assn1

May 22, 2023

[2]:
```
#Aishwarya kelgandre Roll no.73 batch T3
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
s1 =pd.Series(range(1,10,1))
s1
```

[2]:
```
0    1
1    2
2    3
3    4
4    5
5    6
6    7
7    8
8    9
dtype: int64
```

[5]:
```
s3 = pd.Series({1:21, 2:13,3:45})
s3
```

[5]:
```
1    21
2    13
3    45
dtype: int64
```

[6]:
```
s2 = pd.Series([1, 2, 3, 4], index=['p', 'q', 'r','s'], name='one')
s2
```

[6]:
```
p    1
q    2
r    3
s    4
Name: one, dtype: int64
```

[7]:
```
df1 = pd.DataFrame(s2)
df1
```

```
[7]:    one
     p    1
     q    2
     r    3
     s    4
```

```
[10]: df2 = pd.read_csv("E:\\TRINITY ACADEMY OF ENGINEERING PUNE\\TE␣
       ↪2022-23\\assignment\\dsbda\\csv\\housing.csv\\housing.csv")

      df2.head(10)
```

```
[10]:    longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
      0    -122.23     37.88                41.0        880.0           129.0
      1    -122.22     37.86                21.0       7099.0          1106.0
      2    -122.24     37.85                52.0       1467.0           190.0
      3    -122.25     37.85                52.0       1274.0           235.0
      4    -122.25     37.85                52.0       1627.0           280.0
      5    -122.25     37.85                52.0        919.0           213.0
      6    -122.25     37.84                52.0       2535.0           489.0
      7    -122.25     37.84                52.0       3104.0           687.0
      8    -122.26     37.84                42.0       2555.0           665.0
      9    -122.25     37.84                52.0       3549.0           707.0

         population  households  median_income  median_house_value ocean_proximity
      0       322.0       126.0         8.3252            452600.0        NEAR BAY
      1      2401.0      1138.0         8.3014            358500.0        NEAR BAY
      2       496.0       177.0         7.2574            352100.0        NEAR BAY
      3       558.0       219.0         5.6431            341300.0        NEAR BAY
      4       565.0       259.0         3.8462            342200.0        NEAR BAY
      5       413.0       193.0         4.0368            269700.0        NEAR BAY
      6      1094.0       514.0         3.6591            299200.0        NEAR BAY
      7      1157.0       647.0         3.1200            241400.0        NEAR BAY
      8      1206.0       595.0         2.0804            226700.0        NEAR BAY
      9      1551.0       714.0         3.6912            261100.0        NEAR BAY
```

```
[11]: df2.tail(3)
```

```
[11]:        longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
      20637    -121.22     39.43                17.0       2254.0           485.0
      20638    -121.32     39.43                18.0       1860.0           409.0
      20639    -121.24     39.37                16.0       2785.0           616.0

             population  households  median_income  median_house_value  \
      20637      1007.0       433.0         1.7000             92300.0
      20638       741.0       349.0         1.8672             84700.0
      20639      1387.0       530.0         2.3886             89400.0
      ocean_proximity
20637          INLAND
20638          INLAND
20639          INLAND
```

[14]:
```
df2['median_house_value_new']=df2['median_house_value']+111
df2.tail(3)
```

[14]:
```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
20637    -121.22     39.43                17.0       2254.0           485.0   
20638    -121.32     39.43                18.0       1860.0           409.0   
20639    -121.24     39.37                16.0       2785.0           616.0   

       population  households  median_income  median_house_value  \
20637      1007.0       433.0         1.7000             92300.0   
20638       741.0       349.0         1.8672             84700.0   
20639      1387.0       530.0         2.3886             89400.0   

      ocean_proximity  median_house_value_new
20637          INLAND                 92411.0
20638          INLAND                 84811.0
20639          INLAND                 89511.0
```

[15]:
```
df2.to_json('data1.json')
```

[16]:
```
len(df2['total_rooms'])
```

[16]:
```
20640
```

[17]:
```
df2['total_rooms'].count()
```

[17]:
```
20640
```

[18]:
```
df2['total_rooms'].mean()
```

[18]:
```
2635.7630813953488
```

[19]:
```
df2['total_rooms'].describe()
```

[19]:
```
count    20640.000000
mean      2635.763081
std       2181.615252
min          2.000000
25%       1447.750000
50%       2127.000000
75%       3148.000000
max      39320.000000
Name: total_rooms, dtype: float64
```

[20]:

```
df2['total_rooms'].cumsum()
```

[20]:

```
0             880.0
1            7979.0
2            9446.0
3           10720.0
4           12347.0
             …
20635    54394554.0
20636    54395251.0
20637    54397505.0
20638    54399365.0
20639    54402150.0
Name: total_rooms, Length: 20640, dtype: float64
```

[23]:

```
df2.describe()
```

[23]:

```
          longitude      latitude  housing_median_age   total_rooms  \
count  20640.000000  20640.000000        20640.000000  20640.000000
mean    -119.569704     35.631861           28.639486   2635.763081
std        2.003532      2.135952           12.585558   2181.615252
min     -124.350000     32.540000            1.000000      2.000000
25%     -121.800000     33.930000           18.000000   1447.750000
50%     -118.490000     34.260000           29.000000   2127.000000
75%     -118.010000     37.710000           37.000000   3148.000000
max     -114.310000     41.950000           52.000000  39320.000000

       total_bedrooms    population    households  median_income  \
count    20433.000000  20640.000000  20640.000000   20640.000000
mean       537.870553   1425.476744    499.539680       3.870671
std        421.385070   1132.462122    382.329753       1.899822
min          1.000000      3.000000      1.000000       0.499900
25%        296.000000    787.000000    280.000000       2.563400
50%        435.000000   1166.000000    409.000000       3.534800
75%        647.000000   1725.000000    605.000000       4.743250
max       6445.000000  35682.000000   6082.000000      15.000100

       median_house_value  median_house_value_new
count        20640.000000            20640.000000
mean        206855.816909           206966.816909
std         115395.615874           115395.615874
min          14999.000000            15110.000000
25%         119600.000000           119711.000000
50%         179700.000000           179811.000000
75%         264725.000000           264836.000000
max         500001.000000           500112.000000
```

[26]:
```
df = pd.read_csv("E:\\TRINITY ACADEMY OF ENGINEERING PUNE\\TE␣
 ↪2022-23\\assignment\\dsbda\\csv\\housing.csv\\housing.csv")
df.columns
```

[26]:
```
Index(['longitude', 'latitude', 'housing_median_age', 'total_rooms',
       'total_bedrooms', 'population', 'households', 'median_income',
       'median_house_value', 'ocean_proximity'],
      dtype='object')
```

[27]:
```
df.longitude
```

[27]:
```
0       -122.23
1       -122.22
2       -122.24
3       -122.25
4       -122.25
          …
20635   -121.09
20636   -121.21
20637   -121.22
20638   -121.32
20639   -121.24
Name: longitude, Length: 20640, dtype: float64
```

[28]:
```
df.iloc[:,1:3]
```

[28]:
```
       latitude  housing_median_age
0         37.88                41.0
1         37.86                21.0
2         37.85                52.0
3         37.85                52.0
4         37.85                52.0
…         …                 …
20635     39.48                25.0
20636     39.49                18.0
20637     39.43                17.0
20638     39.43                18.0
20639     39.37                16.0

[20640 rows x 2 columns]
```

[33]:
```
import pandas as pd

data = pd.read_csv("Downloads\\attendance_19-11-22.csv")
data.head(10)
```

[33]:
```
   Roll no     PRN no                      name  attended lectures  lectures  
0        1  72157089F         AJAY ANIL GAIKWAD                  5        19   \
1        2  72157090K          APRE OMKAR DILIP                  9        19
2        3  72157091H       AWACHAR SNEHAL SUHAS                 9        19
3        4  72157092F          BADE SHRIRAM ANIL                10        19
4        5  72157094B     BHOJANE APOORVA VASANT                10        19
5        6  72157095L  BHUTADA UTKARSH HIMANSHU                10        19
6        7  72157096J     BIJJARGI MAHESH PANDIT                10        19
7        8  72157097G        BOBADE ROHAN RAJESH                10        19
8        9  72157098E      BORATE SHWETA AMBADAS                10        19
9       10  72157099C      CHAVAN ATHARVA SANJAY                10        19

   percentage  
0     26.3158  
1     47.3684  
2     47.3684  
3     52.6316  
4     52.6316  
5     52.6316  
6     52.6316  
7     52.6316  
8     52.6316  
9     52.6316  
```

[34]:
```
data.describe()
```

[34]:
```
         Roll no  attended lectures  lectures  percentage
count  79.000000          79.000000      79.0   79.000000
mean   40.000000           9.898734      19.0   52.098620
std    22.949219           0.590414       0.0    3.107443
min     1.000000           5.000000      19.0   26.315800
25%    20.500000          10.000000      19.0   52.631600
50%    40.000000          10.000000      19.0   52.631600
75%    59.500000          10.000000      19.0   52.631600
max    79.000000          10.000000      19.0   52.631600
```

[35]:
```
data.isnull()
```

[35]:
```
    Roll no  PRN no   name  attended lectures  lectures  percentage
0     False   False  False              False     False       False
1     False   False  False              False     False       False
2     False   False  False              False     False       False
3     False   False  False              False     False       False
4     False   False  False              False     False       False
..      …      …    …              …       …         …
74    False   False  False              False     False       False
75    False   False  False              False     False       False
```

```
76    False   False  False               False     False       False
77    False   False  False               False     False       False
78    False   False  False               False     False       False

[79 rows x 6 columns]
```

[36]:
```
data.isnull().sum()
```

[36]:
```
Roll no              0
PRN no               0
name                 0
attended lectures    0
lectures             0
percentage           0
dtype: int64
```

[53]:
```
data["Roll no"].fillna("Roll no", inplace = False)
data.isnull().sum()
```

[53]:
```
Roll no              0
PRN no               0
name                 0
attended lectures    0
lectures             0
percentage           0
dtype: int64
```

[46]:
```
import numpy as np
data.replace(to_replace = np.nan, value = -99)
```

[46]:
```
    Roll no     PRN no                    name  attended lectures  lectures  \
0         1  72157089F       AJAY ANIL GAIKWAD                  5        19
1         2  72157090K        APRE OMKAR DILIP                  9        19
2         3  72157091H    AWACHAR SNEHAL SUHAS                  9        19
3         4  72157092F       BADE SHRIRAM ANIL                 10        19
4         5  72157094B  BHOJANE APOORVA VASANT                 10        19
..      …          …                       …                  …         …
74       75  72214538B  KURHADE PRITI RAJENDRA                 10        19
75       76  72214544G       PATIL MAMTA DINKAR                10        19
76       77  72214545E  SATALKAR GAURI NAVNATH                 10        19
77       78  72214546C     SHIRASKAR ISHA RAHUL                10        19
78       79  72214539L      ALEENA HANIF BAGWAN                 9        19

    percentage
0      26.3158
1      47.3684
2      47.3684
```

```
3     52.6316
4     52.6316
..        …
74    52.6316
75    52.6316
76    52.6316
77    52.6316
78    47.3684

[79 rows x 6 columns]
```

[47]:

```
data.fillna(method ='pad')
```

[47]:

```
    Roll no     PRN no                   name  attended lectures  lectures  \
0         1  72157089F       AJAY ANIL GAIKWAD                  5        19
1         2  72157090K        APRE OMKAR DILIP                  9        19
2         3  72157091H     AWACHAR SNEHAL SUHAS                 9        19
3         4  72157092F        BADE SHRIRAM ANIL                10        19
4         5  72157094B   BHOJANE APOORVA VASANT                10        19
..      …         …                       …                 …        …
74       75  72214538B   KURHADE PRITI RAJENDRA                10        19
75       76  72214544G       PATIL MAMTA DINKAR                10        19
76       77  72214545E   SATALKAR GAURI NAVNATH                10        19
77       78  72214546C      SHIRASKAR ISHA RAHUL               10        19
78       79  72214539L       ALEENA HANIF BAGWAN                9        19

    percentage
0      26.3158
1      47.3684
2      47.3684
3      52.6316
4      52.6316
..         …
74     52.6316
75     52.6316
76     52.6316
77     52.6316
78     47.3684

[79 rows x 6 columns]
```

[54]:

```
data['Roll no'].fillna(int(data['Roll no'].mean()), inplace=False)
```

[54]:

```
0      1
1      2
2      3
3      4
```

```
4      5
      ..
74    75
75    76
76    77
77    78
78    79
Name: Roll no, Length: 79, dtype: int64
```

[51]:
```
data.dropna(axis=1)
```

[51]:
```
    Roll no     PRN no                     name  attended lectures  lectures  \
0         1  72157089F        AJAY ANIL GAIKWAD                  5        19
1         2  72157090K         APRE OMKAR DILIP                  9        19
2         3  72157091H     AWACHAR SNEHAL SUHAS                  9        19
3         4  72157092F        BADE SHRIRAM ANIL                 10        19
4         5  72157094B  BHOJANE APOORVA VASANT                 10        19
..      …          …                      …                …         …
74       75  72214538B  KURHADE PRITI RAJENDRA                 10        19
75       76  72214544G       PATIL MAMTA DINKAR                 10        19
76       77  72214545E  SATALKAR GAURI NAVNATH                 10        19
77       78  72214546C     SHIRASKAR ISHA RAHUL                 10        19
78       79  72214539L      ALEENA HANIF BAGWAN                  9        19

    percentage
0      26.3158
1      47.3684
2      47.3684
3      52.6316
4      52.6316
..         …
74     52.6316
75     52.6316
76     52.6316
77     52.6316
78     47.3684

[79 rows x 6 columns]
```

[55]:
```
import pandas as pd

df = pd.DataFrame({"A":[12, 4, 5, None, 1],
 "B":[None, 2, 54, 3, None],
 "C":[20, 16, None, 3, 8],
 "D":[14, 3, None, None, 6]})

df
```

[55]:

```
      A     B     C     D
0  12.0   NaN  20.0  14.0
1   4.0   2.0  16.0   3.0
2   5.0  54.0   NaN   NaN
3   NaN   3.0   3.0   NaN
4   1.0   NaN   8.0   6.0
```

[56]:

```python
df.interpolate(method ='linear', limit_direction ='forward')
```

[56]:

```
      A     B     C     D
0  12.0   NaN  20.0  14.0
1   4.0   2.0  16.0   3.0
2   5.0  54.0   9.5   4.0
3   3.0   3.0   3.0   5.0
4   1.0   3.0   8.0   6.0
```

[58]:

```python
text="today is Monday"

text.replace(' ','')
```

[58]:

```
'todayisMonday'
```

[65]:

```python
text=' Today'
text.lstrip()
```

[65]:

```
'Today'
```

[60]:

```python
text='Today '
text.rstrip()
```

[60]:

```
'Today'
```

[66]:

```python
text=' Today '
text.strip()
```

[66]:

```
'Today'
```

[70]:

```python
import pandas
import scipy
import numpy
from sklearn.preprocessing import MinMaxScaler

X = [ [110, 200], [120, 800], [310, 400], [140, 900], [510, 200], [653, 400]␣
 ↪,[310, 880
] ]

scaler = MinMaxScaler(feature_range=(0,5))
rescaledX = scaler.fit_transform(X)
```

```
X
```

```
[70]: [[110, 200],
       [120, 800],
       [310, 400],
       [140, 900],
       [510, 200],
       [653, 400],
       [310, 880]]
```

```
[ ]:
```