Summary

Solution will be provided based on a Logistic Regression model which the below steps will be followed

- Data cleaning i.e., treat missing values by the following ways
    - Replace select value as null value.
    - Removing columns where there is more than 30% missing values
    - Filling the missing values with mode values where there less than 30% missing values
    - Removing the rows where the missing percentage is minimal.
- Outlier treatment
    - For continuous values, check if there are outliers and if ant remove the ones greater than 0.99 percentile and less than 0.01 percentile.
- EDA
    - Performance exploratory data analysis to get insights from the data
    - Also, understand which columns which useless for model building i.e. data imbalance etc
    - Analyze where there are multiple levels out of which some of them have minimal data belonging to it .
- Preprocessing for model building
    - Drop the columns which are unnecessary and will not contribute towards Model building for examples like Country where all the prospects are mostly based in India and hence there is no relation to lead conversion.
    - Club the unique values/levels of categorical variables where there is minimal data into Others category. This will reduce the number of dummy variables needed.
    - Create dummy variables for the categorical variables.
    - Split the data into train and test data.
    - Scale the continuous variables.
    - Proceed with model building.
- Model Building – Logistic Regresssion
    - Use RFE to select the 5 features
    - Fit the model
    - Drop the variables with high p-value and fit the model again.
    - Check the VIF values, drop the variables with high VIF (in this case it was not unnecessary as all VIFs were low enough)

- Model Evaluation
    - Check the accuracy, sensitivity and specificity with different probability cut off to find an optimal value of the probability cutoff using the ROC curve
    - Using an optimal cut off evaluate the model on the train data and check the accuracy, sensitivity and specificity
    - Repeat the above step with test data and check if the metrics are under acceptable limits.
    - Check the model with other metrics like Positive predictive value , negative predictive value, precision and recall.