In [2]:
```
import pandas as pd
```

In [3]:
```
df=pd.read_csv('g:/dataset/analysis/restaurant.csv')
```

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

Out[4]:

	total_bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

In [5]:
```
df['tax']=df.total_bill*.05
```

In [6]:
```
df
```

Out[6]:

	total_bill	tip	gender	smoker	day	time	size	tax
0	16.99	1.01	Female	No	Sun	Dinner	2	0.8495
1	10.34	1.66	Male	No	Sun	Dinner	3	0.5170
2	21.01	3.50	Male	No	Sun	Dinner	3	1.0505
3	23.68	3.31	Male	No	Sun	Dinner	2	1.1840
4	24.59	3.61	Female	No	Sun	Dinner	4	1.2295
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	1.4515
240	27.18	2.00	Female	Yes	Sat	Dinner	2	1.3590
241	22.67	2.00	Male	Yes	Sat	Dinner	2	1.1335
242	17.82	1.75	Male	No	Sat	Dinner	2	0.8910
243	18.78	3.00	Female	No	Thur	Dinner	2	0.9390

244 rows × 8 columns

In [7]:
```
df.drop('tax',axis=1,inplace=True)
```

In [8]:
```
df
```

Out[8]:

	total_bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

In [9]:
```
def cal_tax(bill):
    if (bill<10):
        return bill*.05
    elif(bill>=10 and bill<20):
        return bill*.10
    else:
        return bill*.15
```

In [10]:
```
cal_tax(9)
```

Out[10]:
```
0.45
```

In [11]:
```
cal_tax(12)
```

Out[11]:
```
1.2000000000000002
```

In [12]:
```
cal_tax(50)
```

Out[12]:
```
7.5
```

In [13]:
```
df['tax']=df.total_bill.apply(cal_tax)
```

In [14]:
```
df
```

Out[14]:

	total_bill	tip	gender	smoker	day	time	size	tax
0	16.99	1.01	Female	No	Sun	Dinner	2	1.6990
1	10.34	1.66	Male	No	Sun	Dinner	3	1.0340
2	21.01	3.50	Male	No	Sun	Dinner	3	3.1515
3	23.68	3.31	Male	No	Sun	Dinner	2	3.5520
4	24.59	3.61	Female	No	Sun	Dinner	4	3.6885
...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	4.3545
240	27.18	2.00	Female	Yes	Sat	Dinner	2	4.0770
241	22.67	2.00	Male	Yes	Sat	Dinner	2	3.4005
242	17.82	1.75	Male	No	Sat	Dinner	2	1.7820
243	18.78	3.00	Female	No	Thur	Dinner	2	1.8780

244 rows × 8 columns

In [15]:
```
df['final_bill']=df.total_bill+df.tax
```

In [16]:
```
df
```

Out[16]:

	total_bill	tip	gender	smoker	day	time	size	tax	final_bill
0	16.99	1.01	Female	No	Sun	Dinner	2	1.6990	18.6890
1	10.34	1.66	Male	No	Sun	Dinner	3	1.0340	11.3740
2	21.01	3.50	Male	No	Sun	Dinner	3	3.1515	24.1615
3	23.68	3.31	Male	No	Sun	Dinner	2	3.5520	27.2320
4	24.59	3.61	Female	No	Sun	Dinner	4	3.6885	28.2785
...	...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	4.3545	33.3845
240	27.18	2.00	Female	Yes	Sat	Dinner	2	4.0770	31.2570
241	22.67	2.00	Male	Yes	Sat	Dinner	2	3.4005	26.0705
242	17.82	1.75	Male	No	Sat	Dinner	2	1.7820	19.6020
243	18.78	3.00	Female	No	Thur	Dinner	2	1.8780	20.6580

244 rows × 9 columns

In [17]:
```
df.final_bill.sum()
```

Out[17]:
```
5441.539
```

In [18]:
```
df['final_bill']=df.total_bill+df.tax+df.tip
```

In [19]:
```
df
```

Out[19]:

	total_bill	tip	gender	smoker	day	time	size	tax	final_bill
0	16.99	1.01	Female	No	Sun	Dinner	2	1.6990	19.6990
1	10.34	1.66	Male	No	Sun	Dinner	3	1.0340	13.0340
2	21.01	3.50	Male	No	Sun	Dinner	3	3.1515	27.6615
3	23.68	3.31	Male	No	Sun	Dinner	2	3.5520	30.5420
4	24.59	3.61	Female	No	Sun	Dinner	4	3.6885	31.8885
...	...	...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3	4.3545	39.3045
240	27.18	2.00	Female	Yes	Sat	Dinner	2	4.0770	33.2570
241	22.67	2.00	Male	Yes	Sat	Dinner	2	3.4005	28.0705
242	17.82	1.75	Male	No	Sat	Dinner	2	1.7820	21.3520
243	18.78	3.00	Female	No	Thur	Dinner	2	1.8780	23.6580

244 rows × 9 columns

In [20]:
```
df.final_bill.sum()
```

Out[20]:
```
6173.119000000001
```

In [21]:
```
df.loc[500]=[20,5,'Male','No','Fri','Lunch',3,2,22]
```

In [22]:
```
df
```

Out[22]:

	total_bill	tip	gender	smoker	day	time	size	tax	final_bill
0	16.99	1.01	Female	No	Sun	Dinner	2	1.6990	19.6990
1	10.34	1.66	Male	No	Sun	Dinner	3	1.0340	13.0340
2	21.01	3.50	Male	No	Sun	Dinner	3	3.1515	27.6615
3	23.68	3.31	Male	No	Sun	Dinner	2	3.5520	30.5420
4	24.59	3.61	Female	No	Sun	Dinner	4	3.6885	31.8885
...	...	...	...	...	...	...	...	...	...
240	27.18	2.00	Female	Yes	Sat	Dinner	2	4.0770	33.2570
241	22.67	2.00	Male	Yes	Sat	Dinner	2	3.4005	28.0705
242	17.82	1.75	Male	No	Sat	Dinner	2	1.7820	21.3520
243	18.78	3.00	Female	No	Thur	Dinner	2	1.8780	23.6580
500	20.00	5.00	Male	No	Fri	Lunch	3	2.0000	22.0000

245 rows × 9 columns

In [23]:
```
df=pd.read_csv('g:/dataset/analysis/emp.txt')
df
```

Out[23]:

	empid	empname	empsal
0	1	sonu	10000
1	2	monu	12000
2	3	chintu	8000
3	4	pintu	11000

In [24]:
```
df.index=df.empid
```

In [25]:
```
df
```

Out[25]:

empid	empid	empname	empsal
1	1	sonu	10000
2	2	monu	12000
3	3	chintu	8000
4	4	pintu	11000

In [26]:
```
df.drop('empid',axis=1,inplace=True)
```

In [27]:
```
df
```

Out[27]:

empid	empname	empsal
1	sonu	10000
2	monu	12000
3	chintu	8000
4	pintu	11000

In [28]:
```
df.loc[[3]]
```

Out[28]:

empid	empname	empsal
3	chintu	8000

In [29]:
```
eid=int(input('enter empid: '))
df.loc[[eid]]
```

Out[29]:

empid	empname	empsal
4	pintu	11000

In [30]:
```
eid=int(input('enter empid: '))
df.loc[[eid]]
```

Out[30]:

empid	empname	empsal
2	monu	12000

In [31]:
```
try:
    eid=int(input('enter empid: '))
    print(df.loc[[eid]])
except:
    print('eid does not exist')
```

```
      empname  empsal
empid
2        monu   12000
```

In [32]:
```
df=pd.read_csv('g:/dataset/analysis/restaurant.csv')
df
```

Out[32]:

	total_bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

In [33]:
```
df.groupby('time')[['day']].count()
```

Out[33]:

time	day
Dinner	176
Lunch	68

In [34]:
```
df3=df.groupby('time')[['tip']].agg(['min','max','sum','mean'])
```

In [35]:
```
df3
```

Out[35]:

	tip			
	min	max	sum	mean
time				
Dinner	1.00	10.0	546.07	3.102670
Lunch	1.25	6.7	185.51	2.728088

In [36]:
```
df3.loc[['Dinner']]
```

Out[36]:

	tip			
	min	max	sum	mean
time				
Dinner	1.0	10.0	546.07	3.10267

In [37]:
```
df3.loc[['Lunch']]
```

Out[37]:

	tip			
	min	max	sum	mean
time				
Lunch	1.25	6.7	185.51	2.728088

In [38]:
```
df_cs=pd.read_csv('g:/dataset/analysis/cs.txt')
df_cs
```

Out[38]:

	roll	name	age
0	105	dhoni	22
1	106	pant	19

In [39]:
```
df_ec=pd.read_csv('g:/dataset/analysis/ec.txt')
df_ec
```

Out[39]:

	roll	name	age
0	108	dhoni	23
1	106	shaw	18

In [40]:
```
df_res=pd.concat([df_cs,df_ec],axis=0,ignore_index=True)
df_res
```

Out[40]:

	roll	name	age
0	105	dhoni	22
1	106	pant	19
2	108	dhoni	23
3	106	shaw	18

In [41]:
```
df_res=pd.concat([df_cs,df_ec],axis=0,ignore_index=True,sort=True)
df_res
```

Out[41]:

	age	name	roll
0	22	dhoni	105
1	19	pant	106
2	23	dhoni	108
3	18	shaw	106

In [42]:
```
df_res=pd.concat([df_cs,df_ec],axis=0)
df_res
```

Out[42]:

	roll	name	age
0	105	dhoni	22
1	106	pant	19
0	108	dhoni	23
1	106	shaw	18

In [43]:
```
df_res.reset_index()
```

Out[43]:

	index	roll	name	age
0	0	105	dhoni	22
1	1	106	pant	19
2	0	108	dhoni	23
3	1	106	shaw	18

In [44]:
```
df_res.reset_index(drop=True)
```

Out[44]:

	roll	name	age
0	105	dhoni	22
1	106	pant	19
2	108	dhoni	23
3	106	shaw	18

In [45]:
```
df1=pd.read_csv('g:/dataset/analysis/emp.txt')
```

In [47]:
```
df1
```

Out[47]:

	empid	empname	empsal
0	1	sonu	10000
1	2	monu	12000
2	3	chintu	8000
3	4	pintu	11000

In [48]:
```
df2=pd.read_csv('g:/dataset/analysis/emp_details.txt')
```

In [49]:
```
df2
```

Out[49]:

	empid	city	mob
0	1	Noida	123
1	2	Delhi	456
2	3	Noida	789
3	5	Noida	555

In [61]:
```
df1.merge(df2,on='empid',how='inner')
```

Out[61]:

	empid	empname	empsal	city	mob
0	1	sonu	10000	Noida	123
1	2	monu	12000	Delhi	456
2	3	chintu	8000	Noida	789

In [62]:
```
df1.merge(df2,on='empid',how='left')
```

Out[62]:

	empid	empname	empsal	city	mob
0	1	sonu	10000	Noida	123.0
1	2	monu	12000	Delhi	456.0
2	3	chintu	8000	Noida	789.0
3	4	pintu	11000	NaN	NaN

In [63]:
```
df1.merge(df2,on='empid',how='right')
```

Out[63]:

	empid	empname	empsal	city	mob
0	1	sonu	10000.0	Noida	123
1	2	monu	12000.0	Delhi	456
2	3	chintu	8000.0	Noida	789
3	5	NaN	NaN	Noida	555

In [64]:
```
df1.merge(df2,on='empid',how='outer')
```

Out[64]:

	empid	empname	empsal	city	mob
0	1	sonu	10000.0	Noida	123.0
1	2	monu	12000.0	Delhi	456.0
2	3	chintu	8000.0	Noida	789.0
3	4	pintu	11000.0	NaN	NaN
4	5	NaN	NaN	Noida	555.0

In [66]:
```
df=pd.read_csv("g:/dataset/analysis/ufo.csv")
df
```

Out[66]:

	City	Colors Reported	Shape Reported	State	Time
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:00
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:00
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:00
3	Abilene	NaN	DISK	KS	6/1/1931 13:00
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:00
...	...	...	...	...	...
18236	Grant Park	NaN	TRIANGLE	IL	12/31/2000 23:00
18237	Spirit Lake	NaN	DISK	IA	12/31/2000 23:00
18238	Eagle River	NaN	NaN	WI	12/31/2000 23:45
18239	Eagle River	RED	LIGHT	WI	12/31/2000 23:45
18240	Ybor	NaN	OVAL	FL	12/31/2000 23:59

18241 rows × 5 columns

In [ ]:
```
```