Day_34_291123

January 23, 2024

[110]:
```
import pandas as pd
```

[111]:
```
import numpy as np
```

1 There are two types of indexs

- Explicit index - User can see
- Implicit index - Users cannot see but the computer will assign indices

[112]:
```
temp = pd.DataFrame([['a','b',1,3.0,4]],columns = ['a','b','c','d','e'])
temp
```

[112]:

	a	b	c	d	e
0	a	b	1	3.0	4

[113]:
```
df = pd.read_csv("mckinsey (1).csv")
df.head(4)
```

[113]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138

2 Index values of a dataframe

[114]:
```
df.index.values
```

[114]:
```
array([   0,    1,    2, …, 1701, 1702, 1703], dtype=int64)
```

3 Changing the index values of a dataframe

[115]:
```
df.index = np.arange(1,1705,dtype='int')
```

[116]:
```
df.index.values
```

[116]:
```
array([   1,    2,    3, …, 1702, 1703, 1704])
```

[117]:
```
df
```

[117]:
```
          country  year  population continent  life_exp     gdp_cap
1     Afghanistan  1952     8425333      Asia    28.801  779.445314
2     Afghanistan  1957     9240934      Asia    30.332  820.853030
3     Afghanistan  1962    10267083      Asia    31.997  853.100710
4     Afghanistan  1967    11537966      Asia    34.020  836.197138
5     Afghanistan  1972    13079460      Asia    36.088  739.981106
…             …     …           …         …         …           …
1700     Zimbabwe  1987     9216418    Africa    62.351  706.157306
1701     Zimbabwe  1992    10704340    Africa    60.377  693.420786
1702     Zimbabwe  1997    11404948    Africa    46.809  792.449960
1703     Zimbabwe  2002    11926563    Africa    39.989  672.038623
1704     Zimbabwe  2007    12311143    Africa    43.487  469.709298

[1704 rows x 6 columns]
```

[118]:
```
df.index[1]
```

[118]:
```
2
```

4 loc (location -> Explicit) and iloc (integer location -> Implicit)

[119]:
```
df.iloc[1]
```

[119]:
```
country       Afghanistan
year                 1957
population        9240934
continent            Asia
life_exp           30.332
gdp_cap         820.85303
Name: 2, dtype: object
```

[120]:
```
df.loc[5]
```

[120]:
```
country       Afghanistan
year                 1972
population       13079460
continent            Asia
life_exp           36.088
gdp_cap        739.981106
Name: 5, dtype: object
```

[121]:
```
df.loc[889]
```

[121]:
```
country          Liberia
year                1952
population         863308
continent          Africa
life_exp            38.48
gdp_cap        575.572996
Name: 889, dtype: object
```

[122]:
```
df.iloc[888]
```

[122]:
```
country          Liberia
year                1952
population         863308
continent          Africa
life_exp            38.48
gdp_cap        575.572996
Name: 889, dtype: object
```

[123]:
```
df.iloc[[2,3,1,7]]
```

[123]:

	country	year	population	continent	life_exp	gdp_cap
3	Afghanistan	1962	10267083	Asia	31.997	853.100710
4	Afghanistan	1967	11537966	Asia	34.020	836.197138
2	Afghanistan	1957	9240934	Asia	30.332	820.853030
8	Afghanistan	1987	13867957	Asia	40.822	852.395945

[124]:
```
df.loc[[10,18,1056]]
```

[124]:

	country	year	population	continent	life_exp	gdp_cap
10	Afghanistan	1997	22227415	Asia	41.763	635.341351
18	Albania	1977	2509048	Europe	68.930	3533.003910
1056	Myanmar	2007	47761980	Asia	62.069	944.000000

[125]:
```
df.iloc[-1]
```

[125]:
```
country         Zimbabwe
year                2007
population      12311143
continent         Africa
life_exp           43.487
gdp_cap       469.709298
Name: 1704, dtype: object
```

[126]:
```
df.iloc[df.iloc[0:10:2].index]
```

[126]:

	country	year	population	continent	life_exp	gdp_cap
2	Afghanistan	1957	9240934	Asia	30.332	820.853030
4	Afghanistan	1967	11537966	Asia	34.020	836.197138

```
6   Afghanistan  1977    14880372      Asia    38.438  786.113360
8   Afghanistan  1987    13867957      Asia    40.822  852.395945
10  Afghanistan  1997    22227415      Asia    41.763  635.341351
```

5 Change the index of a dataframe

[127]:
```
temp = df.set_index('country')
```

[128]:
```
temp
```

[128]:
```
             year  population continent  life_exp     gdp_cap
country
Afghanistan  1952     8425333      Asia    28.801  779.445314
Afghanistan  1957     9240934      Asia    30.332  820.853030
Afghanistan  1962    10267083      Asia    31.997  853.100710
Afghanistan  1967    11537966      Asia    34.020  836.197138
Afghanistan  1972    13079460      Asia    36.088  739.981106
…            …          …       …        …          …
Zimbabwe     1987     9216418    Africa    62.351  706.157306
Zimbabwe     1992    10704340    Africa    60.377  693.420786
Zimbabwe     1997    11404948    Africa    46.809  792.449960
Zimbabwe     2002    11926563    Africa    39.989  672.038623
Zimbabwe     2007    12311143    Africa    43.487  469.709298

[1704 rows x 5 columns]
```

[129]:
```
temp2 = temp.set_index('continent')
```

[130]:
```
temp2.loc['Asia']
```

[130]:
```
           year  population  life_exp      gdp_cap
continent
Asia       1952     8425333    28.801   779.445314
Asia       1957     9240934    30.332   820.853030
Asia       1962    10267083    31.997   853.100710
Asia       1967    11537966    34.020   836.197138
Asia       1972    13079460    36.088   739.981106
…           …          …       …          …
Asia       1987    11219340    52.922  1971.741538
Asia       1992    13367997    55.599  1879.496673
Asia       1997    15826497    58.020  2117.484526
Asia       2002    18701257    60.308  2234.820827
Asia       2007    22211743    62.698  2280.769906

[396 rows x 4 columns]
```

6 Reset index to original index in a dataframe

[131]:
```
temp.reset_index(inplace=True)
```

[132]:
```
temp
```

[132]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
…	…	…	…	…	…	…
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

[1704 rows x 6 columns]

7 Add a new row

[227]:
```
df = pd.read_csv("mckinsey (1).csv")
```

[228]:
```
df.head()
```

[228]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106

Creating a dict of new values

[229]:
```
new_row = {'country':'India','year':1988,'population':94343233,'continent':
 ↪'Asia','life_exp':89.99,'gdp_cap':679.90}
```

[230]:
```
new_row
```

[230]:
```
{'country': 'India',
 'year': 1988,
 'population': 94343233,
 'continent': 'Asia',
 'life_exp': 89.99,
 'gdp_cap': 679.9}
```

[189]:
```
df.loc[1704] = new_row
```

[190]:
```
df.tail(3)
```

[190]:

	country	year	population	continent	life_exp	gdp_cap
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298
1704	India	1988	94343233	Asia	89.990	679.900000

[224]:
```
df
```

[224]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
…	…	…	…	…	…	…
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

[1704 rows x 6 columns]

8 Update the single cell value

data_frame.at[row_number, column_name] = value_to_update

[ ]:
```
df.at[1703,'population'] = 29601212.324530516
```

[231]:
```
n = int(input('Enter how many extra row you want to add:'))
for i in range(0,n):
    df.loc[len(df.index)+i] = new_row
    print(len(df.index)+i)
df.reset_index(inplace=True)
```

```
Enter how many extra row you want to add: 10

1705
1707
1709
1711
1713
1715
1717
1719
```

1721
1723

9 Dropping the rows

[233]:
```
df.drop(1704,axis=0)
```

[233]:

	index	country	year	population	continent	life_exp	gdp_cap
0	0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	4	Afghanistan	1972	13079460	Asia	36.088	739.981106
…	…	…	…	…	…	…	…
1709	1714	India	1988	94343233	Asia	89.990	679.900000
1710	1716	India	1988	94343233	Asia	89.990	679.900000
1711	1718	India	1988	94343233	Asia	89.990	679.900000
1712	1720	India	1988	94343233	Asia	89.990	679.900000
1713	1722	India	1988	94343233	Asia	89.990	679.900000

[1713 rows x 7 columns]

[237]:
```
df.drop([1709,1710,1711,1712,1713],axis=0)
```

[237]:

	index	country	year	population	continent	life_exp	gdp_cap
0	0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	4	Afghanistan	1972	13079460	Asia	36.088	739.981106
…	…	…	…	…	…	…	…
1704	1704	India	1988	94343233	Asia	89.990	679.900000
1705	1706	India	1988	94343233	Asia	89.990	679.900000
1706	1708	India	1988	94343233	Asia	89.990	679.900000
1707	1710	India	1988	94343233	Asia	89.990	679.900000
1708	1712	India	1988	94343233	Asia	89.990	679.900000

[1709 rows x 7 columns]

[234]:
```
df.iloc[1:5]
```

[234]:

	index	country	year	population	continent	life_exp	gdp_cap
1	1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	4	Afghanistan	1972	13079460	Asia	36.088	739.981106

10 In loc the end range becomes inclusive

[235]:
```
df.loc[10:20]
```

[235]:

	index	country	year	population	continent	life_exp	gdp_cap
10	10	Afghanistan	2002	25268405	Asia	42.129	726.734055
11	11	Afghanistan	2007	31889923	Asia	43.828	974.580338
12	12	Albania	1952	1282697	Europe	55.230	1601.056136
13	13	Albania	1957	1476505	Europe	59.280	1942.284244
14	14	Albania	1962	1728137	Europe	64.820	2312.888958
15	15	Albania	1967	1984060	Europe	66.220	2760.196931
16	16	Albania	1972	2263554	Europe	67.690	3313.422188
17	17	Albania	1977	2509048	Europe	68.930	3533.003910
18	18	Albania	1982	2780097	Europe	70.420	3630.880722
19	19	Albania	1987	3075321	Europe	72.000	3738.932735
20	20	Albania	1992	3326498	Europe	71.581	2497.437901

[244]:
```
df2 = df.tail(10)
```

[251]:
```
df2
```

[251]:

	country	year	population	continent	life_exp	gdp_cap
1704	India	1988	94343233	Asia	89.99	679.9
1705	India	1988	94343233	Asia	89.99	679.9
1706	India	1988	94343233	Asia	89.99	679.9
1707	India	1988	94343233	Asia	89.99	679.9
1708	India	1988	94343233	Asia	89.99	679.9
1709	India	1988	94343233	Asia	89.99	679.9
1710	India	1988	94343233	Asia	89.99	679.9
1711	India	1988	94343233	Asia	89.99	679.9
1712	India	1988	94343233	Asia	89.99	679.9
1713	India	1988	94343233	Asia	89.99	679.9

11 To check how many duplicates are there in data

[252]:
```
df2.duplicated()
```

[252]:
```
1704    False
1705     True
1706     True
1707     True
1708     True
1709     True
1710     True
1711     True
1712     True
1713     True
```

```
dtype: bool
```

[ ]: