# Lead Scoring-Summary

**Objective:**

The lead scoring dataset serves as a valuable resource for assessing the quality and potential conversion of leads generated by a company. By analysing lead attributes and behaviours, the dataset enables the prioritization of sales and marketing efforts to maximize conversion rates and optimize resource allocation.

**Key Attributes:**

The lead scoring dataset comprises a wide range of attributes These attributes include demographic information such as age, gender, and location, lead source information indicating how leads were acquired (e.g., website, referral, advertising), engagement metrics such as email opens, website behaviour such as pages visited and time spent, and historical data encompassing previous purchases and interactions.

**Data Preparation:**

To ensure accurate analysis, the dataset undergoes several data preparation steps. This includes data cleaning to handle missing values and outliers, pre-processing to transform and normalize variables, and feature engineering to derive new features that capture valuable information for lead scoring.

**Predictive Modelling:**

Logistic regression is the chosen machine learning algorithm. It analyses the relationships between lead attributes and conversion outcomes to generate lead scores or probabilities of conversion. These scores help prioritize leads based on their conversion potential.

**Evaluation Metrics:**

The performance of lead scoring models is evaluated using a range of metrics to assess their effectiveness and predictive capabilities. Common evaluation metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC).

**Applications:**

The lead scoring dataset finds application in various sales and marketing strategies aimed at maximizing conversion rates and optimizing resource allocation. The lead score enables companies to:

Prioritize leads: Companies can identify leads with a higher likelihood of conversion and focus their sales efforts on these leads to improve conversion rates.

Personalize marketing campaigns: Segmenting leads based on lead scores allows companies to tailor marketing messages and offers to specific segments, increasing the effectiveness of marketing campaigns and improving lead engagement.

Resource optimization: By focusing on leads with higher conversion potential, companies can allocate their resources more efficiently, reducing costs and improving return on investment (ROI).

**Continuous Improvement:**

The lead scoring dataset requires regular monitoring and updates to maintain its accuracy and relevance. As new data becomes available, the dataset should be retrained or fine-tuned. Continuous evaluation and iteration are essential to ensure the lead scoring system remains effective and aligned with market dynamics.

**Recommendations:**

1. Lead Score: Leads with a lead score of 35 or higher can be treated as "hot leads" for achieving a target lead conversion rate of 80%.

2. Critical Fields for Prediction: Special attention should be given to reducing missing values in the following fields:

- Tags
- Lead Source
- Last Activity
- Last Notable Activity
- Occupation

3. Adjusting the Cut off Score: For higher efficiency, cutoff score should be increased beyond 35. This will focus the resources on most promising leads.

4. Prioritizing High Lead Conversion: In cases where maximizing lead conversion is of utmost importance, even at the expense of resource costs, it is recommended to lower the cutoff score below 35.