In [4]:
```
# Day 58 - Loading & Exploring Data with Pandas
# ---------------------------------------------
import pandas as pd
```

In [5]:
```
# Step 1:  Checking Pandas version
print("Pandas version:", pd.__version__)
```

```
Pandas version: 2.3.3
```

In [6]:
```
# Step 2: Loading CSV file (Example: student.csv)
student_data = pd.read_csv("students.csv")
print("\n--- Student Data Loaded Successfully ---")
print(student_data.head())
```

```
--- Student Data Loaded Successfully ---
         id first_name last_name date_of_birth               ethnicity gender
\
0  111111.0       John       Doe       01/2000                Hispanic      M
1  111112.0       Jane     Smith       05/2001                Hispanic      F
2  111113.0      Sarah    Thomas       21/2002                Hispanic      M
3  111114.0      Frank     Brown       13/2002  Race/ethnicity unknown      M
4  111115.0       Mike     Davis       31/2001                   White      F

     status entry_academic_period  exclusion_type  act_composite  ...  \
0        FT             Fall 2008             NaN            NaN  ...
1  TRANSFER             Fall 2006             NaN            NaN  ...
2      FTFT             Fall 2006             NaN           14.0  ...
3      FTFT             Fall 2006             NaN            NaN  ...
4      FTFT             Fall 2007             NaN           22.0  ...

   sat_reading hs_gpa      hs_city    hs_state   hs_zip               email  \
0          NaN   2.71  Albuquerque  New Mexico  87112.0     jdoe@example.com
1          NaN   3.73     New York    New York  10009.0   jsmith@example.com
2          NaN   2.64      Pheonix     Arizona  85006.0  sthomas@example.com
3        210.0   3.68      Pheonix     Arizona  85015.0   fbrown@example.com
4          NaN   3.46      Seattle  Washington  98106.0   mdavis@example.com

   entry_age    ged english_2nd_language first_generation
0       17.9  False                False             True
1       18.1  False                False             True
2       17.6  False                False            False
3       19.0   True                False             True
4       18.2  False                 True            False

[5 rows x 26 columns]
```

In [7]:
```
# Step 3: Exploring the data
print("\nShape of the dataset:", student_data.shape)
print("\nColumns in dataset:", student_data.columns.tolist())
print("\nInformation about dataset:")
print(student_data.info())
```

```
Shape of the dataset: (57, 26)

Columns in dataset: ['id', 'first_name', 'last_name', 'date_of_birth', 'ethnici
ty', 'gender', 'status', 'entry_academic_period', 'exclusion_type', 'act_compos
ite', 'act_math', 'act_english', 'act_reading', 'sat_combined', 'sat_math', 'sa
t_verbal', 'sat_reading', 'hs_gpa', 'hs_city', 'hs_state', 'hs_zip', 'email',
'entry_age', 'ged', 'english_2nd_language', 'first_generation']

Information about dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57 entries, 0 to 56
Data columns (total 26 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   id                     9 non-null      float64
 1   first_name             9 non-null      object
 2   last_name              9 non-null      object
 3   date_of_birth          9 non-null      object
 4   ethnicity              9 non-null      object
 5   gender                 9 non-null      object
 6   status                 9 non-null      object
 7   entry_academic_period  9 non-null      object
 8   exclusion_type         0 non-null      float64
 9   act_composite          5 non-null      float64
 10  act_math               5 non-null      float64
 11  act_english            6 non-null      object
 12  act_reading            5 non-null      float64
 13  sat_combined           2 non-null      float64
 14  sat_math               2 non-null      float64
 15  sat_verbal             2 non-null      float64
 16  sat_reading            2 non-null      float64
 17  hs_gpa                 8 non-null      float64
 18  hs_city                9 non-null      object
 19  hs_state               9 non-null      object
 20  hs_zip                 9 non-null      float64
 21  email                  9 non-null      object
 22  entry_age              9 non-null      float64
 23  ged                    9 non-null      object
 24  english_2nd_language   9 non-null      object
 25  first_generation       9 non-null      object
dtypes: float64(12), object(14)
memory usage: 11.7+ KB
None
```

In [8]:

```python
# Step 4: Adding a new column dynamically
student_data.insert(0, "ID", range(1, len(student_data) + 1))
print("\nAfter inserting new 'ID' column:")
print(student_data.head())
```

```
After inserting new 'ID' column:
   ID        id first_name last_name date_of_birth               ethnicity  \
0   1  111111.0       John       Doe       01/2000                Hispanic
1   2  111112.0       Jane     Smith       05/2001                Hispanic
2   3  111113.0      Sarah    Thomas       21/2002                Hispanic
3   4  111114.0      Frank     Brown       13/2002  Race/ethnicity unknown
4   5  111115.0       Mike     Davis       31/2001                   White

  gender    status entry_academic_period  exclusion_type  ...  sat_reading  \
0      M        FT             Fall 2008             NaN  ...          NaN
1      F  TRANSFER             Fall 2006             NaN  ...          NaN
2      M      FTFT             Fall 2006             NaN  ...          NaN
3      M      FTFT             Fall 2006             NaN  ...        210.0
4      F      FTFT             Fall 2007             NaN  ...          NaN

   hs_gpa      hs_city    hs_state   hs_zip               email  entry_age  \
0    2.71  Albuquerque  New Mexico  87112.0     jdoe@example.com       17.9
1    3.73     New York    New York  10009.0   jsmith@example.com       18.1
2    2.64      Pheonix     Arizona  85006.0  sthomas@example.com       17.6
3    3.68      Pheonix     Arizona  85015.0   fbrown@example.com       19.0
4    3.46      Seattle  Washington  98106.0   mdavis@example.com       18.2

     ged  english_2nd_language first_generation
0  False                 False            True
1  False                 False            True
2  False                 False           False
3   True                 False            True
4  False                  True           False

[5 rows x 27 columns]
```

In [9]:
```
# Step 5: Loading another dataset (Titanic)
titanic = pd.read_csv("titanic.csv")
print("\n--- Titanic Dataset Loaded Successfully ---")
print(titanic.head())
```

```
--- Titanic Dataset Loaded Successfully ---
   survived  pclass                                               name  \
0         0       3                            Braund, Mr. Owen Harris
1         1       1  Cumings, Mrs. John Bradley (Florence Briggs Th...
2         1       3                             Heikkinen, Miss. Laina
3         1       1       Futrelle, Mrs. Jacques Heath (Lily May Peel)
4         0       3                           Allen, Mr. William Henry

      sex   age     fare  sibsp  parch
0    male  22.0   7.2500      1      0
1  female  38.0  71.2833      1      0
2  female  26.0   7.9250      0      0
3  female  35.0  53.1000      1      0
4    male  35.0   8.0500      0      0
```

In [10]:
```
# Step 6: Basic EDA (Exploratory Data Analysis)
print("\nShape of Titanic dataset:", titanic.shape)
print("\nColumn names:", titanic.columns.tolist())
```

```
print("\nDataset info:")
print(titanic.info())
```

```
Shape of Titanic dataset: (714, 8)

Column names: ['survived', 'pclass', 'name', 'sex', 'age', 'fare', 'sibsp', 'pa
rch']

Dataset info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 714 entries, 0 to 713
Data columns (total 8 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   survived  714 non-null    int64
 1   pclass    714 non-null    int64
 2   name      714 non-null    object
 3   sex       714 non-null    object
 4   age       714 non-null    float64
 5   fare      714 non-null    float64
 6   sibsp     714 non-null    int64
 7   parch     714 non-null    int64
dtypes: float64(2), int64(4), object(2)
memory usage: 44.8+ KB
None
```

In [11]:
```
# Step 7: Checking few rows from bottom
print("\nLast 5 rows of Titanic dataset:")
print(titanic.tail())
```

```
Last 5 rows of Titanic dataset:
     survived  pclass                                     name     sex   age  \
709         0       3  Rice, Mrs. William (Margaret Norton)  female  39.0
710         0       2                 Montvila, Rev. Juozas    male  27.0
711         1       1          Graham, Miss. Margaret Edith  female  19.0
712         1       1                 Behr, Mr. Karl Howell    male  26.0
713         0       3                   Dooley, Mr. Patrick    male  32.0

       fare  sibsp  parch
709  29.125      0      5
710  13.000      0      0
711  30.000      0      0
712  30.000      0      0
713   7.750      0      0
```

In [12]:
```
# Step 8: Checking summary statistics
print("\nStatistical Summary:")
print(titanic.describe())
```

Statistical Summary:

	survived	pclass	age	fare	sibsp	parch
count	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000
mean	0.406162	2.236695	29.699118	34.694514	0.512605	0.431373
std	0.491460	0.838250	14.526497	52.918930	0.929783	0.853289
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	1.000000	20.125000	8.050000	0.000000	0.000000
50%	0.000000	2.000000	28.000000	15.741700	0.000000	0.000000
75%	1.000000	3.000000	38.000000	33.375000	1.000000	1.000000
max	1.000000	3.000000	80.000000	512.329200	5.000000	6.000000