In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from pandas import Series, DataFrame
```

In [3]:
```
df1 = sns.load_dataset('tips')
```

In [4]:
```
df1.head()
```

Out[4]:

	total_bill	tip	sex	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

EDA on Tips data - Univariate Analysis.

Numerical Data

1. Histogram
2. KDE plot
3. Distplot
4. Boxplt
5. Violinplot

Categorical Data

1. Bar Graph
2. Pie Chart

In [5]:
```
df1.head()
```

Out[5]:

	total_bill	tip	sex	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

In [6]:

```
df1.shape
```

Out[6]:

```
(244, 7)
```

In [7]:

```
df1.info() #we can get the info about numerical and categorical
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype   
---  ------      --------------  -----   
 0   total_bill  244 non-null    float64 
 1   tip         244 non-null    float64 
 2   sex         244 non-null    category
 3   smoker      244 non-null    category
 4   day         244 non-null    category
 5   time        244 non-null    category
 6   size        244 non-null    int64   
dtypes: category(4), float64(2), int64(1)
memory usage: 7.3 KB
```

There are 2 numerical and 4 Categorical Columns.

In [8]:

```
df1.nunique()
```

Out[8]:

```
total_bill    229
tip           123
sex             2
smoker          2
day             4
time            2
size            6
dtype: int64
```

In [9]:

```
df1['day'].unique()
```

Out[9]:

```
['Sun', 'Sat', 'Thur', 'Fri']
Categories (4, object): ['Sun', 'Sat', 'Thur', 'Fri']
```

In [10]:

```
df1['day'].value_counts()
```

Out[10]:

```
Sat     87
Sun     76
Thur    62
Fri     19
Name: day, dtype: int64
```

In [11]:

```
df1['day'].value_counts(normalize=True)*100
```

Out[11]:

```
Sat     35.655738
Sun     31.147541
Thur    25.409836
Fri      7.786885
Name: day, dtype: float64
```

In [12]:

```
df1.head()
```

Out[12]:

	total_bill	tip	sex	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

In [13]:

```
mean = df1['total_bill'].mean()
print('Averge bill paid is',mean)
```

```
Averge bill paid is 19.785942622950824
```

In [14]:

```
median = df1['total_bill'].median()
print('Median bill paid is',median)
```

```
Median bill paid is 17.795
```

Mean value is greater than Median, hence Right Skewed!

In [15]:

```
print('Skewness of Total Bill Column is',df1['total_bill'].skew())
```

```
Skewness of Total Bill Column is 1.1332130376158205
```

Value of skewness is Positive, hence Right Skewed!

In [17]:

```
mode = df1['total_bill'].mode()
mode
```

Out[17]:

```
0    13.42
dtype: float64
```

In [20]:

```
print('Mode of Total Bill column is',mode[0])
```

```
Mode of Total Bill column is 13.42
```

In [21]:
```
var = df1['total_bill'].var()
std = df1['total_bill'].std()
mad = df1['total_bill'].mad()
```

In [22]:
```
print('Variance of Total Bill column is',var)
print('Standard Deviation of Total Bill column is',std)
print('Mean Absolute Deviation of Total Bill column is',mad)
```

```
Variance of Total Bill column is 79.25293861397826
Standard Deviation of Total Bill column is 8.902411954856856
Mean Absolute Deviation of Total Bill column is 6.869440002687455
```

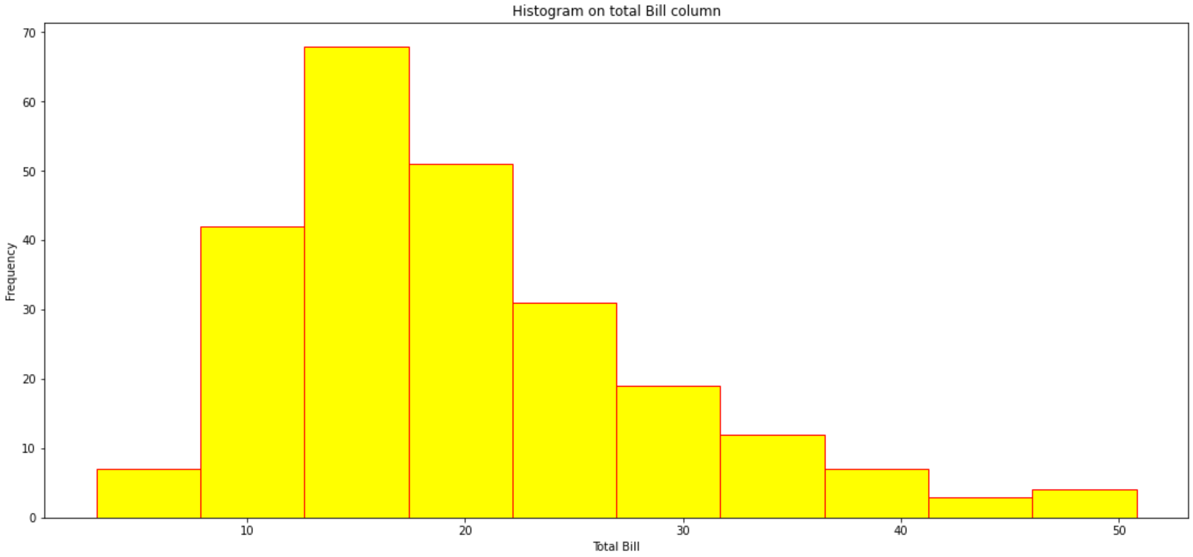
In [23]:
```
print('Kurtosis of Total Bill column',df1['total_bill'].kurt())
```

```
Kurtosis of Total Bill column 1.2184840156638854
```

Positive value suggest's that the total bill column is Leptokurtic (Few Outliers)

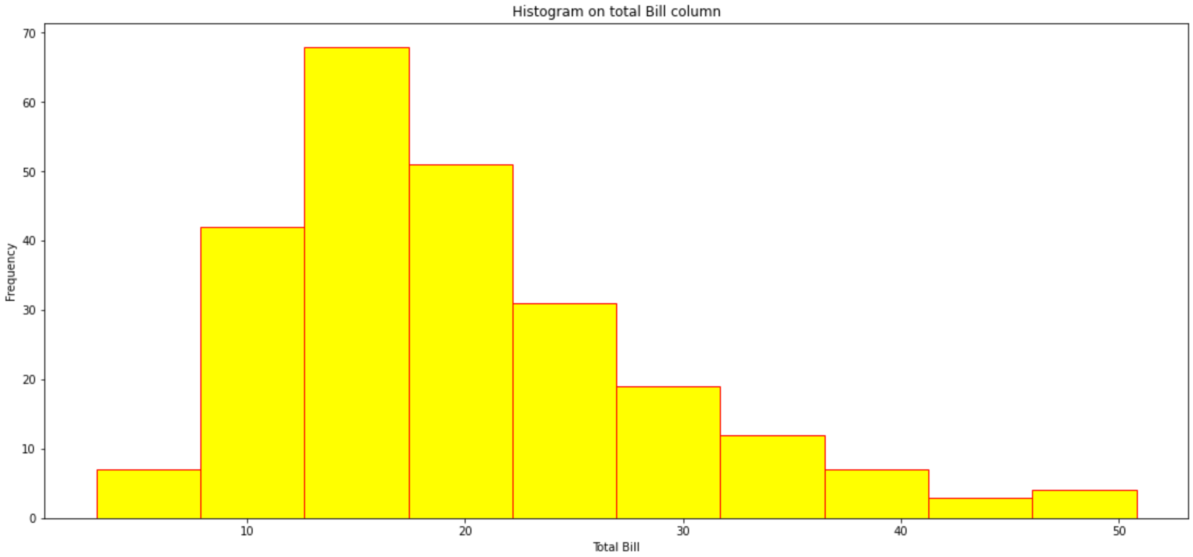
Histogram

In [30]:
```
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='hist',color='yellow',edgecolor='red')
plt.xlabel('Total Bill')
plt.title('Histogram on total Bill column')
plt.show()
```



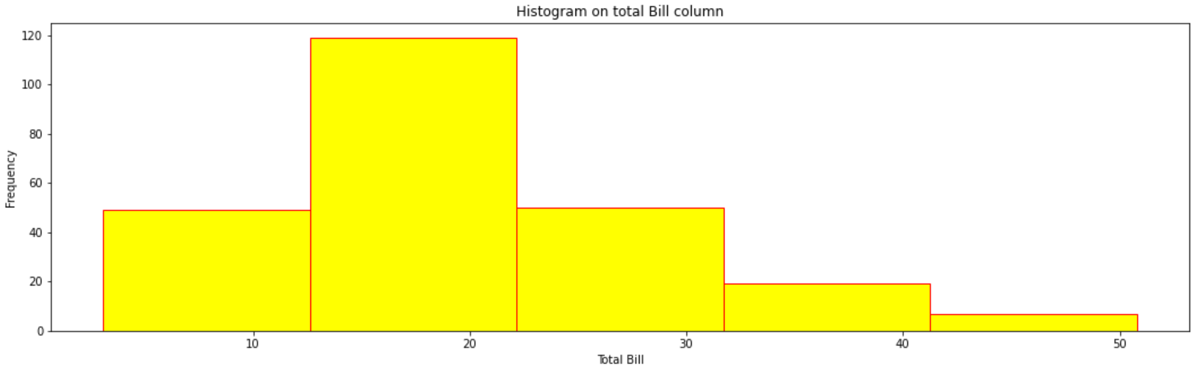
In [31]:

```python
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='hist',color='yellow',edgecolor='red',bins=10)
plt.xlabel('Total Bill')
plt.title('Histogram on total Bill column')
plt.show()
```



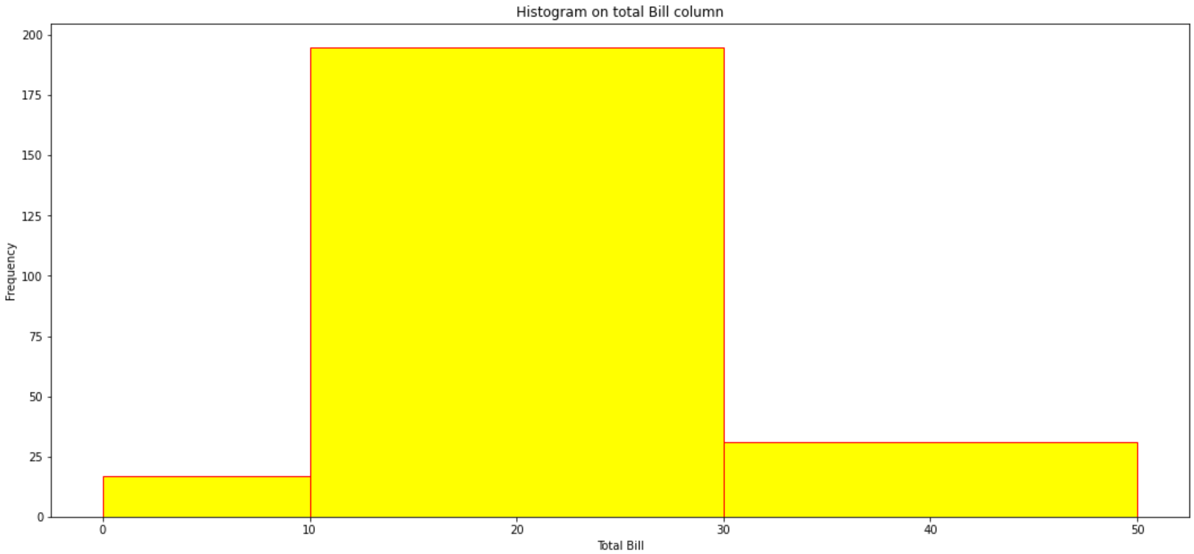
In [33]:

```python
plt.figure(figsize=(18,5))
df1['total_bill'].plot(kind='hist',color='yellow',edgecolor='red',bins=5)
plt.xlabel('Total Bill')
plt.title('Histogram on total Bill column')
plt.show()
```



In [35]:

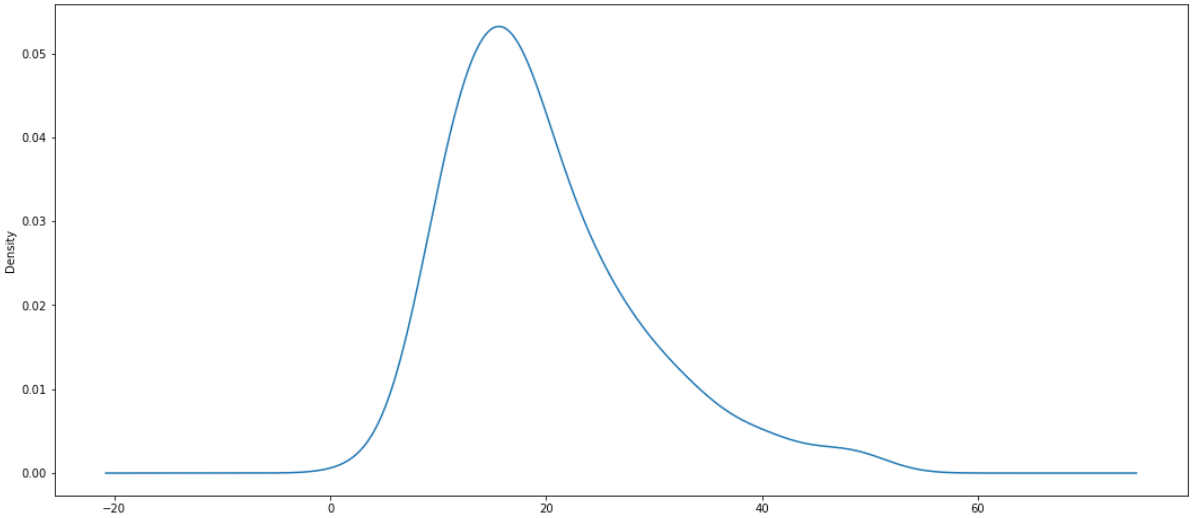
```
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='hist',color='yellow',edgecolor='red',bins=[0,10,30,50])
plt.xlabel('Total Bill')
plt.title('Histogram on total Bill column')
plt.show()
```



KDEplot or Density plot

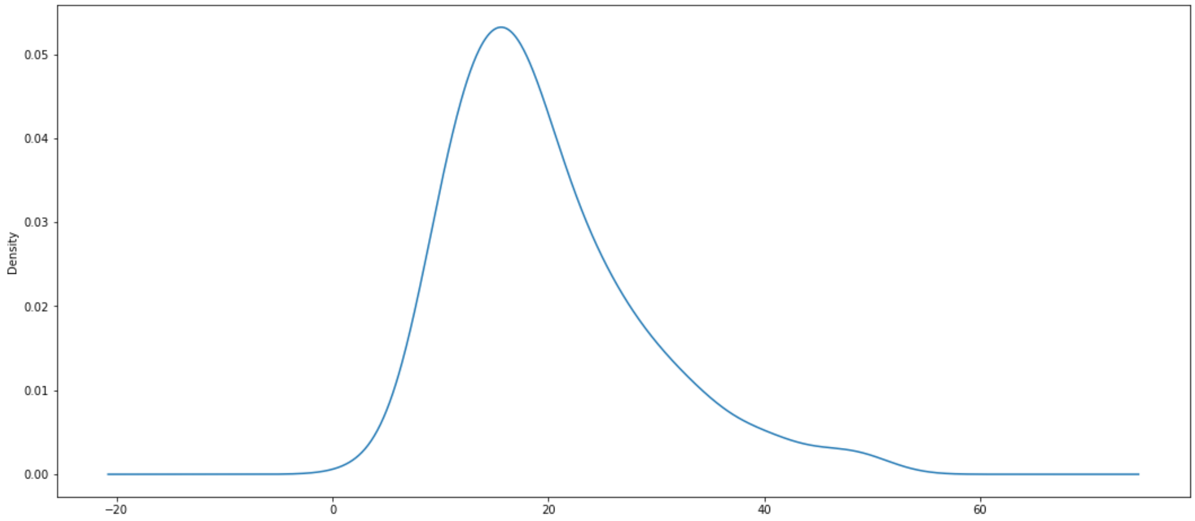
In [36]:

```
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='density')
plt.show()
```



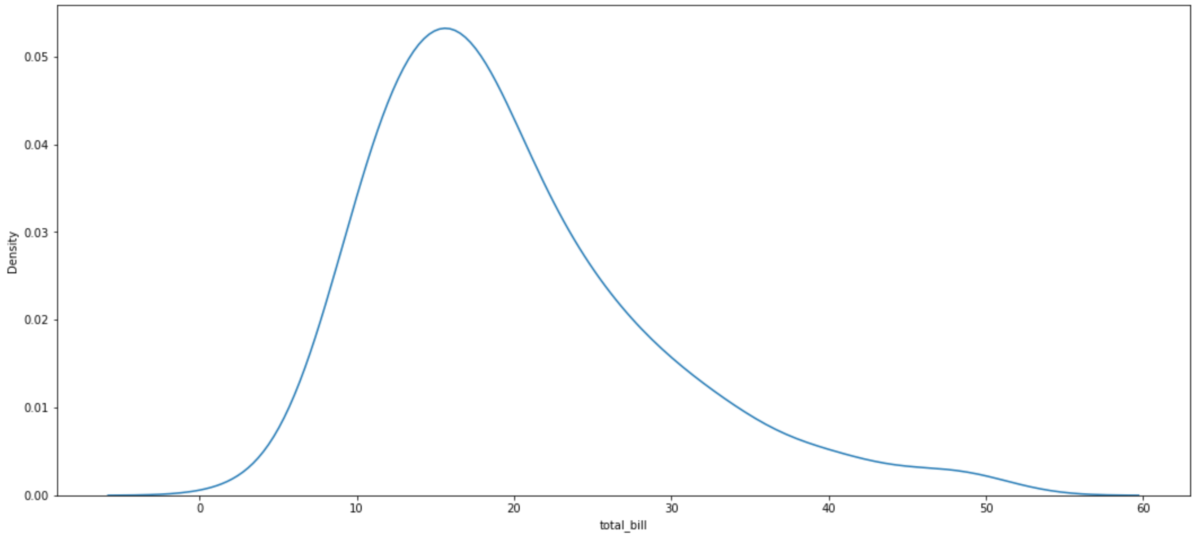
In [37]:

```
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='kde')
plt.show()
```



In [39]:

```
plt.figure(figsize=(18,8))
sns.kdeplot(df1['total_bill'])
plt.show()
```

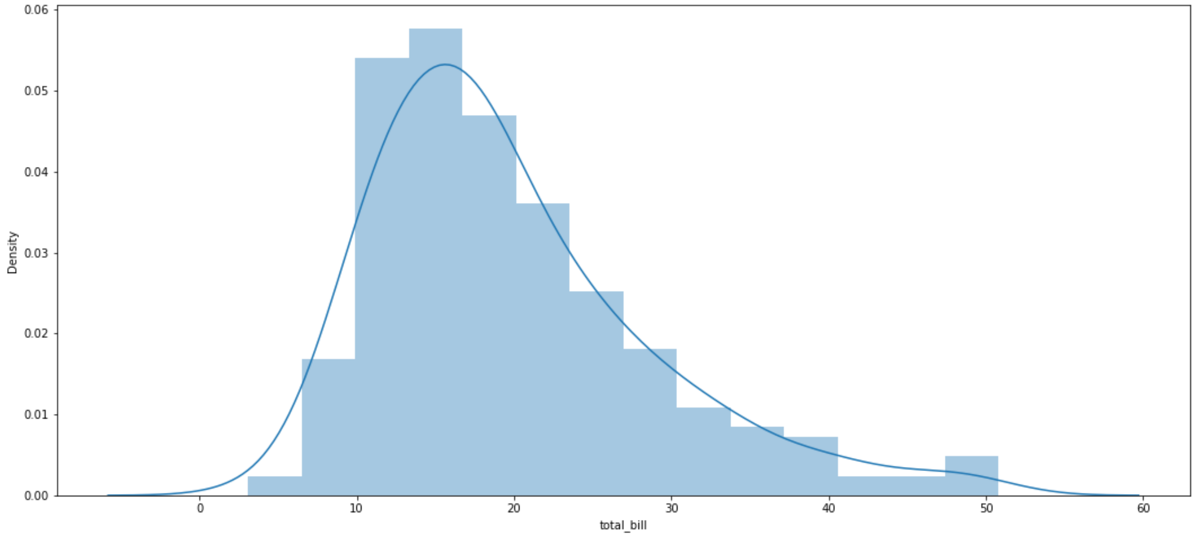


Distplot - (Hist + Density)

In [41]:

```
plt.figure(figsize=(18,8))
sns.distplot(df1['total_bill'])
plt.show()
```

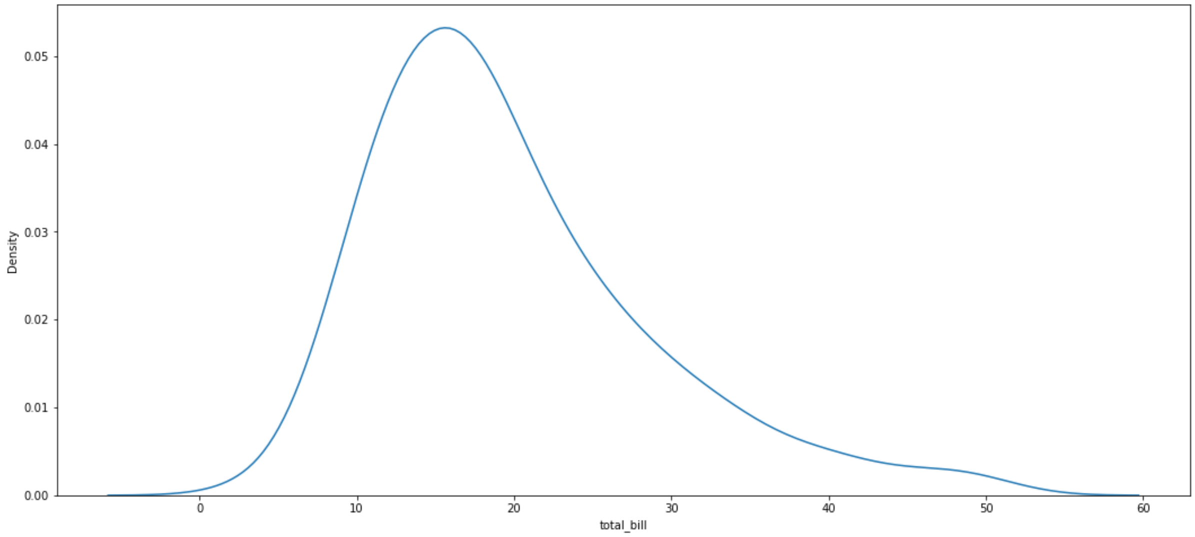
```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



In [42]:

```
plt.figure(figsize=(18,8))
sns.distplot(df1['total_bill'],hist=False)
plt.show()
```

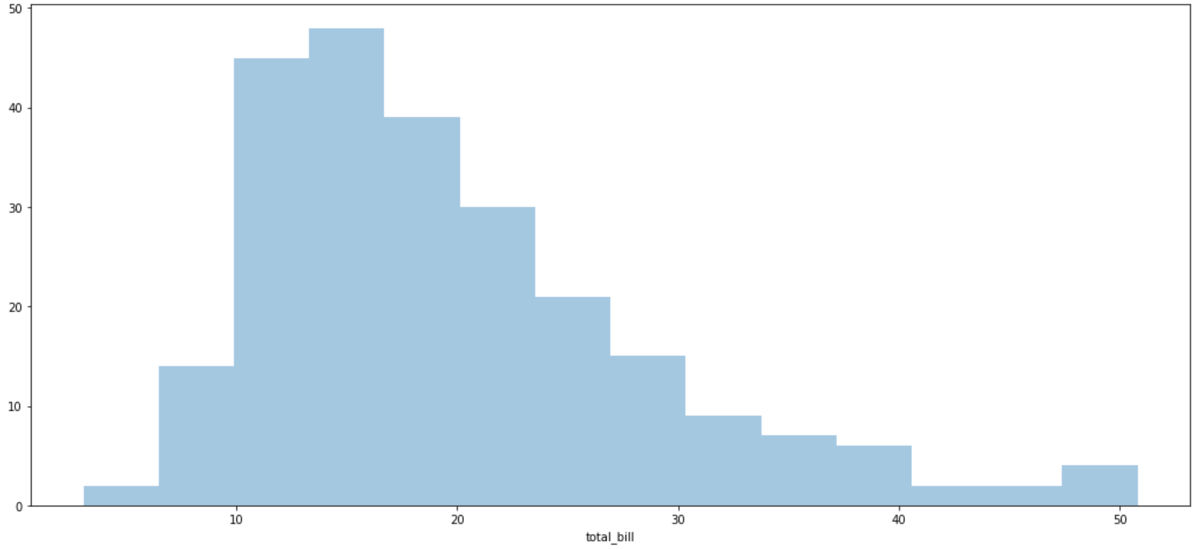
```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `kdeplot` (an axe
s-level function for kernel density plots).
  warnings.warn(msg, FutureWarning)
```



In [43]:

```
plt.figure(figsize=(18,8))
sns.distplot(df1['total_bill'],kde=False)
plt.show()
```

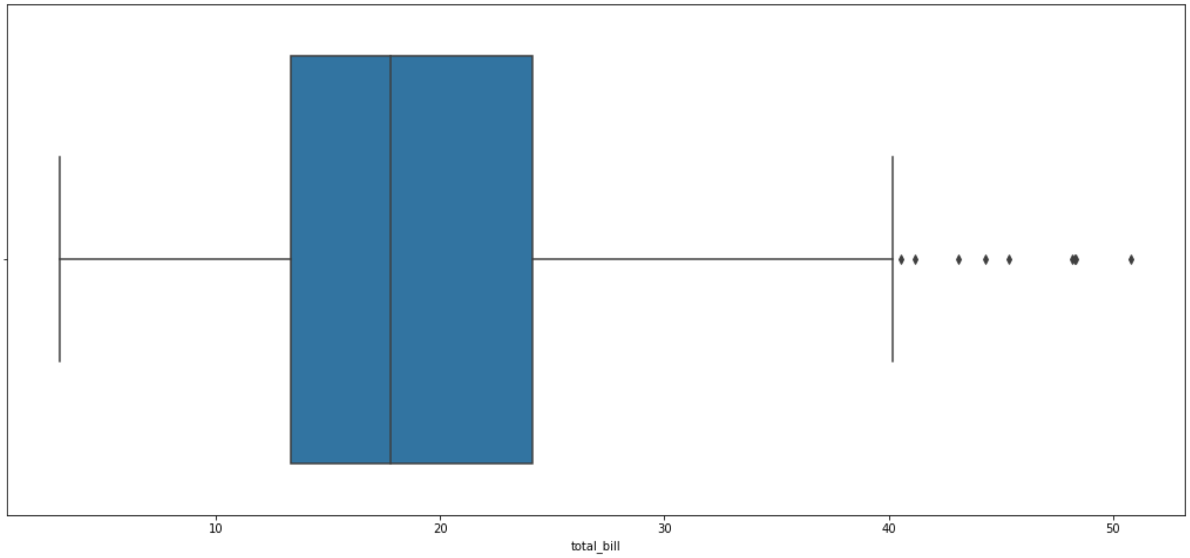
```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



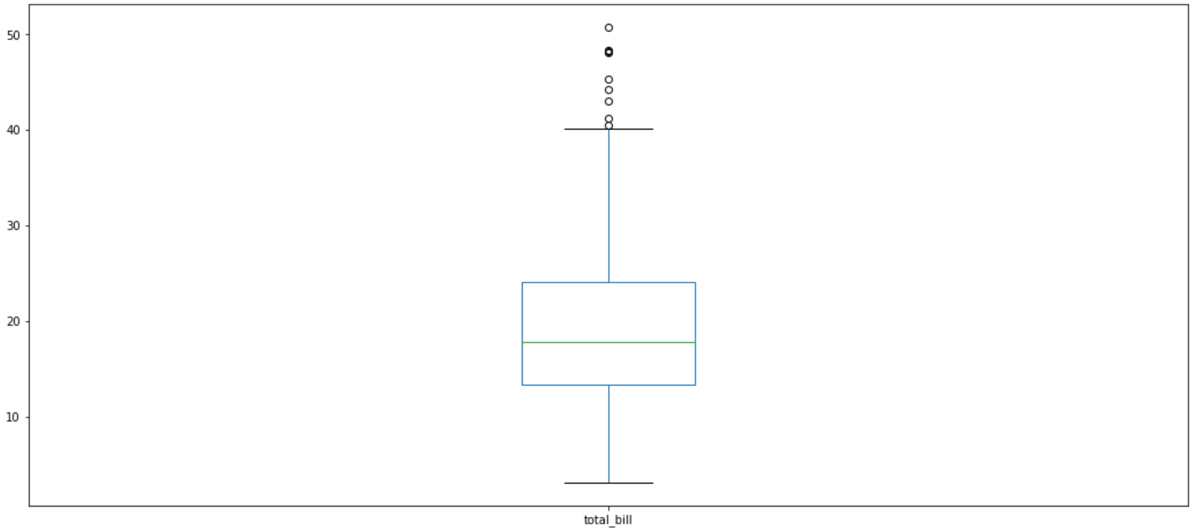
Boxplot

In [45]:

```
plt.figure(figsize=(18,8))
sns.boxplot(x='total_bill',data=df1)
plt.show()
```

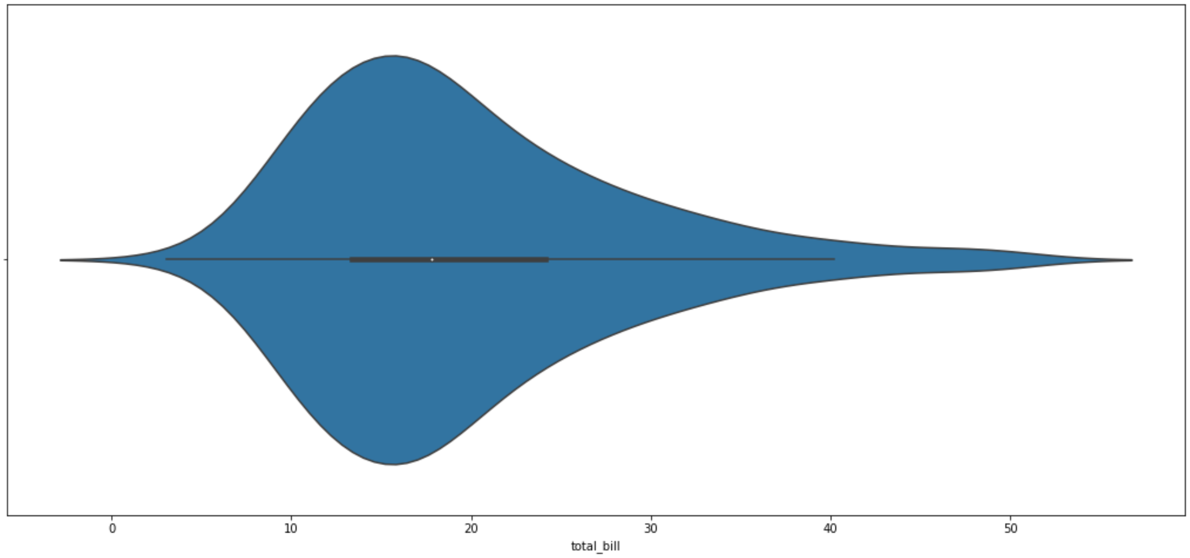


In [46]:
```
plt.figure(figsize=(18,8))
df1['total_bill'].plot(kind='box')
plt.show()
```



Violinplot

In [47]:
```
plt.figure(figsize=(18,8))
sns.violinplot(data=df1,x='total_bill')
plt.show()
```

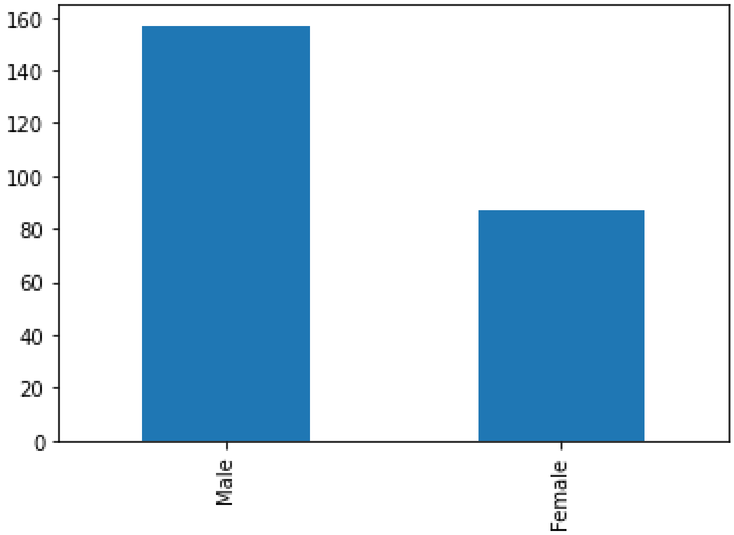


Bar Graph

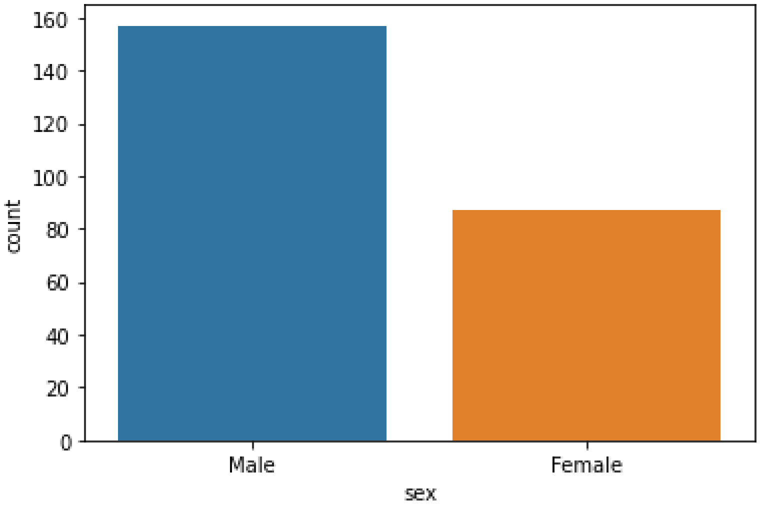
In [48]:
```
df1['sex'].value_counts()
```

Out[48]:
```
Male      157
Female     87
Name: sex, dtype: int64
```

In [49]:
```
df1['sex'].value_counts().plot(kind='bar')
plt.show()
```



In [50]:
```
sns.countplot(x='sex',data=df1)
plt.show()
```



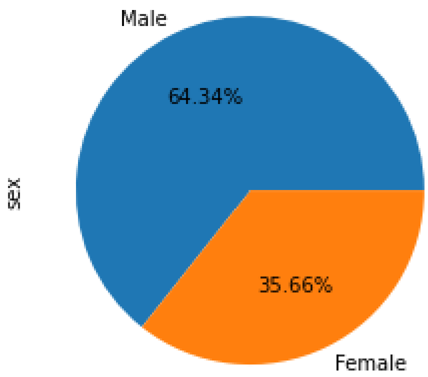
Pie Chart

In [52]:
```
df1['sex'].value_counts()
```

Out[52]:
```
Male      157
Female     87
Name: sex, dtype: int64
```

In [61]:

```
df1['sex'].value_counts().plot(kind='pie',autopct='%1.2f%%')
plt.show()
```



In [64]:

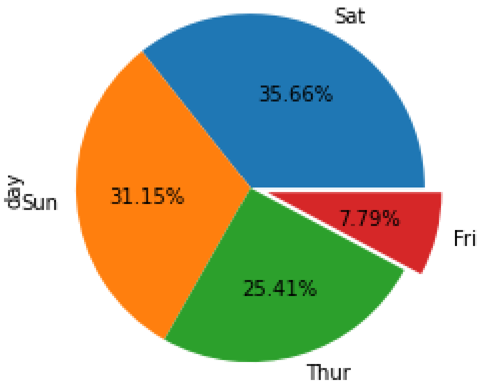
```
df1['day'].value_counts()
```

Out[64]:

```
Sat     87
Sun     76
Thur    62
Fri     19
Name: day, dtype: int64
```

In [65]:

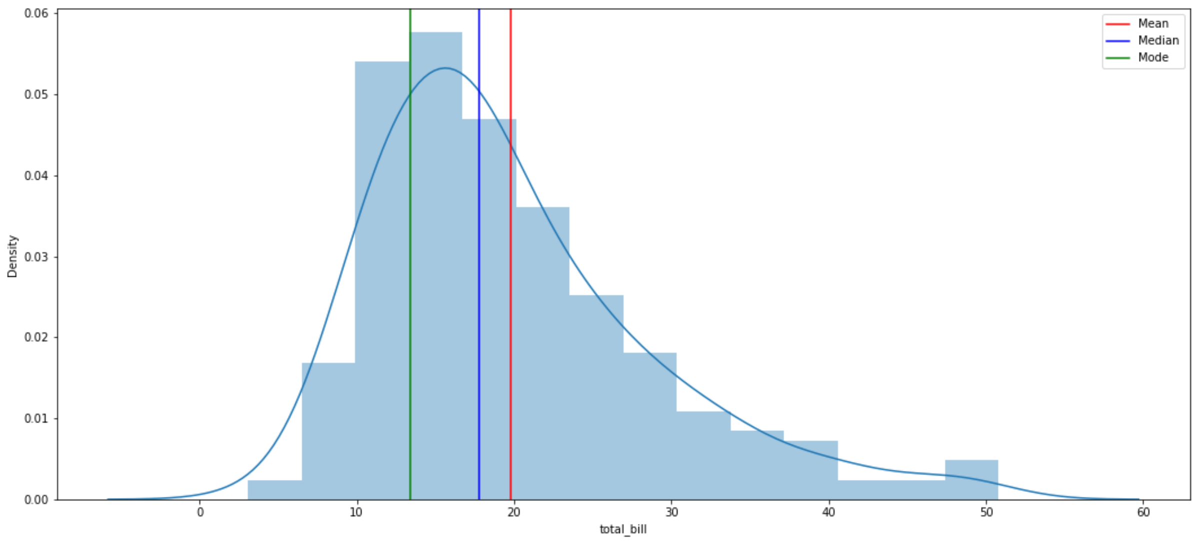
```
df1['day'].value_counts().plot(kind='pie',autopct='%1.2f%%',explode=[0,0,0,0.1
])
plt.show()
```



In [69]:

```
plt.figure(figsize=(18,8))
sns.distplot(df1['total_bill'])
plt.axvline(mean,label='Mean',color='red')
plt.axvline(median,label='Median',color='blue')
plt.axvline(mode[0],label='Mode',color='green')
plt.legend()
plt.show()
```

```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



Scaling the numerical Data.

1. Zscore Scaling
2. Min Max Scaling

In [71]:

```
df1.head()
```

Out[71]:

	total_bill	tip	sex	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

In [72]:

```
from scipy.stats import zscore
```

In [73]:

```
df1['ZTB'] = zscore(df1['total_bill'])
```

In [74]:

```
df1['MMTB'] = (df1['total_bill'] - df1['total_bill'].min()) / (df1['total_bil
l'].max() - df1['total_bill'].min())
```

In [75]:

```
df1.head()
```

Out[75]:

	total_bill	tip	sex	smoker	day	time	size	ZTB	MMTB
0	16.99	1.01	Female	No	Sun	Dinner	2	-0.314711	0.291579
1	10.34	1.66	Male	No	Sun	Dinner	3	-1.063235	0.152283
2	21.01	3.50	Male	No	Sun	Dinner	3	0.137780	0.375786
3	23.68	3.31	Male	No	Sun	Dinner	2	0.438315	0.431713
4	24.59	3.61	Female	No	Sun	Dinner	4	0.540745	0.450775

In [76]:

```
df1['MMTB'].max()
```

Out[76]:

```
1.0
```

In [77]:

```
df1['MMTB'].min()
```

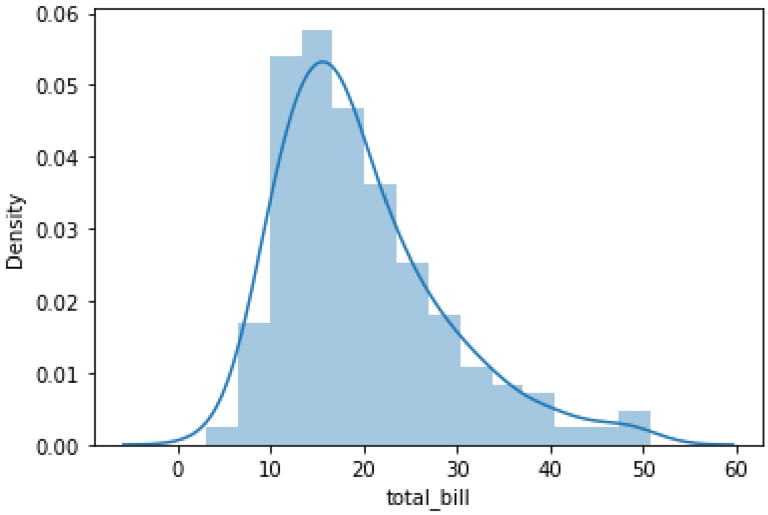
Out[77]:

```
0.0
```

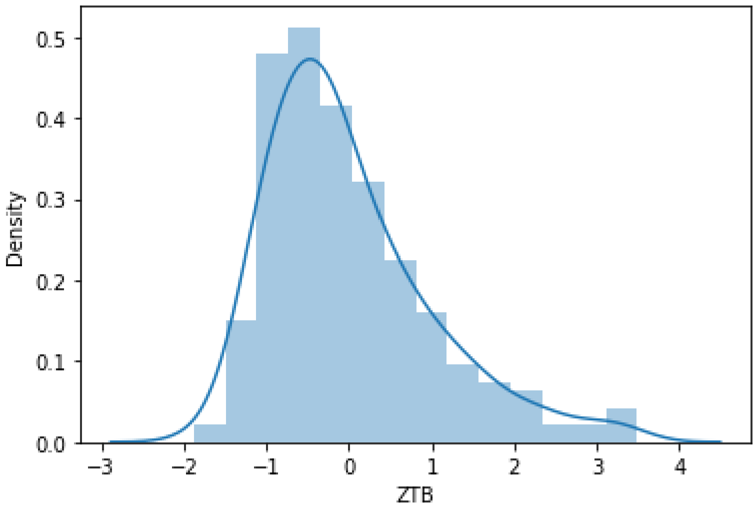
In [78]:

```
sns.distplot(df1['total_bill'])
plt.show()
sns.distplot(df1['ZTB'])
plt.show()
sns.distplot(df1['MMTB'])
plt.show()
```

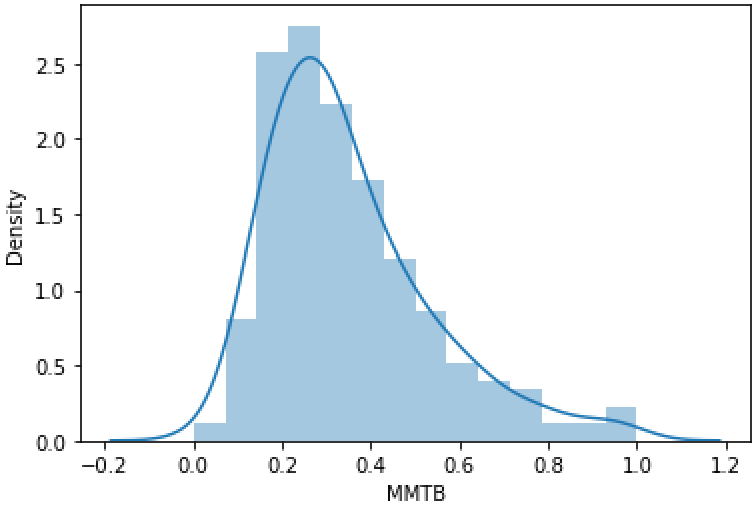
```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



```
/Users/aniruddhakalbande/opt/anaconda3/lib/python3.8/site-packages/seaborn/di
stributions.py:2551: FutureWarning: `distplot` is a deprecated function and w
ill be removed in a future version. Please adapt your code to use either `dis
plot` (a figure-level function with similar flexibility) or `histplot` (an ax
es-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



In [79]:

```
print('Skewness of Total Bill is',df1['total_bill'].skew())
print('Skewness of Zscore Scaled Total Bill is',df1['ZTB'].skew())
print('Skewness of Min Max Scaled Total Bill is',df1['MMTB'].skew())
```

```
Skewness of Total Bill is 1.1332130376158205
Skewness of Zscore Scaled Total Bill is 1.1332130376158205
Skewness of Min Max Scaled Total Bill is 1.1332130376158203
```

In [80]:

```
print('Kurtosis of Total Bill is',df1['total_bill'].kurt())
print('Kurtosis of Zscore Scaled Total Bill is',df1['ZTB'].kurt())
print('Kurtosis of Min Max Scaled Total Bill is',df1['MMTB'].kurt())
```

```
Kurtosis of Total Bill is 1.2184840156638854
Kurtosis of Zscore Scaled Total Bill is 1.2184840156638836
Kurtosis of Min Max Scaled Total Bill is 1.2184840156638836
```