

## Handle Missing Values in Categorical Variable?

1. Consider binary classification problem
2. Steps:-

A) **Delete the rows**- But the other records in the row can be useful specially in the case of less number of rows in the data

B) **Replace with the most frequent values**:- Replace the missing value with the most frequent one but it can create imbalanced data set within the categories not with respect to output variable

C) **Apply classifier algorithm to predict**:- Make blank value variable as output feature and include other independent variables including target as independent and train the data set and model predicts missing value as data and start filling all missing values with that. Eg: KNN

D) **Apply unsupervised ML**:- check other features other than missing value one and check which groups the data cluster and accordingly check according to the no. of categories in missing data feature, which groups this row data lies and assign that group value to the missing data. Eg:- Clustering