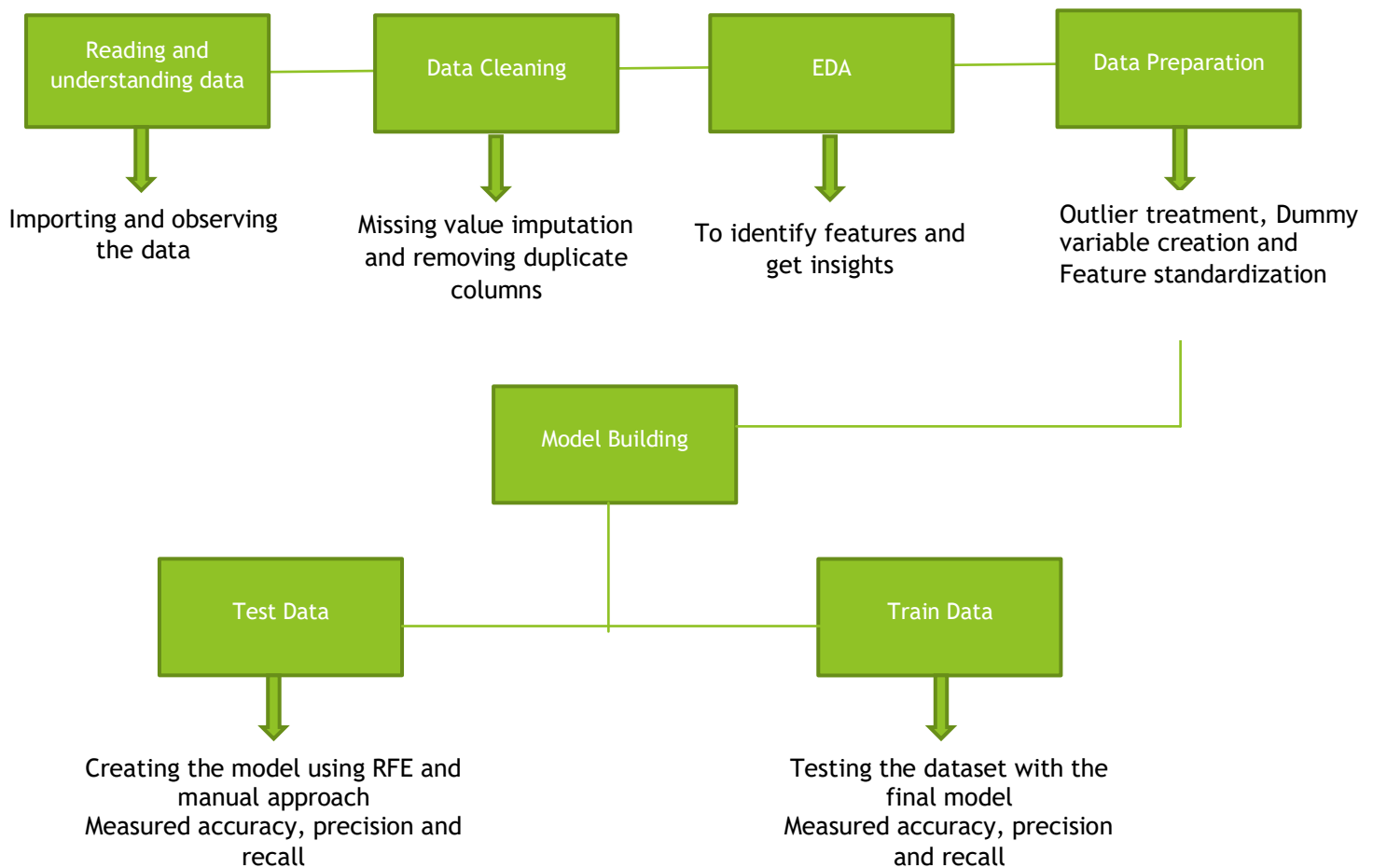# LEAD SCORING SUMMARY

PROBLEM STATEMENT

To help X Education identify the potential lead for high conversion rate and focus more on communication with the potential leads rather than making calls to every customer.

ANALYSIS APPROACH

To build a Logistic Regression model that assigns lead scores to all leads such that the customers with higher lead score have a higher conversion chance and lower lead score have low conversion rate.

Flowchart of data analysis:

Reading and understanding data	Data Cleaning	EDA	Data Preparation

Importing and observing the data	Missing value imputation and removing duplicate columns	To identify features and get insights	Outlier treatment, Dummy variable creation and Feature standardization

Model Building

Test Data	Train Data

Creating the model using RFE and manual approach Measured accuracy, precision and recall	Testing the dataset with the final model Measured accuracy, precision and recall

CONCLUSION

The model created has good accuracy.

Through the model we obtained top three variables which contribute most towards the probability of a lead conversion and they are:

1. Lead Origin   2. Total Time Spent on Website   3. Last Activity