Data Cleaning:

● "SELECT" level represents null values for some categorical variables, as customers did not choose any option from the list.
● There are few columns with over 40 % null values were dropped
● We handled the missing values for categorical columns were based on value Counts
● We have dropped certain columns from our analysis which do not add insights useful for analysis i.e., City, tags, Country, Specialization, Newspaper.
● Columns with no use for modelling (Prospect ID, lead Number) can we dropped.
● Other Cleaning activities were performed to ensure data quality and accuracy.
● Skewed category columns were checked and dropped for example ""Other Cleaning activities were performed to ensure data quality and accuracy.

EDA:

● Univariate and bivariate analysis for categorical and numerical variables.
● Time spend on website, Total Visits shows positive impact on lead conversion.
● Data was imbalanced while analysing the target variable. We cleaned the data and dropped unnecessary columns from analysis

Data Preparation:

● Created dummy features for categorical variables- Lead Origin, Lead Source, Last Activity, Current Occupation
● Splitting the dataset into 70% Train & 30 % test
● Some Categorical columns i.e. "Do not Email" were mapped to 1/0
● There are few numeric variables present in the dataset which have different Scales
● Variables with positive coefficient were used for Predictive Analytics
● We have approximately 68% columns after data cleaning

Model Building:

● We have used RFE to reduce variables to 15. This will make data frame more manageable.
● Variables with VIF values greater than 0.5 are neglected.
● We have built 4 models were built before reaching final Model 5 which was stable with (p-values < 0.05).
● Final model with 11 variables, we used it for making prediction on train and test set.

Model Evaluation:

● Confusion matrix was made. This cut off gave accuracy, specificity and precision approximately 80%.
● Lead score was assigned to train data using 0.345 as cut off.

● Top 3 features are:

1) Total Visits
2) Total Time Spent on Website
3) Lead Origin_Lead Add Form

Recommendations:

1. Focus on features with positive coefficients for marketing strategies.
2. Develop strategies to attract leads from top-performing lead sources.
3. Sales team should engage working professionals with SMS,messaging and Emails.
4. Leads who have spent time on sending messages and opening the emails are also potential hot leads and hence, effective Communication channels should be used.
5. Retention rate of existing leads needs to be identified from the time of enrolment.
6. Social Media can be used as a potential source for understanding and interacting. It can be effectively used for marketing purpose as well by encouraging people to take action.
7. Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.
8.   Existing customers feedback should be taken into consideration to improve the conversion rate.
9.   Discounts can be one way of the important measure to motivate them
10. The company should make calls to the leads coming from the lead sources "Welingak Websites", "What is your current occupation_Working Professional" and "Last Activity_Had a Phone Conversation" as these are more likely to get converted.
11. Focus on leads with high potential : Based on our analysis coefficient factor was high for following :
    A) Total Visits
    B) Total Time Spent on Website
    C) Lead Origin_Lead Add Form
    D)What is your current occupation_Working Professional
    E)Lead Origin_Lead Add Form
    F)Lead Source_Welingak Website
    G)Last Activity_Had a Phone Conversation