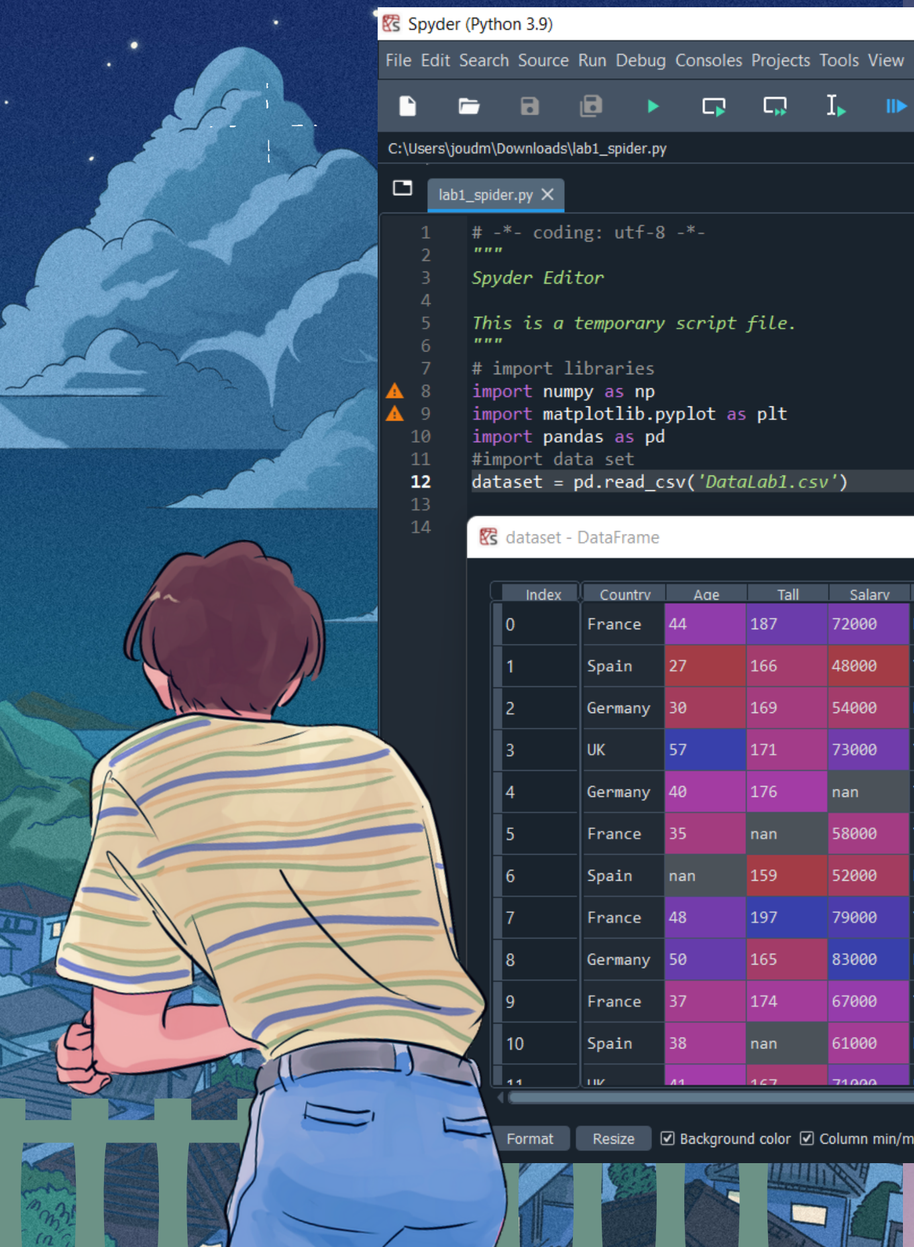
LAB 1

Applied Machine Learning

Student Name : JOUD ALAHMARI
Student ID: 2008071

1. IMPORT THE DATASET AND SHOW THE TABLE.



Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\joudm\Downloads

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
1  # -*- coding: utf-8 -*-
2  """
3  Spyder Editor
4
5  This is a temporary script file.
6  """
7  # import libraries
8  import numpy as np
9  import matplotlib.pyplot as plt
10 import pandas as pd
11 #import data set
12 dataset = pd.read_csv('DataLab1.csv')
13
14
```

dataset - DataFrame

Index	Country	Age	Tall	Salary	Purchased
0	France	44	187	72000	No
1	Spain	27	166	48000	Yes
2	Germany	30	169	54000	No
3	UK	57	171	73000	Yes
4	Germany	40	176	nan	Yes
5	France	35	nan	58000	Yes
6	Spain	nan	159	52000	No
7	France	48	197	79000	Yes
8	Germany	50	165	83000	No
9	France	37	174	67000	Yes
10	Spain	38	nan	61000	No
11	UK	41	167	71000	No

Format Resize Background color Column min/max Save and Close Close

Name	Type	Size	Value
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased

Help Variable Explorer Plots Files

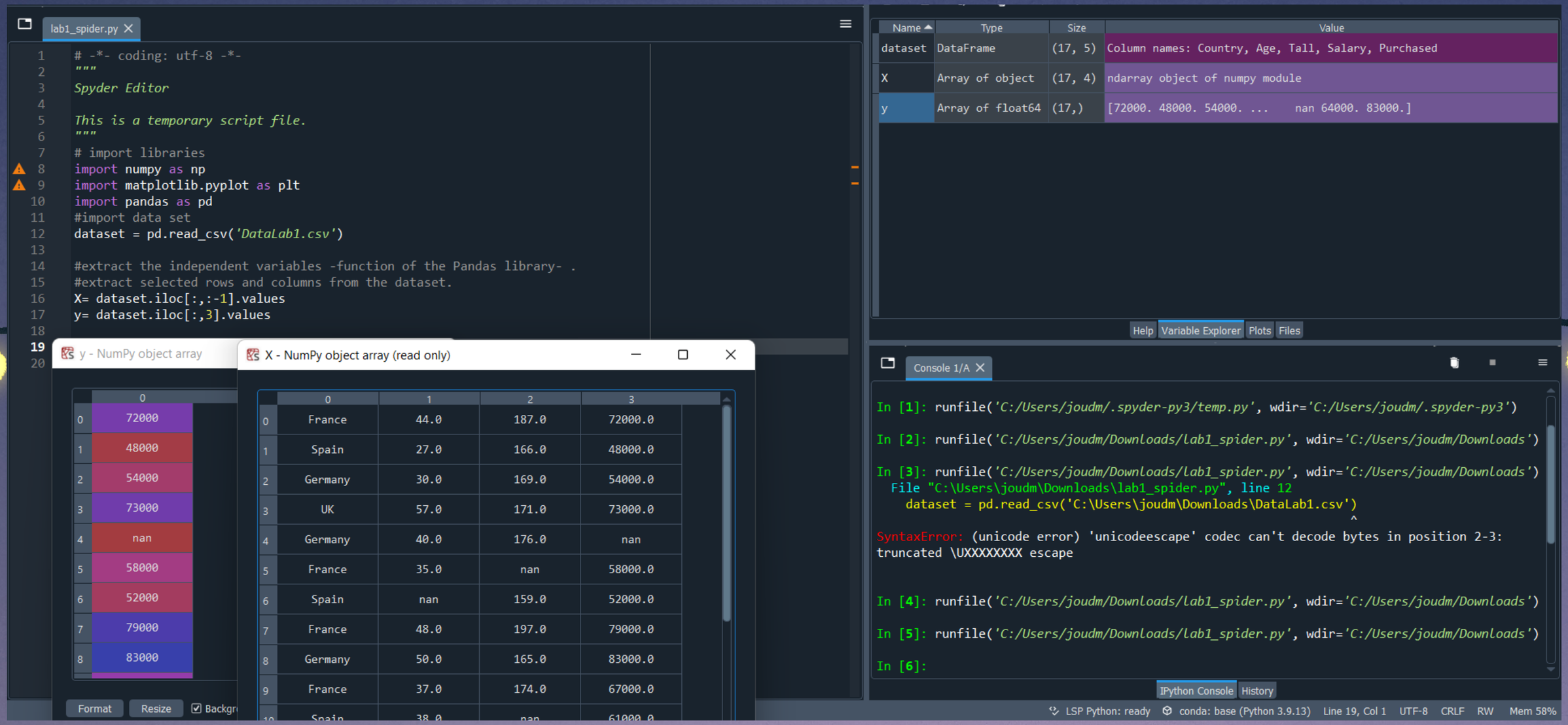
Console 1/A

```
IPython 7.31.1 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/joudm/.spyder-py3/temp.py', wdir='C:/Users/joudm/.spyder-py3')

In [2]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [3]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')
  File "C:\Users\joudm\Downloads\lab1_spider.py", line 12
    dataset = pd.read_csv('C:\Users\joudm\Downloads\DataLab1.csv')
                                                                  ^
SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in position 2-3:
truncated \UXXXXXXXX escape


In [4]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [5]:
```

IPython Console History

LSP Python: ready conda: base (Python 3.9.13) Line 12, Col 24 UTF-8 CRLF RW Mem 58%

2. SEPARATE THE INDEPENDENT AND DEPENDENT VARIABLES.



	0
0	72000
1	48000
2	54000
3	73000
4	nan
5	58000
6	52000
7	79000
8	83000

	0	1	2	3
0	France	44.0	187.0	72000.0
1	Spain	27.0	166.0	48000.0
2	Germany	30.0	169.0	54000.0
3	UK	57.0	171.0	73000.0
4	Germany	40.0	176.0	nan
5	France	35.0	nan	58000.0
6	Spain	nan	159.0	52000.0
7	France	48.0	197.0	79000.0
8	Germany	50.0	165.0	83000.0
9	France	37.0	174.0	67000.0

3. CALCULATE THE MEDIAN FOR EACH FEATURE OR COLUMN THAT CONTAINS A MISSING VALUE AND REPLACE THE RESULT FOR THE MISSING VALUE.

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
#2. Separate the Independent and dependent variables.
#extract the independent variables -function of the Pandas library- .
#extract selected rows and columns from the dataset.
X= dataset.iloc[:,:-1].values
y= dataset.iloc[:,3].values

#3. Calculate the median for each feature or column that contains a missing value and replace the result for the missing value.
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values= np.nan, strategy='median')
imputer = imputer.fit(X[:, 1:4])
X[:, 1:4] = imputer.transform(X[:, 1:4])
```

X - NumPy object array (read only)

	0	1	2	3
0	France	44.0	187.0	72000.0
1	Spain	27.0	166.0	48000.0
2	Germany	30.0	169.0	54000.0
3	UK	57.0	171.0	73000.0
4	Germany	40.0	176.0	71000.0
5	France	35.0	170.0	58000.0
6	Spain	39.0	159.0	52000.0
7	France	48.0	197.0	79000.0
8	Germany	50.0	165.0	83000.0
9	France	37.0	174.0	67000.0
10	Spain	38.0	170.0	61000.0
11	UK	41.0	167.0	71000.0

Format Resize Background color Close

Name	Type	Size	Value
ct	compose._column_transformer.ColumnTransformer	1	ColumnTransfo…
dataset	DataFrame	(17, 5)	Column names:…
imputer	impute._base.SimpleImputer	1	SimpleImputer…
X	Array of object	(17, 4)	ndarray objec…
y	Array of float64	(17,)	[72000. 48000…

Help Variable Explorer Plots Files

Console 1/A

```
In [4]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [5]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [6]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [7]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [8]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/Downloads')

In [9]:
```

IPython Console History

4.ENCODING THE CATEGORICAL DATA

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
15  #extract the independent variables -function of the Pandas library- .
16  #extract selected rows and columns from the dataset.
17  X= dataset.iloc[:,:-1].values
18  y= dataset.iloc[:,3].values
19
20  #3. Calculate the median for each feature or column that contains a missing value and
21  from sklearn.impute import SimpleImputer
22  imputer = SimpleImputer(missing_values= np.nan, strategy='median')
23  imputer = imputer.fit(X[:, 1:4])
24  X[:, 1:4] = imputer.transform(X[:, 1:4])
25
26
27  # 4.Encoding the categorical data
28  from sklearn.preprocessing import LabelEncoder, OneHotEncoder
29  from sklearn.compose import ColumnTransformer
30
31  ct = ColumnTransformer([('encoder', OneHotEncoder(), [0])], remainder ='passthrough')
32  X = np.array(ct.fit_transform(X), dtype= float)
33  y = LabelEncoder().fit_transform(y)
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
```

Name	Type	Size	Value
ct	compose._column_transformer.ColumnTransformer	1	ColumnTransformer object of sklearn.compose._column_transformer module
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased
imputer	impute._base.SimpleImputer	1	SimpleImputer object of sklearn.impute._base module
X	Array of float64	(17, 7)	[[1.00e+00 0.00e+00 0.00e+00 ... 4.40e+01 1.87e+02 7.20e+04] [0.00e+0 ...
y	Array of int64	(17,)	[8 0 2 ... 13 5 12]

X - NumPy object array

	0	1	2	3	4	5	6
0	1	0	0	0	44	187	72000
1	0	0	1	0	27	166	48000
2	0	1	0	0	30	169	54000
3	0	0	0	1	57	171	73000
4	0	1	0	0	40	176	71000
5	1	0	0	0	35	170	58000
6	0	0	1	0	39	159	52000
7	1	0	0	0	48	197	79000
8	0	1	0	0	50	165	83000
9	1	0	0	0	37	174	67000
10	0	0	1	0	38	170	61000

Format | Resize | Background color | Save and Close | Close

IPython Console | History

5. SPLITTING THE DATASET INTO THE TRAINING SET AND TEST SET AND ASSIGN 30% FOR TEST SET. (RANDOM_STATE = 0).

lab1_spider.py

```
32  X = np.array(ct.fit_transform(X), dtype= float)
33  y = LabelEncoder().fit_transform(y)
34
35  #5. Splitting the dataset into the Training set and Test Set and assign 30% for test set. (random_state = 0).
36  from sklearn.model_selection import train_test_split
37  X_train,X_test, y_train, y_test = train_test_split(X,y, test_size=0.3, random_state=0)
38
```

Name	Type	Size	Value
ct	compose._column_transformer.ColumnTransformer	1	ColumnTransformer object of sklear…
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased
imputer	impute._base.SimpleImputer	1	SimpleImputer object of sklearn.impute._base module
X	Array of float64	(17, 7)	[[1.00e+00 0.00e+00 0.00e+00 ... 4… [0.00e+0 ...
X_test	Array of float64	(6, 7)	[[0.00e+00 0.00e+00 1.00e+00 ... 2… [0.00e+0 ...
X_train	Array of float64	(11, 7)	[[0.00e+00 1.00e+00 0.00e+00 ... 3… [0.00e+0 ...
y	Array of int64	(17,)	[8 0 2 ... 13 5 12]
y_test	Array of int64	(6,)	[0 1 12 6 10 13]
y_train	Array of int64	(11,)	[2 13 4 ... 3 5 8]

X_train - NumPy object array

	0	1	2	3	4	5	6
0	0	1	0	0	30	169	54000
1	0	0	1	0	32	170	71000
2	0	0	1	0	38	170	61000
3	1	0	0	0	48	197	79000
4	1	0	0	0	56	168	83000
5	0	0	0	1	41	167	71000
6	0	0	0	1	57	171	73000
7	1	0	0	0	44	187	72000
8	1	0	0	0	35	170	58000
9	0	0	0	1	38	183	64000
10	0	0	0	1	35	178	72000

Format Resize Background color Save and Close Close

X_test - NumPy object array

	0	1	2	3	4	5	6
0	0	0	1	0	27	166	48000
1	0	0	1	0	39	159	52000
2	0	1	0	0	50	165	83000
3	1	0	0	0	37	174	67000
4	0	1	0	0	46	169	77000
5	0	1	0	0	40	176	71000

Format Resize Background color Save and Close Close

```
69
70
```

Downloads')

In [12]:

IPython Console History

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
32  X = np.array(ct.fit_transform(X), dtype= float)
33  y = LabelEncoder().fit_transform(y)
34
35  #5. Splitting the dataset into the Training set and Test Set and assign 30% for test set. (random_state = 0).
36  from sklearn.model_selection import train_test_split
37  X_train,X_test, y_train, y_test = train_test_split(X,y, test_size=0.3, random_state=0)
38
39
40
41
```

y_test - NumPy object array

	0
0	0
1	1
2	12
3	6
4	10
5	13

Format | Resize | Background color | Save and Close | Close

y_train - NumPy object array

	0
0	2
1	13
2	4
3	11
4	12
5	7
6	9
7	8
8	3
9	5
10	8

Format | Resize | Background color | Save and Close | Close

Name	Type	Size	Value
ct	compose._column_transformer.ColumnTransformer	1	ColumnTransformer object of sklear…
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased
imputer	impute._base.SimpleImputer	1	SimpleImputer object of sklearn.impute._base module
X	Array of float64	(17, 7)	[[1.00e+00 0.00e+00 0.00e+00 ... 4… [0.00e+0 ...
X_test	Array of float64	(6, 7)	[[0.00e+00 0.00e+00 1.00e+00 ... 2… [0.00e+0 ...
X_train	Array of float64	(11, 7)	[[0.00e+00 1.00e+00 0.00e+00 ... 3… [0.00e+0 ...
y	Array of int64	(17,)	[8 0 2 ... 13 5 12]
y_test	Array of int64	(6,)	[0 1 12 6 10 13]
y_train	Array of int64	(11,)	[2 13 4 ... 3 5 8]

Help | Variable Explorer | Plots | Files

Console 1/A

```
Downloads')

In [7]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [8]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [9]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [10]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [11]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [12]:
```

6. APPLY THE FEATURE SCALING

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
32  X = np.array(ct.fit_transform(X), dtype= float)
33  y = LabelEncoder().fit_transform(y)
34
35  #5. Splitting the dataset into the Training set and Test Set and assign 30% for test set. (random_state = 0).
36  from sklearn.model_selection import train_test_split
37  X_train,X_test, y_train, y_test = train_test_split(X,y, test_size=0.3, random_state=0)
38
39
40  #6.Feature Scaling
41  from sklearn.preprocessing import StandardScaler
42  sc_X = StandardScaler()
43  X_train = sc_X.fit_transform(X_train)
44  X_test = sc_X.fit_transform(X_test)
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
```

sc_X - Object

Name	Type	Size	Value
sc_X	preprocessing._data.StandardScaler	1	StandardScaler object of sklearn.preprocessing.
_check_feature_names	method	1	method object
_check_n_features	method	1	method object
_get_param_names	method	1	method object
_get_tags	method	1	method object
_more_tags	method	1	method object
_repr_html_inner	method	1	method object
_repr_mimebundle_	method	1	method object
_reset	method	1	method object
_validate_data	method	1	method object
copy	bool	1	True
fit	method	1	method object
fit_transform	method	1	method object

Details

Documentation / Source code / File

```
Standardize features by removing the mean and scaling to unit variance.

The standard score of a sample `x` is calculated as:

    z = (x - u) / s

```

Name	Type	Size	Value
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased
imputer	impute._base.SimpleImputer	1	SimpleImputer object of sklearn.impute._base module
sc_X	preprocessing._data.StandardScaler	1	StandardScaler object of sklearn…
X	Array of float64	(17, 7)	[[1.00e+00 0.00e+00 0.00e+00 ..… [0.00e+0 ...
X_test	Array of float64	(6, 7)	[[-0.4472136 -1. 1.414… -1.4 ...
X_train	Array of float64	(11, 7)	[[-0.75592895 3.16227766 -0.471… -1 ...
y	Array of int64	(17,)	[8 0 2 ... 13 5 12]
y_test	Array of int64	(6,)	[0 1 12 6 10 13]
y_train	Array of int64	(11,)	[2 13 4 ... 3 5 8]

Help | Variable Explorer | Plots | Files

Console 1/A

```
Downloads')

In [8]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [9]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [10]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [11]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [12]: runfile('C:/Users/joudm/Downloads/lab1_spider.py', wdir='C:/Users/joudm/
Downloads')

In [13]:
```

IPython Console | History

6. APPLY THE FEATURE SCALING

C:\Users\joudm\Downloads\lab1_spider.py

lab1_spider.py

```
32  X = np.array(ct.fit_transform(X), dtype= float)
33  y = LabelEncoder().fit_transform(y)
34
35  #5. Splitting the dataset into the Training set and Test Set and assign 30% for test set. (random_state = 0).
36  from sklearn.model_selection import train_test_split
37  X_train,X_test, y_train, y_test = train_test_split(X,y, test_size=0.3, random_state=0)
38
39
40  #6.Feature Scaling
41  from sklearn.preprocessing import StandardScaler
42  sc_X = StandardScaler()
43  X_train = sc_X.fit_transform(X_train)
44  X_test = sc_X.fit_transform(X_test)
45
```

Name	Type	Size	Value
dataset	DataFrame	(17, 5)	Column names: Country, Age, Tall, Salary, Purchased
imputer	impute._base.SimpleImputer	1	SimpleImputer object of sklearn.impute._base module
sc_X	preprocessing._data.StandardScaler	1	StandardScaler object of sklearn…
X	Array of float64	(17, 7)	[[1.00e+00 0.00e+00 0.00e+00 …… [0.00e+0 ...
X_test	Array of float64	(6, 7)	[[-0.4472136 -1. 1.414… -1.4 ...
X_train	Array of float64	(11, 7)	[[-0.75592895 3.16227766 -0.471… -1 ...
y	Array of int64	(17,)	[8 0 2 ... 13 5 12]

X_test - NumPy object array

	0	1	2	3	4	5	6
0	-0.447214	-1	1.41421	0	-1.77164	-0.380221	-1.45291
1	-0.447214	-1	1.41421	0	-0.115041	-1.60863	-1.13591
2	-0.447214	1	-0.707107	0	1.4035	-0.555708	1.32083
3	2.23607	-1	-0.707107	0	-0.391141	1.02367	0.0528332
4	-0.447214	1	-0.707107	0	0.851306	0.146239	0.845331
5	-0.447214	1	-0.707107	0	0.0230083	1.37465	0.369832

X_train - NumPy object array

	0	1	2	3	4	5	6
0	-0.755929	3.16228	-0.471405	-0.755929	-1.29759	-0.698702	-1.77862
1	-0.755929	-0.316228	2.12132	-0.755929	-1.06738	-0.590453	0.249441
2	-0.755929	-0.316228	2.12132	-0.755929	-0.376721	-0.590453	-0.943536
3	1.32288	-0.316228	-0.471405	-0.755929	0.77437	2.33229	1.20382
4	1.32288	-0.316228	-0.471405	-0.755929	1.69524	-0.806952	1.68101
5	-0.755929	-0.316228	-0.471405	1.32288	-0.0313934	-0.915202	0.249441
6	-0.755929	-0.316228	-0.471405	1.32288	1.81035	-0.482203	0.488036
7	1.32288	-0.316228	-0.471405	-0.755929	0.313934	1.24979	0.368738
8	1.32288	-0.316228	-0.471405	-0.755929	-0.722048	-0.590453	-1.30143
9	-0.755929	-0.316228	-0.471405	1.32288	-0.376721	0.816793	-0.585643
10	-0.755929	-0.316228	-0.471405	1.32288	-0.722048	0.275545	0.368738