

DSc

February 24, 2024

1 Initialisation

```
[1]: import pandas as pd
from IPython.display import HTML, Markdown, Latex

def display_df(tp_df=None, index=False):
    tp_df = tp_df if isinstance(tp_df, pd.DataFrame) else df
    display(Markdown(tp_df.to_markdown(index=index)))

# def display_df(tp_df=None, index=False):
#     tp_df = tp_df if isinstance(tp_df, pd.DataFrame) else df
#     display(HTML(tp_df.to_html(index=index)))
```

2 15/02/24

Given a dataset, print the following:

- 1) Records of index 1 & 3
- 2) Records where age ≥ 15
- 3) Records where age ≥ 12 and gender = Male
- 4) City and gender of people with age ≥ 12

```
[2]: data = {
    'age': [10, 22, 13, 21, 12, 11, 17],
    'section': ['A', 'B', 'C', 'B', 'B', 'A', 'A'],
    'city': ['Gurgaon', 'Delhi', 'Mumbai', 'Delhi', 'Mumbai', 'Delhi', 'Mumbai'],
    'gender': ['M', 'F', 'F', 'M', 'M', 'M', 'F'],
    'favorite_color': ['red', 'black', 'yellow', 'pink', 'black', 'green', 'red']
}
df = pd.DataFrame(data)

print('\nOriginal data:')
display_df()

print('\n1) Records of index 1 & 3')
display_df(df.iloc[[1, 3], :])
```

```

print('\n2) Records where age >= 15:')
display_df( df.query('age >= 15') )

print('\n3) Records where age >= 12 and gender = Male:')
display_df( df.query('age >= 12 and gender == "M"') )

print('\n4) City and gender of people with age >= 12:')
display_df( df.query('age >= 12')[['city','gender']] )

```

Original data:

age	section	city	gender	favorite_color
10	A	Gurgaon	M	red
22	B	Delhi	F	black
13	C	Mumbai	F	yellow
21	B	Delhi	M	pink
12	B	Mumbai	M	black
11	A	Delhi	M	green
17	A	Mumbai	F	red

1) Records of index 1 & 3

age	section	city	gender	favorite_color
22	B	Delhi	F	black
21	B	Delhi	M	pink

2) Records where age >= 15:

age	section	city	gender	favorite_color
22	B	Delhi	F	black
21	B	Delhi	M	pink
17	A	Mumbai	F	red

3) Records where age >= 12 and gender = Male:

age	section	city	gender	favorite_color
21	B	Delhi	M	pink
12	B	Mumbai	M	black

4) City and gender of people with age ≥ 12 :

city	gender
Delhi	F
Mumbai	F
Delhi	M
Mumbai	M
Mumbai	F

3 22/02/24

Create a dataframe to store data of 10 students, with the columns being “Name”, “Age”, “Semester I marks out of 600”, “Semester II marks out of 500”, and “Attendance”

- 1) Display details of students who scored more than 560 marks in sem I
- 2) Display details of students who scored less than 250 marks in sem II
- 3) Display details of student who scored minimum marks in sem II
- 4) Display details of student who scored maximum marks in sem II
- 5) Display details of students whose attendance is more than 75
- 6) Display details of students whose attendance is less than 50
- 7) Insert 2 new records
- 8) Add a column corresponding to percentage of marks of both semesters
- 9) Add a new column corresponding to grades:

Percentage	Grade
≥ 90	O
≥ 75 and < 90	A+
≥ 60 and < 75	A
≥ 50 and < 60	B+
≥ 40 and < 50	B
> 40	F

```
[3]: data = {
    'Name': ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J'],
    'Age': [20, 21, 20, 22, 23, 20, 21, 22, 20, 21],
    'Semester I marks out of 600': [213, 31, 57, 406, 417, 45, 217, 200, 588, 319],
    'Semester II marks out of 500': [198, 378, 133, 450, 283, 485, 193, 283, 236, 191],
    'Attendance': [76, 26, 53, 32, 50, 67, 92, 62, 44, 85]
}
df = pd.DataFrame(data)
```

```

print('\nOriginal data:')
display_df()

print('\n1) Students who scored more than 560 marks in sem I:')
ans = df.query('`Semester I marks out of 600` > 560')
display_df(ans, index=True)

print('\n2) Students who scored less than 250 marks in sem II:')
ans = df.query('`Semester II marks out of 500` < 250')
display_df(ans, index=True)

print('\n3) Student who scored minimum marks in sem II:')
min_marks = min(df['Semester II marks out of 500'])
ans = df.query('`Semester II marks out of 500` == @min_marks')
display_df( ans , index=True )

print('\n4) Student who scored maximum marks in sem II:')
ans = df.sort_values(by='Semester II marks out of 500',ascending=False).head(1)
display_df(ans, index=True)

print('\n5) Students whose attendance is more than 75:')
ans = df.query('Attendance > 75')
display_df(ans, index=True)

print('\n6) Students whose attendance is less than 50:')
ans = df.query('Attendance < 50')
display_df(ans, index=True)

print('\n7) Inserted two new records:')
new_data = {
    'Name':          ['K', 'L'],
    'Age':           [22, 23],
    'Semester I marks out of 600': [300, 400],
    'Semester II marks out of 500': [400, 300],
    'Attendance':    [80, 40]
}
new_df = pd.DataFrame(new_data)
df = pd.concat([df,new_df], ignore_index=True)
display_df(index=True)

print('\n8) Added the percentage column:')
df['Percentage'] = (df['Semester I marks out of 600'] + df['Semester II marks_
↳out of 500']) / 11
df['Percentage'] = df['Percentage'].apply(lambda x: round(x,2))
display_df()

print('\n9) Added the grade column:')

```

```
def get_grade(x: float):
    if x >= 90: return 'O'
    elif x >= 75: return 'A+'
    elif x >= 60: return 'A'
    elif x >= 50: return 'B+'
    elif x >= 40: return 'B'
    else: return 'F'
df['Grade'] = df['Percentage'].apply(get_grade)
display_df()
```

Original data:

Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
A	20	213	198	76
B	21	31	378	26
C	20	57	133	53
D	22	406	450	32
E	23	417	283	50
F	20	45	485	67
G	21	217	193	92
H	22	200	283	62
I	20	588	236	44
J	21	319	191	85

1) Students who scored more than 560 marks in sem I:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
8	I	20	588	236	44

2) Students who scored less than 250 marks in sem II:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
0	A	20	213	198	76
2	C	20	57	133	53
6	G	21	217	193	92
8	I	20	588	236	44
9	J	21	319	191	85

3) Student who scored minimum marks in sem II:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
2	C	20	57	133	53

4) Student who scored maximum marks in sem II:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
5	F	20	45	485	67

5) Students whose attendance is more than 75:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
0	A	20	213	198	76
6	G	21	217	193	92
9	J	21	319	191	85

6) Students whose attendance is less than 50:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
1	B	21	31	378	26
3	D	22	406	450	32
8	I	20	588	236	44

7) Inserted two new records:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
0	A	20	213	198	76
1	B	21	31	378	26
2	C	20	57	133	53
3	D	22	406	450	32
4	E	23	417	283	50
5	F	20	45	485	67
6	G	21	217	193	92
7	H	22	200	283	62
8	I	20	588	236	44
9	J	21	319	191	85

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance
10	K	22	300	400	80
11	L	23	400	300	40

8) Added the percentage column:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance	Percentage
	A	20	213	198	76	37.36
	B	21	31	378	26	37.18
	C	20	57	133	53	17.27
	D	22	406	450	32	77.82
	E	23	417	283	50	63.64
	F	20	45	485	67	48.18
	G	21	217	193	92	37.27
	H	22	200	283	62	43.91
	I	20	588	236	44	74.91
	J	21	319	191	85	46.36
	K	22	300	400	80	63.64
	L	23	400	300	40	63.64

9) Added the grade column:

	Name	Age	Semester I marks out of 600	Semester II marks out of 500	Attendance	Percentage	Grade
	A	20	213	198	76	37.36	F
	B	21	31	378	26	37.18	F
	C	20	57	133	53	17.27	F
	D	22	406	450	32	77.82	A+
	E	23	417	283	50	63.64	A
	F	20	45	485	67	48.18	B
	G	21	217	193	92	37.27	F
	H	22	200	283	62	43.91	B
	I	20	588	236	44	74.91	A
	J	21	319	191	85	46.36	B
	K	22	300	400	80	63.64	A
	L	23	400	300	40	63.64	A