Pandas Guide

Add or remove rows & columns

ihiteshmodi

Add or remove rows & columns

```
In [1]: import pandas as pd
```

```
In [51]: df = pd.read_csv("Heart_disease_details.csv")
```

This is how our dataframe looks like, (Heart Disease Dataset from kaggle)

```
In [52]: df.head(3)
```

	Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	...	Cardiac CT	Obesity	Murmur	Chest x-ray
0	Jane Doe	Female	55	1	1	1	140	90	100	1	...	Shows a 50% blockage in the left anterior desc...	0	1	None
1	Mark Johnson	Male	57	1	1	1	150	80	110	1	...	Shows a 60% blockage in the right coronary artery	0	1	None
2	Emily Davis	Female	60	1	1	1	130	85	95	1	...	Shows a 75% blockage in the left anterior desc...	0	1	None

3 rows x 49 columns

ihiteshmodi

Add or remove rows & columns

Lets start with learning how to drop columns as its straight forward ¶

All we have to do is use the drop method using a list

In [53]:
```
df2 = df.copy() # creating a copy dataframe from which we will drop the columns
```

In [55]:
```
df2.drop(columns= ["Name", "Gender", "Age"], inplace = True)
```

In [56]:
```
df2.head(2)
```

	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	Cholesterol level (mg/dL)	LDL level (mg/dL)	HDL level (mg/dL)	...	Cardiac CT	Obesity	Murmur
0	1	1	1	140	90	100	1	220	150	40	...	Shows a 50% blockage in the left anterior desc...	0	1
1	1	1	1	150	80	110	1	210	130	50	...	Shows a 60% blockage in the right coronary artery	0	1

2 rows × 46 columns

And the three columns are dropped

ihiteshmodi

Add or remove rows & columns

When we want to drop rows, We do that by using the index number of the respective row

In [57]:
```
df3 = df.copy(2)
```

Say we want to drop index no 1 "Mark Johnson" we do this and the row will be dropped

In [58]:
```
df3.drop(index = 1, inplace = True)
```

In [59]:
```
df3.head(2)
```

	Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	...	Cardiac CT	Obesity	Murmur	Chest x-ray
0	Jane Doe	Female	55	1	1	1	140	90	100	1	...	Shows a 50% blockage in the left anterior desc...	0	1	None
2	Emily Davis	Female	60	1	1	1	130	85	95	1	...	Shows a 75% blockage in the left anterior desc...	0	1	None

2 rows x 49 columns

We can see that index no 1 is gone and after no 0, we directly have 2 now

ihiteshmodi

Add or remove rows & columns

Can we also drop multiple indexes? Ofcourse, Using a list of indexes

In [60]:
```
df3.drop(index = [0,2,3], inplace = True)
```

In [61]:
```
df3.head(3)
```

	Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	...	Cardiac CT	Obesity	Murmur	Chest x-ray
4	Ashley Johnson	Female	58	1	1	1	135	80	105	1	...	Shows a 90% blockage in the right coronary artery	0	1	None
5	Brian Brown	Male	55	1	1	1	150	95	110	1	...	Shows a 70% blockage in the left anterior desc...	0	1	None
6	Emily Davis	Female	60	1	1	1	145	90	110	1	...	Shows a 50% blockage in the left anterior desc...	0	1	None

3 rows x 49 columns

Now, our dataset starts from index no 4 indicating rest were dropped :)

ihiteshmodi

But what if we want to drop multiple rows using mask? Why not. Lets see how

Lets run through a basic concept first

In [63]:
```
mask = df3["Age"] >= 50 #so we are filtering for only the rows where age is greter than or equal to 50 years
```

In [64]:
```
df3[mask].shape #after applying the mask, we have 165 rows
```

```
(165, 49)
```

In [ ]:

In [ ]:

In [65]:
```
df3[mask].index
```

```
Int64Index([  4,   5,   6,   7,   8,   9,  10,  11,  12,  13,
            ...
            322, 323, 324, 325, 326, 327, 328, 329, 331, 332],
           dtype='int64', length=165)
```

The moment we call .index method, we get a list of indexes from the dataset where Age is greater than or equal to 55 Years, We can now use these indexes to drop all these rows from the dataset

Add or remove rows & columns

```
In [67]: df3.drop(index = df3[mask].index, inplace = True)
```

```
In [68]: df3.head(3)
```

	Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	...	Cardiac CT	Obesity	Murmur	Ch x-r
16	Jacob Smith	Male	45	1	1	1	140	80	90	1	...	Shows a severe regurgitation in the mitral valve	0	1	No
17	Emily Davis	Female	32	0	1	1	120	80	90	1	...	Shows a mild regurgitation in the mitral valve	0	1	No
18	Madison Johnson	Female	42	0	1	1	130	80	90	1	...	Shows a moderate regurgitation in the mitral v...	0	1	No

3 rows x 49 columns

Now, our dataset is starting from 16 indicating all the values where Age is less than 50 years have been dropped

ihiteshmodi

Add or remove rows & columns

What if our index is not basic numbers, instead we have set some string value as our index?

```
In [36]: df4 = df.copy()
```

```
In [37]: df4.set_index("Name", inplace = True)
```

As we can see now, "Name" column is our index now

```
In [42]: df4.index
```

```
Index(['Ashley Johnson', 'Brian Brown', 'Emily Davis', 'John Smith',
       'Jane Doe', 'Mark Johnson', 'Jane Smith', 'John Doe', 'Michael Brown',
       'Jessica Davis',
       ...
       'John Smith', 'Jane Doe', 'Emily Johnson', 'John Smith', 'Jane Doe',
       'Emily Wilson', 'Jacob Smith', 'Jane Smith', 'David Johnson',
       'Emily Smith'],
      dtype='object', name='Name', length=330)
```

```
In [39]: df4.head(3###  As we can see now, "Name" column is our index now)
```

Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	Cholesterol level (mg/dL)	...	Cardiac CT	Obesity	Murn
Ashley Johnson	Female	58	1	1	1	135	80	105	1	220	...	Shows a 90% blockage in the right coronary artery	0	1
Brian Brown	Male	55	1	1	1	150	95	110	1	200	...	Shows a 70% blockage in the left anterior desc...	0	1
Emily Davis	Female	60	1	1	1	145	90	110	1	220	...	Shows a 50% blockage in the left anterior desc...	0	1

3 rows x 48 columns

Add or remove rows & columns

Lets say we want to drop first row, "Ashley Johnson" from the dataset, we do this

```
In [ ]:
```

```
In [43]: df4.drop(index = "Ashley Johnson", inplace = True)
```

```
In [43]: df4.drop(index = "Ashley Johnson", inplace = True)
```

```
In [44]: df4.head(3)
```

Name	Gender	Age	Chest pain	Shortness of breath	Fatigue	Systolic	Diastolic	Heart rate (bpm)	Lung sounds	Cholesterol level (mg/dL)	...	Cardiac CT	Obesity	Murmu
Brian Brown	Male	55	1	1	1	150	95	110	1	200	...	Shows a 70% blockage in the left anterior desc...	0	1
Emily Davis	Female	60	1	1	1	145	90	110	1	220	...	Shows a 50% blockage in the left anterior desc...	0	1
John Smith	Male	70	1	1	1	140	90	100	1	240	...	Shows a 70% blockage in the right coronary artery	0	0

3 rows x 48 columns

As we can see, "Ashley Johnson" has been dropped meaning even if we have set some column as our index, we can stil use the values to drop those rows from our dataset

ihiteshmodi