```
1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
```

```
1 data = pd.read_excel('drive/MyDrive/QVI_transaction_data.xlsx')
```

```
1 data.head()
```

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	T
0	43390	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	
1	43599	1	1307	348	66	CCs Nacho Cheese 175g	3	
2	43605	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	
3	43329	2	2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	5	
4	43330	2	2426	1038	108	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	

SUMMARIZATION

```
1 data.describe()
```

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	
count	264836.000000	264836.00000	2.648360e+05	2.648360e+05	264836.000000	2648
mean	43464.036260	135.08011	1.355495e+05	1.351583e+05	56.583157	
std	105.389282	76.78418	8.057998e+04	7.813303e+04	32.826638	
min	43282.000000	1.00000	1.000000e+03	1.000000e+00	1.000000	
25%	43373.000000	70.00000	7.002100e+04	6.760150e+04	28.000000	
50%	43464.000000	130.00000	1.303575e+05	1.351375e+05	56.000000	
75%	43555.000000	203.00000	2.030942e+05	2.027012e+05	85.000000	
max	43646.000000	272.00000	2.373711e+06	2.415841e+06	114.000000	2

```
1 data.isnull().sum()
```

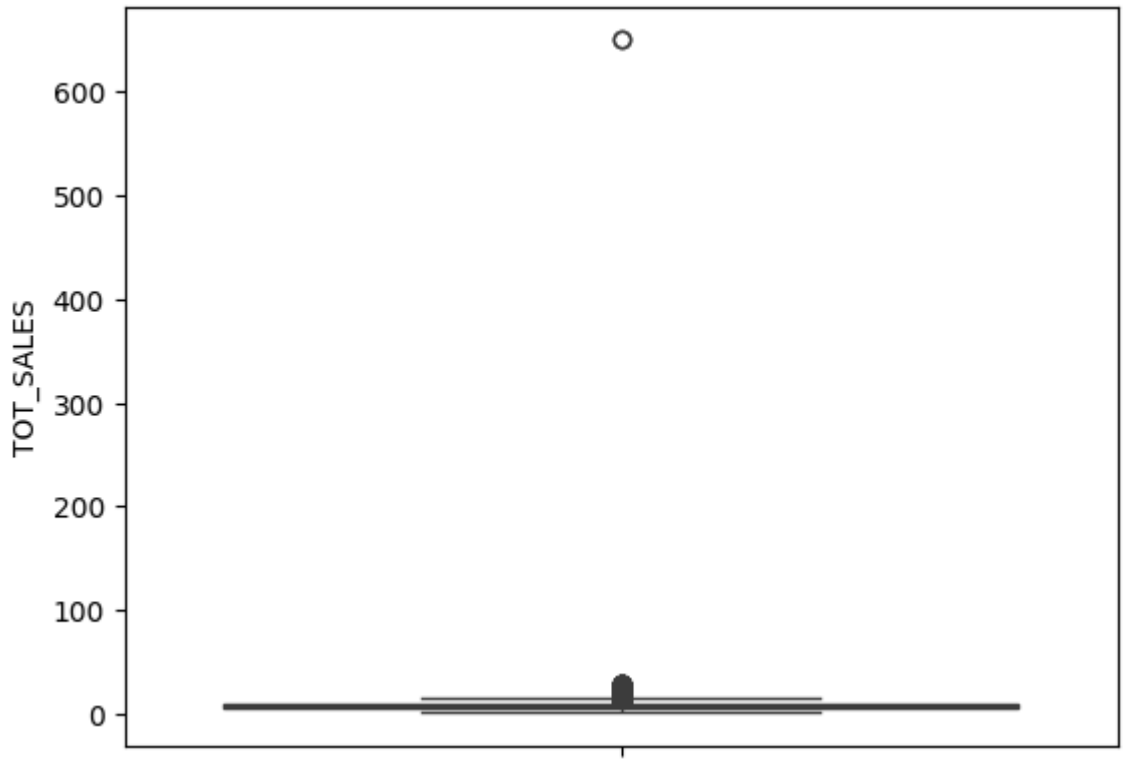
	0
DATE	0
STORE_NBR	0
LYLTY_CARD_NBR	0
TXN_ID	0
PROD_NBR	0
PROD_NAME	0
PROD_QTY	0
TOT_SALES	0

dtype: int64

CHECK FOR OUTLIERS

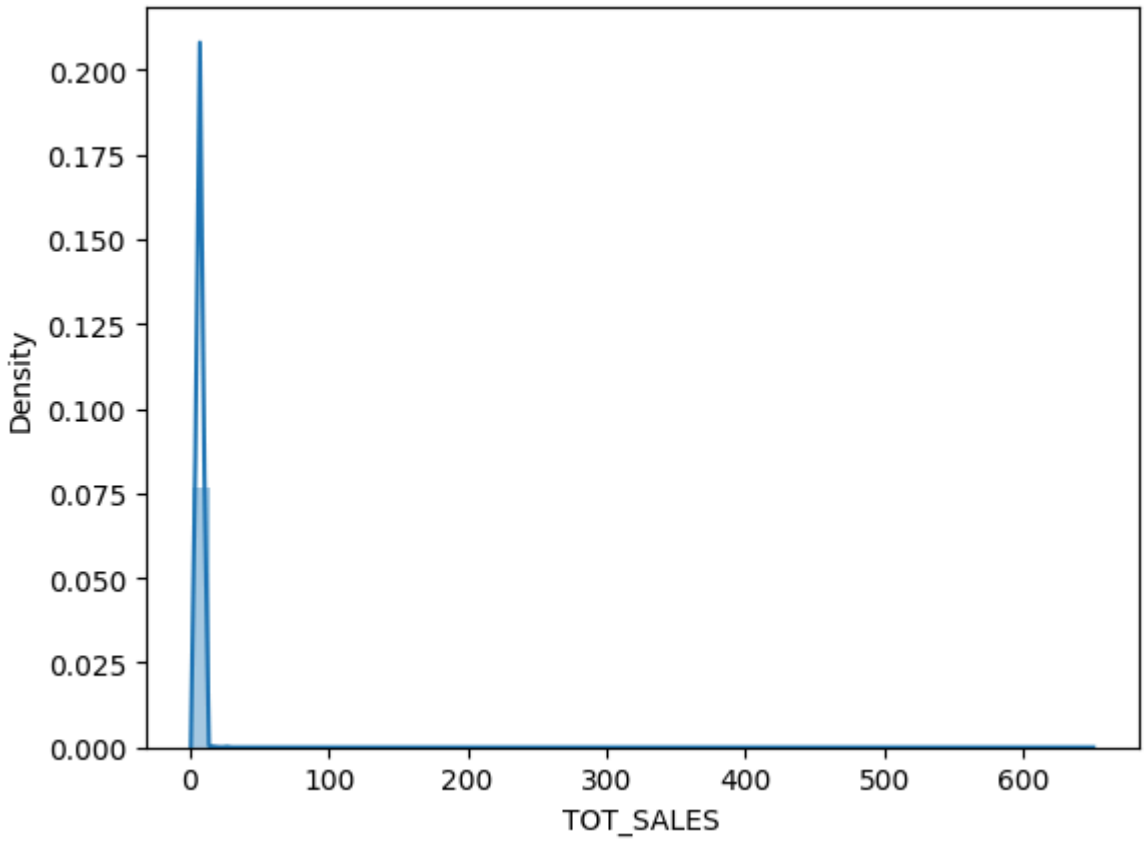
```
1 sns.boxplot(data.TOT_SALES)
```

```
<Axes: ylabel='TOT_SALES'>
```




```
1 sns.distplot(data.TOT_SALES,kde=True)
```

```
<ipython-input-11-67f0fbf7b018>: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(data.TOT_SALES,kde=True)
<Axes: xlabel='TOT_SALES', ylabel='Density'>
```




```
num_data=data.select_dtypes(['float','int'])
```

```
num_data.head()
```

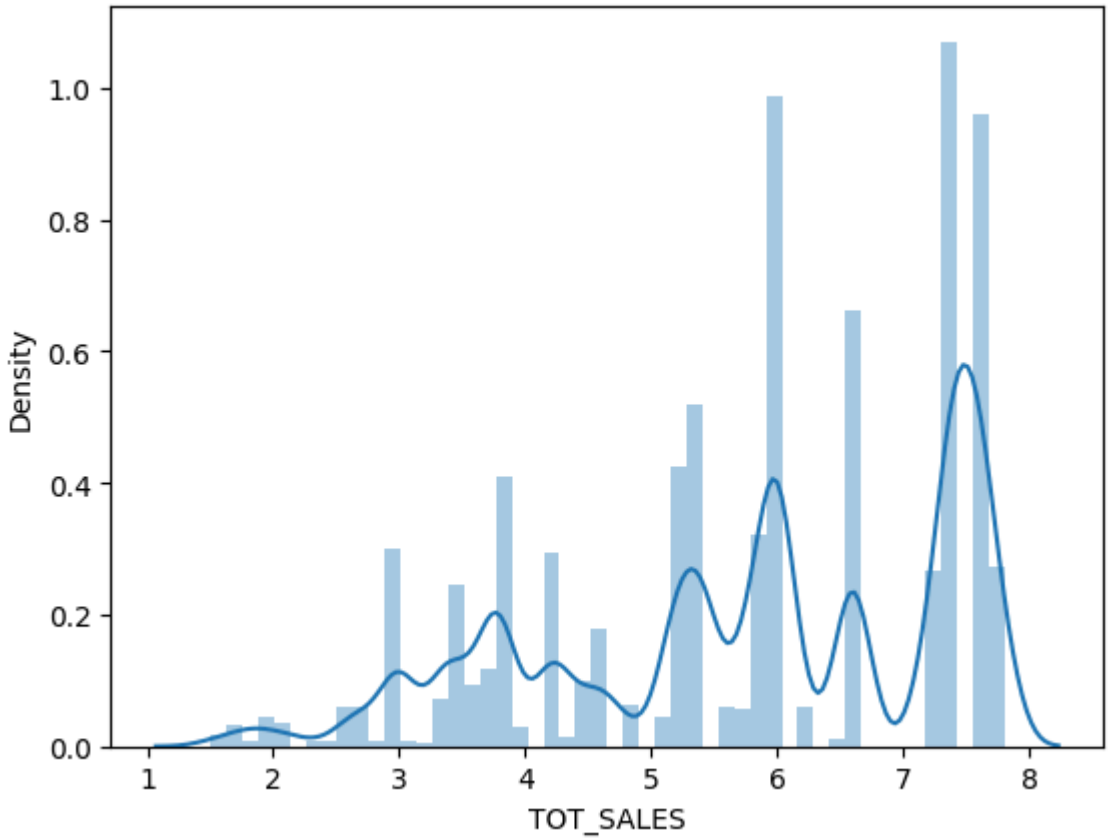
	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_QTY	TOT_SALES
0	43390	1	1000	1	5	2	6.0
1	43599	1	1307	348	66	3	6.3
2	43605	1	1343	383	61	2	2.9
3	43329	2	2373	974	69	5	15.0
4	43330	2	2426	1038	108	3	13.8

REMOVING OUTLIERS

```
1 x=num_data[num_data['TOT_SALES']<8.000]
```

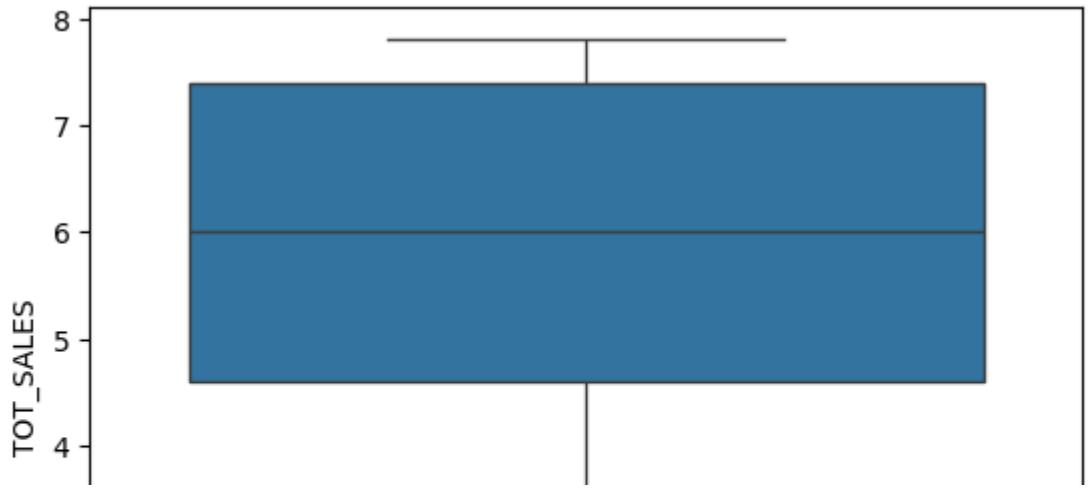
```
1 sns.distplot(x.TOT_SALES,kde=True)
```

```
<ipython-input-17-378e3485a1ab>: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(x.TOT_SALES,kde=True)
<Axes: xlabel='TOT_SALES', ylabel='Density'>
```




```
1 sns.boxplot(x.TOT_SALES)
```

```
<Axes: ylabel='TOT_SALES'>
```



DATA FORMATS

```
1 data.dtypes
```

	0
DATE	int64
STORE_NBR	int64
LYLTY_CARD_NBR	int64
TXN_ID	int64
PROD_NBR	int64
PROD_NAME	object
PROD_QTY	int64
TOT_SALES	float64

dtype: object