Feature Scaling

1.Standardization

2.Nomalization Min-Max Scaler

3.Unit Vector

In [1]:
```
import seaborn as sns
```

In [2]:
```
df = sns.load_dataset('tips')
```

In [3]:
```
df.head()
```

Out[3]:

	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 [4]:
```
import numpy as np
mean=np.mean(df['total_bill'])
std=np.std(df['total_bill'])
print(mean,std)
```

```
19.78594262295082 8.88415057777113
```

In [5]:
```
normalized_data=[]
for i in list(df['total_bill']):
    z_score=(i-mean)/std
    normalized_data.append(z_score)
```

In [6]:
```
normalized_data
```

Out[6]:
```
[-0.31471130509049433,
 -1.0632353132988694,
 0.13777989987156147,
 0.4383151031672548,
 0.5407447042905061,
 0.6195367051545456,
 -1.2399545152367868,
 0.7985071071171497,
 -0.5342033074974615,
 -0.5634689078183904,
 -1.0711145133852735,
 1.7417599174609368,
 -0.49143050702841135,
 -0.1526249033130415,
 -0.5578409077566732,
 0.20193910057513614,
 -1.0643609133112129,
 -0.39350330595453376,
 -0.31696250511518115,
 0.09725829942719796,
 -0.21003050394255576,
 0.05673669898283485,
 -0.4520345065963916,
 2.2100095225958007,
 0.0038334984026941687,
 -0.22241210407833373,
 -0.7221785095588128,
 -0.7987193103981655,
 0.2154463007232573,
 -0.01530170180714419,
 -1.1521577142739996,
 -0.16162970341178867,
 -0.5319521074727744,
 0.10176069947657196,
 -0.22578890411536373,
 0.4810879036363047,
 -0.39125210592984694,
 -0.3214649051645547,
 -0.12335930299211235,
 1.2926455125359118,
 -0.4216433062631193,
 -0.26180810451035325,
 -0.6580193088552377,
 -1.1375249141135353,
 1.1947183114620339,
 -0.16838330348584946,
 0.27510310137745875,
 1.4198383139307182,
 0.9864823091785012,
 -0.196523303794435,
 -0.8156033105833168,
 -1.0688633133605867,
 1.6911079169054832,
 -1.1082593137926064,
 0.649927905487818,
 -0.03331130200463895,
 2.0512999208553775,
 0.7456039065370089,
 -0.9619313121879615,
 3.2061655335197288,
```

```
0.05673669898283485,
-0.6726521090157022,
-0.9866945124595169,
-0.16838330348584946,
-0.24717530434988888,
0.033099098723622905,
-0.375493705757039,
-1.881546522272537,
0.04998309890877447,
-0.5375801075344916,
-0.8741345112251747,
-0.30570650499174673,
0.7962559070924629,
0.6184111051422024,
-0.5690969078801074,
-1.0441001130890313,
-0.21003050394255576,
0.8345263075121392,
0.33475990203166023,
-0.28094330472019163,
-0.038939302066355735,
-0.35185610549782703,
-1.093626513632142,
1.4513551142763341,
-0.4283969063371797,
1.6933591169301696,
-0.7604489099784891,
-0.16950890349819264,
0.5542519044386274,
0.15466390005671263,
1.0337575096969247,
0.30436870169838753,
-1.5798857189645,
-0.39012650591750336,
0.33363430201931665,
2.2944295235215573,
0.8435311076108867,
-0.8730089112128313,
0.13777989987156147,
-0.824608110682064,
-0.949549712052184,
-0.4959329070777849,
2.75930232861939,
0.29648950161198395,
0.12764949976047069,
-0.4981841071024719,
0.0792486992297032,
0.6105319050557985,
-0.17401130354756664,
-0.6163721083985311,
-0.6512657087811772,
-1.4110457171129867,
2.0580535209294384,
0.46870630350052717,
0.6668119056729696,
-0.2786921046955048,
1.1418151108818932,
-1.0283417129162233,
-0.8279849107190944,
0.48333910366099153,
```

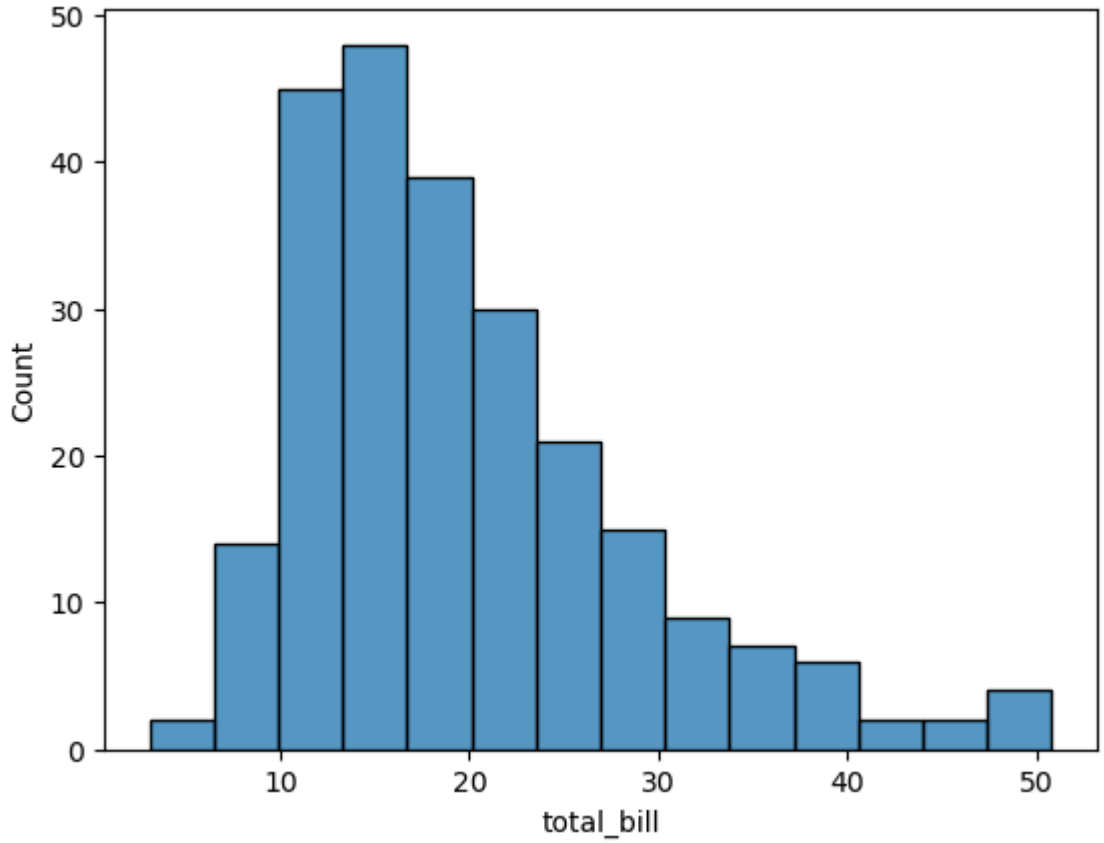
```
-0.9112793116325076,
-0.7165505094970956,
-0.6220001084602482,
-0.4317737063742101,
-0.8223569106573773,
1.127182310721429,
-1.2680945155453722,
-0.5927345081393193,
-0.9461729120151535,
0.3415135021057206,
-0.07946090251071924,
0.054485498958148054,
-0.9698105122743655,
-0.8471201109289326,
-0.17176010352287946,
-1.2692201155577156,
-1.0643609133112129,
-0.6343817085960258,
-0.4261457063124929,
-0.7458161098180246,
-0.26068250449801006,
1.6337023162759683,
2.409240724780586,
0.817642307326988,
-0.3777449057817258,
-1.2872297157552104,
-0.12898730305382955,
-0.8910185114103261,
-1.1262689139901012,
-1.3817801167920578,
-0.6433865086947732,
-0.7491929098550548,
-0.2843201047572216,
0.5362423042411326,
-0.001794501659023019,
1.1328103107831462,
3.1949095333962942,
0.5868943047965864,
-0.7199273095341259,
-0.3709913057076654,
0.19293430047638896,
-0.8020961104351957,
-0.40250810605328097,
-0.6726521090157022,
-0.25618010444863604,
0.5328655042041022,
0.10963989956297594,
1.3421719130790224,
-1.035095312990284,
-1.0305929129409102,
3.4920679366549576,
-0.44753210654701786,
-1.4110457171129867,
1.3579303132518303,
-0.3338465053003323,
1.4761183145478893,
-0.21340730397958615,
-0.5972369081886929,
-1.1465297142122828,
1.6708471166833017,
```

```
1.673098316707988,
0.3989191027352349,
2.8774903299154495,
0.38090950253774053,
2.337202323990607,
0.10176069947657196,
0.1253982997357835,
1.2014719115360948,
-0.18414170365865742,
0.37303030245133656,
-0.461039306695139,
0.002707898390350571,
0.9741007090427236,
-0.4846769069543508,
-0.36086090559657463,
-1.3761521167303408,
-1.0632353132988694,
2.625355927150523,
-0.7638257100155194,
-0.7064201093860049,
-0.12110810296742555,
-0.7930913103364483,
-0.7638257100155194,
-0.3811217058187562,
0.08375109927907719,
-0.3732425057323522,
0.7658647067591905,
2.132343121744104,
0.5047255038955166,
-0.7908401103117615,
1.1564479110423576,
0.687072705895151,
3.212919133593789,
-0.733434509682247,
0.9437095087094514,
-0.7750817101389536,
0.9414583086847641,
-0.9225353117559418,
-1.355891316508159,
1.1654527111411053,
-0.8583761110523668,
-0.7165505094970956,
-1.2613409154713116,
-0.4283969063371797,
-0.7165505094970956,
-0.39575450597922057,
-1.091375313607455,
0.07474629918032961,
-0.7323089096699036,
0.2627215012416812,
0.47545990357458795,
-0.461039306695139,
-0.920284111731255,
-1.0148345127681024,
-0.4790489068926338,
-1.093626513632142,
-0.8088497105092562,
1.4682391144614853,
1.8059191181645118,
1.0405111097709856,
```

```
0.8322751074874524,
0.32462950192056944,
-0.22128650406599013,
-0.11322890288102158]
```

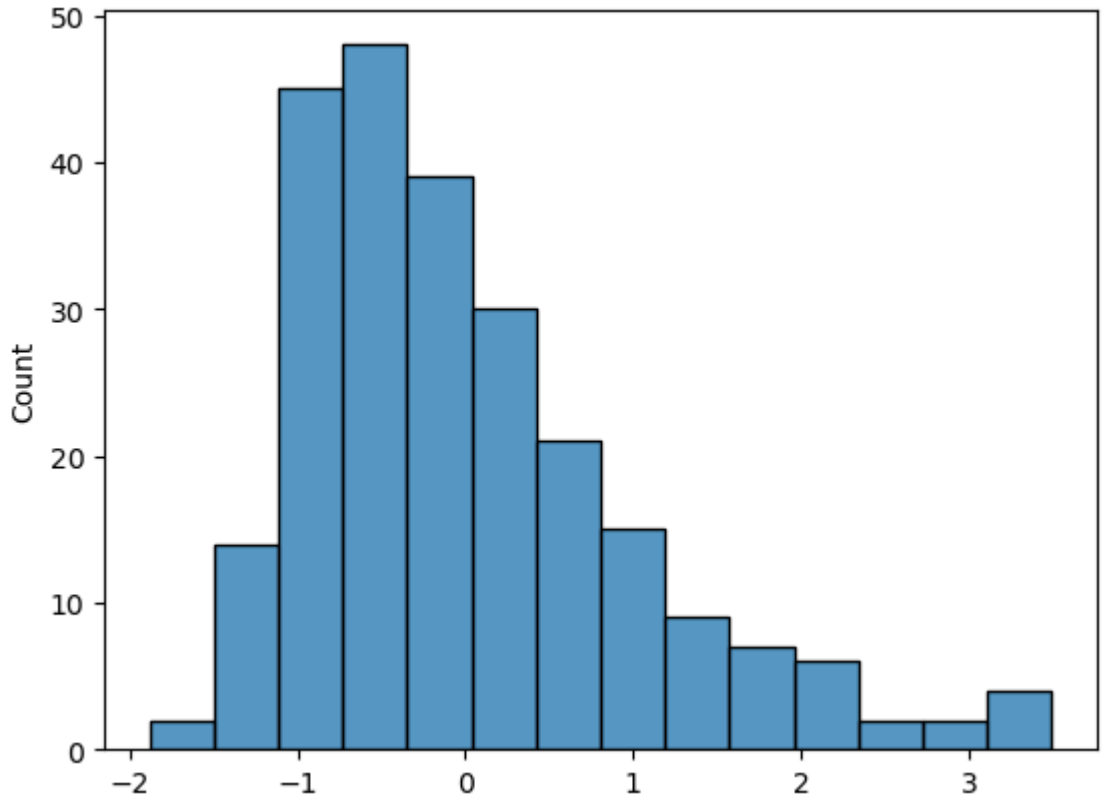
In [7]:
```
sns.histplot(df['total_bill'])
```

Out[7]:
```
<AxesSubplot: xlabel='total_bill', ylabel='Count'>
```



In [8]:
```
sns.histplot(normalized_data)
```

Out[8]:
```
<AxesSubplot: ylabel='Count'>
```



In [9]:
```
from sklearn.preprocessing import StandardScaler
```

In [10]:
```
scaler = StandardScaler()
```

In [13]:
```
scaler
```

Out[13]:
```
StandardScaler
StandardScaler()
```

In [14]:
```
scaler.fit(df[['total_bill']])
```

Out[14]:
```
StandardScaler
StandardScaler()
```

In [15]:
```
scaler.transform(df[['total_bill']])
```

Out[15]:
```
array([[-3.14711305e-01],
       [-1.06323531e+00],
       [ 1.37779900e-01],
       [ 4.38315103e-01],
       [ 5.40744704e-01],
       [ 6.19536705e-01],
       [-1.23995452e+00],
       [ 7.98507107e-01],
       [-5.34203307e-01],
       [-5.63468908e-01],
       [-1.07111451e+00],
       [ 1.74175992e+00],
       [-4.91430507e-01],
       [-1.52624903e-01],
       [-5.57840908e-01],
       [ 2.01939101e-01],
       [-1.06436091e+00],
       [-3.93503306e-01],
       [-3.16962505e-01],
       [ 9.72582994e-02],
       [-2.10030504e-01],
       [ 5.67366990e-02],
       [-4.52034507e-01],
       [ 2.21000952e+00],
       [ 3.83349840e-03],
       [-2.22412104e-01],
       [-7.22178510e-01],
       [-7.98719310e-01],
       [ 2.15446301e-01],
       [-1.53017018e-02],
       [-1.15215771e+00],
       [-1.61629703e-01],
       [-5.31952107e-01],
       [ 1.01760699e-01],
       [-2.25788904e-01],
       [ 4.81087904e-01],
       [-3.91252106e-01],
       [-3.21464905e-01],
       [-1.23359303e-01],
       [ 1.29264551e+00],
       [-4.21643306e-01],
       [-2.61808105e-01],
       [-6.58019309e-01],
       [-1.13752491e+00],
       [ 1.19471831e+00],
       [-1.68383303e-01],
       [ 2.75103101e-01],
       [ 1.41983831e+00],
       [ 9.86482309e-01],
       [-1.96523304e-01],
       [-8.15603311e-01],
       [-1.06886331e+00],
       [ 1.69110792e+00],
       [-1.10825931e+00],
       [ 6.49927905e-01],
       [-3.33113020e-02],
       [ 2.05129992e+00],
       [ 7.45603907e-01],
       [-9.61931312e-01],
       [ 3.20616553e+00],
```

```
[ 5.67366990e-02],
[-6.72652109e-01],
[-9.86694512e-01],
[-1.68383303e-01],
[-2.47175304e-01],
[ 3.30990987e-02],
[-3.75493706e-01],
[-1.88154652e+00],
[ 4.99830989e-02],
[-5.37580108e-01],
[-8.74134511e-01],
[-3.05706505e-01],
[ 7.96255907e-01],
[ 6.18411105e-01],
[-5.69096908e-01],
[-1.04410011e+00],
[-2.10030504e-01],
[ 8.34526308e-01],
[ 3.34759902e-01],
[-2.80943305e-01],
[-3.89393021e-02],
[-3.51856105e-01],
[-1.09362651e+00],
[ 1.45135511e+00],
[-4.28396906e-01],
[ 1.69335912e+00],
[-7.60448910e-01],
[-1.69508903e-01],
[ 5.54251904e-01],
[ 1.54663900e-01],
[ 1.03375751e+00],
[ 3.04368702e-01],
[-1.57988572e+00],
[-3.90126506e-01],
[ 3.33634302e-01],
[ 2.29442952e+00],
[ 8.43531108e-01],
[-8.73008911e-01],
[ 1.37779900e-01],
[-8.24608111e-01],
[-9.49549712e-01],
[-4.95932907e-01],
[ 2.75930233e+00],
[ 2.96489502e-01],
[ 1.27649500e-01],
[-4.98184107e-01],
[ 7.92486992e-02],
[ 6.10531905e-01],
[-1.74011304e-01],
[-6.16372108e-01],
[-6.51265709e-01],
[-1.41104572e+00],
[ 2.05805352e+00],
[ 4.68706304e-01],
[ 6.66811906e-01],
[-2.78692105e-01],
[ 1.14181511e+00],
[-1.02834171e+00],
[-8.27984911e-01],
[ 4.83339104e-01],
```

```
[-9.11279312e-01],
[-7.16550509e-01],
[-6.22000108e-01],
[-4.31773706e-01],
[-8.22356911e-01],
[ 1.12718231e+00],
[-1.26809452e+00],
[-5.92734508e-01],
[-9.46172912e-01],
[ 3.41513502e-01],
[-7.94609025e-02],
[ 5.44854990e-02],
[-9.69810512e-01],
[-8.47120111e-01],
[-1.71760104e-01],
[-1.26922012e+00],
[-1.06436091e+00],
[-6.34381709e-01],
[-4.26145706e-01],
[-7.45816110e-01],
[-2.60682504e-01],
[ 1.63370232e+00],
[ 2.40924072e+00],
[ 8.17642307e-01],
[-3.77744906e-01],
[-1.28722972e+00],
[-1.28987303e-01],
[-8.91018511e-01],
[-1.12626891e+00],
[-1.38178012e+00],
[-6.43386509e-01],
[-7.49192910e-01],
[-2.84320105e-01],
[ 5.36242304e-01],
[-1.79450166e-03],
[ 1.13281031e+00],
[ 3.19490953e+00],
[ 5.86894305e-01],
[-7.19927310e-01],
[-3.70991306e-01],
[ 1.92934300e-01],
[-8.02096110e-01],
[-4.02508106e-01],
[-6.72652109e-01],
[-2.56180104e-01],
[ 5.32865504e-01],
[ 1.09639900e-01],
[ 1.34217191e+00],
[-1.03509531e+00],
[-1.03059291e+00],
[ 3.49206794e+00],
[-4.47532107e-01],
[-1.41104572e+00],
[ 1.35793031e+00],
[-3.33846505e-01],
[ 1.47611831e+00],
[-2.13407304e-01],
[-5.97236908e-01],
[-1.14652971e+00],
[ 1.67084712e+00],
```

```
[ 1.67309832e+00],
[ 3.98919103e-01],
[ 2.87749033e+00],
[ 3.80909503e-01],
[ 2.33720232e+00],
[ 1.01760699e-01],
[ 1.25398300e-01],
[ 1.20147191e+00],
[-1.84141704e-01],
[ 3.73030302e-01],
[-4.61039307e-01],
[ 2.70789839e-03],
[ 9.74100709e-01],
[-4.84676907e-01],
[-3.60860906e-01],
[-1.37615212e+00],
[-1.06323531e+00],
[ 2.62535593e+00],
[-7.63825710e-01],
[-7.06420109e-01],
[-1.21108103e-01],
[-7.93091310e-01],
[-7.63825710e-01],
[-3.81121706e-01],
[ 8.37510993e-02],
[-3.73242506e-01],
[ 7.65864707e-01],
[ 2.13234312e+00],
[ 5.04725504e-01],
[-7.90840110e-01],
[ 1.15644791e+00],
[ 6.87072706e-01],
[ 3.21291913e+00],
[-7.33434510e-01],
[ 9.43709509e-01],
[-7.75081710e-01],
[ 9.41458309e-01],
[-9.22535312e-01],
[-1.35589132e+00],
[ 1.16545271e+00],
[-8.58376111e-01],
[-7.16550509e-01],
[-1.26134092e+00],
[-4.28396906e-01],
[-7.16550509e-01],
[-3.95754506e-01],
[-1.09137531e+00],
[ 7.47462992e-02],
[-7.32308910e-01],
[ 2.62721501e-01],
[ 4.75459904e-01],
[-4.61039307e-01],
[-9.20284112e-01],
[-1.01483451e+00],
[-4.79048907e-01],
[-1.09362651e+00],
[-8.08849711e-01],
[ 1.46823911e+00],
[ 1.80591912e+00],
[ 1.04051111e+00],
```

```
       [ 8.32275107e-01],
       [ 3.24629502e-01],
       [-2.21286504e-01],
       [-1.13228903e-01]])
```

In [16]:
```
scaler.fit_transform(df[['total_bill']])
```

Out[16]:
```
array([[-3.14711305e-01],
       [-1.06323531e+00],
       [ 1.37779900e-01],
       [ 4.38315103e-01],
       [ 5.40744704e-01],
       [ 6.19536705e-01],
       [-1.23995452e+00],
       [ 7.98507107e-01],
       [-5.34203307e-01],
       [-5.63468908e-01],
       [-1.07111451e+00],
       [ 1.74175992e+00],
       [-4.91430507e-01],
       [-1.52624903e-01],
       [-5.57840908e-01],
       [ 2.01939101e-01],
       [-1.06436091e+00],
       [-3.93503306e-01],
       [-3.16962505e-01],
       [ 9.72582994e-02],
       [-2.10030504e-01],
       [ 5.67366990e-02],
       [-4.52034507e-01],
       [ 2.21000952e+00],
       [ 3.83349840e-03],
       [-2.22412104e-01],
       [-7.22178510e-01],
       [-7.98719310e-01],
       [ 2.15446301e-01],
       [-1.53017018e-02],
       [-1.15215771e+00],
       [-1.61629703e-01],
       [-5.31952107e-01],
       [ 1.01760699e-01],
       [-2.25788904e-01],
       [ 4.81087904e-01],
       [-3.91252106e-01],
       [-3.21464905e-01],
       [-1.23359303e-01],
       [ 1.29264551e+00],
       [-4.21643306e-01],
       [-2.61808105e-01],
       [-6.58019309e-01],
       [-1.13752491e+00],
       [ 1.19471831e+00],
       [-1.68383303e-01],
       [ 2.75103101e-01],
       [ 1.41983831e+00],
       [ 9.86482309e-01],
       [-1.96523304e-01],
       [-8.15603311e-01],
       [-1.06886331e+00],
       [ 1.69110792e+00],
       [-1.10825931e+00],
       [ 6.49927905e-01],
       [-3.33113020e-02],
       [ 2.05129992e+00],
       [ 7.45603907e-01],
       [-9.61931312e-01],
       [ 3.20616553e+00],
```

```
[ 5.67366990e-02],
[-6.72652109e-01],
[-9.86694512e-01],
[-1.68383303e-01],
[-2.47175304e-01],
[ 3.30990987e-02],
[-3.75493706e-01],
[-1.88154652e+00],
[ 4.99830989e-02],
[-5.37580108e-01],
[-8.74134511e-01],
[-3.05706505e-01],
[ 7.96255907e-01],
[ 6.18411105e-01],
[-5.69096908e-01],
[-1.04410011e+00],
[-2.10030504e-01],
[ 8.34526308e-01],
[ 3.34759902e-01],
[-2.80943305e-01],
[-3.89393021e-02],
[-3.51856105e-01],
[-1.09362651e+00],
[ 1.45135511e+00],
[-4.28396906e-01],
[ 1.69335912e+00],
[-7.60448910e-01],
[-1.69508903e-01],
[ 5.54251904e-01],
[ 1.54663900e-01],
[ 1.03375751e+00],
[ 3.04368702e-01],
[-1.57988572e+00],
[-3.90126506e-01],
[ 3.33634302e-01],
[ 2.29442952e+00],
[ 8.43531108e-01],
[-8.73008911e-01],
[ 1.37779900e-01],
[-8.24608111e-01],
[-9.49549712e-01],
[-4.95932907e-01],
[ 2.75930233e+00],
[ 2.96489502e-01],
[ 1.27649500e-01],
[-4.98184107e-01],
[ 7.92486992e-02],
[ 6.10531905e-01],
[-1.74011304e-01],
[-6.16372108e-01],
[-6.51265709e-01],
[-1.41104572e+00],
[ 2.05805352e+00],
[ 4.68706304e-01],
[ 6.66811906e-01],
[-2.78692105e-01],
[ 1.14181511e+00],
[-1.02834171e+00],
[-8.27984911e-01],
[ 4.83339104e-01],
```

```
[-9.11279312e-01],
[-7.16550509e-01],
[-6.22000108e-01],
[-4.31773706e-01],
[-8.22356911e-01],
[ 1.12718231e+00],
[-1.26809452e+00],
[-5.92734508e-01],
[-9.46172912e-01],
[ 3.41513502e-01],
[-7.94609025e-02],
[ 5.44854990e-02],
[-9.69810512e-01],
[-8.47120111e-01],
[-1.71760104e-01],
[-1.26922012e+00],
[-1.06436091e+00],
[-6.34381709e-01],
[-4.26145706e-01],
[-7.45816110e-01],
[-2.60682504e-01],
[ 1.63370232e+00],
[ 2.40924072e+00],
[ 8.17642307e-01],
[-3.77744906e-01],
[-1.28722972e+00],
[-1.28987303e-01],
[-8.91018511e-01],
[-1.12626891e+00],
[-1.38178012e+00],
[-6.43386509e-01],
[-7.49192910e-01],
[-2.84320105e-01],
[ 5.36242304e-01],
[-1.79450166e-03],
[ 1.13281031e+00],
[ 3.19490953e+00],
[ 5.86894305e-01],
[-7.19927310e-01],
[-3.70991306e-01],
[ 1.92934300e-01],
[-8.02096110e-01],
[-4.02508106e-01],
[-6.72652109e-01],
[-2.56180104e-01],
[ 5.32865504e-01],
[ 1.09639900e-01],
[ 1.34217191e+00],
[-1.03509531e+00],
[-1.03059291e+00],
[ 3.49206794e+00],
[-4.47532107e-01],
[-1.41104572e+00],
[ 1.35793031e+00],
[-3.33846505e-01],
[ 1.47611831e+00],
[-2.13407304e-01],
[-5.97236908e-01],
[-1.14652971e+00],
[ 1.67084712e+00],
```

```
       [ 1.67309832e+00],
       [ 3.98919103e-01],
       [ 2.87749033e+00],
       [ 3.80909503e-01],
       [ 2.33720232e+00],
       [ 1.01760699e-01],
       [ 1.25398300e-01],
       [ 1.20147191e+00],
       [-1.84141704e-01],
       [ 3.73030302e-01],
       [-4.61039307e-01],
       [ 2.70789839e-03],
       [ 9.74100709e-01],
       [-4.84676907e-01],
       [-3.60860906e-01],
       [-1.37615212e+00],
       [-1.06323531e+00],
       [ 2.62535593e+00],
       [-7.63825710e-01],
       [-7.06420109e-01],
       [-1.21108103e-01],
       [-7.93091310e-01],
       [-7.63825710e-01],
       [-3.81121706e-01],
       [ 8.37510993e-02],
       [-3.73242506e-01],
       [ 7.65864707e-01],
       [ 2.13234312e+00],
       [ 5.04725504e-01],
       [-7.90840110e-01],
       [ 1.15644791e+00],
       [ 6.87072706e-01],
       [ 3.21291913e+00],
       [-7.33434510e-01],
       [ 9.43709509e-01],
       [-7.75081710e-01],
       [ 9.41458309e-01],
       [-9.22535312e-01],
       [-1.35589132e+00],
       [ 1.16545271e+00],
       [-8.58376111e-01],
       [-7.16550509e-01],
       [-1.26134092e+00],
       [-4.28396906e-01],
       [-7.16550509e-01],
       [-3.95754506e-01],
       [-1.09137531e+00],
       [ 7.47462992e-02],
       [-7.32308910e-01],
       [ 2.62721501e-01],
       [ 4.75459904e-01],
       [-4.61039307e-01],
       [-9.20284112e-01],
       [-1.01483451e+00],
       [-4.79048907e-01],
       [-1.09362651e+00],
       [-8.08849711e-01],
       [ 1.46823911e+00],
       [ 1.80591912e+00],
       [ 1.04051111e+00],
```

```
       [ 8.32275107e-01],
       [ 3.24629502e-01],
       [-2.21286504e-01],
       [-1.13228903e-01]])
```

In [17]:
```
df.head()
```

Out[17]:

	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 [18]:
```
import pandas as pd
pd.DataFrame(scaler.fit_transform(df[['total_bill','tip']]),columns=['total_bill
```

Out[18]:

	total_bill	tip
0	-0.314711	-1.439947
1	-1.063235	-0.969205
2	0.137780	0.363356
3	0.438315	0.225754
4	0.540745	0.443020
...	...	...
239	1.040511	2.115963
240	0.832275	-0.722971
241	0.324630	-0.722971
242	-0.221287	-0.904026
243	-0.113229	0.001247

244 rows × 2 columns

In [19]:
```
scaler.transform([[13,4]])
```

```
/opt/conda/lib/python3.10/site-packages/sklearn/base.py:409: UserWarning: X doe
s not have valid feature names, but StandardScaler was fitted with feature name
s
  warnings.warn(
```

Out[19]:
```
array([[-0.76382571,  0.72546447]])
```

Normalization--Min_max Scaler

In [20]:
```
df = sns.load_dataset('taxis')
df.head()
```

Out[20]:

	pickup	dropoff	passengers	distance	fare	tip	tolls	total	color	payment	pickup_zo
0	2019-03-23 20:21:09	2019-03-23 20:27:24	1	1.60	7.0	2.15	0.0	12.95	yellow	credit card	Lenox W
1	2019-03-04 16:11:55	2019-03-04 16:19:00	1	0.79	5.0	0.00	0.0	9.30	yellow	cash	Upper W Side Sou
2	2019-03-27 17:53:01	2019-03-27 18:00:25	1	1.37	7.5	2.36	0.0	14.16	yellow	credit card	Alphal C
3	2019-03-10 01:23:59	2019-03-10 01:49:51	1	7.70	27.0	6.15	0.0	36.95	yellow	credit card	Hudson
4	2019-03-30 13:27:42	2019-03-30 13:37:14	3	2.16	9.0	1.10	0.0	13.40	yellow	credit card	Midto E

In [21]:
```
from sklearn.preprocessing import MinMaxScaler
```

In [22]:
```
min_max = MinMaxScaler()
```

In [23]:
```
min_max.fit_transform(df[['distance','fare','tip']])
```

Out[23]:
```
array([[0.04359673, 0.04026846, 0.06475904],
       [0.02152589, 0.02684564, 0.        ],
       [0.0373297 , 0.04362416, 0.07108434],
       ...,
       [0.11280654, 0.10067114, 0.        ],
       [0.03051771, 0.03355705, 0.        ],
       [0.10490463, 0.09395973, 0.10120482]])
```

In [24]:
```
min_max.fit(df[['distance','fare','tip']])
```

Out[24]:
```
MinMaxScaler()
```

In [25]:
```
min_max.transform(df[['distance','fare','tip']])
```

Out[25]:
```
array([[0.04359673, 0.04026846, 0.06475904],
       [0.02152589, 0.02684564, 0.        ],
       [0.0373297 , 0.04362416, 0.07108434],
       ...,
       [0.11280654, 0.10067114, 0.        ],
       [0.03051771, 0.03355705, 0.        ],
       [0.10490463, 0.09395973, 0.10120482]])
```

In [26]:
```
min_max.transform([[1.60,7.0,2.15]])
```

```
/opt/conda/lib/python3.10/site-packages/sklearn/base.py:409: UserWarning: X doe
s not have valid feature names, but MinMaxScaler was fitted with feature names
  warnings.warn(
```

Out[26]:
```
array([[0.04359673, 0.04026846, 0.06475904]])
```

Unit Vector

In [27]:
```
from sklearn.preprocessing import normalize
```

In [28]:
```
normalize(df[['distance','fare','tip']])
```

Out[28]:
```
array([[0.21346145, 0.93389384, 0.28683882],
       [0.15606402, 0.98774694, 0.        ],
       [0.1716575 , 0.93973084, 0.29570197],
       ...,
       [0.25050016, 0.96811656, 0.        ],
       [0.18349711, 0.98302025, 0.        ],
       [0.2429556 , 0.94658025, 0.21203398]])
```

In [29]:
```
pd.DataFrame(normalize(df[['distance','fare','tip']]))
```

Out[29]:

	0	1	2
0	0.213461	0.933894	0.286839
1	0.156064	0.987747	0.000000
2	0.171657	0.939731	0.295702
3	0.267899	0.939386	0.213971
4	0.231742	0.965592	0.118017
...	...	...	...
6428	0.160133	0.960800	0.226322
6429	0.307453	0.951563	0.000000
6430	0.250500	0.968117	0.000000
6431	0.183497	0.983020	0.000000
6432	0.242956	0.946580	0.212034

6433 rows × 3 columns