1. **Steps for Logistic regression**
2. Import data
3. EDA (Exploratory Data Analysis)
4. Univariate (Summary, check Outlier & Missing value)
5. View, Head functions
6. Data Pre-processing
7. Missing Value Treatment
8. If required create dependent variable (Y)
9. Take a subset if required.
10. Data conversion (Dependent variable should be categorical and independent can be any type.)
11. Outlier Treatment
12. Check Bad rate:

* If bad rate between 3% to 30% move next.
* If bad rate is less than 3% , data is underfit (biased) , under sampling
* If bad rate is more than 30% , data is overfit (biased), over sampling
* Again check bad rate.

1. Again EDA
2. Data Partition
3. Training Data (should be 70%)
4. Testing Data (Should be 30%)
5. Check Bad rate on training data
6. Bad rate should be between 3% to 26% or 30%.
7. If bad rate is less than 3% or more than 30% : Data sampling method (Random, cluster, systematic, stratified) for data partition.
8. Again check bad rate.
9. Create model
10. Check P value should be less than α (0.05). If there is more than 1 level in categorical data(variable) and p value of any of them is less than α , means the p value for that variable is less than α.
11. Check AIC value.
12. Will get model coefficient (β0, β1)
13. Apply model (β0, β1) on training data and predict probability more than 70%.
14. Run confusion matrix and check:

* Accuracy should be more than 70%.
* Sensitivity & Specificity should be more than 70%.

1. Accuracy check ACC value

* If accuracy is more than 92 it means data is biased, undersampling or oversampling

1. Odds ratio for comparing the levels of categorical variable.
2. Create ROC/AUC curve for training data.
3. Value should be more than 70%.
4. Validation (apply model(β0, β1) on testing data
5. Predict probability more than 70%.
6. Run confusion matrix and check:

* Accuracy should be more than 70%.
* Sensitivity & Specificity should be more than 70%.

1. Create ROC/AUC curve testing data.
2. Value should be more than 70%.
3. Conclusion