Untitled11

April 21, 2024

1 Python Data Cleaning and File Handling

```
[1]: import pandas as pd
    # Define the data
    data = {
        'A': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
        'B': ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j'],
        'C': [True, False, True, False, True, False, True, False, True, False],
        'D': [10.5, 20.5, 30.5, 40.5, 50.5, 60.5, 70.5, 80.5, 90.5, 100.5],
        'E': ['apple', 'banana', 'cherry', 'date', 'elderberry', 'fig', 'grape', |
      'F': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
        'G': [1.1, 2.2, 3.3, 4.4, 5.5, 6.6, 7.7, 8.8, 9.9, 10.1],
        'H': ['red', 'green', 'blue', 'yellow', 'orange', 'purple', 'pink', 'cyan', 

¬'magenta', 'black'],
        'I': [True, True, False, False, True, True, False, False, True, True],
        'J': ['dog', 'cat', 'fish', 'bird', 'rabbit', 'hamster', 'turtle', 'snake', 
     # Create the DataFrame
    df = pd.DataFrame(data)
    # Display the DataFrame
    print(df)
```

```
F
                                                                    Ι
    Α
               C
                       D
                                                  G
                                                            Η
                                                                              J
       В
                                     Ε
    1
            True
                    10.5
0
       a
                                apple
                                         100
                                                1.1
                                                          red
                                                                 True
                                                                            dog
    2
           False
                    20.5
                                         200
                                                                 True
1
       b
                               banana
                                                2.2
                                                        green
                                                                            cat
2
    3
       С
            True
                    30.5
                               cherry
                                         300
                                                3.3
                                                         blue
                                                               False
                                                                           fish
3
    4
       d False
                    40.5
                                 date
                                         400
                                                4.4
                                                       yellow
                                                               False
                                                                           bird
4
    5
                    50.5
       е
            True
                          elderberry
                                         500
                                                5.5
                                                       orange
                                                                 True
                                                                         rabbit
5
    6
       f
           False
                    60.5
                                  fig
                                         600
                                                6.6
                                                       purple
                                                                 True
                                                                       hamster
                    70.5
6
    7
            True
                                grape
                                         700
                                                7.7
                                                         pink False
                                                                         turtle
7
    8
       h
           False
                    80.5
                             honeydew
                                         800
                                                8.8
                                                         cyan
                                                                False
                                                                          snake
8
    9
       i
            True
                    90.5
                                 kiwi
                                         900
                                                9.9
                                                     magenta
                                                                 True
                                                                           frog
```

```
9 10 j False 100.5 lemon 1000 10.1 black True lizard
```

1.0.1 How many rows and columns are there in the DataFrame?

```
[2]: print("number of rows is ",df.shape[0])
print("number of columns is ",df.shape[1])
```

number of rows is 10 number of columns is 10

1.0.2 What are the column names of the DataFrame?

```
[3]: df.columns
```

```
[3]: Index(['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J'], dtype='object')
```

1.0.3 What are the data types of each column?

```
[13]: df.dtypes
```

```
[13]: A
              int64
      В
             object
      С
               bool
      D
            float64
      Ε
             object
      F
              int64
            float64
      G
      Η
             object
      Ι
               bool
      J
             object
      K
              int64
      dtype: object
```

1.0.4 How can you access the first 5 rows of the DataFrame?

```
[4]: df.head(5)
```

```
[4]:
                    C
                          D
                                              F
        Α
            В
                                        Ε
                                                   G
                                                            Η
                                                                    Ι
                                                                             J
     0
        1
            a
                True
                       10.5
                                           100
                                                 1.1
                                                                 True
                                   apple
                                                          red
                                                                           dog
     1
        2
            b
               False
                       20.5
                                  banana
                                           200
                                                 2.2
                                                        green
                                                                 True
                                                                           cat
     2
        3
                True
                       30.5
                                           300
                                                 3.3
                                                         blue
            С
                                  cherry
                                                                False
                                                                          fish
        4
     3
            d
              False
                       40.5
                                     date
                                           400
                                                 4.4
                                                       yellow
                                                                False
                                                                          bird
        5
                True
                       50.5
                              elderberry
                                           500
                                                 5.5
                                                       orange
                                                                 True
                                                                       rabbit
```

1.0.5 How can you access the 'C' column of the DataFrame?

```
[5]: df['C']
[5]: 0
            True
     1
          False
     2
            True
     3
          False
     4
            True
     5
          False
     6
            True
     7
          False
     8
            True
     9
          False
     Name: C, dtype: bool
    1.0.6 What is the average value of column 'F'?
[6]: df["F"].mean()
[6]: 550.0
    1.0.7 How can you filter the DataFrame to show only rows where column 'I' is True?
    df[df["I"]==True]
[7]:
            В
                     C
                            D
                                         Ε
                                                F
                                                       G
                                                                 Η
                                                                       Ι
                                                                                 J
         Α
     0
         1
                 True
                         10.5
                                              100
                                                     1.1
                                                               red
                                                                    True
                                     apple
                                                                               dog
         2
     1
                False
                         20.5
                                                     2.2
            b
                                    banana
                                              200
                                                            green
                                                                    True
                                                                               cat
     4
         5
                 True
                         50.5
                                                     5.5
                                                           orange
                                elderberry
                                              500
                                                                    True
                                                                            rabbit
     5
         6
             f
                False
                         60.5
                                       fig
                                              600
                                                     6.6
                                                           purple
                                                                    True
                                                                           hamster
     8
         9
             i
                 True
                         90.5
                                      kiwi
                                              900
                                                     9.9
                                                          magenta
                                                                    True
                                                                              frog
                False
                        100.5
                                             1000
                                                            black
        10
             j
                                     lemon
                                                   10.1
                                                                    True
                                                                            lizard
    1.0.8 How can you sort the DataFrame based on column 'D' in descending order?
[8]: df.sort_values("D",ascending=False)
[8]:
            В
                     С
                            D
                                         Ε
                                                F
                                                       G
                                                                         Ι
                                                                                   J
         Α
                                                                 Η
                        100.5
                                                    10.1
                                                                     True
     9
        10
                False
                                     lemon
                                             1000
                                                            black
                                                                             lizard
     8
         9
             i
                 True
                         90.5
                                      kiwi
                                              900
                                                     9.9
                                                          magenta
                                                                     True
                                                                               frog
     7
         8
            h
                False
                         80.5
                                  honeydew
                                              800
                                                     8.8
                                                                    False
                                                                              snake
                                                             cyan
     6
         7
                 True
                         70.5
                                     grape
                                              700
                                                    7.7
                                                             pink
                                                                    False
                                                                             turtle
             g
     5
         6
            f
                False
                         60.5
                                              600
                                                     6.6
                                                           purple
                                       fig
                                                                     True
                                                                            hamster
     4
         5
                 True
                         50.5
                                elderberry
                                              500
                                                     5.5
                                                           orange
                                                                     True
                                                                             rabbit
     3
         4
                False
                         40.5
                                      date
                                              400
                                                     4.4
                                                           yellow
                                                                               bird
                                                                    False
```

```
2
    3
       С
            True
                    30.5
                                cherry
                                          300
                                                 3.3
                                                           blue
                                                                 False
                                                                             fish
1
    2
                                                 2.2
           False
                    20.5
                                banana
                                          200
                                                                              cat
       b
                                                         green
                                                                   True
0
    1
       a
            True
                    10.5
                                 apple
                                          100
                                                 1.1
                                                            red
                                                                   True
                                                                              dog
```

1.0.9 How can you add a new column 'K' to the DataFrame with values [1, 2, 3, ..., 15]?

```
[9]: df["K"]=[1,2,3,4,5,6,7,8,9,10] df
```

```
[9]:
             В
                     С
                              D
                                           Ε
                                                  F
                                                         G
                                                                    Η
                                                                            Ι
                                                                                       J
                                                                                           K
          Α
                                                                        True
     0
          1
             a
                  True
                          10.5
                                       apple
                                                100
                                                       1.1
                                                                  red
                                                                                    dog
                                                                                           1
          2
     1
             b
                 False
                          20.5
                                      banana
                                                200
                                                       2.2
                                                               green
                                                                         True
                                                                                           2
                                                                                    cat
     2
          3
             С
                  True
                          30.5
                                      cherry
                                                300
                                                       3.3
                                                                blue
                                                                       False
                                                                                   fish
                                                                                           3
     3
          4
                                                400
                                                       4.4
                                                                                           4
             d
                 False
                          40.5
                                        date
                                                              yellow
                                                                       False
                                                                                   bird
     4
          5
                  True
                          50.5
                                                500
                                                       5.5
                                                                                           5
             е
                                 elderberry
                                                              orange
                                                                         True
                                                                                 rabbit
     5
          6
             f
                 False
                          60.5
                                         fig
                                                600
                                                       6.6
                                                              purple
                                                                         True
                                                                               hamster
     6
          7
                  True
                          70.5
                                                700
                                                       7.7
                                                                                           7
                                       grape
                                                                pink
                                                                       False
                                                                                 turtle
             g
     7
          8
             h
                 False
                          80.5
                                   honeydew
                                                800
                                                       8.8
                                                                cyan
                                                                       False
                                                                                  snake
                                                                                           8
                                                             magenta
     8
          9
                  True
                          90.5
                                        kiwi
                                                900
                                                       9.9
                                                                                           9
             i
                                                                         True
                                                                                   frog
                                               1000
     9
         10
             j
                 False
                         100.5
                                       lemon
                                                      10.1
                                                               black
                                                                         True
                                                                                 lizard
                                                                                          10
```

1.0.10 How can you drop the 'E' column from the DataFrame?

```
[10]: df.drop(columns="E")
```

```
G
                                                               Ι
[10]:
              В
                       С
                               D
                                      F
                                                       Η
                                                                          J
                                                                              K
           Α
      0
           1
               a
                   True
                           10.5
                                    100
                                           1.1
                                                     red
                                                            True
                                                                        dog
                                                                               1
                                                  green
           2
              b
                  False
                           20.5
                                    200
                                           2.2
                                                            True
                                                                               2
      1
                                                                        cat
      2
           3
              С
                   True
                           30.5
                                    300
                                          3.3
                                                    blue
                                                           False
                                                                      fish
                                                                               3
      3
           4
              d
                  False
                           40.5
                                                  yellow
                                                                      bird
                                                                              4
                                    400
                                          4.4
                                                          False
      4
                           50.5
           5
                   True
                                    500
                                          5.5
                                                  orange
                                                            True
                                                                    rabbit
                                                                              5
               е
      5
           6
               f
                  False
                           60.5
                                    600
                                          6.6
                                                 purple
                                                            True
                                                                   hamster
                                                                               6
      6
           7
                   True
                           70.5
                                    700
                                          7.7
                                                    pink
                                                         False
                                                                              7
                                                                    turtle
              g
      7
           8
                  False
                           80.5
                                          8.8
                                                          False
                                    800
                                                    cyan
                                                                     snake
                                                                              8
      8
           9
               i
                   True
                           90.5
                                    900
                                          9.9
                                                magenta
                                                            True
                                                                      frog
                                                                              9
          10
                  False
                          100.5
                                  1000
                                         10.1
                                                  black
              j
                                                            True
                                                                    lizard
                                                                             10
```

[]: