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
```

```
# load the dataset
df=pd.read_csv("/content/data-one-hot-encoder.csv")
df.head()
```

	Id	Colour	Country
0	1	Red	USA
1	2	Red	USA
2	3	Blue	UK
3	4	Green	Canada
4	5	Green	Canada

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

```
df.tail()
```

	Id	Colour	Country
295	296	Blue	USA
296	297	Green	USA
297	298	Green	UK
298	299	Green	Canada
299	300	Blue	USA

```
df.dtypes
```

	0
Id	int64
Colour	object
Country	object

dtype: object

```
df["Colour"].unique()
```

```
array(['Red', 'Blue', 'Green'], dtype=object)
```

```
df["Country"].unique()
```

```
array(['USA', 'UK', 'Canada'], dtype=object)
```

```
ohe= OneHotEncoder()
```

```
ohe
```

```
▾ OneHotEncoder
OneHotEncoder()
```

```
feature_array = ohe.fit_transform(df[["Colour","Country"]]).toarray()
```

```
feature_array
```

```
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(['Blue', 'Green', 'Red'], dtype=object),
 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)
```

	Blue	Green	Red	Canada	UK	USA
0	0.0	0.0	1.0	0.0	0.0	1.0
1	0.0	0.0	1.0	0.0	0.0	1.0
2	1.0	0.0	0.0	0.0	1.0	0.0
3	0.0	1.0	0.0	1.0	0.0	0.0
4	0.0	1.0	0.0	1.0	0.0	0.0
...	...	...	...	...	...	...
295	1.0	0.0	0.0	0.0	0.0	1.0
296	0.0	1.0	0.0	0.0	0.0	1.0
297	0.0	1.0	0.0	0.0	1.0	0.0
298	0.0	1.0	0.0	1.0	0.0	0.0
299	1.0	0.0	0.0	0.0	0.0	1.0

300 rows × 6 columns

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

```
features.head()
```

	Blue	Green	Red	Canada	UK	USA
0	0.0	0.0	1.0	0.0	0.0	1.0
1	0.0	0.0	1.0	0.0	0.0	1.0
2	1.0	0.0	0.0	0.0	1.0	0.0
3	0.0	1.0	0.0	1.0	0.0	0.0
4	0.0	1.0	0.0	1.0	0.0	0.0

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

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

```
   Id  Colour  Country  Blue  Green  Red  Canada   UK  USA
0   1     Red      USA   0.0    0.0  1.0     0.0  0.0  1.0
1   2     Red      USA   0.0    0.0  1.0     0.0  0.0  1.0
2   3    Blue       UK   1.0    0.0  0.0     0.0  1.0  0.0
3   4   Green   Canada   0.0    1.0  0.0     1.0  0.0  0.0
```

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

```
df_new.head()
```

```
   Id  Colour  Country  Blue  Green  Red  Canada   UK  USA
0   1     Red      USA   0.0    0.0  1.0     0.0  0.0  1.0
1   2     Red      USA   0.0    0.0  1.0     0.0  0.0  1.0
2   3    Blue       UK   1.0    0.0  0.0     0.0  1.0  0.0
3   4   Green   Canada   0.0    1.0  0.0     1.0  0.0  0.0
4   5   Green   Canada   0.0    1.0  0.0     1.0  0.0  0.0
```

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

Start coding or generate with AI.