In [6]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
pwd
```

Out[2]:
```
'C:\\Users\\DELL'
```

In [7]:
```
df = pd.read_csv('tips.csv')
```

In [8]:
```
df.head()
```

Out[8]:

	total_bill	tip	sex	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	356032
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	447807
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	467613
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	483273

In [9]:
```
df.columns
```

Out[9]:
```
Index(['total_bill', 'tip', 'sex', 'smoker', 'day', 'time', 'size',
       'price_per_person', 'Payer Name', 'CC Number', 'Payment ID'],
      dtype='object')
```

In [10]:
```
df.info
```

Out[10]:
```
<bound method DataFrame.info of      total_bill   tip     sex smoker   da
y    time  size  price_per_person  \
0          16.99  1.01  Female     No   Sun  Dinner     2              8.4
9
1          10.34  1.66    Male     No   Sun  Dinner     3              3.4
5
2          21.01  3.50    Male     No   Sun  Dinner     3              7.0
0
3          23.68  3.31    Male     No   Sun  Dinner     2             11.8
4
4          24.59  3.61  Female     No   Sun  Dinner     4              6.1
5
..           ...   ...     ...    ...   ...     ...   ...
...
239        29.03  5.92    Male     No   Sat  Dinner     3              9.6
8
240        27.18  2.00  Female    Yes   Sat  Dinner     2             13.5
9
241        22.67  2.00    Male    Yes   Sat  Dinner     2             11.3
4
242        17.82  1.75    Male     No   Sat  Dinner     2              8.9
1
243        18.78  3.00  Female     No  Thur  Dinner     2              9.3
9

              Payer Name         CC Number Payment ID
0    Christy Cunningham  3560325168603410    Sun2959
1        Douglas Tucker  4478071379779230    Sun4608
2        Travis Walters  6011812112971322    Sun4458
3      Nathaniel Harris  4676137647685994    Sun5260
4          Tonya Carter  4832732618637221    Sun2251
..                  ...               ...        ...
239       Michael Avila  5296068606052842    Sat2657
240      Monica Sanders  3506806155565404    Sat1766
241          Keith Wong  6011891618747196    Sat3880
242        Dennis Dixon     4375220550950      Sat17
243     Michelle Hardin  3511451626698139    Thur672

[244 rows x 11 columns]>
```

In [11]:
```
def last_four(num):
    return str(num)[-4:]
```

In [12]:
```
last_four(345678)
```

Out[12]:
```
'5678'
```

In [13]:
```
last_four(456)
```

Out[13]:
```
'456'
```

In [14]:
```
df['last_four'] = df['CC Number'].apply(last_four)
```

In [15]:
```
type(df['last_four'])
```

Out[15]:
```
pandas.core.series.Series
```

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

Out[16]:

	total_bill	tip	sex	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	356032
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	447807
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	467613
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	483273

In [18]:
```
df.rename(columns={'sex':'Gender'}, inplace=True)
```

In [19]:
```
df.describe()
```

Out[19]:

	total_bill	tip	size	price_per_person	CC Number
count	244.000000	244.000000	244.000000	244.000000	2.440000e+02
mean	19.785943	2.998279	2.569672	7.888197	2.563496e+15
std	8.902412	1.383638	0.951100	2.914234	2.369340e+15
min	3.070000	1.000000	1.000000	2.880000	6.040679e+10
25%	13.347500	2.000000	2.000000	5.800000	3.040731e+13
50%	17.795000	2.900000	2.000000	7.255000	3.525318e+15
75%	24.127500	3.562500	3.000000	9.390000	4.553675e+15
max	50.810000	10.000000	6.000000	20.270000	6.596454e+15

In [20]:
```
df.head()
```

Out[20]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35603
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	44780
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	46761
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	48327

In [21]:
```
df['total_bill'].mean()
```

Out[21]:
```
19.785942622950824
```

In [22]:
```
df['tip'].max()
```

Out[22]:
```
10.0
```

In [29]:
```
def yess(price):
    if price < 10:
        return 'cheap'
    elif price >= 10 and price < 30:
        return 'Moderate'
    else:
        return 'Very Expensive'
```

In [30]:
```
#pd.set_option('display.float_format')abs
```

In [31]:
```
df['Nature_Bill'] = df['total_bill'].apply(yess)
```

In [32]:
```
df.tail(10)
```

Out[32]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
234	15.53	3.00	Male	Yes	Sat	Dinner	2	7.76	Tracy Douglas	40979
235	10.07	1.25	Male	No	Sat	Dinner	2	5.04	Sean Gonzalez	35340
236	12.60	1.00	Male	Yes	Sat	Dinner	2	6.30	Matthew Myers	35436
237	32.83	1.17	Male	Yes	Sat	Dinner	2	16.42	Thomas Brown	42847
238	35.83	4.67	Female	No	Sat	Dinner	3	11.94	Kimberly Crane	6
239	29.03	5.92	Male	No	Sat	Dinner	3	9.68	Michael Avila	52960
240	27.18	2.00	Female	Yes	Sat	Dinner	2	13.59	Monica Sanders	35068
241	22.67	2.00	Male	Yes	Sat	Dinner	2	11.34	Keith Wong	60118
242	17.82	1.75	Male	No	Sat	Dinner	2	8.91	Dennis Dixon	43
243	18.78	3.00	Female	No	Thur	Dinner	2	9.39	Michelle Hardin	35114

In [33]:
```
def simple(num):
    return num*2
```

In [34]:
```
simple(12)
```

Out[34]:
```
24
```

In [35]:
```
f = lambda num: num*2
```

In [36]:
```
print (f(10))
```

```
20
```

In [37]:
```
x = lambda a,b:a+b
```

In [38]:
```
print (x(23,34))
```

```
57
```

In [39]:
```
y = lambda a,b:a if a>b else b
```

In [40]:
```
print (y(4,5))
```

```
5
```

In [42]:
```
print(y(10,8))
```

```
10
```

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

Out[43]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35603
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	44780
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	46761
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	48327

In [44]:
```
GST = lambda a,b:a*b
```

In [46]:
```
print(GST(10,100))
```

```
1000
```

In [50]:
```
df.head()
```

Out[50]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35603
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	44780
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	46761
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	48327

In [53]:
```
df["GST"]=df['total_bill'].apply(lambda total_bill:total_bill*0.1)
```

In [57]:
```
df.sort_values('total_bill',ascending = False).head()
```

Out[57]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
170	50.81	10.00	Male	Yes	Sat	Dinner	3	16.94	Gregory Clark	5473
212	48.33	9.00	Male	No	Sat	Dinner	4	12.08	Alex Williamson	
59	48.27	6.73	Male	No	Sat	Dinner	4	12.07	Brian Ortiz	6596
156	48.17	5.00	Male	No	Sun	Dinner	6	8.03	Ryan Gonzales	3523
182	45.35	3.50	Male	Yes	Sun	Dinner	3	15.12	Jose Parsons	4112

In [60]:
```
df.sort_values(['tip','total_bill'])
```

Out[60]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
67	3.07	1.00	Female	Yes	Sat	Dinner	1	3.07	Tiffany Brock	43
92	5.75	1.00	Female	Yes	Fri	Dinner	2	2.88	Leah Ramirez	35
111	7.25	1.00	Female	No	Sat	Dinner	1	7.25	Terri Jones	35
236	12.60	1.00	Male	Yes	Sat	Dinner	2	6.30	Matthew Myers	35
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35
...	...	...	...	...	...	...	...	...	...	
141	34.30	6.70	Male	No	Thur	Lunch	6	5.72	Steven Carlson	35
59	48.27	6.73	Male	No	Sat	Dinner	4	12.07	Brian Ortiz	65
23	39.42	7.58	Male	No	Sat	Dinner	4	9.86	Lance Peterson	35
212	48.33	9.00	Male	No	Sat	Dinner	4	12.08	Alex Williamson	
170	50.81	10.00	Male	Yes	Sat	Dinner	3	16.94	Gregory Clark	54

244 rows × 14 columns

In [61]:
```
df[df['tip']>5.0]
```

Out[61]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
23	39.42	7.58	Male	No	Sat	Dinner	4	9.86	Lance Peterson	354
44	30.40	5.60	Male	No	Sun	Dinner	4	7.60	Todd Cooper	
47	32.40	6.00	Male	No	Sun	Dinner	4	8.10	James Barnes	355
52	34.81	5.20	Female	No	Sun	Dinner	4	8.70	Emily Daniel	429
59	48.27	6.73	Male	No	Sat	Dinner	4	12.07	Brian Ortiz	659
85	34.83	5.17	Female	No	Thur	Lunch	4	8.71	Shawna Cook	601
88	24.71	5.85	Male	No	Thur	Lunch	2	12.36	Roger Taylor	
116	29.93	5.07	Male	No	Sun	Dinner	4	7.48	Shawn Blake	468
141	34.30	6.70	Male	No	Thur	Lunch	6	5.72	Steven Carlson	352
155	29.85	5.14	Female	No	Sun	Dinner	5	5.97	Madison Wilson	421
170	50.81	10.00	Male	Yes	Sat	Dinner	3	16.94	Gregory Clark	547
172	7.25	5.15	Male	Yes	Sun	Dinner	2	3.62	Larry White	3
181	23.33	5.65	Male	Yes	Sun	Dinner	2	11.66	Jason Cox	655
183	23.17	6.50	Male	Yes	Sun	Dinner	4	5.79	Dr. Michael James	
211	25.89	5.16	Male	Yes	Sat	Dinner	4	6.47	Christopher Li	601
212	48.33	9.00	Male	No	Sat	Dinner	4	12.08	Alex Williamson	
214	28.17	6.50	Female	Yes	Sat	Dinner	3	9.39	Marissa Jackson	492
239	29.03	5.92	Male	No	Sat	Dinner	3	9.68	Michael Avila	529

In [68]:
```
df.loc[df['day'].isin(['Sat','Sun'])]
```

Out[68]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	356
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	447
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	601
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	467
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	483
...	...	...	...	...	...	...	...	...	...	
238	35.83	4.67	Female	No	Sat	Dinner	3	11.94	Kimberly Crane	
239	29.03	5.92	Male	No	Sat	Dinner	3	9.68	Michael Avila	529
240	27.18	2.00	Female	Yes	Sat	Dinner	2	13.59	Monica Sanders	350
241	22.67	2.00	Male	Yes	Sat	Dinner	2	11.34	Keith Wong	601
242	17.82	1.75	Male	No	Sat	Dinner	2	8.91	Dennis Dixon	

163 rows × 14 columns

In [64]:
```
def calculate_weekend_percentage(total_visitors, weekend_visitors):
    if total_visitors <= 0:
        return "Total number of visitors must be greater than zero."

    percentage = (weekend_visitors / total_visitors) * 100
    return f"The percentage of people who visited on weekends is: {percenta
```

In [69]:
```
len(df.loc[df['day'].isin(['Sat','Sun'])])/len(df)
```

Out[69]:
```
0.6680327868852459
```

In [70]:
```
df.corr()
```

Out[70]:

	total_bill	tip	size	price_per_person	CC Number	GST
total_bill	1.000000	0.675734	0.598315	0.647554	0.104576	1.000000
tip	0.675734	1.000000	0.489299	0.347405	0.110857	0.675734
size	0.598315	0.489299	1.000000	-0.175359	-0.030239	0.598315
price_per_person	0.647554	0.347405	-0.175359	1.000000	0.135240	0.647554
CC Number	0.104576	0.110857	-0.030239	0.135240	1.000000	0.104576
GST	1.000000	0.675734	0.598315	0.647554	0.104576	1.000000

In [71]:
```
df[['total_bill','tip']].corr()
```

Out[71]:

	total_bill	tip
total_bill	1.000000	0.675734
tip	0.675734	1.000000

In [72]:
```
df.head()
```

Out[72]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35603
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	44780
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	46761
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	48327

In [73]:
```
df['total_bill'].max()
```

Out[73]:
```
50.81
```

In [77]:
```
df['total_bill'].idxmax()
```

Out[77]:
```
170
```

In [75]:
```
df.iloc[170]
```

Out[75]:
```
total_bill                    50.81
tip                            10.0
Gender                         Male
smoker                          Yes
day                             Sat
time                         Dinner
size                              3
price_per_person              16.94
Payer Name           Gregory Clark
CC Number       5473850968388236
Payment ID                Sat1954
last_four                    8236
Nature_Bill       Very Expensive
GST                          5.081
Name: 170, dtype: object
```

In [76]:
```
df['total_bill'].min()
```

Out[76]:
```
3.07
```

In [78]:
```
df['total_bill'].idxmin()
```

Out[78]:
```
67
```

In [79]:
```
df.iloc[67]
```

Out[79]:
```
total_bill                      3.07
tip                              1.0
Gender                        Female
smoker                           Yes
day                              Sat
time                          Dinner
size                               1
price_per_person                3.07
Payer Name             Tiffany Brock
CC Number          4359488526995267
Payment ID                  Sat3455
last_four                      5267
Nature_Bill                   cheap
GST                           0.307
Name: 67, dtype: object
```

In [80]:
```
df.head()
```

Out[80]:

	total_bill	tip	Gender	smoker	day	time	size	price_per_person	Payer Name	
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	35603
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	44780
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	60118
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	46761
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	48327

In [83]:
```
df['Gender'].value_counts()
```

Out[83]:
```
Male      157
Female     87
Name: Gender, dtype: int64
```

In [84]:
```
df['smoker'].value_counts()
```

Out[84]:
```
No     151
Yes     93
Name: smoker, dtype: int64
```

In [86]:
```
len(df['smoker'].value_counts())/len(df)*100
```

Out[86]:
```
0.819672131147541
```

In [90]:
```
df['Gender'].value_counts()
```

Out[90]:
```
Male      157
Female     87
Name: Gender, dtype: int64
```

In [92]:
```
len(df[(df['smoker'] == 'Yes') & (df['Gender'] == 'Female')])
```

Out[92]:
```
33
```

In [93]:
```
c=len(df[df['smoker']=='Yes'])/244*100
print('smoker perecentage:',round(c,3),'%')
```

```
smoker perecentage: 38.115 %
```

In [ ]:
```

```