AIML_Tutorial_2_pandas

December 30, 2023

0.1 # Tutorial 2 - Pandas

Name: Thanki Naman Rajeshbhai

Roll Number / Enrolment No: 40

Division: BTech - 6CEE

Subject: Artificial Intelligence and Machine Learning

1. Import NumPy and pandas library.

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

2. Define series with 3 elements.

[2]:
```
s1 = pd.Series([1, 2, 3])
print(s1)
```

```
0    1
1    2
2    3
dtype: int64
```

3. Define series with custom index.

[3]:
```
s2 = pd.Series([1,2,3], ['A', 'B', 'C'])
print(s2)
```

```
A    1
B    2
C    3
dtype: int64
```

4. Fetch Series value and index.

[4]:
```
s2.values
```

[4]:
```
array([1, 2, 3])
```

[5]:
```
s2.index
```

[5]:

```
Index(['A', 'B', 'C'], dtype='object')
```

5. Create DataFrame from list of Dict.

[9]:

```
data = [
    {'Name': 'Naman', 'Marks': 33},
    {'Name': 'Mohil', 'Marks': 40},
    {'Name': 'Manav', 'Marks': 101},
    {'Name': 'Rushi', 'Marks': 50},
    {'Name': 'Jay', 'Marks': 100}
]

f1 = pd.DataFrame(data)
f1
```

[9]:

```
    Name  Marks
0  Naman     33
1  Mohil     40
2  Manav    101
3  Rushi     50
4    Jay    100
```

6. Create dataframe with some missing value - NaN(Not a Number) represents missing values.

[10]:

```
f2 = pd.DataFrame([
    {'John': 20, 'Ram': 25, 'Arya': 50},
    {'Arya': 29, 'Alice': 24}
], index = ['Mathematics', 'Physics'])

f2
```

[10]:

```
             John   Ram  Arya  Alice
Mathematics  20.0  25.0    50    NaN
Physics       NaN   NaN    29   24.0
```

[11]:

```
f2.isna().sum()
```

[11]:

```
John     1
Ram      1
Arya     0
Alice    1
dtype: int64
```

7. Read excel file.

[12]:

```
xlsx_data = pd.read_excel('/content/drive/MyDrive/Machine Learning/temp.xlsx')
xlsx_data
```

```
[12]:    rollno    name
      0      20     jay
      1      24   mohil
      2      34   manav
      3      39  ashish
      4      40   naman
```

8. Explore kaggle.com for dataset. Search iris dataset and download.

9. Read iris dataset from csv file format and findout no. of rows and columns.

```
[13]: csv_data = pd.read_csv('/content/drive/MyDrive/Machine Learning/IRIS.csv')
      csv_data
```

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

      [150 rows x 5 columns]
```

10. Perform following operation on iris dataset.

```
1. Fetch firat 5 and last 5 rows.
2. Find index
3. remane columns
4. replace "Iris-versicolor" with "versicolor"
5. Find Statistical description
6. extract specific column
7. add new column with condition
8. drop added column
9. select first row data
10. select 51 to 55 rows
11. Apply conditional formatting
12. replace "Iris-versicolor" with "versicolor"
13. print "virginica" with "Petal_Length" more then 6
14. find duplicate values
15. Find basic information
```

10. 1 Fetch firat 5 and last 5 rows.

```
[14]: csv_data.head()
```

```
[14]:    sepal_length  sepal_width  petal_length  petal_width      species
      0           5.1          3.5           1.4          0.2  Iris-setosa
      1           4.9          3.0           1.4          0.2  Iris-setosa
      2           4.7          3.2           1.3          0.2  Iris-setosa
      3           4.6          3.1           1.5          0.2  Iris-setosa
      4           5.0          3.6           1.4          0.2  Iris-setosa
```

```
[15]: csv_data.tail()
```

```
[15]:      sepal_length  sepal_width  petal_length  petal_width         species
      145           6.7          3.0           5.2          2.3  Iris-virginica
      146           6.3          2.5           5.0          1.9  Iris-virginica
      147           6.5          3.0           5.2          2.0  Iris-virginica
      148           6.2          3.4           5.4          2.3  Iris-virginica
      149           5.9          3.0           5.1          1.8  Iris-virginica
```

10.2 Find index

```
[16]: csv_data.index
```

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

10.3 remane columns

```
[17]: csv_data = csv_data.rename(columns = {
          'sepal_length': 'Sepal_Length',
          'sepal_width': 'Sepal_Width',
          'petal_length': 'Petal_Length',
          'petal_width': 'Petal_Width',
          'species': 'Species'
      })

      csv_data
```

```
[17]:      Sepal_Length  Sepal_Width  Petal_Length  Petal_Width         Species
      0             5.1          3.5           1.4          0.2     Iris-setosa
      1             4.9          3.0           1.4          0.2     Iris-setosa
      2             4.7          3.2           1.3          0.2     Iris-setosa
      3             4.6          3.1           1.5          0.2     Iris-setosa
      4             5.0          3.6           1.4          0.2     Iris-setosa
      ..              …            …             …            …               …
      145           6.7          3.0           5.2          2.3  Iris-virginica
      146           6.3          2.5           5.0          1.9  Iris-virginica
      147           6.5          3.0           5.2          2.0  Iris-virginica
      148           6.2          3.4           5.4          2.3  Iris-virginica
      149           5.9          3.0           5.1          1.8  Iris-virginica

      [150 rows x 5 columns]
```

10.4 replace “Iris-versicolor” with “versicolor”

[18]:

```
csv_data.replace('Iris-versicolor', 'Versicolor', inplace=True)
csv_data.replace('Iris-setosa', 'Setosa', inplace=True)
csv_data.replace('Iris-virginica', 'Verginica', inplace=True)

csv_data
```

[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
4             5.0          3.6           1.4          0.2     Setosa
..            …            …             …            …          …
145           6.7          3.0           5.2          2.3  Verginica
146           6.3          2.5           5.0          1.9  Verginica
147           6.5          3.0           5.2          2.0  Verginica
148           6.2          3.4           5.4          2.3  Verginica
149           5.9          3.0           5.1          1.8  Verginica

[150 rows x 5 columns]
```

10.5 Find Statistical description

[19]:

```
csv_data.describe()
```

[19]:

```
       Sepal_Length  Sepal_Width  Petal_Length  Petal_Width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000
```

10.6 extract specific column

[20]:

```
csv_data[['Sepal_Length', 'Sepal_Width']].head()
```

[20]:

```
   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
```

10.7 add new column with condition

[21]:

```
def sepal_length_requirement(sl):
  return 'YES' if sl >= 5 else 'NO'

csv_data['Result'] = csv_data['Sepal_Length'].apply(sepal_length_requirement)
csv_data
```

[21]:

```
     Sepal_Length  Sepal_Width  Petal_Length  Petal_Width    Species Result
0             5.1          3.5           1.4          0.2     Setosa    YES
1             4.9          3.0           1.4          0.2     Setosa     NO
2             4.7          3.2           1.3          0.2     Setosa     NO
3             4.6          3.1           1.5          0.2     Setosa     NO
4             5.0          3.6           1.4          0.2     Setosa    YES
..            …            …             …            …          …      …
145           6.7          3.0           5.2          2.3  Verginica    YES
146           6.3          2.5           5.0          1.9  Verginica    YES
147           6.5          3.0           5.2          2.0  Verginica    YES
148           6.2          3.4           5.4          2.3  Verginica    YES
149           5.9          3.0           5.1          1.8  Verginica    YES

[150 rows x 6 columns]
```

10.8 drop added column

[22]:

```
del csv_data['Result']
```

10.9 select first row data

[23]:

```
csv_data.iloc[0]
```

[23]:

```
Sepal_Length       5.1
Sepal_Width        3.5
Petal_Length       1.4
Petal_Width        0.2
Species         Setosa
Name: 0, dtype: object
```

10.10 select 51 to 55 rows

[24]:

```
csv_data.iloc[51:56]
```

[24]:

```
    Sepal_Length  Sepal_Width  Petal_Length  Petal_Width     Species
51           6.4          3.2           4.5          1.5  Versicolor
52           6.9          3.1           4.9          1.5  Versicolor
53           5.5          2.3           4.0          1.3  Versicolor
54           6.5          2.8           4.6          1.5  Versicolor
55           5.7          2.8           4.5          1.3  Versicolor
```

10.11 Apply conditional formatting

```
[25]: csv_data[csv_data['Petal_Length'] > 5]
```

```
[25]:      Sepal_Length  Sepal_Width  Petal_Length  Petal_Width     Species
      83            6.0          2.7           5.1          1.6  Versicolor
      100           6.3          3.3           6.0          2.5   Verginica
      101           5.8          2.7           5.1          1.9   Verginica
      102           7.1          3.0           5.9          2.1   Verginica
      103           6.3          2.9           5.6          1.8   Verginica
      104           6.5          3.0           5.8          2.2   Verginica
      105           7.6          3.0           6.6          2.1   Verginica
      107           7.3          2.9           6.3          1.8   Verginica
      108           6.7          2.5           5.8          1.8   Verginica
      109           7.2          3.6           6.1          2.5   Verginica
      110           6.5          3.2           5.1          2.0   Verginica
      111           6.4          2.7           5.3          1.9   Verginica
      112           6.8          3.0           5.5          2.1   Verginica
      114           5.8          2.8           5.1          2.4   Verginica
      115           6.4          3.2           5.3          2.3   Verginica
      116           6.5          3.0           5.5          1.8   Verginica
      117           7.7          3.8           6.7          2.2   Verginica
      118           7.7          2.6           6.9          2.3   Verginica
      120           6.9          3.2           5.7          2.3   Verginica
      122           7.7          2.8           6.7          2.0   Verginica
      124           6.7          3.3           5.7          2.1   Verginica
      125           7.2          3.2           6.0          1.8   Verginica
      128           6.4          2.8           5.6          2.1   Verginica
      129           7.2          3.0           5.8          1.6   Verginica
      130           7.4          2.8           6.1          1.9   Verginica
      131           7.9          3.8           6.4          2.0   Verginica
      132           6.4          2.8           5.6          2.2   Verginica
      133           6.3          2.8           5.1          1.5   Verginica
      134           6.1          2.6           5.6          1.4   Verginica
      135           7.7          3.0           6.1          2.3   Verginica
      136           6.3          3.4           5.6          2.4   Verginica
      137           6.4          3.1           5.5          1.8   Verginica
      139           6.9          3.1           5.4          2.1   Verginica
      140           6.7          3.1           5.6          2.4   Verginica
      141           6.9          3.1           5.1          2.3   Verginica
      142           5.8          2.7           5.1          1.9   Verginica
      143           6.8          3.2           5.9          2.3   Verginica
      144           6.7          3.3           5.7          2.5   Verginica
      145           6.7          3.0           5.2          2.3   Verginica
      147           6.5          3.0           5.2          2.0   Verginica
      148           6.2          3.4           5.4          2.3   Verginica
      149           5.9          3.0           5.1          1.8   Verginica
```

10.12 replace “Iris-versicolor” with “versicolor”

[36]:

```
csv_data['Species'].replace('Versicolor', 'Iris-Versicolor', inplace=True)
csv_data['Species'].iloc[99]
```

[36]:

```
'Iris-Versicolor'
```

10.13 print "virginica" with "Petal_Length" more then 6

[26]:

```
csv_data[(csv_data['Petal_Length'] > 6) & (csv_data['Species'] == 'Verginica')]
```

[26]:

```
     Sepal_Length  Sepal_Width  Petal_Length  Petal_Width    Species
105           7.6          3.0           6.6          2.1  Verginica
107           7.3          2.9           6.3          1.8  Verginica
109           7.2          3.6           6.1          2.5  Verginica
117           7.7          3.8           6.7          2.2  Verginica
118           7.7          2.6           6.9          2.3  Verginica
122           7.7          2.8           6.7          2.0  Verginica
130           7.4          2.8           6.1          1.9  Verginica
131           7.9          3.8           6.4          2.0  Verginica
135           7.7          3.0           6.1          2.3  Verginica
```

10.14 find duplicate values

[28]:

```
csv_data.duplicated()
```

[28]:

```
0      False
1      False
2      False
3      False
4      False
        …
145    False
146    False
147    False
148    False
149    False
Length: 150, dtype: bool
```

10.15 Find basic information

[27]:

```
csv_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Sepal_Length  150 non-null    float64
 1   Sepal_Width   150 non-null    float64
 2   Petal_Length  150 non-null    float64
 3   Petal_Width   150 non-null    float64
 4   Species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```