```
In [1]: import pandas as pd
In [2]: #연습문제 가져오기
        df = pd.read_csv("연습데이터셋1.csv", encoding = 'cp949')
In [2]:
        df.head()
Out[2]:
          student_id name
                          age score grade city registration_date
        0
               S001 이민준 23.0
                                85.5
                                        B 서울
                                                    2023-01-15
        1
               S002 박서연 25.0
                                92.0
                                        A 부산
                                                    2023-02-20
               S003 김지훈 NaN
                                78.0
                                        C 서울
                                                    2023-03-10
        3
               S004 최예은 22.0
                                NaN
                                        B 인처
                                                    2024-04-05
               S005 정하준 28.0
                                95.5
                                        A 미상
                                                    2024-05-12
In [3]: df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 8 entries, 0 to 7
        Data columns (total 7 columns):
                                Non-Null Count Dtype
         #
            Column
        - - -
         0
             student id
                                8 non-null
                                                 object
                                8 non-null
                                                 object
         1
             name
                                7 non-null
                                                 float64
         3
                                7 non-null
                                                 float64
             score
             arade
                                7 non-null
                                                 object
         5
                                8 non-null
                                                 object
             citv
         6
             registration_date 7 non-null
                                                 object
        dtypes: float64(2), object(5)
        memory usage: 576.0+ bytes
        df.isnull(), df.dropna(), df.fillna ()
In [4]: #null값 총합
        df.isnull().sum()
        student id
Out[4]:
        name
        age
                             1
        score
                             1
        grade
                             0
        city
        registration_date
        dtype: int64
In [ ]:
In [5]:
        #행 전체 삭제 (dropna)
        drop = df.dropna()
        drop.info()
        <class 'pandas.core.frame.DataFrame'>
        Index: 5 entries, 0 to 7
        Data columns (total 7 columns):
                                Non-Null Count Dtype
            Column
         #
         0
                                5 non-null
             student_id
                                                 object
                                5 non-null
         1
             name
                                                 object
         2
             age
                                5 non-null
                                                 float64
         3
             score
                                5 non-null
                                                 float64
             grade
                                5 non-null
                                                 object
             city
                                5 non-null
                                                 object
             registration_date 5 non-null
                                                 object
        dtypes: float64(2), object(5)
        memory usage: 320.0+ bytes
In [6]: # 전체 열 삭제 (dropna)
        co drop = df.dropna(axis=1)
        co_drop.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 8 entries, \theta to 7
        Data columns (total 3 columns):
                                         Dtype
         # Column
                         Non-Null Count
             student_id 8 non-null
         0
                                         obiect
         1
             name
                         8 non-null
                                         object
```

8 non-null

2 city
dtypes: object(3)

memory usage: 320.0+ bytes

object

```
In [ ]:
In [7]: df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 8 entries, 0 to 7
        Data columns (total 7 columns):
                                Non-Null Count Dtype
            Column
                                 -----
         0
             student_id
                               8 non-null
                                                 object
         1
             name
                                 8 non-null
                                                 object
         2
                                 7 non-null
                                                 float64
             age
         3
                                 7 non-null
             score
                                                 float64
         4
             grade
                                 7 non-null
                                                 object
            city
                                8 non-null
                                                 object
         6 registration_date 7 non-null
                                                 object
        dtypes: float64(2), object(5)
        memory usage: 576.0+ bytes
In [8]: df.head()
           student_id name age score grade city registration_date
Out[8]:
        0
               S001 이민준 23.0
                                85.5
                                         B 서울
                                                     2023-01-15
        1
               S002 박서연 25.0
                                92.0
                                         A 부산
                                                     2023-02-20
        2
               S003 김지훈 NaN
                                                     2023-03-10
                                78.0
                                         C 서울
        3
               S004 최예은 22.0
                                NaN
                                         B 인천
                                                     2024-04-05
               S005 정하준 28.0
                                         A 미상
                                                     2024-05-12
                                95.5
        Fillna 예제
        예제 1: 전체 컬럼
          • df2=df.fillna('None')
        예제 2: 컬럼 하나
          • df2['Discount'] = df['Discount'].fillna(0)
        예제 3: 그룹 컬럼
          • df2[['Discount','Fee']] = df[['Discount','Fee']].fillna(0)
        예제 4: 그룹 컬럼인데 다른 값
          df2 = df.fillna(value={'Discount':0,'Fee':10000})
        예제 5: 제한(limit)
          • df2=df.fillna(value={'Discount':0,'Fee':0},limit=1)
In [9]: df_fill = df.fillna(value = {'age':30, 'score':80.0, 'grade': 'None', 'registration_date' :'2023-06-30' })
        df_fill.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 8 entries, 0 to 7
        Data columns (total 7 columns):
                                Non-Null Count Dtype
         #
             Column
        - - -
         0
             student id
                                 8 non-null
                                                  object
         1
             name
                                 8 non-null
                                                  object
         2
             age
                                 8 non-null
                                                  float64
         3
             score
                                 8 non-null
                                                  float64
         4
             grade
                                 8 non-null
                                                 object
         5
                                 8 non-null
                                                 object
             citv
            registration_date 8 non-null
         6
                                                 object
        dtypes: float64(2), object(5)
        memory usage: 576.0+ bytes
```

Astype

In [10]: df_fill.dtypes

```
Out[10]: student_id
          name
                                 object
                                float64
          age
          score
                                float64
          grade
                                 obiect
          city
                                 object
          registration date
                                 object
          dtype: object
           • object ('Kelly','James')
           • float64 (3.14,2.5)
           • int(4,5,6,7,8,9)
In [11]: da = df_fill.astype({'score':'object', 'age':'object'})
          da.dtypes
          student_id
                                object
                                object
          name
                                object
          age
          score
                                object
          grade
                                object
          city
                                object
          registration_date
                                object
          dtype: object
         df.replace
In [12]: df.head()
            student_id name
                                        grade city registration_date
Out[12]:
                             age
                                  score
          0
                                           B 서울
                                                        2023-01-15
                 S001 이민준
                            23.0
                                   85.5
                 S002
                     박서연
                             25.0
                                   92.0
                                           A 부산
                                                        2023-02-20
          2
                 S003
                      김지훈
                            NaN
                                   78.0
                                           C 서울
                                                        2023-03-10
          3
                 S004
                      최예은
                            22.0
                                   NaN
                                           B 인천
                                                        2024-04-05
                 S005 정하준
                            28.0
                                   95.5
                                           A 미상
                                                        2024-05-12
In [13]: df['city'].value_counts()
          city
Out[13]:
          서울
                 2
          부산
          인천
                 1
          미상
                 1
          광주
          Name: count, dtype: int64
In [14]:
         df['city'] = df['city'].replace({'서울':'Seoul', '부산': 'Busan', '인천':'Incheon','미상':'None','광주':'Gwangju'})
          df.head()
Out[14]:
            student_id name
                             age score grade
                                                 city
                                                    registration_date
          0
                 S001 이민준
                            23.0
                                   85.5
                                           В
                                               Seoul
                                                          2023-01-15
          1
                 S002
                      박서연
                             25.0
                                   92.0
                                               Busan
                                                          2023-02-20
                 S003
                      김지훈
                            NaN
                                   78.0
                                           С
                                               Seoul
                                                          2023-03-10
                                                          2024-04-05
          3
                 S004 최예은 22.0
                                   NaN
                                           B Incheon
                 S005 정하준 28.0
                                   95.5
                                                None
                                                          2024-05-12
          필터링
In [15]: df_fill[df_fill['score']>80]
Out[15]:
            student_id name
                             age score
                                        grade
                                              city registration_date
          0
                                   85.5
                                                        2023-01-15
                 S001 이민준
                            23.0
                                             서울
                 S002 박서연
                            25.0
                                   92.0
                                           A 부산
                                                        2023-02-20
          4
                 S005 정하준
                            28.0
                                   95.5
                                           A 미상
                                                        2024-05-12
                                                        2024-07-21
                 S007 한도윤 29.0
                                   88.0
                                           B 광주
In [16]:
          #And
          df_fill[(df_fill['score'] > 80) \& (df_fill['age'] >= 25)]
```

object

```
student_id name age score grade city registration_date
Out[16]:
                 S002 박서연
                             25.0
                                   92.0
                                           A 부산
                                                        2023-02-20
                 S005
                      정하준
                            28.0
                                           A 미상
                                                        2024-05-12
                 S007 한도윤 29.0
                                   88.0
                                           B 광주
                                                        2024-07-21
In [17]:
          #0r
          df_fill[(df_fill['score'] > 80) | (df_fill['age'] >= 25)]
Out[17]:
            student_id name
                             age score grade city registration_date
                 S001
                      이민준
                             23.0
                                   85.5
                                           B 서울
                                                        2023-01-15
                 S002 박서연
                             25.0
                                           A 부산
                                                        2023-02-20
          2
                      김지훈
                                                        2023-03-10
                 S003
                             30.0
                                   78.0
                                           C 서울
                 S005
                      정하준
                             28.0
                                   95.5
                                           A 미상
                                                        2024-05-12
                                                        2023-06-30
                 S006
                      윤채원
                            25.0
                                   -1.0
                                        None 부산
                                                        2024-07-21
                 S007 한도윤 29.0
                                   88.0
                                           B 광주
```

인덱스 설정 및 초기화

```
In [16]: df.head()
Out[16]:
             student_id name
                                                    city registration_date
                              age score grade
                  S001
                       이민준
                               23.0
                                     85.5
                                                  Seoul
                                                              2023-01-15
                  S002 박서연
                              25.0
                                     92.0
                                                  Busan
                                                              2023-02-20
                  S003
                        김지훈
                                              С
                                                              2023-03-10
                              NaN
                                     78.0
                                                  Seoul
                  S004
                       최예은
                              22.0
                                     NaN
                                                              2024-04-05
                  S005 정하준 28.0
                                     95.5
                                                              2024-05-12
                                                   None
In [18]:
          ds = df.set_index('student_id')
                                                  city registration_date
Out[18]:
                     name age score grade
          student_id
               S001 이민준 23.0
                                                 Seoul
                                                            2023-01-15
                                                            2023-02-20
               S002 박서연
                            25.0
                                  92.0
                                                Busan
               S003
                     김지훈
                            NaN
                                   78.0
                                                 Seoul
                                                            2023-03-10
               S004
                    최예은
                            22.0
                                  NaN
                                              Incheon
                                                            2024-04-05
               S005
                     정하준
                            28.0
                                   95.5
                                                            2024-05-12
                                                 None
               S006
                     윤채원
                            25.0
                                   -1.0
                                         NaN
                                                Busan
                                                                 NaN
                           29.0
                                   88.0
                                              Gwangju
                                                            2024-07-21
               S007
                     하도윤
               S008 신유나 21.0
                                                            2024-08-30
                                   76.5
                                                 Seoul
          dr = df.reset_index(drop = True)
In [19]:
             student_id
                                                     city registration_date
Out[19]:
                        name
                               age
                                    score
                                          grade
                                                               2023-01-15
                  S001 이민준
                              23.0
                                     85.5
                                                   Seoul
                  S002
                       박서연
                              25.0
                                     92.0
                                                   Busan
                                                               2023-02-20
                  S003
                                              С
                                                               2023-03-10
                        김지훈
                              NaN
                                     78.0
                                                   Seoul
                  S004
                       최예은
                              22.0
                                                  Incheon
                                                               2024-04-05
                  S005 정하준
                              28.0
                                                               2024-05-12
                                     95.5
                                              Α
                                                   None
                  S006 윤채원
                              25.0
                                     -1.0
                                           NaN
                                                   Busan
                                                                    NaN
          6
                  S007
                        한도윤
                               29.0
                                     88.0
                                                               2024-07-21
                                                 Gwangju
```

시계열 데이터 처리

신유나 21.0

76.5

S008

In [20]: df_fill.info()

2024-08-30

Seoul

```
<class 'pandas.core.frame.DataFrame'>
         RangeIndex: 8 entries, 0 to 7
         Data columns (total 7 columns):
                                Non-Null Count Dtype
         # Column
                                 -----
         0
              student_id
                               8 non-null
                                                object
              name
                                8 non-null
                                                object
          2
                                8 non-null
                                                float64
              age
          3
              score
                                8 non-null
                                                float64
          4
              grade
                                8 non-null
                                                object
          5
              city
                                8 non-null
                                                object
          6 registration_date 8 non-null
                                                object
         dtypes: float64(2), object(5)
         memory usage: 576.0+ bytes
In [21]: df_fill['registration_date'] = pd.to_datetime(df_fill['registration_date'])
         df_fill.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 8 \text{ entries}, 0 \text{ to } 7
         Data columns (total 7 columns):
                                Non-Null Count Dtype
         # Column
         ---
         0
              student_id
                                8 non-null
                                                object
                                8 non-null
          1
              name
                                                object
          2
                                8 non-null
                                                float64
              age
          3
                                8 non-null
                                                float64
              score
          4
              grade
                                8 non-null
                                                object
          5
             city
                                8 non-null
                                                object
          6 registration_date 8 non-null
                                                datetime64[ns]
         dtypes: datetime64[ns](1), float64(2), object(4)
         memory usage: 576.0+ bytes
```

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js