```
import pandas as pd
import numpy as np
```

```
!wget "https://drive.google.com/uc?export=download&id=1E3bwvYGf1ig32RmcYiWc0IXPN-mD_bI " -O mckinsey.csv
```

```
--2023-01-19 15:30:52--  https://drive.google.com/uc?export=download&id=1E3bwvYGf1ig32RmcYiWc0IXPN-mD_bI_
Resolving drive.google.com (drive.google.com)... 173.194.192.100, 173.194.192.139, 173.194.192.113, ...
Connecting to drive.google.com (drive.google.com)|173.194.192.100|:443... connected.
HTTP request sent, awaiting response... 303 See Other
Location: https://doc-0s-68-docs.googleusercontent.com/docs/securesc/ha0ro937gcuc7l7deffksulhg5h7mbp1/629fa4p8sf9ulcghmrk
Warning: wildcards not supported in HTTP.
--2023-01-19 15:30:53--  https://doc-0s-68-docs.googleusercontent.com/docs/securesc/ha0ro937gcuc7l7deffksulhg5h7mbp1/629f
Resolving doc-0s-68-docs.googleusercontent.com (doc-0s-68-docs.googleusercontent.com)... 209.85.146.132, 2607:f8b0:4001:c
Connecting to doc-0s-68-docs.googleusercontent.com (doc-0s-68-docs.googleusercontent.com)|209.85.146.132|:443... connecte
HTTP request sent, awaiting response... 200 OK
Length: 83785 (82K) [text/csv]
Saving to: 'mckinsey.csv'

mckinsey.csv        100%[===================>]  81.82K  --.-KB/s    in 0.001s

2023-01-19 15:30:53 (68.0 MB/s) - 'mckinsey.csv' saved [83785/83785]
```

```
df = pd.read_csv('mckinsey.csv')
```

```
df
```

	country	year	population	continent	life_exp	gdp_cap
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

Automatic saving failed. This file was updated remotely or in another tab. Show diff

1704 rows × 6 columns

```
df.iloc[1:5, 1:4]
```

	year	population	continent
1	1957	9240934	Asia
2	1962	10267083	Asia
3	1967	11537966	Asia
4	1972	13079460	Asia

```
df.loc[1:5, 1:4]
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-6-494208dc7680> in <module>
----> 1 df.loc[1:5, 1:4]

8 frames
/usr/local/lib/python3.8/dist-packages/pandas/core/indexes/base.py in _maybe_cast_slice_bound(self, label, side, kind)
   5747         # reject them, if index does not contain label
   5748         if (is_float(label) or is_integer(label)) and label not in self._values:
-> 5749             raise self._invalid_indexer("slice", label)
   5750
   5751         return label

TypeError: cannot do slice indexing on Index with these indexers [1] of type int
```

SEARCH STACK OVERFLOW

```
df.loc[1:5, ["country", "continent"]]
```

	country	continent
1	Afghanistan	Asia
2	Afghanistan	Asia
3	Afghanistan	Asia
4	Afghanistan	Asia
5	Afghanistan	Asia

```
df.loc[1:5, "year":"life_exp"]
```

	year	population	continent	life_exp
1	1957	9240934	Asia	30.332
2	1962	10267083	Asia	31.997
3	1967	11537966	Asia	34.020
4	1972	13079460	Asia	36.088
5	1977	14880372	Asia	38.438

```
df.iloc[[0, 10, 100], [0, 2, 3]]
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

10	Afghanistan	25268405	Asia
100	Bangladesh	70759295	Asia

```
df.iloc[1:10:2]
```

	country	year	population	continent	life_exp	gdp_cap
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
5	Afghanistan	1977	14880372	Asia	38.438	786.113360
7	Afghanistan	1987	13867957	Asia	40.822	852.395945
9	Afghanistan	1997	22227415	Asia	41.763	635.341351

```
df.loc[1:10:3]
```

	country	year	population	continent	life_exp	gdp_cap
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
7	Afghanistan	1987	13867957	Asia	40.822	852.395945
10	Afghanistan	2002	25268405	Asia	42.129	726.734055

```
df.loc[1:10, "year":"life_exp":2]
```

	year	continent
2	1962	Asia

```
le = df["life_exp"]
```

```
le.mean()
```

```
59.474439366197174
```

```
le.sum()
```

```
101344.44467999999
```

```
le.count()
```

```
1704
```

	year	continent
10	2002	Asia

```
le.sum() / le.count()
```

```
59.474439366197174
```

```
df
```

	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
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 × 6 columns


```
df.sort_values(["life_exp"])
```

	country	year	population	continent	life_exp	gdp_cap
1292	Rwanda	1992	7290203	Africa	23.599	737.068595
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
552	Gambia	1952	284320	Africa	30.000	485.230659
36	Angola	1952	4232095	Africa	30.015	3520.610273
1344	Sierra Leone	1952	2143249	Africa	30.331	879.787736
...	...	...	...	...	...	...
1487	Switzerland	2007	7554661	Europe	81.701	37506.419070
695	Iceland	2007	301931	Europe	81.757	36180.789190
802	Japan	2002	127065841	Asia	82.000	28604.591900
671	Hong Kong, China	2007	6980412	Asia	82.208	39724.978670
803	Japan	2007	127467972	Asia	82.603	31656.068060

1704 rows × 6 columns

```
df.sort_values(["life_exp"], ascending=False)
```

	country	year	population	continent	life_exp	gdp_cap
803	Japan	2007	127467972	Asia	82.603	31656.068060
671	Hong Kong, China	2007	6980412	Asia	82.208	39724.978670
802	Japan	2002	127065841	Asia	82.000	28604.591900
695	Iceland	2007	301931	Europe	81.757	36180.789190
1487	Switzerland	2007	7554661	Europe	81.701	37506.419070
...	...	...	...	...	...	...
1344	Sierra Leone	1952	2143249	Africa	30.331	879.787736
36	Angola	1952	4232095	Africa	30.015	3520.610273

```
df.sort_values(["year", "life_exp"], ascending=[True, False])
```

	country	year	population	continent	life_exp	gdp_cap
1140	Norway	1952	3327728	Europe	72.670	10095.421720
684	Iceland	1952	147962	Europe	72.490	7267.688428
1080	Netherlands	1952	10381988	Europe	72.130	8941.571858
1464	Sweden	1952	7124673	Europe	71.860	8527.844662
408	Denmark	1952	4334000	Europe	70.780	9692.385245
...	...	...	...	...	...	...
1691	Zambia	2007	11746035	Africa	42.384	1271.211593
1043	Mozambique	2007	19951656	Africa	42.082	823.685621
1463	Swaziland	2007	1133066	Africa	39.613	4513.480643

1704 rows × 6 columns


```
users = pd.DataFrame({"userid":[1, 2, 3], "name":["sharadh", "shahid", "khusalli"]})
users
```

	userid	name
0	1	sharadh
1	2	shahid
2	3	khusalli

```
msgs = pd.DataFrame({"userid":[1, 1, 2, 4], "msg":['hmm', "acha", "theek hai", "nice"]})
msgs
```

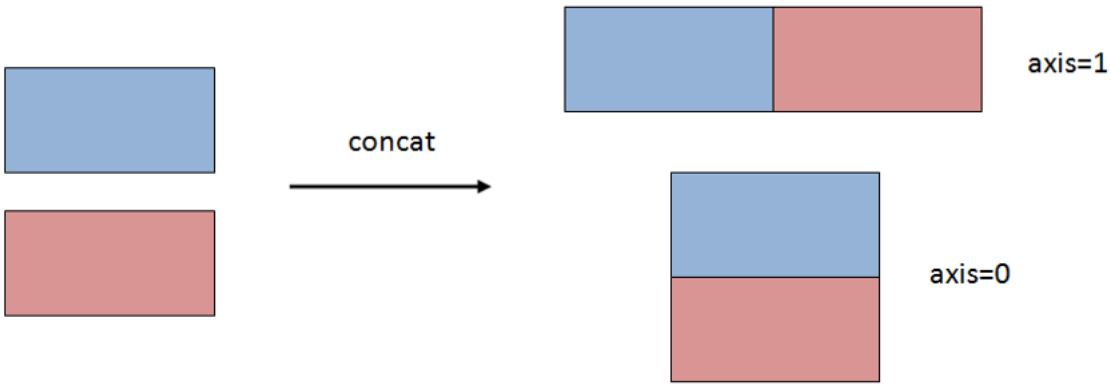
	userid	msg
0	1	hmm
1	1	acha
2	2	theek hai
3	4	nice

```
pd.concat([users, msgs], axis=0, ignore_index=True)
```

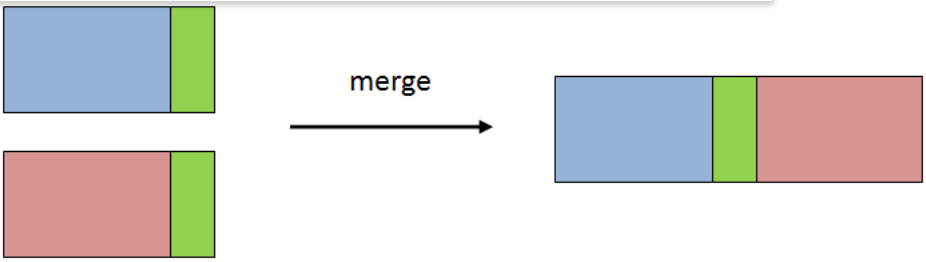
	userid	name	msg
0	1	sharadh	NaN
1	2	shahid	NaN
2	3	khusalli	NaN
3	1	NaN	hmm
4	1	NaN	acha
5	2	NaN	theek hai
6	4	NaN	nice

```
pd.concat([users, msgs], axis=1)
```

	userid	name	userid	msg
0	1.0	sharadh	1	hmm
1	2.0	shahid	1	acha
2	3.0	khusalli	2	theek hai
3	NaN	NaN	4	nice

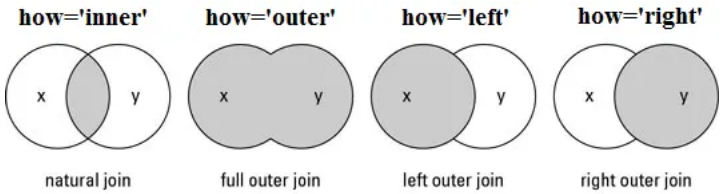


Automatic saving failed. This file was updated remotely or in another tab. Show diff




```
users.merge(msgs, on="userid")
```

	userid	name	msg
0	1	sharadh	hmm
1	1	sharadh	acha
2	2	shahid	theek hai




```
# get info of all ther users and all the msgs
users.merge(msgs, on="userid",how="outer")
```

	userid	name	msg
0	1	sharadh	hmm
1	1	sharadh	acha
2	2	shahid	theek hai
3	3	khusalli	NaN
4	4	NaN	nice

```
# what if I want to presever the details of all the users?
users.merge(msgs, on="userid", how="left")
```

	userid	name	msg
0	1	sharadh	hmm
1	1	sharadh	acha
2	2	shahid	theek hai
3	3	khusalli	NaN

```
# all the messages and omly the users who have sent the messages
users.merge(msgs, on="userid", how="right")
```

	userid	name	msg
0	1	sharadh	hmm
1	1	sharadh	acha
2	2	shahid	theek hai
3	4	NaN	nice

```
users.rename({"userid":"id"}, axis=1, inplace=True)
```

```
users
```

	id	name
2	3	khusalli

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
temp = users.merge(msgs, left_on="id", right_on="userid")
temp
```

	id	name	userid	msg
0	1	sharadh	1	hmm
1	1	sharadh	1	acha
2	2	shahid	2	theek hai

```
temp.drop("userid", axis=1, inplace=True)
```

```
temp
```

	id	name	msg
0	1	sharadh	hmm
1	1	sharadh	acha
2	2	shahid	theek hai

✓ 0s completed at 22:10

Automatic saving failed. This file was updated remotely or in another tab. Show diff