

```
1 import pandas as pd
2 import numpy as np
```

```
1 from google.colab import files
2 uploaded = files.upload()
```

파일 선택 friend_list.csv

- **friend_list.csv**(application/vnd.ms-excel) - 113 bytes, last modified: 2018. 3. 30. - 100% done
Saving friend_list.csv to friend_list.csv

```
1 dict_data2 = { '이름': ['서준', '우현', '인아'], '수학': [90, 80, 70], '영어': [98, 89, 95], '음악': [85, 95, 100], '체육': [100, 90, 90] }
2 df = pd.DataFrame(dict_data2)
3 df
```

	이름	수학	영어	음악	체육
0	서준	90	98	85	100
1	우현	80	89	95	90
2	인아	70	95	100	90

```
1 df.loc[3]=['상민', 100, 90, 100, 100]
2 df
```

	이름	수학	영어	음악	체육
0	서준	90	98	85	100
1	우현	80	89	95	90
2	인아	70	95	100	90
3	상민	100	90	100	100

• 조건주기

```
1 df[df.수학>=90] #수학점수가 90점 이상인 사람만 출력
```

	이름	수학	영어	음악	체육
0	서준	90	98	85	100
3	상민	100	90	100	100

```
1 df[(df.수학>=90) & (df.영어>=90)]
```

이름 수학 영어 음악 체육

```
1 df.query('음악==100') #문자열로 조건 주기 (변수명.query)
```

	이름	수학	영어	음악	체육
2	인아	70	95	100	90
3	상민	100	90	100	100

```
1 df['미술']=[50,60,50,60] #미술 열 추가 =(50,60,50,60)
```

```
2 df.query('미술>=60')
```

	이름	수학	영어	음악	체육	미술
1	우현	80	89	95	90	60
3	상민	100	90	100	100	60

```
1 df['평가'] = np.where(df['수학']>=90, '합격', '불합격') #조건을 주고 where을 기준으로 조건별로 출력
```

```
2 df
```

	이름	수학	영어	음악	체육	미술	평가
0	서준	90	98	85	100	50	합격
1	우현	80	89	95	90	60	불합격
2	인아	70	95	100	90	50	불합격
3	상민	100	90	100	100	60	합격

```
1 from google.colab import files
```

```
2 uploaded = files.upload()
```

파일 선택 major_list.csv

- **major_list.csv**(application/vnd.ms-excel) - 304 bytes, last modified: 2021. 5. 13. - 100% done
Saving major_list.csv to major_list.csv

```
1 mj = pd.read_csv('major_list.csv') #mj 변수에 pd.read_csv('삽입할 파일명')
```

```
2 mj
```

	name	major	sex
0	John	Computer Science	male
1	Nate	Computer Science	male
2	Abraham	Physics	male
3	Brian	Psychology	male
4	Janny	Economics	female

- 그룹

• Jennifer Computer Science female

```
1 mj_g=mj.groupby('major') # major 별로 그룹화
```

```
2 mj_g.groups
```

```
{'Computer Science': [0, 1, 6, 7], 'Economics': [4, 5, 9], 'Physics': [2], 'Psychology': [3,
```

10 Sera Psychology female

```
1 df_major_cnt = pd.DataFrame({'count':mj_g.size()}).reset_index()
```

```
2 df_major_cnt
```

	major	count
0	Computer Science	4
1	Economics	3
2	Physics	1
3	Psychology	3

```
1 ms = mj.groupby('sex')
```

```
2 ms = pd.DataFrame({'count':ms.size()}).reset_index()
```

```
3 ms
```

	sex	count
0	female	6
1	male	5

```
1 groupby_sex = mj.groupby('sex')
```

```
1 for name,group in groupby_sex:
```

```
2     print(name+' : '+str(len(group)))
```

```
3     print(group)
```

```
4     print()
```

```
5
```

```
☞ female:6
```

	name	major	sex
4	Janny	Economics	female

5	Yuna	Economics	female
6	Jeniffer	Computer Science	female
8	Zara	Psychology	female
9	Wendy	Economics	female
10	Sera	Psychology	female

male:5

	name	major	sex
0	John	Computer Science	male
1	Nate	Computer Science	male
2	Abraham	Physics	male
3	Brian	Psychology	male
7	Edward	Computer Science	male