قسم علوم الحاسوب وتقنية المعلومات



الجمهورية اليمنية

جامعة إب كلية العلوم

تطبيق محاضرة مقرر

تنقيب بيانات - عملي

Data Mining

الثالثة

عمل الطالب:

أسامة سعيد محمد حمود سعيد - مجموعة A

إشراف:

أ مالك المصنف

2024 - 2025

```
import pandas as pd
import numpy as np
temp=[4,5,6,78,2]
 ser=pd.Series(temp)
 print(f'The Table is {ser}')
print(f'The Data Type of The Df is {type(ser)}')
The Table is 0 4
1
      5
2
      6
3
     78
      2
4
dtype: int64
The Data Type of The Df is <class 'pandas.core.series.Series'>
data={
    "Name":["Malek","Ali","Ahmed"],
    "Age":[23,25,28]
 }
df=pd.DataFrame(data)
 print(f'The Table is {df}')
print(f'The Data Type of The Df is {type(df)}')
# df.iloc[0]
# df.iloc[0,1]
The Table is
                 Name Age
  Malek
           23
1
     Ali
           25
2 Ahmed
           28
The Data Type of The Df is <class 'pandas.core.frame.DataFrame'>
myJsonFile=pd.read json('output.json')
myJsonFile.head()
        Age
                   City
    Name
0
    John
           25
             New York
1 Alice
           30
                London
2
     Bob
         35
                 Paris
myXMLFile=pd.read xml('output.xml')
myJsonFile.head()
    Name Age
                   City
           25
0
    John
              New York
1 Alice
           30
                 London
     Bob
         35 Paris
myTxtFile = pd.read fwf('data.txt')
 print(myTxtFile)
```

```
John
          25
               170
   Alice
          28
               165
0
     Bob
          30
              180
 dataset=pd.read csv('students.csv')
print(dataset)
          sex race ethnicity parental level of education
                                                                      lunch
0
       female
                                          bachelor's degree
                                                                   standard
                      group B
1
       female
                      group C
                                               some college
                                                                   standard
       female
                                            master's degree
                                                                   standard
                           NaN
         male
                      group A
                                         associate's degree
                                                              free/reduced
         male
                      group C
                                                         NaN
                                                                   standard
20999
      female
                      group E
                                            master's degree
                                                                   standard
21000
         male
                      group C
                                                high school free/reduced
21001
       female
                      group C
                                                high school free/reduced
21002
       female
                      group D
                                               some college
                                                                   standard
21003
       female
                                               some college free/reduced
                      group D
      test preparation course math digree
                                               reading digree writing
score
                                                          72.0
0
                           none
                                           72
74
                                           69
                                                          90.0
1
                     completed
88
                                                          95.0
                                           90
2
                           none
93
3
                           none
                                           47
                                                          57.0
44
                                                          78.0
4
                                           76
                           none
75
. . .
                                                           . . .
20999
                     completed
                                           58
                                                          99.0
95
21000
                           none
                                           62
                                                          55.0
55
21001
                     completed
                                           49
                                                          71.0
```

```
65
21002
                     completed
                                           68
                                                          48.0
77
21003
                                           77
                                                          89.0
                           none
86
[21004 rows x 8 columns]
dataset.shape
(21004, 8)
dataset.head()
      sex race ethnicity parental level of education
                                                                  lunch \
0
   female
                  group B
                                     bachelor's degree
                                                              standard
   female
1
                  group C
                                           some college
                                                              standard
2
   female
                      NaN
                                       master's degree
                                                              standard
3
     male
                  group A
                                    associate's degree
                                                          free/reduced
4
     male
                                                              standard
                  group C
                                                     NaN
  test preparation course math digree reading digree writing score
0
                                                      72.0
                                                                        74
                                      72
                      none
                                      69
1
                 completed
                                                      90.0
                                                                        88
                                      90
                                                      95.0
                                                                        93
                      none
3
                                      47
                                                      57.0
                                                                        44
                      none
                      none
                                      76
                                                      78.0
                                                                        75
dataset.head(20)
       sex race ethnicity parental level of education
                                                                   lunch \
    female
0
                   group B
                                      bachelor's degree
                                                               standard
1
    female
                   group C
                                            some college
                                                               standard
2
    female
                                         master's degree
                       NaN
                                                               standard
3
                                                           free/reduced
      male
                   group A
                                     associate's degree
4
      male
                   group C
                                                      NaN
                                                               standard
5
    female
                                     associate's degree
                                                               standard
                   group B
6
    female
                                            some college
                                                               standard
                   group B
7
      male
                   group B
                                            some college
                                                           free/reduced
8
      male
                   group D
                                             high school
                                                           free/reduced
9
    female
                                             high school
                                                           free/reduced
                   group B
10
      male
                   group C
                                     associate's degree
                                                               standard
11
      male
                   group D
                                     associate's degree
                                                               standard
12
    female
                                             high school
                                                               standard
                   group B
13
      male
                   group A
                                            some college
                                                               standard
```

14	female	group A		ster's degree	standard
15 16	female male	group C group C	SOM	e high school high school	standard standard
17 18	female male	group B group C		e high school ster's degree	free/reduced free/reduced
19	female	group C		iate's degree	free/reduced
	test_prepara	ation_course	math digree	reading digre	e writing score
0		none	72	72.	0 74
1		completed	69	90.	0 88
2		none	90	95.	93
3		none	47	57.	0 44
4		none	76	78.	0 75
5		none	71	83.	0 78
6		completed	88	95.	92
7		none	40	43.	0 39
8		completed	64	64.	0 67
9		none	38	60.	0 50
10		none	58	54.	0 52
11		none	40	52.	0 43
12		none	65	81.	0 73
13		completed	78	72.	0 70
14		NaN	50	53.	0 58
15		none	69	75.	0 78
16		none	88	Na	N 86
17		none	18	32.	0 28
18		completed	46	42.	0 46
19		none	54	58.	0 61
	1				
da	ntaset.tail()				

	sex	race_ethnicity p	arental_level	_of_education	lunch
20999	female	group E	ma	ster's degree	standard
21000	male	group C		high school	free/reduced
21001	female	group C		high school	free/reduced
21002	female	group D		some college	standard
21003	female	group D		some college	free/reduced
	+ +			mandina diama	
score	test_pre	eparation_course	math digree	reading digre	e writing
20999		completed	58	99.	0
95		P			
21000 55		none	62	55.	0
21001		completed	49	71.	0
65 21002		completed	68	48.	0
77					-1
21003		none	77	89.	0
86					

dataset.tail(20)

	sex	race_ethnicity	<pre>parental_level_of_education</pre>	lunch
\				
20984	male	group B	some high school	standard
20985	انثى	group A	some college	standard
20986	انثی	group A	some college	standard
20987	انثى	group A	some college	standard
20988	NaN	group C	some high school	standard
20989	male	aroup A	high school	standard
20909	mate	group A	nigh school	Stalluaru
20990	female	group C	associate's degree	standard
20991	male	group E	some high school	standard
20992	female	group A	some high school	free/reduced
20993	female	group D	some college	free/reduced
20994	male	group E	high school	free/reduced
20995	female	group B	some high school	standard

20996	female	group D	associate's	s degree	free/reduced
20997	female	group D	bachelor's	s degree	free/reduced
20998	male	group A	higl	n school	standard
20999	female	group E	master's	s degree	standard
21000	male	group C	higl	n school	free/reduced
21001	female	group C	higl	n school	free/reduced
21002	female	group D	some	college	standard
21003	female	group D		college	free/reduced
	. 0	g. 0 ap 1	J 00	0011090	
	test_prep	aration_course	math digree read:	ing digre	e writing
score 20984		completed	79	85.	0
86 20985		completed	77	87.	0
91 20986		completed	77	87.	Θ
91 20987		completed	77	87.	
91		·			
20988 72		none	74	75.	
20989 57		none	57	51.	0
20990 51		none	40	57.	0
20991 76		completed	81	76.	0
20992		none	44	75.	0
45 20993		completed	67	86.	0
83 20994		completed	86	87.	Θ
75 20995		completed	67	82.	
78 20996		·	75	76.	
76		none			
20997 74		none	67	72.	Θ
20998 65		none	63	63.	0
20999		completed	58	99.	0

```
95
21000
                                         62
                                                       55.0
                         none
55
                                         49
21001
                    completed
                                                       71.0
65
21002
                    completed
                                         68
                                                       48.0
77
21003
                                         77
                                                       89.0
                         none
86
 dataset.columns
Index(['sex', 'race_ethnicity', 'parental_level_of_education',
'lunch',
       'test preparation course', 'math digree', 'reading digree',
       'writing score'],
      dtype='object')
 dataset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21004 entries, 0 to 21003
Data columns (total 8 columns):
 #
     Column
                                   Non-Null Count
                                                   Dtype
     -----
 0
                                   21002 non-null
                                                   object
     sex
 1
     race ethnicity
                                   20962 non-null
                                                   object
 2
     parental level of education
                                  20962 non-null
                                                   object
 3
     lunch
                                  21004 non-null
                                                   object
 4
     test_preparation_course
                                  20983 non-null
                                                   object
 5
                                  21004 non-null
     math digree
                                                   int64
 6
     reading digree
                                  20983 non-null
                                                   float64
 7
                                  21004 non-null int64
     writing score
dtypes: float64(1), int64(2), object(5)
memory usage: 1.3+ MB
 dataset.describe()
        math digree
                     reading digree
                                     writing score
count
       21004.000000
                       20983.000000
                                       21004.000000
          66.105408
                          69.151027
                                          68.068511
mean
std
          15.169625
                          14.598066
                                          15.190079
min
           0.000000
                          17.000000
                                          10.000000
25%
          57,000000
                                          58,000000
                          59.000000
50%
          66.000000
                          70.000000
                                          69.000000
75%
          77.000000
                          79.000000
                                          79.000000
         100.000000
                         100.000000
                                         100.000000
max
dataset.select dtypes('int64').columns
Index(['math digree', 'writing score'], dtype='object')
```

```
dataset.sex.unique()
array(['female', 'male', 'ذكر', 'انثى', nan], dtype=object)
 dataset.sex.value counts()
sex
female
          10875
          10119
male
6
       انثى
2
       ذکر
Name: count, dtype: int64
 dataset.lunch.unique()
array(['standard', 'free/reduced'], dtype=object)
dataset.lunch.value counts()
lunch
standard
                13549
free/reduced
                 7455
Name: count, dtype: int64
 dataset.rename(columns={'math digree':'math digree'}, inplace=True)
 dataset.rename(columns={'reading digree':'reading digree'},
inplace=True)
 dataset.rename(columns={'writing score':'writing score'},
inplace=True)
 dataset
          sex race ethnicity parental_level_of_education
                                                                   lunch
/
       female
                     group B
                                        bachelor's degree
                                                                standard
                                                                standard
1
       female
                                             some college
                     group C
       female
                         NaN
                                          master's degree
                                                                standard
3
         male
                     group A
                                       associate's degree free/reduced
         male
                     group C
                                                       NaN
                                                                standard
20999
      female
                                          master's degree
                     group E
                                                                standard
21000
         male
                                              high school free/reduced
                     group C
21001
      female
                                              high school free/reduced
                     group C
21002
      female
                     group D
                                             some college
                                                                standard
```

21003	female	group D		some college	free/reduced
	test prepa	aration_course	math digree	reading_digre	e
	g_score		<u>_</u> g.		_
0	3_	none	72	72.	0
74					
1		completed	69	90.	0
88					_
2		none	90	95.	Θ
93 3		nono	47	57	0
3 44		none	47	57.	ט
44		none	76	78.	O.
7 5		Hone	70	701	O
					_
20999		completed	58	99.	9
95		nono	62	55	0
21000 55		none	62	55.	U
21001		completed	49	71.	Θ
65		compreced	15	, 21	•
21002		completed	68	48.	0
77		•			
21003		none	77	89.	0
86					
[21004	rows x 8	columns]			

Dataset Cleaning

```
dataset.isnull().sum()
                                    2
race_ethnicity
                                    42
parental_level_of_education
                                    42
lunch
                                     0
test_preparation_course
                                    21
math_digree
                                     0
reading_digree
                                    21
writing_score
dtype: int64
data = {
    'A': [1, 2, 3, None, 5], 
'B': [None, 2, 3, 4, 5],
```

```
'C': [1, 2, None, None, 5]
df = pd.DataFrame(data)
print("Original Data:\n",df)
print()
df cleaned = df.dropna()
print("Cleaned Data:\n",df cleaned)
Original Data:
          В
             C
     Α
   1.0 NaN 1.0
1
  2.0
       2.0
            2.0
2
  3.0
       3.0 NaN
3
  NaN 4.0 NaN
4 5.0 5.0 5.0
Cleaned Data:
                C
          В
   2.0 2.0 2.0
4 5.0 5.0 5.0
myFirestCopy=dataset.copy()
myFirestCopy
          sex race ethnicity parental level of education
                                                                 lunch
0
       female
                                       bachelor's degree
                                                              standard
                     group B
       female
                     group C
                                            some college
                                                              standard
       female
                         NaN
                                         master's degree
                                                              standard
         male
                     group A
                                      associate's degree free/reduced
         male
                     group C
                                                     NaN
                                                              standard
                                         master's degree
20999
      female
                     group E
                                                              standard
21000
         male
                     group C
                                             high school free/reduced
21001
      female
                     group C
                                             high school free/reduced
21002
      female
                     group D
                                            some college
                                                              standard
                                            some college free/reduced
21003 female
                     group D
      test_preparation_course math_digree reading_digree
writing score
```

0		none	72	72.0)		
74 1		completed	69	90.0)		
88							
2 93		none	90	95.0)		
3		none	47	57.0)		
44		nono	76	70 0	.		
4 75		none	70	78.0)		
20999 95		completed	58	99.0)		
21000 55		none	62	55.0)		
21001 65		completed	49	71.0)		
21002		completed	68	48.0)		
77 21003		none	77	89.0)		
86							
<pre>[21004 rows x 8 columns] df_cleaned = myFirestCopy.dropna() print("Cleaned Data:\n",df_cleaned)</pre>							
print	("Cleaned D						
print	("Cleaned Ded Data:		al_level_of_e	ducation			
print Cleane lunch	c("Cleaned Ded Data: sex ra	<pre>ata:\n",df_cleaned) ce_ethnicity parenta</pre>			standard		
print Cleane	d Data: sex ra	ata:\n",df_cleaned)	al_level_of_ed bachelor's		standard		
print Cleane lunch	c("Cleaned Ded Data: sex ra	<pre>ata:\n",df_cleaned) ce_ethnicity parenta</pre>	bachelor's		standard standard		
print Cleane lunch 0	c("Cleaned Ded Data: sex ra female	ata:\n",df_cleaned) ce_ethnicity parenta group B	bachelor's	degree college			
print Cleane lunch 0	c("Cleaned Ded Data: sex ra female female	ata:\n",df_cleaned) ce_ethnicity parenta group B group C	bachelor's	degree college degree	standard		
print Cleane lunch 0 1	c("Cleaned Deta: sex ra female female male	ata:\n",df_cleaned) ce_ethnicity parenta group B group C group A	bachelor's some of associate's	degree college degree	standard free/reduced		
print Cleane lunch 0 1 3	d Data: sex ra female female male female	ata:\n",df_cleaned) ce_ethnicity parenta group B group C group A group B	bachelor's some of associate's	degree college degree degree	standard free/reduced standard		
print Cleane lunch 0 1 3 5	d Data: sex ra female female male female	ata:\n",df_cleaned) ce_ethnicity parenta group B group C group A group B	bachelor's some of associate's	degree college degree degree college	standard free/reduced standard		
print Cleane lunch 0 1 3 5	female female female female female female	ata:\n",df_cleaned) ce_ethnicity parenta group B group C group A group B group B	bachelor's some of associate's associate's some of	degree college degree degree college	standard free/reduced standard standard		
print Cleane lunch 0 1 3 5 6 20999	female female female female female female female female female	ata:\n",df_cleaned) ce_ethnicity parenta group B group C group A group B group B group E	bachelor's some of associate's associate's some of master's high	degree college degree college degree	standard free/reduced standard standard standard		

21003	female	group D		some college	free/reduced
	test_prepara g_score	ation_course	math_digree	reading_digree	
0	9_500.0	none	72	72.0)
74 1		completed	69	90.0)
88		·			
3 44		none	47	57.0)
5		none	71	83.6)
78 6		completed	88	95.0	a e
92		Completed	88	95.0	
20999		completed	58	99.6)
95		·			
21000 55		none	62	55.0)
21001		completed	49	71.6)
65 21002		completed	68	48.0	1
77		completed	00		
21003		none	77	89.0)
86					
[20897	rows x 8 c	olumns]			
df_cl	eaned.isnul	l(). <mark>sum</mark> ()			
parent lunch test_p math_d	thnicity al_level_of reparation_d igree g_digree	_	0 0 0 0 0 0		
	g_score		0		
'B	': [1, 2, 3 ': [None, 2				
df = print	pd.DataFram ("Original <i>ling NaN va</i>	Data:\n", df)			

```
df.fillna(0, inplace=True)
 print("\nData after filling NaN with 0:\n", df)
Original Data:
           В
                C
      Α
   1.0 NaN 1.0
1
   2.0
        2.0
             2.0
2
  3.0
       3.0
            NaN
3
  NaN 4.0
             NaN
  5.0
        5.0 5.0
Data after filling NaN with 0:
           В
      Α
                C
   1.0 0.0 1.0
   2.0
        2.0
             2.0
  3.0
       3.0
             0.0
3
  0.0
       4.0
             0.0
4 5.0 5.0 5.0
dataset.isnull().sum()
                                 2
sex
                                42
race ethnicity
parental level of education
                                42
lunch
                                 0
test preparation course
                                21
math_digree
                                 0
reading digree
                                21
writing_score
                                 0
dtype: int64
mis=dataset[dataset['reading digree'].isna()]
mis
        sex race ethnicity parental level of education
                                                             lunch \
16
       male
                   group C
                                            high school
                                                          standard
1016
       male
                   group C
                                            high school
                                                          standard
                                            high school
2016
       male
                   group C
                                                          standard
3016
       male
                   group C
                                            high school
                                                          standard
                                            high school
4016
       male
                   group C
                                                          standard
5016
                                                          standard
       male
                   group C
                                            high school
6016
       male
                                            high school
                                                          standard
                   group C
7016
       male
                                            high school
                                                          standard
                   group C
8016
       male
                   group C
                                            high school
                                                          standard
9016
       male
                   group C
                                            high school
                                                          standard
10016
       male
                   group C
                                            high school
                                                          standard
11016
                                            high school
                                                          standard
       male
                   group C
12016
       male
                   group C
                                            high school
                                                          standard
13016
       male
                                            high school
                                                          standard
                   group C
14016
       male
                                            high school
                                                          standard
                   group C
```

15016 16016 17016 18016 19018 20018	male male male male male male	group C group C group C group C group C		high school high school high school high school high school	standard standard standard standard
	test_prepara	tion_course	math_digree	reading_dig	ıree
16	g_score	none	88		NaN
86					
1016		none	88		NaN
86 2016		none	88		NaN
86		none	00		Ivaiv
3016		none	88		NaN
86					
4016		none	88		NaN
86					
5016		none	88		NaN
86					
6016		none	88		NaN
86					
7016		none	88		NaN
86			0.0		NaN
8016		none	88		NaN
86 9016		nono	88		NaN
86		none	00		Ivaiv
10016		none	88		NaN
86		Hone	00		IVAIV
11016		none	88		NaN
86					
12016		none	88		NaN
86					
13016		none	88		NaN
86					
14016		none	88		NaN
86					
15016		none	88		NaN
86					
16016		none	88		NaN
86			0.0		NeN
17016 86		none	88		NaN
18016		nono	88		NaN
86		none	00		IVAIV
19018		none	88		NaN
86		110116	30		

```
20018
                                          88
                                                          NaN
                          none
86
mySecondCopy=dataset.copy()
mySecondCopy.loc[:, "reading digree"] =
mySecondCopy["reading digree"].fillna(20.0)
mySecondCopy.isnull().sum()
sex
                                 2
                                 42
race ethnicity
parental level of education
                                 42
lunch
                                  0
test preparation course
                                 21
                                  0
math digree
                                  0
reading digree
writing score
                                  0
dtype: int64
mis=mySecondCopy[mySecondCopy['reading digree']==20.0]
mis
        sex race ethnicity parental level of education
                                                              lunch
                                                                      1
16
                                             hiah school
       male
                    group C
                                                           standard
1016
       male
                    group C
                                             high school
                                                           standard
2016
       male
                                             high school
                                                           standard
                    group C
3016
       male
                    group C
                                             high school
                                                           standard
4016
       male
                    group C
                                             high school
                                                           standard
5016
       male
                    group C
                                             high school
                                                           standard
6016
       male
                    group C
                                             high school
                                                           standard
                    group C
7016
                                             high school
                                                           standard
       male
8016
       male
                    group C
                                             high school
                                                           standard
9016
       male
                    group C
                                             high school
                                                           standard
       male
                                             high school
                                                           standard
10016
                    group C
11016
       male
                    group C
                                             high school
                                                           standard
12016
                                             high school
                                                           standard
       male
                    group C
13016
       male
                    group C
                                             high school
                                                           standard
14016
                                             high school
                                                           standard
       male
                    group C
15016
                                             high school
       male
                    group C
                                                           standard
16016
       male
                    group C
                                             high school
                                                           standard
17016
       male
                    group C
                                             high school
                                                           standard
18016
       male
                    group C
                                             high school
                                                           standard
19018
       male
                                             high school
                                                           standard
                    group C
20018
       male
                    group C
                                             high school
                                                           standard
      test preparation course math digree reading digree
writing score
16
                          none
                                          88
                                                         20.0
86
                                          88
1016
                          none
                                                         20.0
86
```

2016	none	88	20.0
86			
3016	none	88	20.0
86			
4016	none	88	20.0
86			
5016	none	88	20.0
86			
6016	none	88	20.0
86			
7016	none	88	20.0
86			
8016	none	88	20.0
86			
9016	none	88	20.0
86			
10016	none	88	20.0
86			
11016	none	88	20.0
86			
12016	none	88	20.0
86			
13016	none	88	20.0
86			
14016	none	88	20.0
86			
15016	none	88	20.0
86			
16016	none	88	20.0
86			
17016	none	88	20.0
86			
18016	none	88	20.0
86			
19018	none	88	20.0
86			
20018	none	88	20.0
86			

mySecondCopy.iloc[16,1]='group A'
mySecondCopy.head(20)

	sex	<pre>race_ethnicity</pre>	<pre>parental_level_of_education</pre>	lunch	\
0	female	group B	bachelor's degree	standard	
1	female	group C	some college	standard	
2	female	NaN	master's degree	standard	
3	male	group A	associate's degree	free/reduced	
4	male	group C	NaN	standard	
5	female	group B	associate's degree	standard	
6	female	group B	some college	standard	

7 8 9 10 11 12 13 14 15 16 17 18 19	male male female male female female female female female female	group B group D group C group D group B group A group A group C group A group C group C group C	associ mas some some mas	some college high school high school tate's degree high school some college ster's degree high school high school high school thigh school ster's degree tate's degree tate's degree	free/reduced free/reduced free/reduced standard standard standard standard standard free/reduced free/reduced free/reduced	
	test_prepar	ration_course	math_digree	reading_digre	e writing_sc	ore
0		none	72	72.	0	74
1		completed	69	90.	0	88
2		none	90	95.	0	93
3		none	47	57.	0	44
4		none	76	78.	0	75
5		none	71	83.	0	78
6		completed	88	95.	0	92
7		none	40	43.	0	39
8		completed	64	64.	0	67
9		none	38	60.	0	50
10		none	58	54.	0	52
11		none	40	52.	0	43
12		none	65	81.	0	73
13		completed	78	72.	0	70
14		NaN	50	53.	0	58
15		none	69	75.	0	78
16		none	88	20.	0	86
17		none	18	32.		28

```
18
                  completed
                                       46
                                                       42.0
                                                                         46
19
                                       54
                                                                         61
                       none
                                                       58.0
data = {
    'A': [1, 2, np.nan, 4, 5],
    'B': [np.nan, 2, 3, 4, 5],
    'C': [1, 2, 3, np.nan, 5],
    'D': [1, 2, 3, 4, 5]
 }
df = pd.DataFrame(data)
 df copy=df.copy()
print(f'The Data Before \n{df}')
 print('-----
The Data Before
                C
                   D
     Α
          В
   1.0
                   1
0
       NaN
             1.0
1
  2.0
        2.0
             2.0
                   2
  NaN
        3.0
             3.0
                   3
3
  4.0
       4.0
             NaN
                   4
   5.0
        5.0
                   5
             5.0
dataset.describe()
        math digree
                      reading digree
                                       writing score
       21004.000000
                                        21004.\overline{0}00000
                        20983.000000
count
          66.105408
                            69.151027
                                            68.068511
mean
          15.169625
                            14.598066
                                            15.190079
std
min
           0.000000
                            17.000000
                                            10.000000
25%
          57,000000
                            59.000000
                                            58.000000
          66.000000
                            70.000000
                                            69.000000
50%
75%
          77.000000
                           79.000000
                                            79.000000
         100.000000
                          100.000000
                                           100.000000
max
mis=dataset[dataset["reading digree"].isna()]
mis
        sex race_ethnicity parental_level_of_education
                                                               lunch
16
                                              high school
                                                            standard
       male
                    group C
1016
       male
                    group C
                                              high school
                                                            standard
2016
       male
                    group C
                                              high school
                                                            standard
3016
       male
                    group C
                                              high school
                                                            standard
4016
       male
                    group C
                                              high school
                                                            standard
5016
       male
                    group C
                                              high school
                                                            standard
6016
       male
                    group C
                                              high school
                                                            standard
7016
       male
                    group C
                                              high school
                                                            standard
8016
       male
                    group C
                                              high school
                                                            standard
```

9016 10016 11016 12016 13016 14016 15016 16016 17016 18016 19018 20018	male male male male male male male male	group C		high school	standard standard standard standard standard standard standard
	test_prepara	tion_course	math_digree	reading_dig	gree
	g_score				
16		none	88		NaN
86		nana	0.0		NoN
1016 86		none	88		NaN
2016		none	88		NaN
86		Hone	00		IVAIV
3016		none	88		NaN
86					
4016		none	88		NaN
86					
5016		none	88		NaN
86					
6016		none	88		NaN
86 7016		nono	88		NaN
86		none	00		Ivaiv
8016		none	88		NaN
86					T.G.T.
9016		none	88		NaN
86					
10016		none	88		NaN
86					
11016		none	88		NaN
86			00		NeN
12016 86		none	88		NaN
13016		none	88		NaN
86		Hone	00		IVAIV
14016		none	88		NaN
86					
15016		none	88		NaN
86					
16016		none	88		NaN
86					

17016 86			none		88		NaN	
18016			none		88		NaN	
86 19018			none		88		NaN	
86 20018			none		88		NaN	
86			Horic		00		Null	
mis["r readir	<pre>mis=dataset.copy() mis["reading_digree"]=dataset["reading_digree"].fillna(value=dataset[" reading_digree"].min()) mis.head(20)</pre>							
0 fe	sex race_e emale	thnicity group l	y parental_			ucation degree	lunch standard	-
$1 ext{ f} \epsilon$	emale	group	С		some	college	standard	
2 fe	emale male	Nal group <i>l</i>				degree degree	standard free/reduced	
4	male	group (C			NaN	standard	
	emale emale	group l group l		assoc		degree college	standard standard	
7	male	group I				college	free/reduced	
8	male	group I	ס		high	school	free/reduced	
9 fe	emale male	group (2000		school degree	free/reduced standard	
11	male	group (degree	standard	
12 fe	emale	group I	В		high	school	standard	
13 14 fe	male emale	group /		m		college	standard	
	emale	group (degree school	standard standard	
16	male	group	С		high	school	standard	
	emale	group I				school	free/reduced	
18 19 fe	male emale	group (degree degree	<pre>free/reduced free/reduced</pre>	
test_preparation_course math_digree reading_digree writing_score								
0		noi	ne	72		72.	0	74
1		complete	ed	69		90.	0	88
2		noı	ne	90		95.	0	93
3		noı	ne	47		57.	0	44
4		noı	ne	76		78.	0	75
5		noı	ne	71		83.	0	78

6 completed 88 95.0 92 7 none 40 43.0 39 8 completed 64 64.0 67 9 none 38 60.0 50 10 none 58 54.0 52 11 none 40 52.0 43 12 none 65 81.0 73 13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard						
8 completed 64 64.0 67 9 none 38 60.0 50 10 none 58 54.0 52 11 none 40 52.0 43 12 none 65 81.0 73 13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female group A as	6		completed	88	95.0	92
9	7		none	40	43.0	39
10 none 58 54.0 52 11 none 40 52.0 43 12 none 65 81.0 73 13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree standard 3 male group C NaN standard	8		completed	64	64.0	67
11 none 40 52.0 43 12 none 65 81.0 73 13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group B bachelor's degree standard 2 female group C some college standard 3 male group A associate's degree free/reduced 20999	9		none	38	60.0	50
12 none 65 81.0 73 13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree free/reduced 4 male group C NaN standard 20999 female group E master's degree standard 21000	10		none	58	54.0	52
13 completed 78 72.0 70 14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree standard 3 male group C NaN standard 20999 female group E master's degree standard 1000 male group C high school free/reduced	11		none	40	52.0	43
14 NaN 50 53.0 58 15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree standard 3 male group C NaN standard 20999 female group E master's degree standard 20999 female group C high school free/reduced 21001	12		none	65	81.0	73
15 none 69 75.0 78 16 none 88 17.0 86 17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree standard 3 male group A associate's degree free/reduced 4 male group C NaN standard 20999 female group E master's degree standard 21000 male group C high school free/reduced	13		completed	78	72.0	70
none 88 17.0 86 none 18 32.0 28 scompleted 46 42.0 46 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard male group C NaN standard male group C high school free/reduced	14		NaN	50	53.0	58
17 none 18 32.0 28 18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch 0 female group B bachelor's degree standard 1 female group C some college standard 2 female NaN master's degree standard 3 male group A associate's degree free/reduced 4 male group C NaN standard 20999 female group E master's degree standard 21000 male group C high school free/reduced	15		none	69	75.0	78
18 completed 46 42.0 46 19 none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard male group E master's degree standard proup C NaN standard high school free/reduced free/reduced	16		none	88	17.0	86
none 54 58.0 61 dataset sex race_ethnicity parental_level_of_education lunch female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard group E master's degree standard high school free/reduced female group C high school free/reduced	17		none	18	32.0	28
sex race_ethnicity parental_level_of_education lunch female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard male group C NaN standard male group E master's degree standard high school free/reduced female group C high school free/reduced	18		completed	46	42.0	46
sex race_ethnicity parental_level_of_education lunch female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard male group E master's degree standard male group E high school free/reduced proup C high school free/reduced female group C high school free/reduced	19		none	54	58.0	61
female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard male group E master's degree standard male group E high school free/reduced free/reduced free/reduced free/reduced	datase	t				
female group B bachelor's degree standard female group C some college standard female NaN master's degree standard male group A associate's degree free/reduced male group C NaN standard group E master's degree standard high school free/reduced free/reduced free/reduced free/reduced		sex	race_ethnicity	parental_leve	el_of_education	lunch
2 female NaN master's degree standard 3 male group A associate's degree free/reduced 4 male group C NaN standard 20999 female group E master's degree standard 21000 male group C high school free/reduced 21001 female group C high school free/reduced	0	female	group B	bac	chelor's degree	standard
3 male group A associate's degree free/reduced 4 male group C NaN standard 20999 female group E master's degree standard 21000 male group C high school free/reduced 21001 female group C high school free/reduced	1	female	group C		some college	standard
4 male group C NaN standard 20999 female group E master's degree standard 21000 male group C high school free/reduced 21001 female group C high school free/reduced	2	female	NaN	r	master's degree	standard
20999 female group E master's degree standard 21000 male group C high school free/reduced 21001 female group C high school free/reduced	3	male	group A	asso	ociate's degree	free/reduced
21000 male group C high school free/reduced 21001 female group C high school free/reduced	4	male	group C		NaN	standard
21000 male group C high school free/reduced 21001 female group C high school free/reduced						
21001 female group C high school free/reduced	20999	female	group E	ſ	master's degree	standard
	21000	male	group C		high school	free/reduced
21002 female group D some college standard	21001	female	group C		high school	free/reduced
	21002	female	group D		some college	standard

21003	female	group D	s.c	ome college	free/reduced	
21005	Tellia ce	group b	30	me correge	Tree/reduced	
weitin		eparation_course	math_digree re	eading_digre	е	
0	ig_score	none	72	72.0	9	
74 1		completed	69	90.0	9	
88 2		none	90	95.0	9	
93 3		none	47	57.0	a 1	
44						
4 75		none	76	78.0	9	
20999 95		completed	58	99.0	9	
21000 55		none	62	55.0	9	
21001		completed	49	71.0	9	
65 21002		completed	68	48.0	9	
77 21003		none	77	89.0	9	
86						
[21004	rows x	8 columns]				
<pre>dup=dataset[dataset.duplicated()] dup</pre>						
\	sex	race_ethnicity p	arental_level_of	f_education	lunch	
1000	female	group B	bachelo	or's degree	standard	
1001	female	group C	so	ome college	standard	
1002	female	NaN	maste	er's degree	standard	
1003	male	group A	associat	te's degree	free/reduced	
1004	male	group C		NaN	standard	
20984	male	group B	some h	nigh school	standard	

some college

standard

انثی

group A

20986

20987 20993	انثی female	group A group D	some colle some	ege college	standard free/reduced	
21000	male	group C	high	school	free/reduced	
	test_prep g score	paration_course ma	th_digree readi	ng_digre	e	
1000 74	9_50010	none	72	72.	0	
1001 88		completed	69	90.		
1002 93		none	90	95.		
1003 44		none	47	57.		
1004 75		none	76	78.	0	
20984 86		completed	79	85.	0	
20986 91		completed	77	87.	9	
20987 91		completed	77	87.	0	
20993		completed	67	86.	0	
83 21000		none	62	55.	9	
[19858 rows x 8 columns]						
<pre>dropDu=dataset.drop_duplicates() dropDu</pre>						
G G		race_ethnicity pare	ntal_level_of_ed	ucation	lunch	
0	female	group B	bachelor's	degree	standard	
1	female	group C	some	college	standard	
2	female	NaN	master's	degree	standard	
3	male	group A	associate's	degree	free/reduced	
4	male	group C		NaN	standard	
20998	male	group A	high	school	standard	

```
20999 female
                     group E
                                         master's degree standard
      female
21001
                     group C
                                             high school free/reduced
21002 female
                                            some college
                     group D
                                                               standard
                                            some college free/reduced
21003 female
                     group D
      test preparation course math digree
                                            reading digree
writing score
                         none
                                        72
                                                       72.0
74
                                        69
                                                       90.0
1
                    completed
88
                                                       95.0
2
                                        90
                         none
93
3
                         none
                                        47
                                                       57.0
44
4
                                        76
                                                       78.0
                         none
75
. . .
                                                       . . .
20998
                                        63
                                                       63.0
                         none
65
20999
                    completed
                                        58
                                                       99.0
95
                                        49
21001
                    completed
                                                       71.0
65
                                                       48.0
21002
                    completed
                                        68
77
21003
                                        77
                                                       89.0
                         none
86
[1146 rows x 8 columns]
data = {
    'Country': ['USA', 'Canada', 'Australia', 'Germany', 'Japan'],
    'Date': ['2023-07-20', '2023-07-21', '2023-07-22', '2023-07-23',
'2023-07-24'1.
    'Temperature': [25.5, '28.0', 30.2, 22.8, 26.3]
df = pd.DataFrame(data)
print(df.Temperature.unique())
print('-----
[25.5 '28.0' 30.2 22.8 26.3]
```

```
df['Temperature'] = df['Temperature'].astype(float)
print(df.Temperature.unique())
 print('-----')
 print(df.info())
print('----')
# calculate the mean temperature
mean temperature = df['Temperature'].mean()
print(mean temperature)
[25.5 28. 30.2 22.8 26.3]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 3 columns):
    Column Non-Null Count
                                 Dtype
    Country 5 non-null Date 5 non-null
0
                                 object
1
                                 object
    Temperature 5 non-null float64
2
dtypes: float64(1), object(2)
memory usage: 252.0+ bytes
None
26.560000000000000
df = pd.DataFrame({'date': ['2022-12-01', '01/02/2022', '2022-03-23',
'03/02/2022', '3 4 2023', '2023.9.30']})
df['date'] = pd.to datetime(df['date'], format='mixed',
dayfirst=True)
print(df)
       date
0 2022-12-01
1 2022-02-01
2 2022-03-23
3 2022-02-03
4 2023-04-03
5 2023-09-30
 data = {
    'Name': ['John', 'Michael', 'Tom', 'Alex', 'Ryan'],
    'Age': [8, 9, 7, 80, 100],
'Gender': ['M', 'M', 'M', 'F', 'M'],
    'Standard': [3, 4, 12, 3, 5]
df = pd.DataFrame(data)
 df.loc[3, 'Gender'] = 'M'
 print(df)
```

```
Name Age Gender
                          Standard
0
      John
               8
                                 3
1
   Michael
               9
                      М
                                 4
2
                                 12
       Tom
              7
                      М
3
      Alex
              80
                      М
                                 3
                                 5
      Ryan 100
                      М
 data = {
    'Name': ['John', 'Michael', 'Tom', 'Alex', 'Ryan'],
    'Age': [8, 9, 7, 80, 100],
'Gender': ['M', 'M', 'M', 'M', 'M'],
    'Standard': [3, 4, 12, 3, 5]
 }
 df = pd.DataFrame(data)
 for i in df.index:
    age val = df.loc[i, 'Age']
    if (age val > \frac{14}{10}) and (age val%\frac{10}{10} == 0):
        df.loc[i, 'Age'] = age val/10
 print(df)
      Name
             Age Gender
                          Standard
0
              8
      John
                      М
                                 3
1
                                 4
   Michael
               9
                      М
2
               7
                                 12
                      М
       Tom
3
              8
                                 3
      Alex
                      Μ
                                 5
      Ryan
            10
                      М
 len(dataset)
21004
 dataset.sex.unique()
array(['female', 'male', 'ذكر', 'انثى', nan], dtype=object)
 for i in range(0,len(dataset)):
    if dataset.loc[i, 'sex']=='شنا':
          dataset.loc[i,'sex']='female'
dataSex=dataset[dataset['sex']=='شنا']
 dataSex
Empty DataFrame
Columns: [sex, race_ethnicity, parental_level_of_education, lunch,
test_preparation_course, math_digree, reading_digree, writing_score]
Index: []
dataSex=dataset[dataset['sex']=='ركذ']
dataSex
Empty DataFrame
Columns: [sex, race ethnicity, parental level of education, lunch,
```

```
test_preparation_course, math_digree, reading_digree, writing_score]
Index: []

for i in range(0,len(dataset)):
    if dataset.loc[i,'sex']=='\(\frac{1}{2}\)':
        dataset.loc[i,'sex']='Male'

dataset.to_csv('datanew2.csv')
```