## Table of content

* What is Categorical Data?
* Label Encoding or Ordinal Encoding
* One hot Encoding
* Dummy Encoding
* Effect Encoding
* Binary Encoding
* BaseN Encoding
* Hash Encoding
* Target Encoding

## What is categorical data?

Categorical variables are usually represented as ‘strings’ or ‘categories’ and are finite in number. **Here are a few examples:**

1. The city where a person lives: Delhi, Mumbai, Ahmedabad, Bangalore, etc.
2. The department a person works in: Finance, Human resources, IT, Production.
3. The highest degree a person has: High school, Diploma, Bachelors, Masters, PhD.
4. The grades of a student: A+, A, B+, B, B- etc.

For encoding categorical data, we have a python package category\_encoders. The following code helps you install easily.

**pip install category\_encoders**

## 1. Label Encoding or Ordinal Encoding

We use this categorical data encoding technique when the categorical feature is ordinal. In this case, retaining the order is important. Hence encoding should reflect the sequence.

In Label encoding, each label is converted into an integer value. We will create a variable that contains the categories representing the education qualification of a person.

import category\_encoders as ce

import pandas as pd

train\_df=pd.DataFrame({'Degree':['High school','Masters','Diploma','Bachelors','Bachelors','Masters','Phd','High school','High school']})

# create object of Ordinalencoding

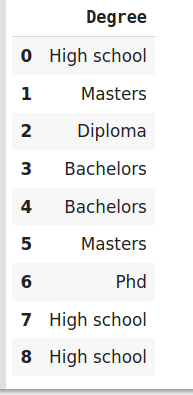
encoder= ce.OrdinalEncoder(cols=['Degree'],return\_df=True,

mapping=[{'col':'Degree',

'mapping':{'None':0,'High school':1,'Diploma':2,'Bachelors':3,'Masters':4,'phd':5}}])

#Original data

train\_df



#fit and transform train data

df\_train\_transformed = encoder.fit\_transform(train\_df)



# Types of Encoding

# 1. Nominal Encoding :- The categories do not have an inherent order. We have a feature, where we dont have to worry about the arrangement, So i will call as Norminal Categorial Variables.

# Example:- Gender, State

# Types of Nominal Categorial variables :-

# a. One Hot encoding

# b. One Hot encoding with many categorial variables

# c. Mean Encoding

# 2. Ordinal Encoding:- The categories have an inherent order. Ranking, Suppose we are having a dataset with the salary column based on the education.

# Education:- PHD[1], Master[2], BE[3] and Diploma[4]

# Types of Ordinal Categorial variables :-

# a. Label encoding

# b. Target Guided Ordinal Encoding

# <https://www.youtube.com/watch?v=OTPz5plKb40&list=PLZoTAELRMXVPwYGE2PXD3x0bfKnR0cJjN&index=3>