

## [과제] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기

### 1. excel > csv 파일 형식 변환

> LMS에 제시된 엑셀 파일 Test\_result.xlsx를 CSV 타입 파일로 변환

### 2. csv > Data Frame 데이터 변환

> 변환한 CSV 파일을 읽어 데이터 프레임으로 변환

> 변환된 데이터 프레임의 정보 확인

> Index 컬럼 'Unnamed: 0' 제거

### 3. Data Frame에 'Total' 파생변수를 추가

> 'Total'은 총점으로 (Attendance + Homework + Midterm + Final)

### 4. Data Frame에 'grade' 파생변수를 추가

> 'Grade'는 총점에 대한 등급 : "A+" : 100-95, "A" : 94-90, "B+" : 89-85, "B" : 84-80,  
"C+" : 79-75, "C" : 74-70, "D+" : 69-65, "D" : 64-60, "F" : 59-0

### 5. 파생 변수로 그래프 그리기

> 'Grade'별 인원 수 분포 구하고, 확인하기

> 'Grade'별 인원 수 막대 그래프 그리기

> x-축을 높은 등급 순으로 그리기

>>[도전] 막대에 빈도 값, 표 타이틀, 축 레이블 표시하기

### 6. 데이터 프레임을 Test\_result2.csv 파일로 변환

> index는 제외하고 변환

```
In [1]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
##### 1. excel > csv 파일 형식 변환
##### > LMS에 제시된 엑셀 파일 Test_result.xlsx를 CSV 타입 파일로 변환
```

```
In [2]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
##### 2. csv > Data Frame 데이터 변환
```

```
##### > 변환한 CSV 파일을 읽어 데이터 프레임으로 변환
```

```
df_test
```

Out[2]:

	Unnamed: 0	ID	Name	Class	Attendance	Homework	Midterm	Final
0	0	2022001	고ㅇㅇ	B	14	17	22	22
1	1	2022002	석ㅇㅇ	A	18	20	27	25
2	2	2022003	강ㅇㅇ	A	20	19	22	25
3	3	2022004	민ㅇㅇ	A	19	19	24	26
4	4	2022005	지ㅇㅇ	A	20	19	24	23
...	...	...	...	...	...	...	...	...
69	69	2022070	유ㅇㅇ	B	18	19	20	23
70	70	2022071	김ㅇㅇ	B	20	19	23	19
71	71	2022072	이ㅇㅇ	B	18	17	23	24
72	72	2022073	은ㅇㅇ	A	14	18	29	26
73	73	2022074	김ㅇㅇ	B	20	17	28	27

74 rows × 8 columns

```
In [3]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
##### 2. csv > Data Frame 데이터 변환
##### > 변환된 데이터 프레임의 정보 확인
df_test.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74 entries, 0 to 73
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0   74 non-null    int64
1   ID           74 non-null    int64
2   Name         74 non-null    object
3   Class        74 non-null    object
4   Attendance   74 non-null    int64
5   Homework     74 non-null    int64
6   Midterm      74 non-null    int64
7   Final        74 non-null    int64
dtypes: int64(6), object(2)
memory usage: 4.8+ KB

```

```

In [4]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
##### 2. csv > Data Frame 데이터 변환
## Index 컬럼 'Unnamed: 0' 제거

df_test

```

Out[4]:

	ID	Name	Class	Attendance	Homework	Midterm	Final
0	2022001	고ㅇㅇ	B	14	17	22	22
1	2022002	석ㅇㅇ	A	18	20	27	25
2	2022003	강ㅇㅇ	A	20	19	22	25
3	2022004	민ㅇㅇ	A	19	19	24	26
4	2022005	지ㅇㅇ	A	20	19	24	23
...	...	...	...	...	...	...	...
69	2022070	유ㅇㅇ	B	18	19	20	23
70	2022071	김ㅇㅇ	B	20	19	23	19
71	2022072	이ㅇㅇ	B	18	17	23	24
72	2022073	은ㅇㅇ	A	14	18	29	26
73	2022074	김ㅇㅇ	B	20	17	28	27

74 rows × 7 columns

In [5]:

```
## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
#### 3. Data Frame에 'total' 파생변수를 추가
##### > 'total'은 총점으로 (Attendance + Homework + Midterm + Final)

df_test
```

Out[5]:

	ID	Name	Class	Attendance	Homework	Midterm	Final	Total
0	2022001	고ㅇㅇ	B	14	17	22	22	75
1	2022002	석ㅇㅇ	A	18	20	27	25	90
2	2022003	강ㅇㅇ	A	20	19	22	25	86
3	2022004	민ㅇㅇ	A	19	19	24	26	88
4	2022005	지ㅇㅇ	A	20	19	24	23	86
...	...	...	...	...	...	...	...	...
69	2022070	유ㅇㅇ	B	18	19	20	23	80
70	2022071	김ㅇㅇ	B	20	19	23	19	81
71	2022072	이ㅇㅇ	B	18	17	23	24	82
72	2022073	은ㅇㅇ	A	14	18	29	26	87
73	2022074	김ㅇㅇ	B	20	17	28	27	92

74 rows × 8 columns

In [6]:

```
## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
#### 4. Data Frame에 'grade' 파생변수를 추가
##### > 'Grade'는 총점에 대한 등급 : "A+" : 100-95, "A" : 94-90, "B+" : 89-85, "B" : 84-80,
#####      "C+" : 79-75, "C" : 74-70, "D" : 69-60, "F" : 59-0
```

```
df_test
```

Out[6]:

	ID	Name	Class	Attendance	Homework	Midterm	Final	Total	Grade
0	2022001	고ㅇㅇ	B	14	17	22	22	75	C+
1	2022002	석ㅇㅇ	A	18	20	27	25	90	A
2	2022003	강ㅇㅇ	A	20	19	22	25	86	B+
3	2022004	민ㅇㅇ	A	19	19	24	26	88	B+
4	2022005	지ㅇㅇ	A	20	19	24	23	86	B+
...	...	...	...	...	...	...	...	...	...
69	2022070	유ㅇㅇ	B	18	19	20	23	80	B
70	2022071	김ㅇㅇ	B	20	19	23	19	81	B
71	2022072	이ㅇㅇ	B	18	17	23	24	82	B
72	2022073	은ㅇㅇ	A	14	18	29	26	87	B+
73	2022074	김ㅇㅇ	B	20	17	28	27	92	A

74 rows × 9 columns

In [7]:

```
## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
#### 5. 파생 변수로 그래프 그리기
##### > 'Grade'별 인원 수 분포 확인하기

count_grade
```

Out[7]:

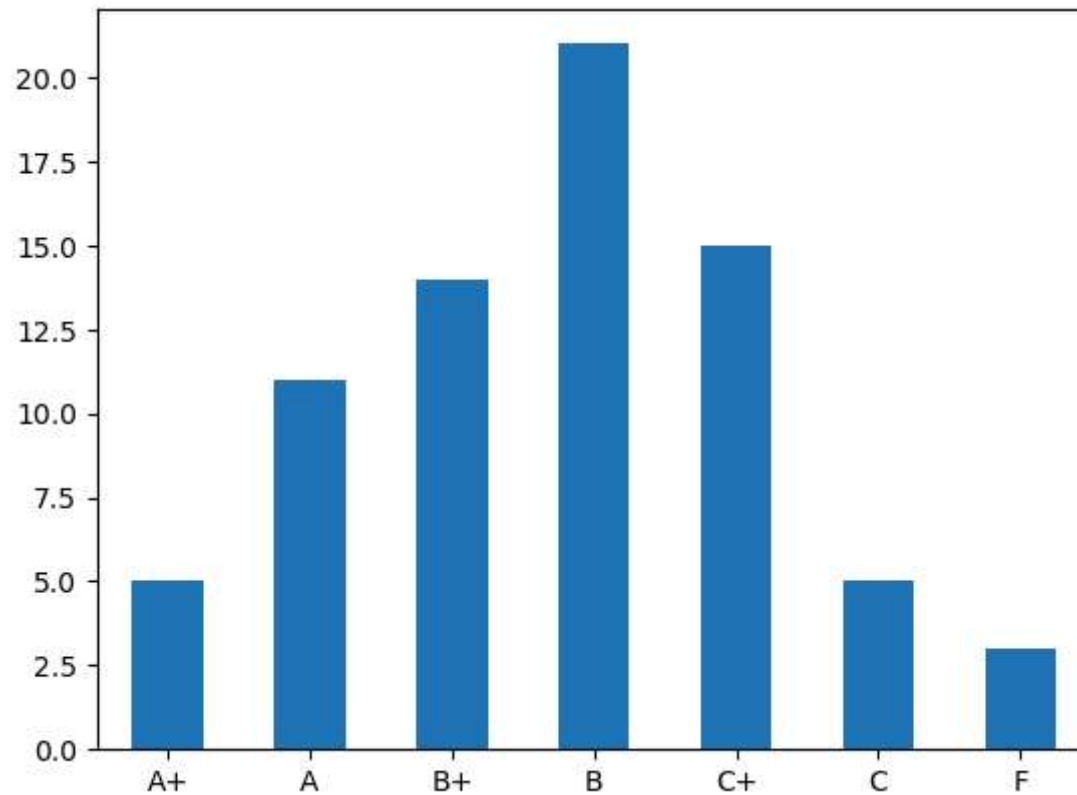
```
B      21
C+     15
B+     14
A       11
A+      5
C        5
F        3
Name: Grade, dtype: int64
```

In [8]:

```
## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
#### 5. 파생 변수로 그래프 그리기
##### > x-축을 높은 등급 순으로 막대 그리기
```

```
count_grade.plot.bar(rot = 0)
```

Out[8]: <AxesSubplot:>



```
In [9]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기
##### 5. 파생 변수로 그래프 그리기
##### > x-축을 높은 등급 순으로 막대 그리기
##### >>[도전] 막대에 빈도 값 표시하기
```

```
#데이터 추출하기
```

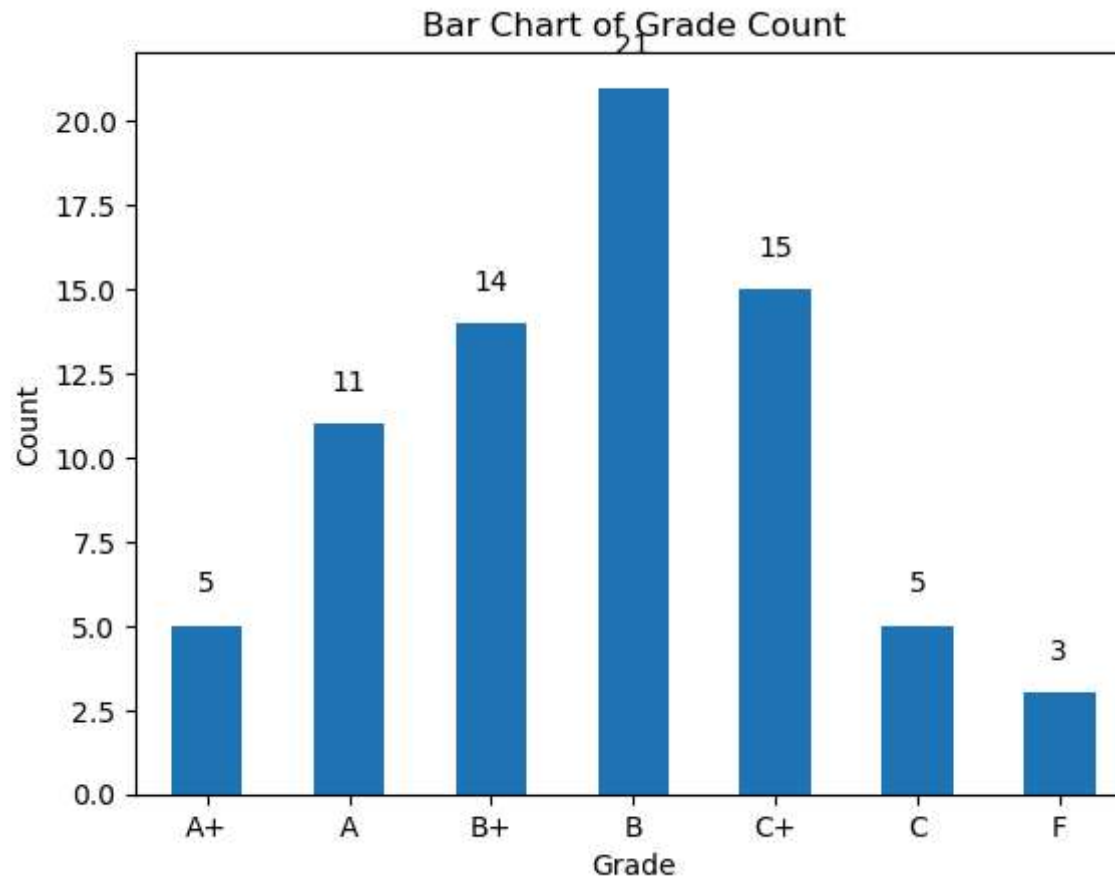
```
# 막대 그래프 그리기
```

```
ax = count_grade.plot.bar(rot = 0)
```

```
# 막대 위에 값 표기
```

```
# 그래프에 제목과 레이블 추가
```

```
# 그래프 표시  
plt.show()
```



```
In [ ]: ## [응용 실습] excel > csv > Data Frame > Data Frame 파생변수 추가 > 그래프 그리기  
##### 6. 데이터 프레임을 csv 파일로 변환  
##### > index는 제외하고 변환
```



```
df_test
```

```
Out[ ]:
```

	ID	Name	Class	Attendance	Homework	Midterm	Final	Total	Grade
<b>0</b>	2022001	고ㅇㅇ	B	14	17	22	22	75	C+
<b>1</b>	2022002	석ㅇㅇ	A	18	20	27	25	90	A
<b>2</b>	2022003	강ㅇㅇ	A	20	19	22	25	86	B+
<b>3</b>	2022004	민ㅇㅇ	A	19	19	24	26	88	B+
<b>4</b>	2022005	지ㅇㅇ	A	20	19	24	23	86	B+
...	...	...	...	...	...	...	...	...	...
<b>69</b>	2022070	유ㅇㅇ	B	18	19	20	23	80	B
<b>70</b>	2022071	김ㅇㅇ	B	20	19	23	19	81	B
<b>71</b>	2022072	이ㅇㅇ	B	18	17	23	24	82	B
<b>72</b>	2022073	은ㅇㅇ	A	14	18	29	26	87	B+
<b>73</b>	2022074	김ㅇㅇ	B	20	17	28	27	92	A

74 rows × 9 columns

```
In [ ]:
```