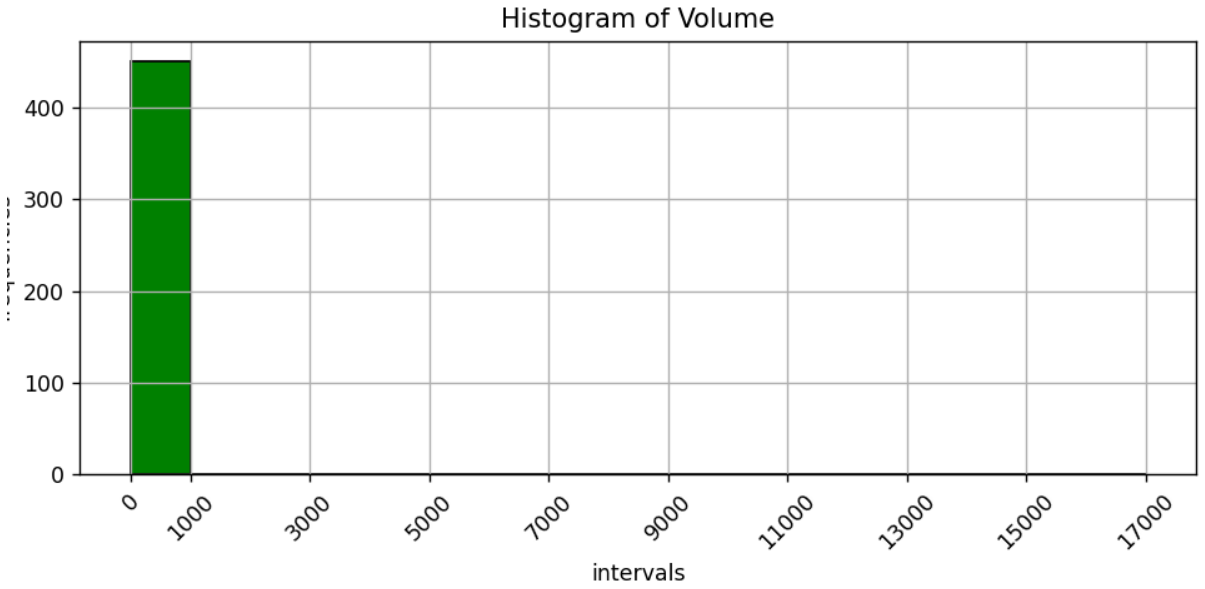
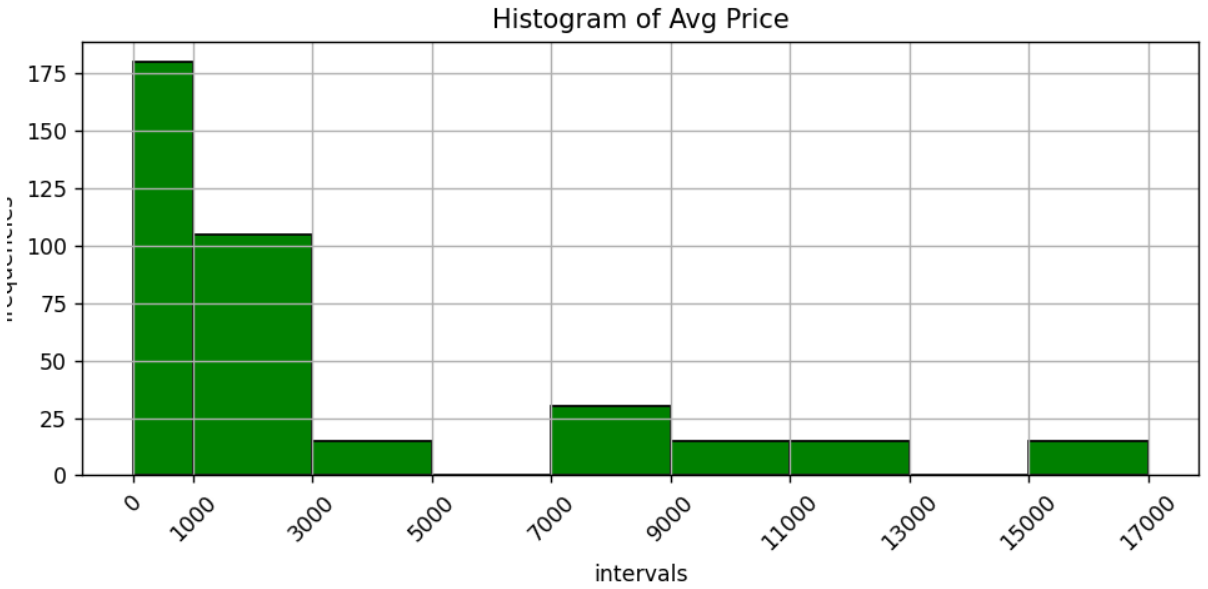
Task 1

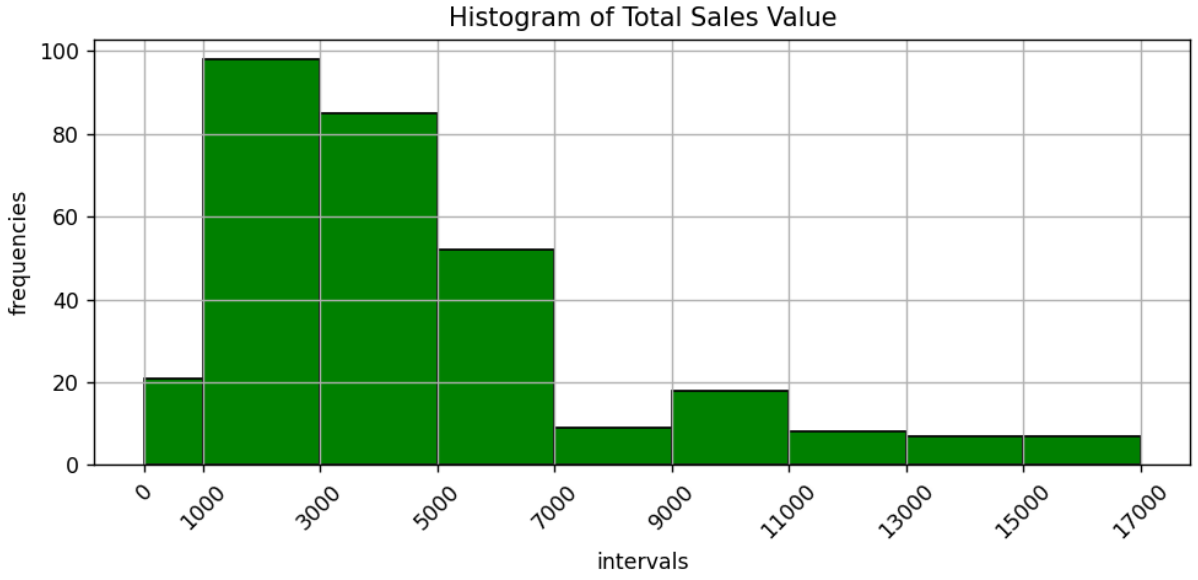
```
================== RESTART: D:\python\spr
volume_mean       5.066667
volume_median     4.000000
volume_mode       3.000000
volume_std        4.231602
Name: Volume, dtype: float64

price_mean      10453.433333
price_median     1450.000000
price_mode        400.000000
price_std       18079.904840
Name: Avg Price, dtype: float64

sales_mean      33812.835556
sales_median     5700.000000
sales_mode      24300.000000
sales_std       50535.074173
Name: Total Sales Value, dtype: float64

discount_rate_mean      15.155242
discount_rate_median    16.577766
discount_rate_mode       5.007822
discount_rate_std        4.220602
Name: Discount Rate (%), dtype: float64

discount_mean      3346.499424
discount_median     988.933733
discount_mode        69.177942
discount_std       4509.902963
Name: Discount Amount, dtype: float64

net_sales_mean      30466.336131
net_sales_median     4677.788059
net_sales_mode        326.974801
net_sales_std       46358.656624
Name: Net Sales Value, dtype: float64
```

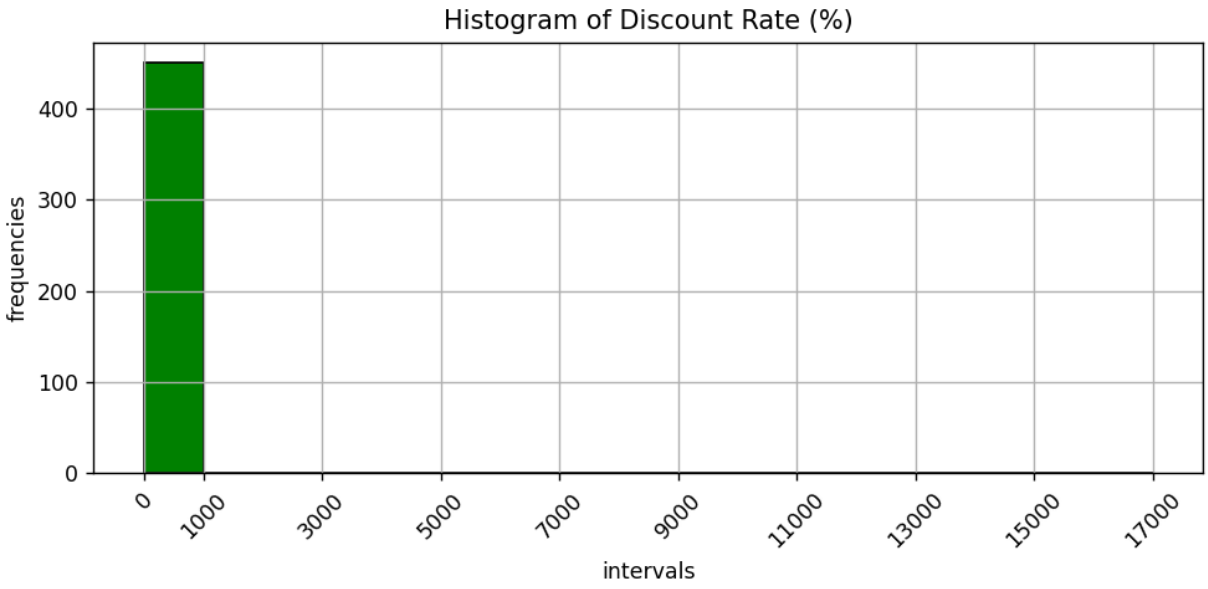
Data Visualization

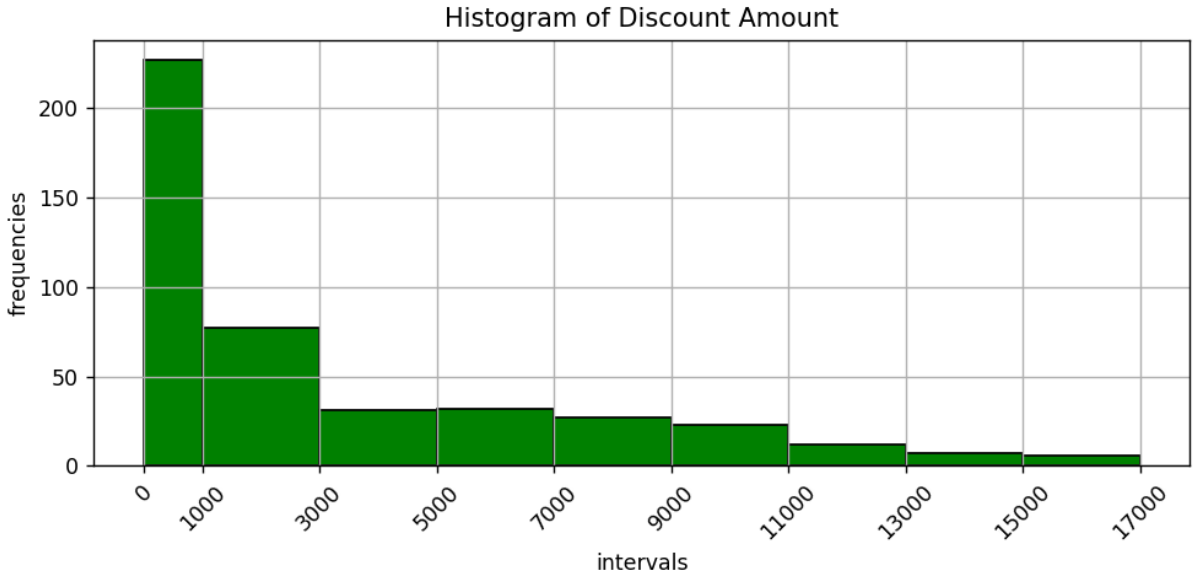
Histogram

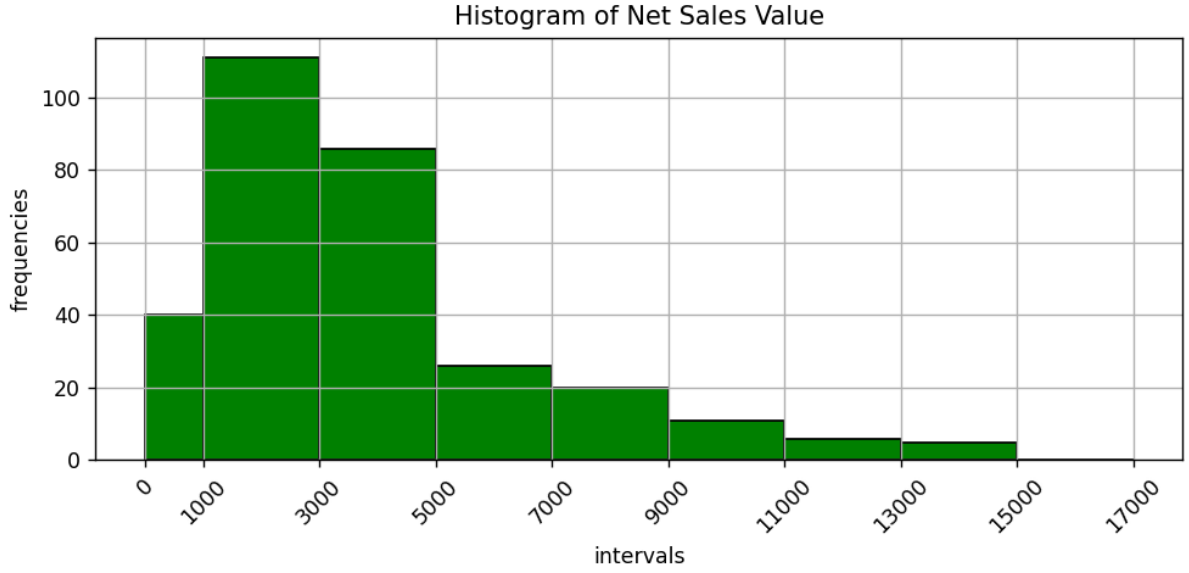

Histogram of Volume
400
300
200
100
0
0
1000
3000
5000
7000
9000
11000
13000
15000
17000
intervals


Histogram of Avg Price
175
150
125
100
75
50
25
0
0
1000
3000
5000
7000
9000
11000
13000
15000
17000
intervals


Histogram of Total Sales Value
frequencies
100
80
60
40
20
0
0
1000
3000
5000
7000
9000
11000
13000
15000
17000
intervals

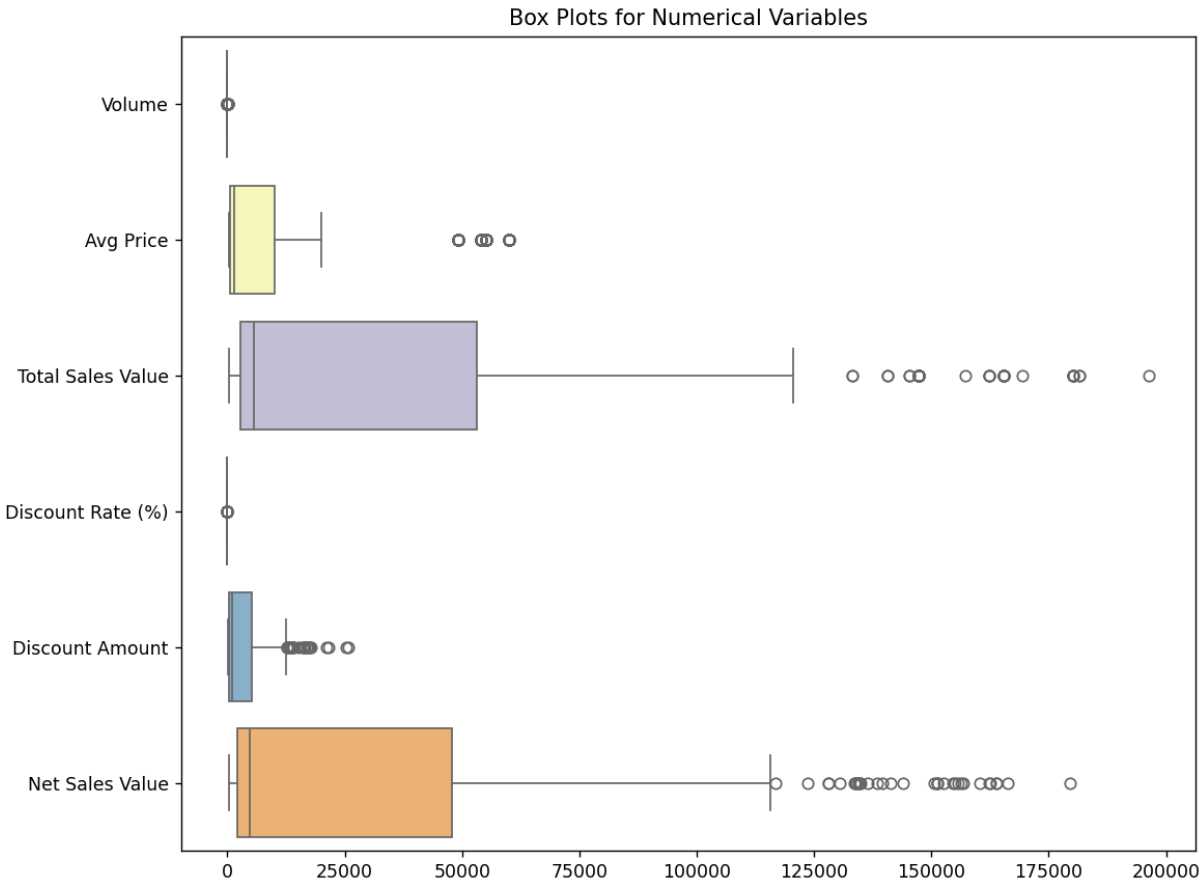

Histogram of Discount Rate (%)
frequencies
intervals
0
100
200
300
400
0
1000
3000
5000
7000
9000
11000
13000
15000
17000


Histogram of Discount Amount
frequencies
intervals
0
50
100
150
200
0
1000
3000
5000
7000
9000
11000
13000
15000
17000

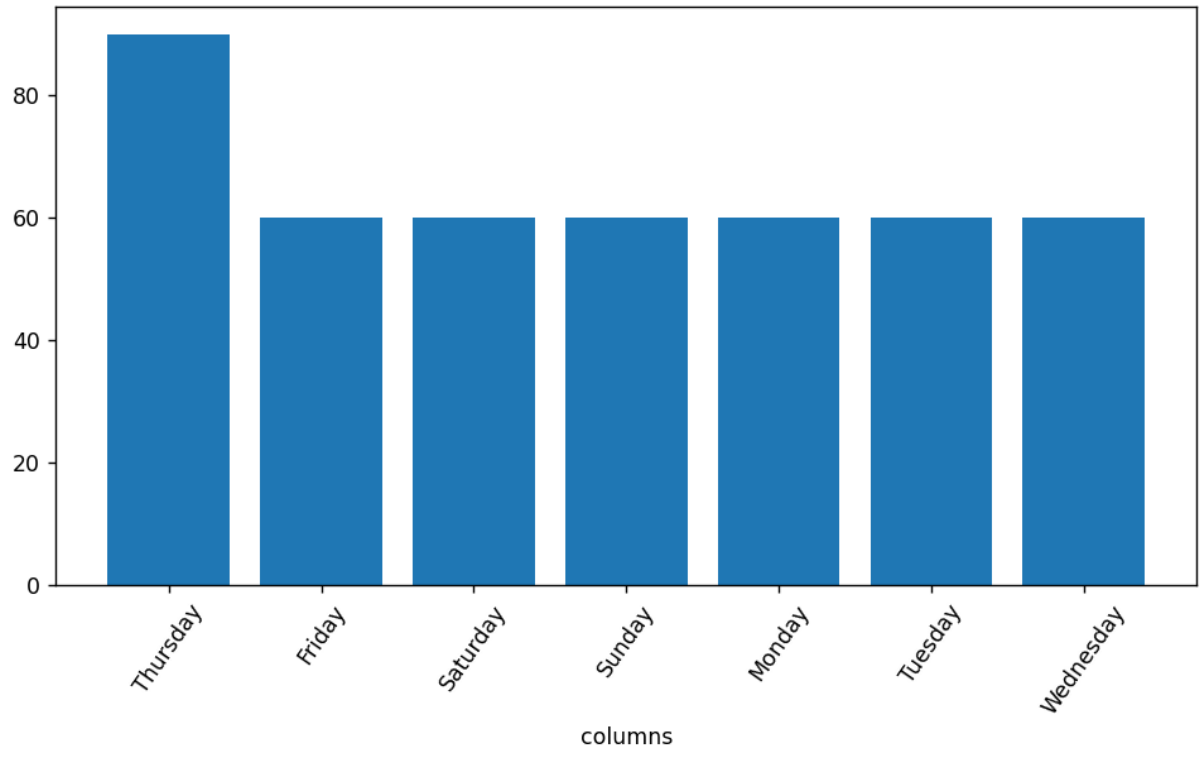

Histogram of Net Sales Value
frequencies
intervals
0
20
40
60
80
100
0
1000
3000
5000
7000
9000
11000
13000
15000
17000

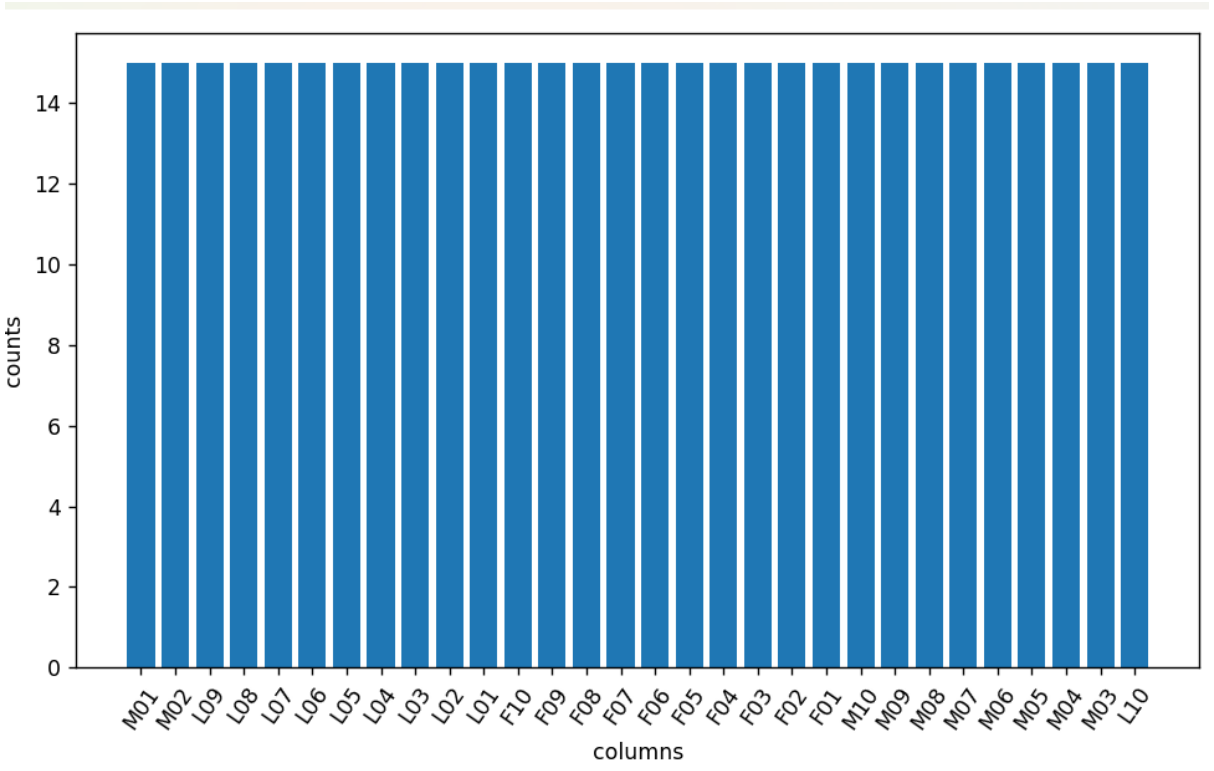
Box plots

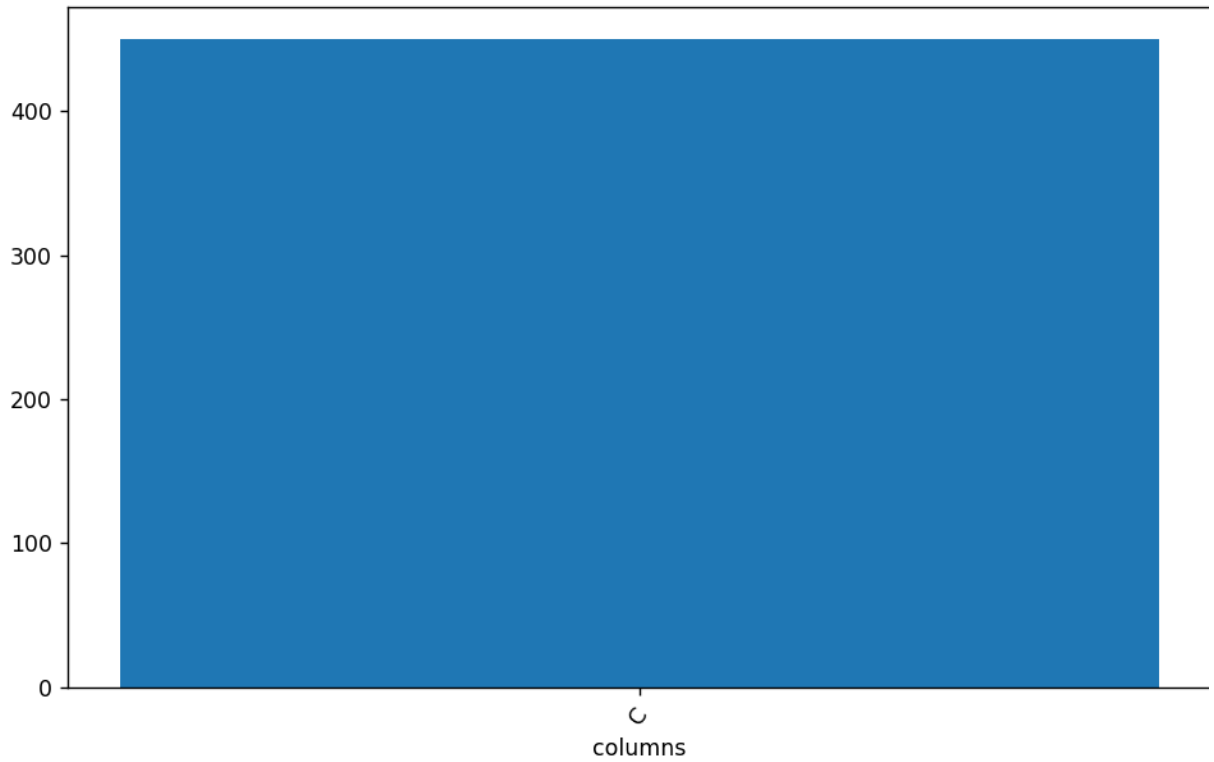
```
Outliers detected in each numerical column:

Volume: 44 outlier(s)
    Volume
0       15
10      13
30      11
40      29
50      13

Avg Price: 60 outlier(s)
    Avg Price
6       49100
7       54100
8       55100
9       60100
36      49100

Total Sales Value: 36 outlier(s)
    Total Sales Value
0              181500
6              147300
9              180300
30             133100
36             147300

Discount Rate (%): 45 outlier(s)
    Discount Rate (%)
3            6.935385
7            5.553719
8            7.410104
33           6.214888
36           5.252113

Discount Amount: 24 outlier(s)
    Discount Amount
0      21153.498820
6      13594.039719
9      17900.983733
30     17445.603828
38     13951.660194

Net Sales Value: 35 outlier(s)
    Net Sales Value
0     160346.501180
6     133705.960281
```

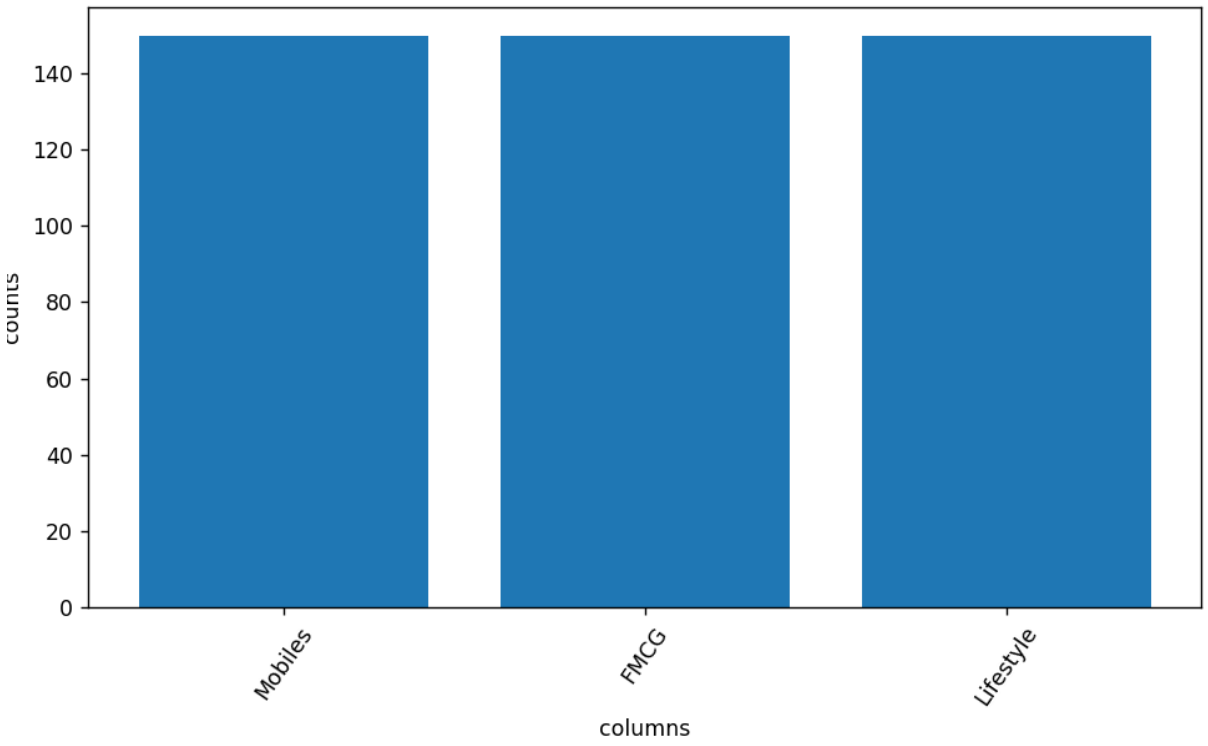

Box Plots for Numerical Variables
Volume
Avg Price
Total Sales Value
Discount Rate (%)
Discount Amount
Net Sales Value
0
25000
50000
75000
100000
125000
150000
175000
200000

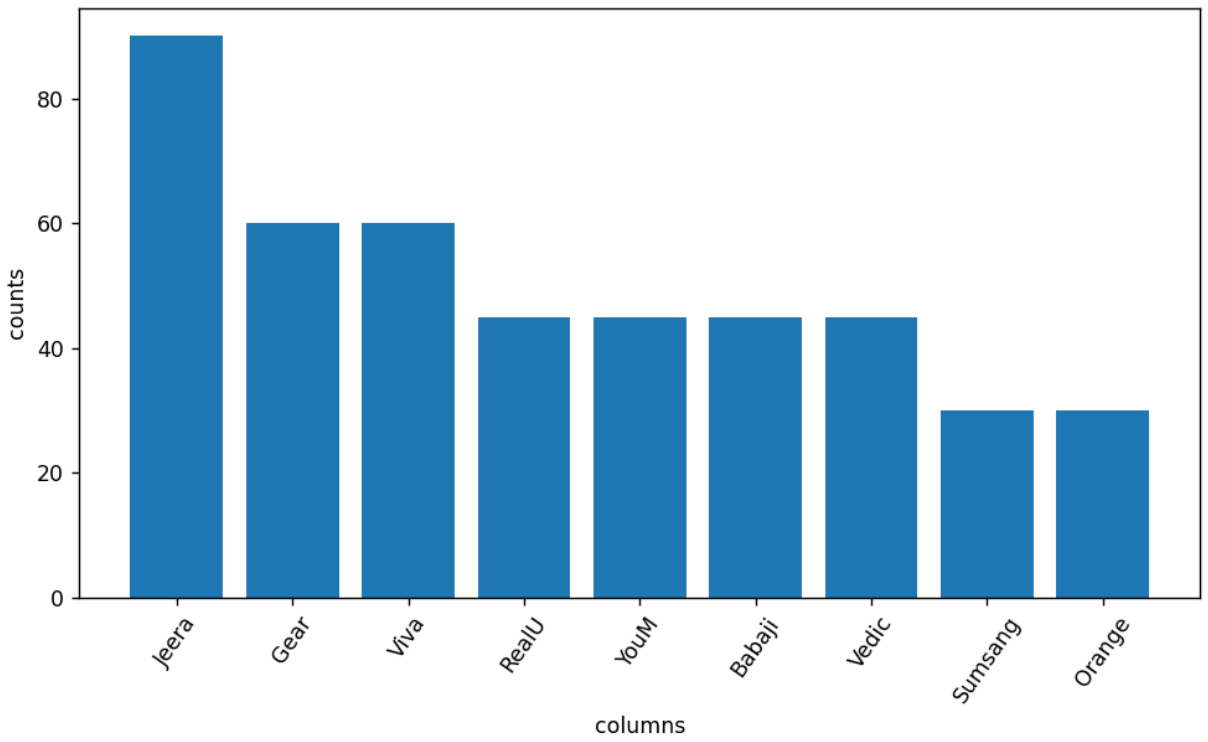
Bar chart

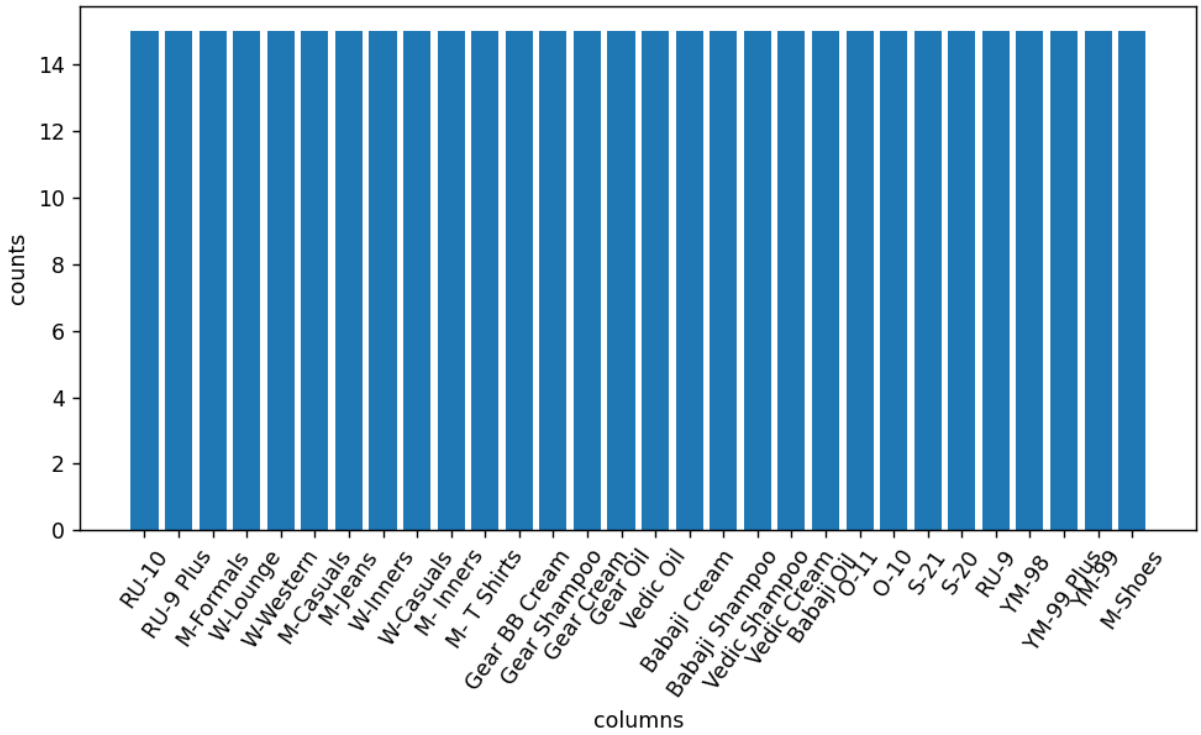

80
60
40
20
0
Thursday
Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
columns


counts
0
2
4
6
8
10
12
14
M01 M02 L09 L08 L07 L06 L05 L04 L03 L02 L01 F10 F09 F08 F07 F06 F05 F04 F03 F02 F01 M10 M09 M08 M07 M06 M05 M04 M03 L10
columns


0
100
200
300
400
C
columns


counts
0
20
40
60
80
100
120
140
Mobiles
FMCG
Lifestyle
columns


counts
0
20
40
60
80
Jeera
Gear
Viva
RealU
YouM
Babaji
Vedic
Sumsang
Orange
columns



Standardization of numerical varisbles

```
================== RESTART: D:\python\sprint3\jobsimulation_1.py ================
Original Data:
                      Date      Volume  ...  Discount Amount  Net Sales Value
count                  450  450.000000  ...       450.000000       450.000000
mean   2021-04-08 00:00:00    5.066667  ...      3346.499424     30466.336131
min    2021-04-01 00:00:00    1.000000  ...        69.177942       326.974801
25%    2021-04-04 00:00:00    3.000000  ...       460.459304      2202.208645
50%    2021-04-08 00:00:00    4.000000  ...       988.933733      4677.788059
75%    2021-04-12 00:00:00    6.000000  ...      5316.495427     47847.912852
max    2021-04-15 00:00:00   31.000000  ...     25738.022194    179507.479049
std                    NaN    4.231602  ...      4509.902963     46358.656624

[8 rows x 7 columns]

z_score Data:
 count    4.500000e+02
mean     4.736952e-17
std      1.001113e+00
min     -6.508581e-01
25%     -6.103625e-01
50%     -5.569025e-01
75%      3.753543e-01
max      3.218536e+00
Name: Net Sales Value, dtype: float64
```

Conversion of Categorical Data into Dummy Variables

```
================ RESTART: D:\python\sprint3\jobsimulation_1.py ================
          Date  Volume  Avg Price  ...  Model_YM-98  Model_YM-99  Model_YM-99 Plus
0   2021-04-01      15      12100  ...            0            0                 0
1   2021-04-01      10      10100  ...            0            0                 0
2   2021-04-01       7      16100  ...            0            1                 0
3   2021-04-01       6      20100  ...            0            0                 1
4   2021-04-01       3       8100  ...            1            0                 0
..         ...     ...        ...  ...          ...          ...               ...
445 2021-04-15       2       1300  ...            0            0                 0
446 2021-04-15       6       2600  ...            0            0                 0
447 2021-04-15       2       1600  ...            0            0                 0
448 2021-04-15       3       1900  ...            0            0                 0
449 2021-04-15       1       3100  ...            0            0                 0

[450 rows x 87 columns]
```

Task2

```
================ RESTART: D:\python\sprint3\jobsimulation_1.py ================
t_critical: -1.7958848187036696
Reject the null hypothesis, there is sufficient evidence to support the analyst
```