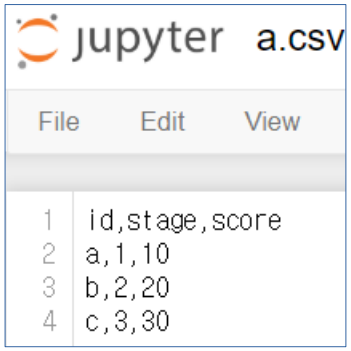
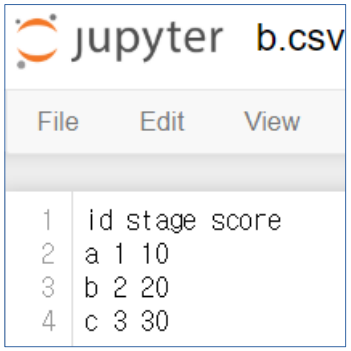
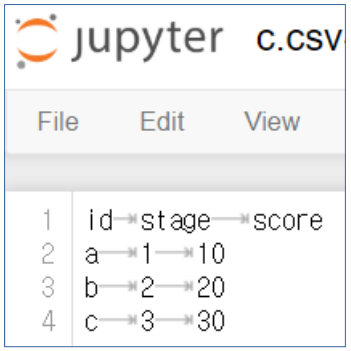
■ CSV

- 몇가지 필드를 comma(,)로 구분한 텍스트 데이터 또는 파일
- Comma Separated Variables
- MS Excel과 같은 스프레드시트, 데이터베이스 소프트웨어에서 많이 사용
- Tab Separated Values(TSV), Space Separated Values(SSV) 등을 합쳐서 Character Separated Values(CSV)라고 부르기도 한다





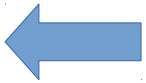


■ 데이터 불러오기

● comma(,) 구분

```
df = pd.read_csv('data/friend_list.csv')
df
```

	name	age	job
0	John	20	student
1	Jenny	30	developer
2	Nate	30	teacher
3	Julia	40	dentist
4	Brian	45	manager
5	Chris	25	intern



jupyter friend_list.csv 03/30/2018

File Edit View Language

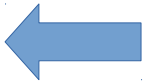
```
1 name,age,job
2 John,20,student
3 Jenny,30,developer
4 Nate,30,teacher
5 Julia,40,dentist
6 Brian,45,manager
7 Chris,25,intern
```

■ 데이터 불러오기

● comma(,) 구분

```
df = pd.read_csv('data/friend_list.txt')  # 확장자 txt
df
```

	name	age	job
0	John	20	student
1	Jenny	30	developer
2	Nate	30	teacher
3	Julia	40	dentist
4	Brian	45	manager
5	Chris	25	intern



jupyter friend_list.txt 03/30/2018

File Edit View Language

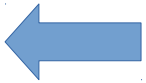
```
1 name,age,job
2 John,20,student
3 Jenny,30,developer
4 Nate,30,teacher
5 Julia,40,dentist
6 Brian,45,manager
7 Chris,25,intern
```

■ 데이터 불러오기

● tab(₩t) 구분

```
df = pd.read_csv('data/friend_list_tab.txt', delimiter='₩t')
df
```

	name	age	job
0	John	20	student
1	Jenny	30	developer
2	Nate	30	teacher
3	Julia	40	dentist
4	Brian	45	manager
5	Chris	25	intern



jupyter friend_list_tab.txt 03/30/2(

File Edit View Language

```
1 name→age→job
2 John→20→student
3 Jenny→30→developer
4 Nate→30→teacher
5 Julia→40→dentist
6 Brian→45→manager
7 Chris→25→intern
```

delimiter 미사용

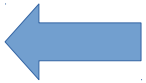
	name age job
0	John\t20\tstudent
1	Jenny\t30\tdeveloper
2	Nate\t30\tteacher
3	Julia\t40\tdentist
4	Brian\t45\tmanager
5	Chris\t25\tintern

■ 데이터 불러오기

● 공백 구분

```
df = pd.read_csv('data/friend_list_space.txt', delimiter=' ')
df
```

	name	age	job
0	John	20	student
1	Jenny	30	developer
2	Nate	30	teacher
3	Julia	40	dentist
4	Brian	45	manager
5	Chris	25	intern



jupyter friend_list_space.txt Last

File Edit View Language

```
1 name age job
2 John 20 student
3 Jenny 30 developer
4 Nate 30 teacher
5 Julia 40 dentist
6 Brian 45 manager
7 Chris 25 intern
```

delimiter 미사용

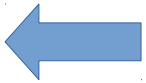
	name age job
0	John 20 student
1	Jenny 30 developer
2	Nate 30 teacher
3	Julia 40 dentist
4	Brian 45 manager
5	Chris 25 intern

■ 데이터 불러오기

● 제목이 없는 경우

```
df = pd.read_csv('data/friend_list_no_head.csv', header=None)
df
```

	0	1	2
0	John	20	student
1	Jenny	30	developer
2	Nate	30	teacher
3	Julia	40	dentist
4	Brian	45	manager
5	Chris	25	intern



jupyter friend_list_no_head.csv

File Edit View Language

```
1 John,20,student
2 Jenny,30,developer
3 Nate,30,teacher
4 Julia,40,dentist
5 Brian,45,manager
6 Chris,25,intern
7
```

header 미사용

	John	20	student
0	Jenny	30	developer
1	Nate	30	teacher
2	Julia	40	dentist
3	Brian	45	manager
4	Chris	25	intern

■ 데이터 저장하기

● 저장할 DataFrame 생성

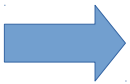
```
list = [
    ['kim', 20, 'designer'],
    ['lee', 21, 'programmer'],
    ['park', 22, 'dba']
]
df = pd.DataFrame(list, columns=['name', 'age', 'job'], index=[1, 2, 3])
df
```

	name	age	job
1	kim	20	designer
2	lee	21	programmer
3	park	22	dba

■ 데이터 저장하기

● 기본 사용

```
df.to_csv('data/save.csv')
```



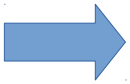
jupyter save.csv

File Edit View Langua

```
1 ,name,age,job
2 1,kim,20,designer
3 2,lee,21,programmer
4 3,park,22,dba
5
```

● header(colums) 및 index 미사용

```
df.to_csv('data/save_no_head_index.csv',
          header=False, index=False)
```



jupyter save_no_h

File Edit View Langua

```
1 kim,20,designer
2 lee,21,programmer
3 park,22,dba
4
```

■ 데이터 저장하기

● None Value 처리

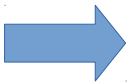
```
list = [
    ['kim', None, 'designer'],
    ['lee', 21, 'programmer'],
    ['park', 22, None]
]
df = pd.DataFrame(list, columns=['name', 'age', 'job'], index=[1, 2, 3])
df
```

	name	age	job
1	kim	NaN	designer
2	lee	21.0	programmer
3	park	22.0	None

■ 데이터 저장하기

● 기본 사용

```
df.to_csv('data/save_none_value1.csv')
```



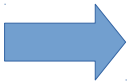
jupyter save_none

File Edit View Langua

```
1 ,name,age,job
2 1,kim,,designer
3 2,lee,21.0,programmer
4 3,park,22.0,
5
```

● header(colums) 및 index 미사용

```
df.to_csv('data/save_none_value2.csv', na_rep='-')
```



jupyter save_none

File Edit View Langua

```
1 ,name,age,job
2 1,kim,-,designer
3 2,lee,21.0,programmer
4 3,park,22.0,-
5
```