practical-01

April 26, 2024

[1]:
```
import pandas as pd
```

[2]:
```
df = pd.read_csv("C:/Users/gugal/Desktop/THIRD 2/PRACTICALS/DS/CODES/DATASETS/
↪iris.csv")
df
```

[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
..	…	…	…	…	…
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

[150 rows x 5 columns]

[4]:
```
df.head(5)
```

[4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

[5]:
```
df.tail(5)
```

[5]:

	sepal_length	sepal_width	petal_length	petal_width	species
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica

```
149          5.9          3.0           5.1          1.8  virginica
```

[6]:
```
df.index
```

[6]:
```
RangeIndex(start=0, stop=150, step=1)
```

[7]:
```
df.columns
```

[7]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

[8]:
```
df.shape
```

[8]:
```
(150, 5)
```

[9]:
```
df.dtypes
```

[9]:
```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

[10]:
```
df.columns.values
```

[10]:
```
array(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'], dtype=object)
```

[11]:
```
df.describe(include='all' )
```

[11]:

	sepal_length	sepal_width	petal_length	petal_width	species
count	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	setosa
freq	NaN	NaN	NaN	NaN	50
mean	5.843333	3.054000	3.758667	1.198667	NaN
std	0.828066	0.433594	1.764420	0.763161	NaN
min	4.300000	2.000000	1.000000	0.100000	NaN
25%	5.100000	2.800000	1.600000	0.300000	NaN
50%	5.800000	3.000000	4.350000	1.300000	NaN
75%	6.400000	3.300000	5.100000	1.800000	NaN
max	7.900000	4.400000	6.900000	2.500000	NaN

[12]:
```
df['sepal_length']
```

[12]:
```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
      ...
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: sepal_length, Length: 150, dtype: float64
```

[13]:
```
df.sort_index(axis=1, ascending=False)
```

[13]:

	species	sepal_width	sepal_length	petal_width	petal_length
0	setosa	3.5	5.1	0.2	1.4
1	setosa	3.0	4.9	0.2	1.4
2	setosa	3.2	4.7	0.2	1.3
3	setosa	3.1	4.6	0.2	1.5
4	setosa	3.6	5.0	0.2	1.4
..	...	...	...	...	...
145	virginica	3.0	6.7	2.3	5.2
146	virginica	2.5	6.3	1.9	5.0
147	virginica	3.0	6.5	2.0	5.2
148	virginica	3.4	6.2	2.3	5.4
149	virginica	3.0	5.9	1.8	5.1

[150 rows x 5 columns]

[14]:
```
df.sort_values(by="sepal_width")
```

[14]:

	sepal_length	sepal_width	petal_length	petal_width	species
60	5.0	2.0	3.5	1.0	versicolor
62	6.0	2.2	4.0	1.0	versicolor
119	6.0	2.2	5.0	1.5	virginica
68	6.2	2.2	4.5	1.5	versicolor
41	4.5	2.3	1.3	0.3	setosa
..	...	...	...	...	...
16	5.4	3.9	1.3	0.4	setosa
14	5.8	4.0	1.2	0.2	setosa
32	5.2	4.1	1.5	0.1	setosa
33	5.5	4.2	1.4	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa

[150 rows x 5 columns]

[15]:
```
df.iloc[4]
```

[15]:
```
sepal_length       5.0
sepal_width        3.6
petal_length       1.4
petal_width        0.2
species         setosa
Name: 4, dtype: object
```

[16]:
```
df[0:3]
```

[16]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa

[17]:
```
df.loc[:, ["sepal_length",
"sepal_width"]]
```

[17]:

	sepal_length	sepal_width
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6
..	…	…
145	6.7	3.0
146	6.3	2.5
147	6.5	3.0
148	6.2	3.4
149	5.9	3.0

[150 rows x 2 columns]

[18]:
```
df.iloc[:4, :]
```

[18]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa

[19]:
```
df.iloc[:, :1]
```

[19]:

	sepal_length
0	5.1
1	4.9

```
2             4.7
3             4.6
4             5.0
..            …
145           6.7
146           6.3
147           6.5
148           6.2
149           5.9

[150 rows x 1 columns]
```

[20]:
```
df.iloc[:4, :4]
```

[20]:
```
   sepal_length  sepal_width  petal_length  petal_width
0           5.1          3.5           1.4          0.2
1           4.9          3.0           1.4          0.2
2           4.7          3.2           1.3          0.2
3           4.6          3.1           1.5          0.2
```

[21]:
```
df.isnull()
```

[21]:
```
     sepal_length  sepal_width  petal_length  petal_width  species
0           False        False         False        False    False
1           False        False         False        False    False
2           False        False         False        False    False
3           False        False         False        False    False
4           False        False         False        False    False
..            …            …             …            …        …
145         False        False         False        False    False
146         False        False         False        False    False
147         False        False         False        False    False
148         False        False         False        False    False
149         False        False         False        False    False

[150 rows x 5 columns]
```

[22]:
```
df.isnull().any()
```

[22]:
```
sepal_length    False
sepal_width     False
petal_length    False
petal_width     False
species         False
dtype: bool
```

[23]:
```
df.isnull().sum()
```

[23]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

[24]:
```
df.sepal_length.isnull().sum()
```

[24]:
```
0
```

[26]:
```
from sklearn import preprocessing
x = df[['sepal_length']].values.astype(float)
min_max_scaler = preprocessing.MinMaxScaler()
x_Scaled = min_max_scaler.fit_transform(x)
df_normalized = pd.DataFrame(x_Scaled)
df_normalized
```

[26]:
```
            0
0    0.222222
1    0.166667
2    0.111111
3    0.083333
4    0.194444
..        …
145  0.666667
146  0.555556
147  0.611111
148  0.527778
149  0.444444

[150 rows x 1 columns]
```

[27]:
```
# Label Encoding:
df['species'].unique()
le = preprocessing.LabelEncoder()
df['species'] = le.fit_transform(df['species'])
df
```

[27]:
```
     sepal_length  sepal_width  petal_length  petal_width  species
0             5.1          3.5           1.4          0.2        0
1             4.9          3.0           1.4          0.2        0
2             4.7          3.2           1.3          0.2        0
3             4.6          3.1           1.5          0.2        0
4             5.0          3.6           1.4          0.2        0
..            …            …             …            …          …
145           6.7          3.0           5.2          2.3        2
```

```
146          6.3          2.5          5.0          1.9        2
147          6.5          3.0          5.2          2.0        2
148          6.2          3.4          5.4          2.3        2
149          5.9          3.0          5.1          1.8        2

[150 rows x 5 columns]
```