Pandas-intro

March 15, 2020

0.0.1 Step 4

[2]:
```
import pandas as pd
import numpy as np
```

[4]:
```
students_performance = pd.read_csv("../data/StudentsPerformance.csv")
```

[5]:
```
students_performance
```

[5]:
```
     gender 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 B            master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      …            …                          …           …
995  female        group E            master's degree      standard
996    male        group C                high school  free/reduced
997  female        group C                high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test preparation course  math score  reading score  writing score
0                      none          72             72             74
1                 completed          69             90             88
2                      none          90             95             93
3                      none          47             57             44
4                      none          76             78             75
..                      …         …           …            …
995               completed          88             99             95
996                    none          62             55             55
997               completed          59             71             65
998               completed          68             78             77
999                    none          77             86             86

[1000 rows x 8 columns]
```

[7]:
```
students_performance.head(10)
```

[7]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50

[8]:
```
students_performance.tail(10)
```

[8]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
990	male	group E	high school	free/reduced	completed	86	81	75
991	female	group B	some high school	standard	completed	65	82	78
992	female	group D	associate's degree	free/reduced	none	55	76	76
993	female	group D	bachelor's degree	free/reduced	none	62	72	74
994	male	group A	high school	standard	none	63	63	62
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced				
997	female	group C	high school	free/reduced				
998	female	group D	some college	standard				
999	female	group D	some college	free/reduced				

```
996                  none          62             55             55
997             completed          59             71             65
998             completed          68             78             77
999                  none          77             86             86
```

[9]:
```
students_performance.describe()
```

[9]:
```
       math score  reading score  writing score
count  1000.00000    1000.000000    1000.000000
mean     66.08900      69.169000      68.054000
std      15.16308      14.600192      15.195657
min       0.00000      17.000000      10.000000
25%      57.00000      59.000000      57.750000
50%      66.00000      70.000000      69.000000
75%      77.00000      79.000000      79.000000
max     100.00000     100.000000     100.000000
```

[10]:
```
students_performance.dtypes
```

[10]:
```
gender                        object
race/ethnicity                object
parental level of education   object
lunch                         object
test preparation course       object
math score                     int64
reading score                  int64
writing score                  int64
dtype: object
```

[11]:
```
students_performance.shape
```

[11]:
```
(1000, 8)
```

[12]:
```
students_performance.groupby('gender').aggregate({'writing score' : 'mean'})
```

[12]:
```
        writing score
gender
female      72.467181
male        63.311203
```

[13]:
```
students_performance.size
```

[13]:
```
8000
```

0.0.2 Step 5

[14]:
```
students_performance.iloc[0:5, 0:3]
```

[14]:
```
   gender race/ethnicity parental level of education
0  female        group B           bachelor's degree
1  female        group C                some college
2  female        group B             master's degree
3    male        group A          associate's degree
4    male        group C                some college
```

[15]:
```
students_performance.iloc[[0, 3, 10], [0, 5, -1]]
```

[15]:
```
    gender  math score  writing score
0   female          72             74
3     male          47             44
10    male          58             52
```

0.0.3 Step 6

[16]:
```
students_performance_with_names = students_performance.iloc[[0, 3, 4, 7, 8]]
```

[17]:
```
students_performance_with_names
```

[17]:
```
   gender race/ethnicity parental level of education         lunch  \
0  female        group B           bachelor's degree      standard
3    male        group A          associate's degree  free/reduced
4    male        group C                some college      standard
7    male        group B                some college  free/reduced
8    male        group D                 high school  free/reduced

  test preparation course  math score  reading score  writing score
0                    none          72             72             74
3                    none          47             57             44
4                    none          76             78             75
7                    none          40             43             39
8               completed          64             64             67
```

[18]:
```
students_performance_with_names.index = ["Cersei", "Tywin", "Gregor",␣
↪"Joffrey", "Ilyn Payne"]
```

[19]:
```
students_performance_with_names
```

[19]:
```
            gender race/ethnicity parental level of education         lunch  \
Cersei      female        group B           bachelor's degree      standard
Tywin         male        group A          associate's degree  free/reduced
```

```
Gregor        male        group C               some college      standard
Joffrey       male        group B               some college  free/reduced
Ilyn Payne    male        group D                high school  free/reduced

           test preparation course  math score  reading score  writing score
Cersei                        none          72             72             74
Tywin                         none          47             57             44
Gregor                        none          76             78             75
Joffrey                       none          40             43             39
Ilyn Payne               completed          64             64             67
```

[20]:
```
students_performance_with_names.loc[['Cersei', 'Joffrey'], ['gender', 'writing score']]
```

[20]:
```
          gender  writing score
Cersei    female             74
Joffrey     male             39
```

0.0.4 Step 7

[21]:
```
students_performance_with_names.iloc[: , 1]
```

[21]:
```
Cersei        group B
Tywin         group A
Gregor        group C
Joffrey       group B
Ilyn Payne    group D
Name: race/ethnicity, dtype: object
```

[22]:
```
type(students_performance_with_names.iloc[: , 1])
```

[22]:
```
pandas.core.series.Series
```

[24]:
```
pd.Series([1, 2, 3], index=["Cersei", "Tywin", "Gregor"])
```

[24]:
```
Cersei    1
Tywin     2
Gregor    3
dtype: int64
```

[25]:
```
my_series1 = pd.Series([1, 2, 3], index=["Cersei", "Tywin", "Gregor"])
my_series2 = pd.Series([4, 5, 6], index=["Cersei", "Tywin", "Gregor"])
```

[26]:
```
pd.DataFrame({'col_name_1' : my_series1, 'col_name_2' : my_series2})
```

[26]:
```
         col_name_1  col_name_2
Cersei            1           4
Tywin             2           5
Gregor            3           6
```

[27]:
```
students_performance_with_names["gender"]
```

[27]:
```
Cersei        female
Tywin           male
Gregor          male
Joffrey         male
Ilyn Payne      male
Name: gender, dtype: object
```

[28]:
```
students_performance_with_names[["gender"]]
```

[28]:
```
            gender
Cersei      female
Tywin         male
Gregor        male
Joffrey       male
Ilyn Payne    male
```

[29]:
```
students_performance_with_names[["gender"]].shape
```

[29]:
```
(5, 1)
```

[31]:
```
students_performance_with_names["gender"].shape
```

[31]:
```
(5,)
```

[35]:
```
students_performance_with_names.loc[:'Joffrey']
```

[35]:
```
          gender race/ethnicity parental level of education         lunch  \
Cersei    female        group B           bachelor's degree      standard
Tywin       male        group A          associate's degree  free/reduced
Gregor      male        group C                some college      standard
Joffrey     male        group B                some college  free/reduced

         test preparation course  math score  reading score  writing score
Cersei                      none          72             72             74
Tywin                       none          47             57             44
Gregor                      none          76             78             75
Joffrey                     none          40             43             39
```