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

In [2]: ```python
import xlrd
```

In [3]: ```python
df=pd.read_csv('emp.csv')
```

In [4]: ```python
df
```

Out[4]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25	20
1	1002	Lakshmi	F	22	20
2	1003	Vinod Kumar	M	30	22
3	1004	Neelam	F	33	30
4	1005	Reethu	F	21	19
5	1006	Gaurav Pandey	M	40	25
6	1007	Kulkarni	M	30	24
7	1008	Vishal Muhammad	M	29	19
8	1009	Meena Singh	F	25	21
9	1010	Ankhit Chandra	M	40	26

In [5]: ```python
loc() locate function
iloc() index based function
```

In [6]: ```python
df.loc[:,['name','gender']]
```

Out[6]:

	name	gender
0	Ganesh Rao	M
1	Lakshmi	F
2	Vinod Kumar	M
3	Neelam	F
4	Reethu	F
5	Gaurav Pandey	M
6	Kulkarni	M
7	Vishal Muhammad	M
8	Meena Singh	F
9	Ankhit Chandra	M

In [7]: ```python
df1=df.loc[:,['name','gender']]
```

In [8]: `df1`

Out[8]:

|   | name | gender |
|---|---|---|
| 0 | Ganesh Rao | M |
| 1 | Lakshmi | F |
| 2 | Vinod Kumar | M |
| 3 | Neelam | F |
| 4 | Reethu | F |
| 5 | Gaurav Pandey | M |
| 6 | Kulkarni | M |
| 7 | Vishal Muhammad | M |
| 8 | Meena Singh | F |
| 9 | Ankhit Chandra | M |

In [9]: `df1=df.iloc[:,[1,3]]`

In [10]: `df1`

Out[10]:

|   | name | age |
|---|---|---|
| 0 | Ganesh Rao | 25 |
| 1 | Lakshmi | 22 |
| 2 | Vinod Kumar | 30 |
| 3 | Neelam | 33 |
| 4 | Reethu | 21 |
| 5 | Gaurav Pandey | 40 |
| 6 | Kulkarni | 30 |
| 7 | Vishal Muhammad | 29 |
| 8 | Meena Singh | 25 |
| 9 | Ankhit Chandra | 40 |

In [11]: `df1=df.iloc[0:3,[1,3]]`

In [12]: `df1`

Out[12]:

|   | name | age |
|---|---|---|
| 0 | Ganesh Rao | 25 |
| 1 | Lakshmi | 22 |
| 2 | Vinod Kumar | 30 |

In [13]: `df.shape`

Out[13]: `(10, 5)`

In [14]: `r,c=df.shape`

In [15]: `r`

Out[15]: `10`

In [16]: `c`

Out[16]: `5`

In [17]: `df[2:5]`

Out[17]:

|   | id | name | gender | age | salary |
|---|------|-------------|--------|-----|--------|
| **2** | 1003 | Vinod Kumar | M | 30 | 22 |
| **3** | 1004 | Neelam | F | 33 | 30 |
| **4** | 1005 | Reethu | F | 21 | 19 |

In [19]: `df[0::2]` *#one gap*

Out[19]:

|   | id | name | gender | age | salary |
|---|------|-------------|--------|-----|--------|
| **0** | 1001 | Ganesh Rao | M | 25 | 20 |
| **2** | 1003 | Vinod Kumar | M | 30 | 22 |
| **4** | 1005 | Reethu | F | 21 | 19 |
| **6** | 1007 | Kulkarni | M | 30 | 24 |
| **8** | 1009 | Meena Singh | F | 25 | 21 |

In [20]: `df.columns`

Out[20]: `Index(['id', 'name', 'gender', 'age', 'salary'], dtype='object')`

In [21]: `df.name`

Out[21]:
```
0   Ganesh Rao
1   Lakshmi
2   Vinod Kumar
3   Neelam
4   Reethu
5   Gaurav Pandey
6   Kulkarni
7   Vishal Muhammad
8   Meena Singh
9   Ankhit Chandra
Name: name, dtype: object
```

In [22]: `df[['name','age']]`

Out[22]:

|   | name | age |
|---|------|-----|
| 0 | Ganesh Rao | 25 |
| 1 | Lakshmi | 22 |
| 2 | Vinod Kumar | 30 |
| 3 | Neelam | 33 |
| 4 | Reethu | 21 |
| 5 | Gaurav Pandey | 40 |
| 6 | Kulkarni | 30 |
| 7 | Vishal Muhammad | 29 |
| 8 | Meena Singh | 25 |
| 9 | Ankhit Chandra | 40 |

In [23]: `df.salary`

Out[23]:
```
0   20
1   20
2   22
3   30
4   19
5   25
6   24
7   19
8   21
9   26
Name: salary, dtype: int64
```

In [24]: `df.salary.max()`

Out[24]: 30

In [25]: `df.describe()`

Out[25]:

|       | id | age | salary |
|-------|------|------|--------|
| count | 10.00000 | 10.000000 | 10.000000 |
| mean | 1005.50000 | 29.500000 | 22.600000 |
| std | 3.02765 | 6.687468 | 3.596294 |
| min | 1001.00000 | 21.000000 | 19.000000 |
| 25% | 1003.25000 | 25.000000 | 20.000000 |
| 50% | 1005.50000 | 29.500000 | 21.500000 |
| 75% | 1007.75000 | 32.250000 | 24.750000 |
| max | 1010.00000 | 40.000000 | 30.000000 |

In [26]: `df.salary>12`

Out[26]:
```
0   True
1   True
2   True
3   True
4   True
5   True
6   True
7   True
8   True
9   True
Name: salary, dtype: bool
```

In [27]: `df[df.salary>12]`

Out[27]:

|   | id | name | gender | age | salary |
|---|-----|---------------|--------|-----|--------|
| 0 | 1001 | Ganesh Rao | M | 25 | 20 |
| 1 | 1002 | Lakshmi | F | 22 | 20 |
| 2 | 1003 | Vinod Kumar | M | 30 | 22 |
| 3 | 1004 | Neelam | F | 33 | 30 |
| 4 | 1005 | Reethu | F | 21 | 19 |
| 5 | 1006 | Gaurav Pandey | M | 40 | 25 |
| 6 | 1007 | Kulkarni | M | 30 | 24 |
| 7 | 1008 | Vishal Muhammad | M | 29 | 19 |
| 8 | 1009 | Meena Singh | F | 25 | 21 |
| 9 | 1010 | Ankhit Chandra | M | 40 | 26 |

In [28]: `df[df.salary>21]`

Out[28]:

|   | id | name | gender | age | salary |
|---|-----|---------------|--------|-----|--------|
| 2 | 1003 | Vinod Kumar | M | 30 | 22 |
| 3 | 1004 | Neelam | F | 33 | 30 |
| 5 | 1006 | Gaurav Pandey | M | 40 | 25 |
| 6 | 1007 | Kulkarni | M | 30 | 24 |
| 9 | 1010 | Ankhit Chandra | M | 40 | 26 |

In [29]: `df[df.salary==df.salary.max()]`

Out[29]:

|   | id | name | gender | age | salary |
|---|-----|--------|--------|-----|--------|
| 3 | 1004 | Neelam | F | 33 | 30 |

In [30]: `df.sort_values('salary')`

Out[30]:

| | id | name | gender | age | salary |
|---|---|---|---|---|---|
| **4** | 1005 | Reethu | F | 21 | 19 |
| **7** | 1008 | Vishal Muhammad | M | 29 | 19 |
| **0** | 1001 | Ganesh Rao | M | 25 | 20 |
| **1** | 1002 | Lakshmi | F | 22 | 20 |
| **8** | 1009 | Meena Singh | F | 25 | 21 |
| **2** | 1003 | Vinod Kumar | M | 30 | 22 |
| **6** | 1007 | Kulkarni | M | 30 | 24 |
| **5** | 1006 | Gaurav Pandey | M | 40 | 25 |
| **9** | 1010 | Ankhit Chandra | M | 40 | 26 |
| **3** | 1004 | Neelam | F | 33 | 30 |

In [31]:
```python
df.salary.mean()
```

Out[31]: 22.6

In [32]:
```python
df=pd.read_csv('emp.csv')
```

In [33]:
```python
df
```

Out[33]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25.0	20
1	1002	Lakshmi	F	22.0	20
2	1003	Vinod Kumar	M	30.0	22
3	1004	Neelam	F	33.0	30
4	1005	NaN	F	21.0	19
5	1006	Gaurav Pandey	M	40.0	25
6	1007	Kulkarni	M	30.0	24
7	1008	Vishal Muhammad	M	29.0	19
8	1009	Meena Singh	F	25.0	21
9	1010	Ankhit Chandra	M	NaN	26

In [34]:
```python
# NaN -> default marker for missing values
```

In [35]:
```python
df1=df.fillna(0)
```

In [36]:
```python
df1
```

Out[36]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25.0	20
1	1002	Lakshmi	F	22.0	20
2	1003	Vinod Kumar	M	30.0	22
3	1004	Neelam	F	33.0	30
4	1005	0	F	21.0	19
5	1006	Gaurav Pandey	M	40.0	25
6	1007	Kulkarni	M	30.0	24
7	1008	Vishal Muhammad	M	29.0	19
8	1009	Meena Singh	F	25.0	21
9	1010	Ankhit Chandra	M	0.0	26

In [37]:
```python
df1=df.fillna({'name': 'name is missing here','salary':0})
```

In [38]:
```python
df1
```

Out[38]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25.0	20
1	1002	Lakshmi	F	22.0	20
2	1003	Vinod Kumar	M	30.0	22
3	1004	Neelam	F	33.0	30
4	1005	name is missing here	F	21.0	19
5	1006	Gaurav Pandey	M	40.0	25
6	1007	Kulkarni	M	30.0	24
7	1008	Vishal Muhammad	M	29.0	19
8	1009	Meena Singh	F	25.0	21
9	1010	Ankhit Chandra	M	NaN	26

In [41]:
```python
df1=df.fillna({'name': 'name is missing here','age':0})
```

In [42]:
```python
df1
```

Out[42]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25.0	20
1	1002	Lakshmi	F	22.0	20
2	1003	Vinod Kumar	M	30.0	22
3	1004	Neelam	F	33.0	30
4	1005	name is missing here	F	21.0	19
5	1006	Gaurav Pandey	M	40.0	25
6	1007	Kulkarni	M	30.0	24
7	1008	Vishal Muhammad	M	29.0	19
8	1009	Meena Singh	F	25.0	21
9	1010	Ankhit Chandra	M	0.0	26

In [48]:
```python
# joining dataframe
df1=pd.DataFrame({'name':['rd','vk','sg'],'age':[30,24,31]})
```

In [44]:
```python
df1
```

Out[44]:

	name	age
0	rd	30
1	vk	24
2	sg	31

In [45]:
```python
df2=pd.DataFrame({'name':['rd','vk','sg'],'salary':[40,70,15]})
```

In [46]:
```python
df2
```

Out[46]:

	name	salary
0	rd	40
1	vk	70
2	sg	15

In [47]:
```python
pd.merge(df1,df2,on='name',how='inner')
```

Out[47]:

	name	age	salary
0	rd	30	40
1	vk	24	70
2	sg	31	15

In [49]:
```python
pd.merge(df1,df2,on='name',how='left')
```

Out[49]:

	name	age	salary
0	rd	30	40
1	vk	24	70
2	sg	31	15

In [50]:
```
#visualization
```

In [51]:
```
# A bar graph
```

In [ ]:

In [53]:
```
df
```

Out[53]:

	id	name	gender	age	salary
0	1001	Ganesh Rao	M	25.0	20
1	1002	Lakshmi	F	22.0	20
2	1003	Vinod Kumar	M	30.0	22
3	1004	Neelam	F	33.0	30
4	1005	NaN	F	21.0	19
5	1006	Gaurav Pandey	M	40.0	25
6	1007	Kulkarni	M	30.0	24
7	1008	Vishal Muhammad	M	29.0	19
8	1009	Meena Singh	F	25.0	21
9	1010	Ankhit Chandra	M	NaN	26

In [ ]:

In [55]:
```
y=df['salary']
```

In [56]:
```
y
```

Out[56]:
```
0    20
1    20
2    22
3    30
4    19
5    25
6    24
7    19
8    21
9    26
Name: salary, dtype: int64
```

In [57]:
```
x
```

```
Out[57]:  0          Ganesh Rao
          1             Lakshmi
          2        Vinod Kumar
          3              Neelam
          4                 NaN
          5       Gaurav Pandey
          6            Kulkarni
          7     Vishal Muhammad
          8         Meena Singh
          9      Ankhit Chandra
          Name: name, dtype: object
```

```python
In [66]:  import matplotlib.pyplot as plt
          import pandas as pd
          a=pd.read_csv('ss.csv')
          df=pd.DataFrame(a)
          x=df['name']
          y=df['salary']
          plt.bar(x,y,label='employee data',color='red')
          plt.xlabel('emp name')
          plt.title('my company')
          plt.legend()
          plt.show()
```



```
In [ ]:
```