190031920

A Nikhil Reddy

DS Practical 2

In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
filedata = pd.read_csv("./21B15600.csv")
file2data = pd.read_csv("./74B05600.csv")
```

In [3]:
```
filedata
```

Out[3]:

	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

In [4]:
```
file2data
```

Out[4]:

	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
...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

1000 rows × 8 columns

In [5]:
```
# 1. Gather Columns into rows
pd.melt(filedata)
```

Out[5]:

	variable	value
0	gender	female
1	gender	female
2	gender	female
3	gender	male
4	gender	male
...	...	...
67	writing score	75
68	writing score	78
69	writing score	92
70	writing score	39
71	writing score	67

72 rows × 2 columns

In [6]:
```
#2. Spread rows into columns
filedata.pivot(columns = 'gender', values = 'math score')
```

Out[6]:

gender	female	male
0	72.0	NaN
1	69.0	NaN
2	90.0	NaN
3	NaN	47.0
4	NaN	76.0
5	71.0	NaN
6	88.0	NaN
7	NaN	40.0
8	NaN	64.0

In [7]:
```
# 3. Append rows of Data Frames
pd.concat([filedata,file2data])
```

Out[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
...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

1009 rows × 8 columns

In [8]:
```
#4. Append columns of Data Frames
pd.concat([filedata,file2data], axis=1)
```

Out[8]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	gender	race/ethnicity
0	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0	female	group B
1	female	group C	some college	standard	completed	69.0	90.0	88.0	female	group C
2	female	group B	master's degree	standard	none	90.0	95.0	93.0	female	group B
3	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0	male	group A
4	male	group C	some college	standard	none	76.0	78.0	75.0	male	group C
...	...	...	...	...	...	...	...	...	...	...
995	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	female	group E
996	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	male	group C
997	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	female	group C
998	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	female	group D
999	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	female	group D

1000 rows × 16 columns

In [9]:
```
#5.     Order rows by values of a column
filedata.sort_values('math score')
```

Out[9]:

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

In [10]:
```
#6. Order rows by values of a column
filedata.sort_values('math score', ascending=False)
```

Out[10]:

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

In [11]:
```
#7. Sort the index of a Data Frame
filedata.sort_index()
```

Out[11]:

	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

In [12]:
```
#8. Reset index of Data Frame to row numbers, moving index to columns.
filedata.reset_index()
```

Out[12]:

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

In [13]:
```
#9.     Drop columns from Data Frame Subset observations
filedata.drop(columns=['gender','test preparation course'])
```

Out[13]:

	race/ethnicity	parental level of education	lunch	math score	reading score	writing score
0	group B	bachelor's degree	standard	72	72	74
1	group C	some college	standard	69	90	88
2	group B	master's degree	standard	90	95	93
3	group A	associate's degree	free/reduced	47	57	44
4	group C	some college	standard	76	78	75
5	group B	associate's degree	standard	71	83	78
6	group B	some college	standard	88	95	92
7	group B	some college	free/reduced	40	43	39
8	group D	high school	free/reduced	64	64	67

In [14]:
```
# 10. Extract rows that meet logical criteria.
filedata[filedata['reading score'] > 80]
```

Out[14]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92

In [15]:
```
#11. Remove duplicate rows (only considers columns)
filedata.drop_duplicates()
```

Out[15]:

	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

In [16]:
```
# 12.   Select first n rows
filedata.head(50)
```

Out[16]:

	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

In [17]:
```
# 12.   Select last n rows
filedata.tail(50)
```

Out[17]:

	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

In [18]:
```
# 13. Select single column with specific name.
filedata['parental level of education']
```

Out[18]:
```
0     bachelor's degree
1          some college
2       master's degree
3    associate's degree
4          some college
5    associate's degree
6          some college
7          some college
8           high school
Name: parental level of education, dtype: object
```

In [19]:
```
#14.    Select Multiple column with specific name.
filedata[['gender', 'math score', 'reading score', 'writing score']]
```

Out[19]:

	gender	math score	reading score	writing score
0	female	72	72	74
1	female	69	90	88
2	female	90	95	93
3	male	47	57	44
4	male	76	78	75
5	female	71	83	78
6	female	88	95	92
7	male	40	43	39
8	male	64	64	67

In [20]:
```
#15.    Drop rows with any column having NA/null data.
filedata.dropna()
```

Out[20]:

	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

In [21]:
```
#16.    Replace all NA/null data with value.
filedata.fillna('value')
```

Out[21]:

	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

In [22]:
```
filedata.describe()
```

Out[22]:

	math score	reading score	writing score
count	9.000000	9.000000	9.000000
mean	68.555556	75.222222	72.222222
std	16.629124	17.942810	19.504985
min	40.000000	43.000000	39.000000
25%	64.000000	64.000000	67.000000
50%	71.000000	78.000000	75.000000
75%	76.000000	90.000000	88.000000
max	90.000000	95.000000	93.000000

In [23]:
```
filedata.groupby(['gender']).mean()
```

Out[23]:

gender	math score	reading score	writing score
female	78.00	87.0	85.00
male	56.75	60.5	56.25

In [24]:
```
# Combine Data Sets
pd.merge(filedata, file2data, how='outer', on='gender')
```

Out[24]:

	gender	race/ethnicity_x	parental level of education_x	lunch_x	test preparation course_x	math score_x	reading score_x	writing score_x	race/ethnicity_y
0	female	group B	bachelor's degree	standard	none	72	72	74	group B
1	female	group B	bachelor's degree	standard	none	72	72	74	group C
2	female	group B	bachelor's degree	standard	none	72	72	74	group B
3	female	group B	bachelor's degree	standard	none	72	72	74	group B
4	female	group B	bachelor's degree	standard	none	72	72	74	group B
...	...	...	...	...	...	...	...	...	...
4513	male	group D	high school	free/reduced	completed	64	64	67	group A
4514	male	group D	high school	free/reduced	completed	64	64	67	group E
4515	male	group D	high school	free/reduced	completed	64	64	67	group E
4516	male	group D	high school	free/reduced	completed	64	64	67	group A
4517	male	group D	high school	free/reduced	completed	64	64	67	group C

4518 rows × 15 columns