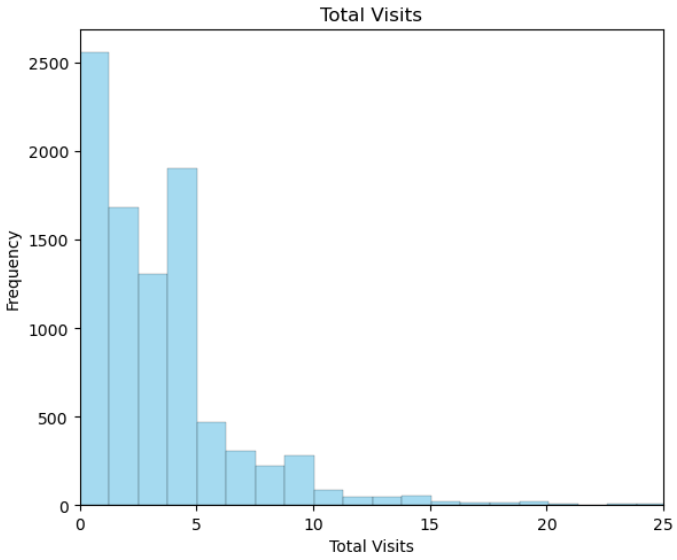


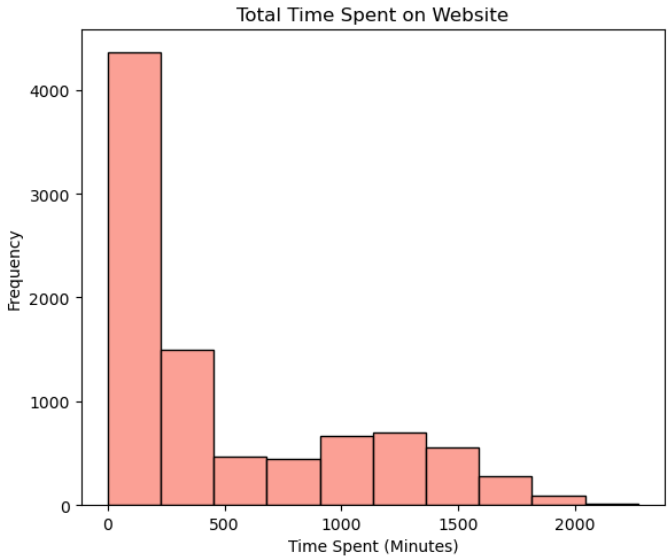


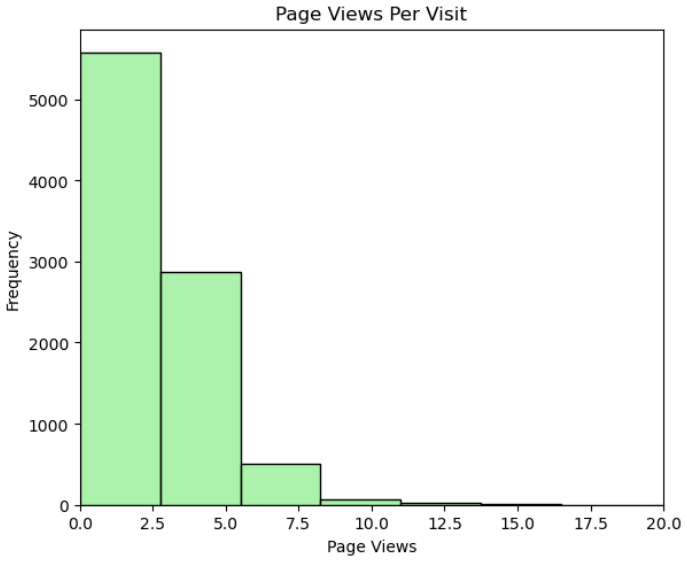


EDA – Univariate Analysis

The top three variables VS Frequency:

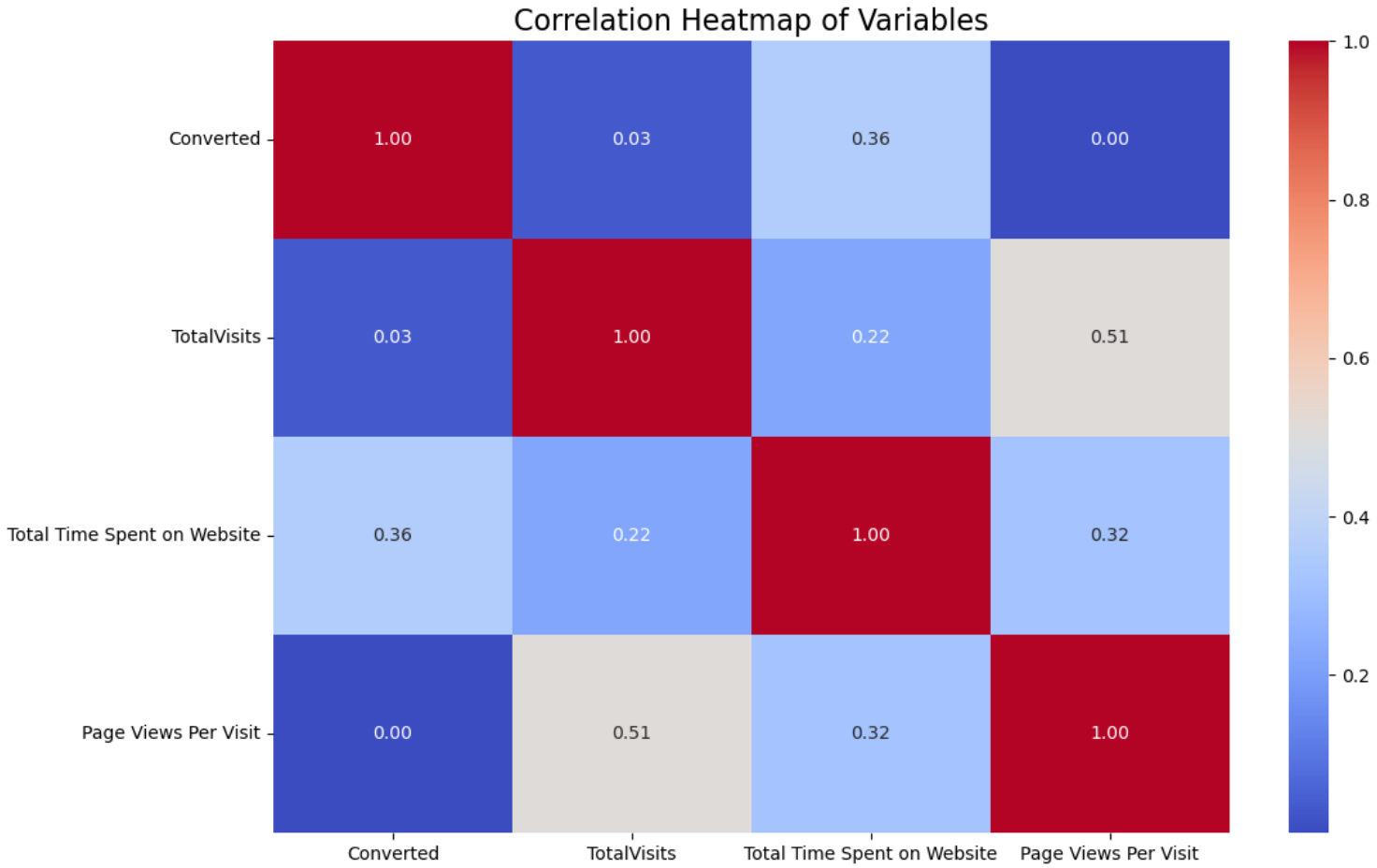






EDA – Bivariate Analysis

Correlation heatmap of the top three variables:



Model Building

Dividing Data into Training and Testing Sets: The initial step for regression involves splitting the data, using a 70:30 ratio for training and testing.

Feature Selection with RFE: Recursive Feature Elimination (RFE) is applied to select the top 15 important features for the model.

Model Building and Refinement: The model is built by excluding variables with a p-value greater than 0.05 and a Variance Inflation Factor (VIF) above 5 to improve efficiency.

Prediction on Test Data: Predictions are made on the test data to evaluate the model's performance.

Achieved Accuracy: The model achieved an overall accuracy of 81%, indicating acceptable performance.

Conclusion

- X Education needs to focus on Total Visits, Total Time Spent on Website, Page Views Per Visit as these are the top variables that contribute the most towards the probability of a lead getting converted.
- The top three lead sources are: Direct Traffic, Google & Organic Search. The sales team can look into potential customer who use the above three to get potential conversions.
- The sales team should also focus on working professionals as they have a high probability of getting converted. Avoiding those users whose last activity was a long time ago and those who spend the least amount of time on the websites can help the sales team focus more on those users that have a higher potential conversion.