installing the 3rd party python library

In [1]:
```
# <syntax> pip install package_name
!pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\admin\anaconda3\envs\gpuenv
\lib\site-packages (1.3.5)
Requirement already satisfied: numpy>=1.17.3 in c:\users\admin\anaconda3\envs
\gpuenv\lib\site-packages (from pandas) (1.18.5)
Requirement already satisfied: pytz>=2017.3 in c:\users\admin\anaconda3\envs
\gpuenv\lib\site-packages (from pandas) (2022.7.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\admin\anaco
nda3\envs\gpuenv\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\gpue
nv\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
```

Importing the library

In [2]:
```
# syntax import package_name
import pandas as pd
```

In [3]:
```
string_data = ['mehul','nilesh','faizan','vishal','chandu']
age = [26, 52,14,18,19]

data = tuple(zip(string_data, age))
```

In [4]:
```
df = pd.DataFrame(data, columns=['name','age'])
```

In [5]:
```
df
```

Out[5]:

	name	age
0	mehul	26
1	nilesh	52
2	faizan	14
3	vishal	18
4	chandu	19

In [6]:
```
# average age
df.age.mean()
```

Out[6]:
```
25.8
```

In [7]:
```
# median
df.age.median()
```

Out[7]:
```
19.0
```

In [8]:
```
df[df['age'] >=20]
```

Out[8]:

	name	age
0	mehul	26
1	nilesh	52

In [9]:
```
# read csv format file using pandas dataframe
df = pd.read_csv('student.csv')
# to view first n records in dataframe
df.head(10)
```

Out[9]:

	id	name	class	mark	gender
0	1	John Deo	Four	75	female
1	2	Max Ruin	Three	85	male
2	3	Arnold	Three	55	male
3	4	Krish Star	Four	60	female
4	5	John Mike	Four	60	female
5	6	Alex John	Four	55	male
6	7	My John Rob	Fifth	78	male
7	8	Asruid	Five	85	male
8	9	Tes Qry	Six	78	male
9	10	Big John	Four	55	female

In [10]:
```
# to view last 10 records in dataframe
df.tail(10)
```

Out[10]:

	id	name	class	mark	gender
25	26	Crelea	Seven	79	male
26	27	Big Nose	Three	81	female
27	28	Rojj Base	Seven	86	female
28	29	Tess Played	Seven	55	male
29	30	Reppy Red	Six	79	female
30	31	Marry Toeey	Four	88	male
31	32	Binn Rott	Seven	90	female
32	33	Kenn Rein	Six	96	female
33	34	Gain Toe	Seven	69	male
34	35	Rows Noump	Six	88	female

In [11]:
```
df.head()
```

Out[11]:

	id	name	class	mark	gender
0	1	John Deo	Four	75	female
1	2	Max Ruin	Three	85	male
2	3	Arnold	Three	55	male
3	4	Krish Star	Four	60	female
4	5	John Mike	Four	60	female

In [12]:
```
# view columns in dataframe
df.columns
```

Out[12]:
```
Index(['id', 'name', 'class', 'mark', 'gender'], dtype='object')
```

In [13]:
```
# check information about dataframe
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   id      35 non-null     int64 
 1   name    35 non-null     object
 2   class   35 non-null     object
 3   mark    35 non-null     int64 
 4   gender  35 non-null     object
dtypes: int64(2), object(3)
memory usage: 1.5+ KB
```

In [14]:
```
df.shape
# (num_rows, num_cols)
```

Out[14]:
```
(35, 5)
```

In [15]:
```
columns_read =['name','gender']
df_gender_info = df[columns_read]
```

In [16]:
```
df_gender_info.head()
```

Out[16]:

	name	gender
0	John Deo	female
1	Max Ruin	male
2	Arnold	male
3	Krish Star	female
4	John Mike	female

In [17]:
```
df.gender.value_counts()
```

Out[17]:
```
male      18
female    17
Name: gender, dtype: int64
```

In [18]:
```
df.name.unique()
```

Out[18]:
```
array(['John Deo', 'Max Ruin', 'Arnold', 'Krish Star', 'John Mike',
       'Alex John', 'My John Rob', 'Asruid', 'Tes Qry', 'Big John',
       'Ronald', 'Recky', 'Kty', 'Bigy', 'Tade Row', 'Gimmy', 'Tumyu',
       'Honny', 'Tinny', 'Jackly', 'Babby John', 'Reggid', 'Herod',
       'Tiddy Now', 'Giff Tow', 'Crelea', 'Big Nose', 'Rojj Base',
       'Tess Played', 'Reppy Red', 'Marry Toeey', 'Binn Rott',
       'Kenn Rein', 'Gain Toe', 'Rows Noump'], dtype=object)
```

In [19]:
```
df[df['mark']>= 90]
```

Out[19]:

	id	name	class	mark	gender
11	12	Recky	Six	94	female
31	32	Binn Rott	Seven	90	female
32	33	Kenn Rein	Six	96	female

In [20]:
```
df.mark.mean()
```

Out[20]:
```
74.65714285714286
```

In [21]:
```
# median
df.mark.median()
```

Out[21]:
```
79.0
```

In [22]:
```
df.mark.mode()
```

Out[22]:
```
0    88
dtype: int64
```

In [23]:
```
df[df.mark == 88]
```

Out[23]:

	id	name	class	mark	gender
12	13	Kty	Seven	88	female
13	14	Bigy	Seven	88	female
14	15	Tade Row	Four	88	male
15	16	Gimmy	Four	88	male
24	25	Giff Tow	Seven	88	male
30	31	Marry Toeey	Four	88	male
34	35	Rows Noump	Six	88	female

In [24]:
```
df[df.mark == 18]['gender']= 'female'
```

```
c:\Users\admin\anaconda3\envs\gpuEnv\lib\site-packages\ipykernel_launcher.py:
1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [25]:
```
df[df.mark == 18]
```

Out[25]:

	id	name	class	mark	gender
18	19	Tinny	Nine	18	male

In [26]:
```
df.iloc[18]['gender'] = 'female'
```

```
c:\Users\admin\anaconda3\envs\gpuEnv\lib\site-packages\pandas\core\series.py:
1056: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  cacher_needs_updating = self._check_is_chained_assignment_possible()
```

In [27]:
```
df.iloc[18]
```

Out[27]:
```
id           19
name      Tinny
class      Nine
mark         18
gender     male
Name: 18, dtype: object
```

In [28]:
```
df.iat[18,4]
```

Out[28]:
```
'male'
```

In [29]:
```
df.iat[18,4] = 'other'
```

In [30]:
```
df.iloc[18]
```

Out[30]:
```
id           19
name      Tinny
class      Nine
mark         18
gender    other
Name: 18, dtype: object
```

In [31]:
```
df.gender.value_counts()
```

Out[31]:
```
female    17
male      17
other      1
Name: gender, dtype: int64
```

Dataframe using dictionary

In [32]:
```
data = {
    'name':string_data,
    'age': age
}
data
```

Out[32]:
```
{'name': ['mehul', 'nilesh', 'faizan', 'vishal', 'chandu'],
 'age': [26, 52, 14, 18, 19]}
```

In [33]:
```
df_dict = pd.DataFrame(data)
df_dict
```

Out[33]:

	name	age
0	mehul	26
1	nilesh	52
2	faizan	14
3	vishal	18
4	chandu	19

In [34]:
```
virat_kohli = pd.read_html(r'https://en.wikipedia.org/wiki/List_of_internation
```

In [40]:
```
virat_kohli[3].to_csv('virat_kohli_centuries.csv',index=False)
```

In [42]:
```
test_centuries = virat_kohli[3]
```

In [43]:
```
test_centuries.head()
```

Out[43]:

	No.	Runs	Against	Pos.	Inn.	Test	Venue	H/A	Date	Result	Ref
0	1	116	Australia	6	2	4/4	Adelaide Oval, Adelaide	Away	24 January 2012	Lost	[26]
1	2	103 †	New Zealand	5	2	2/2	M. Chinnaswamy Stadium, Bangalore	Home	31 August 2012	Won	[27]
2	3	103	England	5	2	4/4	Vidarbha Cricket Association Stadium, Nagpur	Home	13 December 2012	Drawn	[28]
3	4	107	Australia	5	2	1/4	M. A. Chidambaram Stadium, Chennai	Home	22 February 2013	Won	[29]
4	5	119 †	South Africa	4	1	1/2	Wanderers Stadium, Johannesburg	Away	18 December 2013	Drawn	[30]

In [47]:
```
test_centuries['H/A'].value_counts()
```

Out[47]:
```
Away    15
Home    14
Name: H/A, dtype: int64
```

In [49]:
```
test_centuries.shape[0]
```

Out[49]:
```
29
```

In [50]:
```
total_row  = test_centuries.shape[0]
test_centuries['H/A'].value_counts()/ total_row
```

Out[50]:
```
Away    0.517241
Home    0.482759
Name: H/A, dtype: float64
```

In [52]:
```
test_centuries['Against'].value_counts()
```

Out[52]:
```
Australia       8
England         5
Sri Lanka       5
New Zealand     3
South Africa    3
West Indies     3
Bangladesh      2
Name: Against, dtype: int64
```

In [58]:
```
test_centuries[test_centuries['Against'] == 'Australia']['H/A'].value_counts()
```

Out[58]:
```
Away    0.75
Home    0.25
Name: H/A, dtype: float64
```

In [65]:
```
test_centuries['Result'].value_counts()
```

Out[65]:
```
Won      13
Drawn     9
Lost      7
Name: Result, dtype: int64
```

In [ ]:
```

```