

## [ 판다스 실습(4) ]

- (1) "./data/emp.csv"파일을 읽고 df 변수에 저장한다.
- (2) 다음과 같이 df의 모든 행과 열을 출력한다.

	empno	ename	job	mgr	hiredate	sal	comm	deptno
0	7369	SMITH	CLERK	7902.0	1980-12-17	800	NaN	20
1	7499	ALLEN	SALESMAN	7698.0	1981-02-20	1600	300.0	30
2	7521	WARD	SALESMAN	7698.0	1981-02-03	1250	500.0	30
3	7566	JONES	MANAGER	7839.0	1981-03-02	2975	NaN	20
4	7654	MARTIN	SALESMAN	7698.0	1981-10-22	1250	1400.0	30
5	7698	BLAKE	MANAGER	7839.0	1981-05-01	2850	NaN	30
6	7782	CLARK	MANAGER	7839.0	1981-09-06	2450	NaN	10
7	7788	SCOTT	ANALYST	7566.0	1982-12-08	3000	NaN	20
8	7839	KING	PRESIDENT	NaN	1981-11-17	5000	NaN	10
9	7844	TURNER	SALESMAN	7698.0	1984-10-08	1500	NaN	30
10	7876	ADAMS	CLERK	7788.0	1983-01-12	1100	NaN	20
11	7900	JAMES	CLERK	7698.0	1981-12-03	950	NaN	30
12	7902	FORD	ANALYST	7566.0	1981-12-13	3000	NaN	20
13	7934	MILLER	CLERK	7782.0	1982-01-25	1300	NaN	10

또한 다음과 같이 월급이 많은 순으로도 출력하는데 df 를 변경하지는 않는다.

	empno	ename	job	mgr	hiredate	sal	comm	deptno
8	7839	KING	PRESIDENT	NaN	1981-11-17	5000	NaN	10
7	7788	SCOTT	ANALYST	7566.0	1982-12-08	3000	NaN	20
12	7902	FORD	ANALYST	7566.0	1981-12-13	3000	NaN	20
3	7566	JONES	MANAGER	7839.0	1981-03-02	2975	NaN	20
5	7698	BLAKE	MANAGER	7839.0	1981-05-01	2850	NaN	30
6	7782	CLARK	MANAGER	7839.0	1981-09-06	2450	NaN	10
1	7499	ALLEN	SALESMAN	7698.0	1981-02-20	1600	300.0	30
9	7844	TURNER	SALESMAN	7698.0	1984-10-08	1500	NaN	30
13	7934	MILLER	CLERK	7782.0	1982-01-25	1300	NaN	10
2	7521	WARD	SALESMAN	7698.0	1981-02-03	1250	500.0	30
4	7654	MARTIN	SALESMAN	7698.0	1981-10-22	1250	1400.0	30
10	7876	ADAMS	CLERK	7788.0	1983-01-12	1100	NaN	20
11	7900	JAMES	CLERK	7698.0	1981-12-03	950	NaN	30
0	7369	SMITH	CLERK	7902.0	1980-12-17	800	NaN	20

- (3) 다음과 같이 df 에 할당된 데이터 프레임이 몇행 몇열인지 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

(14, 8)

(4) 다음 결과를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   empno       14 non-null    int64
1   ename       14 non-null    object
2   job         14 non-null    object
3   mgr         13 non-null    float64
4   hiredate    14 non-null    object
5   sal         14 non-null    int64
6   comm        3 non-null     float64
7   deptno      14 non-null    int64
dtypes: float64(2), int64(3), object(3)
memory usage: 1.0+ KB
```

(5) 다음 결과를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

	empno	mgr	sal	comm	deptno
count	14.000000	13.000000	14.000000	3.000000	14.000000
mean	7726.571429	7739.307692	2073.214286	733.333333	22.142857
std	178.294361	103.714660	1182.503224	585.946528	8.017837
min	7369.000000	7566.000000	800.000000	300.000000	10.000000
25%	7588.000000	7698.000000	1250.000000	400.000000	20.000000
50%	7785.000000	7698.000000	1550.000000	500.000000	20.000000
75%	7868.000000	7839.000000	2943.750000	950.000000	30.000000
max	7934.000000	7902.000000	5000.000000	1400.000000	30.000000

(6) 다음과 같이 각 변수들의 타입 정보를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```
empno      int64
ename      object
job        object
mgr        float64
hiredate   object
sal        int64
comm       float64
deptno     int64
dtype: object
```

(7) 다음과 같이 변수별 데이터의 갯수를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```
empno      14
ename      14
job        14
mgr        13
hiredate   14
sal        14
comm       3
deptno     14
dtype: int64
```

(8) 다음 결과를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

	empno	ename	job	mgr	hiredate	sal	comm	deptno
<b>count</b>	14.000000	14	14	13.000000	14	14.000000	3.000000	14.000000
<b>unique</b>	NaN	14	5	NaN	14	NaN	NaN	NaN
<b>top</b>	NaN	ALLEN	SALESMAN	NaN	1981-05-01	NaN	NaN	NaN
<b>freq</b>	NaN	1	4	NaN	1	NaN	NaN	NaN
<b>mean</b>	7726.571429	NaN	NaN	7739.307692	NaN	2073.214286	733.333333	22.142857
<b>std</b>	178.294361	NaN	NaN	103.714660	NaN	1182.503224	585.946528	8.017837
<b>min</b>	7369.000000	NaN	NaN	7566.000000	NaN	800.000000	300.000000	10.000000
<b>25%</b>	7588.000000	NaN	NaN	7698.000000	NaN	1250.000000	400.000000	20.000000
<b>50%</b>	7785.000000	NaN	NaN	7698.000000	NaN	1550.000000	500.000000	20.000000
<b>75%</b>	7868.000000	NaN	NaN	7839.000000	NaN	2943.750000	950.000000	30.000000
<b>max</b>	7934.000000	NaN	NaN	7902.000000	NaN	5000.000000	1400.000000	30.000000

(9) 다음 과 같이 각 부서별 재직 인원수를 출력한다. 결과를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```

30      6
20      5
10      3
Name: deptno, dtype: int64

```

(10) 다음과 같이 각 직무별 인원수를 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```

SALESMAN      4
CLERK         4
MANAGER       3
ANALYST       2
PRESIDENT     1
Name: job, dtype: int64

```

(11) 다음과 같이 수치형 변수들의 값의 평균을 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```

empno      7726.571429
mgr         7739.307692
sal         2073.214286
comm        733.333333
deptno      22.142857
dtype: float64

```

(12) 다음과 같이 월급의 평균만 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```

2073.214285714286

```

(13) 다음과 같이 각 변수별 최대값을 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```
empno      7934
ename      WARD
job        SALESMAN
mgr        7902.0
hiredate   1984-10-08
sal        5000
comm       1400.0
deptno     30
dtype: object
```

(14) 다음과 같이 각 변수별 최소값을 출력한다. Part 3 에서 학습한 API 를 검토해서 다음 결과를 출력하는 함수 또는 메서드를 찾아서 해결한다.

```
empno      7369
ename      ADAMS
job        ANALYST
mgr        7566.0
hiredate   1980-12-17
sal        800
comm       300.0
deptno     10
dtype: object
```

(15) df 에서 직원이름, 월급 그리고 커미션 정보만으로 구성된 데이터프레임을 생성하고 df1 변수에 저장하고 앞에서부터 5개 행만 출력한다.

	ename	sal	comm
0	SMITH	800	NaN
1	ALLEN	1600	300.0
2	WARD	1250	500.0
3	JONES	2975	NaN
4	MARTIN	1250	1400.0

(16) df1 에서 ename 컬럼을 인덱스로 변경하는데 df1 에 반영한다.

	sal	comm
ename		
SMITH	800	NaN
ALLEN	1600	300.0
WARD	1250	500.0
JONES	2975	NaN
MARTIN	1250	1400.0

(17) df1 을 가지고 다음과 같이 시각화 결과를 출력해 본다.

```
<AxesSubplot:xlabel='ename'>
```

