NAME: NERELLA VENKATA RADHAKRISHNA

ID: 190031187

PRACTICAL-2

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

In [2]:
```
#1
df=pd.read_csv('21B15600.csv')
df
```

Out[2]:

	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 [3]:
```
#1.1
#Gather Columns into rows
pd.melt(df)
```

Out[3]:

	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 [4]:
```
#1.2
#Spread rows into columns
df.pivot(columns='gender',values='writing score')
```

Out[4]:

gender	female	male
0	74.0	NaN
1	88.0	NaN
2	93.0	NaN
3	NaN	44.0
4	NaN	75.0
5	78.0	NaN
6	92.0	NaN
7	NaN	39.0
8	NaN	67.0

In [5]:
```
#1.2
#Spread rows into columns
df.pivot(index='gender',columns='math score',values='writing score')
```

Out[5]:

math score	40	47	64	69	71	72	76	88	90
gender									
female	NaN	NaN	NaN	88.0	78.0	74.0	NaN	92.0	93.0
male	39.0	44.0	67.0	NaN	NaN	NaN	75.0	NaN	NaN

In [6]:
```
#1.3
#Append rows of Data Frames
df1=pd.read_csv('21B15600.csv')
df2=pd.read_csv('74B05600.csv')
pd.concat([df1,df2])
```

Out[6]:

	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 [7]:
```
#1.4
#Append Columns of Data Frames
df1=pd.read_csv('21B15600.csv')
df2=pd.read_csv('74B05600.csv')
pd.concat([df1,df2],axis=1)
```

Out[7]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	gender	race/ethnicity	ec
0	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0	female	group B	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	as
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 [8]:
```
#1.5
#Order rows by values of a column
df.sort_values('math score')
```

Out[8]:

	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 [9]:
```
#1.6
#Order rows by values of a column
df.sort_values('math score',ascending=False)
```

Out[9]:

	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 [10]:
```
#1.7
#Sort the index of a Data Frame
df.sort_index()
```

Out[10]:

	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 [11]:
```
#1.8
#Reset index of Data Frame to row numbers, moving index to columns.
df.reset_index()
```

Out[11]:

	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 [12]:
```
#1.9
#Drop columns from Data Frame
df.drop(columns=['gender'])
```

Out[12]:

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

In [13]:
```
#1.10
#Extract rows that meet logical criteria
df[df['writing score']>60]
```

Out[13]:

	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
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
8	male	group D	high school	free/reduced	completed	64	64	67

In [14]:
```
#1.11
#Remove duplicate rows (only considers columns)
df.drop_duplicates('lunch')
```

Out[14]:

	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
3	male	group A	associate's degree	free/reduced	none	47	57	44

In [15]:
```
#1.12
#Select first n rows,
df.head(5)
```

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

In [16]:
```
#1.12
#Select last n rows
df.tail(5)
```

Out[16]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
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]:
```
#1.13
#Select single column with specific name
df['parental level of education']
```

Out[17]:
```
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 [18]:
```
#1.14
#Select Multiple column with specific name.
df[['gender','writing score']]
```

Out[18]:

	gender	writing score
0	female	74
1	female	88
2	female	93
3	male	44
4	male	75
5	female	78
6	female	92
7	male	39
8	male	67

In [19]:
```
#2.15
#Drop rows with any column having NA/null data.
df.dropna()
```

Out[19]:

	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 [20]:
```
#2.16
#Replace all NA/null data with value.
df.fillna(5)
```

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]:
```
#3
#Summarize data
df.describe()
```

Out[21]:

	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 [22]:
```
#3
#Summarize data
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   gender                       9 non-null      object
 1   race/ethnicity               9 non-null      object
 2   parental level of education  9 non-null      object
 3   lunch                        9 non-null      object
 4   test preparation course      9 non-null      object
 5   math score                   9 non-null      int64 
 6   reading score                9 non-null      int64 
 7   writing score                9 non-null      int64 
dtypes: int64(3), object(5)
memory usage: 704.0+ bytes
```

In [24]:
```
#3
#Summarize data
df['lunch'].value_counts()
```

Out[24]:
```
standard        6
free/reduced    3
Name: lunch, dtype: int64
```

In [25]:
```
#3
#Summarize data
len(df)
```

Out[25]:
```
9
```

In [26]:
```
#4
#New Columns
#Make New Columns
df=df.assign(total_score=df['math score']+df['reading score']+df['writing score'])
df
```

Out[26]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total_score
0	female	group B	bachelor's degree	standard	none	72	72	74	218
1	female	group C	some college	standard	completed	69	90	88	247
2	female	group B	master's degree	standard	none	90	95	93	278
3	male	group A	associate's degree	free/reduced	none	47	57	44	148
4	male	group C	some college	standard	none	76	78	75	229
5	female	group B	associate's degree	standard	none	71	83	78	232
6	female	group B	some college	standard	completed	88	95	92	275
7	male	group B	some college	free/reduced	none	40	43	39	122
8	male	group D	high school	free/reduced	completed	64	64	67	195

In [27]:
```
#5
#Combine Data Sets
pd.merge(df1,df2,how='left',on='writing score')
```

Out[27]:

	gender_x	race/ethnicity_x	parental level of education_x	lunch_x	test preparation course_x	math score_x	reading score_x	writing score	gender_y	race/et
0	female	group B	bachelor's degree	standard	none	72	72	74	female	
1	female	group B	bachelor's degree	standard	none	72	72	74	female	
2	female	group B	bachelor's degree	standard	none	72	72	74	female	
3	female	group B	bachelor's degree	standard	none	72	72	74	female	
4	female	group B	bachelor's degree	standard	none	72	72	74	female	
...	...	...	...	...	...	...	...	...	...	...
140	male	group D	high school	free/reduced	completed	64	64	67	female	
141	male	group D	high school	free/reduced	completed	64	64	67	male	
142	male	group D	high school	free/reduced	completed	64	64	67	male	
143	male	group D	high school	free/reduced	completed	64	64	67	female	
144	male	group D	high school	free/reduced	completed	64	64	67	female	

145 rows × 15 columns