Lead Conversion Optimization for X Education

Objective:
To improve the lead conversion rate for X Education, an online course provider for industry professionals, by building a predictive model that helps the sales team prioritize leads with a higher probability of conversion.

Process:

1. Data Preparation: Cleaned the dataset by handling missing values, removing duplicates, and fixing inconsistencies.
2. Feature Selection: Analyzed the importance of features and selected the most relevant ones using techniques like Recursive Feature Elimination with Cross-Validation (RFECV) and SelectKBest with chi-squared test.
3. Exploratory Data Analysis (EDA): Examined the distribution of variables and their relationship with the target variable.
4. Model Building: Trained a Logistic Regression model on the selected features, using GridSearchCV for hyperparameter tuning.
5. Performance Evaluation: Assessed the model's performance using accuracy, precision, recall, F1-score, ROC AUC score, and the precision-recall curve.

Key Findings:

1. Most leads originate from the landing page submission, making it crucial to maintain an engaging and informative website.
2. Direct traffic, Google search, and organic search are the most common sources of leads, suggesting the importance of search engine optimization.
3. Finance, HR, and Marketing are the most popular specializations, indicating a potential focus area for course development.
4. API and Lead Add Form have high conversion rates, implying their effectiveness in lead generation.
5. SMS has a high conversion rate (61%), emphasizing its importance as a communication channel.
6. Increased time spent on the website is associated with higher conversion rates, reflecting the user's interest and engagement.

Model Performance:

- Accuracy: 0.779
- Precision: 0.770
- Recall: 0.710
- F1-score: 0.739
- ROC AUC score: 0.772
- Average Precision (AP) score: 0.83
-

Conclusion:

We successfully built a predictive model to help X Education prioritize leads with a higher probability of conversion. The model achieved reasonable performance considering the reduced feature set. By implementing this model, X Education can improve its lead conversion rate and grow its customer base. Regular evaluation and refinement of the model will ensure its ongoing effectiveness.