

C4_S4_Practice

Task 1

In [1]:

```
#Import the required library
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
DF = pd.read_csv('E:\Aishwarya official\Aishwarya Data Scince\Course 4\DS1_C4_S4_Test_Scores_Data_Practice.csv')
DF
```

Out[2]:

	school	school_setting	school_type	classroom	teaching_method	student_id	gender		lunch	pretest	posttest
0	ANKYI	Urban	Non-public	6OL	Standard	2FHT3	Female		Does not qualify	62.0	72.0
1	ANKYI	Urban	Non-public	6OL	Standard	3JIVH	Female		Does not qualify	66.0	79.0
2	ANKYI	Urban	Non-public	6OL	Standard	3XOWE	Male		Does not qualify	64.0	76.0
3	ANKYI	Urban	Non-public	6OL	Standard	556O0	Female		Does not qualify	61.0	77.0
4	ANKYI	Urban	Non-public	6OL	Standard	74LOE	Male		Does not qualify	64.0	76.0
...
2128	ZOWMK	Urban	Public	ZBH	Standard	T8LSK	Female		Does not qualify	39.0	55.0
2129	ZOWMK	Urban	Public	ZBH	Standard	VNP26	Female	Qualifies for reduced/free lunch		38.0	46.0
2130	ZOWMK	Urban	Public	ZBH	Standard	YDR1Z	Female	Qualifies for reduced/free lunch		45.0	51.0
2131	ZOWMK	Urban	Public	ZBH	Standard	YUEIH	Male	Qualifies for reduced/free lunch		46.0	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	ZVCQ8	Male	Qualifies for reduced/free lunch		41.0	48.0

2133 rows × 10 columns

In [4]:

```
df = DF.loc[:,DF.columns != 'student_id']
df
```

Out[4]:

	school	school_setting	school_type	classroom	teaching_method	gender		lunch	pretest	posttest
0	ANKYI	Urban	Non-public	6OL	Standard	Female		Does not qualify	62.0	72.0
1	ANKYI	Urban	Non-public	6OL	Standard	Female		Does not qualify	66.0	79.0
2	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	76.0
3	ANKYI	Urban	Non-public	6OL	Standard	Female		Does not qualify	61.0	77.0
4	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	76.0
...
2128	ZOWMK	Urban	Public	ZBH	Standard	Female		Does not qualify	39.0	55.0
2129	ZOWMK	Urban	Public	ZBH	Standard	Female	Qualifies for reduced/free lunch		38.0	46.0
2130	ZOWMK	Urban	Public	ZBH	Standard	Female	Qualifies for reduced/free lunch		45.0	51.0
2131	ZOWMK	Urban	Public	ZBH	Standard	Male	Qualifies for reduced/free lunch		46.0	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	Male	Qualifies for reduced/free lunch		41.0	48.0

2133 rows × 9 columns

In [5]:

```
dup = df[df.duplicated(keep = 'first')]
dup
```

Out[5]:

	school	school_setting	school_type	classroom	teaching_method	gender		lunch	pretest	posttest
4	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	76.0
19	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	73.0
38	ANKYI	Urban	Non-public	ZNS	Standard	Female		Does not qualify	59.0	69.0
50	CCAAW	Suburban	Non-public	2B1	Experimental	Male	Qualifies for reduced/free lunch		59.0	74.0
106	CCAAW	Suburban	Non-public	PGK	Standard	Female		Does not qualify	73.0	80.0
...
2058	ZOWMK	Urban	Public	Q0E	Experimental	Female	Qualifies for reduced/free lunch		38.0	53.0
2067	ZOWMK	Urban	Public	Q0E	Experimental	Male	Qualifies for reduced/free lunch		37.0	53.0
2070	ZOWMK	Urban	Public	Q0E	Experimental	Female	Qualifies for reduced/free lunch		38.0	53.0
2080	ZOWMK	Urban	Public	QA2	Standard	Female	Qualifies for reduced/free lunch		45.0	51.0
2128	ZOWMK	Urban	Public	ZBH	Standard	Female		Does not qualify	39.0	55.0

79 rows × 9 columns

Task 2

In [6]:

```
dup2 = df[df.duplicated(['school_type', 'teaching_method', 'pretest', 'posttest'], keep=False)]
dup2
```

Out[6]:

	school	school_setting	school_type	classroom	teaching_method	gender		lunch	pretest	posttest
0	ANKYI	Urban	Non-public	6OL	Standard	Female		Does not qualify	62.0	72.0
2	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	76.0
3	ANKYI	Urban	Non-public	6OL	Standard	Female		Does not qualify	61.0	77.0
4	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	64.0	76.0
6	ANKYI	Urban	Non-public	6OL	Standard	Male		Does not qualify	63.0	75.0
...
2128	ZOWMK	Urban	Public	ZBH	Standard	Female		Does not qualify	39.0	55.0
2129	ZOWMK	Urban	Public	ZBH	Standard	Female	Qualifies for reduced/free lunch		38.0	46.0
2130	ZOWMK	Urban	Public	ZBH	Standard	Female	Qualifies for reduced/free lunch		45.0	51.0
2131	ZOWMK	Urban	Public	ZBH	Standard	Male	Qualifies for reduced/free lunch		46.0	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	Male	Qualifies for reduced/free lunch		41.0	48.0

1405 rows × 9 columns

Task 3

In [9]:

```
group = df.groupby(['school'])['pretest'].mean()
group
```

Out[9]:

```
school
ANKYI    61.341463
CCAAW    64.623853
CIMBB    65.067568
CUQAM    53.925234
DNQDD    54.327869
FBUMG    62.891304
GJJHK    53.194915
GOKXL    50.796875
GOOBU    38.248408
IDGFP    75.202128
KFZMY    41.865385
KZKKE    37.261261
LAYPA    62.035088
OJOBUS   56.197531
QQQTS    52.527027
UAGPU    62.390805
UKPGS    78.453125
UUUQX    67.253012
VHDHF    52.666667
VKWQH    52.060000
VVTVA    35.955752
ZMNVA    68.130435
ZOWMK    41.572650
Name: pretest, dtype: float64
```

Task 4

In [12]:

```
group1 = df.groupby(['school_setting', 'gender'])['gender'].count()
group1
```

Out[12]:

```
school_setting  gender
Rural           Female    228
                Male      279
Suburban        Female    368
                Male      348
Urban           Female    456
                Male      444
Name: gender, dtype: int64
```

Task 5

In [14]:

```
pvt = pd.pivot_table(df,index=['school'], values = ['posttest'], aggfunc = ['min'])
pvt
```

Out[14]:

min	
posttest	
school	
ANKYI	63.0
CCAAW	67.0
CIMBB	64.0
CUQAM	56.0
DNQDD	49.0
FBUMG	68.0
GJJHK	49.0
GOKXL	48.0
GOOBU	32.0
IDGFP	74.0
KFZMY	44.0
KZKKE	36.0
LAYPA	63.0
OJOBV	50.0
QQQTS	51.0
UAGPU	62.0
UKPGS	82.0
UUUQX	62.0
VHDHF	52.0
VKWQH	48.0
VVTVA	39.0
ZMNYA	66.0
ZOWMK	43.0

In [87]:

```
pvt1 = pd.pivot_table(df,index=['school'], values = ['posttest'], aggfunc = ['max'])
pvt1
```

Out[87]:

max	
posttest	
school	
ANKYI	79.00000
CCAAW	91.00000
CIMBB	88.00000
CUQAM	76.00000
DNQDD	79.00000
FBUMG	88.00000
GJJHK	85.00000
GOKXL	77.00000
GOOBU	67.11132
IDGFP	100.00000
KFZMY	67.00000
KZKKE	62.00000
LAYPA	84.00000
OJOBU	84.00000
QQQTS	85.00000
UAGPU	83.00000
UKPGS	99.00000
UUUQX	91.00000
VHDHF	77.00000
VKWQH	82.00000
VVTVA	67.11132
ZMNYA	95.00000
ZOWMK	63.00000

In []:

```
###sort = df.sort_values('Sales',ascending = False)
```

Task 6

In [16]:

```
pvt3 = pd.pivot_table(df,index=['school'], values = ['pretest'], aggfunc = ['sum'])
pvt3
```

Out[16]:

sum	
pretest	
school	
ANKYI	2515.0
CCAAW	7044.0
CIMBB	4815.0
CUQAM	5770.0
DNQDD	6628.0
FBUMG	2893.0
GJJHK	6277.0
GOKXL	3251.0
GOOBU	6005.0
IDGFP	7069.0
KFZMY	2177.0
KZKKE	4136.0
LAYPA	3536.0
OJOBU	4552.0
QQQTS	7774.0
UAGPU	5428.0
UKPGS	10042.0
UUUQX	5582.0
VHDHF	2686.0
VKWQH	5206.0
VVTVA	4063.0
ZMNYA	4701.0
ZOWMK	4864.0

In [17]:

```
pvt4 = pd.pivot_table(df,index=['school'], values = ['pretest'], aggfunc = ['mean'])
pvt4
```

Out[17]:

mean	
pretest	
school	
ANKYI	61.341463
CCAAW	64.623853
CIMBB	65.067568
CUQAM	53.925234
DNQDD	54.327869
FBUMG	62.891304
GJJHK	53.194915
GOKXL	50.796875
GOOBU	38.248408
IDGFP	75.202128
KFZMY	41.865385
KZKKE	37.261261
LAYPA	62.035088
OJOBU	56.197531
QQQTS	52.527027
UAGPU	62.390805
UKPGS	78.453125
UUUQX	67.253012
VHDHF	52.666667
VKWQH	52.060000
VVTVA	35.955752
ZMNYA	68.130435
ZOWMK	41.572650

Task 7

In [38]:

```
A = DF[(DF.classroom=='5LQ') & (DF.school=='FBUMG')]
A
```

472	FBUMG	Rural	Non-public	5LQ	Experimental	ATQQJ	Female	Does not qualify	68.0	82.0
473	FBUMG	Rural	Non-public	5LQ	Experimental	AYEU1	Male	Qualifies for reduced/free lunch	68.0	79.0
474	FBUMG	Rural	Non-public	5LQ	Experimental	B9FSU	Male	Does not qualify	69.0	83.0
475	FBUMG	Rural	Non-public	5LQ	Experimental	EV13K	Female	Does not qualify	66.0	83.0
476	FBUMG	Rural	Non-public	5LQ	Experimental	I5H37	Male	Qualifies for reduced/free lunch	65.0	83.0
477	FBUMG	Rural	Non-public	5LQ	Experimental	JC5I9	NaN	Does not qualify	58.0	80.0
478	FBUMG	Rural	Non-public	5LQ	Experimental	JPE2J	Male	Does not qualify	59.0	81.0
479	FBUMG	Rural	Non-public	5LQ	Experimental	MFBYU	Female	Does not qualify	73.0	82.0
480	FBUMG	Rural	Non-public	5LQ	Experimental	O144X	Male	Does not qualify	67.0	86.0
481	FBUMG	Rural	Non-public	5LQ	Experimental	OGKP3	Male	Does not qualify	72.0	83.0
482	FBUMG	Rural	Non-public	5LQ	Experimental	UMFI7	Female	Qualifies for reduced/free lunch	56.0	73.0
483	FBUMG	Rural	Non-public	5LQ	Experimental	V1DNJ	Male	Does not qualify	61.0	76.0
484	FBUMG	Rural	Non-public	5LQ	Experimental	YRN9S	Male	Does not qualify	68.0	87.0

In [39]:

```
st_id = A.loc[:,DF.columns == 'student_id']  
st_id
```

Out[39]:

	student_id
467	04DG5
468	20M2D
469	39KCW
470	5Z1B6
471	6TLU8
472	ATQQJ
473	AYEU1
474	B9FSU
475	EV13K
476	I5H37
477	JC5I9
478	JPE2J
479	MFBYU
480	O144X
481	OGKP3
482	UMFI7
483	V1DNJ
484	YRN9S

In [36]:

```
Scholarship = (10000*A['posttest'])/100  
Scholarship
```

Out[36]:

467	8400.0
468	8500.0
469	8100.0
470	8800.0
471	8100.0
472	8200.0
473	7900.0
474	8300.0
475	8300.0
476	8300.0
477	8000.0
478	8100.0
479	8200.0
480	8600.0
481	8300.0
482	7300.0
483	7600.0
484	8700.0

Name: posttest, dtype: float64

In [37]:

```
Dic = {'Scholarship': Scholarship}
df1 = pd.DataFrame(Dic)
df1
```

Out[37]:

	Scholarship
467	8400.0
468	8500.0
469	8100.0
470	8800.0
471	8100.0
472	8200.0
473	7900.0
474	8300.0
475	8300.0
476	8300.0
477	8000.0
478	8100.0
479	8200.0
480	8600.0
481	8300.0
482	7300.0
483	7600.0
484	8700.0

In [90]:

```
D = pd.concat([st_id,df1],axis=1)
D
```

Out[90]:

	student_id	Scholarship
467	04DG5	8400.0
468	20M2D	8500.0
469	39KCW	8100.0
470	5Z1B6	8800.0
471	6TLU8	8100.0
472	ATQQJ	8200.0
473	AYEU1	7900.0
474	B9FSU	8300.0
475	EV13K	8300.0
476	I5H37	8300.0
477	JC5J9	8000.0
478	JPE2J	8100.0
479	MFBYU	8200.0
480	O144X	8600.0
481	OGKP3	8300.0
482	UMFI7	7300.0
483	V1DNJ	7600.0
484	YRN9S	8700.0

In [92]:

```
##D.set_index('student_id').join(A.set_index('student_id')), on = 'student_id', how = 'outer')
```

Task 8 : ¶

In [58]:

```
Rp= df.replace(to_replace=['Does not qualify','Qualifies for reduced/free lunch'], value=['No','Yes'])
Rp
```

Out[58]:

	school	school_setting	school_type	classroom	teaching_method	gender	lunch	pretest	posttest
0	ANKYI	Urban	Non-public	6OL	Standard	Female	No	62.0	72.0
1	ANKYI	Urban	Non-public	6OL	Standard	Female	No	66.0	79.0
2	ANKYI	Urban	Non-public	6OL	Standard	Male	No	64.0	76.0
3	ANKYI	Urban	Non-public	6OL	Standard	Female	No	61.0	77.0
4	ANKYI	Urban	Non-public	6OL	Standard	Male	No	64.0	76.0
...
2128	ZOWMK	Urban	Public	ZBH	Standard	Female	No	39.0	55.0
2129	ZOWMK	Urban	Public	ZBH	Standard	Female	Yes	38.0	46.0
2130	ZOWMK	Urban	Public	ZBH	Standard	Female	Yes	45.0	51.0
2131	ZOWMK	Urban	Public	ZBH	Standard	Male	Yes	46.0	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	Male	Yes	41.0	48.0

2133 rows × 9 columns

Task 9

In [50]:

```
pd.crosstab(df.school_setting, df.teaching_method, rownames = ['school_setting'] , colnames = ['teaching_method'])
```

Out[50]:

teaching_method	Experimental	Standard
school_setting		
Rural	201	309
Suburban	284	433
Urban	275	631

Task 10

In [52]:

```
B = DF[(DF.gender == 'Male') & (DF.posttest>39)]
B
```

Out[52]:

	school	school_setting	school_type	classroom	teaching_method	student_id	gender	lunch	pretest	posttest
2	ANKYI	Urban	Non-public	6OL	Standard	3XOWE	Male	Does not qualify	64.0	76.0
4	ANKYI	Urban	Non-public	6OL	Standard	74LOE	Male	Does not qualify	64.0	76.0
6	ANKYI	Urban	Non-public	6OL	Standard	9KMZD	Male	Does not qualify	63.0	75.0
8	ANKYI	Urban	Non-public	6OL	Standard	CS5QP	Male	Does not qualify	64.0	77.0
10	ANKYI	Urban	Non-public	6OL	Standard	DZMKU	Male	Does not qualify	61.0	73.0
...
2124	ZOWMK	Urban	Public	ZBH	Standard	Q6QZP	Male	Qualifies for reduced/free lunch	41.0	50.0
2126	ZOWMK	Urban	Public	ZBH	Standard	R805N	Male	Qualifies for reduced/free lunch	38.0	51.0
2127	ZOWMK	Urban	Public	ZBH	Standard	S4I5S	Male	Qualifies for reduced/free lunch	39.0	50.0
2131	ZOWMK	Urban	Public	ZBH	Standard	YUEIH	Male	Qualifies for reduced/free lunch	46.0	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	ZVCQ8	Male	Qualifies for reduced/free lunch	41.0	48.0

1061 rows × 10 columns

In [60]:

```
#DF['Sales'] = DF['Sales'].replace([32961],33200)
B['pretest'] = B['pretest'].replace([135], ['pretest'])
B ['pretest']=135
B
```

Out[60]:

	school	school_setting	school_type	classroom	teaching_method	student_id	gender		lunch	pretest	posttest
2	ANKYI	Urban	Non-public	6OL	Standard	3XOWE	Male		Does not qualify	135	76.0
4	ANKYI	Urban	Non-public	6OL	Standard	74LOE	Male		Does not qualify	135	76.0
6	ANKYI	Urban	Non-public	6OL	Standard	9KMZD	Male		Does not qualify	135	75.0
8	ANKYI	Urban	Non-public	6OL	Standard	CS5QP	Male		Does not qualify	135	77.0
10	ANKYI	Urban	Non-public	6OL	Standard	DZMKU	Male		Does not qualify	135	73.0
...
2124	ZOWMK	Urban	Public	ZBH	Standard	Q6QZP	Male	Qualifies for reduced/free lunch		135	50.0
2126	ZOWMK	Urban	Public	ZBH	Standard	R805N	Male	Qualifies for reduced/free lunch		135	51.0
2127	ZOWMK	Urban	Public	ZBH	Standard	S4I5S	Male	Qualifies for reduced/free lunch		135	50.0
2131	ZOWMK	Urban	Public	ZBH	Standard	YUEIH	Male	Qualifies for reduced/free lunch		135	53.0
2132	ZOWMK	Urban	Public	ZBH	Standard	ZVCQ8	Male	Qualifies for reduced/free lunch		135	48.0

1061 rows × 10 columns

Task 11

In [56]:

```
avg1 = pd.pivot_table(df,index=['school_setting'], values = ['posttest'], aggfunc = ['mean'])
avg1
```

Out[56]:

mean	
posttest	
school_setting	
Rural	64.039293
Suburban	76.020950
Urban	61.784292

Task 12

In [61]:

```
df.isnull().sum()
```

Out[61]:

school	0
school_setting	0
school_type	0
classroom	0
teaching_method	0
gender	10
lunch	0
pretest	4
posttest	4
dtype: int64	

In [62]:

```
df['gender'].value_counts()
```

Out[62]:

Male	1071
Female	1052
Name: gender, dtype: int64	

In [64]:

```
df['gender'] = df['gender'].fillna('Male')  
df['gender'].value_counts()
```

Out[64]:

```
Male      1081  
Female    1052  
Name: gender, dtype: int64
```

In [65]:

```
df['pretest'] = df['pretest'].fillna(df['pretest'].mean())
```

In [66]:

```
df['posttest'] = df['posttest'].fillna(df['posttest'].mean())
```

In [67]:

```
df.isnull().sum()
```

Out[67]:

```
school          0  
school_setting  0  
school_type     0  
classroom       0  
teaching_method 0  
gender          0  
lunch           0  
pretest         0  
posttest        0  
dtype: int64
```