

pandas-2

February 26, 2024

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

```
[ ]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
[ ]: a=['jwalitha','ramya','durga','jahnvi','lahari','sunny','dhanush']
r=pd.Series(a,index=[67,43,44,89,34,45,23])
print(r)
```

```
67    jwalitha
43      ramya
44      durga
89     jahnvi
34     lahari
45      sunny
23     dhanush
dtype: object
```

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/dataset/diabetcsv.csv")
print(df)
```

	preg	plas	pres	skin	insu	mass	pedi	age	class
0	6	148	72	35	0	33.6	0.627	50	tested_positive
1	1	85	66	29	0	26.6	0.351	31	tested_negative
2	8	183	64	0	0	23.3	0.672	32	tested_positive
3	1	89	66	23	94	28.1	0.167	21	tested_negative
4	0	137	40	35	168	43.1	2.288	33	tested_positive
..
763	10	101	76	48	180	32.9	0.171	63	tested_negative
764	2	122	70	27	0	36.8	0.340	27	tested_negative
765	5	121	72	23	112	26.2	0.245	30	tested_negative
766	1	126	60	0	0	30.1	0.349	47	tested_positive
767	1	93	70	31	0	30.4	0.315	23	tested_negative

[768 rows x 9 columns]

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/grades_withnulls.
↪csv")
print(df)
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
0	Joe	K	9.8	10.0	9.9	A+	1
1	Rajesh	M	8.9	9.1	9.3	A	1
2	Kissan	V	9.9	9.8	10.0	A	0
3	Mary	N	7.7	8.0	NaN	B	0
4	Jeen	K	9.8	9.1	9.9	A+	1
5	Raj	M	8.9	9.1	9.3	A	1
6	Hassan	V	9.9	9.0	9.2	A	1
7	Mari	N	7.7	8.0	7.1	B	1
8	Jess	K	NaN	9.1	9.9	A+	1
9	Rajini	M	NaN	9.1	9.3	A	0
10	Kiran	V	NaN	9.3	9.2	A	0
11	Maya	N	7.7	8.0	7.1	B	0
12	Jolin	K	9.8	9.1	9.9	A+	1
13	Rajesh	M	8.9	9.1	9.3	A	1
14	Riya	M	9.3	9.9	10.0	A	1
15	Sana	V	9.9	9.3	9.2	A	0
16	Mark	N	7.7	8.0	7.0	B	0

```
[ ]: df=pd.read_excel("/content/drive/MyDrive/DAP_TEAM 4/datasets/grades_withnulls_
↪(1).xlsx",sheet_name=0)
print(df)
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
0	Joe	K	9.8	10.0	9.9	A+	1
1	Rajesh	M	8.9	9.1	9.3	A	1
2	Kissan	V	9.9	9.8	10.0	A	0
3	Mary	N	7.7	8.0	NaN	B	0
4	Jeen	K	9.8	9.1	9.9	A+	1
5	Raj	M	8.9	9.1	9.3	A	1
6	Hassan	V	9.9	9.0	9.2	A	1
7	Mari	N	7.7	8.0	7.1	B	1
8	Jess	K	NaN	9.1	9.9	A+	1
9	Rajini	M	NaN	9.1	9.3	A	0
10	Kiran	V	NaN	9.3	9.2	A	0
11	Maya	N	7.7	8.0	7.1	B	0
12	Jolin	K	9.8	9.1	9.9	A+	1
13	Rajesh	M	8.9	9.1	9.3	A	1
14	Riya	M	9.3	9.9	10.0	A	1
15	Sana	V	9.9	9.3	9.2	A	0
16	Mark	N	7.7	8.0	7.0	B	0

```
[ ]: df=pd.read_excel("/content/drive/MyDrive/DAP_TEAM 4/datasets/diabetcsv.
↳xlsx",sheet_name=0)
print(df)
```

	preg	plas	pres	skin	insu	mass	pedi	age	class
0	6	148	72	35	0	33.6	0.627	50	tested_positive
1	1	85	66	29	0	26.6	0.351	31	tested_negative
2	8	183	64	0	0	23.3	0.672	32	tested_positive
3	1	89	66	23	94	28.1	0.167	21	tested_negative
4	0	137	40	35	168	43.1	2.288	33	tested_positive
..
763	10	101	76	48	180	32.9	0.171	63	tested_negative
764	2	122	70	27	0	36.8	0.340	27	tested_negative
765	5	121	72	23	112	26.2	0.245	30	tested_negative
766	1	126	60	0	0	30.1	0.349	47	tested_positive
767	1	93	70	31	0	30.4	0.315	23	tested_negative

[768 rows x 9 columns]

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/grades_withnulls.
↳csv")
print(df.head())
print(df.tail())
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
0	Joe	K	9.8	10.0	9.9	A+	1
1	Rajesh	M	8.9	9.1	9.3	A	1
2	Kissan	V	9.9	9.8	10.0	A	0
3	Mary	N	7.7	8.0	NaN	B	0
4	Jeen	K	9.8	9.1	9.9	A+	1
12	Jolin	K	9.8	9.1	9.9	A+	1
13	Rajesh	M	8.9	9.1	9.3	A	1
14	Riya	M	9.3	9.9	10.0	A	1
15	Sana	V	9.9	9.3	9.2	A	0
16	Mark	N	7.7	8.0	7.0	B	0

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/diabetcsv.csv")
print(df.describe)
```

	preg	plas	pres	skin	insu	mass	pedi	age	class
0	6	148	72	35	0	33.6	0.627	50	tested_positive
1	1	85	66	29	0	26.6	0.351	31	tested_negative
2	8	183	64	0	0	23.3	0.672	32	tested_positive
3	1	89	66	23	94	28.1	0.167	21	tested_negative
4	0	137	40	35	168	43.1	2.288	33	tested_positive
..

```

763    10   101    76    48   180   32.9   0.171    63  tested_negative
764     2   122    70    27     0   36.8   0.340    27  tested_negative
765     5   121    72    23   112   26.2   0.245    30  tested_negative
766     1   126    60     0     0   30.1   0.349    47  tested_positive
767     1    93    70    31     0   30.4   0.315    23  tested_negative

```

[768 rows x 9 columns]>

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/diabetcsv.csv")
      print(df.describe().T)
```

	count	mean	std	min	25%	50%	75%	\
preg	768.0	3.845052	3.369578	0.000	1.00000	3.0000	6.00000	
plas	768.0	120.894531	31.972618	0.000	99.00000	117.0000	140.25000	
pres	768.0	69.105469	19.355807	0.000	62.00000	72.0000	80.00000	
skin	768.0	20.536458	15.952218	0.000	0.00000	23.0000	32.00000	
insu	768.0	79.799479	115.244002	0.000	0.00000	30.5000	127.25000	
mass	768.0	31.992578	7.884160	0.000	27.30000	32.0000	36.60000	
pedi	768.0	0.471876	0.331329	0.078	0.24375	0.3725	0.62625	
age	768.0	33.240885	11.760232	21.000	24.00000	29.0000	41.00000	

	max
preg	17.00
plas	199.00
pres	122.00
skin	99.00
insu	846.00
mass	67.10
pedi	2.42
age	81.00

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/diabetcsv.csv")
      print(df.shape)
      print(df.shape[0])
      print(df.shape[1])
```

(768, 9)

768

9

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/DAP_TEAM 4/datasets/diabetcsv.csv")
      print(df.columns)
```

Index(['preg', 'plas', 'pres', 'skin', 'insu', 'mass', 'pedi', 'age', 'class'],
dtype='object')

```
[ ]: df2=df.copy()
      df2.loc[2:5, 'SEM1']=None
```

```
df2.head(7)
```

```
[ ]:      Names Initials SEM1 SEM2 SEM3 Grade Placed
0      Joe          K   9.8  10.0   9.9   A+      1
1  Rajesh          M   8.9   9.1   9.3    A      1
2  Kissan          V   NaN   9.8  10.0    A      0
3    Mary          N   NaN   8.0   NaN    B      0
4    Jeen          K   NaN   9.1   9.9   A+      1
5     Raj          M   NaN   9.1   9.3    A      1
6  Hassan          V   9.9   9.0   9.2    A      1
```

```
[ ]: df2.isnull().head(7)
```

```
[ ]:      Names Initials SEM1 SEM2 SEM3 Grade Placed
0  False      False  False  False  False  False  False
1  False      False  False  False  False  False  False
2  False      False   True  False  False  False  False
3  False      False   True  False   True  False  False
4  False      False   True  False  False  False  False
5  False      False   True  False  False  False  False
6  False      False  False  False  False  False  False
```

```
[ ]: df.isnull().tail(7)
```

```
[ ]:      Names Initials SEM1 SEM2 SEM3 Grade Placed
10  False      False   True  False  False  False  False
11  False      False  False  False  False  False  False
12  False      False  False  False  False  False  False
13  False      False  False  False  False  False  False
14  False      False  False  False  False  False  False
15  False      False  False  False  False  False  False
16  False      False  False  False  False  False  False
```

```
[ ]: print(df2.isnull().sum())
```

```
Names      0
Initials   0
SEM1        7
SEM2        0
SEM3        1
Grade       0
Placed      0
dtype: int64
```

```
[ ]: print(df2['SEM1'])
```

```
0      9.8
1      8.9
```

```

2    NaN
3    NaN
4    NaN
5    NaN
6    9.9
7    7.7
8    NaN
9    NaN
10   NaN
11   7.7
12   9.8
13   8.9
14   9.3
15   9.9
16   7.7
Name: SEM1, dtype: float64

```

```
[ ]: print(df2[['SEM1', 'Names']])
```

```

      SEM1  Names
0    9.8    Joe
1    8.9  Rajesh
2    NaN  Kissan
3    NaN   Mary
4    NaN   Jeen
5    NaN    Raj
6    9.9  Hassan
7    7.7   Mari
8    NaN   Jess
9    NaN  Rajini
10   NaN   Kiran
11   7.7   Maya
12   9.8   Jolin
13   8.9  Rajesh
14   9.3   Riya
15   9.9   Sana
16   7.7   Mark

```

```
[ ]: print(df2[df2.index==1])
```

```

      Names Initials  SEM1  SEM2  SEM3  Grade  Placed
1  Rajesh         M    8.9    9.1    9.3     A        1

```

```
[ ]: print(df2[df2.index.isin(range(2,6))])
```

```

      Names Initials  SEM1  SEM2  SEM3  Grade  Placed
2  Kissan         V   NaN    9.8   10.0     A        0
3   Mary         N   NaN    8.0   NaN     B        0

```

4	Jeen	K	NaN	9.1	9.9	A+	1
5	Raj	M	NaN	9.1	9.3	A	1

```
[ ]: print(df2.loc[3])
```

```
Names      Mary
Initials    N
SEM1       NaN
SEM2       8.0
SEM3       NaN
Grade       B
Placed      0
Name: 3, dtype: object
```

```
[ ]: print(df2.loc[6:11])
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
6	Hassan	V	9.9	9.0	9.2	A	1
7	Mari	N	7.7	8.0	7.1	B	1
8	Jess	K	NaN	9.1	9.9	A+	1
9	Rajini	M	NaN	9.1	9.3	A	0
10	Kiran	V	NaN	9.3	9.2	A	0
11	Maya	N	7.7	8.0	7.1	B	0

```
[ ]: print(df2.loc[3]==None)
```

```
Names      False
Initials    False
SEM1       False
SEM2       False
SEM3       False
Grade       False
Placed      False
Name: 3, dtype: bool
```

```
[ ]: df2.loc[[10,2,7]]
```

```
[ ]:
      Names Initials SEM1 SEM2 SEM3 Grade Placed
10  Kiran         V  NaN  9.3  9.2    A      0
2   Kissan        V  NaN  9.8 10.0    A      0
7   Mari          N  7.7  8.0  7.1    B      1
```

```
[ ]: df2.loc[[1,2,3]]
```

```
[ ]:
      Names Initials SEM1 SEM2 SEM3 Grade Placed
1  Rajesh         M  8.9  9.1  9.3    A      1
2  Kissan         V  NaN  9.8 10.0    A      0
3   Mary         N  NaN  8.0  NaN    B      0
```

```
[ ]: df2.loc[10:15,['Names','Grade'],]
```

```
[ ]:      Names Grade
10   Kiran      A
11    Maya      B
12   Jolin     A+
13 Rajesh      A
14   Riya      A
15   Sana      A
```

```
[ ]: df2.iloc[10:14,:3]
```

```
[ ]:      Names Initials  SEM1
10   Kiran           V   NaN
11    Maya           N   7.7
12   Jolin           K   9.8
13 Rajesh           M   8.9
```

```
[ ]: df2.loc[:7,["Names"]]
```

```
[ ]:      Names
0     Joe
1  Rajesh
2  Kissan
3    Mary
4    Jeen
5     Raj
6  Hassan
7    Mari
```

```
[ ]: df2.loc[df2['Names']=='Rajesh']
```

```
[ ]:      Names Initials  SEM1  SEM2  SEM3 Grade  Placed
1   Rajesh           M   8.9   9.1   9.3     A        1
13  Rajesh           M   8.9   9.1   9.3     A        1
```

```
[ ]: df2.loc[df2['Grade']=='A']
```

```
[ ]:      Names Initials  SEM1  SEM2  SEM3 Grade  Placed
1   Rajesh           M   8.9   9.1   9.3     A        1
2   Kissan           V   NaN   9.8  10.0     A        0
5     Raj           M   NaN   9.1   9.3     A        1
6  Hassan           V   9.9   9.0   9.2     A        1
9  Rajini           M   NaN   9.1   9.3     A        0
10  Kiran           V   NaN   9.3   9.2     A        0
13  Rajesh           M   8.9   9.1   9.3     A        1
14   Riya           M   9.3   9.9  10.0     A        1
```


15	Sana	V	9.9	9.3	9.2	A	0
----	------	---	-----	-----	-----	---	---

```
[ ]: df2[df2.SEM2==9]
```

```
[ ]:      Names Initials SEM1 SEM2 SEM3 Grade Placed
6 Hassan          V   9.9   9.0   9.2     A      1
```

```
[ ]: df2.loc[df2['SEM1']>9,['Names']]
```

```
[ ]:      Names
0      Joe
6 Hassan
12 Jolin
14 Riya
15 Sana
```

```
[ ]: df3=df2.copy()
df3.dropna(inplace=True,axis=1)
df3
```

```
[ ]:      Names Initials SEM2 Grade Placed
0      Joe          K  10.0   A+      1
1 Rajesh          M   9.1    A      1
2 Kissan          V   9.8    A      0
3 Mary           N   8.0    B      0
4 Jeen           K   9.1   A+      1
5 Raj            M   9.1    A      1
6 Hassan          V   9.0    A      1
7 Mari           N   8.0    B      1
8 Jess           K   9.1   A+      1
9 Rajini          M   9.1    A      0
10 Kiran          V   9.3    A      0
11 Maya           N   8.0    B      0
12 Jolin          K   9.1   A+      1
13 Rajesh          M   9.1    A      1
14 Riya           M   9.9    A      1
15 Sana           V   9.3    A      0
16 Mark           N   8.0    B      0
```

```
[ ]: df3.dropna(inplace=True,how='all')
df3
```

```
[ ]:      Names Initials SEM2 Grade Placed
0      Joe          K  10.0   A+      1
1 Rajesh          M   9.1    A      1
2 Kissan          V   9.8    A      0
3 Mary           N   8.0    B      0
```

4	Jeen	K	9.1	A+	1
5	Raj	M	9.1	A	1
6	Hassan	V	9.0	A	1
7	Mari	N	8.0	B	1
8	Jess	K	9.1	A+	1
9	Rajini	M	9.1	A	0
10	Kiran	V	9.3	A	0
11	Maya	N	8.0	B	0
12	Jolin	K	9.1	A+	1
13	Rajesh	M	9.1	A	1
14	Riya	M	9.9	A	1
15	Sana	V	9.3	A	0
16	Mark	N	8.0	B	0

```
[ ]: df3['Names']
```

```
[ ]: 0      Joe
      1    Rajesh
      2    Kissan
      3     Mary
      4     Jeen
      5      Raj
      6    Hassan
      7     Mari
      8     Jess
      9    Rajini
     10    Kiran
     11     Maya
     12    Jolin
     13    Rajesh
     14     Riya
     15     Sana
     16     Mark
Name: Names, dtype: object
```

```
[ ]: mv = df2['SEM1'].mean()
      df2 = df2.fillna(mv)
      print(df2)
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
0	Joe	K	9.80	10.0	9.90	A+	1
1	Rajesh	M	8.90	9.1	9.30	A	1
2	Kissan	V	8.96	9.8	10.00	A	0
3	Mary	N	8.96	8.0	8.96	B	0
4	Jeen	K	8.96	9.1	9.90	A+	1
5	Raj	M	8.96	9.1	9.30	A	1
6	Hassan	V	9.90	9.0	9.20	A	1

7	Mari	N	7.70	8.0	7.10	B	1
8	Jess	K	8.96	9.1	9.90	A+	1
9	Rajini	M	8.96	9.1	9.30	A	0
10	Kiran	V	8.96	9.3	9.20	A	0
11	Maya	N	7.70	8.0	7.10	B	0
12	Jolin	K	9.80	9.1	9.90	A+	1
13	Rajesh	M	8.90	9.1	9.30	A	1
14	Riya	M	9.30	9.9	10.00	A	1
15	Sana	V	9.90	9.3	9.20	A	0
16	Mark	N	7.70	8.0	7.00	B	0

```
[ ]: df2 = df2.drop_duplicates()
      print(df2)
```

	Names	Initials	SEM1	SEM2	SEM3	Grade	Placed
0	Joe	K	9.80	10.0	9.90	A+	1
1	Rajesh	M	8.90	9.1	9.30	A	1
2	Kissan	V	8.96	9.8	10.00	A	0
3	Mary	N	8.96	8.0	8.96	B	0
4	Jeen	K	8.96	9.1	9.90	A+	1
5	Raj	M	8.96	9.1	9.30	A	1
6	Hassan	V	9.90	9.0	9.20	A	1
7	Mari	N	7.70	8.0	7.10	B	1
8	Jess	K	8.96	9.1	9.90	A+	1
9	Rajini	M	8.96	9.1	9.30	A	0
10	Kiran	V	8.96	9.3	9.20	A	0
11	Maya	N	7.70	8.0	7.10	B	0
12	Jolin	K	9.80	9.1	9.90	A+	1
14	Riya	M	9.30	9.9	10.00	A	1
15	Sana	V	9.90	9.3	9.20	A	0
16	Mark	N	7.70	8.0	7.00	B	0

```
[ ]: df2.rename(columns = {'Grade':'GPA'}, inplace = True)
      df2.head()
```

```
[ ]:      Names Initials SEM1 SEM2 SEM3 GPA Placed
0      Joe      K  9.80  10.0  9.90  A+      1
1 Rajesh      M  8.90   9.1   9.30   A      1
2 Kissan      V  8.96   9.8  10.00   A      0
3   Mary      N  8.96   8.0   8.96   B      0
4   Jeen      K  8.96   9.1   9.90  A+      1
```

```
[ ]: df2['Avg_score'] = (df2['1stsem']+df2['2ndsem']+df2['3rdsem'])/3
      df2.head()
```

```
[ ]:      Student_name Mr/Ms 1stsem 2ndsem 3rdsem F_Outcome Placed? Avg_score
0      Joe      K    9.8    10.0    9.9      A+      1      9.9
1 Rajesh      M    8.9     9.1     9.3      A      1      9.1
```

2	Kissan	V	NaN	9.8	10.0	A	0	NaN
3	Mary	N	NaN	8.0	NaN	B	0	NaN
4	Jeen	K	NaN	9.1	9.9	A+	1	NaN

```
[ ]: df2.groupby('Placed?').min()
```

```
[ ]:      Student_name Mr/Ms  1stsem  2ndsem  3rdsem F_Outcome  Avg_score
Placed?
0      Kiran      M      7.7      8.0      7.0      A      7.566667
1      Hassan     K      7.7      8.0      7.1      A      7.600000
```

```
[ ]: import pandas as pd
df2.plot.line(subplots=True)
```

```
[ ]: array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >], dtype=object)
```

