Summary

In order to convert potential consumers, business X Education is constructing models and making predictions. In order to target the right audience and boost conversion rates, we will further analyze and confirm the data. Let's go over the subsequent steps:

1. Initializing and Understanding the data

Importing csv file and checking different parameters like datatypes, shape, info, etc.

2. EDA

Null value checking, dropping the variables above 35 % null value, Null value treatment importing one new category called as 'NA' for categoric variables and dropping rows for Numeric variables.

3. Univariate and Bivariate Analysis

- 'Landing Page Submission' as Lead Origin are converted more than others.
- Among all the variables, 'Google' as Lead Source has the highest conversion rate.
- The variable 'SMS Sent' as last activity has shown a positive conversion rate and is also the highest.
- The target staying in 'Mumbai' as city have the highest conversion rate.
- Target in 'India' as Country are more likely to get converted.
- The 'Unemployed' as what is your occupation are more likely to get converted.

4. Data Preparation

- Mapping of YES/NO variables to 1/0
- Dummy Variables incorporation

5. Test-Train Split

Importing train_test_split from sklearn.model_selection using Train_size=0.7, Test_size=0.3, Random_state=100

6. Feature Scalling

- Importing StandardScaler from scaler.fit_transform
- Dropping Few dummy variables as they are corelated to other variables.

7. Model Building

- Importing LogisticRegression from sklearn.linear_model
- Imporing RFE from sklearn.feature_selection

- Dropping the columns with high p-value
- Calculating VIF (All VIF is less than 5. So we need not drop any more variables and we can proceed with making predictions using this model.)

8. Prediction

Prediction using the Model developed

9. Model Evaluation

- The model seems to be performing well.
- The ROC curve has a value of 0.84, which is very good.
- Accuracy : 77.12%
- Sensitivity :82.51%
- Specificity : 72.57%
- True Postive Rate : 82.51%
- False Postive Rate : 27.42%
- Precision : 64.80%
- Recall : 82.51%

10. Predictions on the test dataset

- Accuracy : 76.82%
- Sensitivity :82.24%
- Specificity : 73.47%
- Precision : 65.67%
- Recall : 82.24%
- True Postive Rate : 82.24%
- False Postive Rate : 26.52%

11. Conclusion

- Lead Source_Welingak Website
- What is your current occupation_Working Professional

The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.