

```
In [ ]: import numpy as np
```

```
In [ ]: pip install pandas
```

Requirement already satisfied: pandas in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (2.2.3)

Requirement already satisfied: numpy>=1.23.2 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (1.26.4)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2024.2)

Requirement already satisfied: tzdata>=2022.7 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2024.2)

Requirement already satisfied: six>=1.5 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run: C:\Users\P00JA\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip

```
In [ ]: pip install seaborn
```

Requirement already satisfied: seaborn in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (0.13.2)

Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from seaborn) (1.26.4)

Requirement already satisfied: pandas>=1.2 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from seaborn) (2.2.3)

Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from seaborn) (3.10.0)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.1)

Requirement already satisfied: cycler>=0.10 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.55.3)

Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)

Requirement already satisfied: packaging>=20.0 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.1)

Requirement already satisfied: pillow>=8 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.1.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.1)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas>=1.2->seaborn) (2024.2)

Requirement already satisfied: tzdata>=2022.7 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas>=1.2->seaborn) (2024.2)

Requirement already satisfied: six>=1.5 in c:\users\pooja\appdata\local\packages\pythonsoftwarefoundation.python.3.11\_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)

Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run: C:\Users\P00JA\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip

```
In [ ]: import pandas as pd
import seaborn as sns
```

```
In [ ]: pip install openpyxl
```

Collecting openpyxl

Downloading openpyxl-3.1.5-py2.py3-none-any.whl.metadata (2.5 kB)

Collecting et\_xmlfile (from openpyxl)

Downloading et\_xmlfile-2.0.0-py3-none-any.whl.metadata (2.7 kB)

Downloading openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)

Downloading et\_xmlfile-2.0.0-py3-none-any.whl (18 kB)

Installing collected packages: et\_xmlfile, openpyxl

Successfully installed et\_xmlfile-2.0.0 openpyxl-3.1.5

Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run: C:\Users\P00JA\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip

In [ ]: `import pandas as pd`

*# Define the file path*

`file_path = "C:\\Users\\P00JA\\Desktop\\Quantium\\Task1"`

`transaction_data = pd.read_excel( file_path + "\\QVI_transaction_data.xlsx")`

*# Preview the data*

`transaction_data.head()`

Out [ ]: 

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY
--	------	-----------	----------------	--------	----------	-----------	----------

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
---	-------	---	------	------	-----	--	---



In [ ]: `file_path1 = "C:\\Users\\P00JA\\Desktop\\Quantium\\Task1"`  
`customer_data= pd.read_csv( file_path1 + "\\QVI_purchase_behaviour.csv")`  
`customer_data.head()`

Out[ ]:	LYLTY_CARD_NBR	LIFESTAGE	PREMIUM_CUSTOMER
0	1000	YOUNG SINGLES/COUPLES	Premium
1	1002	YOUNG SINGLES/COUPLES	Mainstream
2	1003	YOUNG FAMILIES	Budget
3	1004	OLDER SINGLES/COUPLES	Mainstream
4	1005	MIDAGE SINGLES/COUPLES	Mainstream

In [ ]: `transaction_data.describe()`

Out[ ]:	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PR
count	264836.000000	264836.00000	2.648360e+05	2.648360e+05	264836.000000	26483
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	20

In [ ]: `transaction_data.isnull().sum()`

Out[ ]: 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

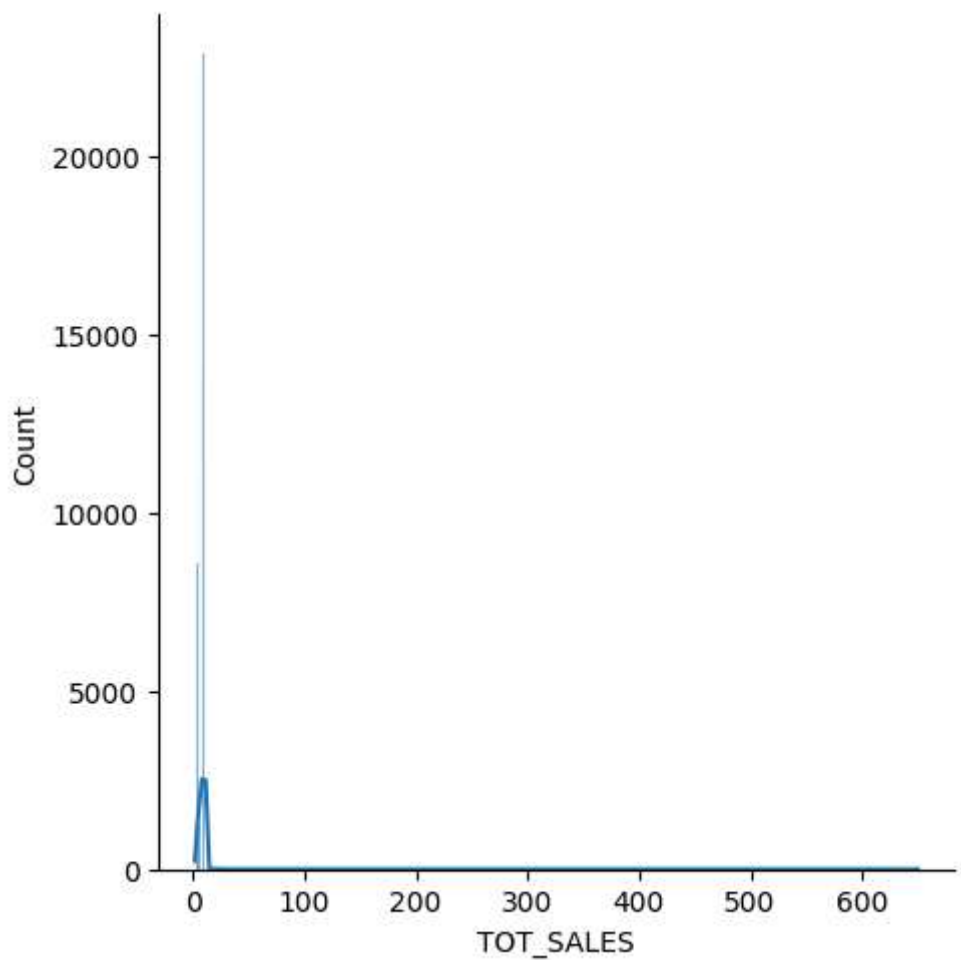
In [ ]: `data_type=transaction_data.dtypes`  
`print(data_type)`

```
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
```

```
In [ ]: import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]: sns.displot(transaction_data.TOT_SALES,kde=True)
```

```
Out[ ]: <seaborn.axisgrid.FacetGrid at 0x1dc5c3b9250>
```



```
In [ ]: numericdata= transaction_data.select_dtypes(['float', 'int'])
numericdata.head()
```

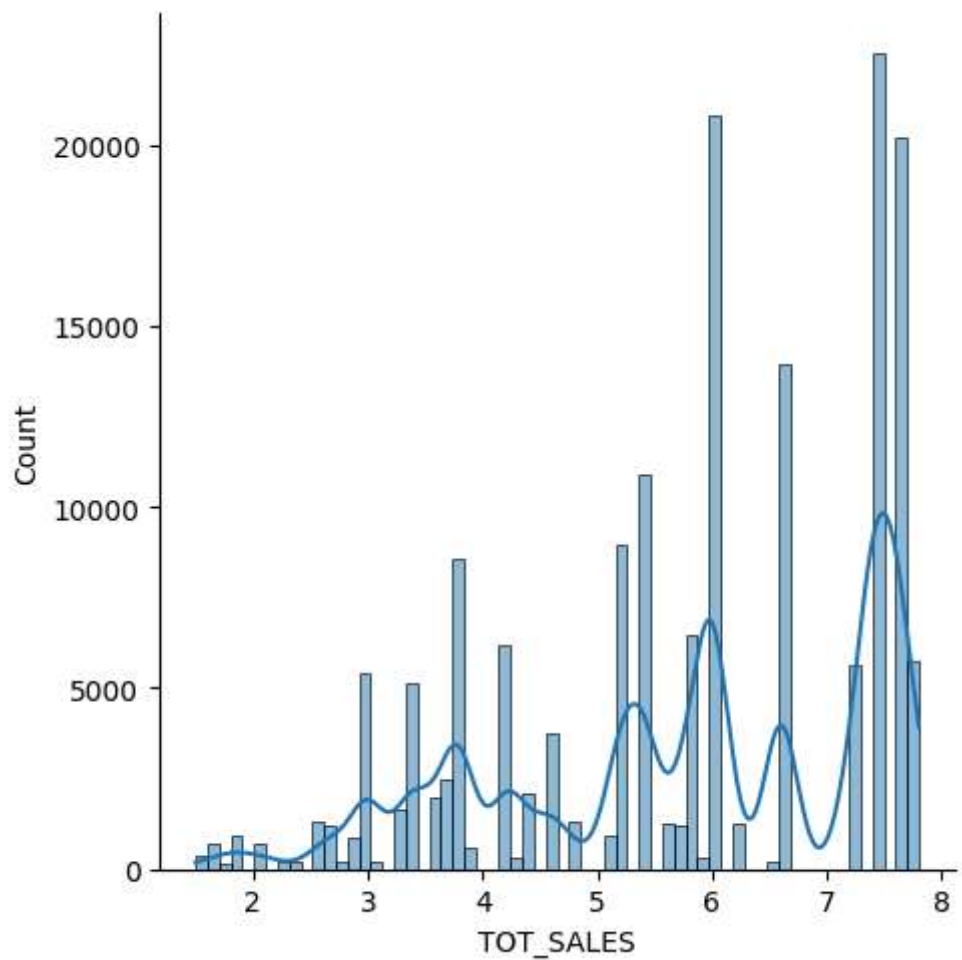
```
Out[ ]:
```

	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 [ ]: x=numericdata[numericdata['TOT_SALES']<8.000]
```

```
In [ ]: sns.displot(x.TOT_SALES,kde=True)
```

```
Out[ ]: <seaborn.axisgrid.FacetGrid at 0x1dc5f75abd0>
```



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

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

