Detailed Summary: Lead Scoring Case Study Using Logistic Regression

1. Objective

- The goal of this study is to identify the key factors influencing lead conversion for an online professional training course.
- Insights from this analysis will help in improving marketing strategies and increasing enrollment rates.

2. Data Preparation and Cleaning

- Missing values were addressed, and unnecessary columns were removed.
- Categorical variables were converted into numerical form using one-hot encoding.
- Some fields, such as **Customer Occupation** and **Specialization**, contained a "Select" option, indicating missing selections. Mandatory selection can improve data quality.
- Outliers were identified and treated to prevent them from affecting model predictions.

3. Exploratory Data Analysis (EDA)

- The majority of leads are from **India**, with **Mumbai** having the highest number of potential customers.
- Leads who spent more time on the platform and had multiple visits showed a higher likelihood of conversion.
- The primary motivation for enrolling in courses is **Better Career Prospects**.
- Specializations in **Finance, HR, and Marketing Management** have a higher chance of conversion.
- A significant portion of leads are **unemployed**, making them a key focus group for marketing efforts.

4. Feature Engineering

- Additional features were created to measure user engagement, such as **visit-to-conversion ratios** and **interaction-based features** (e.g., email and SMS response rates).
- New insights were derived from existing data to improve model performance.

5. Model Building & Selection

- A **Logistic Regression Model** was built to predict lead conversion.
- Two models were compared:
    - **Model 1:** Included all features.
    - **Model 2:** Used a refined feature set selected using **Recursive Feature Elimination (RFE)**.
- RFE helped in identifying the most important predictors, removing less relevant features to enhance model efficiency.

6. Model Evaluation & Key Metrics

- **Accuracy, Precision, Recall, and F1-score** were used to evaluate model performance.
- Findings from the model:
    - **Higher engagement on the website** (more visits and time spent) increased conversion probability.
    - **Specialization and Occupation** significantly impacted conversion, with Finance, HR, and Marketing showing higher conversion rates.
    - **Interaction with emails and SMS** improved conversion likelihood.
    - **Leads from Mumbai and Maharashtra** had a higher probability of conversion.
- The **RFE-optimized model** performed better by eliminating unnecessary features, improving both accuracy and interpretability.

7. Business Insights & Recommendations

- **Improve website engagement** by encouraging repeated visits and increasing time spent on the platform.
- **Focus marketing efforts on unemployed leads**, as they have a higher likelihood of conversion.
- **Enhance customer engagement through emails and SMS**, as interactions with these channels positively impact conversions.
- **Ensure complete data collection** by making certain fields mandatory, such as specialization and occupation.
- **Target leads from Mumbai and Maharashtra** more effectively, as they show a higher conversion rate.

8. Data Visualization Techniques Used

- **Histograms & Boxplots:** Used to understand data distribution and identify outliers.
- **Heatmaps:** Helped in finding correlations between different variables.

- **Feature Importance Charts:** Provided insights into the key drivers of lead conversion.

9. Conclusion

- This study successfully identified **key factors influencing lead conversion**.
- By focusing on **website engagement, relevant user segments, and targeted marketing**, enrollment rates can be improved.
- Future improvements can include testing additional machine learning models (e.g., Decision Trees, Random Forest) for better predictive accuracy.