ONE HOT ENCODING

One-hot encoding is a method used to represent categorical data, where each category is represented by a binary variable. The binary variable takes the value 1 if the category is present and 0 otherwise. The binary variables are also known as dummy variables.

importing the required libraries
import numpy as np
import pandas as pd
from sklearn.preprocessing import OneHotEncoder

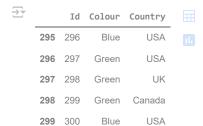


df=pd.read_csv("/content/data-one-hot-encoder.csv")
df.head()

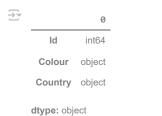


Next steps: Generate code with df View recommended plots New interactive sheet

df.tail()



df.dtypes



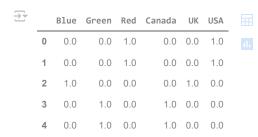
```
df["Colour"].unique()
⇒ array(['Red', 'Blue', 'Green'], dtype=object)
df["Country"].unique()
array(['USA', 'UK', 'Canada'], dtype=object)
ohe= OneHotEncoder()
ohe
\Rightarrow
     ▼ OneHotEncoder
     OneHotEncoder()
feature_array = ohe.fit_transform(df[["Colour","Country"]]).toarray()
feature_array
\Rightarrow array([[0., 0., 1., 0., 0., 1.],
           [0., 0., 1., 0., 0., 1.],
           [1., 0., 0., 0., 1., 0.],
           [0., 1., 0., 0., 1., 0.],
           [0., 1., 0., 1., 0., 0.],
           [1., 0., 0., 0., 0., 1.]]
ohe.categories_
array(['Canada', 'UK', 'USA'], dtype=object)]
feature_labels = ohe.categories_
np.array(feature_labels).ravel()
⇒ array(['Blue', 'Green', 'Red', 'Canada', 'UK', 'USA'], dtype=object)
feature_labels=np.array(feature_labels).ravel()
feature_labels
→ array(['Blue', 'Green', 'Red', 'Canada', 'UK', 'USA'], dtype=object)
pd.DataFrame(feature_array,columns=feature_labels)
```



300 rows × 6 columns

features = pd.DataFrame(feature_array,columns=feature_labels)

features.head()



Next steps: Generate code with features View recommended plots New interactive sheet

pd.concat([df,features],axis=1)



Start coding or generate with AI.