

```
In [13]: import numpy as np  
import pandas as pd
```

```
In [20]: dic1 = { "name":['harry','rohan','skillf','subh'],  
             "marks":[92,34,24,76],  
             "country":['India','Japan','North Korea', 'Iceland']}
```

```
In [21]: df = pd.DataFrame(dic1)
```

```
In [22]: df
```

```
Out[22]:   name  marks      country  
0    harry     92        India  
1    rohan     34        Japan  
2    skillf     24  North Korea  
3    subh      76      Iceland
```

```
In [29]: df.sort_values(["marks"], ascending = False).reset_index()
```

```
Out[29]:   index  name  marks      country  
0       0  harry     92        India  
1       3  subh      76      Iceland  
2       1  rohan     34        Japan  
3       2  skillf     24  North Korea
```

```
In [34]: df.to_csv("friends.csv", index = False, header = True, sep = "|")
```

```
In [35]: df.head(2)
```

```
Out[35]:   name  marks  country
```

| 0 | harry | 92 | India |
|---|-------|----|-------|
| 1 | rohan | 34 | Japan |

```
In [36]: df.tail(2)
```

```
Out[36]:   name  marks  country
```

| 2 | skillf | 24 | North Korea |
|---|--------|----|-------------|
| 3 | subh | 76 | Iceland |

```
In [39]: df.describe()
```

```
Out[39]:      marks
```

| count | 4.000000 |
|-------|-----------|
| mean | 56.500000 |
| std | 32.675169 |
| min | 24.000000 |
| 25% | 31.500000 |
| 50% | 55.000000 |
| 75% | 80.000000 |
| max | 92.000000 |

```
In [42]: trains = pd.read_csv("trains.csv")
```

```
In [43]: trains
```

Out[43]:

| | trains | speed | destination |
|---|--------|-------|-------------|
| 0 | 12332 | 213 | pari |
| 1 | 12332 | 213 | pari |
| 2 | 12332 | 213 | pari |
| 3 | 12332 | 213 | pari |
| 4 | 12332 | 213 | pari |
| 5 | 12332 | 213 | pari |
| 6 | 12332 | 213 | pari |
| 7 | 12332 | 213 | pari |
| 8 | 12332 | 213 | pari |
| 9 | 12332 | 213 | pari |

In [45]:

```
trains["speed"][0] = 50
```

C:\Users\mause\AppData\Local\Temp\ipykernel_23128\2845172488.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
trains["speed"][0] = 50

In [46]:

```
trains
```

Out[46]:

| | trains | speed | destination |
|---|--------|-------|-------------|
| 0 | 12332 | 50 | pari |
| 1 | 12332 | 213 | pari |
| 2 | 12332 | 213 | pari |
| 3 | 12332 | 213 | pari |
| 4 | 12332 | 213 | pari |
| 5 | 12332 | 213 | pari |
| 6 | 12332 | 213 | pari |
| 7 | 12332 | 213 | pari |
| 8 | 12332 | 213 | pari |
| 9 | 12332 | 213 | pari |

In [49]:

```
trains.to_csv("trains.csv")
```

In [61]:

```
trains.index = ["first", "second", "third", "fourth", "fifth", "six", "seven", "eight", "nine", "ten"]
```

In [62]:

```
trains
```

```
Out[62]:
```

| | trains | speed | destination |
|---------------|--------|-------|-------------|
| first | 12332 | 50 | pari |
| second | 12332 | 213 | pari |
| third | 12332 | 213 | pari |
| fourth | 12332 | 213 | pari |
| fifth | 12332 | 213 | pari |
| six | 12332 | 213 | pari |
| seven | 12332 | 213 | pari |
| eight | 12332 | 213 | pari |
| nine | 12332 | 213 | pari |
| ten | 12332 | 213 | pari |

why do we use pandas and not excel ? in pandas we can use power and functions of python. It has speed and robust data operations

Pandas has two types of datastructures series : one dimensional array with indexes . it stores single column or row data in a DataFrame. sereies can hold any one type of data Dataframe : it is a tabular spreadsheet like structure representing rows each of which contains one or multiple columns. data frame can hold multiple kind of data.

```
In [67]: type(trains["trains"])
```

```
Out[67]: pandas.core.series.Series
```

```
In [68]: type(trains)
```

```
Out[68]: pandas.core.frame.DataFrame
```

```
In [70]: ser = pd.Series(np.random.rand(15))
```

```
In [71]: ser
```

```
Out[71]: 0    0.953918
         1    0.913958
         2    0.918622
         3    0.019609
         4    0.078709
         5    0.877305
         6    0.147470
         7    0.080428
         8    0.239136
         9    0.370746
        10   0.849885
        11   0.555321
        12   0.716564
        13   0.345084
        14   0.925324
        dtype: float64
```

```
In [72]: type(ser)
```

```
Out[72]: pandas.core.series.Series
```

```
In [77]: newdf = pd.DataFrame(np.random.rand(334,5),index = np.arange(334))
```

```
In [78]: newdf
```

Out[78]:

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | 0.469295 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.932450 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.430010 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [79]: `type(newdf)`

Out[79]: `pandas.core.frame.DataFrame`

In [80]: `newdf.describe()`

```
Out[80]:
```

| | 0 | 1 | 2 | 3 | 4 |
|--------------|------------|------------|------------|------------|------------|
| count | 334.000000 | 334.000000 | 334.000000 | 334.000000 | 334.000000 |
| mean | 0.492598 | 0.490965 | 0.486031 | 0.480435 | 0.498423 |
| std | 0.290492 | 0.296993 | 0.274524 | 0.287157 | 0.291631 |
| min | 0.000349 | 0.000134 | 0.002409 | 0.004485 | 0.005581 |
| 25% | 0.232086 | 0.239583 | 0.264293 | 0.233407 | 0.255451 |
| 50% | 0.487265 | 0.504747 | 0.496114 | 0.476280 | 0.499167 |
| 75% | 0.745001 | 0.746251 | 0.698652 | 0.718703 | 0.752092 |
| max | 0.998609 | 0.998872 | 0.991687 | 0.994728 | 0.998795 |

```
In [83]: newdf.dtypes
```

```
Out[83]:
```

| | |
|---|---------------|
| 0 | float64 |
| 1 | float64 |
| 2 | float64 |
| 3 | float64 |
| 4 | float64 |
| | dtype: object |

```
In [86]: newdf[0][0] = "harry"
```

```
In [87]: newdf
```

```
Out[87]:
```

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | harry | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

```
In [89]: newdf[0].dtypes
```

```
Out[89]: dtype('O')
```

```
In [90]: newdf[1].dtypes
```

```
Out[90]: dtype('float64')
```

```
In [91]: newdf.index
```

```
Out[91]: Int64Index([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9,
...,
324, 325, 326, 327, 328, 329, 330, 331, 332, 333],
dtype='int64', length=334)
```

```
In [92]: newdf.columns
```

```
Out[92]: RangeIndex(start=0, stop=5, step=1)
```

```
In [93]: newdf.to_numpy()
```

```
Out[93]: array([[ 'harry',  0.7446473007913158,  0.7274665152825885,
   0.0425511657079809,  0.8966007612136686],
   [0.13726944215629422,  0.5851144812068143,  0.9873237889166186,
   0.19181506964647044,  0.5792949193641653],
   [0.42929169186464267,  0.7867051304864211,  0.6132726384161775,
   0.05437587830756718,  0.8152947823449564],
   ...,
   [0.922774061875306,  0.3923388427777281,  0.05071431674186755,
   0.769555642521064,  0.10098805311179637],
   [0.43000976608359254,  0.28123570062464587,  0.516260876183876,
   0.6569162866562084,  0.7767486545424611],
   [0.29273675337133664,  0.00013443288930414798,  0.9613026424442971,
   0.4895632362136326,  0.9526330745510918]], dtype=object)
```

```
In [94]: newdf[0][0] = 0.72349
```

```
In [95]: newdf.head()
```

```
Out[95]:      0      1      2      3      4
 0  0.72349  0.744647  0.727467  0.042551  0.896601
 1  0.137269  0.585114  0.987324  0.191815  0.579295
 2  0.429292  0.786705  0.613273  0.054376  0.815295
 3  0.355345  0.383176  0.916397  0.850284  0.532916
 4  0.873201  0.937285  0.674893  0.828821  0.941144
```

```
In [96]: type(newdf[0])
```

```
Out[96]: pandas.core.series.Series
```

```
In [97]: newdf.T
```

Out[97]:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... | 324 | 325 | 326 |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|----------|----------|----------|
| 0 | 0.72349 | 0.137269 | 0.429292 | 0.355345 | 0.873201 | 0.658502 | 0.509922 | 0.839917 | 0.054864 | 0.194357 | ... | 0.843786 | 0.697065 | 0.438444 |
| 1 | 0.744647 | 0.585114 | 0.786705 | 0.383176 | 0.937285 | 0.58226 | 0.396915 | 0.259683 | 0.570929 | 0.067636 | ... | 0.254827 | 0.393975 | 0.845575 |
| 2 | 0.727467 | 0.987324 | 0.613273 | 0.916397 | 0.674893 | 0.431429 | 0.608272 | 0.191953 | 0.827565 | 0.779275 | ... | 0.289997 | 0.201824 | 0.36842 |
| 3 | 0.042551 | 0.191815 | 0.054376 | 0.850284 | 0.828821 | 0.892262 | 0.691234 | 0.945699 | 0.290621 | 0.295874 | ... | 0.924219 | 0.837462 | 0.802211 |
| 4 | 0.896601 | 0.579295 | 0.815295 | 0.532916 | 0.941144 | 0.357932 | 0.505227 | 0.077954 | 0.257761 | 0.717791 | ... | 0.612772 | 0.197224 | 0.346154 |

5 rows × 334 columns

In [99]: `ndf = newdf.head()`In [101...]: `ndf.sort_values(by = 0, ascending = True)`

Out[101]:

| | 0 | 1 | 2 | 3 | 4 |
|---|----------|----------|----------|----------|----------|
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 0 | 0.72349 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |

In [104...]: `ndf.sort_index(axis = 1, ascending = False) # column`

Out[104]:

| | 4 | 3 | 2 | 1 | 0 |
|---|----------|----------|----------|----------|----------|
| 0 | 0.896601 | 0.042551 | 0.727467 | 0.744647 | 0.72349 |
| 1 | 0.579295 | 0.191815 | 0.987324 | 0.585114 | 0.137269 |
| 2 | 0.815295 | 0.054376 | 0.613273 | 0.786705 | 0.429292 |
| 3 | 0.532916 | 0.850284 | 0.916397 | 0.383176 | 0.355345 |
| 4 | 0.941144 | 0.828821 | 0.674893 | 0.937285 | 0.873201 |

```
In [105...]: ndf.sort_index(axis = 0, ascending = False) #rows
```

```
Out[105]:      0      1      2      3      4
               4  0.873201  0.937285  0.674893  0.828821  0.941144
               3  0.355345  0.383176  0.916397  0.850284  0.532916
               2  0.429292  0.786705  0.613273  0.054376  0.815295
               1  0.137269  0.585114  0.987324  0.191815  0.579295
               0  0.72349   0.744647  0.727467  0.042551  0.896601
```

```
In [109...]: newdf[0].dtypes
```

```
Out[109]: dtype('O')
```

```
In [110...]: ndf
```

```
Out[110]:      0      1      2      3      4
               0  0.72349  0.744647  0.727467  0.042551  0.896601
               1  0.137269  0.585114  0.987324  0.191815  0.579295
               2  0.429292  0.786705  0.613273  0.054376  0.815295
               3  0.355345  0.383176  0.916397  0.850284  0.532916
               4  0.873201  0.937285  0.674893  0.828821  0.941144
```

difference between a copy and a view

```
In [111...]: xyz = ndf
```

```
In [112...]: xyz
```

Out[112]:

| | 0 | 1 | 2 | 3 | 4 |
|---|----------|----------|----------|----------|----------|
| 0 | 0.72349 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |

In [113...]: xyz[0][0] = 72

C:\Users\mause\AppData\Local\Temp\ipykernel_23128\2535501250.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
xyz[0][0] = 72

In [114...]: xyz

Out[114]:

| | 0 | 1 | 2 | 3 | 4 |
|---|----------|----------|----------|----------|----------|
| 0 | 72 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |

In [115...]: newdf

Out[115]:

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | 72 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

xyz is not a copy of newdf here, it is a view or a pointer to the actual dataframe

Copy

In [123...]

xyz = newdf.copy()

In [124...]

xyz

Out[124]:

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | 3423 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [129...]

newdf[0][0] = 34235

```
C:\Users\mause\AppData\Local\Temp\ipykernel_23128\3832573312.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returnin  
g-a-view-versus-a-copy  
newdf[0][0] = 34235
```

In [130...]

newdf

Out[130]:

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | 34235 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [131...]

xyz

Out[131]:

| | 0 | 1 | 2 | 3 | 4 |
|-----|----------|----------|----------|----------|----------|
| 0 | 3423 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

we don't use newdf[0][0] in pandas to change the element as we can't be sure what we will get in return. So we use loc function

In [132...]

newdf.loc[0,0] = 654

In [133...]

newdf.head(2)

Out[133]:

| | 0 | 1 | 2 | 3 | 4 |
|---|----------|----------|----------|----------|----------|
| 0 | 654 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |

In [135...]

newdf.columns = list("ABCDE")

In [136...]

newdf

Out[136]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 0 | 654 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [137...]

newdf.loc[0,0] = 654

In [139...]

newdf.head(2)

Out[139]:

| | A | B | C | D | E | 0 |
|---|----------|----------|----------|----------|----------|-------|
| 0 | 654 | 0.744647 | 0.727467 | 0.042551 | 0.896601 | 654.0 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 | NaN |

In [144...]

newdf.loc[0,"A"] = 765

In [145...]

newdf

Out[145]:

| | A | B | C | D | E | 0 |
|-----|----------|----------|----------|----------|----------|-------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 | 654.0 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 | NaN |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 | NaN |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 | NaN |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 | NaN |
| ... | ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 | NaN |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 | NaN |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 | NaN |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 | NaN |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 | NaN |

334 rows × 6 columns

In [147...]

newdf.loc[0,0] = 7988

In [148...]

newdf

Out[148]:

| | A | B | C | D | E | 0 |
|-----|----------|----------|----------|----------|----------|--------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 | 7988.0 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 | NaN |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 | NaN |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 | NaN |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 | NaN |
| ... | ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 | NaN |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 | NaN |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 | NaN |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 | NaN |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 | NaN |

334 rows × 6 columns

In [150...]

newdf.drop(0, axis = 1)

Out[150]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [151...]

newdf.head()

Out[151]:

| | A | B | C | D | E | 0 |
|---|----------|----------|----------|----------|----------|--------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 | 7988.0 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 | NaN |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 | NaN |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 | NaN |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 | NaN |

In [152...]

newdf = newdf.drop(0, axis = 1)

In [153...]

newdf

Out[153]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [159...]

newdf.loc[[1,2],['C',"D"]] #i wrote df = newdf.loc

Out[159]:

| | C | D |
|---|----------|----------|
| 1 | 0.987324 | 0.191815 |
| 2 | 0.613273 | 0.054376 |

In [161...]

df

Out[161]:

| | C | D |
|---|----------|----------|
| 1 | 0.987324 | 0.191815 |
| 2 | 0.613273 | 0.054376 |

In [162...]

newdf.loc[:,['C',"D"]]

Out[162]:

| | C | D |
|-----|----------|----------|
| 0 | 0.727467 | 0.042551 |
| 1 | 0.987324 | 0.191815 |
| 2 | 0.613273 | 0.054376 |
| 3 | 0.916397 | 0.850284 |
| 4 | 0.674893 | 0.828821 |
| ... | ... | ... |
| 329 | 0.578047 | 0.589086 |
| 330 | 0.243989 | 0.842591 |
| 331 | 0.050714 | 0.769556 |
| 332 | 0.516261 | 0.656916 |
| 333 | 0.961303 | 0.489563 |

334 rows × 2 columns

In [163...]

newdf.loc[[1,2],:]

Out[163]:

| | A | B | C | D | E |
|---|----------|----------|----------|----------|----------|
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |

In [168...]

newdf.loc[(newdf["A"] < 0.3) & (newdf["C"] > 0.1)]

Out[168]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 8 | 0.054864 | 0.570929 | 0.827565 | 0.290621 | 0.257761 |
| 9 | 0.194357 | 0.067636 | 0.779275 | 0.295874 | 0.717791 |
| 11 | 0.113157 | 0.993368 | 0.692144 | 0.214571 | 0.359740 |
| 13 | 0.134097 | 0.779449 | 0.673082 | 0.042873 | 0.964261 |
| ... | ... | ... | ... | ... | ... |
| 308 | 0.052626 | 0.247481 | 0.185500 | 0.916508 | 0.858943 |
| 315 | 0.168138 | 0.830240 | 0.742892 | 0.436291 | 0.874826 |
| 318 | 0.164676 | 0.210732 | 0.957561 | 0.328809 | 0.405957 |
| 323 | 0.218975 | 0.793008 | 0.335533 | 0.432164 | 0.653766 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

93 rows × 5 columns

In [172...]

newdf.head()

Out[172]:

| | A | B | C | D | E |
|---|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |

In [174...]

newdf.iloc[0,0]

Out[174]:

765

In [171...]

newdf.loc[0,"A"]

Out[171]: 765

In [175... newdf.head(3)

Out[175]:

| | A | B | C | D | E |
|---|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |

In [176... newdf.drop(0)

Out[176]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| 5 | 0.658502 | 0.582260 | 0.431429 | 0.892262 | 0.357932 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

333 rows × 5 columns

In [178... newdf.drop(["A","D"], axis = 1)

Out[178]:

| | B | C | E |
|-----|----------|----------|----------|
| 0 | 0.744647 | 0.727467 | 0.896601 |
| 1 | 0.585114 | 0.987324 | 0.579295 |
| 2 | 0.786705 | 0.613273 | 0.815295 |
| 3 | 0.383176 | 0.916397 | 0.532916 |
| 4 | 0.937285 | 0.674893 | 0.941144 |
| ... | ... | ... | ... |
| 329 | 0.504391 | 0.578047 | 0.081438 |
| 330 | 0.679940 | 0.243989 | 0.111000 |
| 331 | 0.392339 | 0.050714 | 0.100988 |
| 332 | 0.281236 | 0.516261 | 0.776749 |
| 333 | 0.000134 | 0.961303 | 0.952633 |

334 rows × 3 columns

this will not change the dataframe. To change the dataframe below thing should be done

In [183...]: df = newdf.drop(["A", "C"], axis = 1) # we need to assign the dataframe

In [181...]: df

Out[181]:

| | B | D | E |
|------------|----------|----------|----------|
| 0 | 0.744647 | 0.042551 | 0.896601 |
| 1 | 0.585114 | 0.191815 | 0.579295 |
| 2 | 0.786705 | 0.054376 | 0.815295 |
| 3 | 0.383176 | 0.850284 | 0.532916 |
| 4 | 0.937285 | 0.828821 | 0.941144 |
| ... | ... | ... | ... |
| 329 | 0.504391 | 0.589086 | 0.081438 |
| 330 | 0.679940 | 0.842591 | 0.111000 |
| 331 | 0.392339 | 0.769556 | 0.100988 |
| 332 | 0.281236 | 0.656916 | 0.776749 |
| 333 | 0.000134 | 0.489563 | 0.952633 |

334 rows × 3 columns

In [182...]

newdf

Out[182]:

| | A | B | C | D | E |
|-----|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 0.383176 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 0.937285 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 0.504391 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 0.679940 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 0.392339 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 0.281236 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 0.000134 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

or, we do inplace

In [185...]: df.drop(["B"], axis = 1, inplace = True)

In [186...]: df

Out[186]:

| | D | E |
|------------|----------|----------|
| 0 | 0.042551 | 0.896601 |
| 1 | 0.191815 | 0.579295 |
| 2 | 0.054376 | 0.815295 |
| 3 | 0.850284 | 0.532916 |
| 4 | 0.828821 | 0.941144 |
| ... | ... | ... |
| 329 | 0.589086 | 0.081438 |
| 330 | 0.842591 | 0.111000 |
| 331 | 0.769556 | 0.100988 |
| 332 | 0.656916 | 0.776749 |
| 333 | 0.489563 | 0.952633 |

334 rows × 2 columns

In [187...]

newdf.head(3)

Out[187]:

| | A | B | C | D | E |
|----------|----------|----------|----------|----------|----------|
| 0 | 765 | 0.744647 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 0.585114 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 0.786705 | 0.613273 | 0.054376 | 0.815295 |

In [190...]

newdf.reset_index(drop = True, inplace = True)

In [191...]

df

Out[191]:

| | D | E |
|------------|----------|----------|
| 0 | 0.042551 | 0.896601 |
| 1 | 0.191815 | 0.579295 |
| 2 | 0.054376 | 0.815295 |
| 3 | 0.850284 | 0.532916 |
| 4 | 0.828821 | 0.941144 |
| ... | ... | ... |
| 329 | 0.589086 | 0.081438 |
| 330 | 0.842591 | 0.111000 |
| 331 | 0.769556 | 0.100988 |
| 332 | 0.656916 | 0.776749 |
| 333 | 0.489563 | 0.952633 |

334 rows × 2 columns

insa

In [194...]

newdf["B"] = None

In [195...]

newdf

Out[195]:

| | A | B | C | D | E |
|-----|----------|------|----------|----------|----------|
| 0 | 765 | None | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | None | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | None | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | None | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | None | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | None | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | None | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | None | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | None | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | None | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [196...]: newdf["B"].isnull()

Out[196]:

| | |
|-----|------|
| 0 | True |
| 1 | True |
| 2 | True |
| 3 | True |
| 4 | True |
| ... | |
| 329 | True |
| 330 | True |
| 331 | True |
| 332 | True |
| 333 | True |

Name: B, Length: 334, dtype: bool

In [199...]: newdf.loc[:, ["B"]] = 56

In [200...]: newdf

Out[200]:

| | A | B | C | D | E |
|-----|----------|-----|----------|----------|----------|
| 0 | 765 | 56 | 0.727467 | 0.042551 | 0.896601 |
| 1 | 0.137269 | 56 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 0.429292 | 56 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 0.355345 | 56 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 0.873201 | 56 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 0.373049 | 56 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 0.93245 | 56 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 0.922774 | 56 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 0.43001 | 56 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 0.292737 | 56 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [204...]

newdf.iloc[0,:] = 56

In [205...]

newdf

Out[205]:

| | A | B | C | D | E |
|------------|----------|----------|-----------|-----------|-----------|
| 0 | 56 | 56 | 56.000000 | 56.000000 | 56.000000 |
| 1 | 56 | 56 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 56 | 56 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 56 | 56 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 56 | 56 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 56 | 56 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 56 | 56 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 56 | 56 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 56 | 56 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 56 | 56 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [215...]

newdf

Out[215]:

| | A | B | C | D | E |
|-----|-----|-----|-----------|-----------|-----------|
| 0 | 56 | 56 | 56.000000 | 56.000000 | 56.000000 |
| 1 | 56 | 56 | 0.987324 | 0.191815 | 0.579295 |
| 2 | 56 | 56 | 0.613273 | 0.054376 | 0.815295 |
| 3 | 56 | 56 | 0.916397 | 0.850284 | 0.532916 |
| 4 | 56 | 56 | 0.674893 | 0.828821 | 0.941144 |
| ... | ... | ... | ... | ... | ... |
| 329 | 56 | 56 | 0.578047 | 0.589086 | 0.081438 |
| 330 | 56 | 56 | 0.243989 | 0.842591 | 0.111000 |
| 331 | 56 | 56 | 0.050714 | 0.769556 | 0.100988 |
| 332 | 56 | 56 | 0.516261 | 0.656916 | 0.776749 |
| 333 | 56 | 56 | 0.961303 | 0.489563 | 0.952633 |

334 rows × 5 columns

In [213...]

```
newdf.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 334 entries, 0 to 333
Data columns (total 5 columns):
 # Column Non-Null Count Dtype

 0 A 334 non-null int64
 1 B 334 non-null int64
 2 C 334 non-null float64
 3 D 334 non-null float64
 4 E 334 non-null float64
dtypes: float64(3), int64(2)
memory usage: 13.2 KB

In [216...]

```
newdf["A"].value_counts(dropna = False)
```

Out[216]:

56 334
Name: A, dtype: int64

In [227...]

```
thisdict = {  
    "brand": ["Ford", "Fre", "jk"],  
    "model": ["Mustang", np.nan, "dsjih"],  
    "year": [1964, pd.NaT, 43432]  
}
```

In [228...]

```
dic = pd.DataFrame(thisdict)
```

In [229...]

```
dic
```

Out[229]:

| | brand | model | year |
|---|-------|---------|-------|
| 0 | Ford | Mustang | 1964 |
| 1 | Fre | NaN | NaT |
| 2 | jk | dsjih | 43432 |

In [230...]

```
dic.notnull()
```

Out[230]:

| | brand | model | year |
|---|-------|-------|-------|
| 0 | True | True | True |
| 1 | True | False | False |
| 2 | True | True | True |

In [231...]

```
dic.isnull()
```

Out[231]:

| | brand | model | year |
|---|-------|-------|-------|
| 0 | False | False | False |
| 1 | False | True | True |
| 2 | False | False | False |

In [232...]

```
#create a dataframe with only integers with 3 rows and 2 columns
```

```
In [238...]: v = pd.DataFrame(np.random.rand(3,2),index = np.arange(3))
```

```
In [239...]: v
```

```
Out[239]:      0      1
0  0.677307  0.695747
1  0.770411  0.117542
2  0.710770  0.105841
```

```
In [240...]: v.describe()
```

```
Out[240]:      0      1
count  3.000000  3.000000
mean   0.719496  0.306377
std    0.047161  0.337255
min   0.677307  0.105841
25%   0.694038  0.111692
50%   0.710770  0.117542
75%   0.740590  0.406645
max   0.770411  0.695747
```

```
In [241...]: v.mean()
```

```
Out[241]: 0    0.719496
1    0.306377
dtype: float64
```

```
In [242...]: v.corr()
```

```
Out[242]:      0      1  
                0  1.000000 -0.763628  
                1  -0.763628  1.000000
```

```
In [243...]: v.count()
```

```
Out[243]: 0      3  
           1      3  
           dtype: int64
```

```
In [244...]: v.max()
```

```
Out[244]: 0      0.770411  
           1      0.695747  
           dtype: float64
```

```
In [245...]: v.min()
```

```
Out[245]: 0      0.677307  
           1      0.105841  
           dtype: float64
```

```
In [246...]: v.median()
```

```
Out[246]: 0      0.710770  
           1      0.117542  
           dtype: float64
```

```
In [247...]: v.std()
```

```
Out[247]: 0      0.047161  
           1      0.337255  
           dtype: float64
```

```
In [249...]: v.max()
```

```
Out[249]: 0      0.770411  
           1      0.695747  
           dtype: float64
```

```
In [250...]: v.min()
```

```
Out[250]: 0    0.677307  
           1    0.105841  
          dtype: float64
```

In []: