Untitled11

February 6, 2024

```
[1]: import pandas as pd
```

```
[2]: sales = pd.read_csv(r"D:\Data\Dataset MeriSKILL\Sales Data.csv")
```

```
[3]: sales
```

```
[3]:         Unnamed: 0  Order ID                Product  Quantity Ordered  \
     0                0    295665      Macbook Pro Laptop                 1
     1                1    295666      LG Washing Machine                 1
     2                2    295667     USB-C Charging Cable                1
     3                3    295668          27in FHD Monitor              1
     4                4    295669     USB-C Charging Cable                1
     ...            ...       ...                     ...               ...
     185945       13617    222905     AAA Batteries (4-pack)             1
     185946       13618    222906          27in FHD Monitor              1
     185947       13619    222907     USB-C Charging Cable               1
     185948       13620    222908     USB-C Charging Cable               1
     185949       13621    222909     AAA Batteries (4-pack)             1

             Price Each          Order Date  \
     0           1700.00  2019-12-30 00:01:00
     1            600.00  2019-12-29 07:03:00
     2             11.95  2019-12-12 18:21:00
     3            149.99  2019-12-22 15:13:00
     4             11.95  2019-12-18 12:38:00
     ...            ...                  ...
     185945        2.99  2019-06-07 19:02:00
     185946      149.99  2019-06-01 19:29:00
     185947       11.95  2019-06-22 18:57:00
     185948       11.95  2019-06-26 18:35:00
     185949        2.99  2019-06-25 14:33:00

                               Purchase Address  Month    Sales  \
     0       136 Church St, New York City, NY 10001     12  1700.00
     1         562 2nd St, New York City, NY 10001     12   600.00
     2        277 Main St, New York City, NY 10001     12    11.95
     3        410 6th St, San Francisco, CA 94016     12   149.99
     4               43 Hill St, Atlanta, GA 30301     12    11.95
```

```
...                                          ...  ...     ...
185945            795 Pine St, Boston, MA 02215      6      2.99
185946    495 North St, New York City, NY 10001      6    149.99
185947    319 Ridge St, San Francisco, CA 94016      6     11.95
185948    916 Main St, San Francisco, CA 94016       6     11.95
185949            209 11th St, Atlanta, GA 30301      6      2.99

                  City  Hour
0         New York City     0
1         New York City     7
2         New York City    18
3         San Francisco    15
4               Atlanta    12
...                 ...   ...
185945           Boston    19
185946    New York City    19
185947    San Francisco    18
185948    San Francisco    18
185949          Atlanta    14

[185950 rows x 11 columns]
```

[4]:
```
sales = sales.drop(columns = ['Unnamed: 0' ,'Order ID'])
```

[5]:
```
sales
```

[5]:
```
                    Product  Quantity Ordered  Price Each  \
0         Macbook Pro Laptop                1     1700.00
1          LG Washing Machine               1      600.00
2         USB-C Charging Cable              1       11.95
3              27in FHD Monitor             1      149.99
4         USB-C Charging Cable              1       11.95
...                      ...              ...         ...
185945    AAA Batteries (4-pack)            1        2.99
185946         27in FHD Monitor             1      149.99
185947    USB-C Charging Cable              1       11.95
185948    USB-C Charging Cable              1       11.95
185949    AAA Batteries (4-pack)            1        2.99

                 Order Date                        Purchase Address  Month  \
0       2019-12-30 00:01:00  136 Church St, New York City, NY 10001     12
1       2019-12-29 07:03:00     562 2nd St, New York City, NY 10001     12
2       2019-12-12 18:21:00    277 Main St, New York City, NY 10001     12
3       2019-12-22 15:13:00     410 6th St, San Francisco, CA 94016     12
4       2019-12-18 12:38:00         43 Hill St, Atlanta, GA 30301      12
...                     ...                                     ...    ...
185945  2019-06-07 19:02:00           795 Pine St, Boston, MA 02215      6
```

```
185946   2019-06-01 19:29:00    495 North St, New York City, NY 10001       6
185947   2019-06-22 18:57:00    319 Ridge St, San Francisco, CA 94016       6
185948   2019-06-26 18:35:00    916 Main St, San Francisco, CA 94016        6
185949   2019-06-25 14:33:00        209 11th St, Atlanta, GA 30301          6

            Sales           City  Hour
0          1700.00   New York City     0
1           600.00   New York City     7
2            11.95   New York City    18
3           149.99   San Francisco    15
4            11.95         Atlanta    12
...            ...             ...    ...
185945        2.99          Boston    19
185946      149.99   New York City    19
185947       11.95   San Francisco    18
185948       11.95   San Francisco    18
185949        2.99         Atlanta    14

[185950 rows x 9 columns]
```

[6]: `sales.describe()`

[6]:
```
       Quantity Ordered     Price Each          Month           Sales  \
count     185950.000000  185950.000000  185950.000000  185950.000000
mean           1.124383     184.399735       7.059140     185.490917
std            0.442793     332.731330       3.502996     332.919771
min            1.000000       2.990000       1.000000       2.990000
25%            1.000000      11.950000       4.000000      11.950000
50%            1.000000      14.950000       7.000000      14.950000
75%            1.000000     150.000000      10.000000     150.000000
max            9.000000    1700.000000      12.000000    3400.000000

                Hour
count  185950.000000
mean       14.413305
std         5.423416
min         0.000000
25%        11.000000
50%        15.000000
75%        19.000000
max        23.000000
```

[7]: `sales.values`

[7]:
```
array([['Macbook Pro Laptop', 1, 1700.0, …, 1700.0, ' New York City',
        0],
       ['LG Washing Machine', 1, 600.0, …, 600.0, ' New York City', 7],
```

```
          ['USB-C Charging Cable', 1, 11.95, …, 11.95, ' New York City',
           18],
          …,
          ['USB-C Charging Cable', 1, 11.95, …, 11.95, ' San Francisco',
           18],
          ['USB-C Charging Cable', 1, 11.95, …, 11.95, ' San Francisco',
           18],
          ['AAA Batteries (4-pack)', 1, 2.99, …, 2.99, ' Atlanta', 14]],
         dtype=object)
```

[8]: `sales.dtypes`

```
[8]: Product             object
     Quantity Ordered     int64
     Price Each         float64
     Order Date          object
     Purchase Address    object
     Month                int64
     Sales              float64
     City                object
     Hour                 int64
     dtype: object
```

[9]: ```python
import datetime as dt
```

[10]: ```python
sales['Order Date'] = pd.to_datetime(sales['Order Date'], format ='%Y-%m-%d %H:
  ↪%M:%S')
```

[11]: `sales.dtypes`

```
[11]: Product                   object
      Quantity Ordered           int64
      Price Each               float64
      Order Date        datetime64[ns]
      Purchase Address          object
      Month                      int64
      Sales                    float64
      City                      object
      Hour                       int64
      dtype: object
```

[12]: `sales`

```
[12]:                     Product  Quantity Ordered  Price Each  \
      0        Macbook Pro Laptop                 1     1700.00
      1         LG Washing Machine                1      600.00
      2        USB-C Charging Cable               1       11.95
```

```
3                27in FHD Monitor                  1      149.99
4            USB-C Charging Cable                  1       11.95
...                         ...                   ...       ...
185945  AAA Batteries (4-pack)                     1        2.99
185946          27in FHD Monitor                   1      149.99
185947     USB-C Charging Cable                    1       11.95
185948     USB-C Charging Cable                    1       11.95
185949  AAA Batteries (4-pack)                     1        2.99

                Order Date                         Purchase Address  Month  \
0      2019-12-30 00:01:00   136 Church St, New York City, NY 10001     12
1      2019-12-29 07:03:00     562 2nd St, New York City, NY 10001      12
2      2019-12-12 18:21:00    277 Main St, New York City, NY 10001      12
3      2019-12-22 15:13:00     410 6th St, San Francisco, CA 94016      12
4      2019-12-18 12:38:00           43 Hill St, Atlanta, GA 30301      12
...                    ...                                    ...      ...
185945 2019-06-07 19:02:00          795 Pine St, Boston, MA 02215       6
185946 2019-06-01 19:29:00   495 North St, New York City, NY 10001      6
185947 2019-06-22 18:57:00   319 Ridge St, San Francisco, CA 94016      6
185948 2019-06-26 18:35:00    916 Main St, San Francisco, CA 94016      6
185949 2019-06-25 14:33:00         209 11th St, Atlanta, GA 30301       6

          Sales           City  Hour
0       1700.00   New York City     0
1        600.00   New York City     7
2         11.95   New York City    18
3        149.99   San Francisco    15
4         11.95         Atlanta    12
...         ...             ...   ...
185945     2.99          Boston    19
185946   149.99   New York City    19
185947    11.95   San Francisco    18
185948    11.95   San Francisco    18
185949     2.99         Atlanta    14

[185950 rows x 9 columns]
```
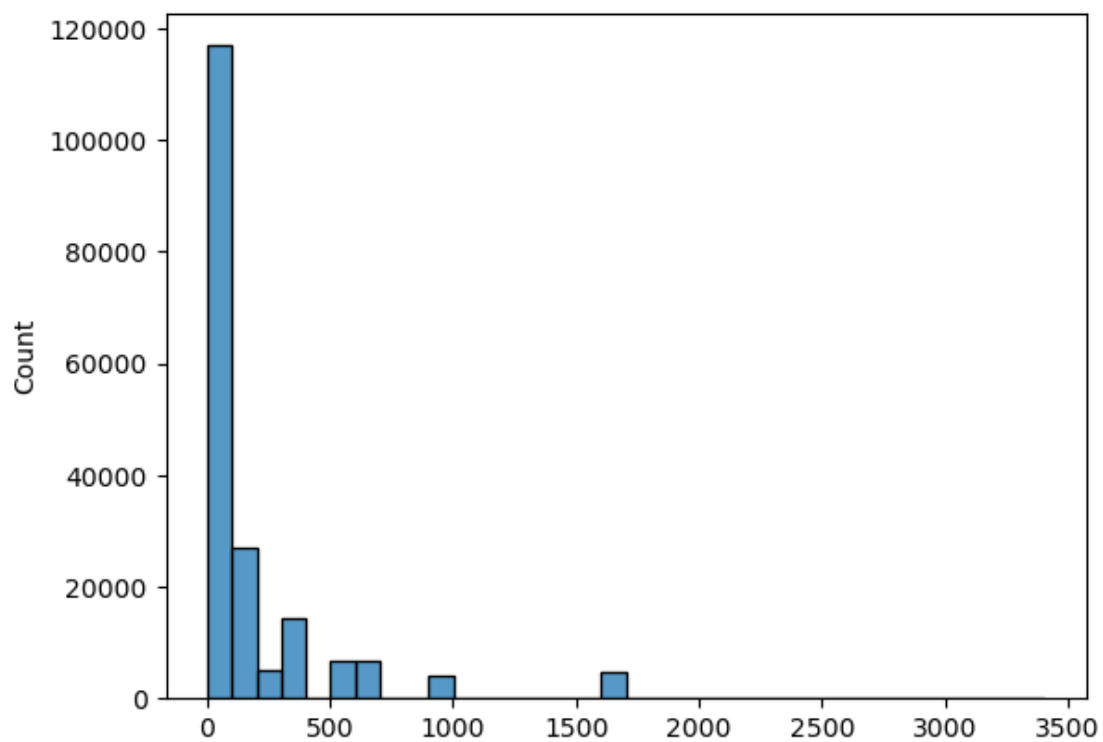
[13]: `sales['Sales'].mean().round()`

[13]: 185.0

[14]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```
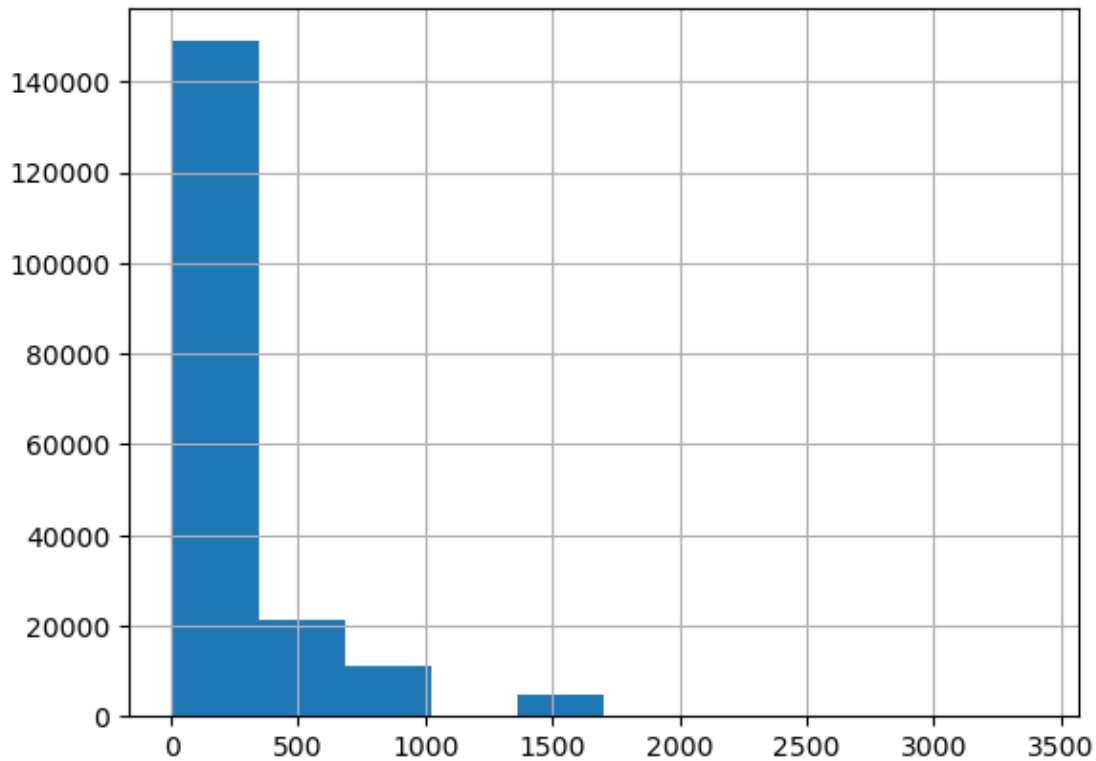
[15]: `sns.histplot(x = sales['Sales'].values , bins = np.arange(2,3500,100))`

[15]: `<Axes: ylabel='Count'>`

[16]: `sales['Sales'].hist()`

[16]: `<Axes: >`

```
[17]: lowerlimit = sales['Sales'].mean() - 1 * sales['Sales'].std()
      upperlimit = sales['Sales'].mean() + 1 * sales['Sales'].std()
      print ( 'lowerlimit is : ' , lowerlimit,'upperlimit is:' ,upperlimit,'drange is␣
       ↪:', upperlimit - lowerlimit )
```

lowerlimit is : -147.4288546346088 upperlimit is: 518.4106881382388 drange is :
665.8395427728476

```
[18]: ((sales['Sales'] >= lowerlimit) & (sales['Sales'] <= upperlimit)).mean()
```

[18]: 0.8785426189835978

```
[19]: lowerlimit = sales['Sales'].mean() - 2 * sales['Sales'].std()
      upperlimit = sales['Sales'].mean() + 2 * sales['Sales'].std()
      print ( 'lowerlimit is : ' , lowerlimit,'upperlimit is:' ,upperlimit,'range is :
       ↪', upperlimit - lowerlimit )
      ((sales['Sales'] >= lowerlimit) & (sales['Sales'] <= upperlimit)).mean()
```

lowerlimit is : -480.3486260210326 upperlimit is: 851.3304595246626 range is :
1331.6790855456952

[19]: 0.9523205162678139

```
[20]: lowerlimit = sales['Sales'].mean() - 3 * sales['Sales'].std()
      upperlimit = sales['Sales'].mean() + 3 * sales['Sales'].std()
      print ( 'lowerlimit is : ' , lowerlimit,'upperlimit is:' ,upperlimit,'range is :
       ↪', upperlimit - lowerlimit )
      ((sales['Sales'] >= lowerlimit) & (sales['Sales'] <= upperlimit)).mean()
```

lowerlimit is : -813.2683974074564 upperlimit is: 1184.2502309110864 range is :
1997.5186283185428

```
[20]: 0.9745092766872815
```

```
[21]: from scipy import stats
```

```
[22]: sales['z_score'] = stats.zscore(sales['Sales'])
```

```
[23]: sales
```

```
[23]:                         Product  Quantity Ordered  Price Each  \
      0              Macbook Pro Laptop                 1     1700.00
      1               LG Washing Machine                1      600.00
      2           USB-C Charging Cable                 1       11.95
      3                27in FHD Monitor                 1      149.99
      4           USB-C Charging Cable                 1       11.95
      ...                          ...               ...         ...
      185945  AAA Batteries (4-pack)                 1        2.99
      185946          27in FHD Monitor                 1      149.99
      185947     USB-C Charging Cable                 1       11.95
      185948     USB-C Charging Cable                 1       11.95
      185949  AAA Batteries (4-pack)                 1        2.99

                     Order Date                        Purchase Address  Month  \
      0      2019-12-30 00:01:00  136 Church St, New York City, NY 10001     12
      1      2019-12-29 07:03:00     562 2nd St, New York City, NY 10001     12
      2      2019-12-12 18:21:00    277 Main St, New York City, NY 10001     12
      3      2019-12-22 15:13:00     410 6th St, San Francisco, CA 94016     12
      4      2019-12-18 12:38:00          43 Hill St, Atlanta, GA 30301     12
      ...                    ...                                     ...    ...
      185945 2019-06-07 19:02:00          795 Pine St, Boston, MA 02215      6
      185946 2019-06-01 19:29:00  495 North St, New York City, NY 10001      6
      185947 2019-06-22 18:57:00  319 Ridge St, San Francisco, CA 94016      6
      185948 2019-06-26 18:35:00   916 Main St, San Francisco, CA 94016      6
      185949 2019-06-25 14:33:00          209 11th St, Atlanta, GA 30301      6

                Sales            City  Hour   z_score
      0       1700.00   New York City     0  4.549184
      1        600.00   New York City     7  1.245075
      2         11.95   New York City    18 -0.521271
      3        149.99   San Francisco    15 -0.106635
```

```
4        11.95            Atlanta    12 -0.521271
...          ...              ...   ...       ...
185945    2.99            Boston    19 -0.548184
185946  149.99    New York City    19 -0.106635
185947   11.95    San Francisco    18 -0.521271
185948   11.95    San Francisco    18 -0.521271
185949    2.99            Atlanta    14 -0.548184

[185950 rows x 10 columns]
```

[24]:
```python
#to detect the outliers
sales[(sales['z_score'] > 3 ) | (sales['z_score'] < -3)]
```

[24]:
```
                  Product   Quantity Ordered   Price Each           Order Date  \
0         Macbook Pro Laptop                 1       1700.0  2019-12-30 00:01:00
52        Macbook Pro Laptop                 1       1700.0  2019-12-10 20:02:00
57        Macbook Pro Laptop                 1       1700.0  2019-12-25 09:51:00
223       Macbook Pro Laptop                 1       1700.0  2019-12-28 11:19:00
302       Macbook Pro Laptop                 1       1700.0  2019-12-17 21:08:00
...                      ...                ...          ...                  ...
185757    Macbook Pro Laptop                 1       1700.0  2019-06-13 21:29:00
185775    Macbook Pro Laptop                 1       1700.0  2019-06-11 07:38:00
185796    Macbook Pro Laptop                 1       1700.0  2019-06-14 21:33:00
185857    Macbook Pro Laptop                 1       1700.0  2019-06-17 00:18:00
185944    Macbook Pro Laptop                 1       1700.0  2019-06-09 22:07:00

                              Purchase Address   Month   Sales  \
0            136 Church St, New York City, NY 10001     12  1700.0
52           331 Madison St, New York City, NY 10001    12  1700.0
57              82 10th St, San Francisco, CA 94016     12  1700.0
223                  661 Park St, Dallas, TX 75001     12  1700.0
302             863 West St, San Francisco, CA 94016    12  1700.0
...                                           ...    ...     ...
185757       657 River St, Los Angeles, CA 90001      6  1700.0
185775       775 11th St, Los Angeles, CA 90001       6  1700.0
185796          319 6th St, Portland, OR 97035        6  1700.0
185857   393 Willow St, San Francisco, CA 94016       6  1700.0
185944          975 2nd St, Los Angeles, CA 90001     6  1700.0

                  City  Hour    z_score
0        New York City     0   4.549184
52       New York City    20   4.549184
57       San Francisco     9   4.549184
223             Dallas    11   4.549184
302      San Francisco    21   4.549184
...                ...   ...        ...
185757     Los Angeles    21   4.549184
```
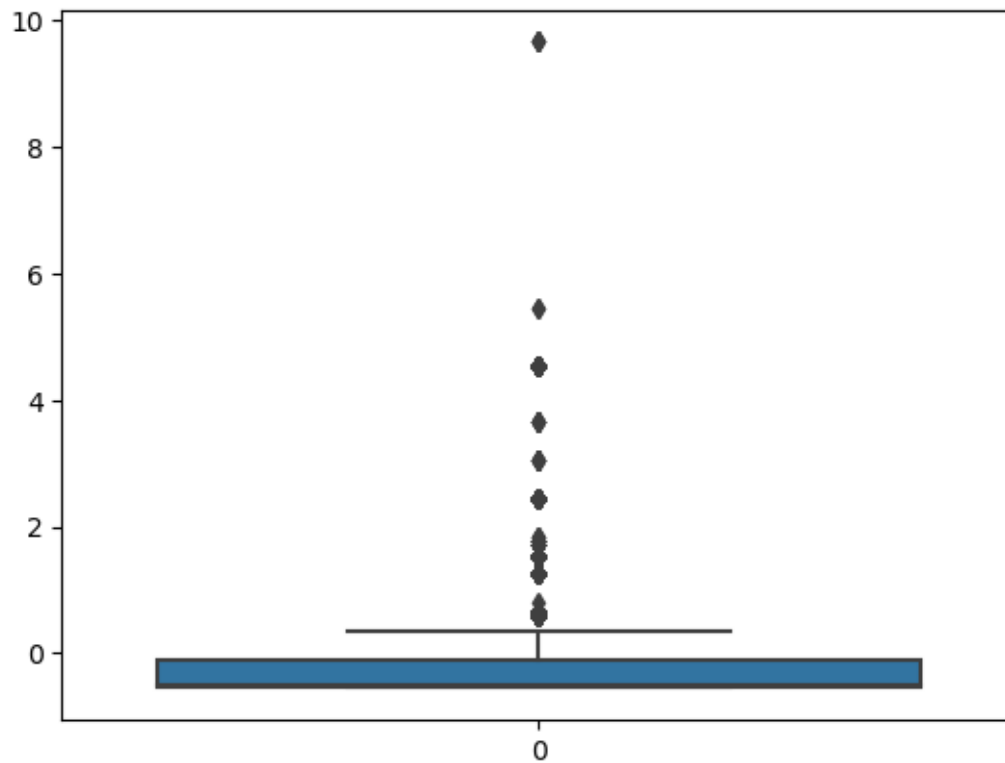
```
185775      Los Angeles     7  4.549184
185796         Portland    21  4.549184
185857    San Francisco     0  4.549184
185944      Los Angeles    22  4.549184

[4740 rows x 10 columns]
```
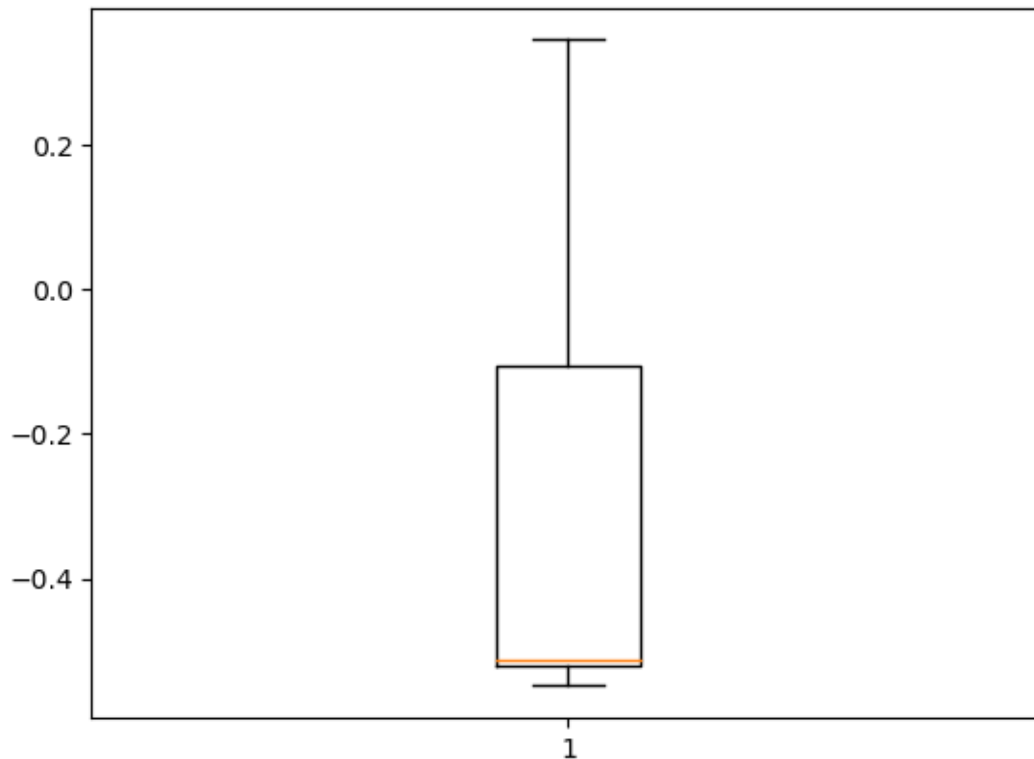
[25]: `sns.boxplot(sales['z_score'])`

[25]: `<Axes: >`

[33]: `plt.boxplot(sales['z_score'],showfliers = False)`

[33]: ```
{'whiskers': [<matplotlib.lines.Line2D at 0x21a91435bd0>,
  <matplotlib.lines.Line2D at 0x21a9143c210>],
 'caps': [<matplotlib.lines.Line2D at 0x21a9143dd90>,
  <matplotlib.lines.Line2D at 0x21a9143fa50>],
 'boxes': [<matplotlib.lines.Line2D at 0x21a913b3050>],
 'medians': [<matplotlib.lines.Line2D at 0x21a91448b50>],
 'fliers': [],
 'means': []}
```

```
[42]: sns.distplot(((sales['z_score'] > 3 ) | (sales['z_score'] < -3)))
```

C:\Users\User\AppData\Local\Temp\ipykernel_2064\2678802000.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(((sales['z_score'] > 3) | (sales['z_score'] < -3)))

```
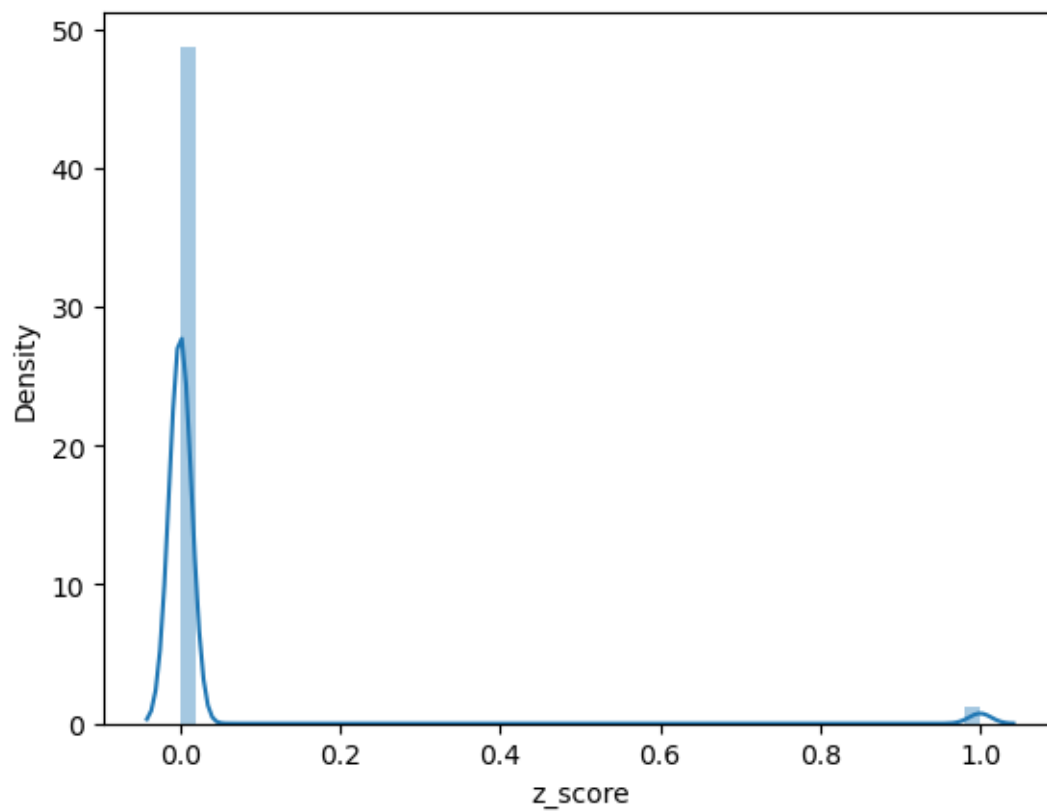[42]: <Axes: xlabel='z_score', ylabel='Density'>
```

[ ]: