CREATED BY
SAI SHRUTHI SRIDHAR
saishruthi21@gmail.com

LEAD CASE STUDY

INTRODUCTION

Welcome to the presentation on lead scoring for enhanced conversion at X Education. The primary objective of this analysis is to build a logistic regression model to assign lead scores, aiding the identification of potential customers with higher conversion probabilities.

DATA EXPLORATION:

- Dataset comprises around 9,240 data points and 37 features.
- Features include lead origin, lead source, total time spent on website, etc.
- Challenges encountered: missing values in lead source, total visits, last activity.
- Categorical variables contain 'Select' level, treated as null values.
- These observations will inform data preprocessing and model building efforts.

DATA PREPROCESSING

Following Data pre-processing steps are taken-

1. **Handling Missing Values**
2. **Encoding Categorical Variables**
3. **Feature Engineering**
4. **Data Balancing**

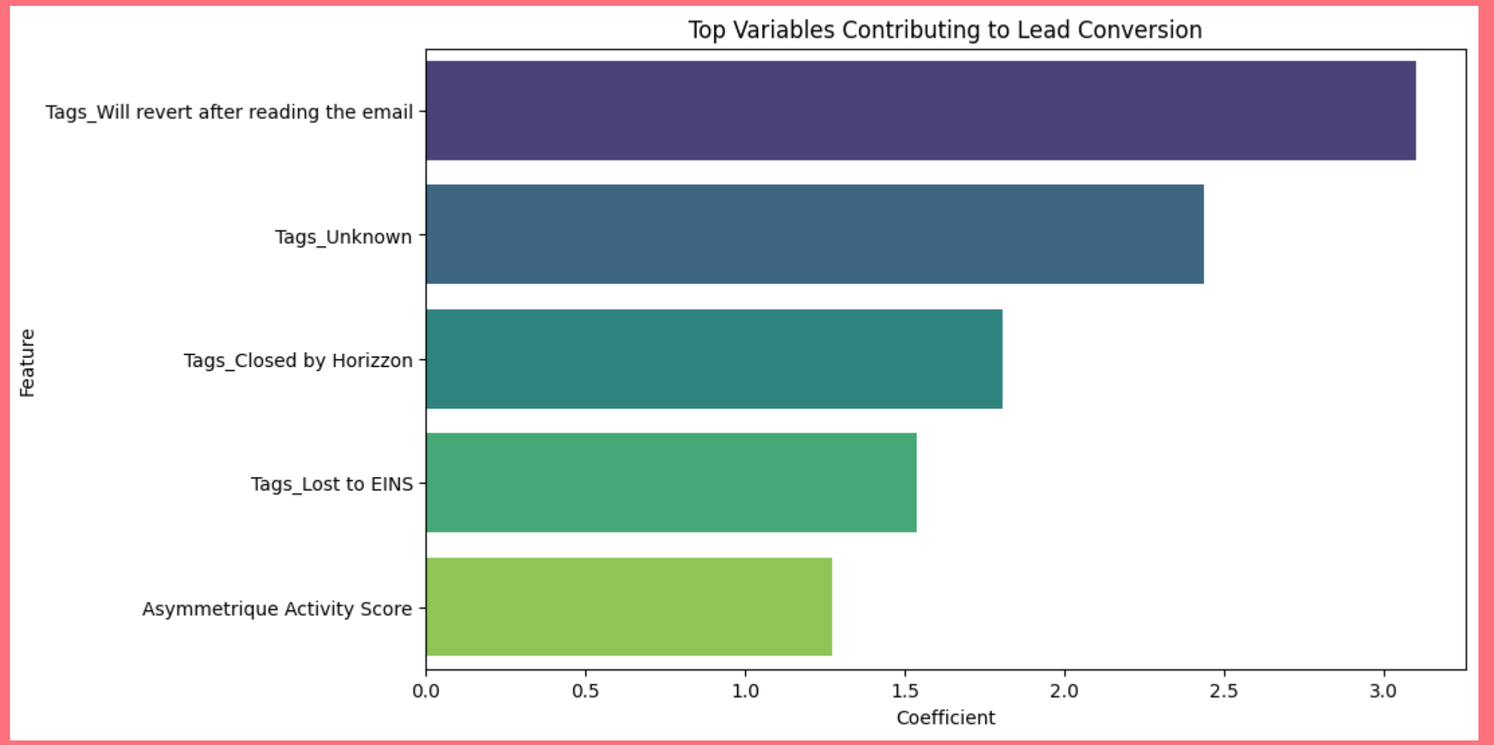
MODEL BUILDING

1. Splitting the Data
   - Divide the dataset into training and testing sets.
   - Training set: Used to train the model.
   - Testing set: Used to evaluate the model's performance on unseen data.
2. Building a Logistic Regression Model
   - Utilize logistic regression due to its effectiveness in binary classification tasks.
   - Logistic regression predicts the probability of a binary outcome based on input features.
3. Hyperparameter Tuning using GridSearchCV
   - Employ GridSearchCV to systematically search for the best combination of hyperparameters.
   - Hyperparameters are parameters whose values are set before the learning process begins.

RESULTS AND INSIGHTS

Recommendations based on logistic regression analysis:

- Focus on optimizing **Tags_Will revert after reading the email, Tags_Unknown, Tags_Closed by Horizzon, Tags_Lost to EINS, Asymmetrique Activity Score** to maximize lead conversion.
- Implement targeted strategies based on the insights gained from the analysis.



THANK YOU