```
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
                16.99 1.01 Female
          0
                                     No Sun Dinner
                                                      2
                          Male
          1
               10.34 1.66
                                     No
                                         Sun Dinner
                21.01 3.50
          2
                          Male
                                    No
                                         Sun Dinner
                                                      3
          3
                23.68 3.31
                                                      2
                          Male
                                     No
                                         Sun Dinner
                24.59 3.61 Female
          4
                                    No
                                         Sun Dinner
                                                      4
               ... ... ...
                                    •••
                                             ... ...
                29.03 5.92
        239
                          Male
                                    No
                                         Sat Dinner
                                                      3
        240
                27.18 2.00 Female
                                    Yes
                                         Sat Dinner
        241
                22.67 2.00
                          Male
                                    Yes
                                         Sat Dinner
                                                      2
        242
                17.82 1.75
                                                      2
                          Male
                                     No
                                          Sat Dinner
        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
                                                      2 0.8495
                                         Sun Dinner
                                     No
               10.34 1.66
                          Male
                                     No
                                         Sun Dinner
                                                      3 0.5170
                                                      3 1.0505
               21.01 3.50
          2
                          Male
                                         Sun Dinner
                                     No
                          Male
                23.68 3.31
                                     No
                                             Dinner
                                                      2 1.1840
                                         Sun
                24.59 3.61 Female
                                         Sun Dinner
                                                      4 1.2295
          4
                                    No
        239
                29.03 5.92
                           Male
                                         Sat Dinner
                                                      3 1.4515
                                    No
                27.18 2.00 Female
        240
                                    Yes
                                          Sat Dinner
                                                      2 1.3590
        241
                22.67 2.00
                            Male
                                          Sat Dinner
                                                      2 1.1335
                                    Yes
        242
                17.82 1.75
                            Male
                                     No
                                          Sat Dinner
                                                      2 0.8910
        243
               18.78 3.00 Female
                                                      2 0.9390
                                    No Thur Dinner
       244 rows × 8 columns
```

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

```
df
 In [8]:
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
                                             Sun Dinner
                                                           3
                              Male
                                        No
            3
                  23.68 3.31
                               Male
                                        No
                                             Sun Dinner
                                                           2
                  24.59 3.61 Female
                                                           4
            4
                                        No
                                             Sun Dinner
                  ... ...
          239
                  29.03 5.92
                               Male
                                             Sat Dinner
                                                           3
                                        No
                  27.18 2.00 Female
          240
                                        Yes
                                              Sat Dinner
                                                           2
          241
                  22.67 2.00
                                                           2
                               Male
                                             Sat Dinner
                                        Yes
                  17.82 1.75
          242
                             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):</pre>
                  return bill*.05
              elif(bill>=10 and bill<20):</pre>
                  return bill*.10
              else:
                  return bill*.15
In [10]:
          cal_tax(9)
          0.45
Out[10]:
          cal_tax(12)
In [11]:
          1.20000000000000002
Out[11]:
In [12]:
          cal_tax(50)
         7.5
Out[12]:
In [13]: df['tax']=df.total_bill.apply(cal_tax)
In [14]:
```

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 r	ows × 8 co	olumr	ıs						
In [15]:	df['	final_bi	11']=	df.tota	1_bill+c	lf.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
	•	10.55	1.01	геппате	NO	Sull	Dillillei	2	1.0550	10.0050
	1	10.34	1.66	Male	No		Dinner		1.0340	11.3740
							Dinner			
	1	10.34	1.66 3.50	Male	No	Sun Sun	Dinner	3	1.0340	11.3740
	1 2	10.34 21.01	1.66 3.50 3.31	Male Male	No No	Sun Sun Sun	Dinner Dinner	3 3 2	1.0340 3.1515	11.3740 24.1615
	1 2 3	10.34 21.01 23.68	1.66 3.50 3.31	Male Male Male	No No No	Sun Sun Sun	Dinner Dinner Dinner	3 3 2	1.0340 3.1515 3.5520	11.3740 24.1615 27.2320
	1 2 3 4	10.34 21.01 23.68 24.59	1.66 3.50 3.31 3.61 	Male Male Male Female	No No No	Sun Sun Sun Sun	Dinner Dinner Dinner Dinner	3 3 2 4 	1.0340 3.1515 3.5520 3.6885	11.3740 24.1615 27.2320 28.2785
	1 2 3 4 	10.34 21.01 23.68 24.59	1.66 3.50 3.31 3.61 5.92	Male Male Male Female	No No No No	Sun Sun Sun Sun Sun Sun	Dinner Dinner Dinner Dinner	3 3 2 4 	1.0340 3.1515 3.5520 3.6885	11.3740 24.1615 27.2320 28.2785
	1 2 3 4 239	10.34 21.01 23.68 24.59 29.03	1.66 3.50 3.31 3.61 5.92	Male Male Male Female Male	No No No No 	Sun Sun Sun Sun Sun Sun Sat	Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3	1.0340 3.1515 3.5520 3.6885 4.3545	11.3740 24.1615 27.2320 28.2785 33.3845
	1 2 3 4 239 240	10.34 21.01 23.68 24.59 29.03 27.18	1.66 3.50 3.31 3.61 5.92 2.00 2.00	Male Male Male Female Male	No No No No Wo Yes	Sun Sun Sun Sun Sun Sun Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570
	1 2 3 4 239 240 241	10.34 21.01 23.68 24.59 29.03 27.18 22.67	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75	Male Male Male Female Male Female Male	No No No No Wo Yes	Sun Sun Sun Sun Sun Sat Sat Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705
	1 2 3 4 239 240 241 242 243	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00	Male Male Male Female Male Female Male Male Female	No No No No No Yes Yes No	Sun Sun Sun Sun Sun Sat Sat Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020
	1 2 3 4 239 240 241 242 243	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82 18.78	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00	Male Male Male Female Male Female Male Male Female	No No No No No Yes Yes No	Sun Sun Sun Sun Sun Sat Sat Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020
In [17]:	1 2 3 4 239 240 241 242 243	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82 18.78	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00 column	Male Male Male Female Male Female Male Male Female	No No No No No Yes Yes No	Sun Sun Sun Sun Sun Sat Sat Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020
	1 2 3 4 239 240 241 242 243	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82 18.78 Ows × 9 co	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00 column	Male Male Male Female Male Female Male Male Female	No No No No No Yes Yes No	Sun Sun Sun Sun Sun Sat Sat Sat Sat	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020
In [17]:	1 2 3 4 239 240 241 242 243 244 rd 5441	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82 18.78 Ows × 9 co	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00 column	Male Male Male Female Male Female Male Male Male Female ()	No No No No No Yes Yes No No	Sun Sun Sun Sun Sat Sat Sat Thur	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020
In [17]: Out[17]:	1 2 3 4 239 240 241 242 243 244 rd 5441	10.34 21.01 23.68 24.59 29.03 27.18 22.67 17.82 18.78 ows × 9 co	1.66 3.50 3.31 3.61 5.92 2.00 2.00 1.75 3.00 column	Male Male Male Female Male Female Male Male Male Female ()	No No No No No Yes Yes No No	Sun Sun Sun Sun Sat Sat Sat Thur	Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner Dinner	3 3 2 4 3 2 2 2 2	1.0340 3.1515 3.5520 3.6885 4.3545 4.0770 3.4005 1.7820	11.3740 24.1615 27.2320 28.2785 33.3845 31.2570 26.0705 19.6020

Out[19]: total_bill tip gender smoker day time size tax final_bill 0 16.99 1.01 Female Sun Dinner 2 1.6990 19.6990 No 1 10.34 1.66 3 1.0340 13.0340 Male No Sun Dinner 2 21.01 3.50 Male Sun Dinner 3 3.1515 27.6615 No 23.68 3.31 2 3.5520 3 Male No Sun Dinner 30.5420 4 24.59 3.61 Female No Sun Dinner 4 3.6885 31.8885 ••• 3 4.3545 239 29.03 5.92 Male No Sat Dinner 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 No 2 1.7820 21.3520 Male Sat Dinner 243 18.78 3.00 Female No Thur Dinner 2 1.8780 23.6580 244 rows × 9 columns In [20]: df.final bill.sum() 6173.119000000001 Out[20]: df.loc[500]=[20,5,'Male','No','Fri','Lunch',3,2,22] In [21]: In [22]: 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
                               10000
                        sonu
          1
                 2
                               12000
                        monu
          2
                 3
                                8000
                       chintu
                 4
          3
                        pintu
                               11000
          df.index=df.empid
In [24]:
In [25]:
          df
Out[25]:
                 empid empname empsal
          empid
              1
                     1
                            sonu
                                   10000
              2
                     2
                            monu
                                   12000
              3
                     3
                            chintu
                                    8000
                     4
                            pintu
                                   11000
          df.drop('empid',axis=1,inplace=True)
In [26]:
In [27]:
          df
Out[27]:
                 empname empsal
          empid
              1
                            10000
                     sonu
              2
                     monu
                            12000
              3
                             8000
                    chintu
                            11000
                     pintu
In [28]:
          df.loc[[3]]
Out[28]:
                 empname empsal
          empid
              3
                             8000
                    chintu
          eid=int(input('enter empid: '))
In [29]:
          df.loc[[eid]]
Out[29]:
                 empname empsal
          empid
              4
                            11000
                     pintu
          eid=int(input('enter empid: '))
In [30]:
          df.loc[[eid]]
```

```
Out[30]:
                 empname empsal
          empid
              2
                           12000
                    monu
          try:
In [31]:
              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')
Out[32]:
              total_bill tip gender smoker day
                                                  time size
            0
                 16.99 1.01 Female
                                            Sun Dinner
                                                          2
                                        No
            1
                 10.34 1.66
                              Male
                                        No
                                            Sun Dinner
                                                          3
            2
                 21.01 3.50
                              Male
                                            Sun Dinner
                                                          3
                                        No
                 23.68 3.31
                              Male
                                        No
                                            Sun Dinner
                                                          2
                 24.59 3.61 Female
                                                          4
            4
                                        No
                                            Sun Dinner
          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
                                             Sat Dinner
                                                          2
                              Male
                                       Yes
                              Male
          242
                 17.82 1.75
                                        No
                                             Sat Dinner
                                                          2
          243
                 18.78 3.00 Female
                                                          2
                                       No Thur Dinner
         244 rows × 7 columns
In [33]: df.groupby('time')[['day']].count()
Out[33]:
                 day
           time
          Dinner 176
          Lunch
                  68
          df3=df.groupby('time')[['tip']].agg(['min','max','sum','mean'])
In [34]:
          df3
In [35]:
```

```
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
                 1.0 10.0 546.07 3.10267
          Dinner
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
         df_ec=pd.read_csv('g:/dataset/analysis/ec.txt')
In [39]:
          df_ec
Out[39]:
            roll name age
          0 108 dhoni
                        23
          1 106
                 shaw
                        18
         df_res=pd.concat([df_cs,df_ec],axis=0,ignore_index=True)
In [40]:
          df_res
Out[40]:
            roll name age
         0 105
                 dhoni
                        22
          1 106
                        19
                 pant
          2 108 dhoni
                        23
          3 106
                 shaw
                        18
```

tip

Out[35]:

```
df_res=pd.concat([df_cs,df_ec],axis=0,ignore_index=True,sort=True)
In [41]:
Out[41]:
            age name roll
             22 dhoni 105
             19
                pant 106
         2
             23 dhoni 108
         3
             18 shaw 106
In [42]: df_res=pd.concat([df_cs,df_ec],axis=0)
            roll name age
Out[42]:
         0 105 dhoni
                       22
         1 106
                 pant
                       19
         0 108 dhoni
                       23
         1 106
                shaw
                       18
In [43]: df_res.reset_index()
            index roll name age
Out[43]:
         0
               0 105 dhoni
                             22
               1 106
                      pant
                             19
         2
               0 108 dhoni
                             23
               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')
         df1
In [47]:
```

```
Out[47]:
             empid empname empsal
          0
                               10000
                 1
                        sonu
          1
                 2
                               12000
                        monu
          2
                 3
                       chintu
                                8000
                 4
                        pintu
          3
                               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
                    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
                 2
                        monu
                               12000
                                      Delhi
                                             456
          2
                 3
                                8000
                                             789
                       chintu
                                     Noida
          df1.merge(df2,on='empid',how='left')
In [62]:
Out[62]:
             empid empname empsal
                                       city
                                            mob
          0
                 1
                               10000 Noida
                                           123.0
                        sonu
                 2
                        monu
                               12000
                                      Delhi 456.0
          2
                 3
                                8000
                                      Noida 789.0
                       chintu
          3
                 4
                        pintu
                               11000
                                       NaN
                                             NaN
          df1.merge(df2,on='empid',how='right')
In [63]:
Out[63]:
             empid empname empsal
                                       city mob
          0
                 1
                              10000.0 Noida
                                             123
                        sonu
          1
                 2
                             12000.0
                                      Delhi
                                             456
                        monu
          2
                 3
                       chintu
                               0.0008
                                      Noida
                                             789
          3
                 5
                         NaN
                                 NaN Noida
                                             555
          df1.merge(df2,on='empid',how='outer')
In [64]:
```

Out[64]: empid empname empsal city mob 0 10000.0 Noida 1 sonu 123.0 1 2 12000.0 Delhi 456.0 monu 2 3 chintu 8000.0 Noida 789.0 3 4 pintu 11000.0 NaN NaN 5 4 NaN NaN Noida 555.0

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

Out[66]: **Colors Reported** Shape Reported **Time** State 0 6/1/1930 22:00 Ithaca NaN **TRIANGLE** NY 1 Willingboro NaN **OTHER** NJ 6/30/1930 20:00 2 Holyoke NaN OVAL 2/15/1931 14:00 CO 3 Abilene NaN DISK KS 6/1/1931 13:00 4 New York Worlds Fair 4/18/1933 19:00 NY NaN LIGHT 18236 Grant Park **TRIANGLE** 12/31/2000 23:00 NaN 18237 Spirit Lake NaN DISK 12/31/2000 23:00 **Eagle River** 12/31/2000 23:45 18238 NaN NaN 18239 **Eagle River** RED LIGHT 12/31/2000 23:45 18240 FL 12/31/2000 23:59 Ybor NaN OVAL

18241 rows × 5 columns

In []: