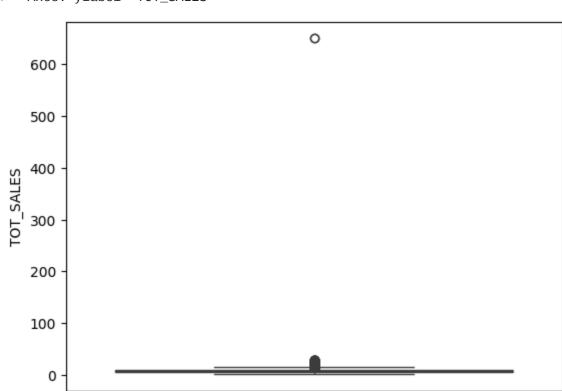
In [38]: **import** numpy **as** np import pandas as pd import seaborn as sns In [46]: import openpyxl as xl dataset = pd.read_excel("Downloads/QVI_transaction_data.xlsx") In [48]: dataset.head() Out[48]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME PROD_QTY TOT_SALES **0** 43390 1000 5 Natural Chip Compny SeaSalt175g 2 6.0 1 43599 1307 348 66 CCs Nacho Cheese 175g 6.3 2 43605 1343 383 Smiths Crinkle Cut Chips Chicken 170g 2.9 2 **3** 43329 69 Smiths Chip Thinly S/Cream&Onion 175g 2373 974 15.0 4 43330 2 2426 1038 108 Kettle Tortilla ChpsHny&Jlpno Chili 150g 3 13.8 In [86]: dataset.describe() #sumarization of dataset Out[86]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_QTY TOT_SALES count 264836.000000 264836.00000 1.355495e+05 1.351583e+05 43464.036260 135.08011 56.583157 1.907309 7.304200 mean 105.389282 76.78418 8.057998e+04 7.813303e+04 32.826638 0.643654 3.083226 std 1.000000e+03 1.000000e+00 43282.000000 1.00000 1.000000 1.000000 1.500000 43373.000000 70.00000 7.002100e+04 6.760150e+04 28.000000 2.000000 5.400000 25% 1.303575e+05 1.351375e+05 43464.000000 130.00000 56.000000 2.000000 7.400000 50% 43555.000000 203.00000 2.030942e+05 2.027012e+05 2.000000 9.200000 75% 85.000000 max 43646.000000 272.00000 114.000000 200.000000 650.000000 2.373711e+06 2.415841e+06

In [53]: dataset.isnull().sum()

Out[53]: **DATE** 0 STORE_NBR 0 LYLTY_CARD_NBR TXN_ID 0 PROD_NBR 0 PROD_NAME 0 PROD_QTY 0 TOT_SALES 0 dtype: int64

In [56]: sns.boxplot(dataset.TOT_SALES)

Out[56]: <Axes: ylabel='TOT_SALES'>



In [61]: sns.distplot(dataset.TOT_SALES, kde = True)

C:\Users\pandeysunny2315\AppData\Local\Temp\ipykernel_11792\386653243.py:1: UserWarning:

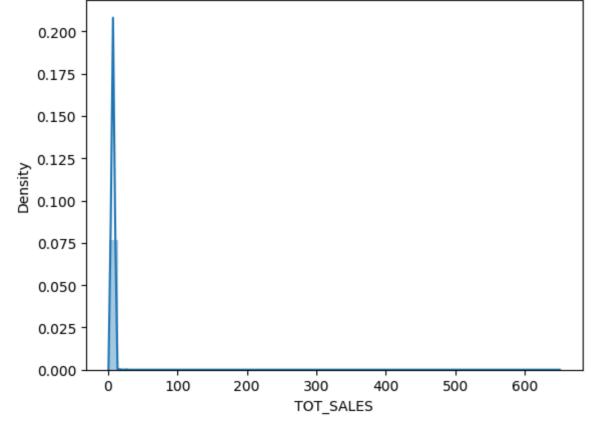
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

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(dataset.TOT_SALES, kde = True)

Out[61]: <Axes: xlabel='TOT_SALES', ylabel='Density'>



In [69]: numericdataset = dataset.select_dtypes(['float','int'])

In [70]: numericdataset.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

In [74]: x = numericdataset[numericdataset["TOT_SALES"] < 8.000]</pre>

In [76]: sns.distplot(x.TOT_SALES,kde = True)

C:\Users\pandeysunny2315\AppData\Local\Temp\ipykernel_11792\712269423.py:1: UserWarning:

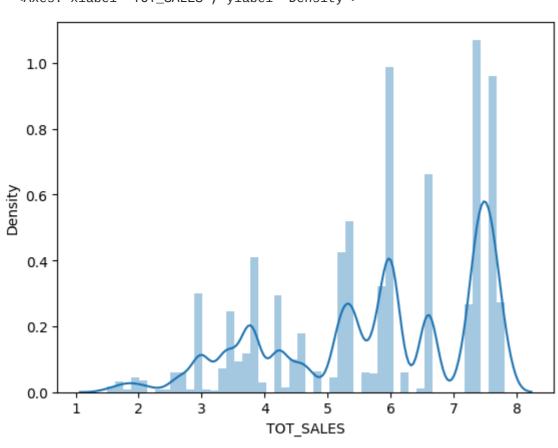
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with

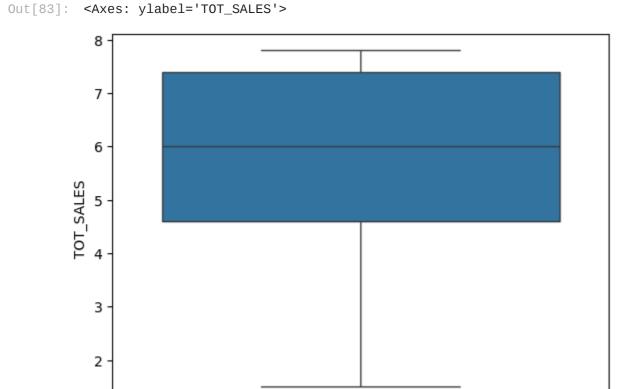
Please adapt your code to use either 'displot' (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

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)
Out[76]: <Axes: xlabel='TOT_SALES', ylabel='Density'>



In [83]: sns.boxplot(x.TOT_SALES)



In [87]: dataset.dtypes

Out[87]: 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