Summary

X Education faces challenges with lead conversion rates, currently hovering at a modest 30%. To enhance this, a robust lead scoring model is proposed, aiming to prioritize leads with higher conversion potential. The CEO sets an ambitious target of 80% conversion rate, necessitating a strategic approach.

Data Cleaning:

- Data cleaning procedures were implemented meticulously. Columns with over 40% null values were dropped to maintain data integrity.
- Categorical columns underwent scrutiny, where imputation strategies were employed based on value counts.
- Imputation, creation of a new category labeled 'Others', or dropping columns altogether were considered based on skewness and value frequency.

EDA:

- EDA shed light on the data imbalance, with only 38.5% of leads converting.
- Univariate and bivariate analyses provided insights into categorical and numerical variables, revealing significant factors influencing lead conversion.
- Notably, lead origin, current occupation, and lead source emerged as pivotal variables. Time spent on the website exhibited a positive correlation with lead conversion.

Data Preparation:

- Data preparation entailed creating dummy features for categorical variables and splitting the dataset into a 70:30 ratio for train and test sets.
- Feature scaling via standardization enhanced model performance.
- Highly correlated columns were dropped to mitigate multicollinearity.

Model Building:

- Model building commenced with Recursive Feature Elimination (RFE), reducing variables from 48 to 15 for improved manageability.
- Manual feature reduction, guided by p-values, refined the model further.
- The final model, logm4, comprising 12 variables, exhibited robustness with p-values below 0.05 and VIF less than 5, indicating no multicollinearity issues.

Model Evaluation:

- Model evaluation culminated in the selection of a cutoff point of 0.345, balancing accuracy, sensitivity, and specificity.
- This decision, informed by sensitivity-specificity view, aligned with the CEO's objective of achieving an 80% conversion rate.
- Predictions on the test data yielded promising results, with evaluation metrics closely aligning with the desired thresholds.

Making Predictions on Test Data:

- Making Predictions on Test: Scaling and predicting using final model.
- Evaluation metrics for train & test are very close to around 80%.
- Lead score was assigned.
- Top 3 features are:
    - Lead Source_Welingak Website
    - Lead Source_Reference
    - Current_occupation_Working Professional

Recommendations:

- More budget/spend can be done on Welingak Website in terms of advertising, etc.
- Incentives/discounts for providing reference that convert to lead, encourage to provide more references.
- Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.
