

```
import pandas as pd
import numpy as np

courses = pd.read_csv('/content/courses.csv')
students = pd.read_csv('/content/students.csv')
nov = pd.read_csv('/content/reg-month1.csv')
dec = pd.read_csv('/content/reg-month2.csv')

matches = pd.read_csv('/content/matches.csv')
delivery = pd.read_csv('/content/deliveries.csv')
```

dec

	student_id	course_id
0	3	5
1	16	7
2	12	10
3	12	1
4	14	9
5	7	7
6	7	2
7	16	3
8	17	10
9	11	8
10	14	6
11	12	5
12	12	7
13	18	8
14	1	10
15	1	9
16	2	5
17	7	6
18	22	5
19	22	6
20	23	9
21	23	5
22	14	4
23	14	1
24	11	10
25	42	9
26	50	8
27	38	1

```
# pd.concat
# df.concat
# ignore_index
# df.append
# mullitindex -> fetch using iloc
# concat dataframes horizontally

regs = pd.concat([nov,dec],ignore_index=True)
regs
```

```

    student_id  course_id
0           23         1
1           15         5
2           18         6
3           23         4
4           16         9
5           18         1
6            1         1
7            7         8
8           22         3
9           15         1
10          19         4
11            1         6
12            7        10
13           11         7
14           13         3
15           24         4
16           21         1
17           16         5
18           23         3
19           17         7
20           23         6
21           25         1
22           19         2
23           25        10
24            3         3
25            3         5
26           16         7
27           12        10
28           12         1
29           14         9
30            7         7

nov.append(dec, ignore_index=True)
```

	student_id	course_id
0	23	1
1	15	5
2	18	6
3	23	4
4	16	9
5	18	1
6	1	1
7	7	8
8	22	3
9	15	1
10	19	4
11	1	6
12	7	10
13	11	7
14	13	3
15	24	4
16	21	1
17	16	5
18	23	3
19	17	7
20	23	6
21	25	1
22	19	2
23	25	10
24	3	3
25	3	5
26	16	7
27	12	10
28	12	1
29	14	9
30	7	7

```
multi = pd.concat([nov,dec],keys=['Nov','Dec'])
# Multiindex DataFrame
multi.loc[('Dec',4)]

student_id    14
course_id      9
Name: (Dec, 4), dtype: int64

pd.concat([nov,dec],axis=1)
```

	student_id	course_id	student_id	course_id
0	23.0	1.0	3	5
1	15.0	5.0	16	7
2	18.0	6.0	12	10
3	23.0	4.0	12	1
4	16.0	9.0	14	9
5	18.0	1.0	7	7
6	1.0	1.0	7	2
7	7.0	8.0	16	3
8	22.0	3.0	17	10
9	15.0	1.0	11	8
10	19.0	4.0	14	6
11	1.0	6.0	12	5
12	7.0	10.0	12	7
13	11.0	7.0	18	8
14	13.0	3.0	1	10
15	24.0	4.0	1	9
16	21.0	1.0	2	5
17	16.0	5.0	7	6
18	23.0	3.0	22	5
19	17.0	7.0	22	6
20	23.0	6.0	23	9
21	25.0	1.0	23	5
22	19.0	2.0	14	4
23	25.0	10.0	14	1
24	3.0	3.0	11	10
25	NaN	NaN	42	9
26	NaN	NaN	50	8
27	NaN	NaN	38	1

```
# inner join
students.merge(regs,how='inner',on='student_id')
```

	student_id	name	partner	course_id
0	1	Kailash Harjo	23	1
1	1	Kailash Harjo	23	6
2	1	Kailash Harjo	23	10
3	1	Kailash Harjo	23	9
4	2	Esha Butala	1	5
5	3	Parveen Bhalla	3	3
6	3	Parveen Bhalla	3	5
7	7	Tarun Thaker	9	8
8	7	Tarun Thaker	9	10
9	7	Tarun Thaker	9	7
10	7	Tarun Thaker	9	2
11	7	Tarun Thaker	9	6
12	11	David Mukhopadhyay	20	7
13	11	David Mukhopadhyay	20	8
14	11	David Mukhopadhyay	20	10
15	12	Radha Dutt	19	10
16	12	Radha Dutt	19	1
17	12	Radha Dutt	19	5
18	12	Radha Dutt	19	7
19	13	Munni Varghese	24	3
20	14	Pranab Natarajan	22	9
21	14	Pranab Natarajan	22	6
22	14	Pranab Natarajan	22	4
23	14	Pranab Natarajan	22	1
24	15	Preet Sha	16	5
25	15	Preet Sha	16	1
26	16	Elias Dodiya	25	9
27	16	Elias Dodiya	25	5
28	16	Elias Dodiya	25	7
29	16	Elias Dodiya	25	3
30	17	Yasmin Palan	7	7

```
# left join
courses.merge(regs,how='left',on='course_id')
```

	course_id	course_name	price	student_id
0	1	python	2499	23.0
1	1	python	2499	18.0
2	1	python	2499	1.0
3	1	python	2499	15.0
4	1	python	2499	21.0
5	1	python	2499	25.0
6	1	python	2499	12.0
7	1	python	2499	14.0
8	1	python	2499	38.0
9	2	sql	3499	19.0
10	2	sql	3499	7.0
11	3	data analysis	4999	22.0
12	3	data analysis	4999	13.0
13	3	data analysis	4999	23.0
14	3	data analysis	4999	3.0
15	3	data analysis	4999	16.0
16	4	machine learning	9999	23.0
17	4	machine learning	9999	19.0
18	4	machine learning	9999	24.0
19	4	machine learning	9999	14.0
20	5	tableau	2499	15.0
21	5	tableau	2499	16.0
22	5	tableau	2499	3.0
23	5	tableau	2499	12.0
24	5	tableau	2499	2.0
25	5	tableau	2499	22.0
26	5	tableau	2499	23.0
27	6	power bi	1899	18.0
28	6	power bi	1899	1.0
29	6	power bi	1899	23.0
30	6	power bi	1899	14.0

```
# right join
temp_df = pd.DataFrame({
    'student_id': [26,27,28],
    'name': ['Nitish', 'Ankit', 'Rahul'],
    'partner': [28,26,17]
})

students = pd.concat([students,temp_df],ignore_index=True)

students.tail()
```

	student_id	name	partner
23	24	Radhika Suri	17
24	25	Shashank D'Alia	2
25	26	Nitish	28
26	27	Ankit	26
27	28	Rahul	17

```
students.merge(regs,how='right',on='student_id')
```

	student_id	name	partner	course_id
0	23	Chhavi Lachman	18.0	1
1	15	Preet Sha	16.0	5
2	18	Fardeen Mahabir	13.0	6
3	23	Chhavi Lachman	18.0	4
4	16	Elias Dodiya	25.0	9
5	18	Fardeen Mahabir	13.0	1
6	1	Kailash Harjo	23.0	1
7	7	Tarun Thaker	9.0	8
8	22	Yash Sethi	21.0	3
9	15	Preet Sha	16.0	1
10	19	Qabeel Raman	12.0	4
11	1	Kailash Harjo	23.0	6
12	7	Tarun Thaker	9.0	10
13	11	David Mukhopadhyay	20.0	7
14	13	Munni Varghese	24.0	3
15	24	Radhika Suri	17.0	4
16	21	Seema Kota	15.0	1
17	16	Elias Dodiya	25.0	5
18	23	Chhavi Lachman	18.0	3
19	17	Yasmin Palan	7.0	7
20	23	Chhavi Lachman	18.0	6
21	25	Shashank D'Alia	2.0	1
22	19	Qabeel Raman	12.0	2
23	25	Shashank D'Alia	2.0	10
24	3	Parveen Bhalla	3.0	3
25	3	Parveen Bhalla	3.0	5
26	16	Elias Dodiya	25.0	7
27	12	Radha Dutt	19.0	10
28	12	Radha Dutt	19.0	1
29	14	Pranab Natarajan	22.0	9
30	7	Tarun Thaker	9.0	7

```
regs.merge(students,how='left',on='student_id')
```

	student_id	course_id	name	partner
0	23	1	Chhavi Lachman	18.0
1	15	5	Preet Sha	16.0
2	18	6	Fardeen Mahabir	13.0
3	23	4	Chhavi Lachman	18.0
4	16	9	Elias Dodiya	25.0
5	18	1	Fardeen Mahabir	13.0
6	1	1	Kailash Harjo	23.0
7	7	8	Tarun Thaker	9.0
8	22	3	Yash Sethi	21.0
9	15	1	Preet Sha	16.0
10	19	4	Qabeel Raman	12.0
11	1	6	Kailash Harjo	23.0
12	7	10	Tarun Thaker	9.0
13	11	7	David Mukhopadhyay	20.0
14	13	3	Munni Varghese	24.0
15	24	4	Radhika Suri	17.0
16	21	1	Seema Kota	15.0
17	16	5	Elias Dodiya	25.0
18	23	3	Chhavi Lachman	18.0
19	17	7	Yasmin Palan	7.0
20	23	6	Chhavi Lachman	18.0
21	25	1	Shashank D'Alia	2.0
22	19	2	Qabeel Raman	12.0
23	25	10	Shashank D'Alia	2.0
24	3	3	Parveen Bhalla	3.0
25	3	5	Parveen Bhalla	3.0
26	16	7	Elias Dodiya	25.0
27	12	10	Radha Dutt	19.0
28	12	1	Radha Dutt	19.0
29	14	9	Pranab Natarajan	22.0
30	7	7	Tarun Thaker	9.0

```
# outer join
students.merge(regs,how='outer',on='student_id').tail(10)
```

	student_id	name	partner	course_id
53	23	Chhavi Lachman	18.0	5.0
54	24	Radhika Suri	17.0	4.0
55	25	Shashank D'Alia	2.0	1.0
56	25	Shashank D'Alia	2.0	10.0
57	26	Nitish	28.0	NaN
58	27	Ankit	26.0	NaN
59	28	Rahul	17.0	NaN
60	42	NaN	NaN	9.0
61	50	NaN	NaN	8.0
62	38	NaN	NaN	1.0
63	22	Yash Sethi	21.0	

```
# 1. find total revenue generated
total = regs.merge(courses,how='inner',on='course_id')['price'].sum()
total
```

154247
48
14
1
Pranab Natarajan
22.0


```
# 2. find month by month revenue
temp_df = pd.concat([nov,dec],keys=['Nov','Dec']).reset_index()
temp_df.merge(courses,on='course_id').groupby('level_0')['price'].sum()
```

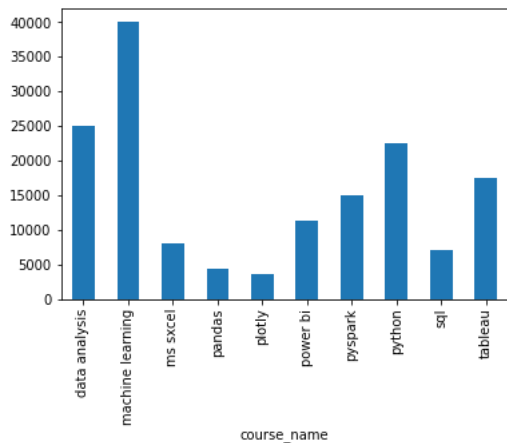
```
level_0
Dec      65072
Nov      89175
Name: price, dtype: int64
```

```
# 3. Print the registration table
# cols -> name -> course -> price
regs.merge(students,on='student_id').merge(courses,on='course_id')[['name','course_name','price']]
```

	name	course_name	price
0	Chhavi Lachman	python	2499
1	Preet Sha	python	2499
2	Fardeen Mahabir	python	2499
3	Kailash Harjo	python	2499
4	Seema Kota	python	2499
5	Shashank D'Alia	python	2499
6	Radha Dutt	python	2499
7	Pranab Natarajan	python	2499
8	Chhavi Lachman	machine learning	9999
9	Qabeel Raman	machine learning	9999
10	Radhika Suri	machine learning	9999
11	Pranab Natarajan	machine learning	9999
12	Chhavi Lachman	data analysis	4999
13	Elias Dodiya	data analysis	4999
14	Yash Sethi	data analysis	4999
15	Munni Varghese	data analysis	4999
16	Parveen Bhalla	data analysis	4999
17	Chhavi Lachman	power bi	1899
18	Fardeen Mahabir	power bi	1899
19	Kailash Harjo	power bi	1899
20	Tarun Thaker	power bi	1899
21	Yash Sethi	power bi	1899
22	Pranab Natarajan	power bi	1899
23	Chhavi Lachman	plotly	699
24	Elias Dodiya	plotly	699
25	Kailash Harjo	plotly	699
26	Pranab Natarajan	plotly	699
27	Chhavi Lachman	tableau	2499
28	Preet Sha	tableau	2499
29	Elias Dodiya	tableau	2499
30	Yash Sethi	tableau	2499

```
# 4. Plot bar chart for revenue/course
regs.merge(courses,on='course_id').groupby('course_name')['price'].sum().plot(kind='bar')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f56b73cb2e0>



```
# 5. find students who enrolled in both the months
common_student_id = np.intersect1d(nov['student_id'],dec['student_id'])
common_student_id

array([ 1,  3,  7, 11, 16, 17, 18, 22, 23])
```

```
students[students['student_id'].isin(common_student_id)]
```

	student_id	name	partner
0	1	Kailash Harjo	23
2	3	Parveen Bhalla	3
6	7	Tarun Thaker	9
10	11	David Mukhopadhyay	20
15	16	Elias Dodiya	25
16	17	Yasmin Palan	7
17	18	Fardeen Mahabir	13
21	22	Yash Sethi	21
22	23	Chhavi Lachman	18

```
# 6. find course that got no enrollment
# courses['course_id']
# regs['course_id']

course_id_list = np.setdiff1d(courses['course_id'],regs['course_id'])
courses[courses['course_id'].isin(course_id_list)]
```

	course_id	course_name	price
10	11	Numpy	699
11	12	C++	1299

```
# 7. find students who did not enroll into any courses
student_id_list = np.setdiff1d(students['student_id'],regs['student_id'])
students[students['student_id'].isin(student_id_list)].shape[0]
```

```
(10/28)*100
```

```
35.714285714285715
```

```
students
```

	student_id	name	partner
0	1	Kailash Harjo	23
1	2	Esha Butala	1
2	3	Parveen Bhalla	3
3	4	Marlo Dugal	14
4	5	Kusum Bahri	6
5	6	Lakshmi Contractor	10
6	7	Tarun Thaker	9
7	8	Radheshyam Dey	5
8	9	Nitika Chatterjee	4
9	10	Aayushman Sant	8
10	11	David Mukhopadhyay	20
11	12	Radha Dutt	19
12	13	Munni Varghese	24
13	14	Pranab Natarajan	22
14	15	Preet Sha	16
15	16	Elias Dodiya	25
16	17	Yasmin Palan	7
17	18	Fardeen Mahabir	13
18	19	Qabeel Raman	12
19	20	Hanuman Hegde	11
20	21	Seema Kota	15
21	22	Yash Sethi	21
22	23	Chhavi Lachman	18
23	24	Radhika Suri	17
24	25	Shashank D'Alia	2
25	26	Nitish	28
26	27	Ankit	26
27	28	Rahul	17

```
# 8. Print student name -> partner name for all enrolled students
# self join
students.merge(students,how='inner',left_on='partner',right_on='student_id')[['name_x','name_y']]
```

	name_x	name_y
0	Kailash Harjo	Chhavi Lachman
1	Esha Butala	Kailash Harjo
2	Parveen Bhalla	Parveen Bhalla
3	Marlo Dugal	Pranab Natarajan
4	Kusum Bahri	Lakshmi Contractor
5	Lakshmi Contractor	Aayushman Sant
6	Tarun Thaker	Nitika Chatterjee
7	Radheshyam Dey	Kusum Bahri
8	Nitika Chatterjee	Marlo Dugal
9	Aayushman Sant	Radheshyam Dey
10	David Mukhopadhyay	Hanuman Hegde
11	Radha Dutt	Qabeel Raman
12	Munni Varghese	Radhika Suri
13	Pranab Natarajan	Yash Sethi
14	Preet Sha	Elias Dodiya
15	Elias Dodiya	Shashank D'Alia
16	Yasmin Palan	Tarun Thaker
17	Fardeen Mahabir	Munni Varghese
18	Qabeel Raman	Radha Dutt
19	Hanuman Hegde	David Mukhopadhyay
20	Seema Kota	Preet Sha
21	Yash Sethi	Seema Kota
22	Chhavi Lachman	Fardeen Mahabir
23	Radhika Suri	Yasmin Palan
24	Rahul	Yasmin Palan
25	Shashank D'Alia	Esha Butala
26	Nitish	Rahul
27	Ankit	Nitish

9. find top 3 students who did most number enrollments

```
regs.merge(students,on='student_id').groupby(['student_id','name'])['name'].count().sort_values(ascending=False).head(3)
```

```
student_id  name
23          Chhavi Lachman    6
7           Tarun Thaker     5
1           Kailash Harjo     4
Name: name, dtype: int64
```

10. find top 3 students who spent most amount of money on courses

```
regs.merge(students,on='student_id').merge(courses,on='course_id').groupby(['student_id','name'])['price'].sum().sort_values
```

```
student_id  name
23          Chhavi Lachman    22594
14          Pranab Natarajan   15096
19          Qabeel Raman      13498
Name: price, dtype: int64
```

Alternate syntax for merge

```
# students.merge(regs)
```

```
pd.merge(students,regs,how='inner',on='student_id')
```

Year	2014	2015	2016	2017	2018
Matches	34	40	40	40	39

	id	season	city	date	team1	team2	toss_winner	toss_decisi
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	f
1	2	2017	Pune	2017-	Mumbai	Rising Pune	Rising Pune	f