

```
from google.colab import drive
drive.mount("/content/drive")
```

Mounted at /content/drive

```
%cd /content/drive/MyDrive/ML LAB/Week 1
```

/content/drive/MyDrive/ML LAB/Week 1

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
dataset = pd.read_csv('Data 5H7 .csv')
X = dataset.iloc[:, :-1].values
Y = dataset.iloc[:, -1].values
```

```
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit(X[:, 1:3])
X[:, 1:3] = imputer.transform(X[:, 1:3])
```

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
ct = ColumnTransformer(transformers = [['encoder', OneHotEncoder(), [0]]], remainder = 'passthrough')
X = np.array(ct.fit_transform(X))
```

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
Y = le.fit_transform(Y)
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=1)
```

```
print(X_train)
print(X_test)
print("")
print("#####")
print(Y_train)
print("")
print(Y_test)
```

```
[[0.0 0.0 1.0 38.77777777777778 52000.0]
 [0.0 1.0 0.0 40.0 63777.77777777778]
 [1.0 0.0 0.0 44.0 72000.0]
 [0.0 0.0 1.0 38.0 61000.0]
 [0.0 0.0 1.0 27.0 48000.0]
 [1.0 0.0 0.0 48.0 79000.0]
 [0.0 1.0 0.0 50.0 83000.0]
 [1.0 0.0 0.0 35.0 58000.0]]
[[0.0 1.0 0.0 30.0 54000.0]
 [1.0 0.0 0.0 37.0 67000.0]]
```

```
#####
[0 1 0 0 1 1 0 1]

[0 1]
```

```
filled_data = pd.DataFrame(X, columns=['Country_France', 'Country_Germany', 'Country_Spain', 'Age', 'Salary'])
filled_data['Purchased'] = dataset['Purchased']
filled_data.to_csv('Filled_Data.csv', index=False)
```