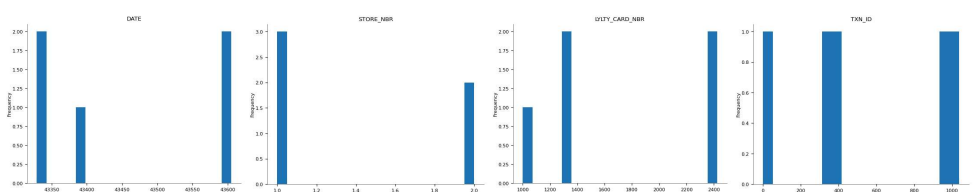
```
import pandas as pd
import numpy as np
import seaborn as sns
```

```
dataset=pd.read_excel('QVI_transaction_data.xlsx')
```

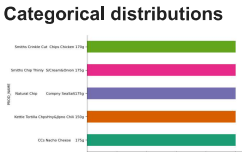
```
dataset.head()
```

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES
0	43390	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	6.0
1	43599	1	1307	348	66	CCs Nacho Cheese 175g	3	6.3
2	43605	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	2.9
3	43329	2	2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	5	15.0
4	43330	2	2426	1038	108	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.8

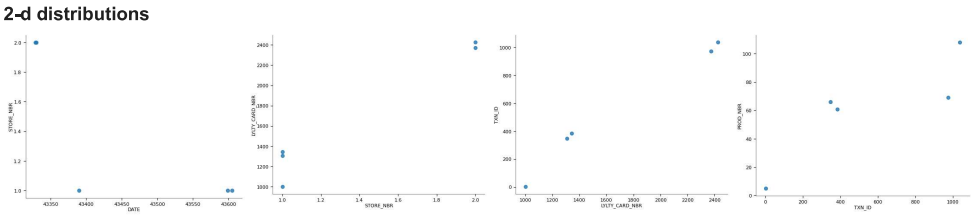
Distributions



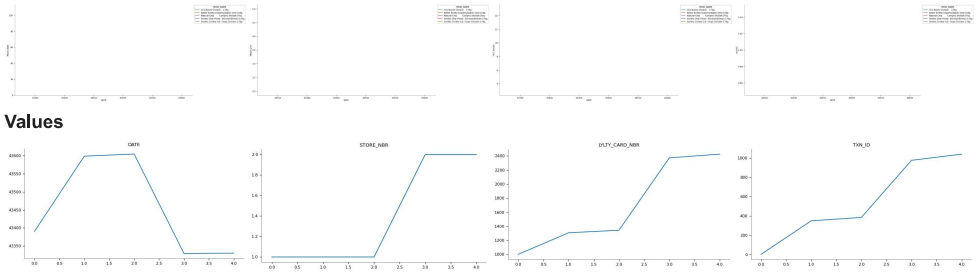
Categorical distributions



2-d distributions



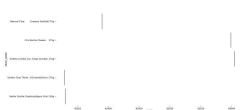
Time series



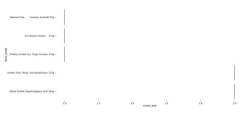
Values

Faceted distributions

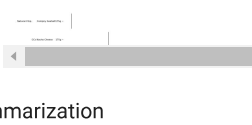
```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari
```



```
                    <string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari
```



```
                    <string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari
```



```
                    <string>:5: FutureWarning:
```

Summarization

```
dataset.describe()
```

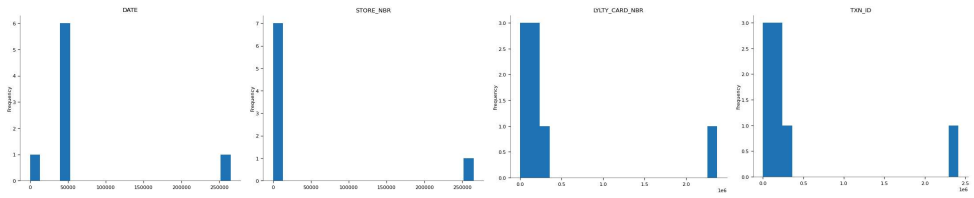
1 to 8 of 8 entries Filter

index	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_
count	264836.0	264836.0	264836.0	264836.0	264836.0	
mean	43464.03626017611	135.08010995483997	135549.47640426527	135158.31081499494	56.58315712365388	1.9073086
std	105.38928199808305	76.78418000108746	80579.9780220639	78133.02602582192	32.826637807097676	0.6436539
min	43282.0	1.0	1000.0	1.0	1.0	
25%	43373.0	70.0	70021.0	67601.5	28.0	
50%	43464.0	130.0	130357.5	135137.5	56.0	
75%	43555.0	203.0	203094.25	202701.25	85.0	
max	43646.0	272.0	2373711.0	2415841.0	114.0	

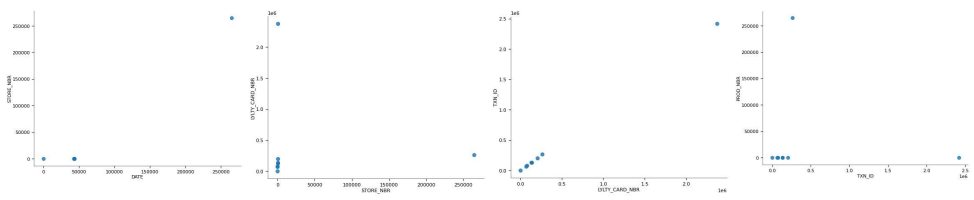
Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

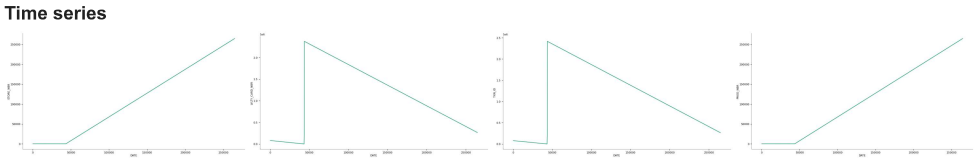
Distributions



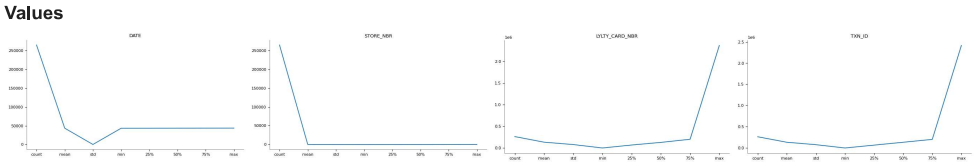
2-d distributions



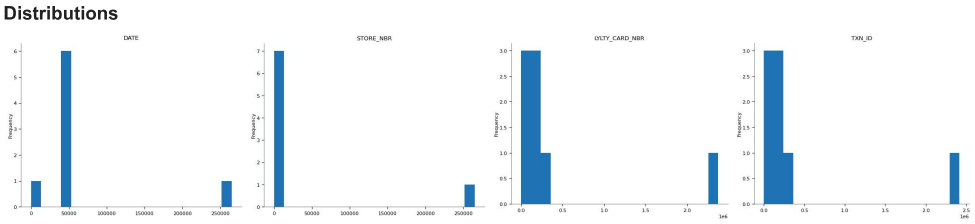
Time series



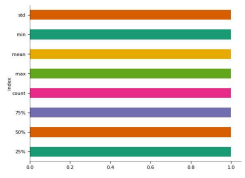
Values



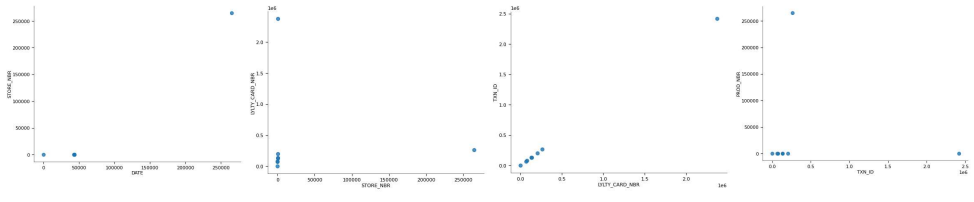
Distributions



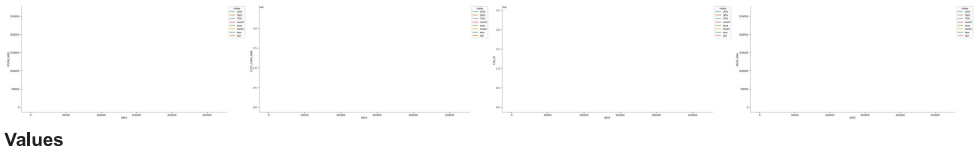
Categorical distributions



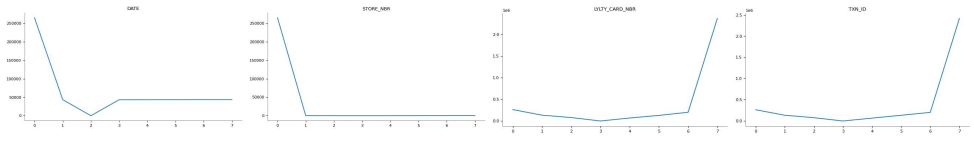
2-d distributions



Time series



Values



Faceted distributions

```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari
```

```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari

<string>:5: FutureWarning:
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` vari

                  <string>:5: FutureWarning:
```

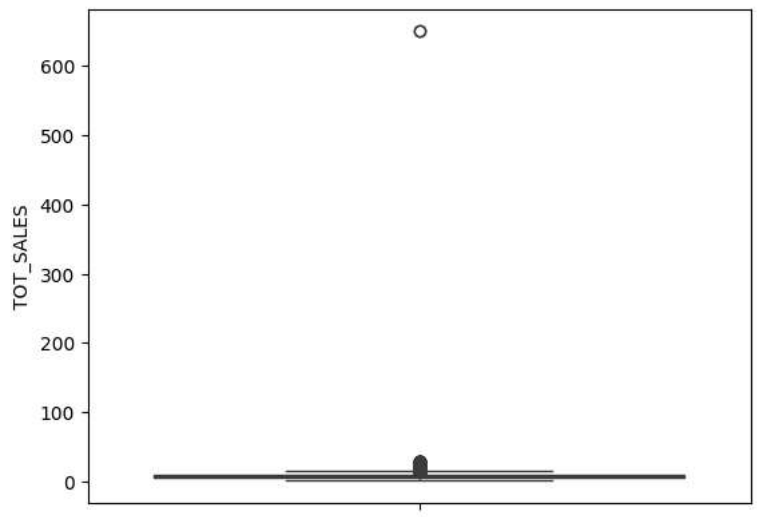
```
dataset.isnull().sum()
```

```
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
```

CHECKING FOR OUTLIERS

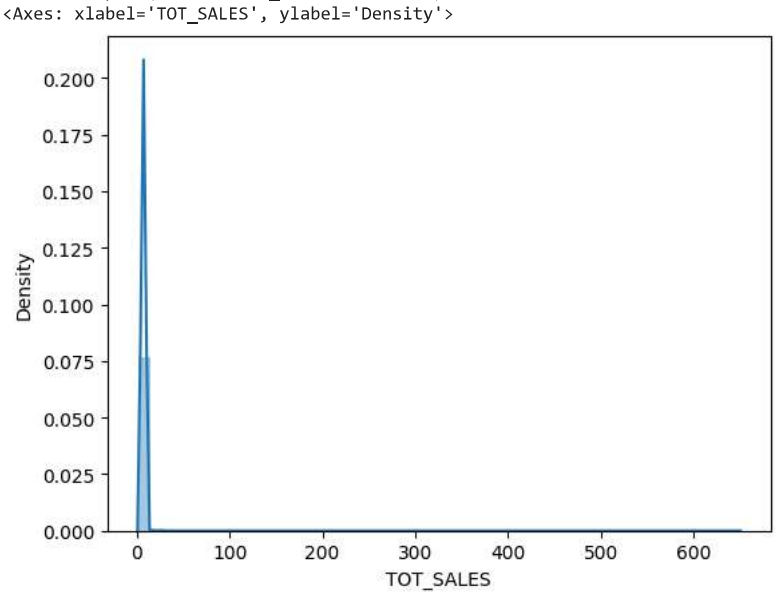
```
sns.boxplot(dataset.TOT_SALES)
```

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




```
sns.distplot(dataset.TOT_SALES, kde=True)
```

```
<ipython-input-12-d8db5eb0d189>: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)
<Axes: xlabel='TOT_SALES', ylabel='Density'>
```




```
numericdata=dataset.select_dtypes(['float','int'])
```

```
numericdata.head()
```

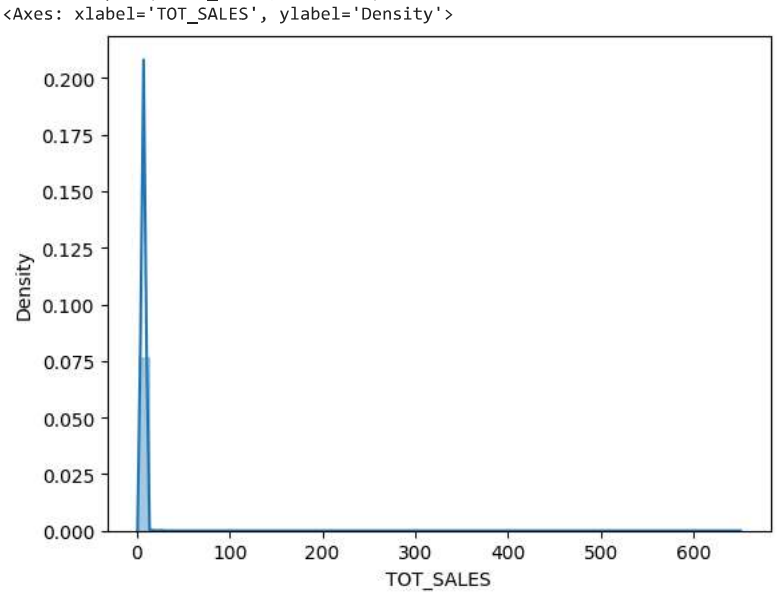
	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_QTY	TOT_SALES
0	43390	1	1000	1	5	2	6.0
1	43599	1	1307	348	66	3	6.3
2	43605	1	1343	383	61	2	2.9
4	43330	2	2426	1038	108	3	13.8

REMOVING OUTLIERS

```
x=numericdata[numericdata['TOT_SALES' ]<8000]
```

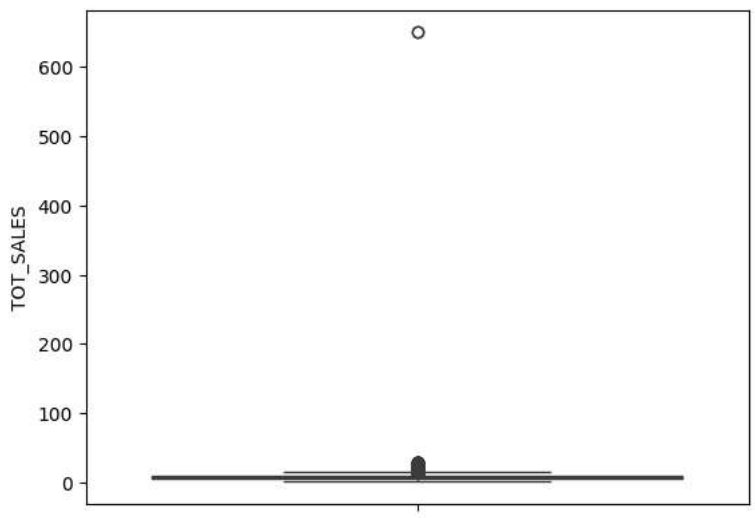
```
sns.distplot(x.TOT_SALES, kde=True)
```

```
<ipython-input-21-3b9dd6de9231>: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'>
```




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

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



DATA FORMATS

```
dataset.dtypes
```

```
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
```

Start coding or generate with AI.