Categorial data(It is much more simpler in R than python )

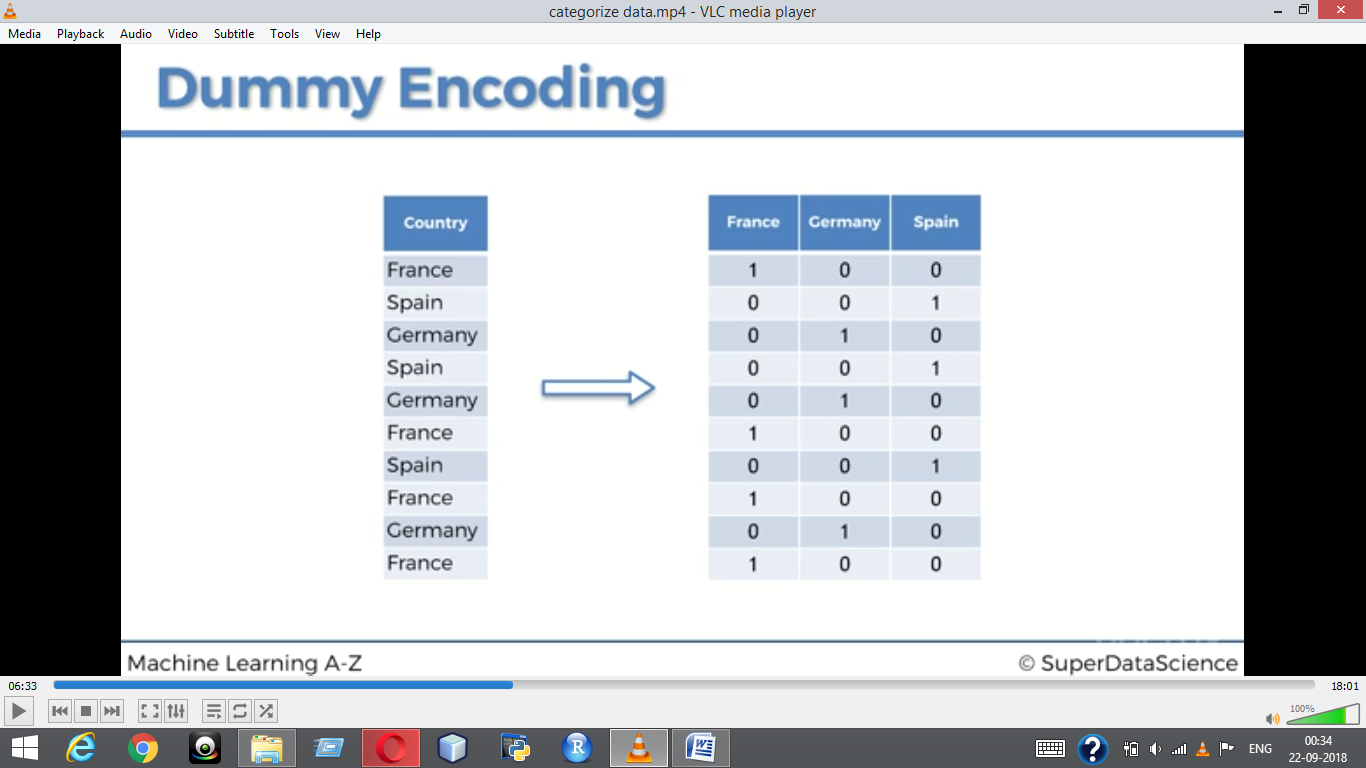
Method 1:

from sklearn.preprocessing import LabelEncoder

labelencoder\_X = LabelEncoder()

labelencoder\_X.fit\_transform(X[:,0])

But this method has a problem that we may have encoded numbers of string but .. let’s say we have france =0, germany=2 , spain =1 . We can’t put these in mathematical equations because it’ll assume that germany is greater than spain and spain is greater than france which is not the case. So, here comes the second method.



From sklearn.preprocessing import LabelEncoder, OneHotEncoder

Onehotencoder = OneHotEncoder(categorical\_features = [0])

X = onehotencoder.fit\_transform(X).toarray()