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

In [2]:
```
dataset=pd.read_csv("example.csv")
```

In [3]:
```
dataset
```

Out[3]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver
3	4	Orazio	Frayling	ofrayling3@economist.com	Male	25.199.143.143	Kungur
4	5	Julianne	Murrison	jmurrison4@cbslocal.com	Female	10.186.243.144	Sainte-Luce-sur-Loire
...	...	...	...	...	...	...	...
995	996	Parrnell	Jeff	pjeffrn@4shared.com	Male	159.55.55.39	Dengok
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário do Sul
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaopingba
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethlehem
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Złoty

1000 rows × 7 columns

In [4]:
```
dataset['id'].sum()
```

Out[4]:
```
500500
```

In [5]:
```
dataset.sum()
```

Out[5]:
```
id                                                       500500
first_name    JosephFreidaNanniOrazioJulianneLucyDyanaKassey...
last_name     ZaniolIniDrillingcourtHerityFraylingMurrisonGa...
email         jzaniolini0@simplemachines.orgfdrillingcourt1@...
gender        MaleFemaleFemaleMaleFemaleFemaleFemaleFemaleMa...
ip_address    163.168.68.13297.212.102.79145.151.178.9825.19...
city          Pedro LeopoldoBuriClaverKungurSainte-Luce-sur-...
dtype: object
```

In [6]:
```
dataset['id'].mean()
```

Out[6]:
```
500.5
```

In [7]:
```
dataset.describe()
```

Out[7]:

	id
count	1000.000000
mean	500.500000
std	288.819436
min	1.000000
25%	250.750000
50%	500.500000
75%	750.250000
max	1000.000000

In [8]:
```
len(dataset)
```

Out[8]:
```
1000
```

In [9]:
```
dataset.head()
```

Out[9]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver
3	4	Orazio	Frayling	ofrayling3@economist.com	Male	25.199.143.143	Kungur
4	5	Julianne	Murrison	jmurrison4@cbslocal.com	Female	10.186.243.144	Sainte-Luce-sur-Loire

In [10]:
```
dataset.tail()
```

Out[10]:

	id	first_name	last_name	email	gender	ip_address	city
995	996	Parrnell	Jeff	pjeffrn@4shared.com	Male	159.55.55.39	Dengok
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário do Sul
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaopingba
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethlehem
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Złoty

In [11]:
```
ds=dataset.iloc[0:1]
ds
```

Out[11]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo

In [12]:
```
ds
```

Out[12]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo

In [13]:
```
ds=pd.DataFrame(ds)
```

In [14]:
```
ds
```

Out[14]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo

In [15]:
```
# Filtering
dataset[dataset.id>500]
```

Out[15]:

	id	first_name	last_name	email	gender	ip_address	c
500	501	Ambur	Holdey	aholdeydw@xinhuanet.com	Female	236.209.51.231	Sawaeng
501	502	Sabine	Lyosik	slyosikdx@360.cn	Female	161.232.137.136	Bayt ‘Īn
502	503	Gladys	Bandt	gbandtdy@ocn.ne.jp	Female	158.14.30.142	Hunk
503	504	Paten	Vasyukhichev	pvasyukhichevdz@illinois.edu	Male	175.216.192.130	Nando
504	505	Hunter	Creggan	hcreggane0@amazon.co.uk	Male	208.173.52.40	Albue
...	...	...	...	...	...	...	
995	996	Parrnell	Jeff	pjeffrn@4shared.com	Male	159.55.55.39	Deng
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário S
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaoping
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethleh
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Zł

500 rows × 7 columns

In [16]:
```
dataset.rename(columns={'ip_address':'ip'})
```

Out[16]:

	id	first_name	last_name	email	gender	ip	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver
3	4	Orazio	Frayling	ofrayling3@economist.com	Male	25.199.143.143	Kungur
4	5	Julianne	Murrison	jmurrison4@cbslocal.com	Female	10.186.243.144	Sainte- Luce-sur- Loire
...	...	...	...	...	...	...	...
995	996	Parrnell	Jeff	pjeffrn@4shared.com	Male	159.55.55.39	Dengok
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário do Sul
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaopingba
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethlehem
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Złoty

1000 rows × 7 columns

In [17]:

```
df=dataset.drop_duplicates()
print(df,id(df),id(dataset))
```

```
       id first_name      last_name                             email  gender  \
0       1     Joseph      Zaniolini  jzaniolini0@simplemachines.org    Male   
1       2     Freida  Drillingcourt         fdrillingcourt1@umich.edu  Female   
2       3      Nanni         Herity          nherity2@statcounter.com  Female   
3       4     Orazio       Frayling          ofrayling3@economist.com    Male   
4       5   Julianne       Murrison           jmurrison4@cbslocal.com  Female   
..    ...        ...            ...                               ...     ...   
995   996   Parrnell           Jeff               pjeffrn@4shared.com    Male   
996   997        Pat       Warstall           pwarstallro@dropbox.com    Male   
997   998     Willyt     Vannuccini       wvannuccinirp@tuttocitta.it  Female   
998   999     Gaelan     Petrashkov       gpetrashkovrq@hostgator.com    Male   
999  1000   Stirling       Francesc       sfrancescrr@cyberchimps.com    Male   

         ip_address                   city  
0    163.168.68.132         Pedro Leopoldo  
1     97.212.102.79                   Buri  
2    145.151.178.98                 Claver  
3    25.199.143.143                 Kungur  
4    10.186.243.144  Sainte-Luce-sur-Loire  
..              ...                    ...  
995    159.55.55.39                 Dengok  
996   169.200.7.131         Rosário do Sul  
997  156.149.18.155             Xiaopingba  
998  219.109.169.75              Bethlehem  
999   108.193.9.206            Potok Złoty  

[1000 rows x 7 columns] 2181652110480 2181651126048
```

In [18]:

```
df=dataset.dropna()
print(df,id(df),id(dataset))
```

```
       id first_name    last_name                          email  gender  \
0       1     Joseph    Zaniolini  jzaniolini0@simplemachines.org    Male   
1       2     Freida  Drillingcourt       fdrillingcourt1@umich.edu  Female   
2       3      Nanni       Herity      nherity2@statcounter.com  Female   
3       4     Orazio     Frayling      ofrayling3@economist.com    Male   
4       5   Julianne     Murrison       jmurrison4@cbslocal.com  Female   
..    ...        ...          ...                            ...     ...   
995   996   Parrnell         Jeff          pjeffrn@4shared.com    Male   
996   997        Pat     Warstall        pwarstallro@dropbox.com    Male   
997   998     Willyt   Vannuccini     wvannuccinirp@tuttocitta.it  Female   
998   999     Gaelan   Petrashkov     gpetrashkovrq@hostgator.com    Male   
999  1000   Stirling     Francesc     sfrancescrr@cyberchimps.com    Male   

         ip_address                   city  
0    163.168.68.132         Pedro Leopoldo  
1     97.212.102.79                   Buri  
2    145.151.178.98                 Claver  
3    25.199.143.143                 Kungur  
4    10.186.243.144  Sainte-Luce-sur-Loire  
..              ...                    ...  
995    159.55.55.39                 Dengok  
996   169.200.7.131         Rosário do Sul  
997  156.149.18.155             Xiaopingba  
998  219.109.169.75              Bethlehem  
999   108.193.9.206            Potok Złoty  

[1000 rows x 7 columns] 2181651133056 2181651126048
```

```
In [19]: dataset.fillna(value="shiva")
```

```
Out[19]:
        id  first_name     last_name                         email  gender      ip_address           city
   0     1      Joseph     Zaniolini jzaniolini0@simplemachines.org   Male  163.168.68.132          Pedro
                                                                                                 Leopoldo
   1     2      Freida  Drillingcourt      fdrillingcourt1@umich.edu Female   97.212.102.79           Buri
   2     3       Nanni        Herity     nherity2@statcounter.com  Female  145.151.178.98         Claver
   3     4      Orazio      Frayling      ofrayling3@economist.com   Male  25.199.143.143         Kungur
                                                                                                   Sainte-
   4     5    Julianne      Murrison      jmurrison4@cbslocal.com  Female  10.186.243.144       Luce-sur-
                                                                                                    Loire
 ...   ...         ...           ...                           ...     ...             ...            ...
 995   996    Parrnell          Jeff          pjeffrn@4shared.com   Male    159.55.55.39         Dengok
 996   997         Pat      Warstall      pwarstallro@dropbox.com   Male   169.200.7.131     Rosário do
                                                                                                      Sul
 997   998      Willyt    Vannuccini   wvannuccinirp@tuttocitta.it Female  156.149.18.155     Xiaopingba
 998   999      Gaelan    Petrashkov  gpetrashkovrq@hostgator.com   Male  219.109.169.75      Bethlehem
 999  1000    Stirling      Francesc sfrancescrr@cyberchimps.com    Male   108.193.9.206    Potok Złoty

1000 rows × 7 columns
```

```
In [20]: pd.isnull(dataset)
```

```
Out[20]:
        id  first_name  last_name  email  gender  ip_address   city
   0  False      False      False  False   False       False  False
   1  False      False      False  False   False       False  False
   2  False      False      False  False   False       False  False
   3  False      False      False  False   False       False  False
   4  False      False      False  False   False       False  False
 ...    ...        ...        ...    ...     ...         ...    ...
 995  False      False      False  False   False       False  False
 996  False      False      False  False   False       False  False
 997  False      False      False  False   False       False  False
 998  False      False      False  False   False       False  False
 999  False      False      False  False   False       False  False

1000 rows × 7 columns
```

In [21]:

```
# viewing data
df.head(4)
```

Out[21]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver
3	4	Orazio	Frayling	ofrayling3@economist.com	Male	25.199.143.143	Kungur

In [22]:

```
df.tail(4)
```

Out[22]:

	id	first_name	last_name	email	gender	ip_address	city
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário do Sul
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaopingba
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethlehem
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Złoty

In [23]:

```
df.dtypes
```

Out[23]:

```
id             int64
first_name    object
last_name     object
email         object
gender        object
ip_address    object
city          object
dtype: object
```

In [24]:

```
df.columns
```

Out[24]:

```
Index(['id', 'first_name', 'last_name', 'email', 'gender', 'ip_address',
       'city'],
      dtype='object')
```

In [25]:
```
df.sort_values(by=['id'])
```

Out[25]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver
3	4	Orazio	Frayling	ofrayling3@economist.com	Male	25.199.143.143	Kungur
4	5	Julianne	Murrison	jmurrison4@cbslocal.com	Female	10.186.243.144	Sainte-Luce-sur-Loire
...	...	...	...	...	...	...	...
995	996	Parrnell	Jeff	pjeffrn@4shared.com	Male	159.55.55.39	Dengok
996	997	Pat	Warstall	pwarstallro@dropbox.com	Male	169.200.7.131	Rosário do Sul
997	998	Willyt	Vannuccini	wvannuccinirp@tuttocitta.it	Female	156.149.18.155	Xiaopingba
998	999	Gaelan	Petrashkov	gpetrashkovrq@hostgator.com	Male	219.109.169.75	Bethlehem
999	1000	Stirling	Francesc	sfrancescrr@cyberchimps.com	Male	108.193.9.206	Potok Złoty

1000 rows × 7 columns

In [26]:
```
df.sort_values(by=['email'])
```

Out[26]:

	id	first_name	last_name	email	gender	ip_address	
219	220	Albrecht	Airds	aairds63@drupal.org	Male	209.213.159.128	Palaic
646	647	Abbie	Allett	aalletthy@accuweather.com	Female	14.20.177.82	D
944	945	Aaron	Antonsen	aantonsenq8@whitehouse.gov	Male	0.192.253.81	
660	661	Alex	Ayce	aayceic@mail.ru	Female	50.148.243.117	
356	357	Aaron	Baker	abaker9w@spiegel.de	Male	218.209.50.211	E
...	...	...	...	...	...	...	
881	882	Yorke	Hurran	yhurranoh@flickr.com	Male	153.15.188.46	
814	815	Yuma	Tattershall	ytattershallmm@spiegel.de	Male	115.219.51.8	
404	405	York	Tockell	ytockellb8@moonfruit.com	Male	193.255.184.166	A
469	470	Zorah	MacMenemy	zmacmenemyd1@seattletimes.com	Female	207.96.93.51	
824	825	Zelig	Sparrowhawk	zsparrowhawkmw@vimeo.com	Male	108.125.27.193	

1000 rows × 7 columns

In [27]:

```
# viewing data
df[0:3]
```

Out[27]:

	id	first_name	last_name	email	gender	ip_address	city
0	1	Joseph	Zaniolini	jzaniolini0@simplemachines.org	Male	163.168.68.132	Pedro Leopoldo
1	2	Freida	Drillingcourt	fdrillingcourt1@umich.edu	Female	97.212.102.79	Buri
2	3	Nanni	Herity	nherity2@statcounter.com	Female	145.151.178.98	Claver

In [29]:

```
df["ip_address"]
```

Out[29]:

```
0       163.168.68.132
1        97.212.102.79
2       145.151.178.98
3       25.199.143.143
4       10.186.243.144
             ...
995       159.55.55.39
996      169.200.7.131
997     156.149.18.155
998     219.109.169.75
999      108.193.9.206
Name: ip_address, Length: 1000, dtype: object
```

In [31]:

```
df.iloc[2,3]
```

Out[31]:

```
'nherity2@statcounter.com'
```

In [36]:

```
df1=df[["id",'email']]
df2=df[["ip_address","email"]]
```

In [35]:
```
pd.merge(df1,df2,on='email')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
Input In [35], in <cell line: 1>()
----> 1 pd.merge(df1,df2,on='email')

File ~\AppData\Roaming\Python\Python310\site-packages\pandas\core\reshape\mer
ge.py:107, in merge(left, right, how, on, left_on, right_on, left_index, righ
t_index, sort, suffixes, copy, indicator, validate)
     90 @Substitution("\nleft : DataFrame or named Series")
     91 @Appender(_merge_doc, indents=0)
     92 def merge(
   (...)
    105     validate: str | None = None,
    106 ) -> DataFrame:
--> 107     op = _MergeOperation(
    108         left,
    109         right,
    110         how=how,
    111         on=on,
    112         left_on=left_on,
    113         right_on=right_on,
    114         left_index=left_index,
    115         right_index=right_index,
    116         sort=sort,
    117         suffixes=suffixes,
    118         copy=copy,
    119         indicator=indicator,
    120         validate=validate,
    121     )
    122     return op.get_result()

File ~\AppData\Roaming\Python\Python310\site-packages\pandas\core\reshape\mer
ge.py:700, in _MergeOperation.__init__(self, left, right, how, on, left_on, r
ight_on, axis, left_index, right_index, sort, suffixes, copy, indicator, vali
date)
    693 self._cross = cross_col
    695 # note this function has side effects
    696 (
    697     self.left_join_keys,
    698     self.right_join_keys,
    699     self.join_names,
--> 700 ) = self._get_merge_keys()
    702 # validate the merge keys dtypes. We may need to coerce
    703 # to avoid incompatible dtypes
    704 self._maybe_coerce_merge_keys()

File ~\AppData\Roaming\Python\Python310\site-packages\pandas\core\reshape\mer
ge.py:1097, in _MergeOperation._get_merge_keys(self)
   1095 if not is_rkey(rk):
   1096     if rk is not None:
-> 1097         right_keys.append(right._get_label_or_level_values(rk))
   1098     else:
   1099         # work-around for merge_asof(right_index=True)
   1100         right_keys.append(right.index)
```

```
File ~\AppData\Roaming\Python\Python310\site-packages\pandas\core\generic.py:
1840, in NDFrame._get_label_or_level_values(self, key, axis)
   1838     values = self.axes[axis].get_level_values(key)._values
   1839 else:
-> 1840     raise KeyError(key)
   1842 # Check for duplicates
   1843 if values.ndim > 1:

KeyError: 'email'
```

In [ ]: