task 39 pandas

July 26, 2022

[1]:
```
!pip install pandas
import pandas as pd
import numpy as np
```

```
Requirement already satisfied: pandas in d:\anakonda\lib\site-packages (1.4.2)
Requirement already satisfied: pytz>=2020.1 in d:\anakonda\lib\site-packages
(from pandas) (2021.3)
Requirement already satisfied: python-dateutil>=2.8.1 in d:\anakonda\lib\site-
packages (from pandas) (2.8.2)
Requirement already satisfied: numpy>=1.18.5 in d:\anakonda\lib\site-packages
(from pandas) (1.21.5)
Requirement already satisfied: six>=1.5 in d:\anakonda\lib\site-packages (from
python-dateutil>=2.8.1->pandas) (1.16.0)
```

[2]:
```
s1=pd.Series([23,24,25,26,27,28,29])
s1
```

[2]:
```
0    23
1    24
2    25
3    26
4    27
5    28
6    29
dtype: int64
```

[3]:
```
s2=pd.Series([23,24,25,26,27,28,29],index=['a','b','c','d','e','f','g'])
s2
```

[3]:
```
a    23
b    24
c    25
d    26
e    27
f    28
g    29
dtype: int64
```

[4]:
```
s3=pd.
  ↪Series([23,24,25,26,27,28,29],index=['a','b','c','d','e','f','g'],dtype='float')
s3
```

[4]:
```
a    23.0
b    24.0
c    25.0
d    26.0
e    27.0
f    28.0
g    29.0
dtype: float64
```

[5]:
```
# dict
```

[6]:
```
s4=pd.Series({'a':65,'f':43,'c':45})
s4
```

[6]:
```
a    65
f    43
c    45
dtype: int64
```

[7]:
```
# data frame
```

[8]:
```
d1=pd.DataFrame([43,54,65,76])
d1
```

[8]:

	0
0	43
1	54
2	65
3	76

[9]:
```
d2=pd.DataFrame([[2,3,4],[4,5,6],[1,2,3]])
d2
```

[9]:

	0	1	2
0	2	3	4
1	4	5	6
2	1	2	3

[10]:
```
d2=pd.DataFrame(s2)
d2
```

[10]:

	0
a	23
b	24

```
c  25
d  26
e  27
f  28
g  29
```

[11]:
```
d3=pd.DataFrame([[2,3,4],[4,5,6],[1,2,3]],columns=['a','b','c'])
d3
```

[11]:
```
   a  b  c
0  2  3  4
1  4  5  6
2  1  2  3
```

[12]:
```
d3=pd.
↪DataFrame([[2,3,4],[4,5,6],[1,2,3]],columns=['a','b','c'],index=['x','y','z'])
d3
```

[12]:
```
   a  b  c
x  2  3  4
y  4  5  6
z  1  2  3
```

[13]:
```
dic=[{'alex':1,'joe':2},{'alex':1,'joe':2,'ball':20},{'alex':1,'joe':2,'hon':
↪20}]
pd.DataFrame(dic,index=['a','b','c'])
```

[13]:
```
   alex  joe  ball   hon
a     1    2   NaN   NaN
b     1    2  20.0   NaN
c     1    2   NaN  20.0
```

[14]:
```
d3
```

[14]:
```
   a  b  c
x  2  3  4
y  4  5  6
z  1  2  3
```

[15]:
```
print(d3['a'])
print(d3['b'])
```

```
x    2
y    4
z    1
Name: a, dtype: int64
x    3
y    5
```

```
z    2
Name: b, dtype: int64
```

[16]:
```
d3['d']=d3['a']*d3['b']
d3['e']=d3['a']+d3['d']
d3
```

[16]:

	a	b	c	d	e
x	2	3	4	6	8
y	4	5	6	20	24
z	1	2	3	2	3

[17]:
```
pop = d3.pop('a')
d3
```

[17]:

	b	c	d	e
x	3	4	6	8
y	5	6	20	24
z	2	3	2	3

[18]:
```
d3
```

[18]:

	b	c	d	e
x	3	4	6	8
y	5	6	20	24
z	2	3	2	3

[19]:
```
del d3['c']
d3
```

[19]:

	b	d	e
x	3	6	8
y	5	20	24
z	2	2	3

[20]:
```
d3.insert(1,'new1',d3['e'])
d3
```

[20]:

	b	new1	d	e
x	3	8	6	8
y	5	24	20	24
z	2	3	2	3

[21]:
```
import numpy as np
d4 = pd.DataFrame({'abc':np.random.randint(2,6,size=(10)),'bcd':np.random.
  ↪randint(4,10,size=(10)),'cde':np.random.randint(3,10,size=(10))})
d4
```

[21]:

	abc	bcd	cde
0	5	9	4
1	3	7	8
2	4	7	3
3	5	6	8
4	4	4	4
5	3	4	3
6	5	8	4
7	3	7	3
8	3	9	4
9	2	8	3

[22]:
```
d={'c':1,'new1':33,'e':22}
df1=d3.append(d,ignore_index=True)
df1
```

```
C:\Users\bbala\AppData\Local\Temp\ipykernel_14104\273345864.py:2: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a
future version. Use pandas.concat instead.
  df1=d3.append(d,ignore_index=True)
```

[22]:

	b	new1	d	e	c
0	3.0	8	6.0	8	NaN
1	5.0	24	20.0	24	NaN
2	2.0	3	2.0	3	NaN
3	NaN	33	NaN	22	1.0

[23]:
```
d4.head()
```

[23]:

	abc	bcd	cde
0	5	9	4
1	3	7	8
2	4	7	3
3	5	6	8
4	4	4	4

[24]:
```
d4.tail()
```

[24]:

	abc	bcd	cde
5	3	4	3
6	5	8	4
7	3	7	3
8	3	9	4
9	2	8	3

[25]:
```
d4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
```

```
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   abc     10 non-null     int32
 1   bcd     10 non-null     int32
 2   cde     10 non-null     int32
dtypes: int32(3)
memory usage: 248.0 bytes
```

[26]:
```
d4
```

[26]:

	abc	bcd	cde
0	5	9	4
1	3	7	8
2	4	7	3
3	5	6	8
4	4	4	4
5	3	4	3
6	5	8	4
7	3	7	3
8	3	9	4
9	2	8	3

[27]:
```
d4.loc[9,'cde']
```

[27]: 3

[28]:
```
d4
```

[28]:

	abc	bcd	cde
0	5	9	4
1	3	7	8
2	4	7	3
3	5	6	8
4	4	4	4
5	3	4	3
6	5	8	4
7	3	7	3
8	3	9	4
9	2	8	3

[29]:
```
d4.loc[4:9,['cde','abc']]
```

[29]:

	cde	abc
4	4	4
5	3	3
6	4	5
7	3	3

```
8    4    3
9    3    2
```

[30]:
```
d4.loc[[3,4,7],['cde','abc']]
```

[30]:
```
   cde  abc
3    8    5
4    4    4
7    3    3
```

[31]:
```
d4
```

[31]:
```
   abc  bcd  cde
0    5    9    4
1    3    7    8
2    4    7    3
3    5    6    8
4    4    4    4
5    3    4    3
6    5    8    4
7    3    7    3
8    3    9    4
9    2    8    3
```

[32]:
```
d4.iloc[9,2]
```

[32]:
```
3
```

[33]:
```
d4.iloc[2:7,[0,2]]
```

[33]:
```
   abc  cde
2    4    3
3    5    8
4    4    4
5    3    3
6    5    4
```

[34]:
```
d4.abc
```

[34]:
```
0    5
1    3
2    4
3    5
4    4
5    3
6    5
7    3
8    3
```

```
9    2
Name: abc, dtype: int32
```

[35]:
```
d4.abc.values
```

[35]:
```
array([5, 3, 4, 5, 4, 3, 5, 3, 3, 2])
```

[36]:
```
d4['sub']= d4.abc.values - d4.cde.values - d4.bcd.values
d4
```

[36]:

	abc	bcd	cde	sub
0	5	9	4	-8
1	3	7	8	-12
2	4	7	3	-6
3	5	6	8	-9
4	4	4	4	-4
5	3	4	3	-4
6	5	8	4	-7
7	3	7	3	-7
8	3	9	4	-10
9	2	8	3	-9

[37]:
```
import numpy as np
```

[38]:
```
a = [['rk',102,15000],['rama',103,20000],['krishna',104,25000]]
df = pd.DataFrame(a,columns = ['name','id','salary'])
df
```

[38]:

	name	id	salary
0	rk	102	15000
1	rama	103	20000
2	krishna	104	25000

[39]:
```
y=df[df.salary>=20000]
print(y)
y[['id','salary']]
```

```
      name   id  salary
1     rama  103   20000
2  krishna  104   25000
```

[39]:

	id	salary
1	103	20000
2	104	25000

[40]:
```
df.append({'name':'sajan','id':105,'salary':30000},ignore_index=True)
```

```
C:\Users\bbala\AppData\Local\Temp\ipykernel_14104\717350958.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a
```

```
future version. Use pandas.concat instead.
  df.append({'name':'sajan','id':105,'salary':30000},ignore_index=True)
```

[40]:

	name	id	salary
0	rk	102	15000
1	rama	103	20000
2	krishna	104	25000
3	sajan	105	30000

[41]:
```
df=df.append({'name':np.nan,'id':105,'salary':30000},ignore_index=True)
df
```

```
C:\Users\bbala\AppData\Local\Temp\ipykernel_14104\2032447596.py:1:
FutureWarning: The frame.append method is deprecated and will be removed from
pandas in a future version. Use pandas.concat instead.
  df=df.append({'name':np.nan,'id':105,'salary':30000},ignore_index=True)
```

[41]:

	name	id	salary
0	rk	102.0	15000.0
1	rama	103.0	20000.0
2	krishna	104.0	25000.0
3	NaN	105.0	30000.0

[42]:
```
df.isnull()
```

[42]:

	name	id	salary
0	False	False	False
1	False	False	False
2	False	False	False
3	True	False	False

[43]:
```
df.isnull().sum()
```

[43]:
```
name      1
id        0
salary    0
dtype: int64
```

[44]:
```
df.dropna()
```

[44]:

	name	id	salary
0	rk	102.0	15000.0
1	rama	103.0	20000.0
2	krishna	104.0	25000.0

[45]:
```
df
```

[45]:

	name	id	salary
0	rk	102.0	15000.0

```
1     rama  103.0  20000.0
2  krishna  104.0  25000.0
3      NaN  105.0  30000.0
```

[46]:
```
df.fillna(value='abc')
```

[46]:
```
      name     id   salary
0       rk  102.0  15000.0
1     rama  103.0  20000.0
2  krishna  104.0  25000.0
3      abc  105.0  30000.0
```

[47]:
```
df = pd.DataFrame({'Animal' : ['Falcon', 'Falcon','Parrot', 'Parrot'],'Max␣
 ↪Speed' : [380., 370., 24., 26.]})
df
```

[47]:
```
   Animal  Max Speed
0  Falcon      380.0
1  Falcon      370.0
2  Parrot       24.0
3  Parrot       26.0
```

[48]:
```
df.groupby(['Animal']).mean()
```

[48]:
```
        Max Speed
Animal
Falcon      375.0
Parrot       25.0
```

[ ]:
```

```

[ ]:
```
# get dummies
```

[49]:
```
data = {'firstname': ['Arun', 'Jebu', 'Venkat', 'Rekha', 'Majid','Mohsin'],
        'lastname': ['Kumar', 'Jacob', 'Raghavan', 'Singh', 'Khan','Khan'],
        'employmenttype': ['Service', 'Business', 'Student', 'Service',␣
 ↪'Business','Business'],
        'country' :['India','USA','USA','Sweden','Australia','Germany']}


df = pd.DataFrame(data, columns =␣
 ↪['firstname','lastname','employmenttype','country'])

print(df)
```

```
  firstname  lastname employmenttype    country
0      Arun     Kumar        Service      India
1      Jebu     Jacob       Business        USA
```

```
2    Venkat  Raghavan        Student        USA
3     Rekha     Singh        Service     Sweden
4     Majid      Khan       Business  Australia
5    Mohsin      Khan       Business    Germany
```

[50]:
```
df1 = pd.get_dummies(df['employmenttype'])
df2 = pd.get_dummies(df['country'])
print(df1)
print(df2)
```

```
   Business  Service  Student
0         0        1        0
1         1        0        0
2         0        0        1
3         0        1        0
4         1        0        0
5         1        0        0
   Australia  Germany  India  Sweden  USA
0          0        0      1       0    0
1          0        0      0       0    1
2          0        0      0       0    1
3          0        0      0       1    0
4          1        0      0       0    0
5          0        1      0       0    0
```

[51]:
```
df
```

[51]:
```
  firstname  lastname employmenttype    country
0      Arun     Kumar        Service      India
1      Jebu     Jacob       Business        USA
2    Venkat  Raghavan        Student        USA
3     Rekha     Singh        Service     Sweden
4     Majid      Khan       Business  Australia
5    Mohsin      Khan       Business    Germany
```

[53]:
```
frames=[df,df1,df2]

result=pd.concat(frames,axis=1)
result
```

[53]:
```
  firstname  lastname employmenttype    country  Business  Service  Student  \
0      Arun     Kumar        Service      India         0        1        0
1      Jebu     Jacob       Business        USA         1        0        0
2    Venkat  Raghavan        Student        USA         0        0        1
3     Rekha     Singh        Service     Sweden         0        1        0
4     Majid      Khan       Business  Australia         1        0        0
5    Mohsin      Khan       Business    Germany         1        0        0
```

	Australia	Germany	India	Sweden	USA
0	0	0	1	0	0
1	0	0	0	0	1
2	0	0	0	0	1
3	0	0	0	1	0
4	1	0	0	0	0
5	0	1	0	0	0

[54]:
```
# concatination
```

[55]:
```
india_weather = pd.DataFrame({
    "city": ["mumbai","delhi","banglore"],
    "temperature": [32,45,30],
    "humidity": [80, 60, 78]})
india_weather
```

[55]:

	city	temperature	humidity
0	mumbai	32	80
1	delhi	45	60
2	banglore	30	78

[56]:
```
us_weather = pd.DataFrame({
    "city": ["new york","chicago","orlando"],
    "temperature": [21,14,35],
    "humidity": [68, 65, 75]})
us_weather
```

[56]:

	city	temperature	humidity
0	new york	21	68
1	chicago	14	65
2	orlando	35	75

[57]:
```
df = pd.concat([india_weather, us_weather])
df
```

[57]:

	city	temperature	humidity
0	mumbai	32	80
1	delhi	45	60
2	banglore	30	78
0	new york	21	68
1	chicago	14	65
2	orlando	35	75

[58]:
```
df = pd.concat([india_weather, us_weather], ignore_index=True)
df
```

[58]:
```
        city  temperature  humidity
0     mumbai           32        80
1      delhi           45        60
2   banglore           30        78
3   new york           21        68
4    chicago           14        65
5    orlando           35        75
```

[59]:
```
df = pd.concat([india_weather, us_weather], keys=["india", "us"])
df
```

[59]:
```
              city  temperature  humidity
india 0     mumbai           32        80
      1      delhi           45        60
      2   banglore           30        78
us    0   new york           21        68
      1    chicago           14        65
      2    orlando           35        75
```

[60]:
```
# relationships
```

[ ]:
```
df=pd.read_csv('data_panda.csv')
df.corr()
```

[62]:
```
import pandas as pd

def calc_sum(x):
    return x+1

data = {
  "x": [50, 40, 30],
  "y": [300, 1112, 42]}

df = pd.DataFrame(data)
print(df)

x = df.apply(calc_sum)

print(x)
```

```
    x     y
0  50   300
1  40  1112
2  30    42
    x     y
0  51   301
1  41  1113
2  31    43
```

[65]:
```
# explore sql joins
import pandas as pd

data1 = {
  "name": ["Sally", "Mary", "John"],
  "age": [50, 40, 30]}

data2 = {
  "name": ["Sally", "Peter", "Micky"],
  "age": [77, 44, 22]}

df1 = pd.DataFrame(data1)
df2 = pd.DataFrame(data2)

newdf = df1.merge(df2, how='right')
print(newdf)
```

```
    name  age
0  Sally   77
1  Peter   44
2  Micky   22
```

[66]:
```
df3=pd.merge(df1,df2,on="name",how="outer")
df3
```

[66]:

	name	age_x	age_y
0	Sally	50.0	77.0
1	Mary	40.0	NaN
2	John	30.0	NaN
3	Peter	NaN	44.0
4	Micky	NaN	22.0

[67]:
```
df3=pd.merge(df1,df2,on="name",how="left")
df3
```

[67]:

	name	age_x	age_y
0	Sally	50	77.0
1	Mary	40	NaN
2	John	30	NaN

[68]:
```
df3=pd.merge(df1,df2,on="name",how="right")
df3
```

[68]:

	name	age_x	age_y
0	Sally	50.0	77
1	Peter	NaN	44
2	Micky	NaN	22

[69]:
```
df3=pd.merge(df1,df2,on="name",how="inner")
df3
```

[69]:

	name	age_x	age_y
0	Sally	50	77

[ ]:
```

```