```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sbs
from scipy.stats import zscore
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.linear_model import LinearRegression
```

```
file=pd.read_csv("Sales_Transactions_Dataset_Weekly.csv")
file.head()
```

	Product_Code	W0	W1	W2	W3	W4	W5	W6	W7	W8	...	Normalized 42	Normalized 43	N
0	P1	11	12	10	8	13	12	14	21	6	...	0.06	0.22	
1	P2	7	6	3	2	7	1	6	3	3	...	0.20	0.40	
2	P3	7	11	8	9	10	8	7	13	12	...	0.27	1.00	
3	P4	12	8	13	5	9	6	9	13	13	...	0.41	0.47	
4	P5	8	5	13	11	6	7	9	14	9	...	0.27	0.53	

5 rows × 107 columns

```
file.describe()
```

	W0	W1	W2	W3	W4	W5	
count	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	811.000
mean	8.902589	9.129470	9.389642	9.717633	9.574599	9.466091	9.720
std	12.067163	12.564766	13.045073	13.553294	13.095765	12.823195	13.347
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
50%	3.000000	3.000000	3.000000	4.000000	4.000000	3.000000	4.000
75%	12.000000	12.000000	12.000000	13.000000	13.000000	12.500000	13.000
max	54.000000	53.000000	56.000000	59.000000	61.000000	52.000000	56.000

8 rows × 106 columns

```
file.columns
```

```
Index(['Product_Code', 'W0', 'W1', 'W2', 'W3', 'W4', 'W5', 'W6', 'W7', 'W8',
       ...
       'Normalized 42', 'Normalized 43', 'Normalized 44', 'Normalized 45',
       'Normalized 46', 'Normalized 47', 'Normalized 48', 'Normalized 49',
       'Normalized 50', 'Normalized 51'],
      dtype='object', length=107)
```

```
file.dtypes
```

```
Product_Code      object
W0                 int64
W1                 int64
W2                 int64
W3                 int64
                  ...
Normalized 47    float64
Normalized 48    float64
Normalized 49    float64
Normalized 50    float64
Normalized 51    float64
Length: 107, dtype: object
```

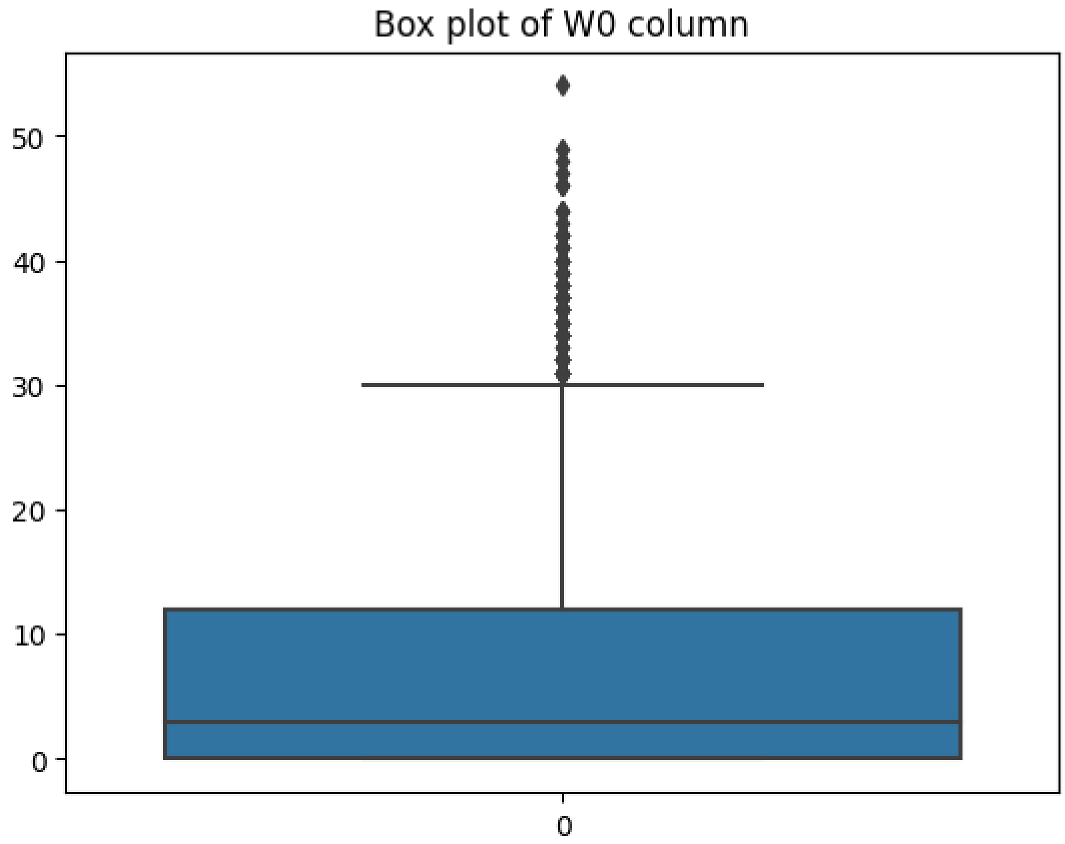
```
file.shape
```

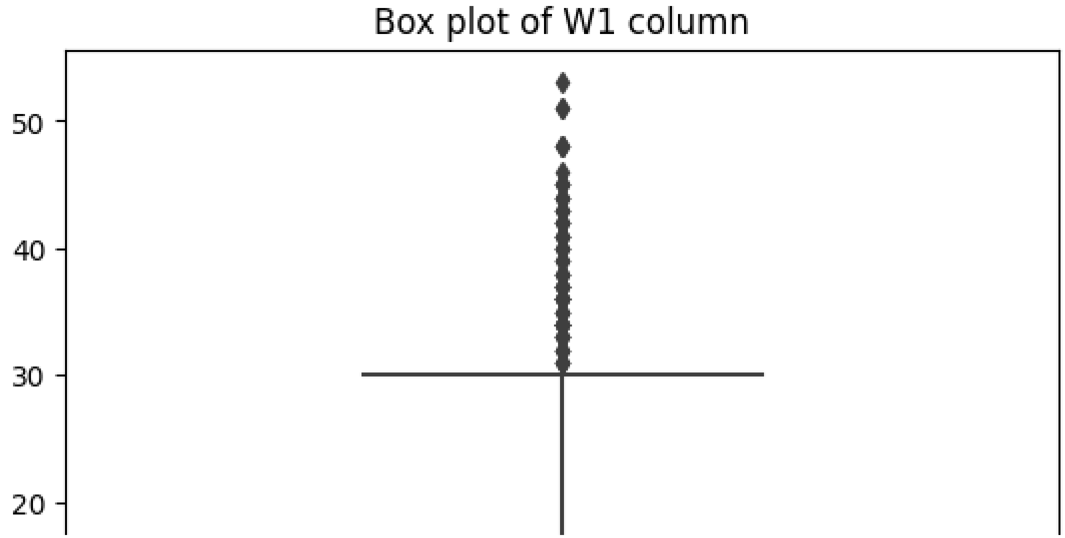
```
(811, 107)
```

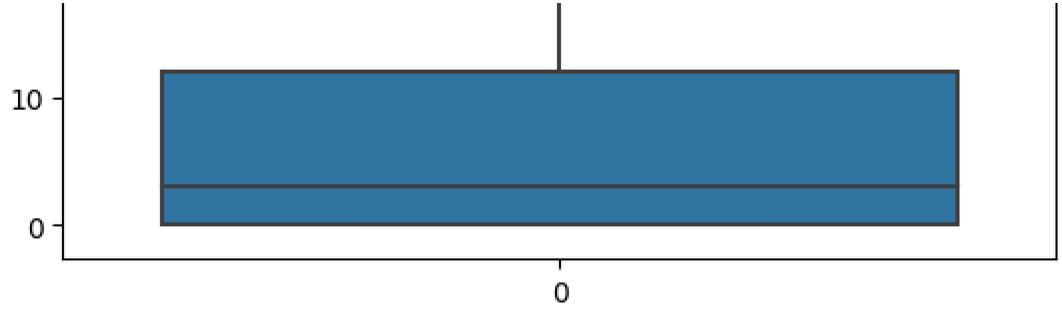
```
file.isnull().sum()
```

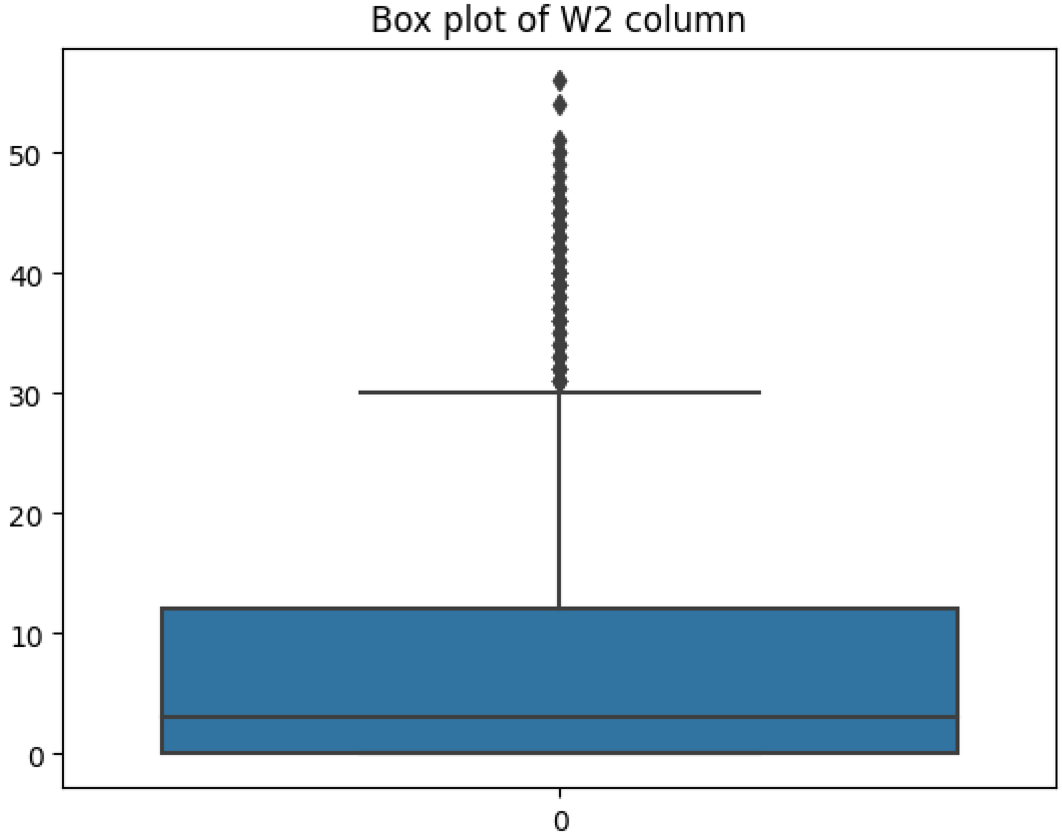
```
Product_Code     0
W0               0
W1               0
W2               0
W3               0
                ..
Normalized 47    0
Normalized 48    0
Normalized 49    0
Normalized 50    0
Normalized 51    0
Length: 107, dtype: int64
```

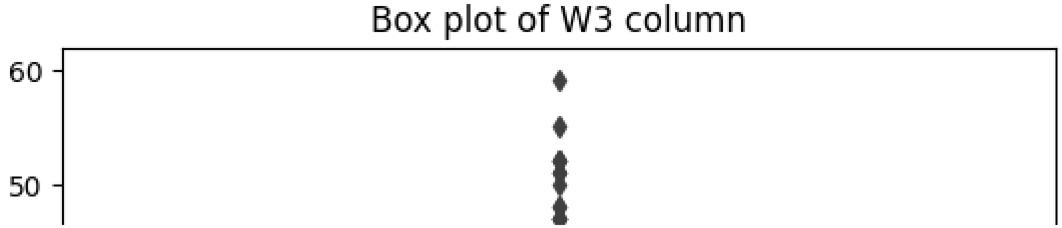
```
for i in file.select_dtypes(['int','float']).columns:
      plt.title("Box plot of "+str(i)+' column')
      sbs.boxplot(file[i])
      plt.show()
```

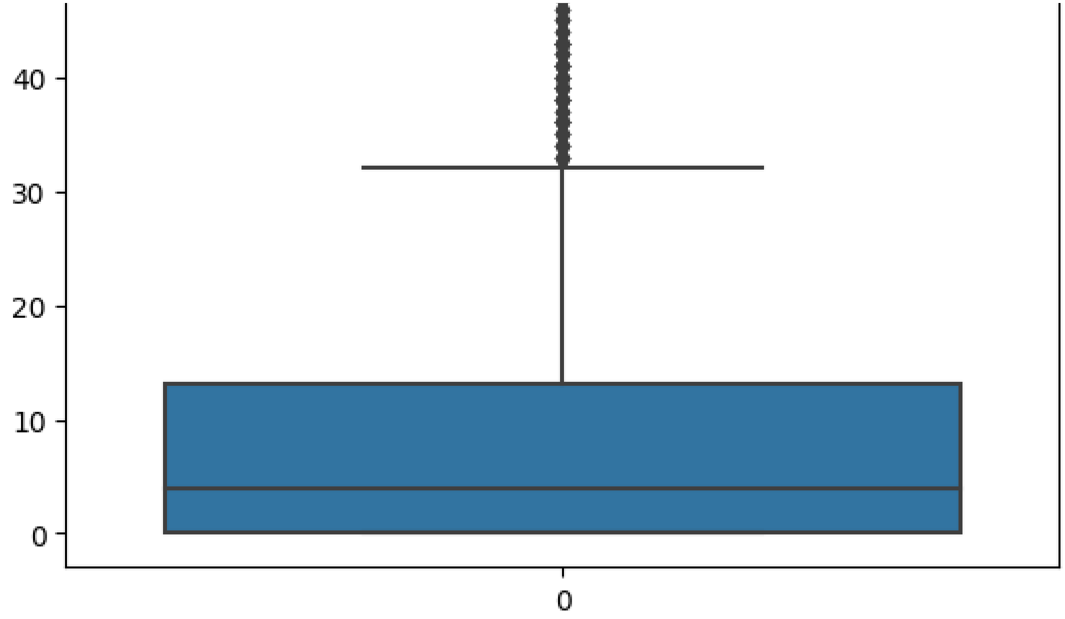

Box plot of W0 column
50
40
30
20
10
0
0

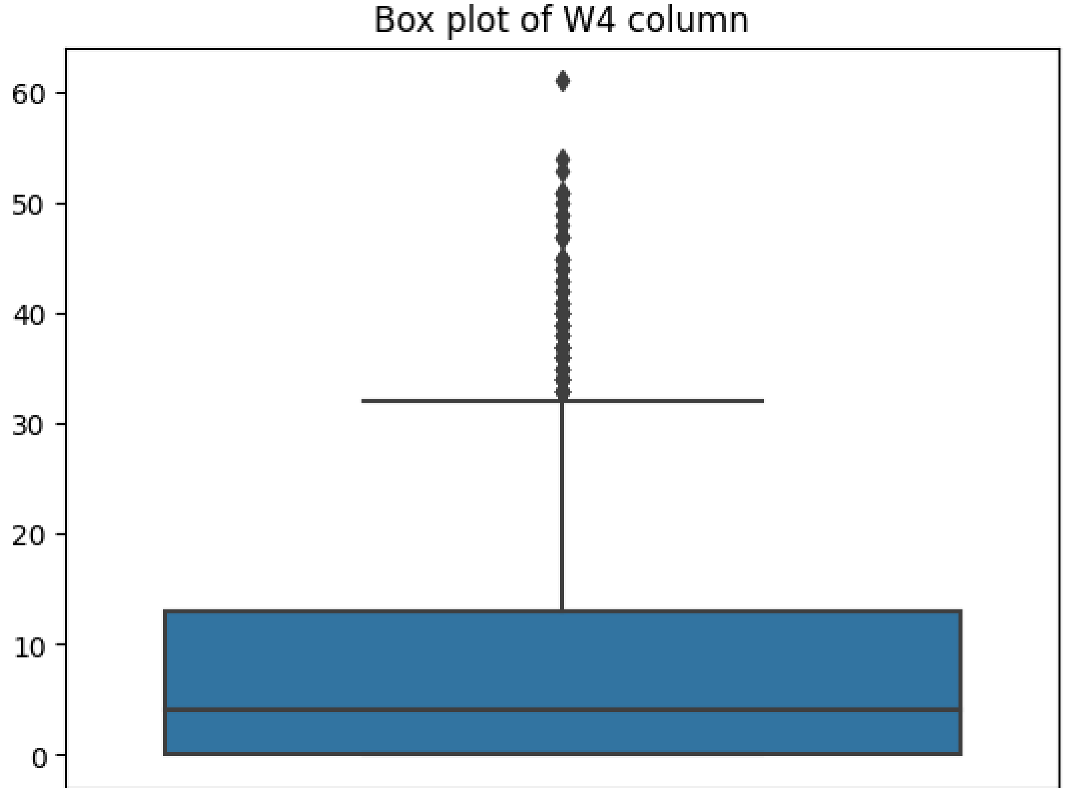

Box plot of W1 column
50
40
30
20

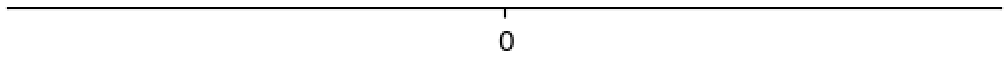

10
0
0

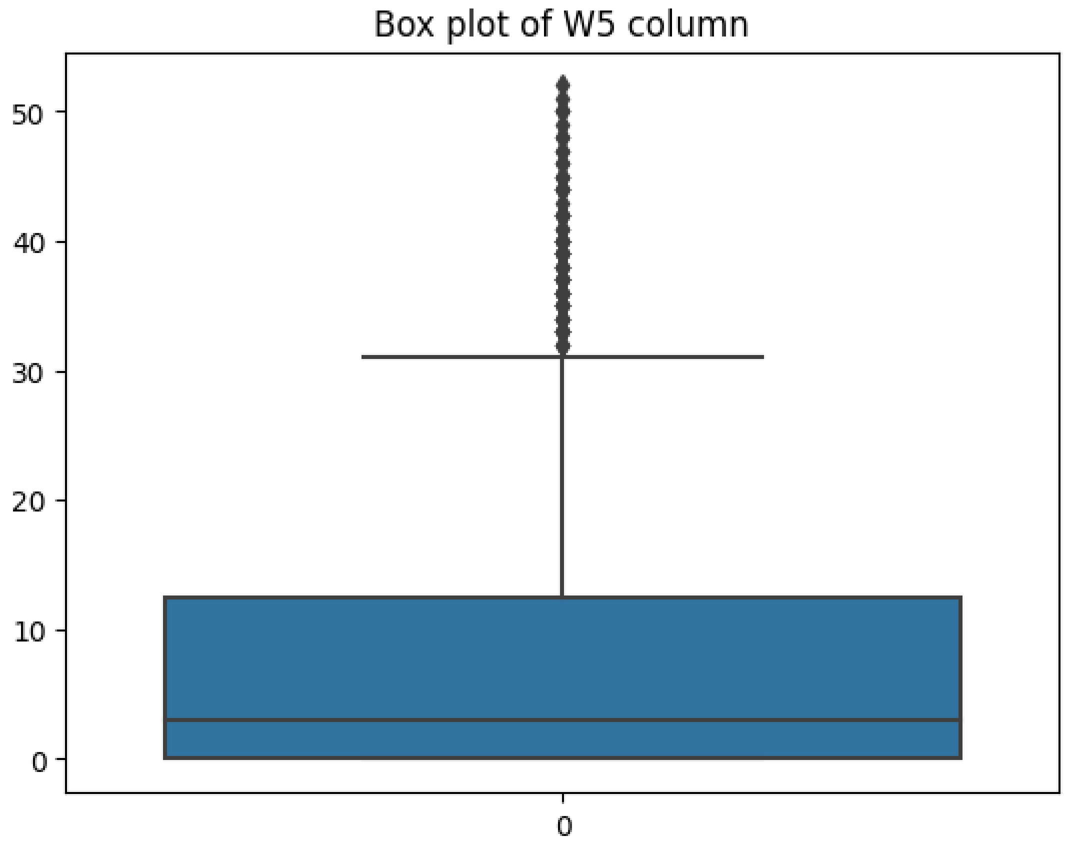

Box plot of W2 column
50
40
30
20
10
0
0

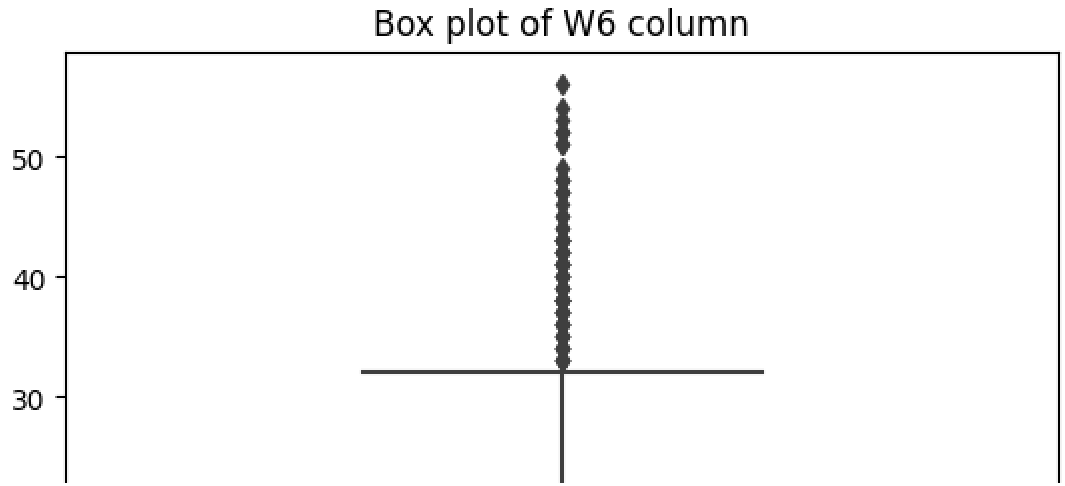

Box plot of W3 column
60
50

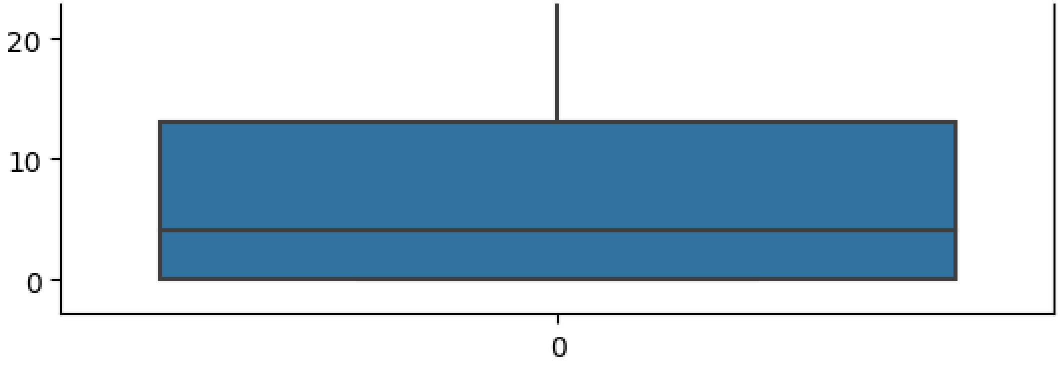

40
30
20
10
0
0

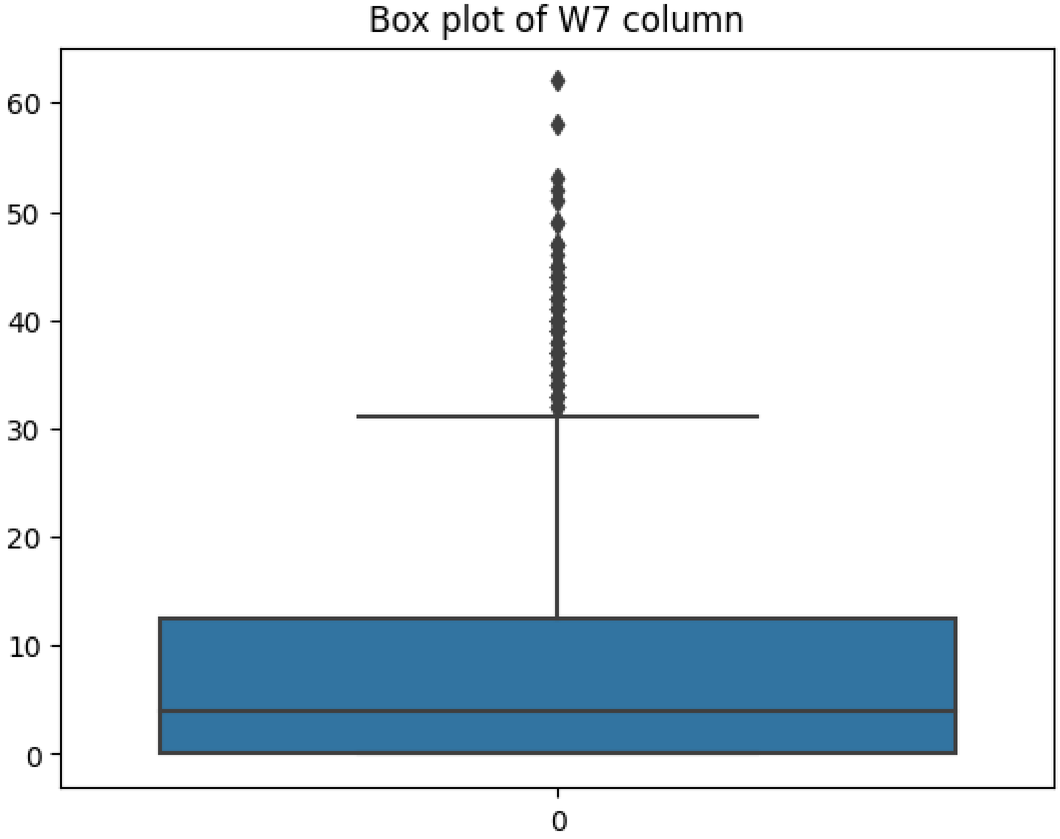

Box plot of W4 column
60
50
40
30
20
10
0

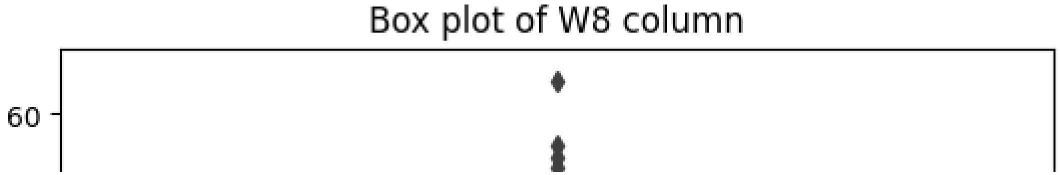


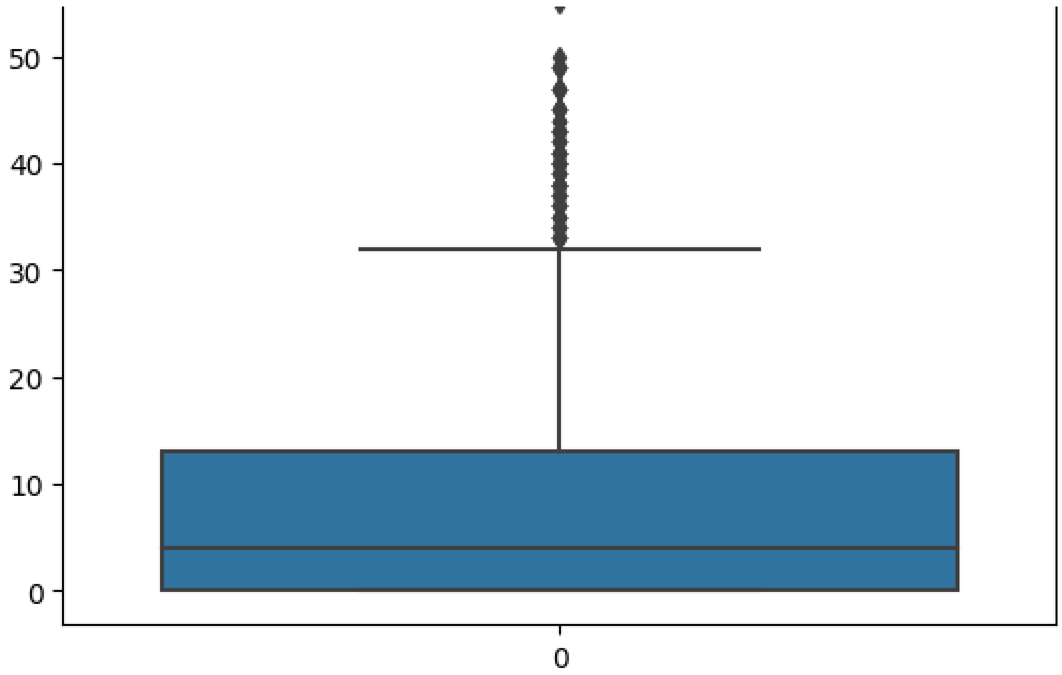

Box plot of W5 column


Box plot of W6 column

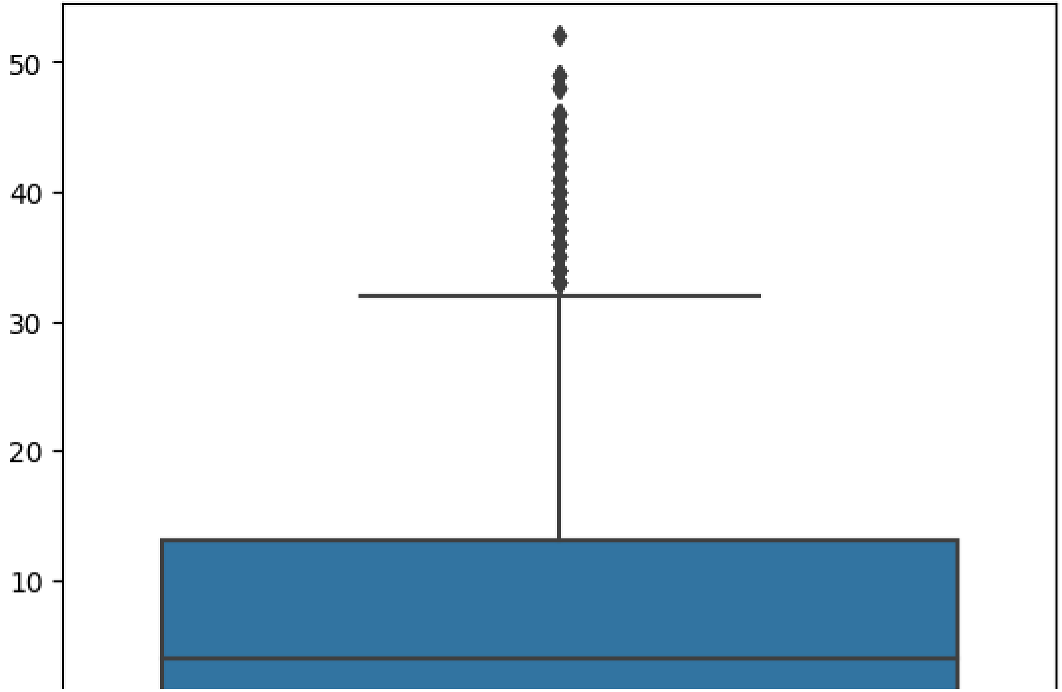

20
10
0
0

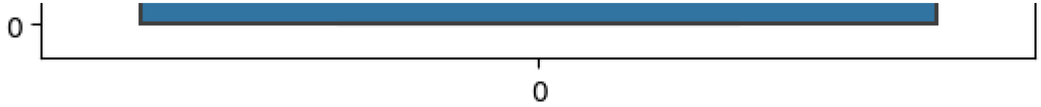

Box plot of W7 column
60
50
40
30
20
10
0
0

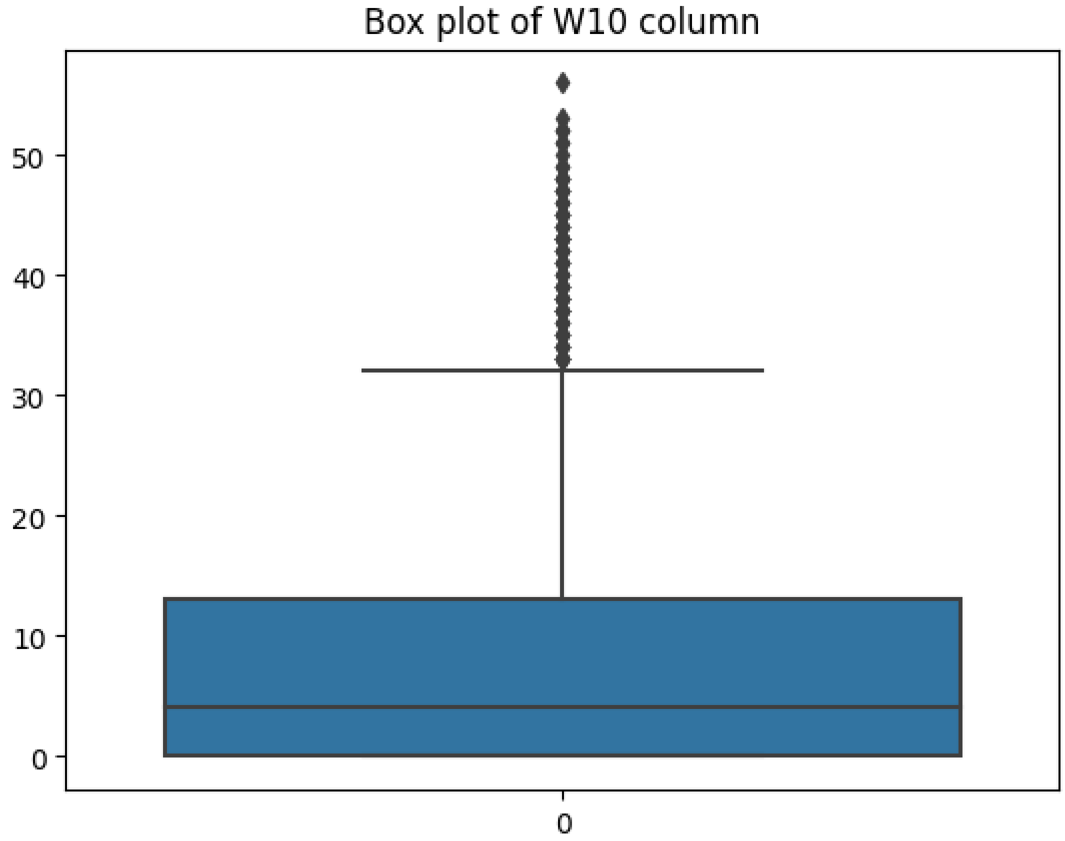

Box plot of W8 column
60

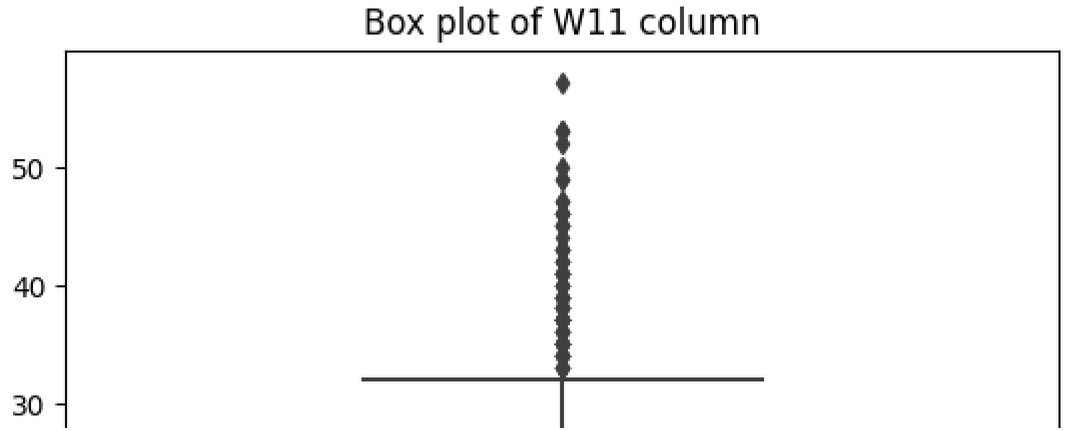


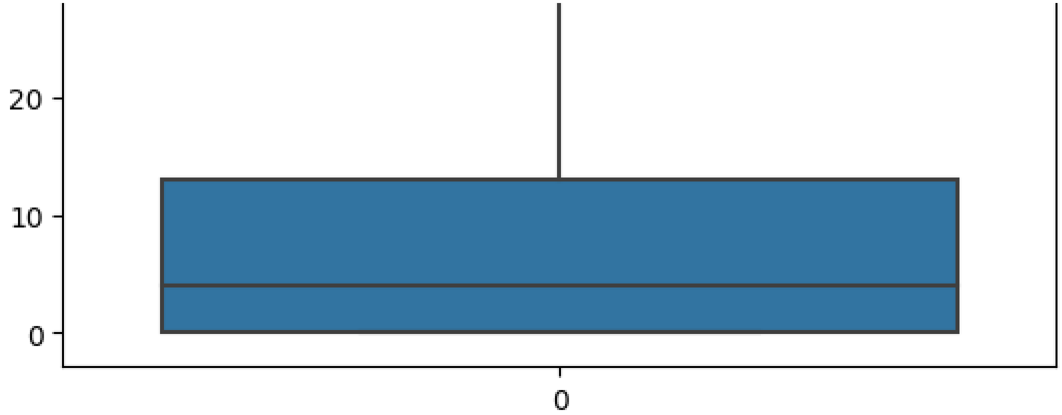
Box plot of W9 column

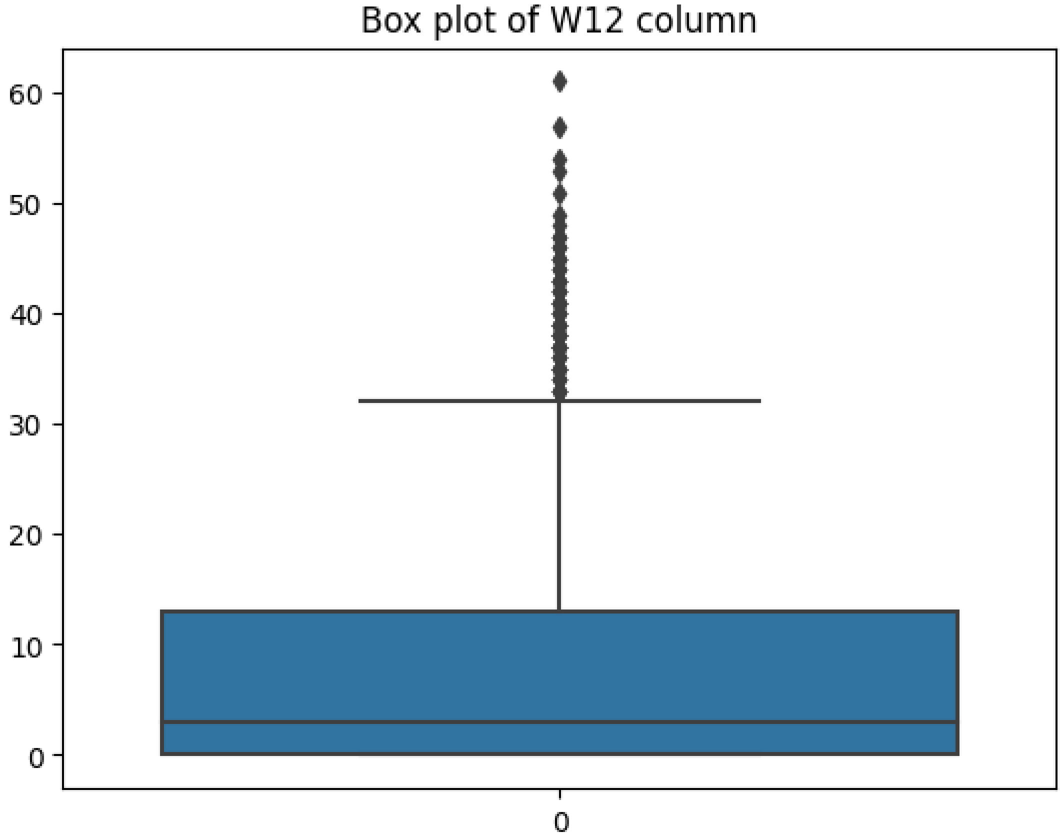


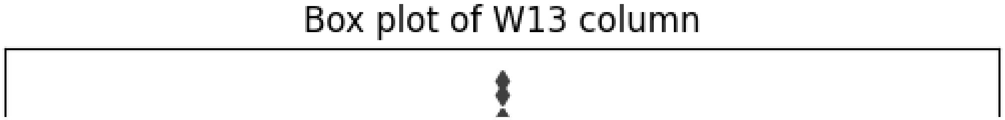

0
0

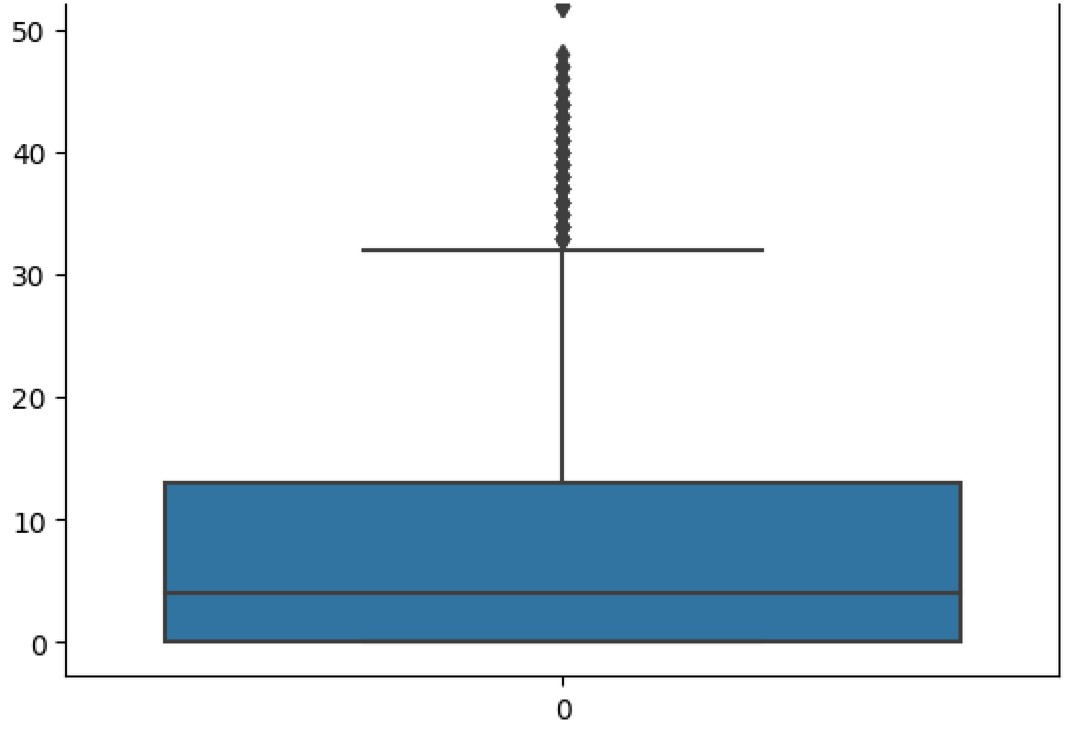

Box plot of W10 column
50
40
30
20
10
0
0

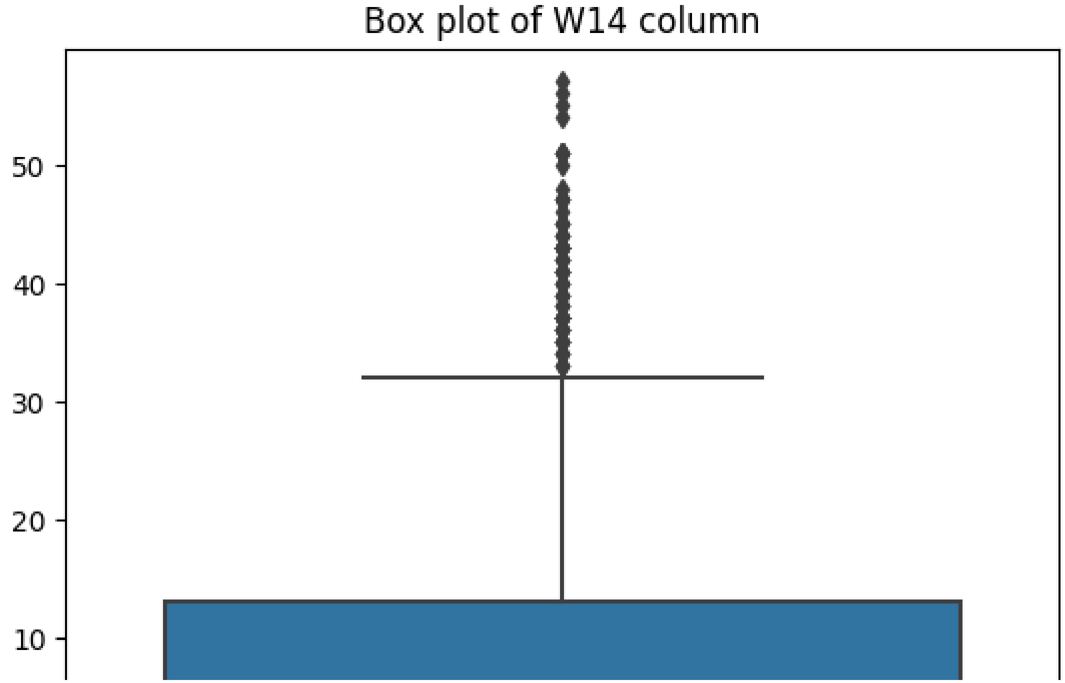

Box plot of W11 column
50
40
30

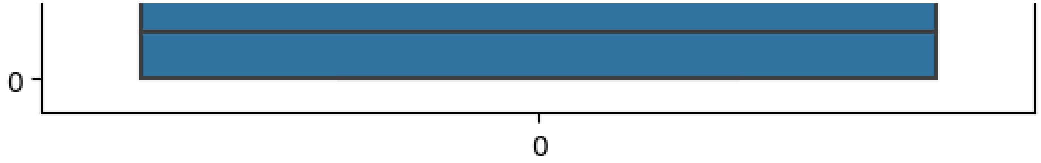

20
10
0
0

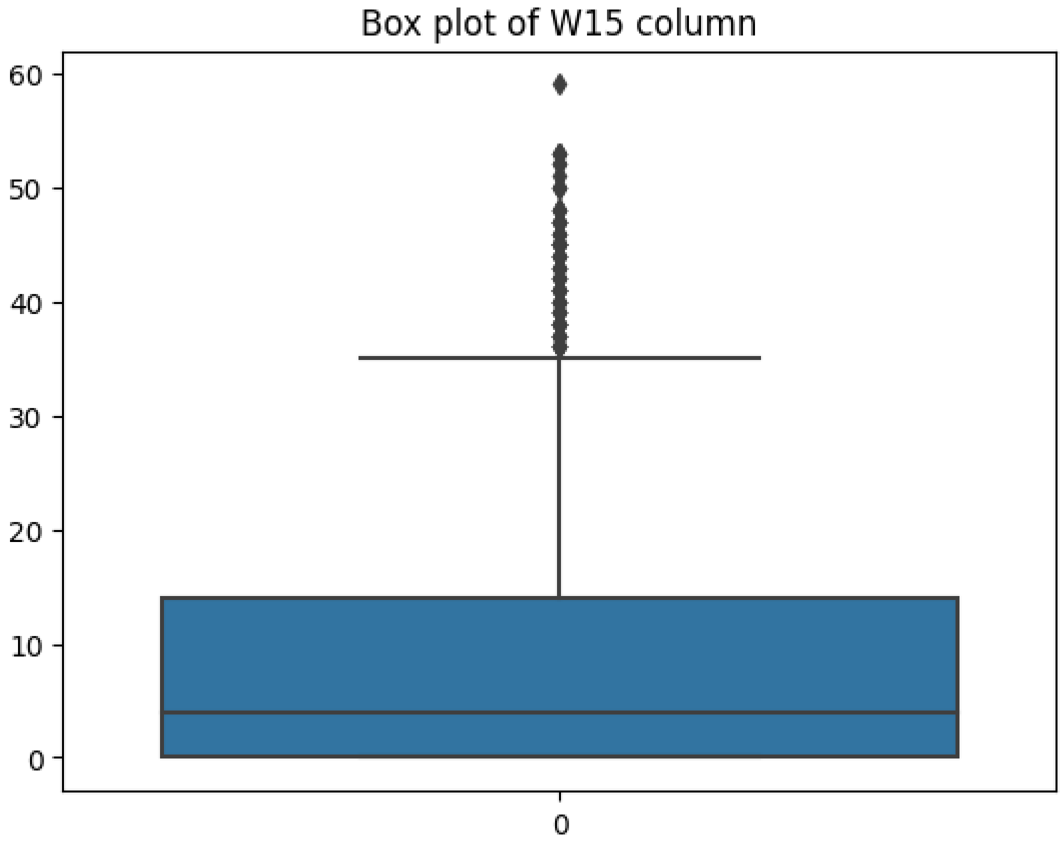

Box plot of W12 column
60
50
40
30
20
10
0
0

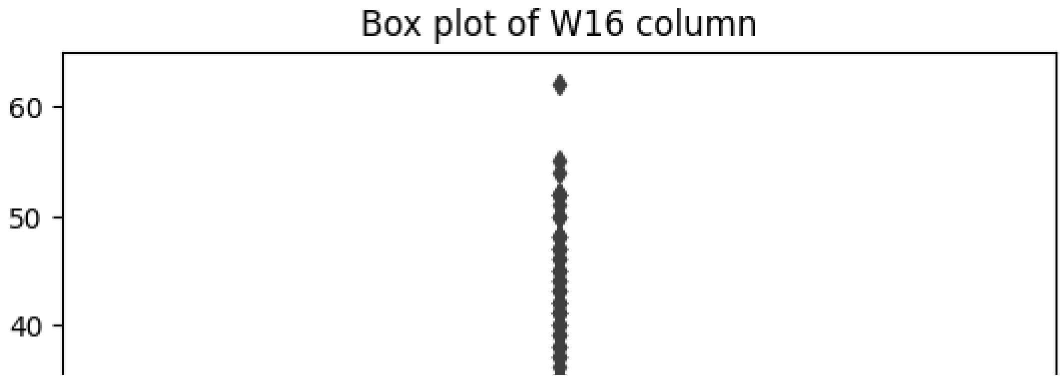

Box plot of W13 column

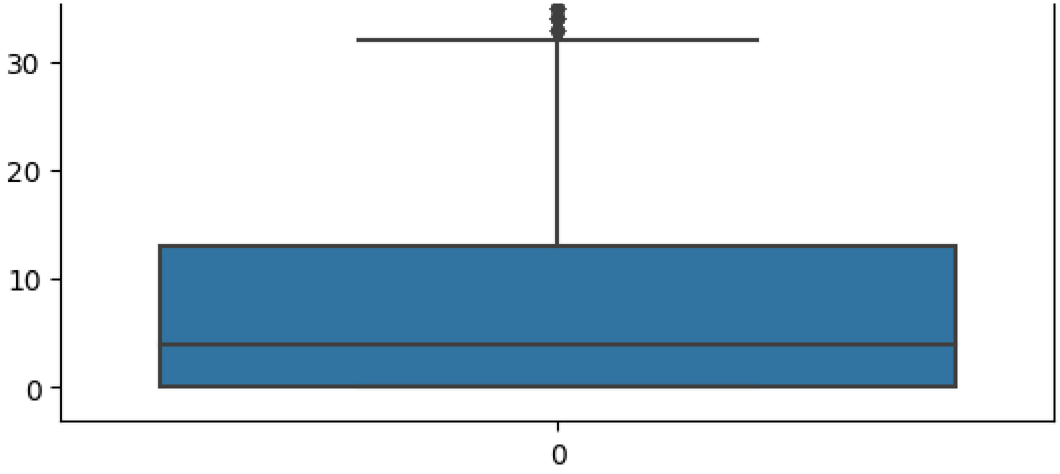

50
40
30
20
10
0
0


Box plot of W14 column
50
40
30
20
10

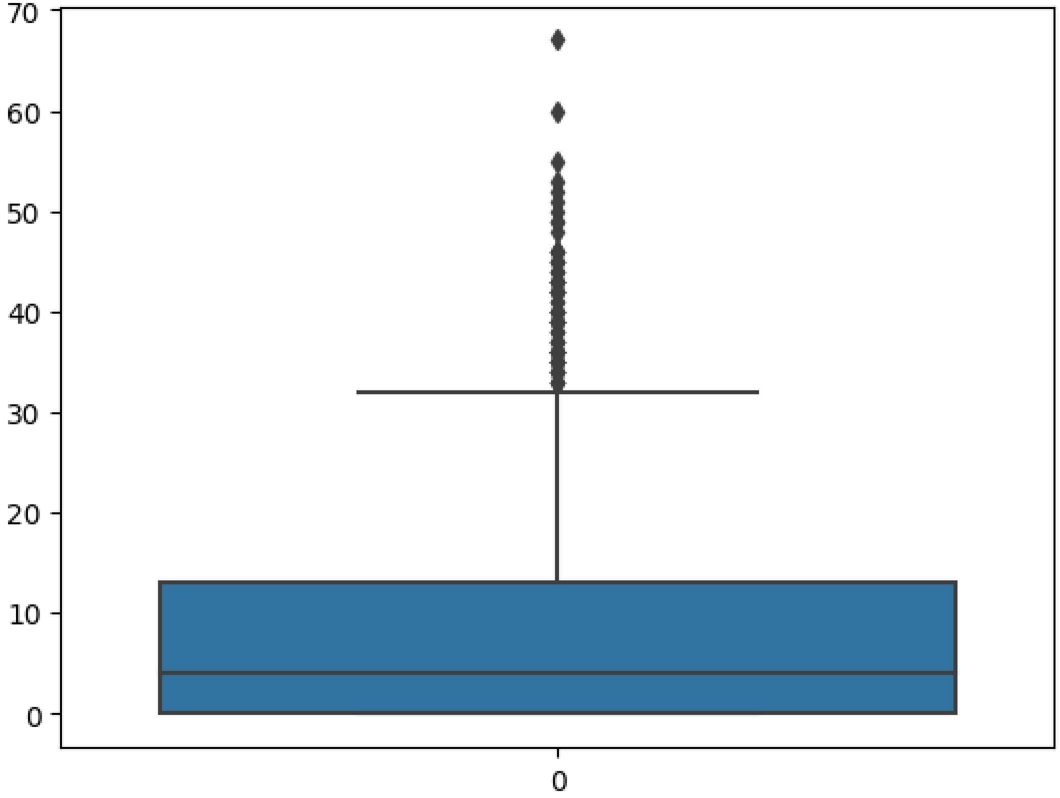

0
0


Box plot of W15 column
60
50
40
30
20
10
0
0

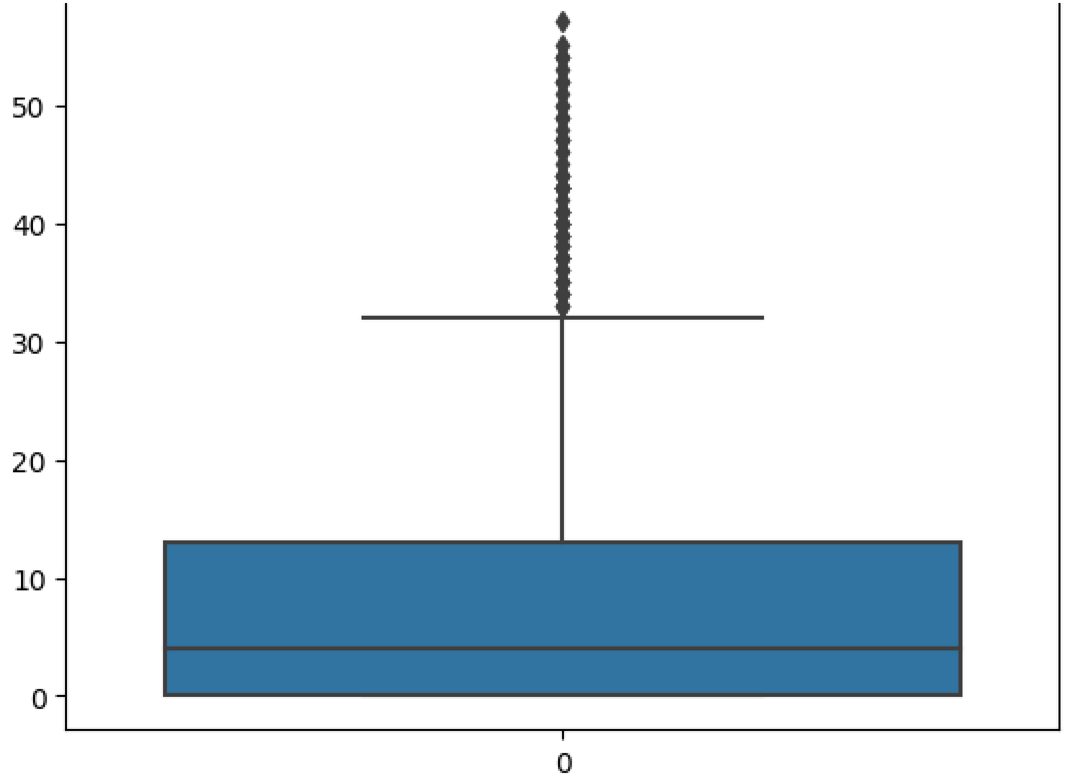

Box plot of W16 column
60
50
40



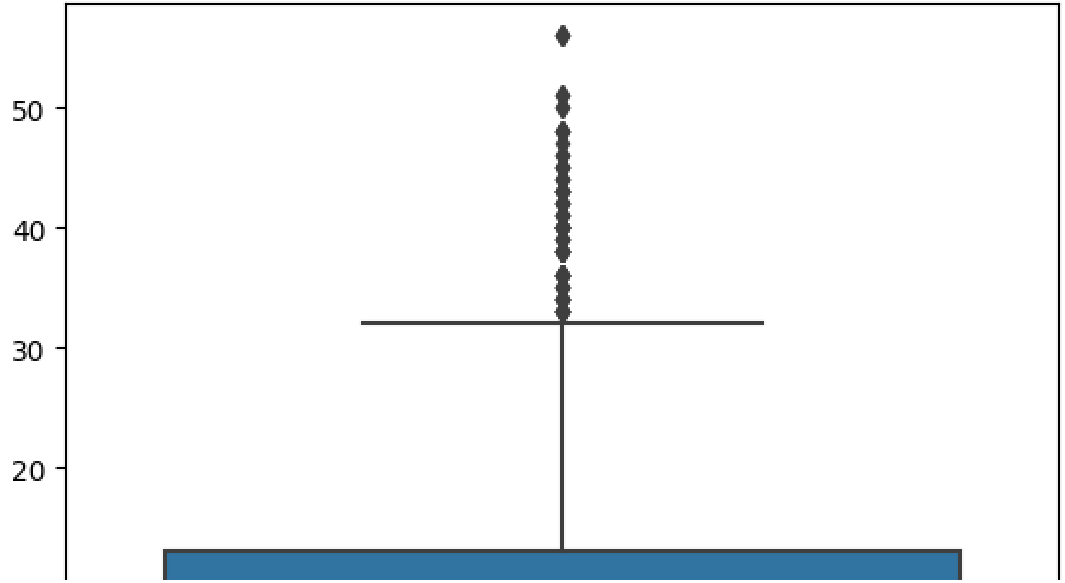
Box plot of W17 column

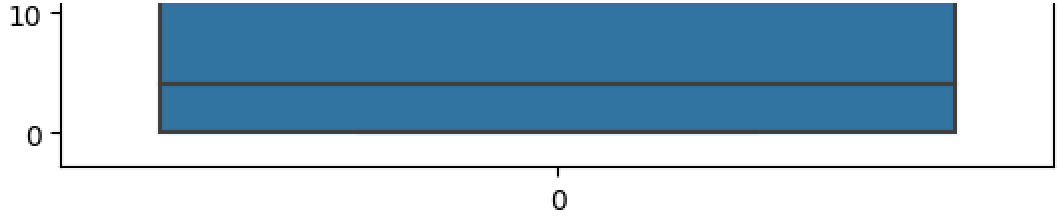


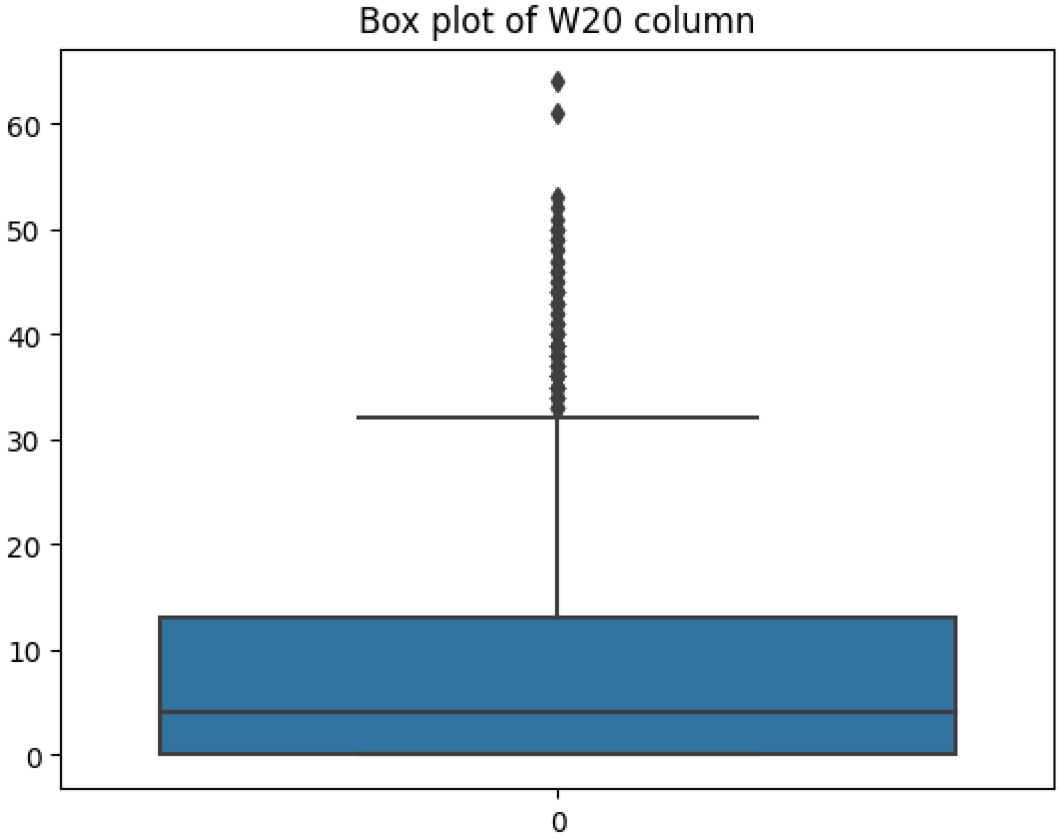
Box plot of W18 column

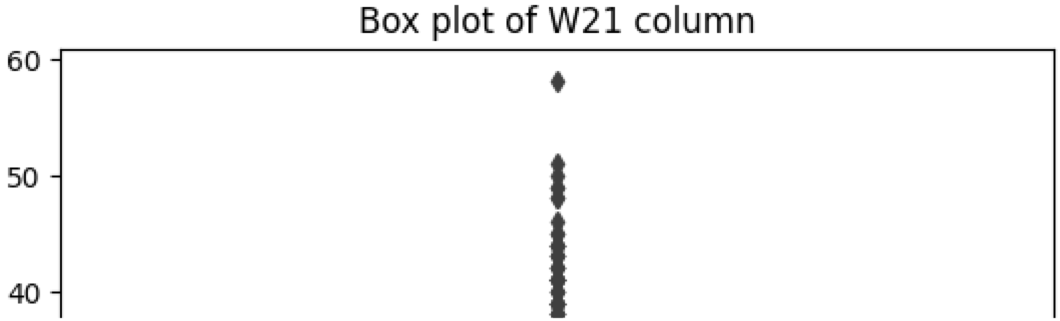


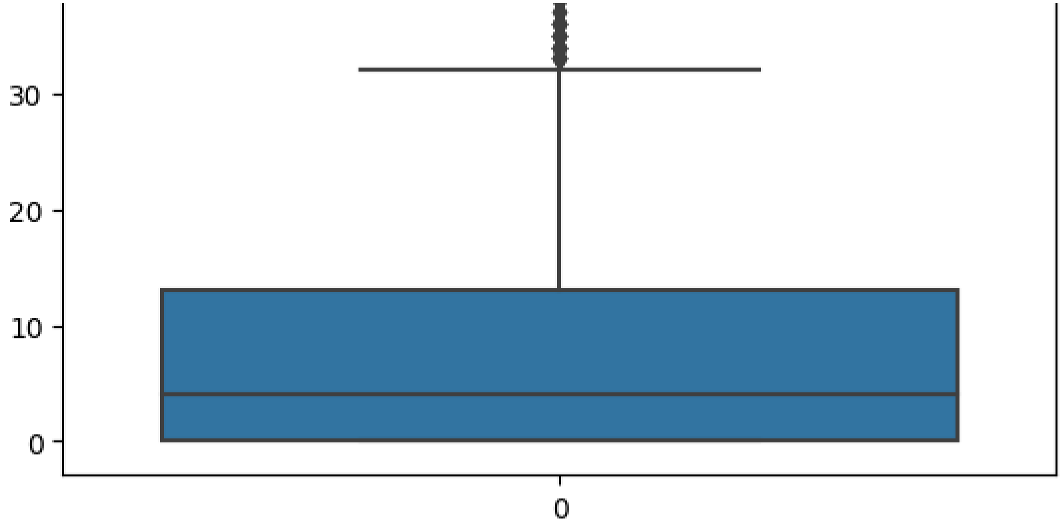
Box plot of W19 column

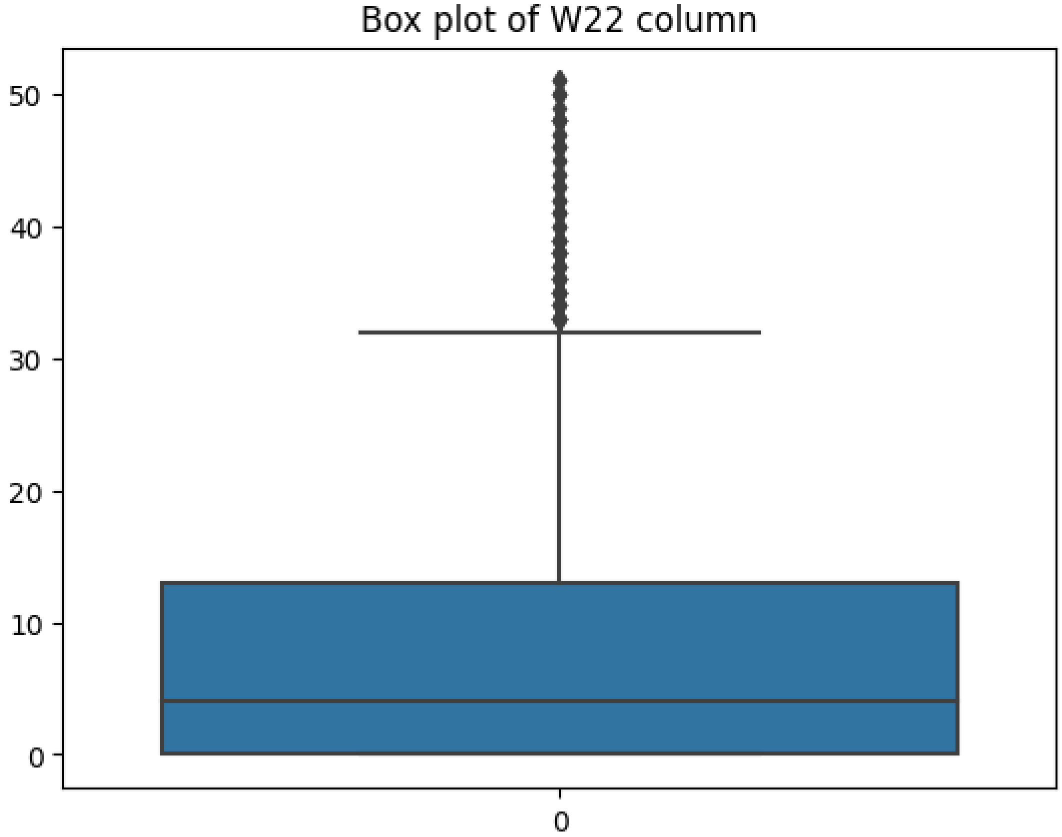


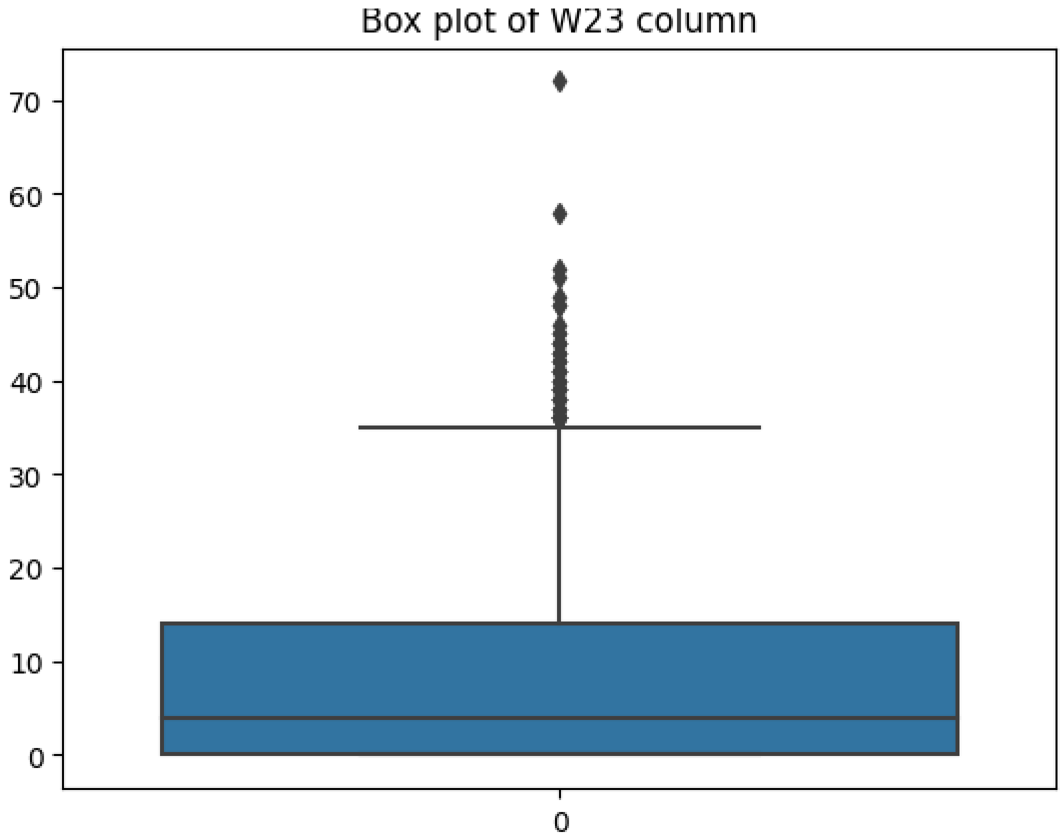


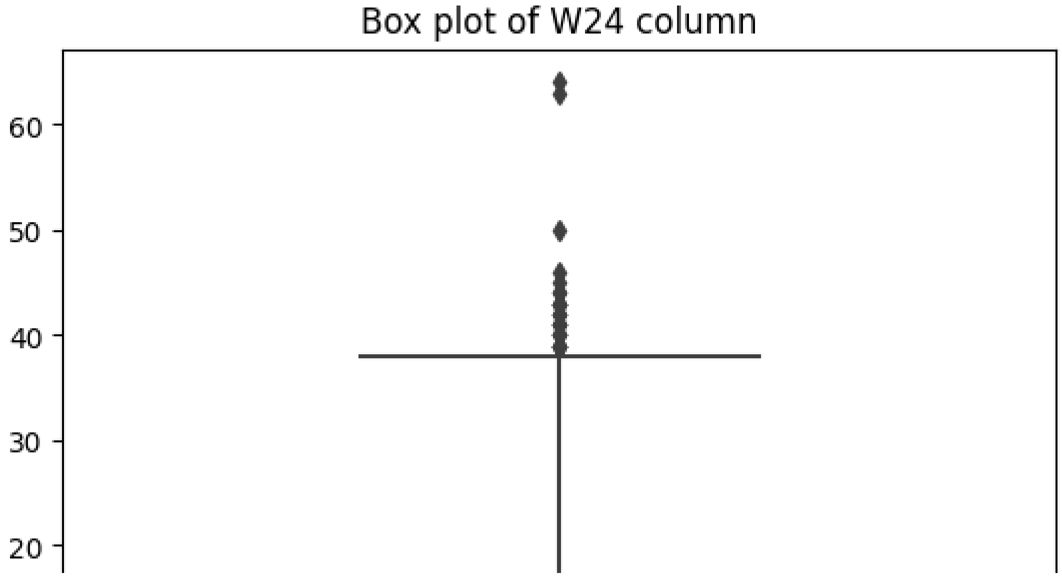

Box plot of W20 column

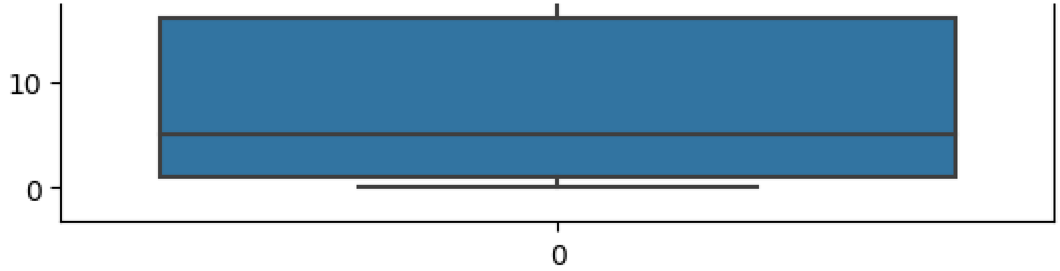

Box plot of W21 column

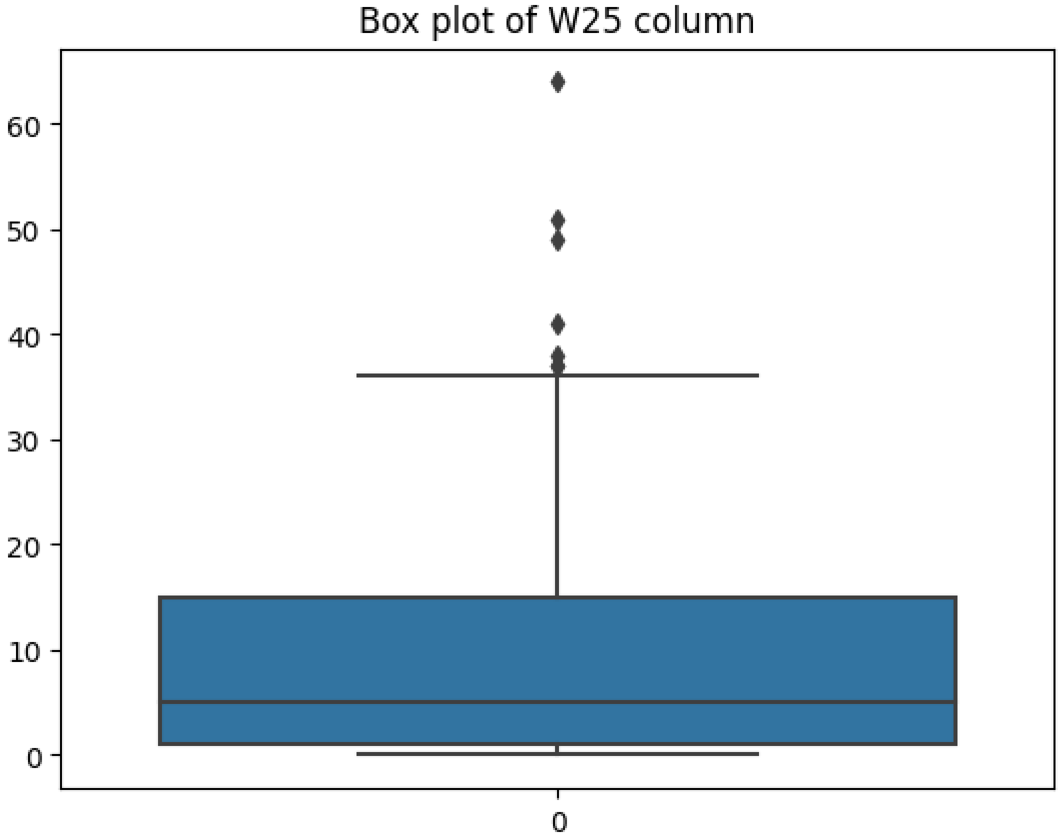

30
20
10
0
0

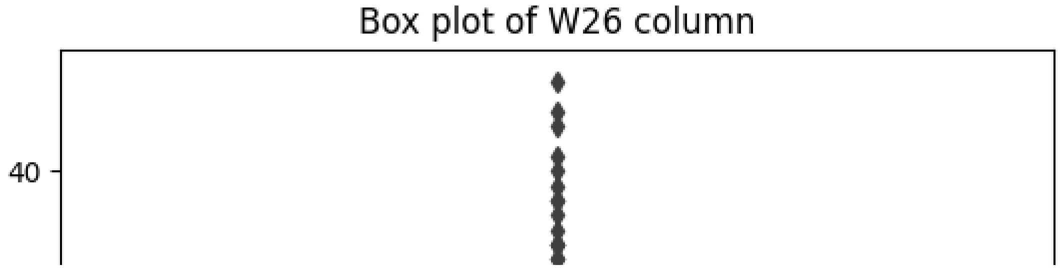

Box plot of W22 column
50
40
30
20
10
0
0

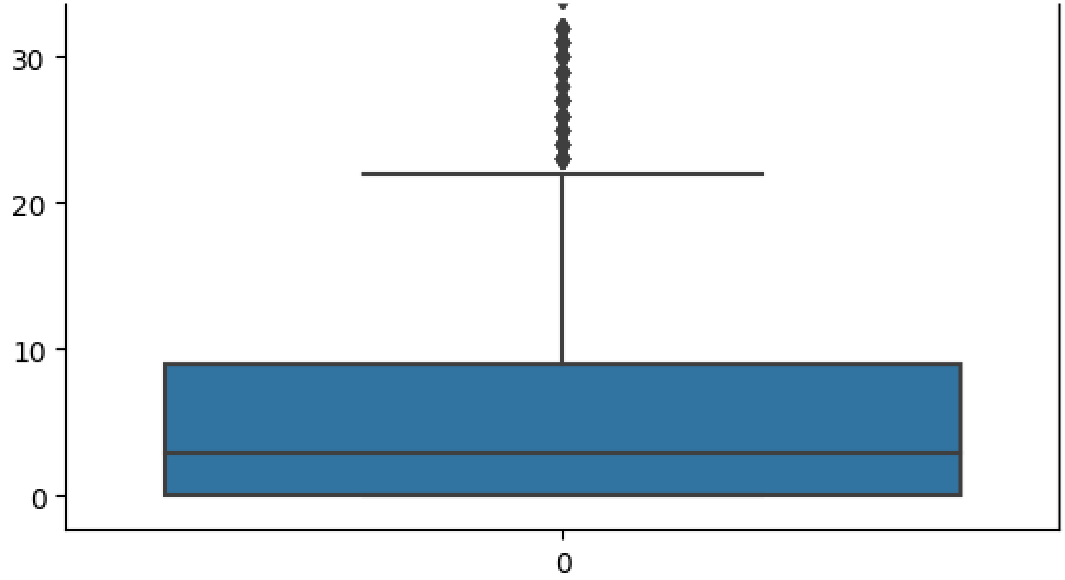

Box plot of W23 column
70
60
50
40
30
20
10
0
0

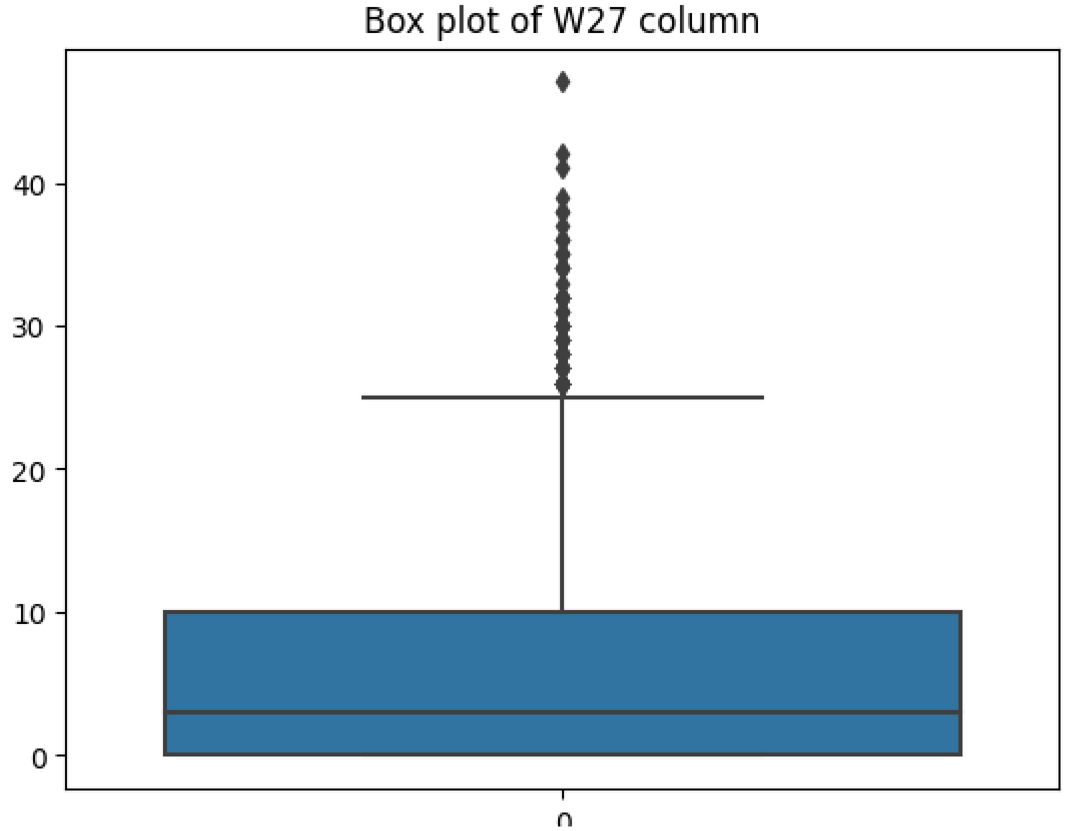

Box plot of W24 column
60
50
40
30
20

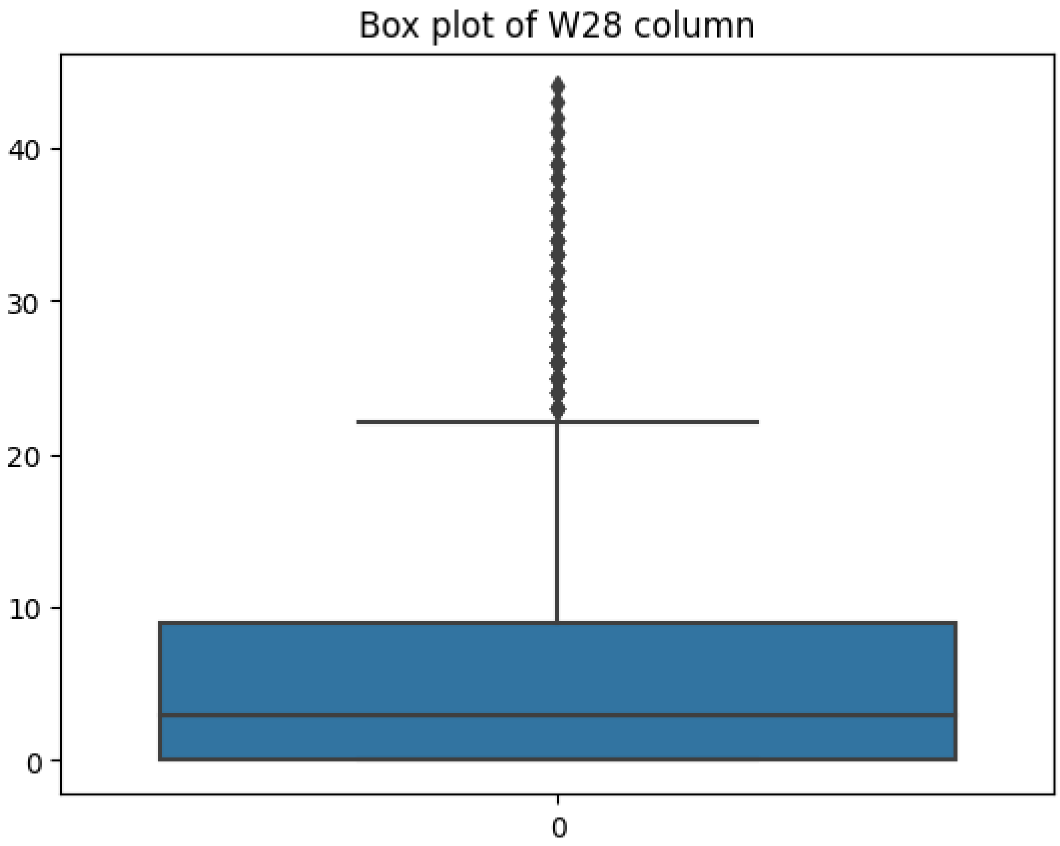

10
0
0

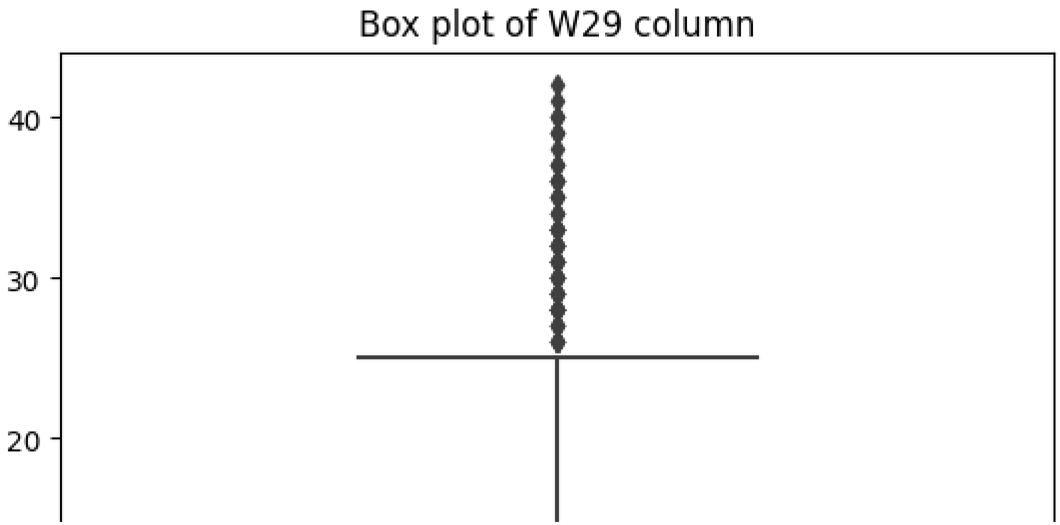

Box plot of W25 column
60
50
40
30
20
10
0
0

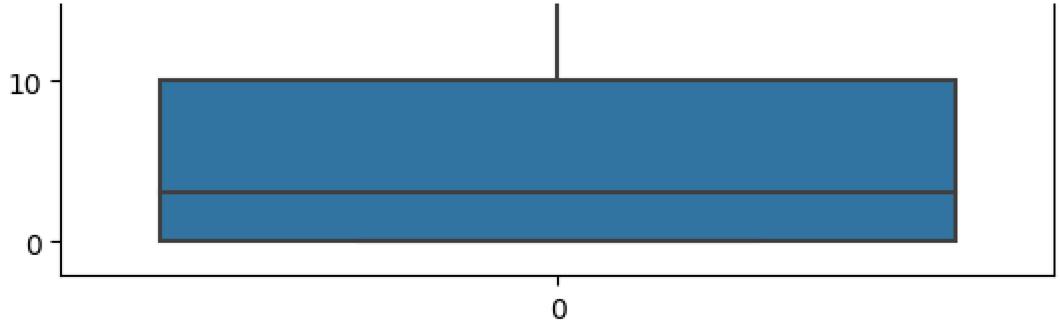

Box plot of W26 column
40

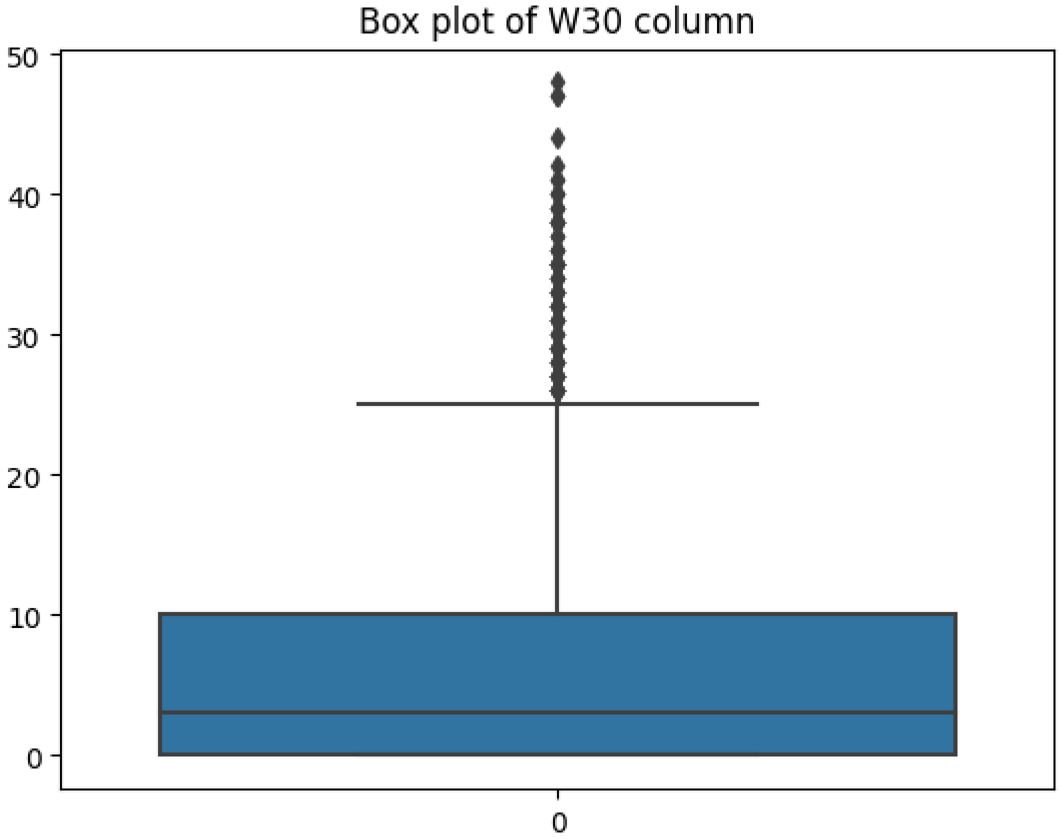

30
20
10
0
0

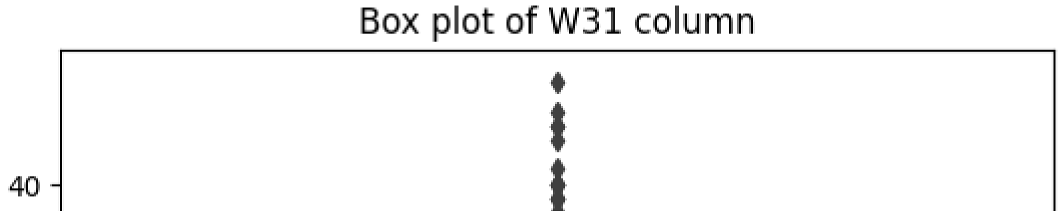

Box plot of W27 column
40
30
20
10
0
0

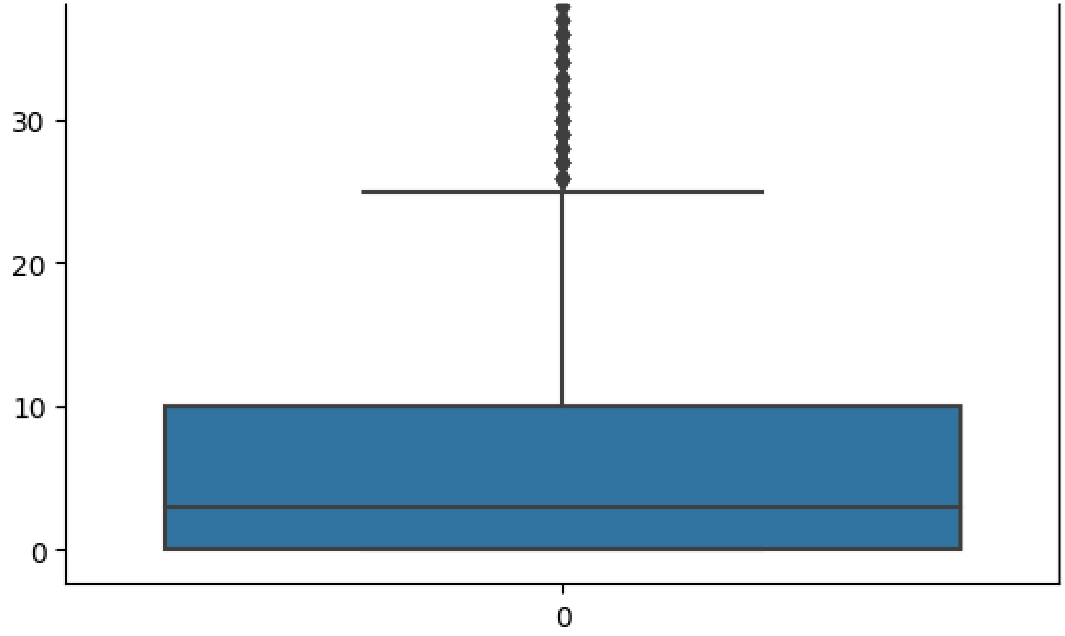

Box plot of W28 column
40
30
20
10
0
0

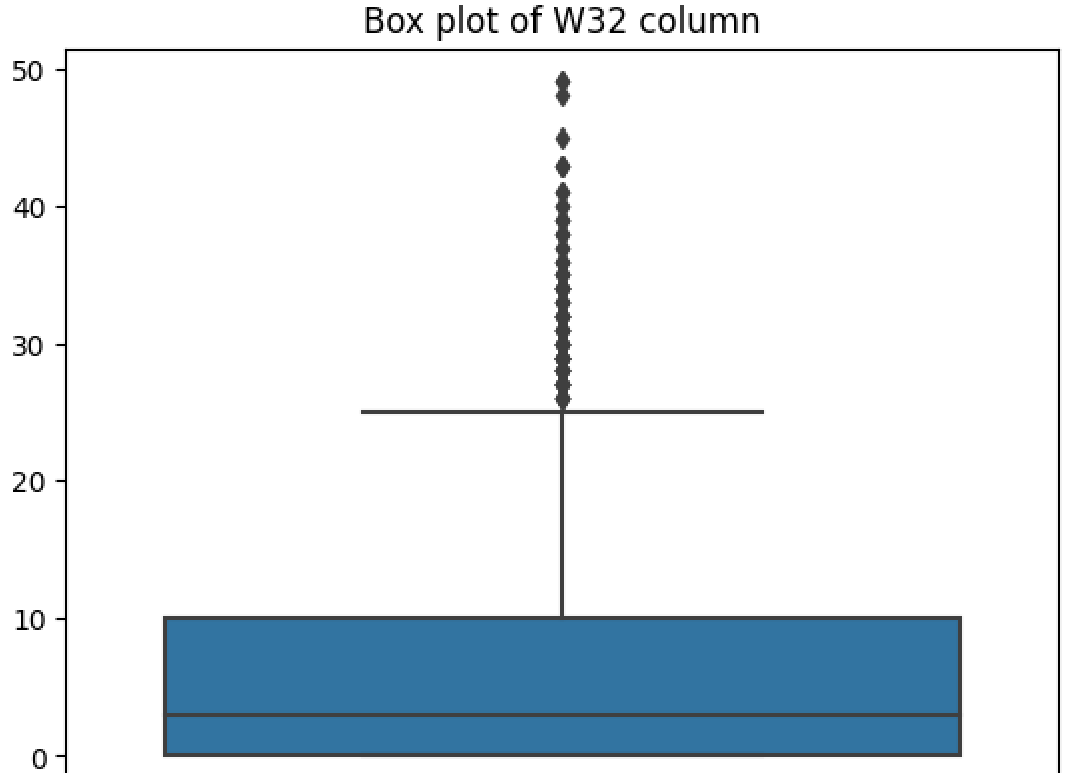

Box plot of W29 column
40
30
20

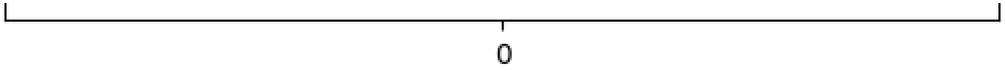

10
0
0

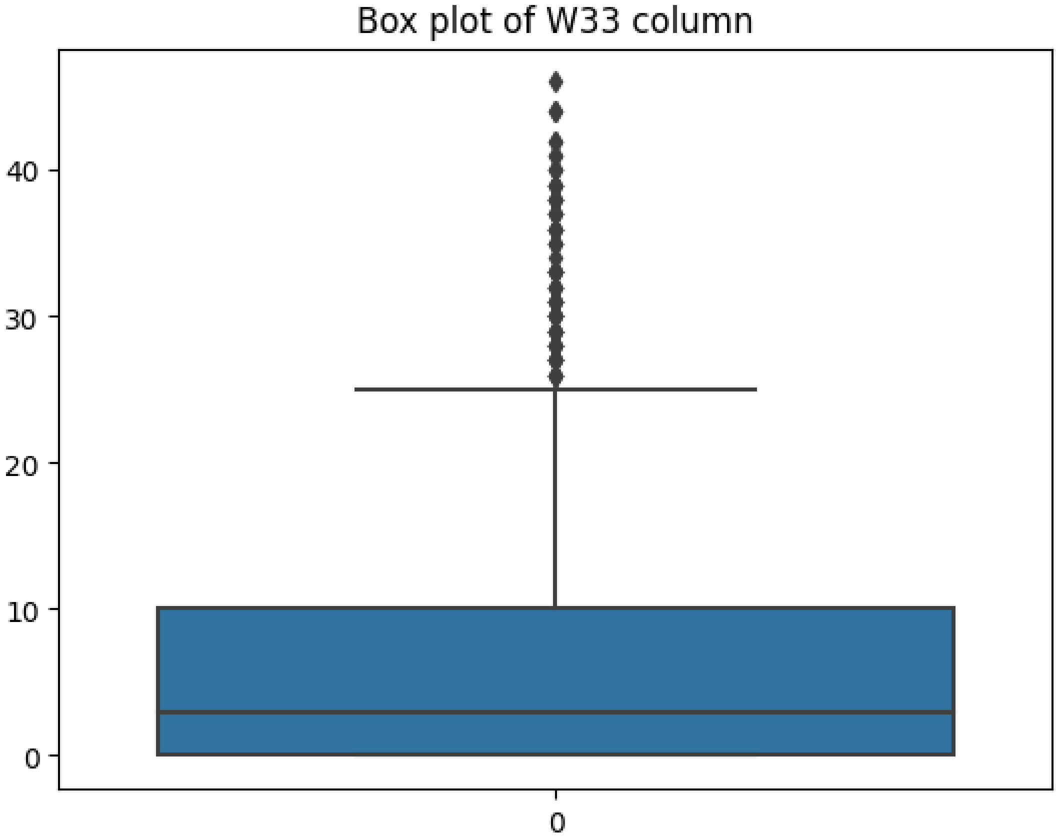

Box plot of W30 column
50
40
30
20
10
0
0

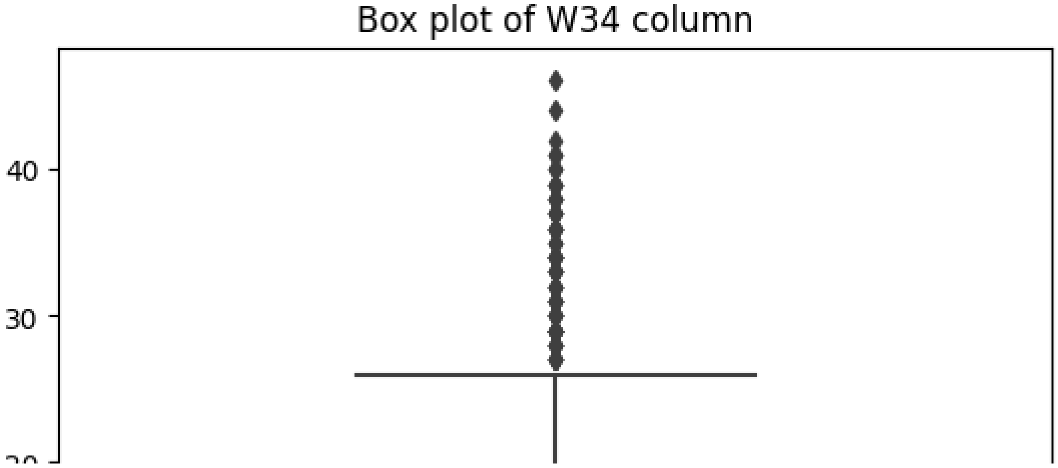

Box plot of W31 column
40

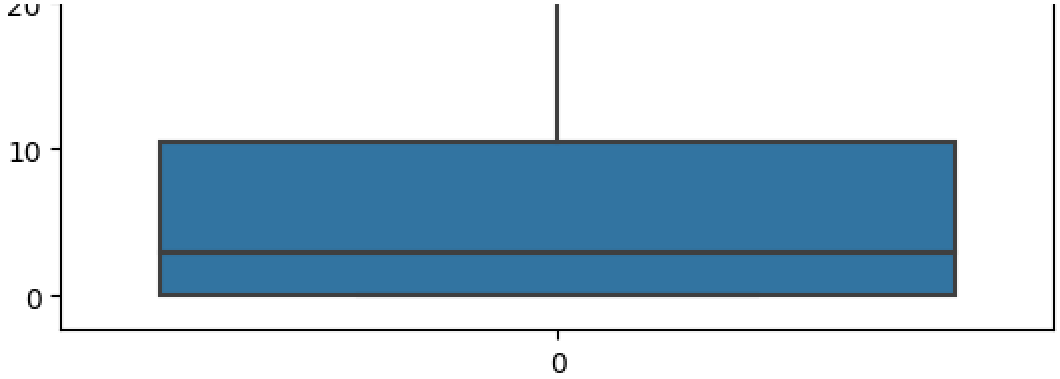

30
20
10
0
0

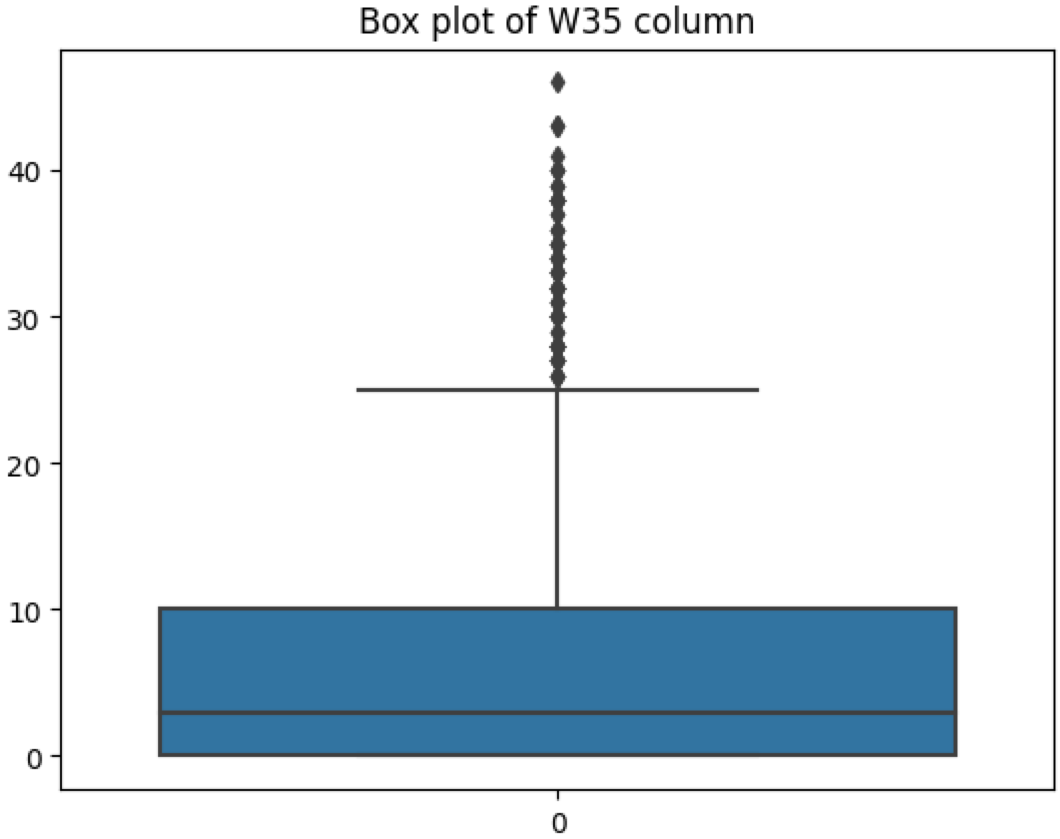

Box plot of W32 column
50
40
30
20
10
0

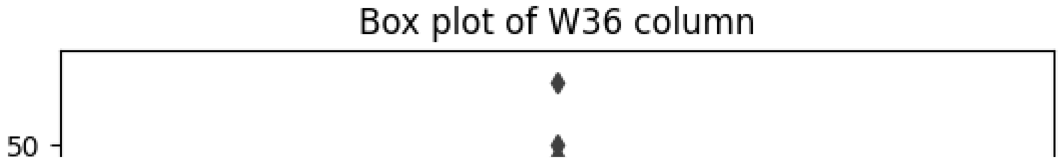

0

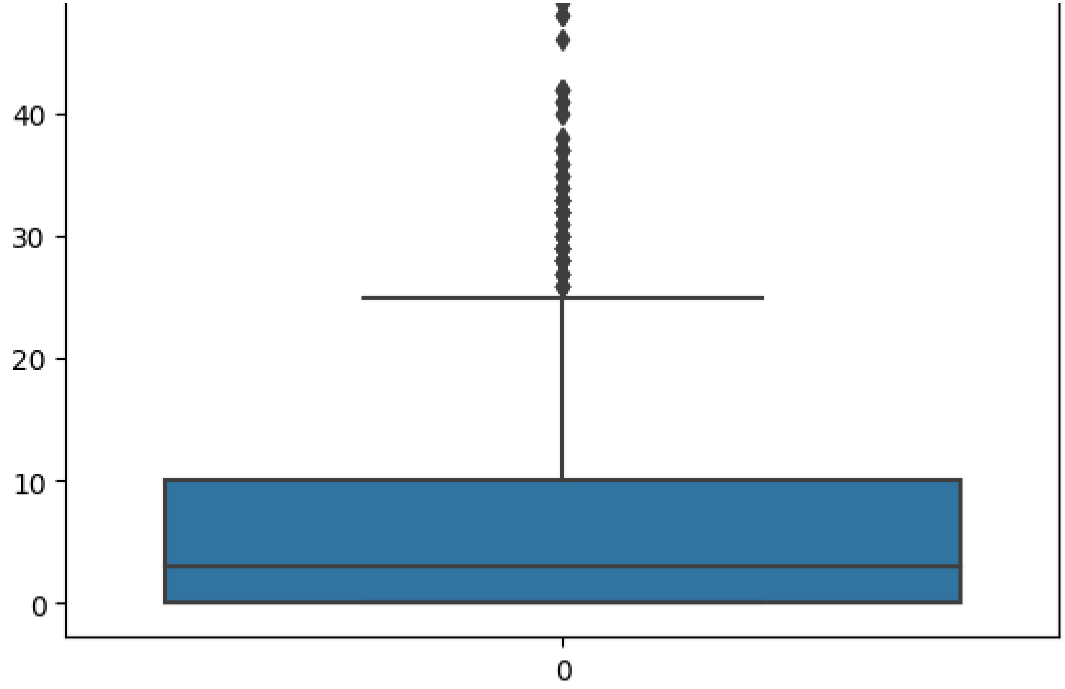

Box plot of W33 column
40
30
20
10
0
0


Box plot of W34 column
40
30

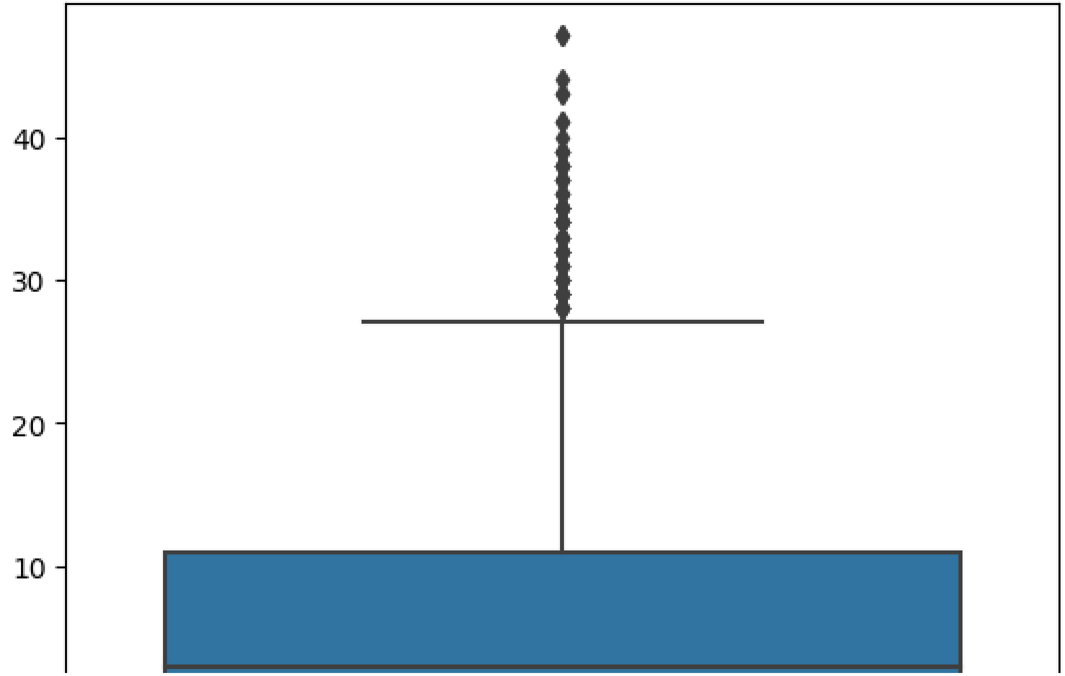

20
10
0
0

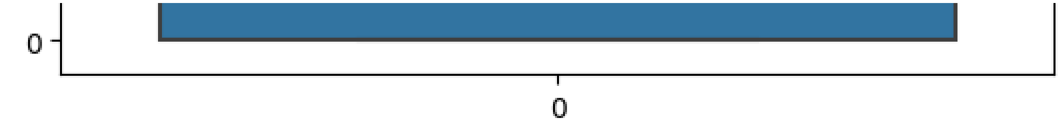

Box plot of W35 column
40
30
20
10
0
0

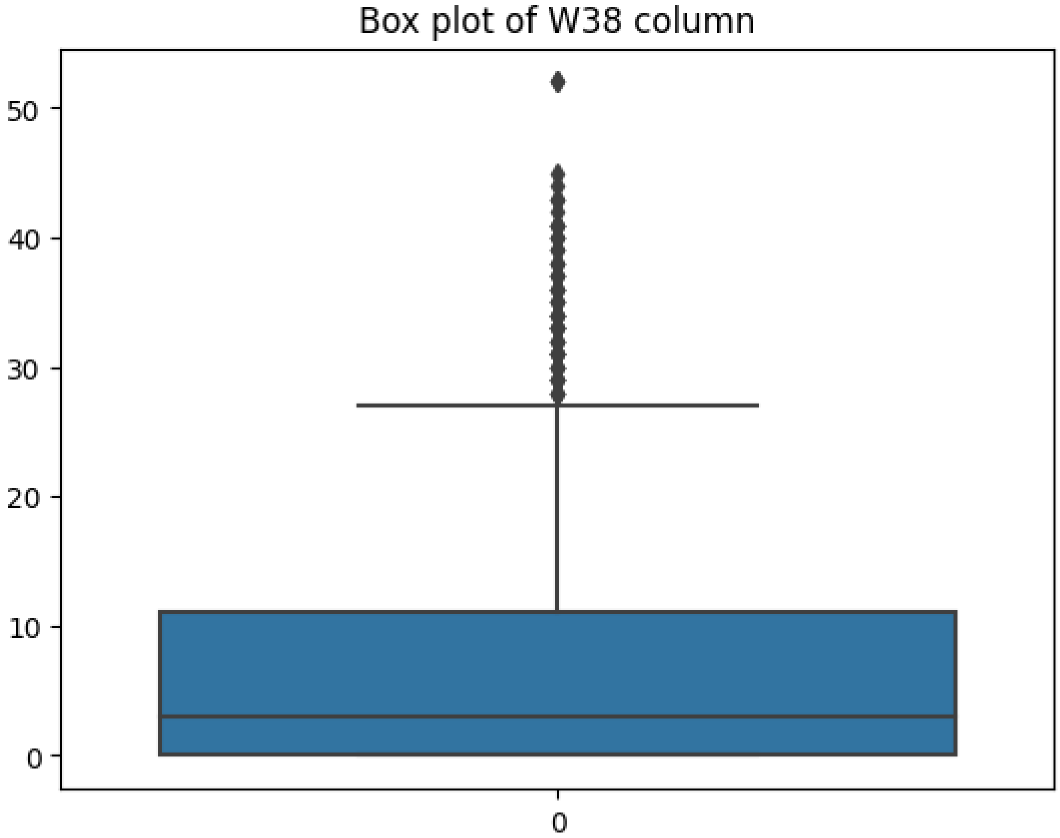

Box plot of W36 column
50

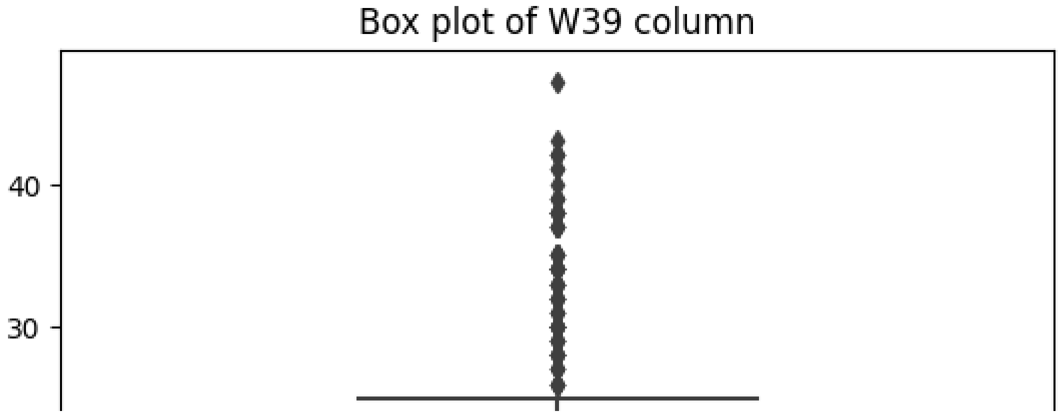


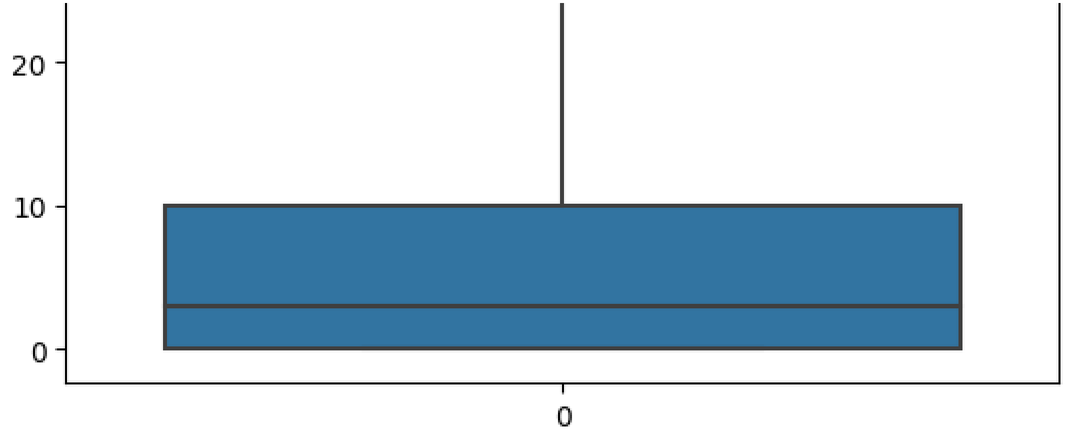
Box plot of W37 column

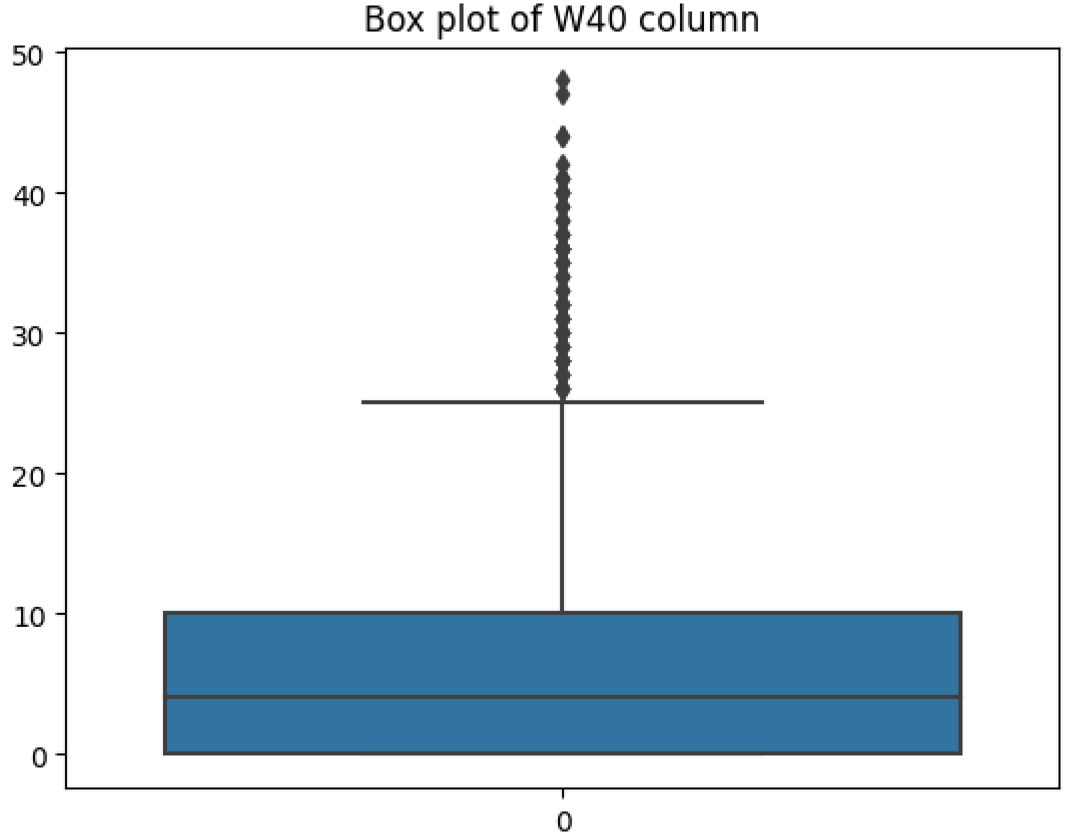


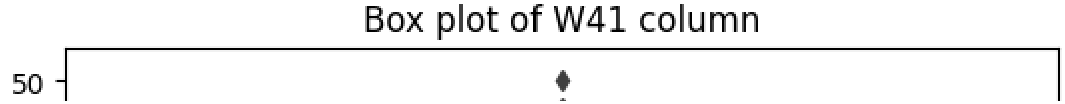

0
0

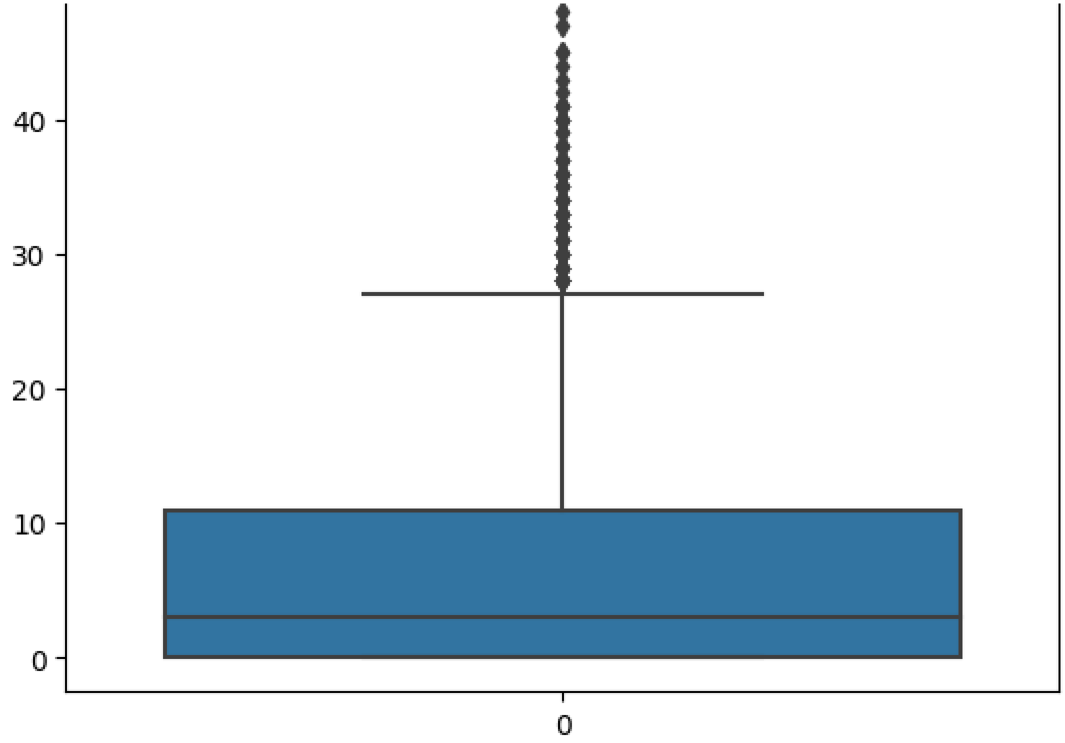

Box plot of W38 column
50
40
30
20
10
0
0


Box plot of W39 column
40
30

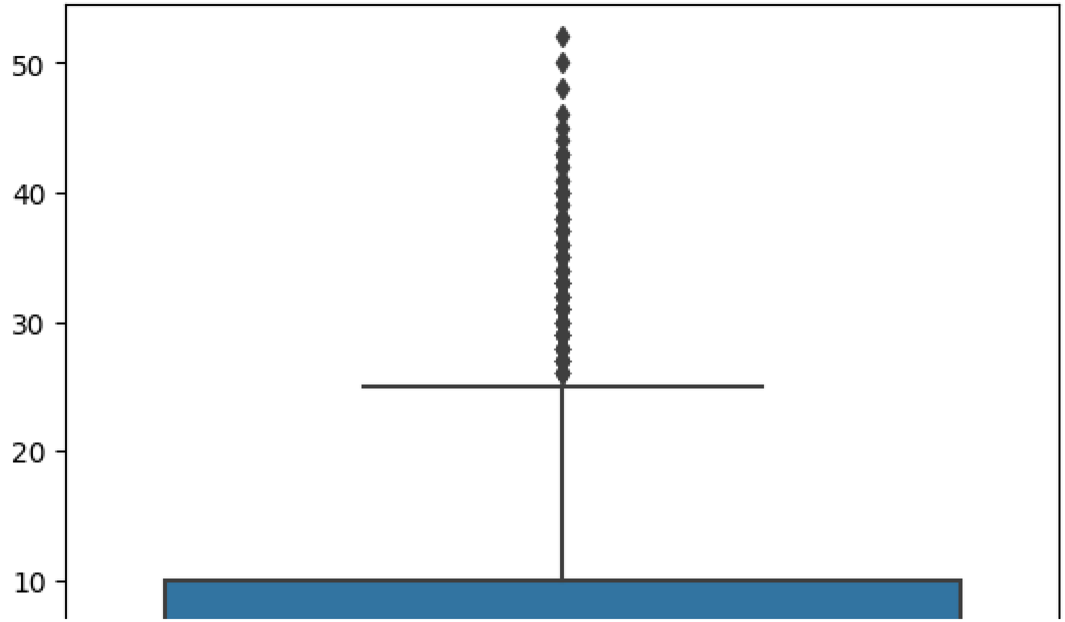

20
10
0
0

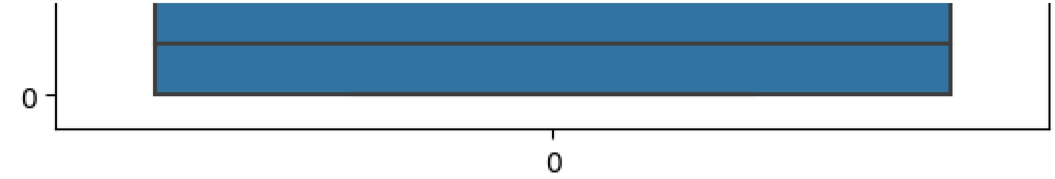

Box plot of W40 column
50
40
30
20
10
0
0

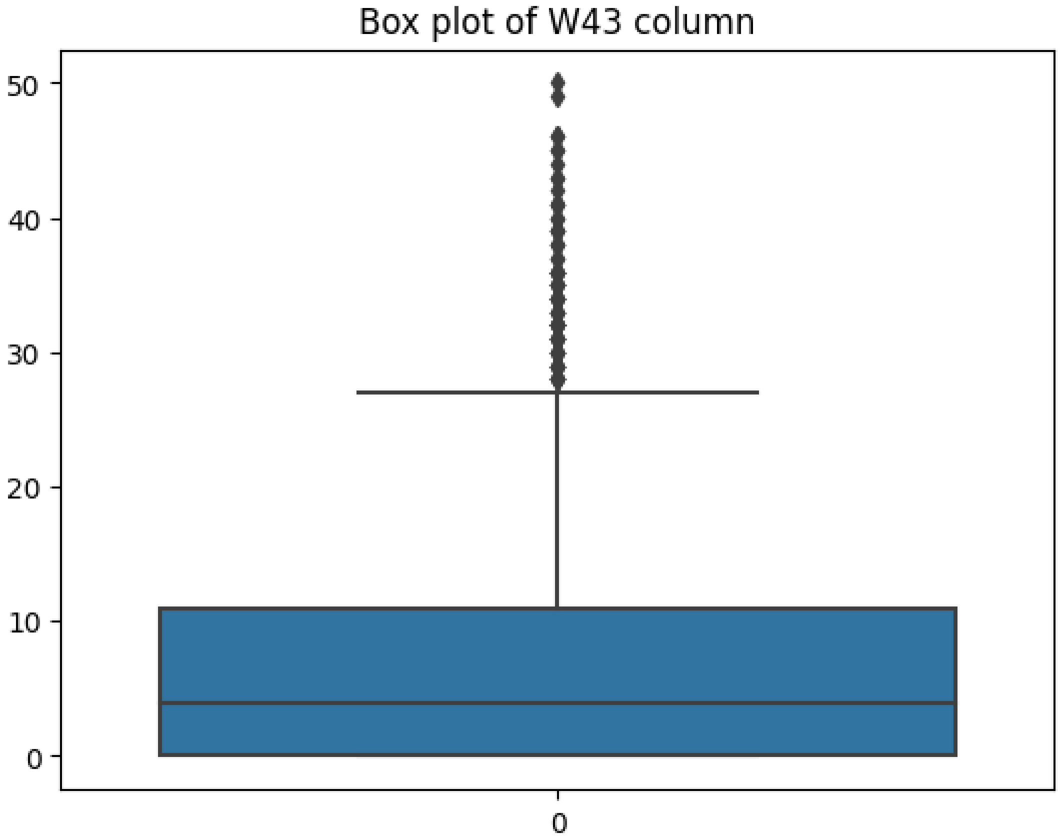

Box plot of W41 column
50

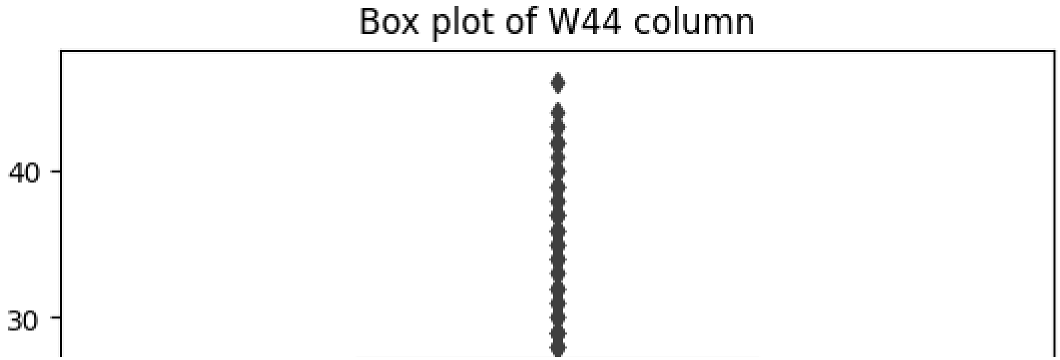


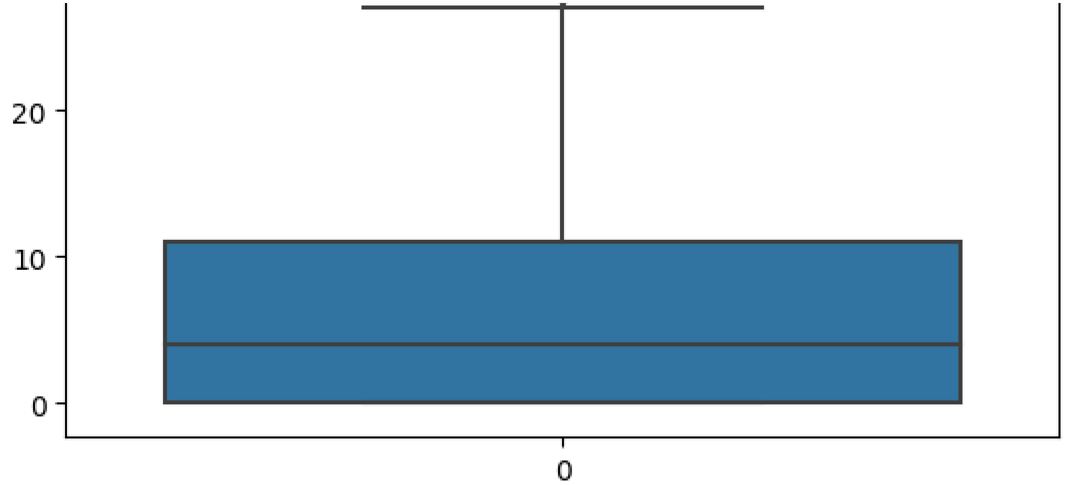
Box plot of W42 column



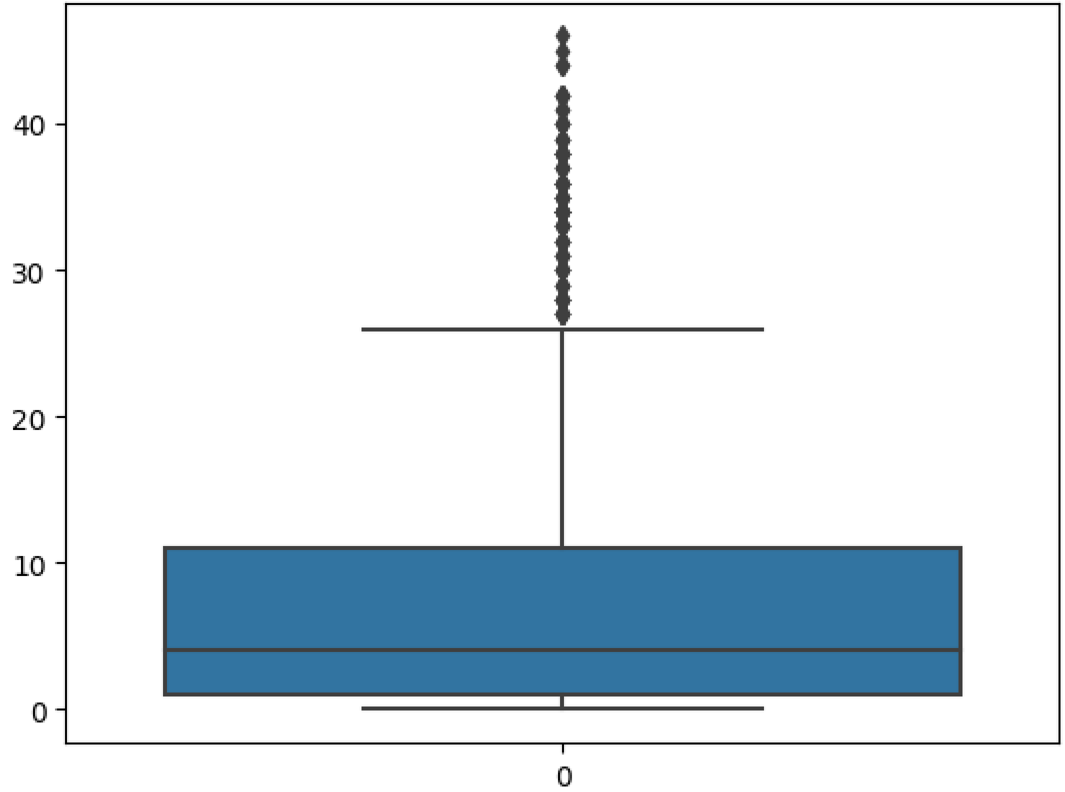

0
0


Box plot of W43 column
50
40
30
20
10
0
0

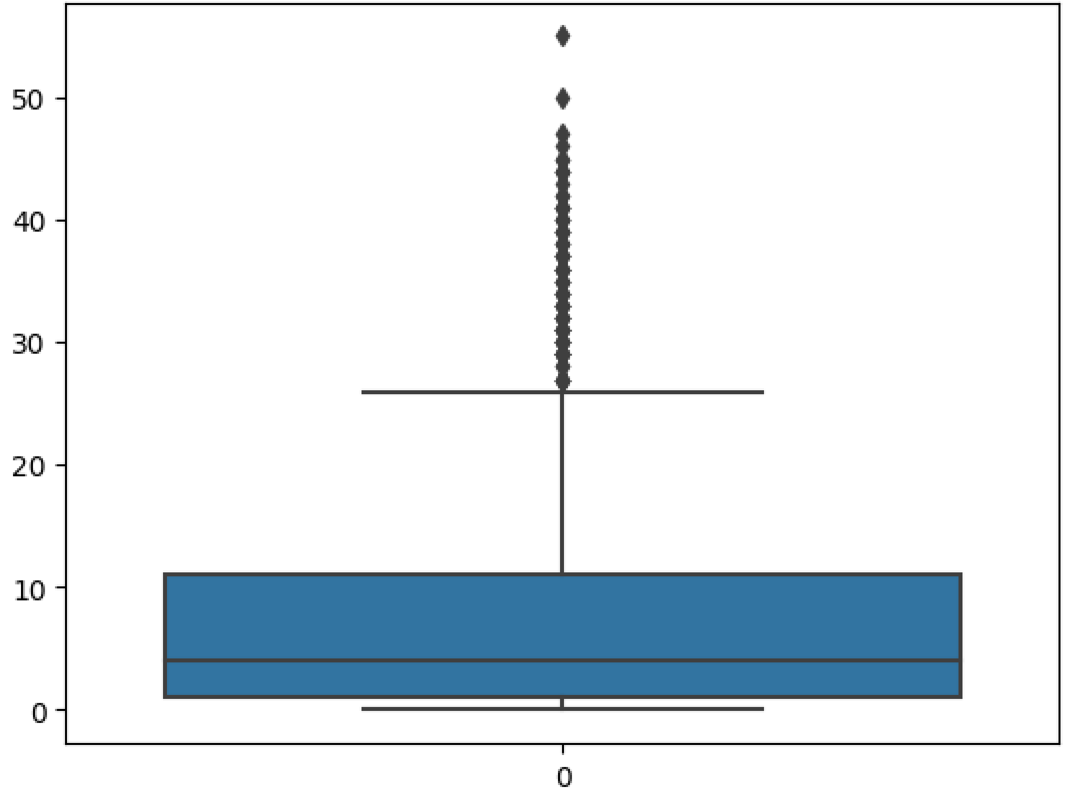

Box plot of W44 column
40
30

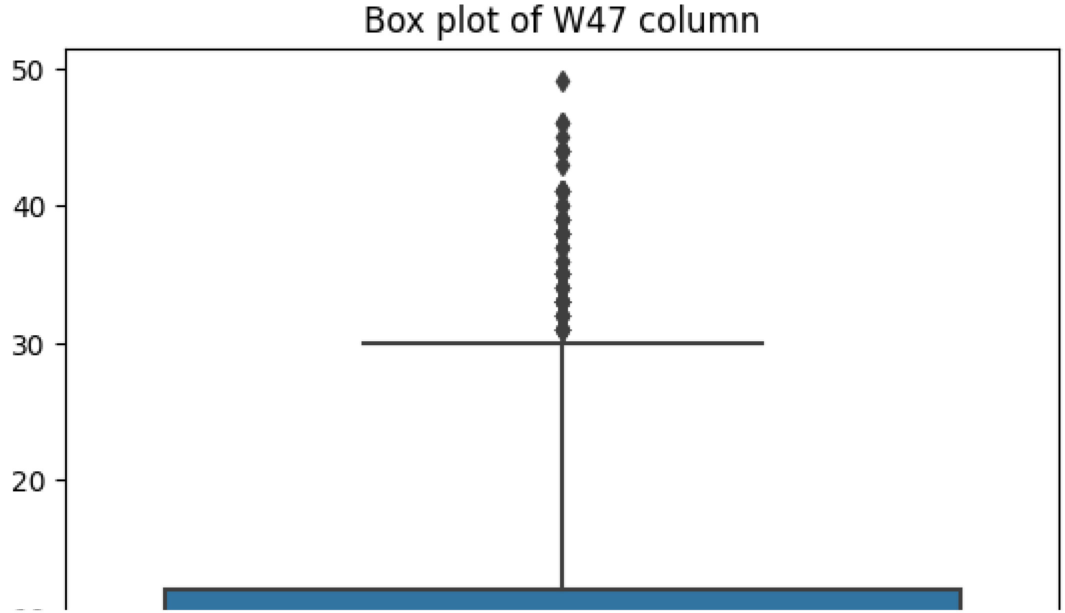


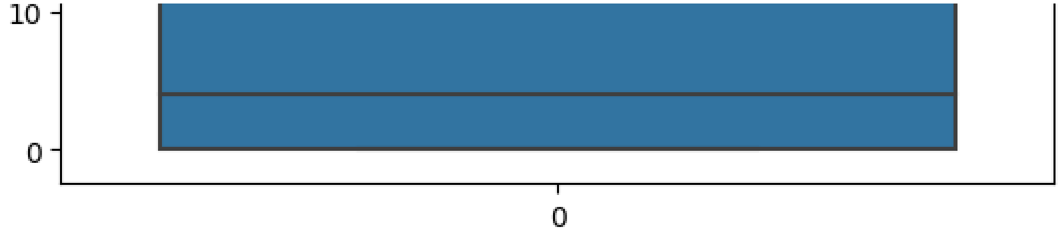
Box plot of W45 column

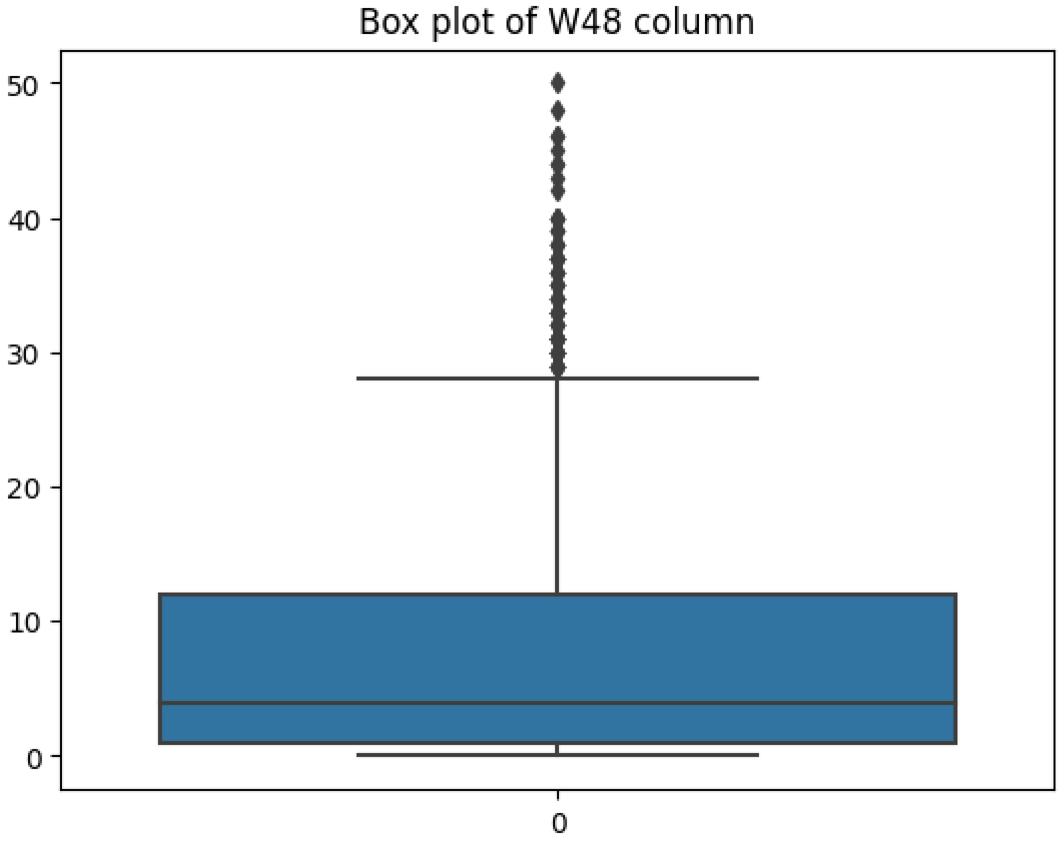


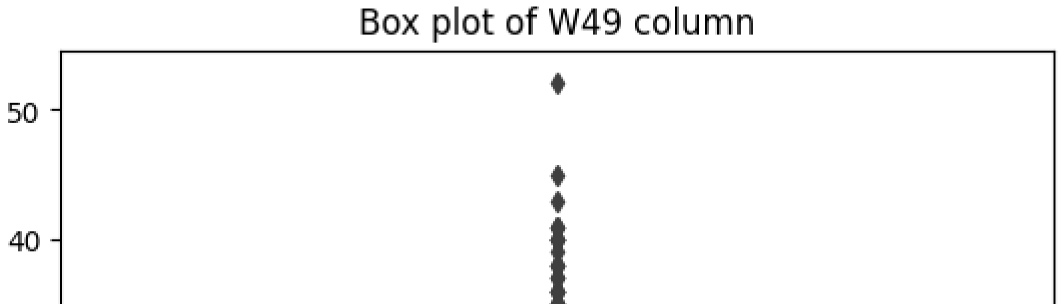
Box plot of W46 column

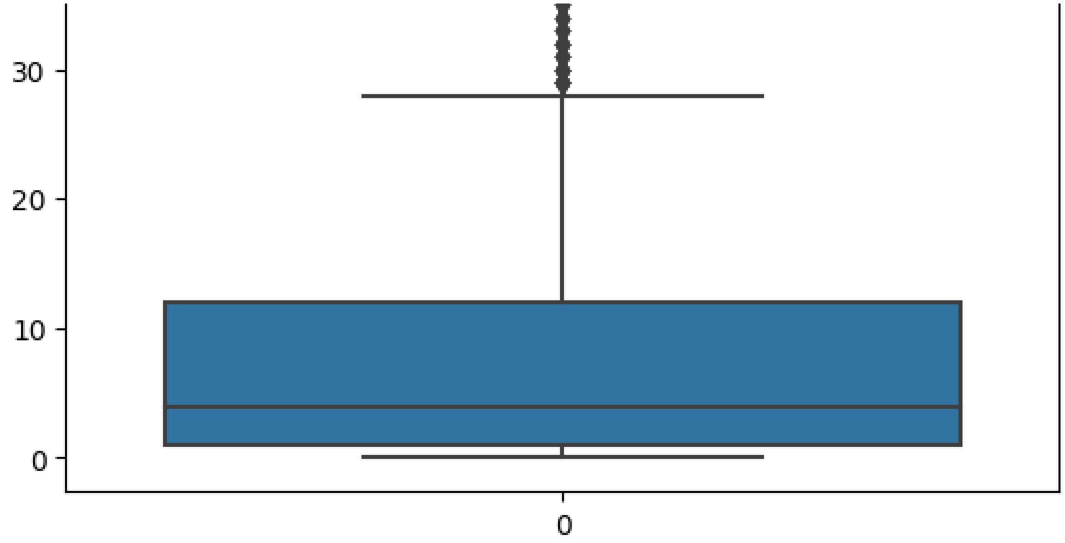

50
40
30
20
10
0
0


Box plot of W47 column
50
40
30
20

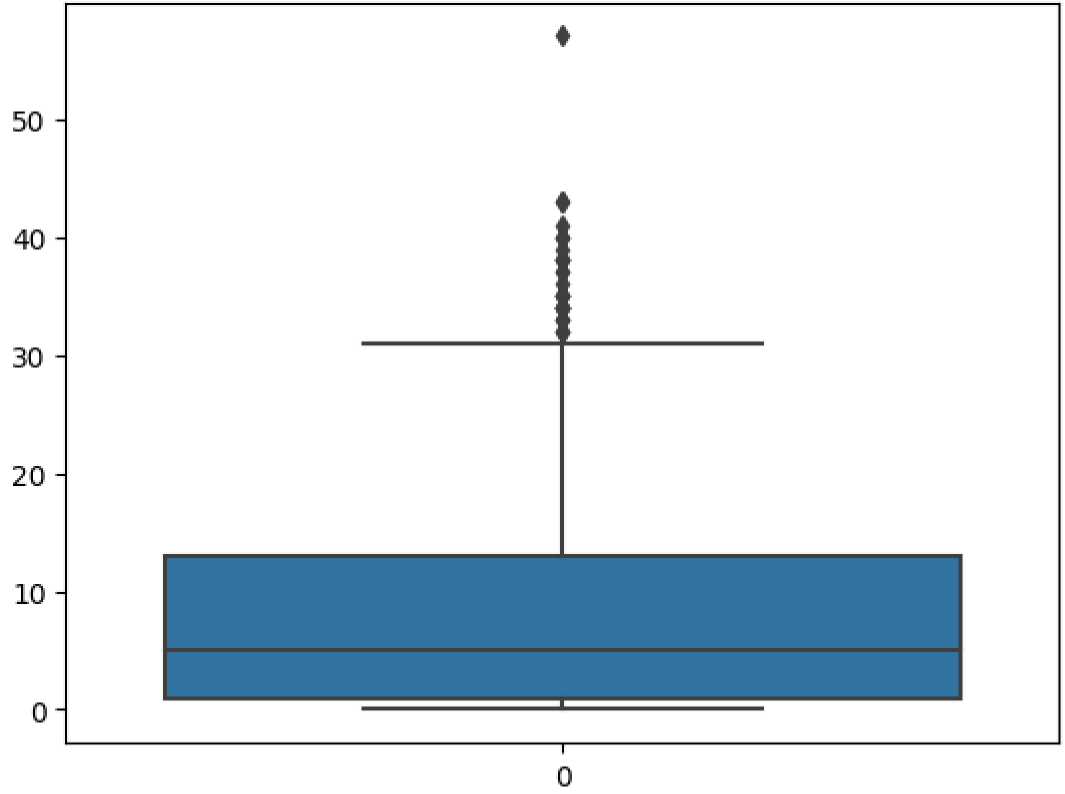

10
0
0

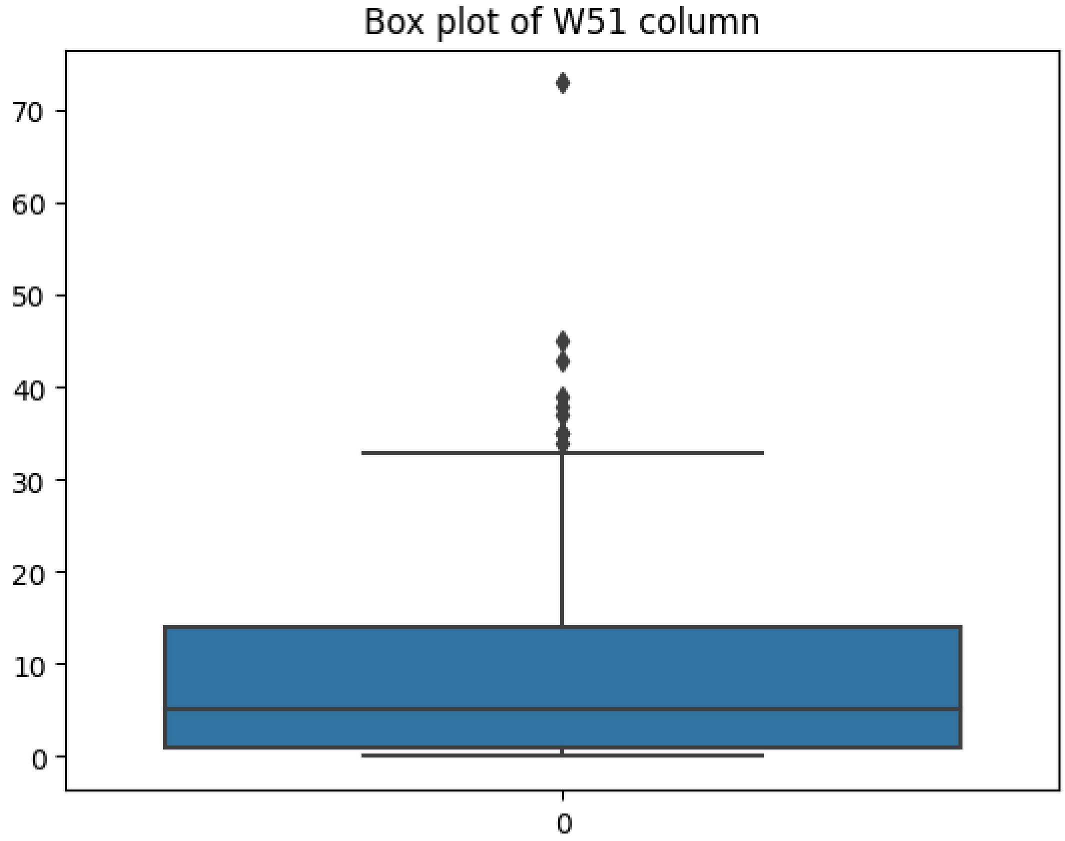

Box plot of W48 column
50
40
30
20
10
0
0

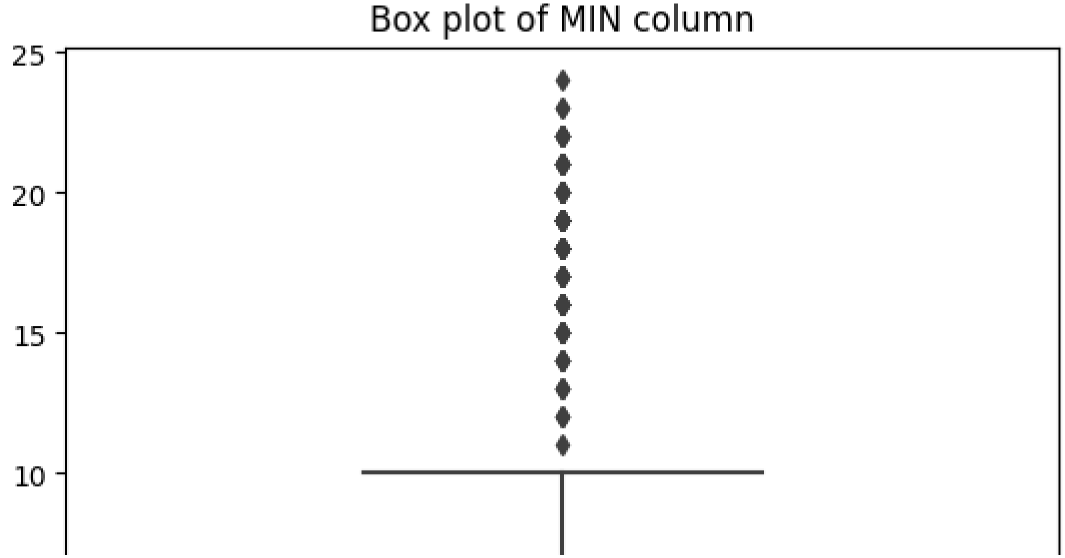

Box plot of W49 column
50
40

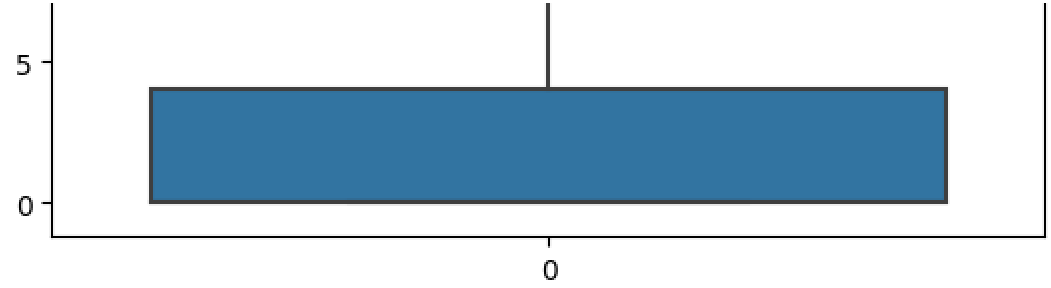


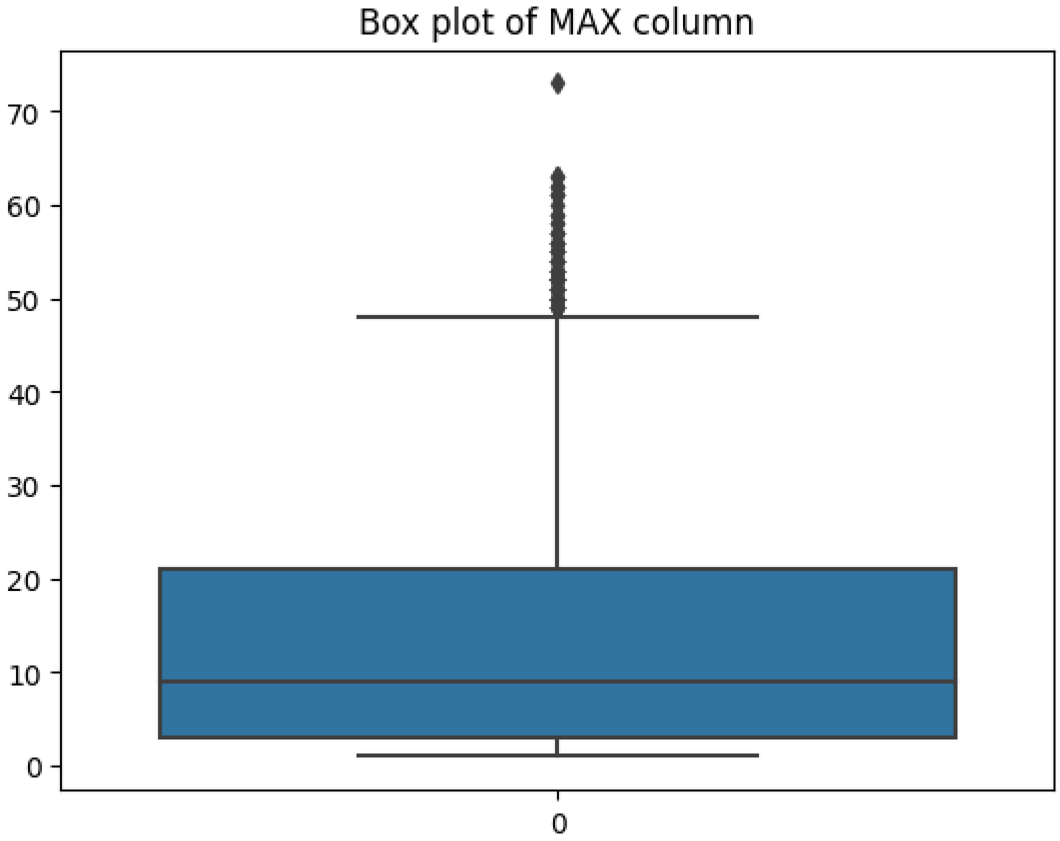
Box plot of W50 column

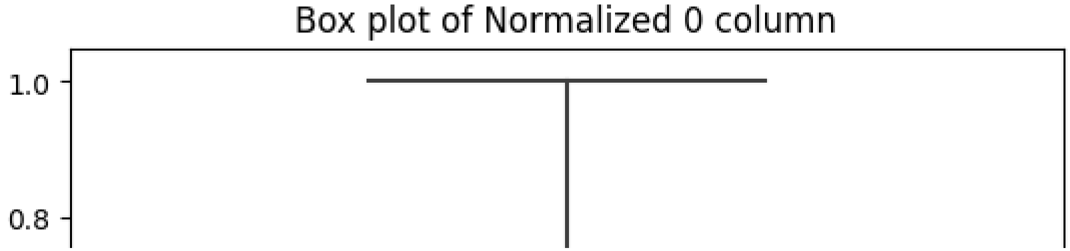


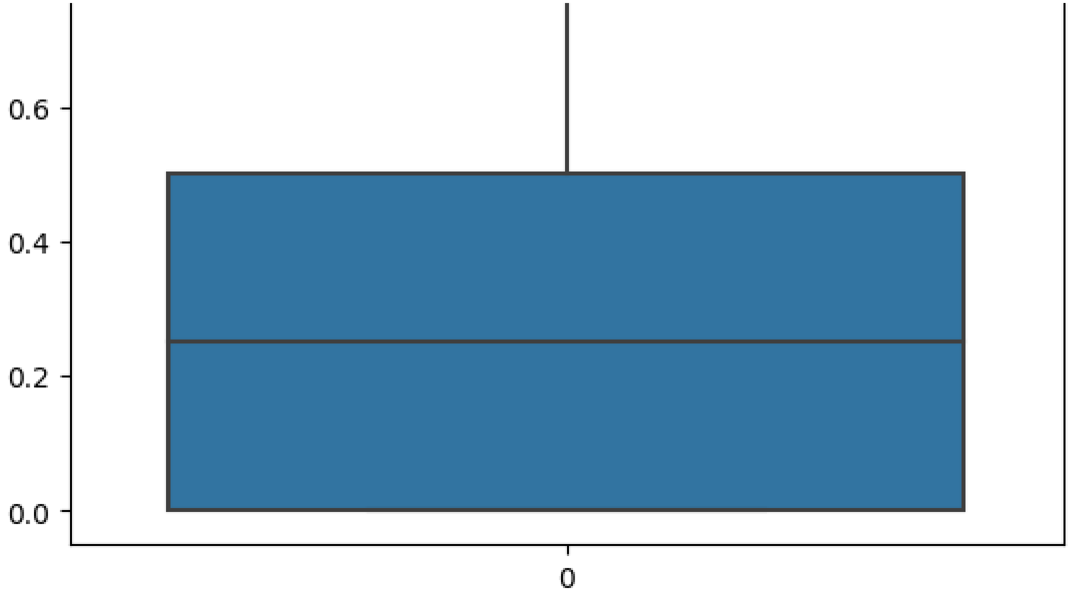

Box plot of W51 column
70
60
50
40
30
20
10
0
0

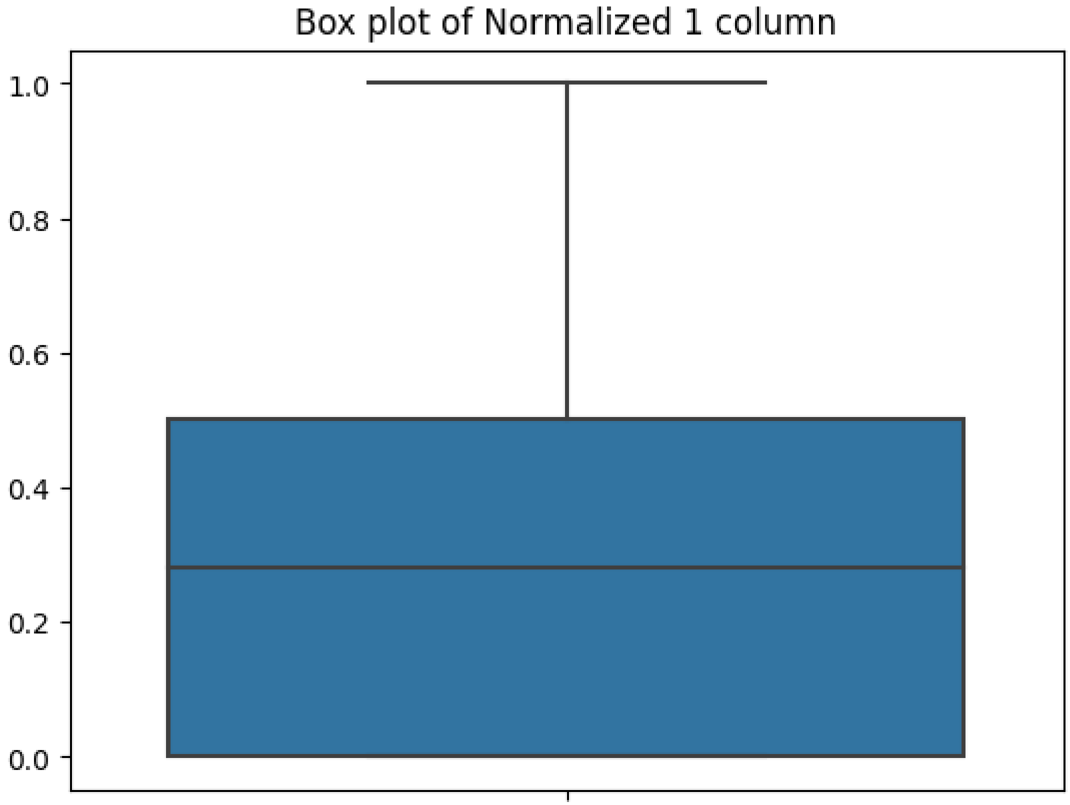

Box plot of MIN column
25
20
15
10


5
0
0


Box plot of MAX column
70
60
50
40
30
20
10
0
0

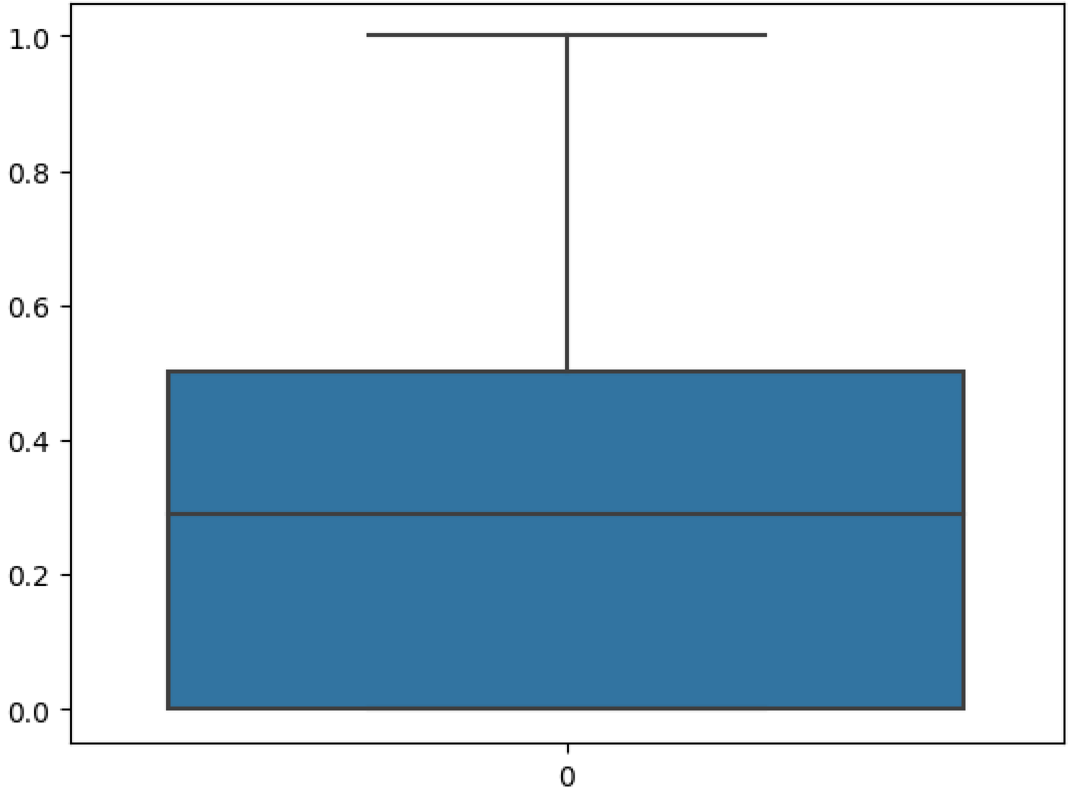

Box plot of Normalized 0 column
1.0
0.8


0.6
0.4
0.2
0.0
0

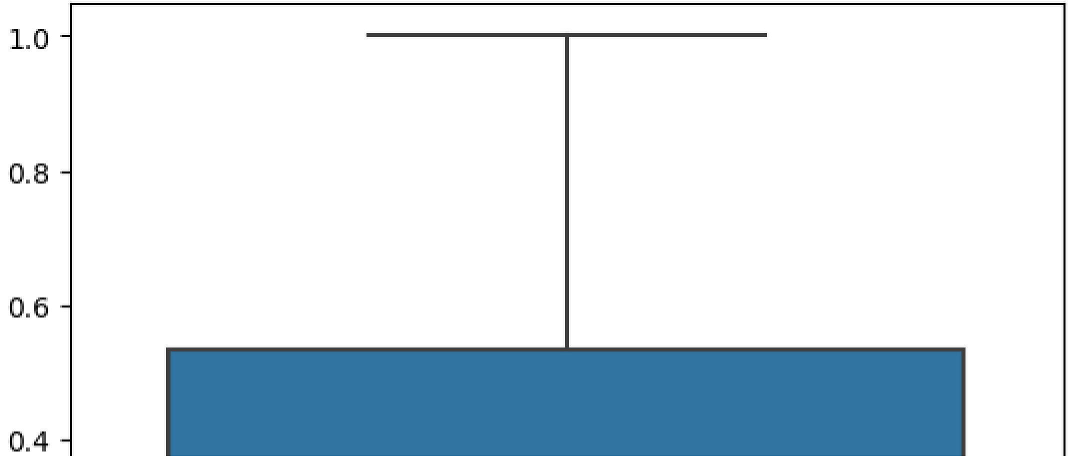

Box plot of Normalized 1 column
1.0
0.8
0.6
0.4
0.2
0.0

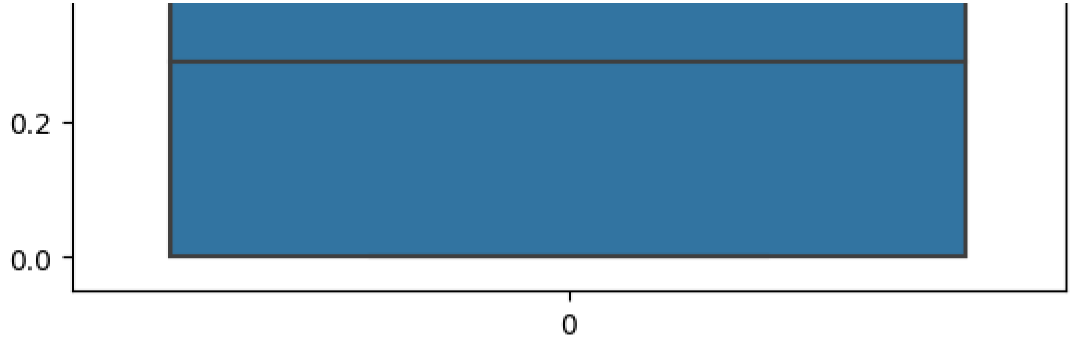
0

Box plot of Normalized 2 column

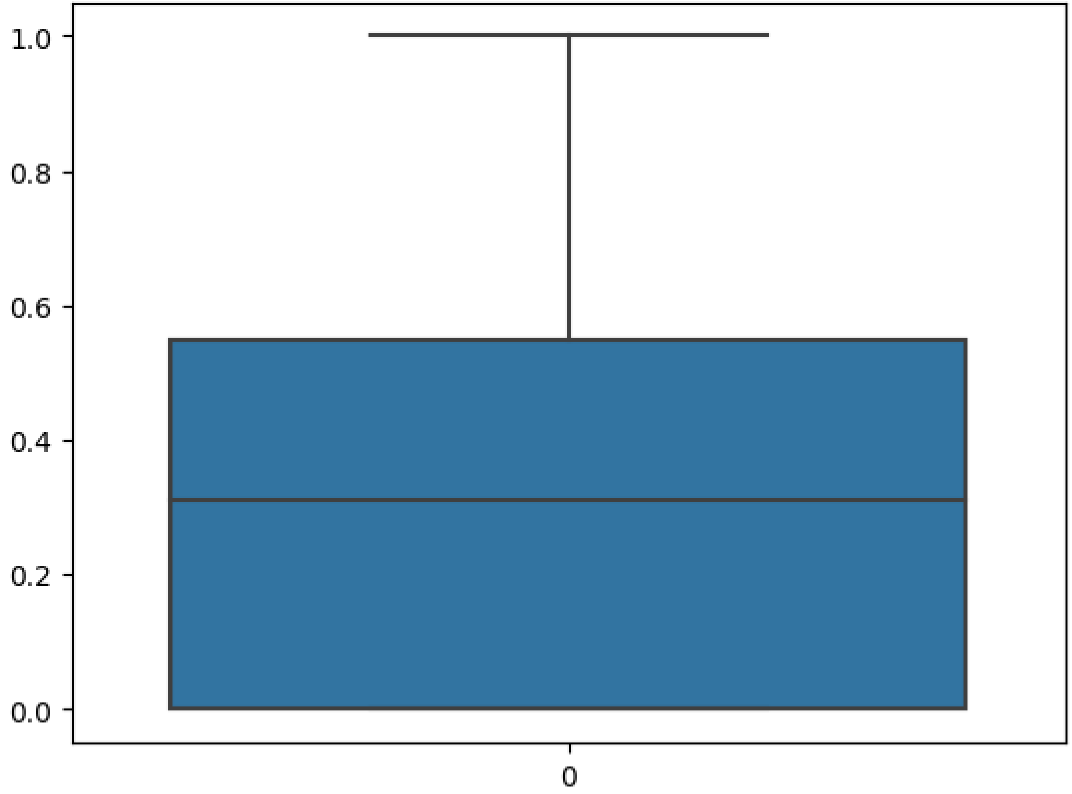


Box plot of Normalized 3 column

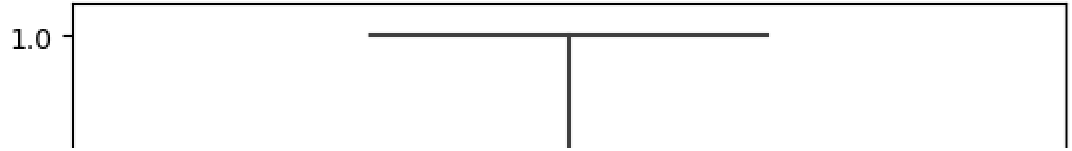


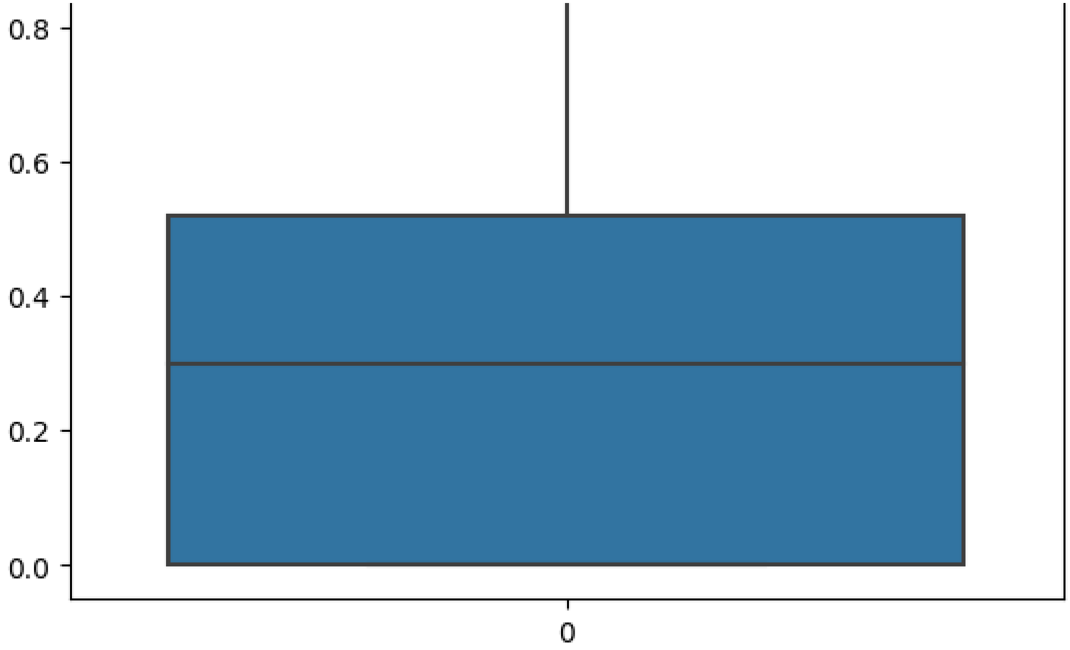


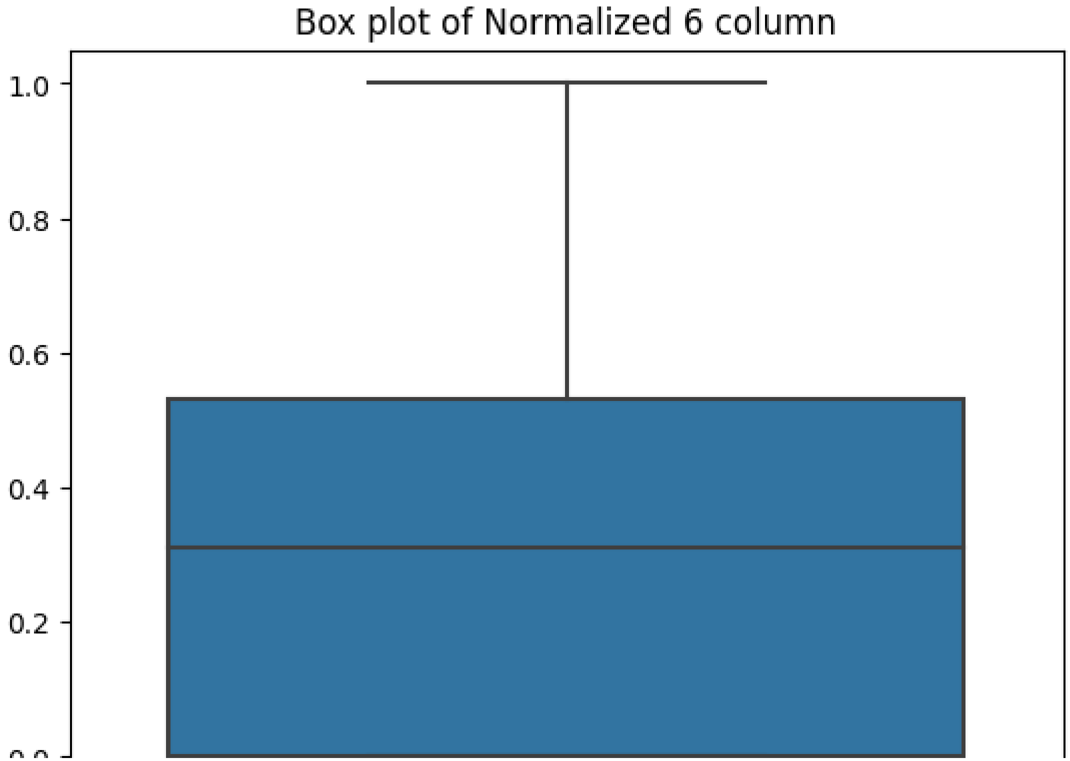
Box plot of Normalized 4 column

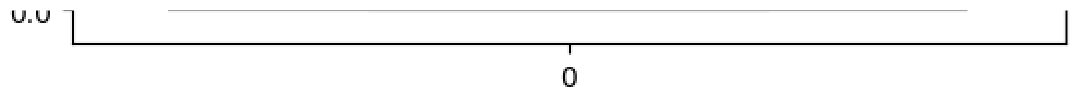


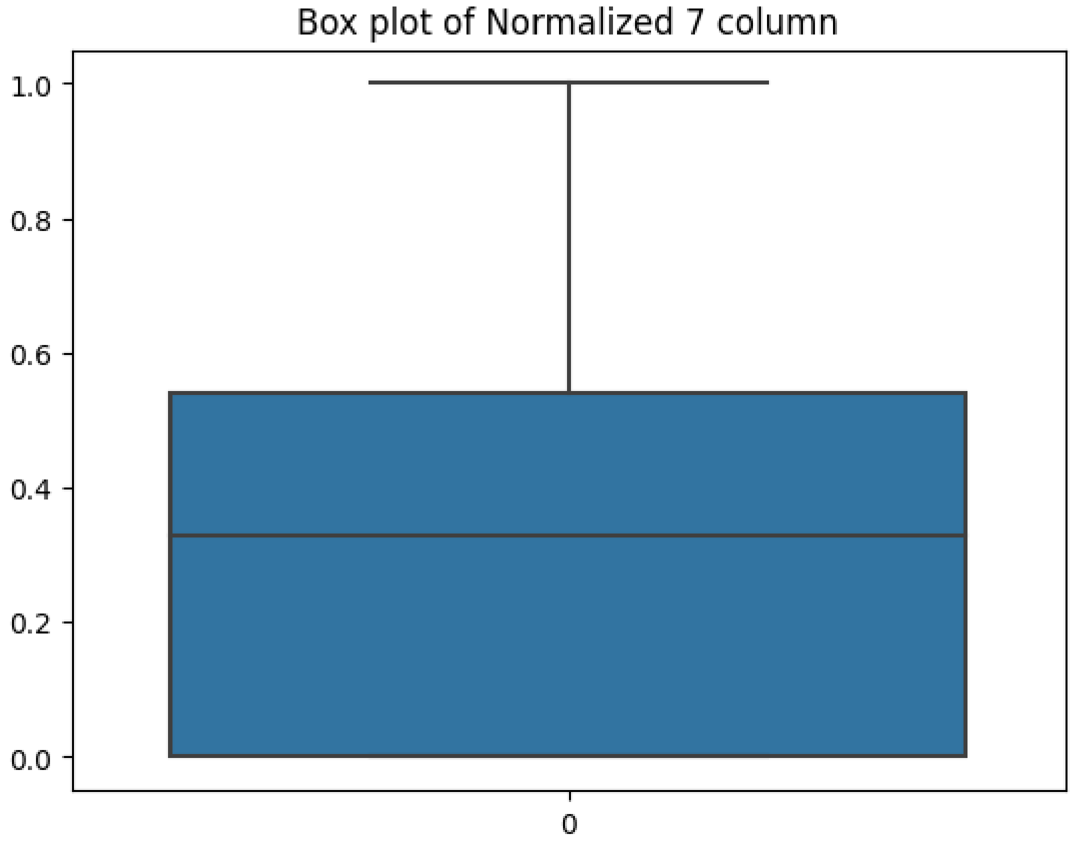
Box plot of Normalized 5 column

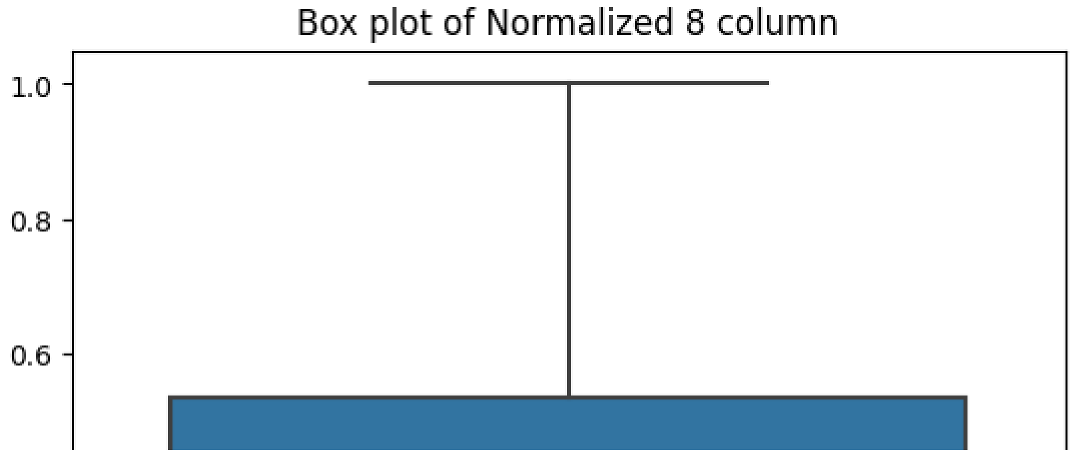


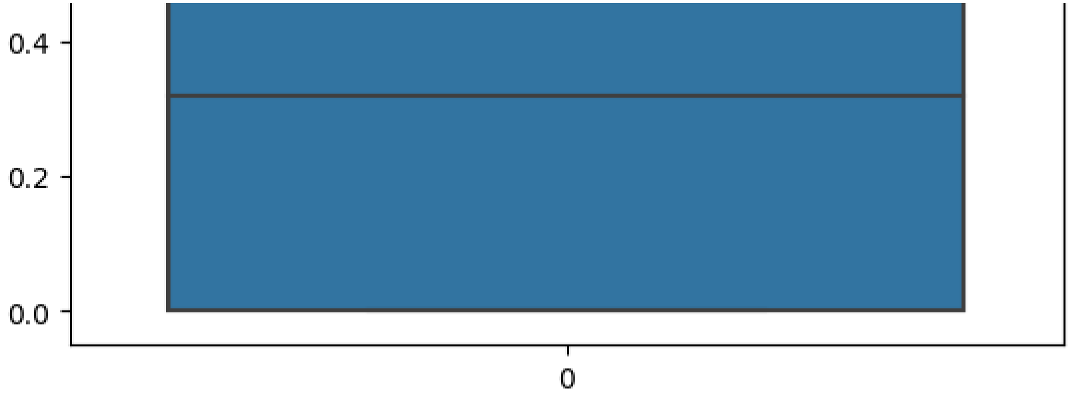

0.8
0.6
0.4
0.2
0.0
0

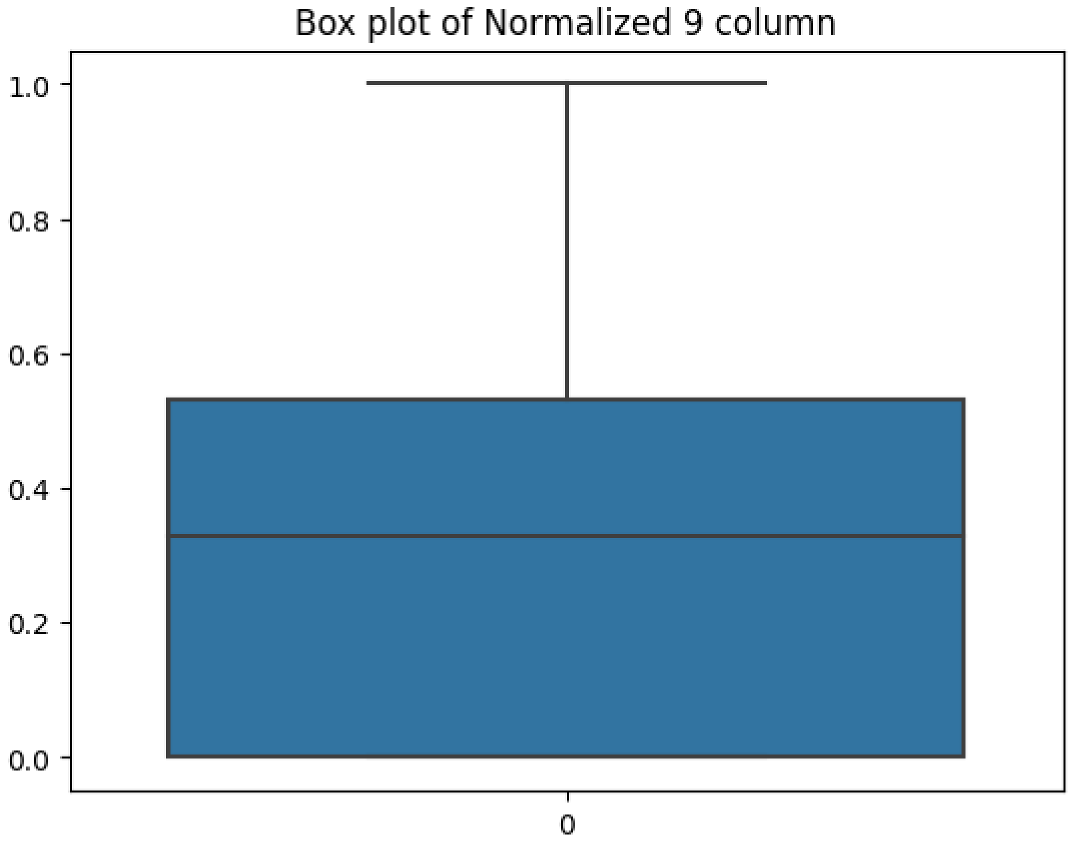

Box plot of Normalized 6 column
1.0
0.8
0.6
0.4
0.2

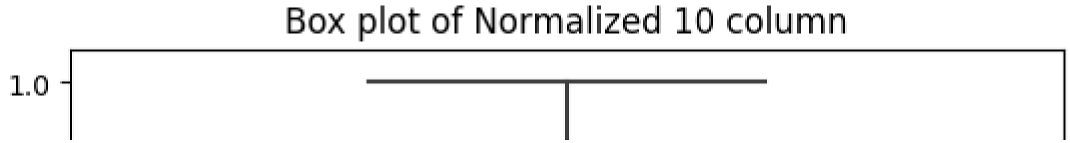

0.0
0

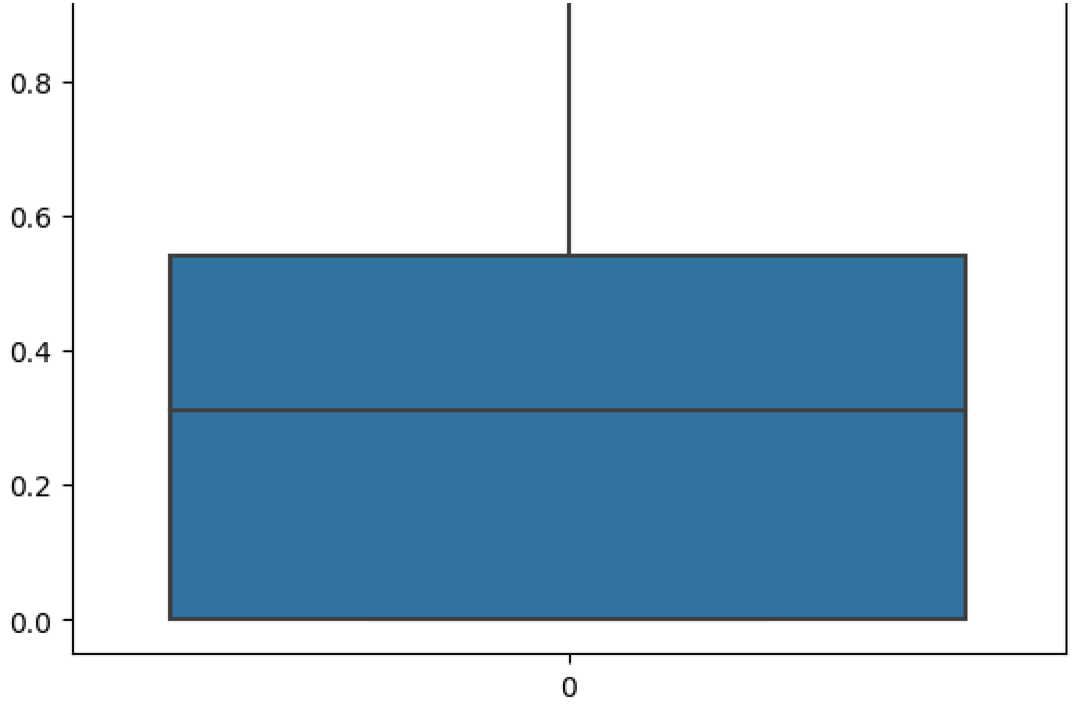

Box plot of Normalized 7 column
1.0
0.8
0.6
0.4
0.2
0.0
0

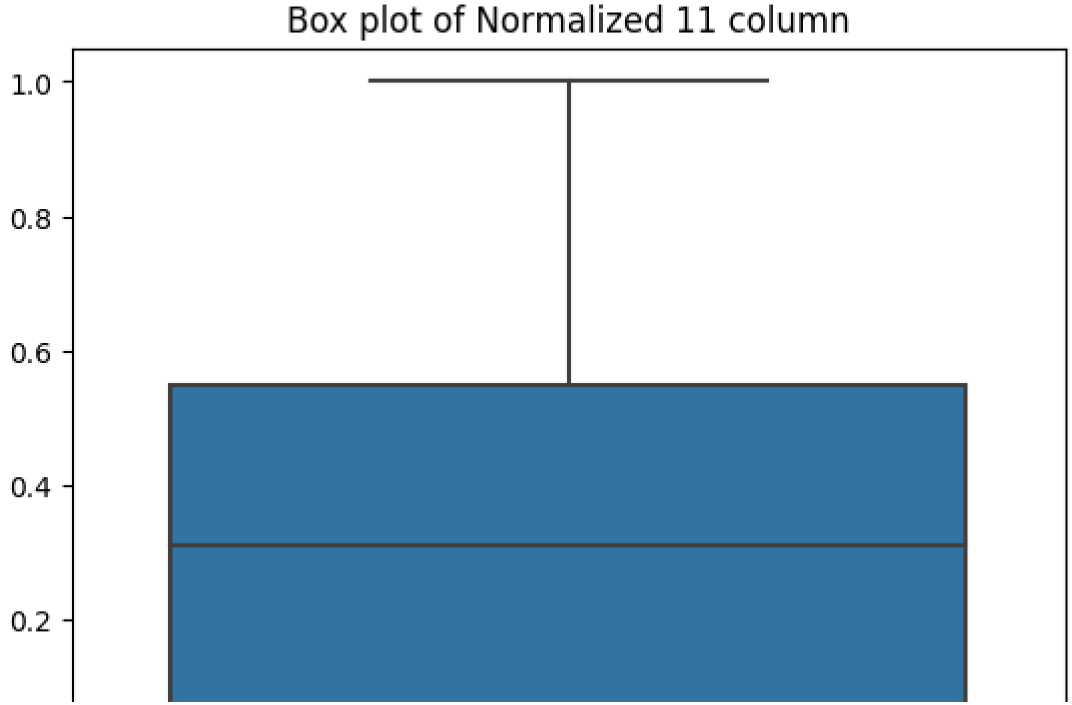

Box plot of Normalized 8 column
1.0
0.8
0.6

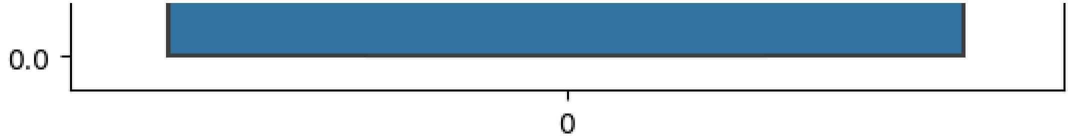

0.4
0.2
0.0
0

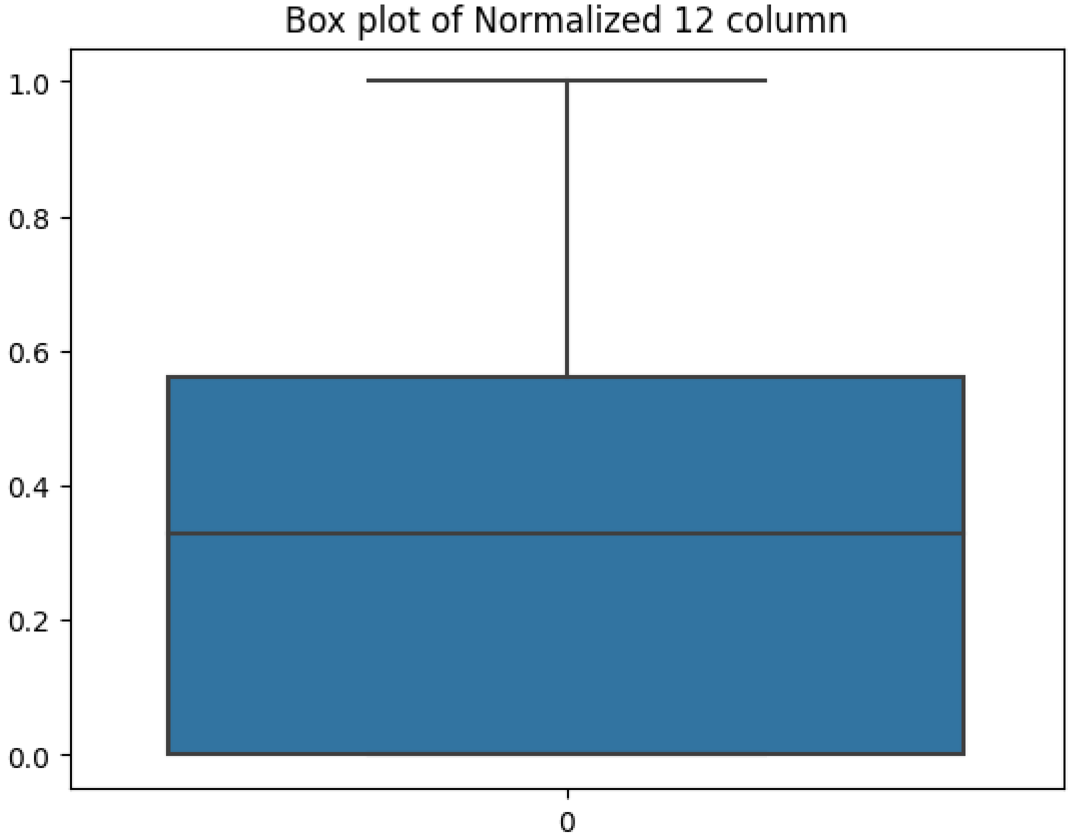

Box plot of Normalized 9 column
1.0
0.8
0.6
0.4
0.2
0.0
0

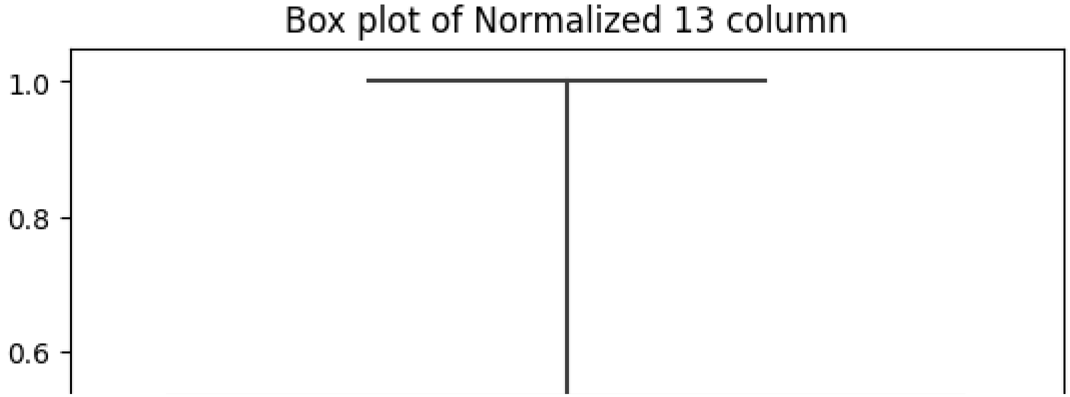

Box plot of Normalized 10 column
1.0

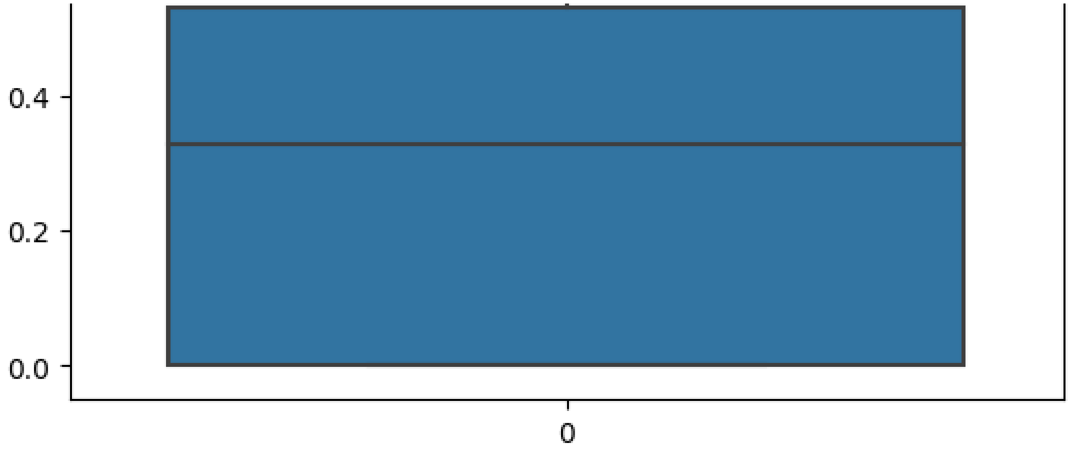

0.8
0.6
0.4
0.2
0.0
0

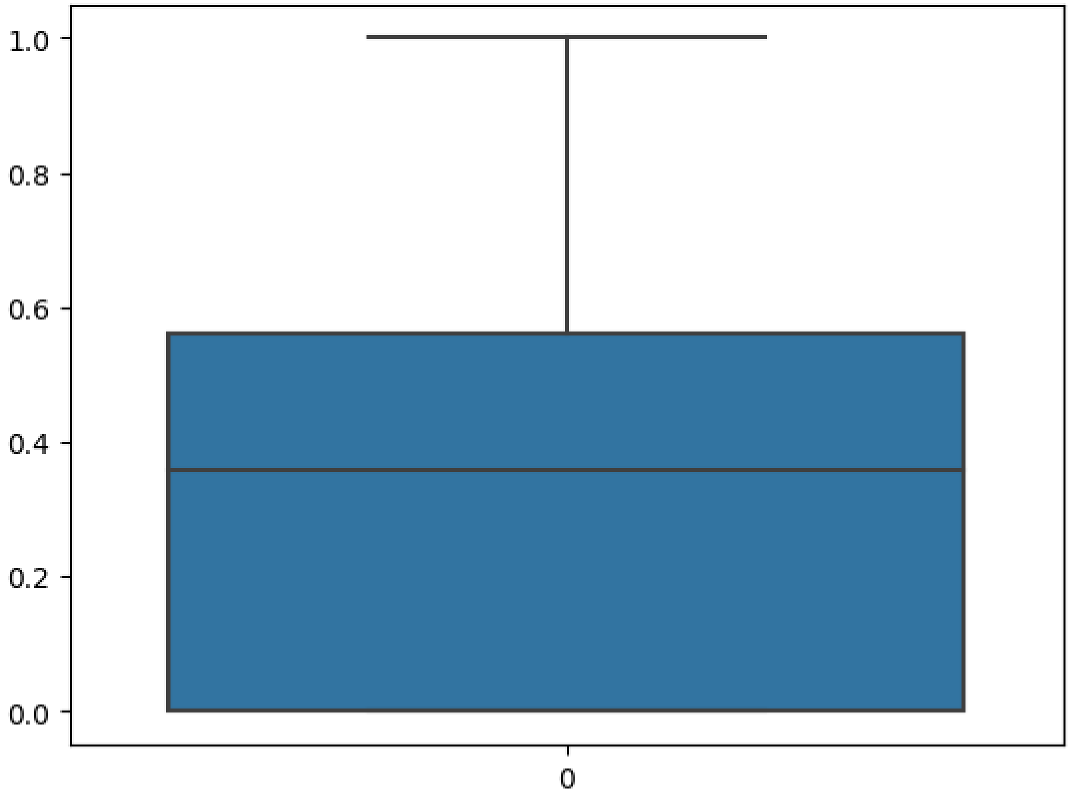

Box plot of Normalized 11 column
1.0
0.8
0.6
0.4
0.2


0.0
0

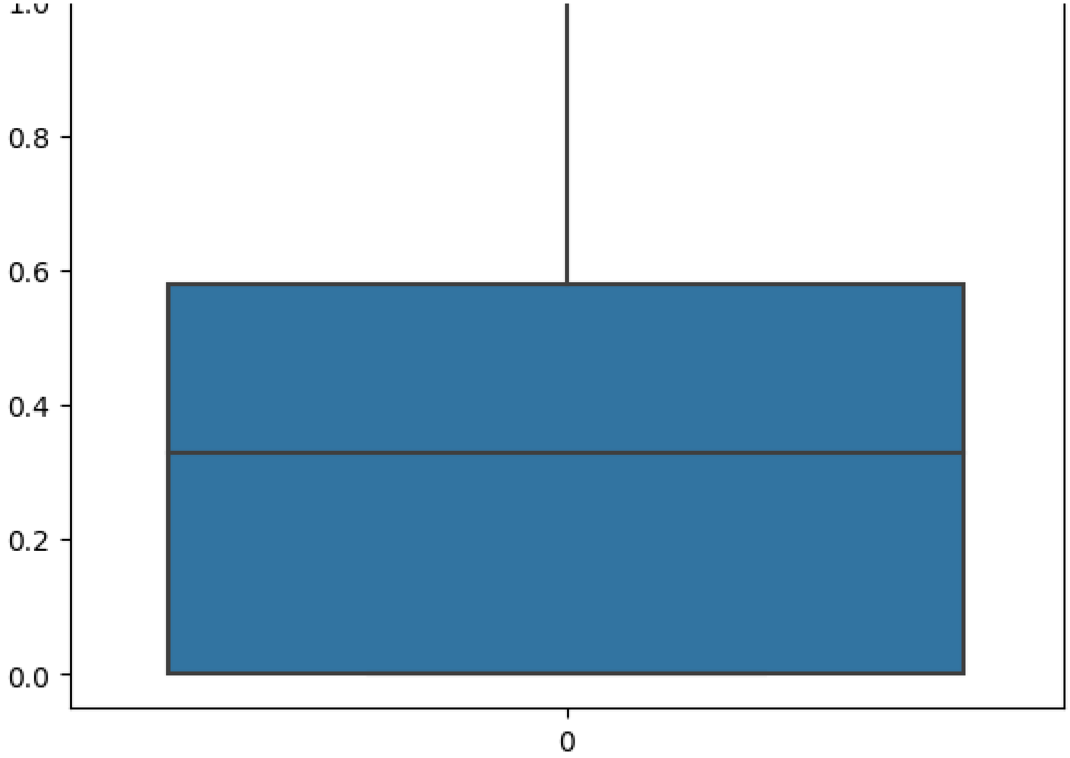

Box plot of Normalized 12 column
1.0
0.8
0.6
0.4
0.2
0.0
0

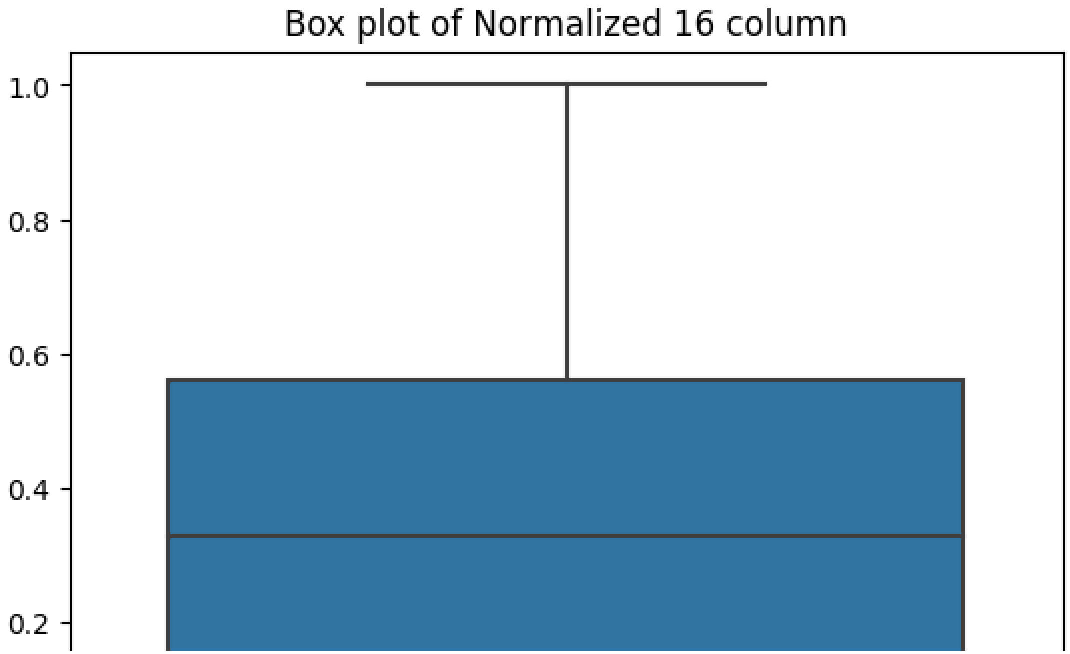

Box plot of Normalized 13 column
1.0
0.8
0.6

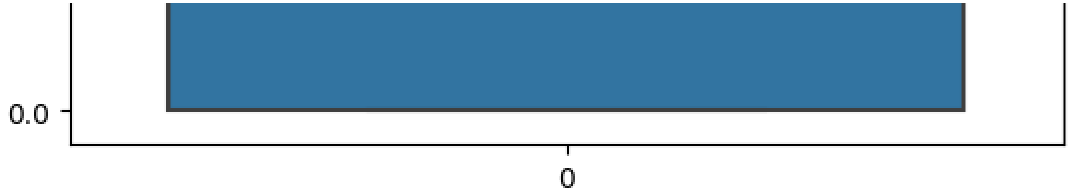


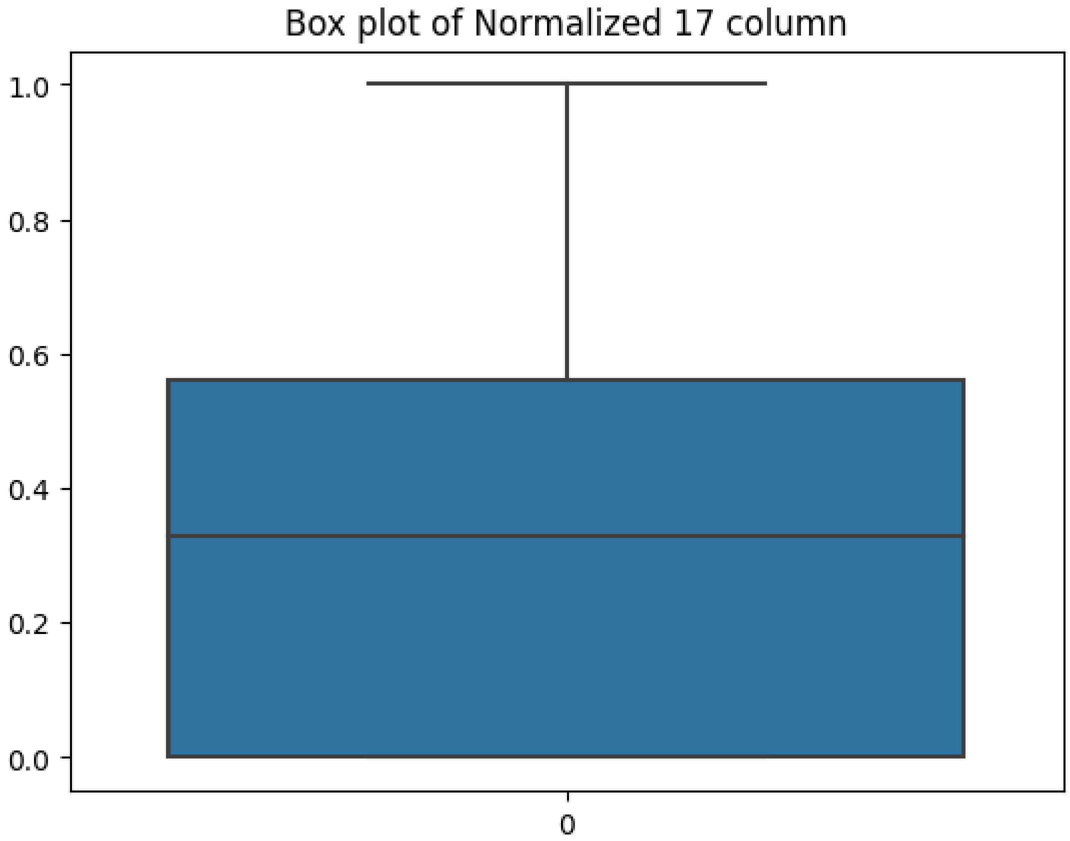

Box plot of Normalized 14 column

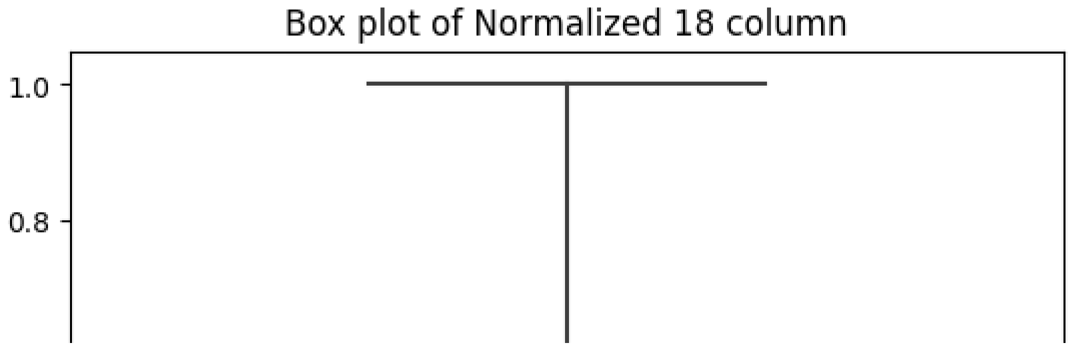
Box plot of Normalized 15 column

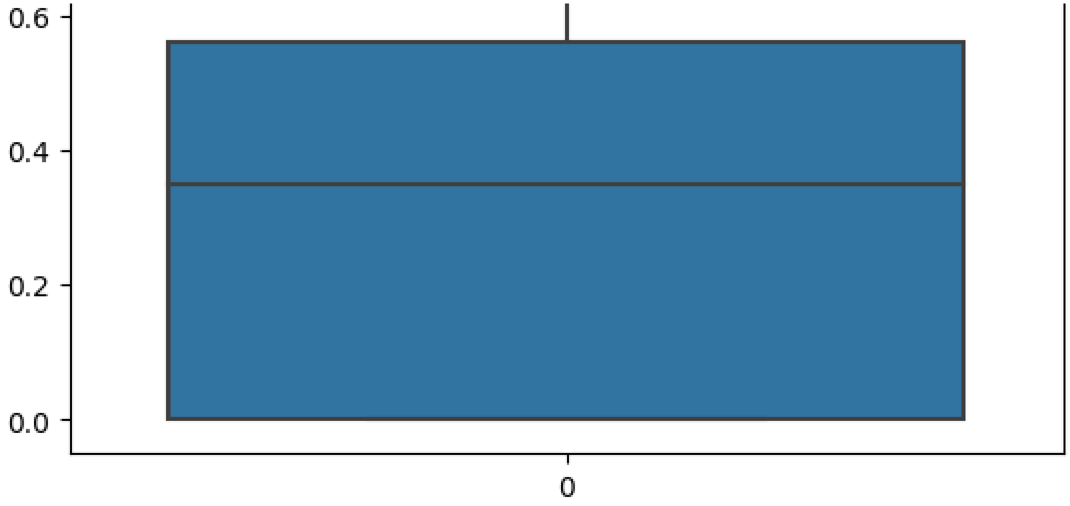

1.0
0.8
0.6
0.4
0.2
0.0
0


Box plot of Normalized 16 column
1.0
0.8
0.6
0.4
0.2

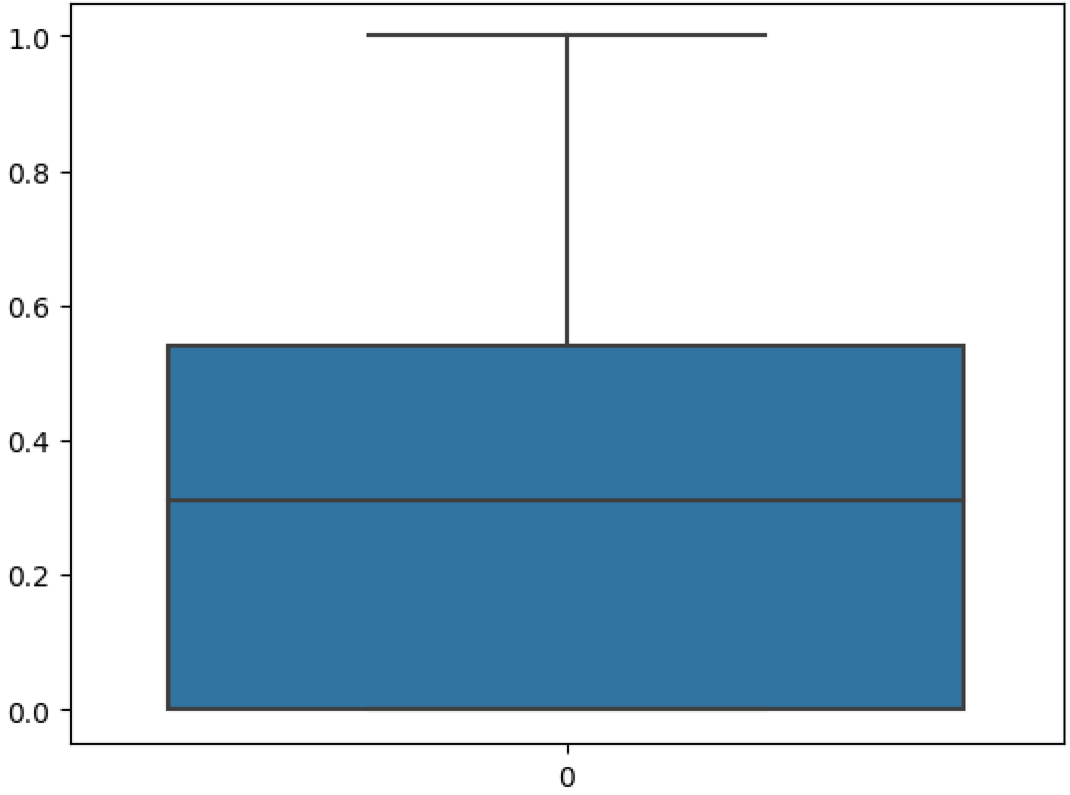

0.0
0


Box plot of Normalized 17 column
1.0
0.8
0.6
0.4
0.2
0.0
0

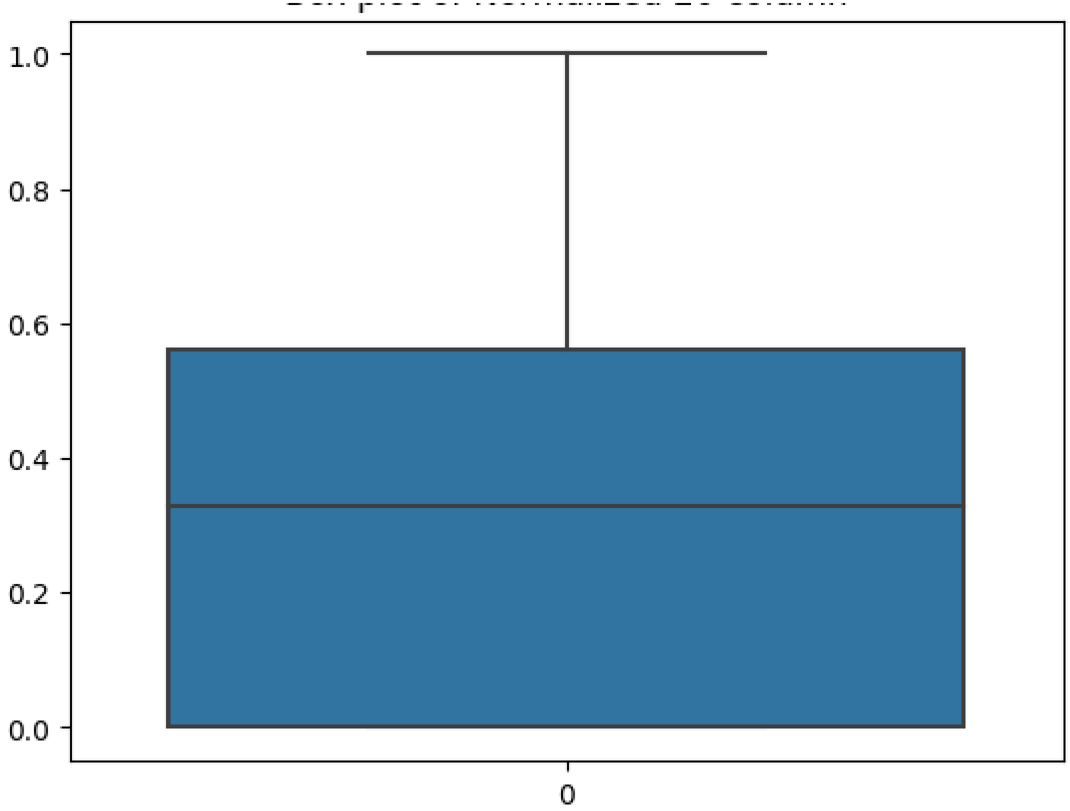

Box plot of Normalized 18 column
1.0
0.8

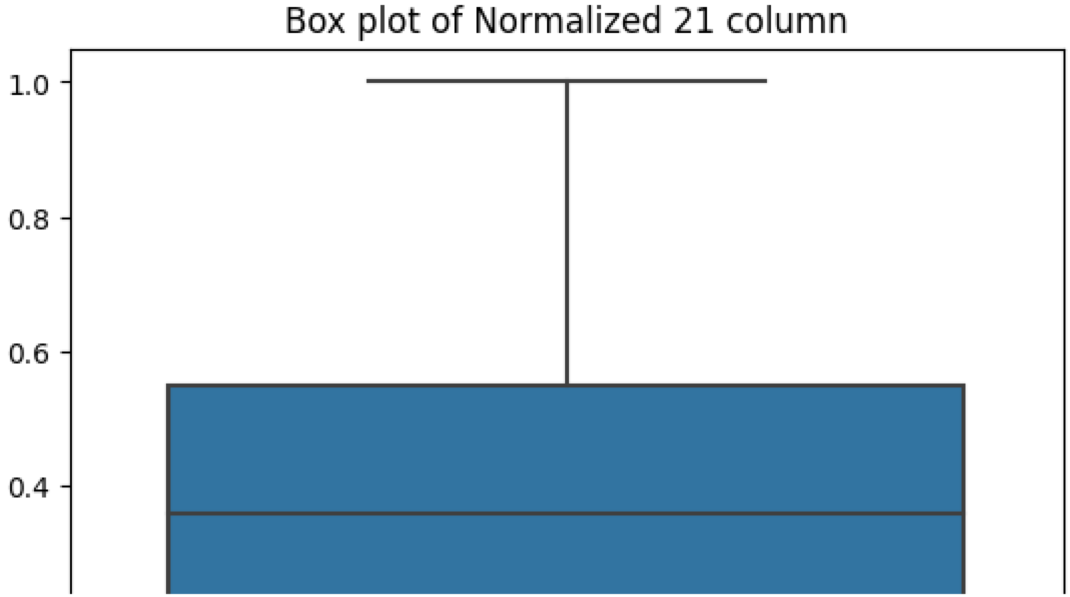


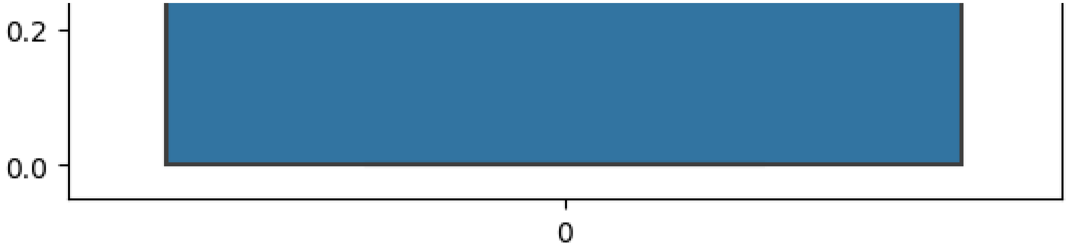
Box plot of Normalized 19 column

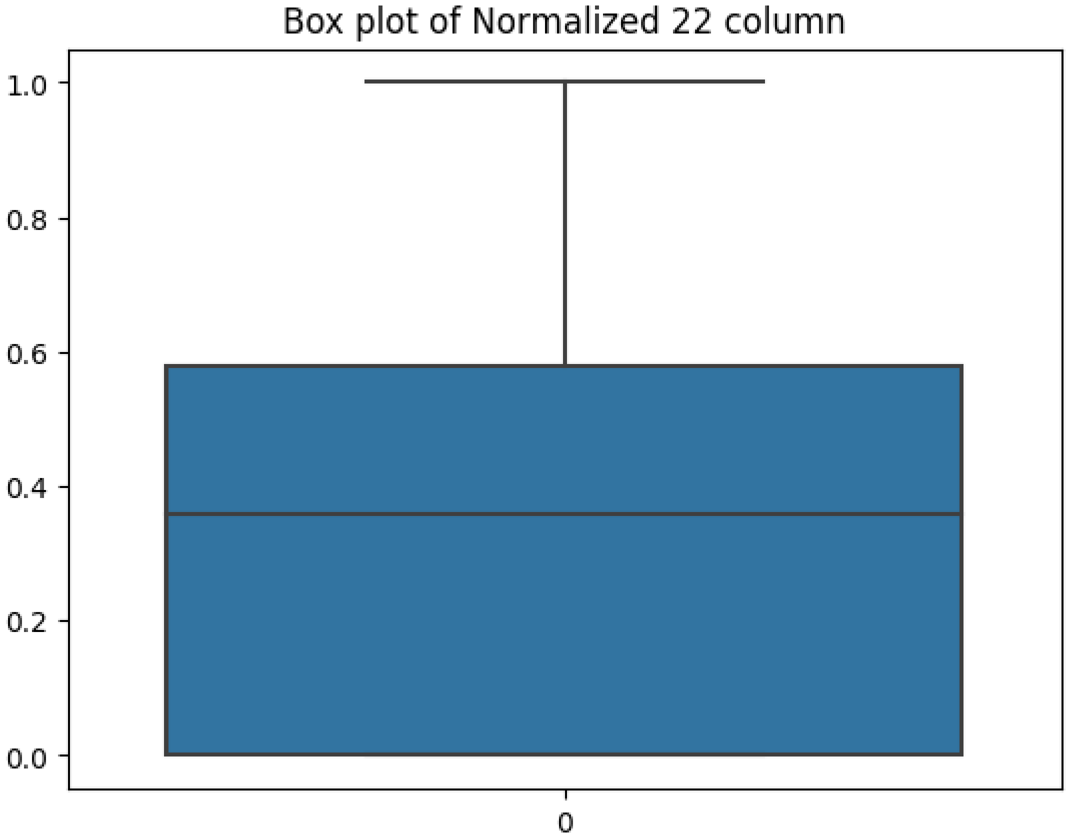


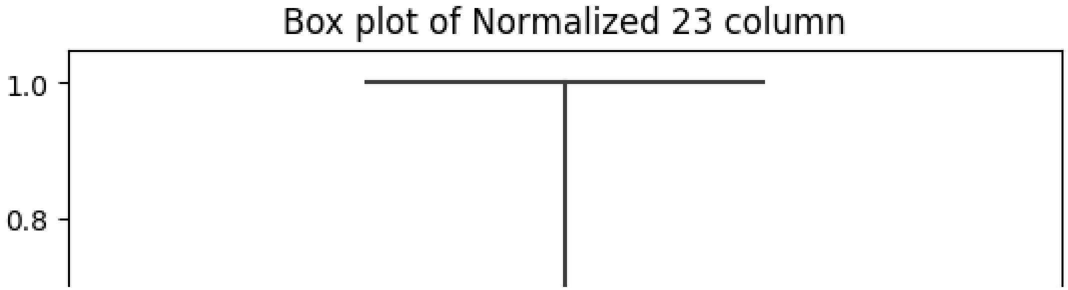
Box plot of Normalized 20 column

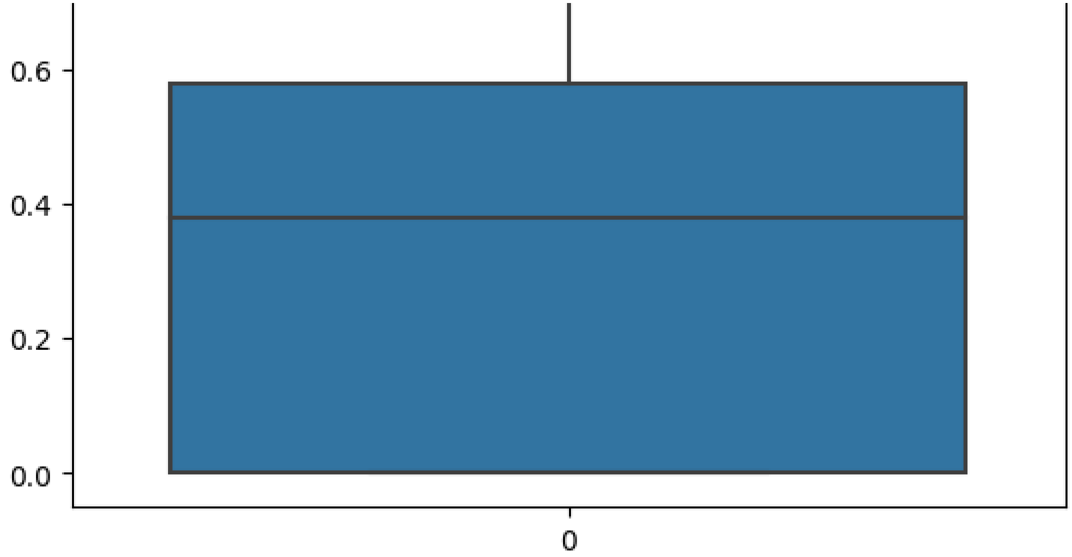

1.0
0.8
0.6
0.4
0.2
0.0
0

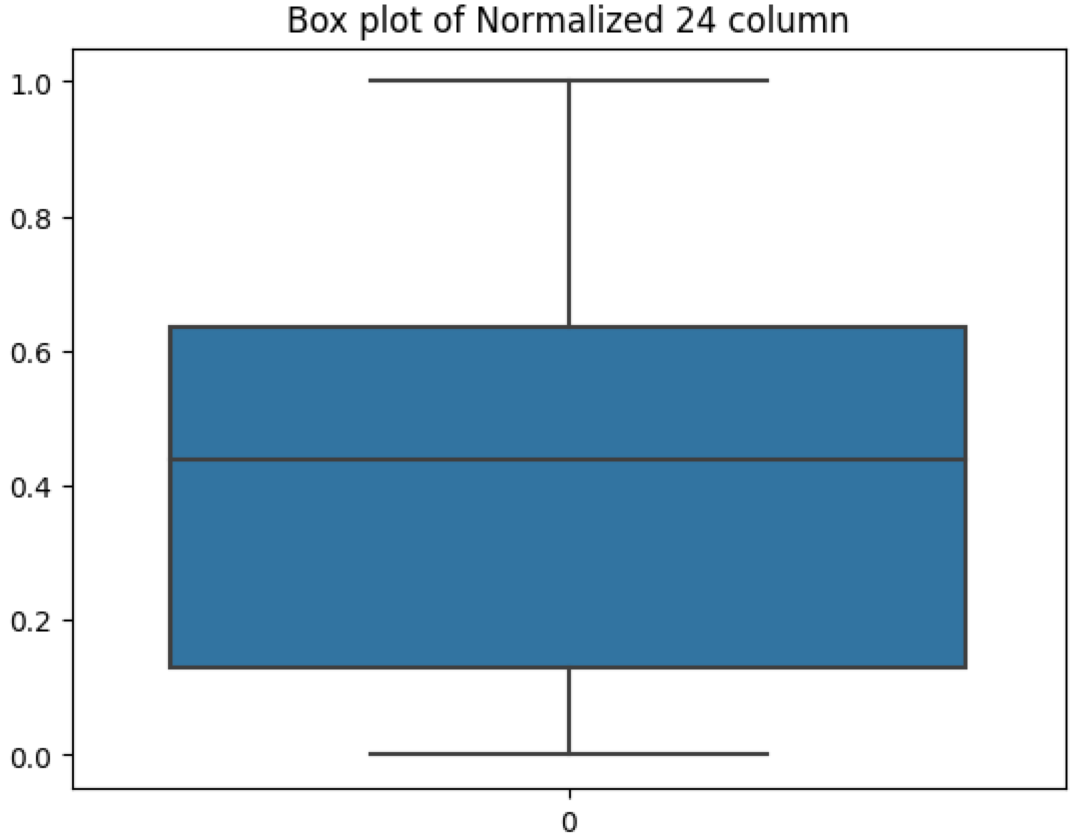

Box plot of Normalized 21 column
1.0
0.8
0.6
0.4

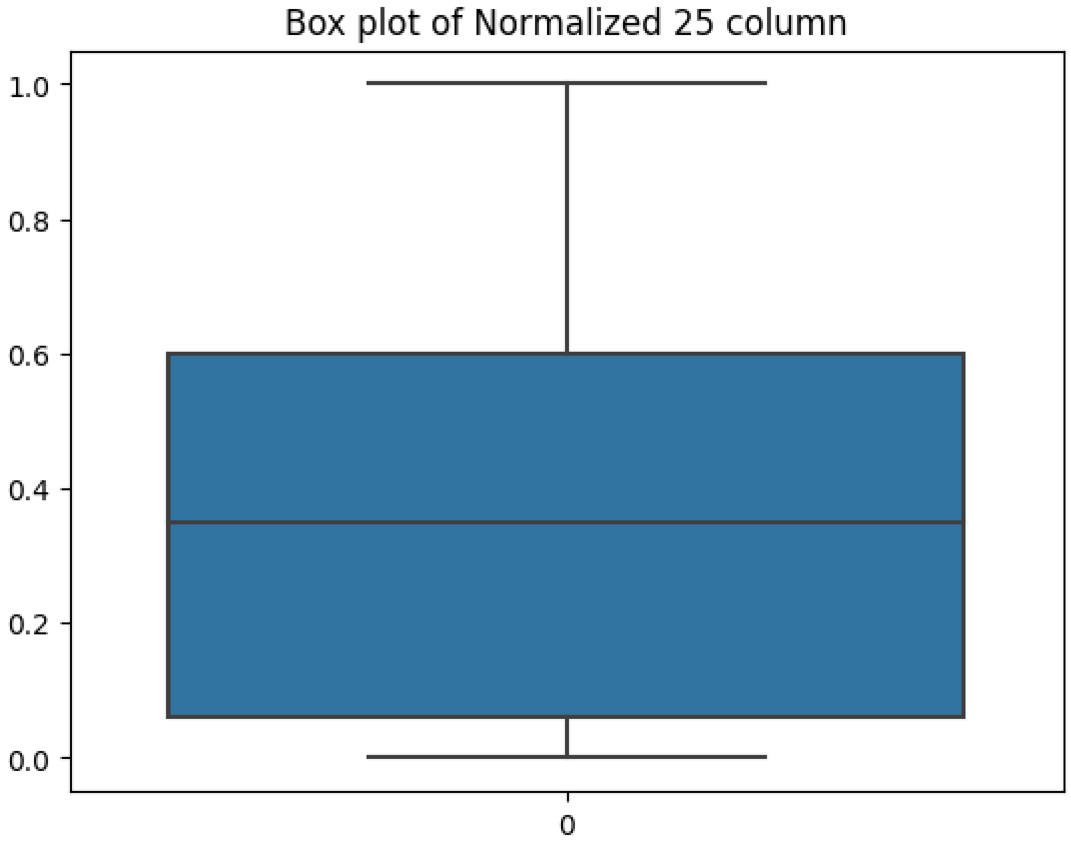

0.2
0.0
0

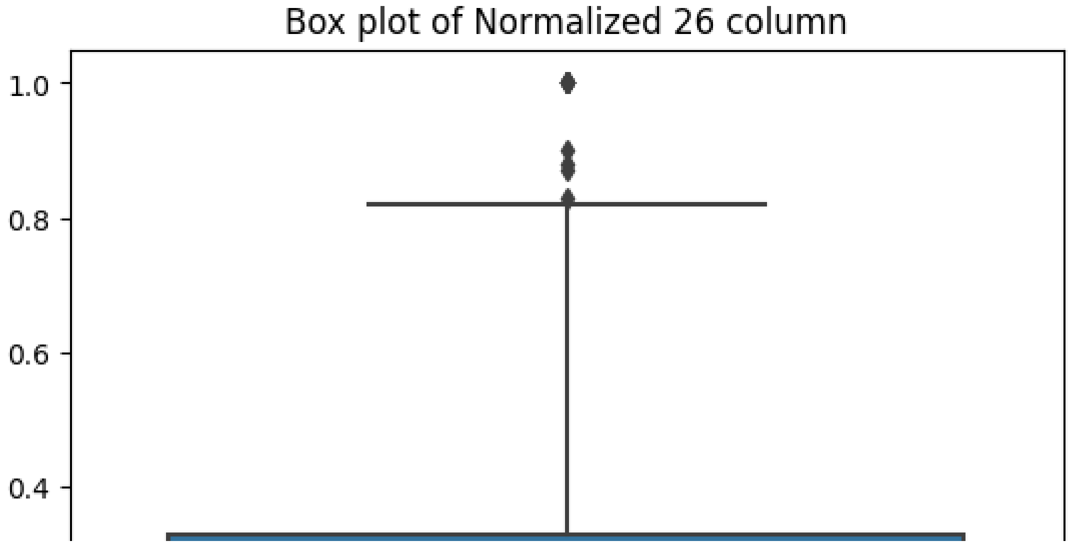

Box plot of Normalized 22 column
1.0
0.8
0.6
0.4
0.2
0.0
0

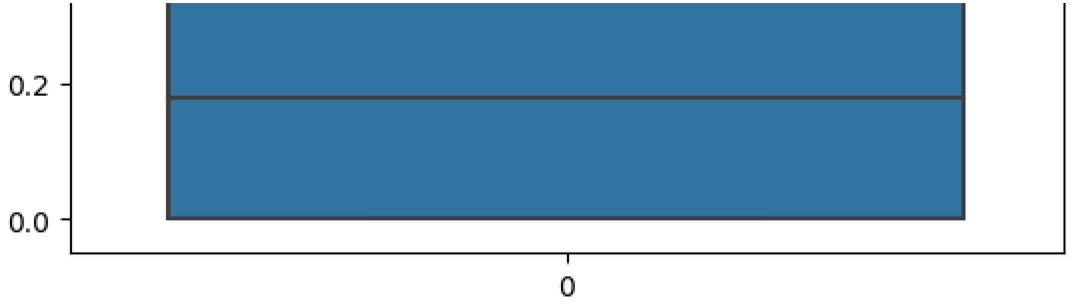

Box plot of Normalized 23 column
1.0
0.8


0.6
0.4
0.2
0.0
0

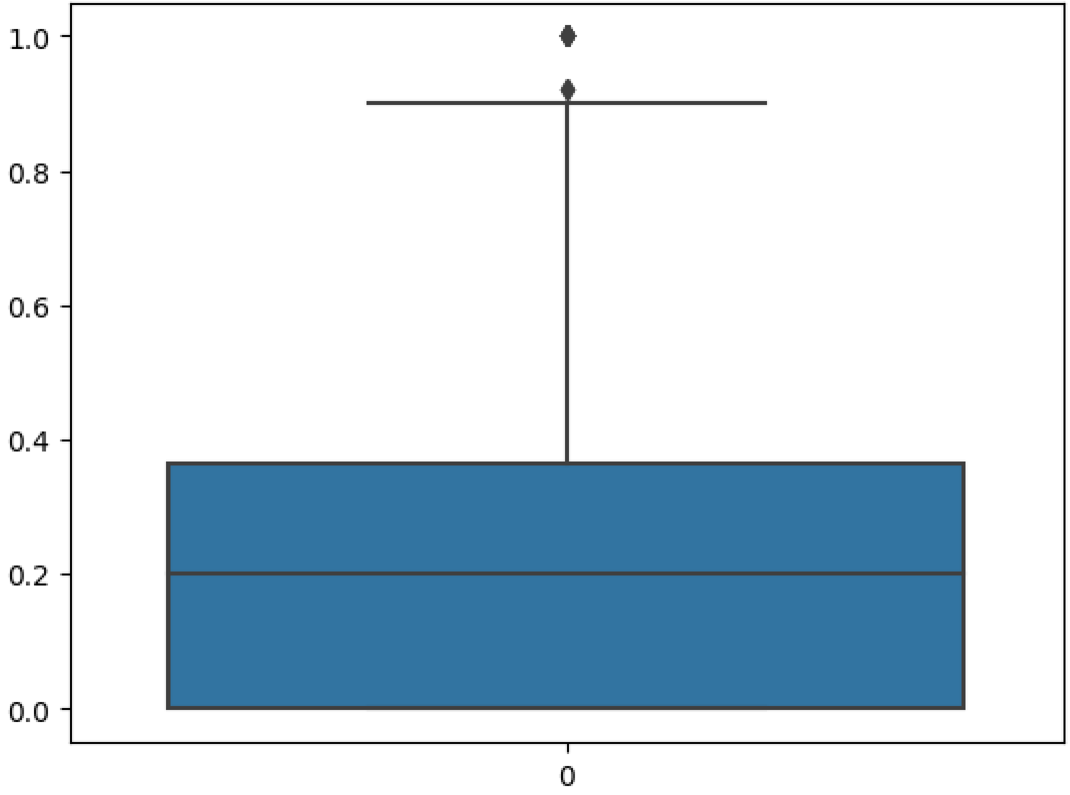

Box plot of Normalized 24 column
1.0
0.8
0.6
0.4
0.2
0.0
0


Box plot of Normalized 25 column
1.0
0.8
0.6
0.4
0.2
0.0
0

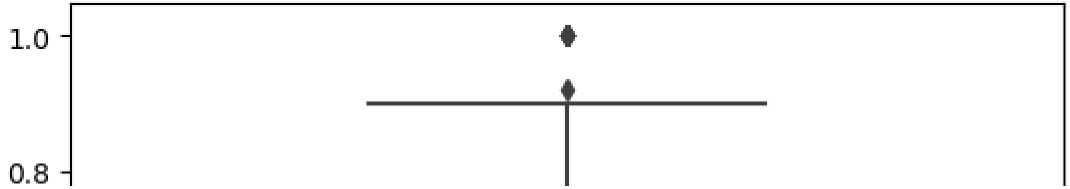

Box plot of Normalized 26 column
1.0
0.8
0.6
0.4

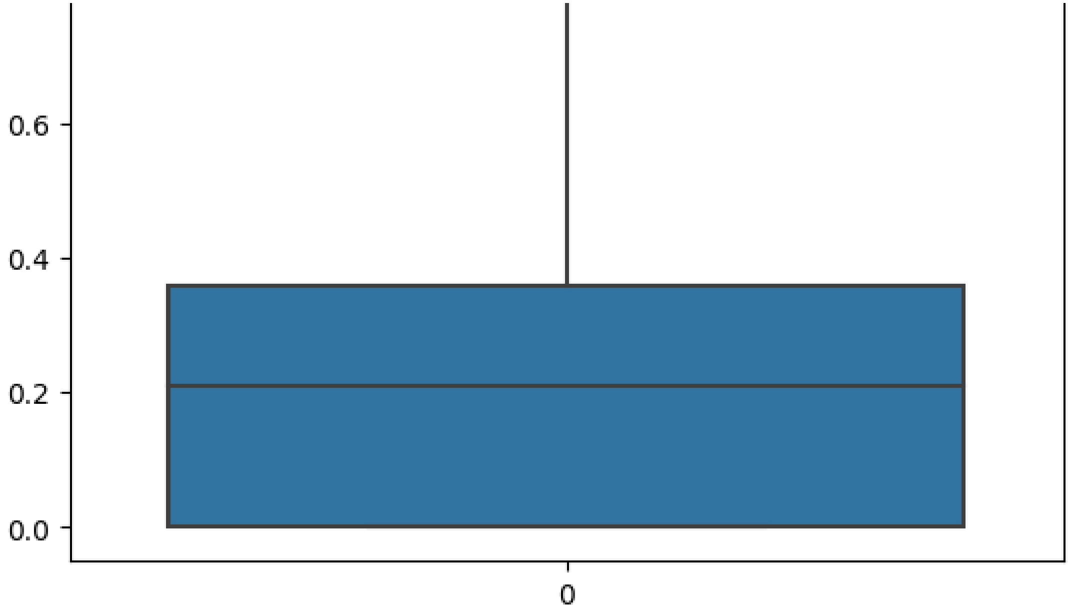


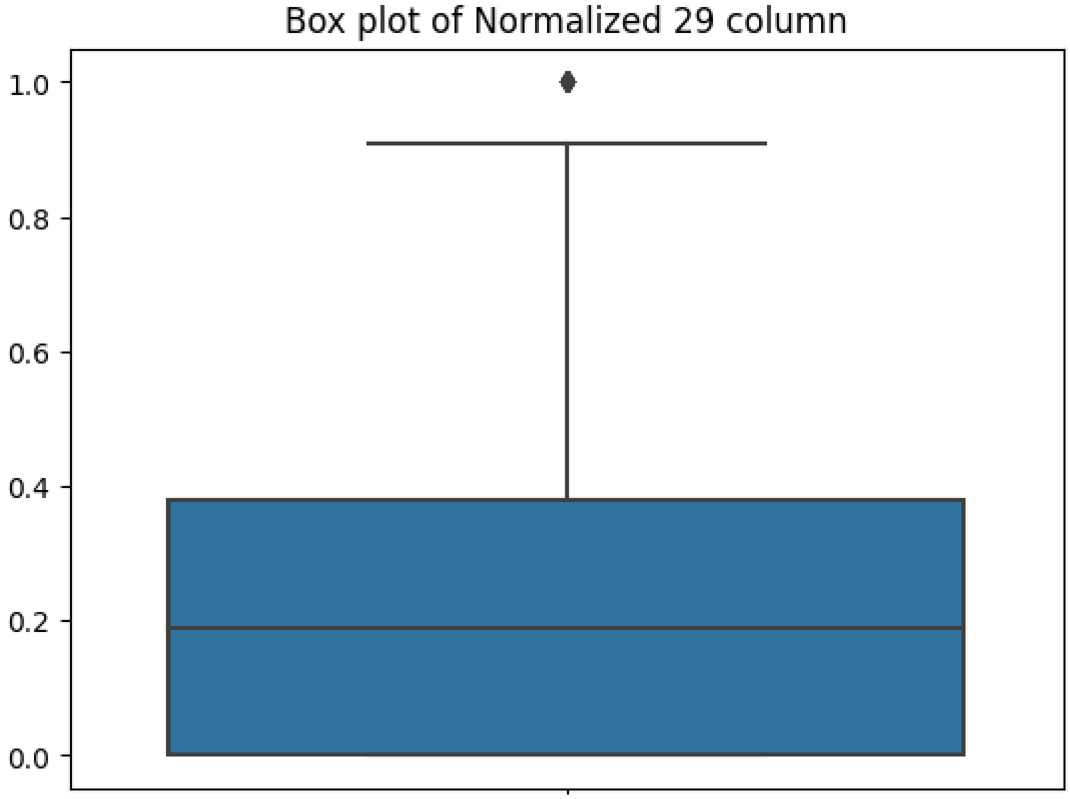
Box plot of Normalized 27 column

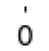


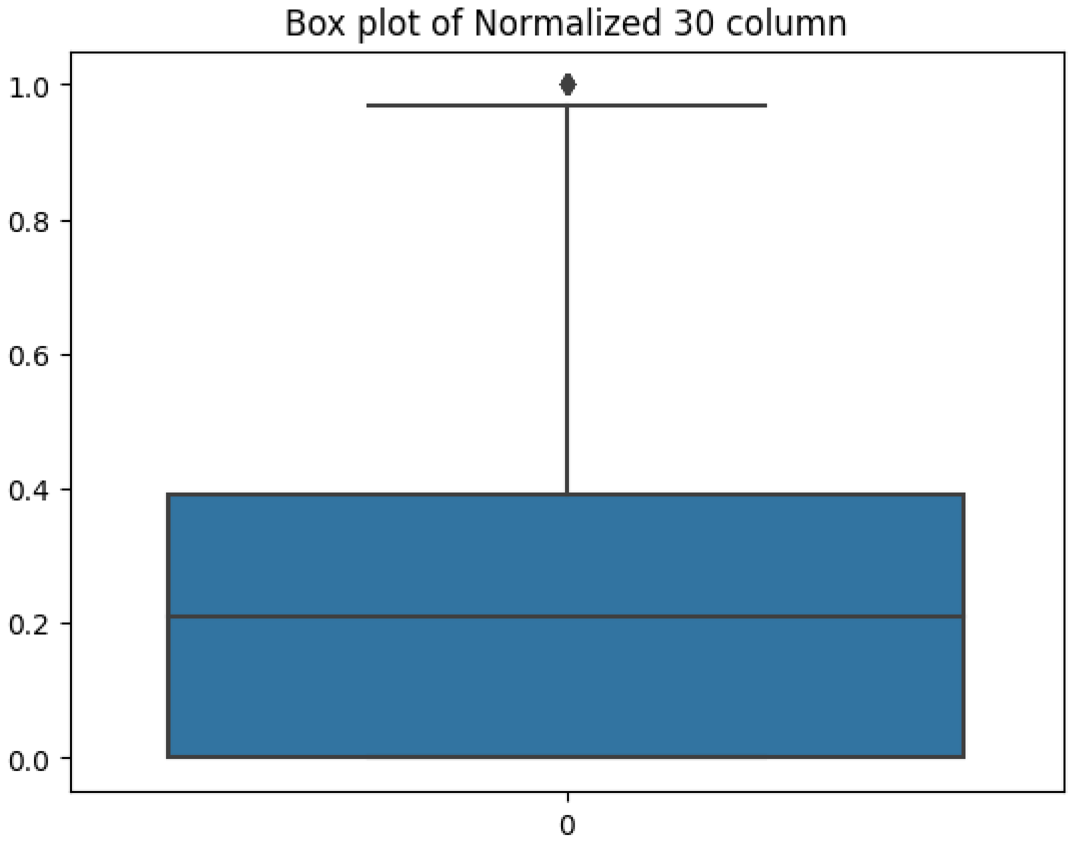
Box plot of Normalized 28 column

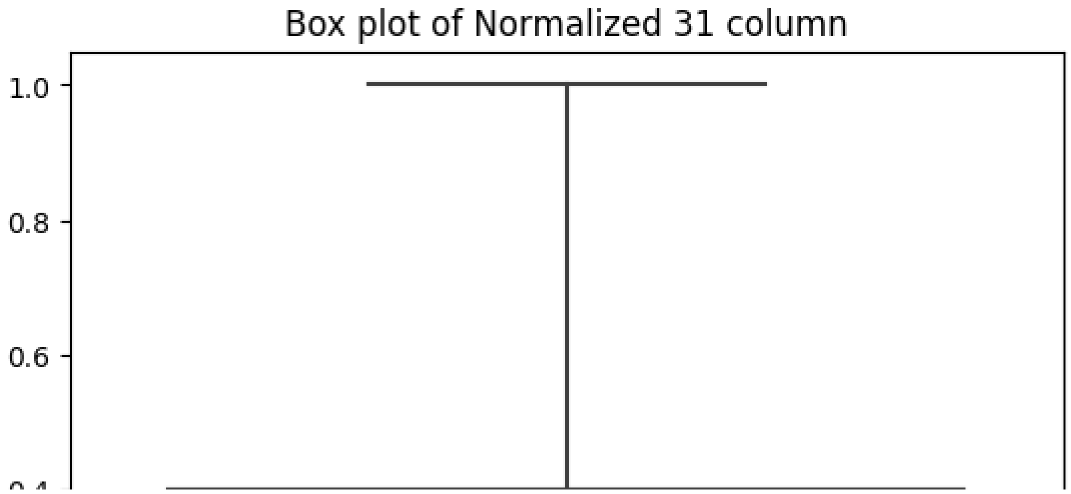


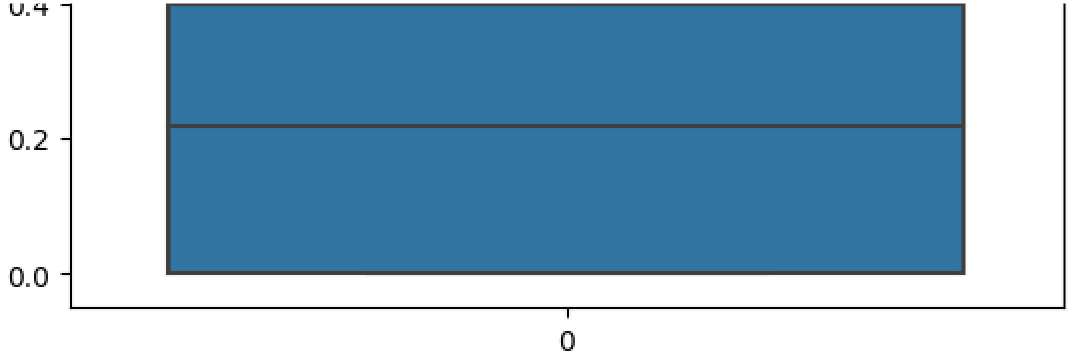

0.6
0.4
0.2
0.0
0

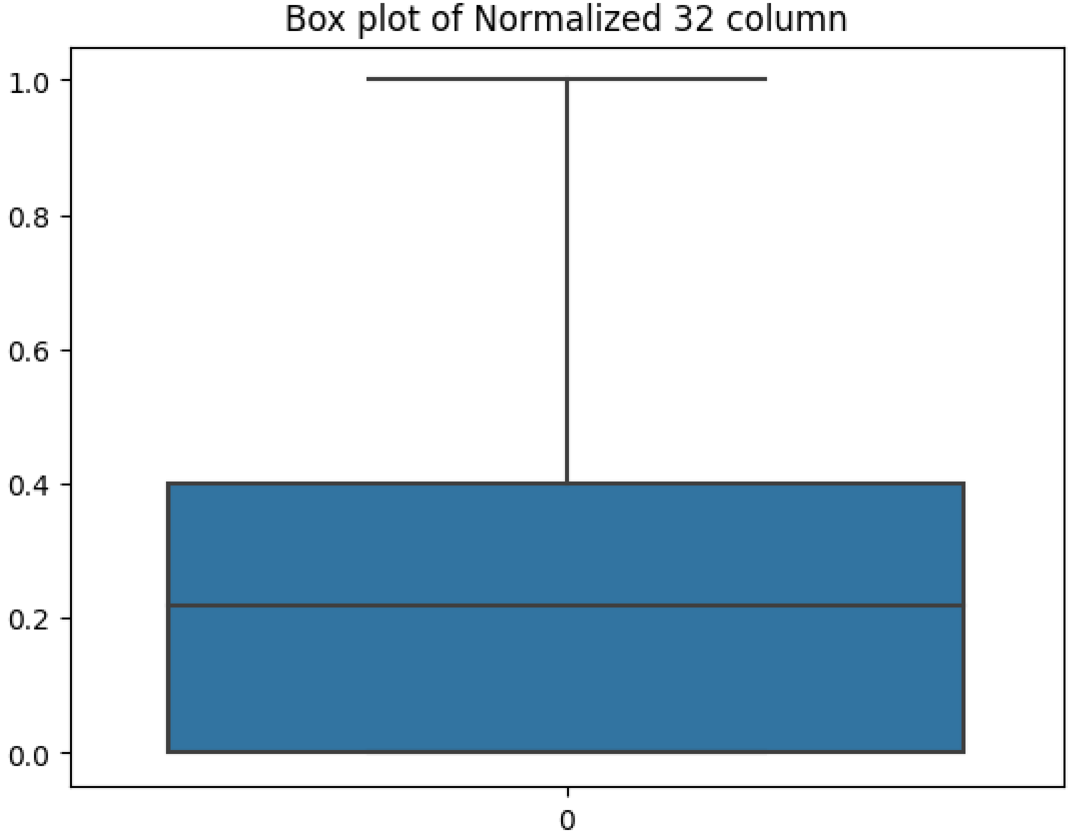

Box plot of Normalized 29 column
1.0
0.8
0.6
0.4
0.2
0.0

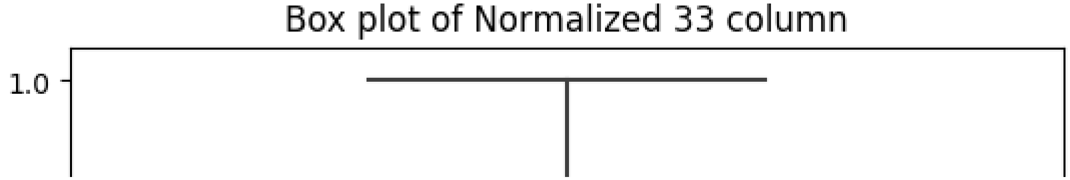

0

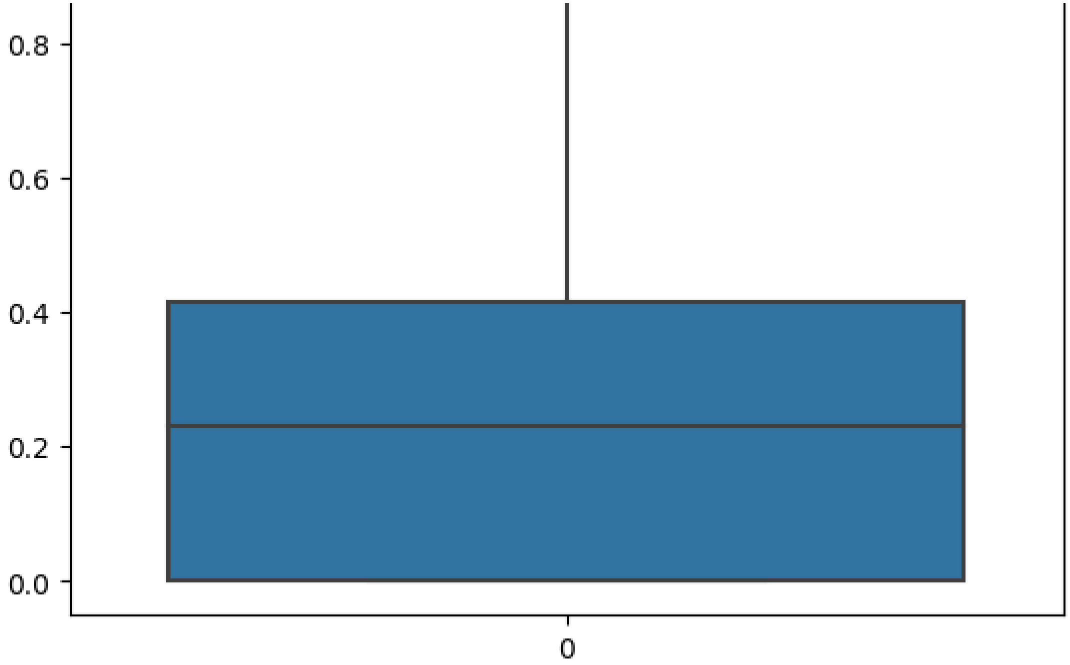

Box plot of Normalized 30 column
1.0
0.8
0.6
0.4
0.2
0.0
0

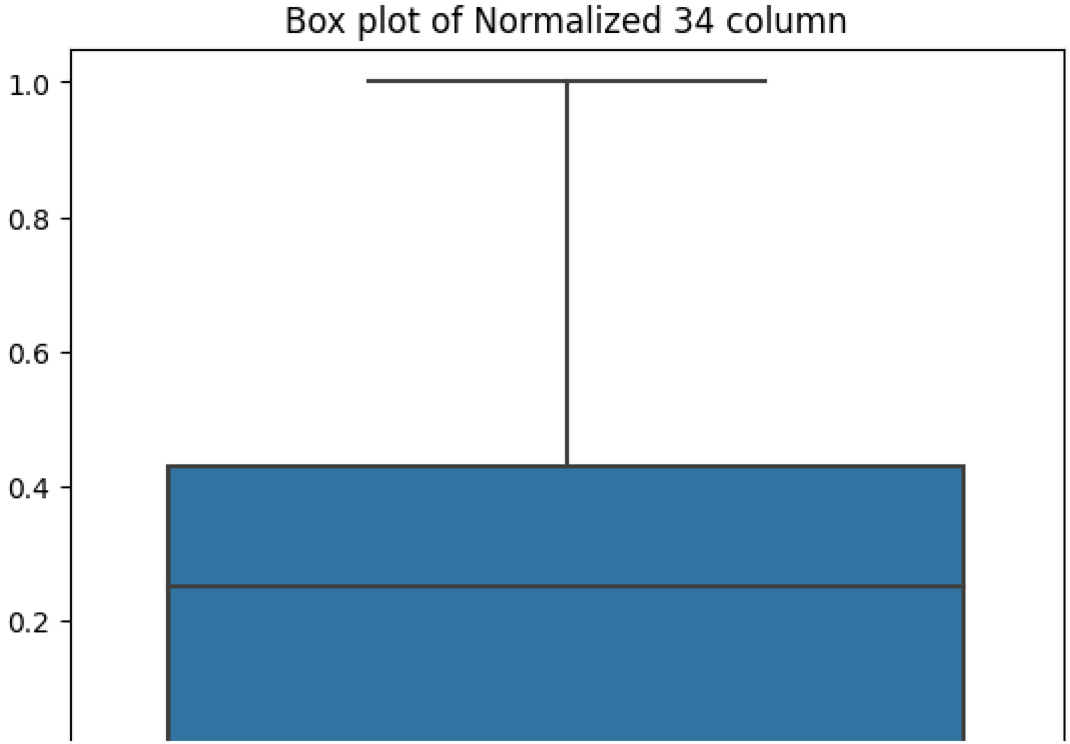

Box plot of Normalized 31 column
1.0
0.8
0.6

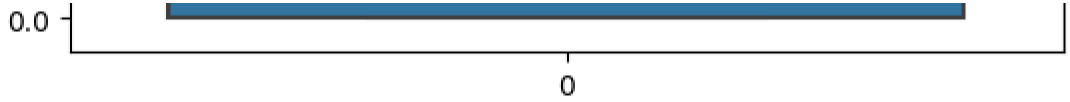

0.4
0.2
0.0
0

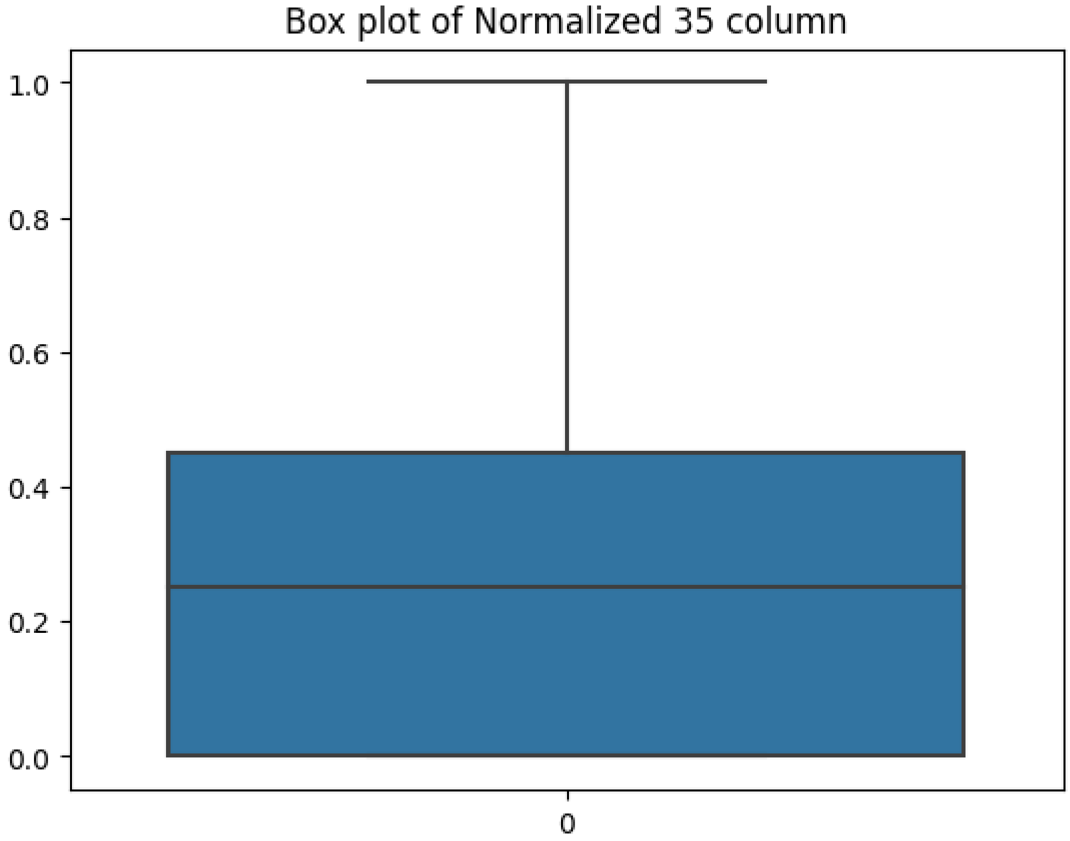

Box plot of Normalized 32 column
1.0
0.8
0.6
0.4
0.2
0.0
0

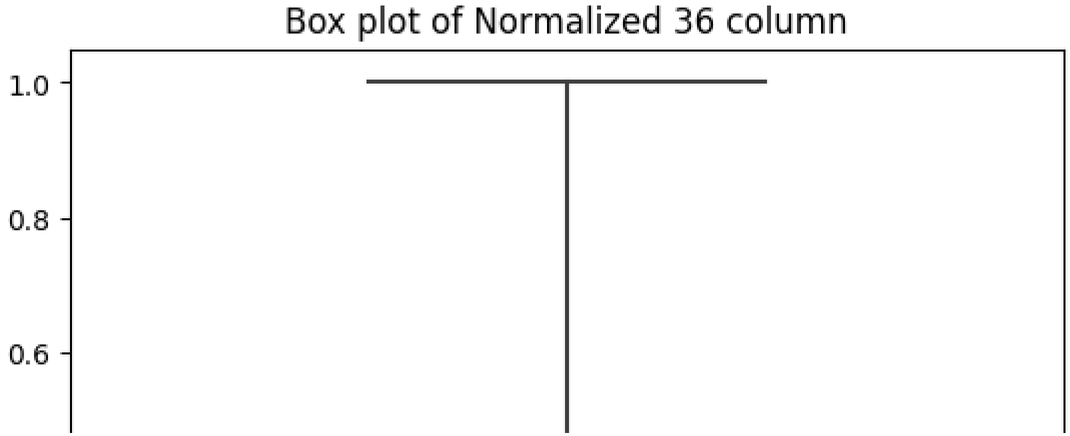

Box plot of Normalized 33 column
1.0

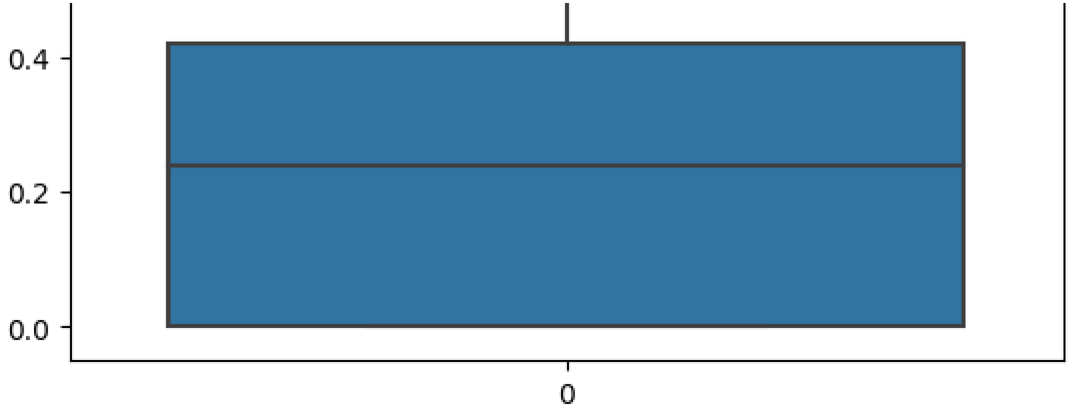

0.8
0.6
0.4
0.2
0.0
0

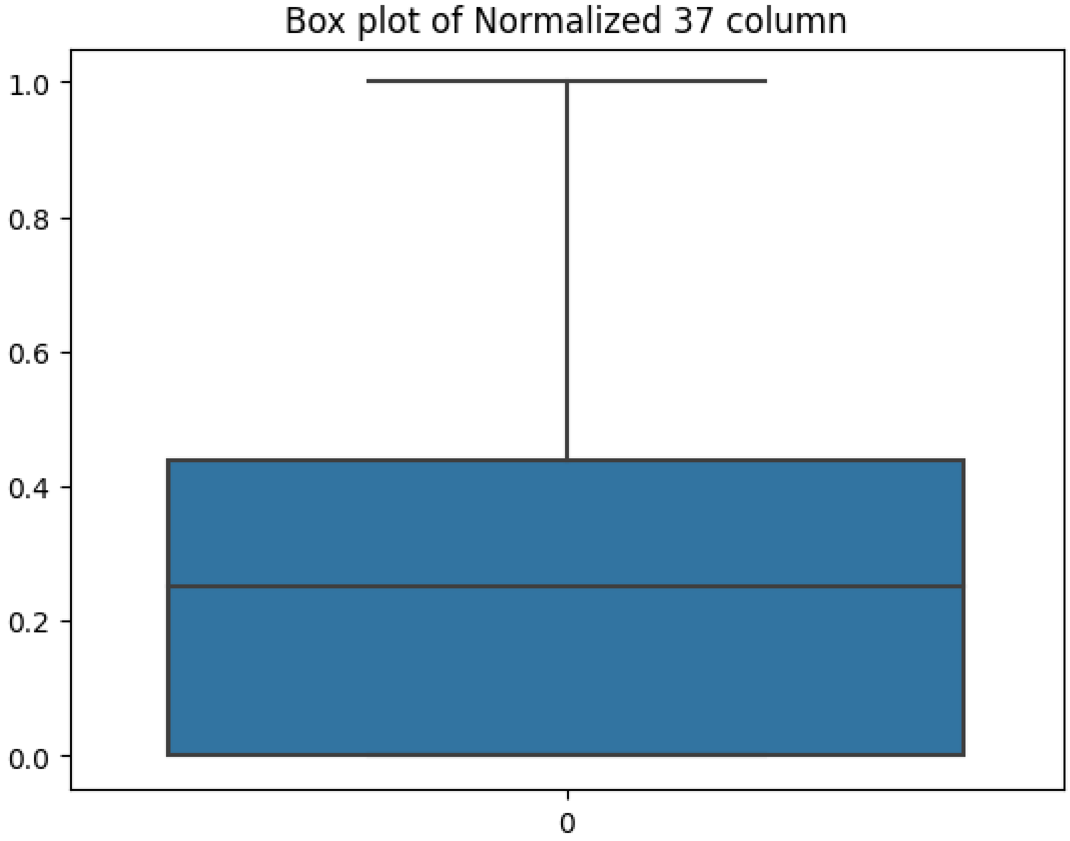

Box plot of Normalized 34 column
1.0
0.8
0.6
0.4
0.2

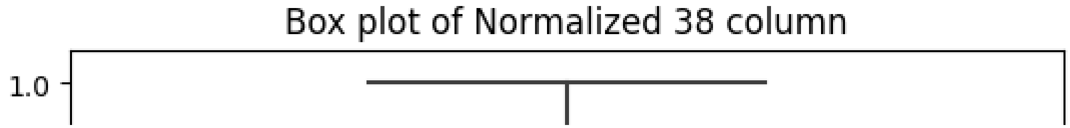

0.0
0

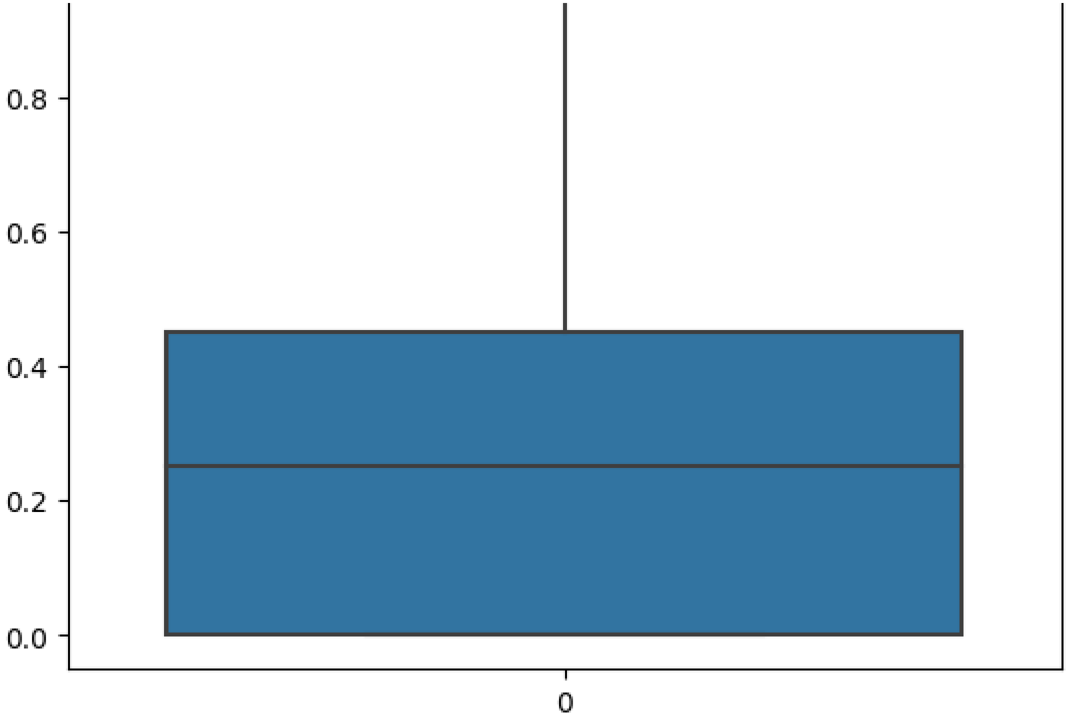

Box plot of Normalized 35 column
1.0
0.8
0.6
0.4
0.2
0.0
0

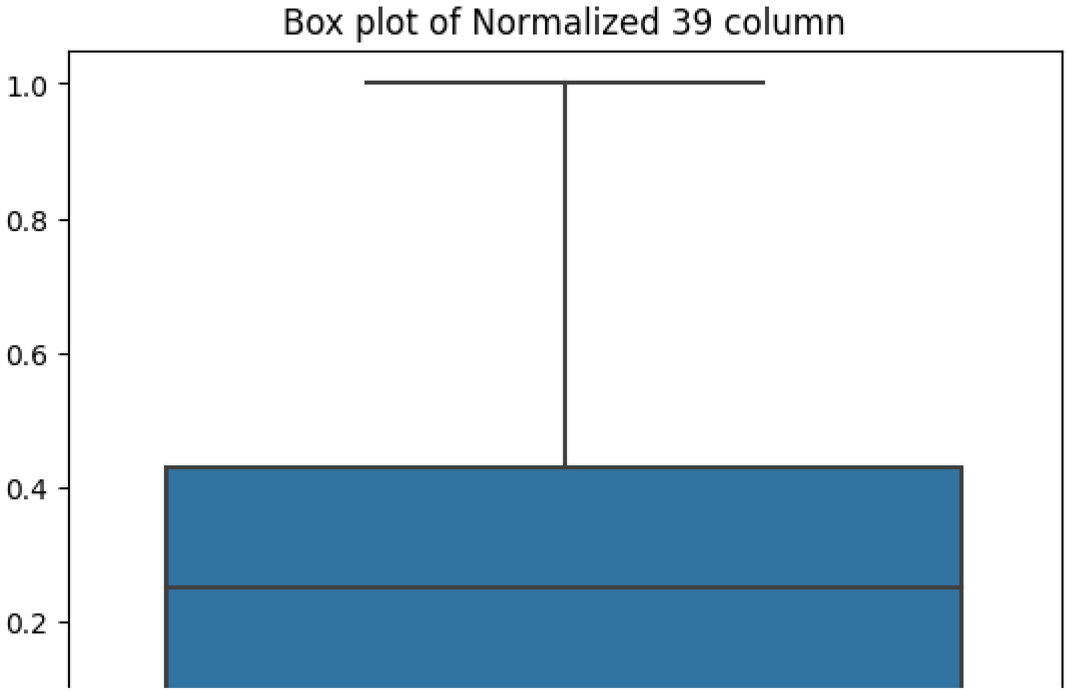

Box plot of Normalized 36 column
1.0
0.8
0.6

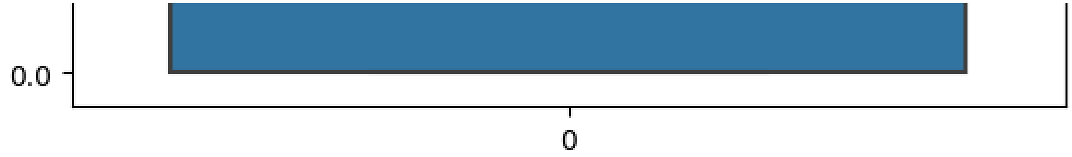

0.4
0.2
0.0
0

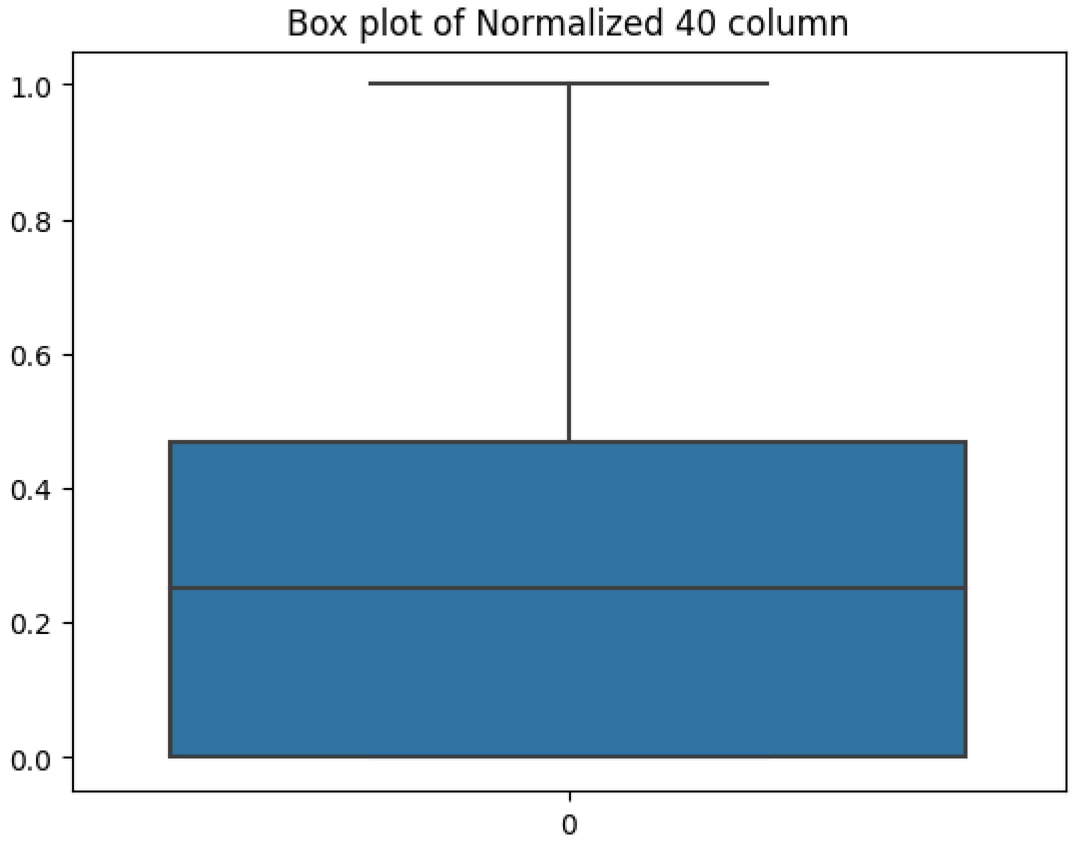

Box plot of Normalized 37 column
1.0
0.8
0.6
0.4
0.2
0.0
0

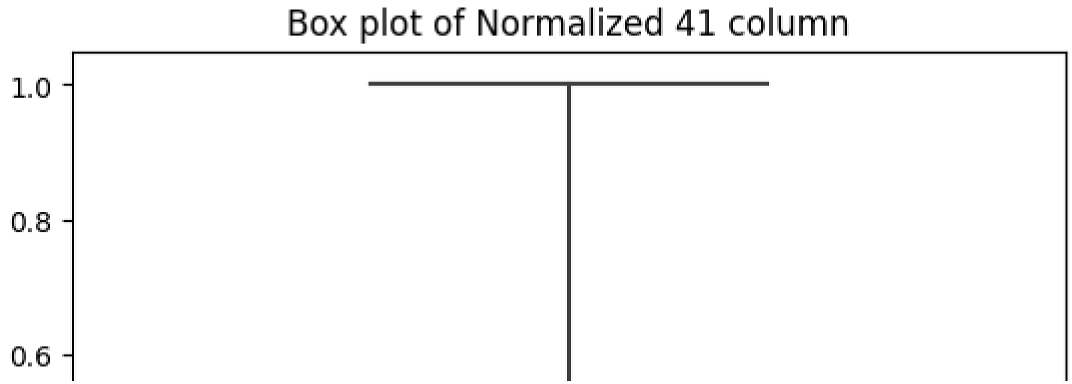

Box plot of Normalized 38 column
1.0

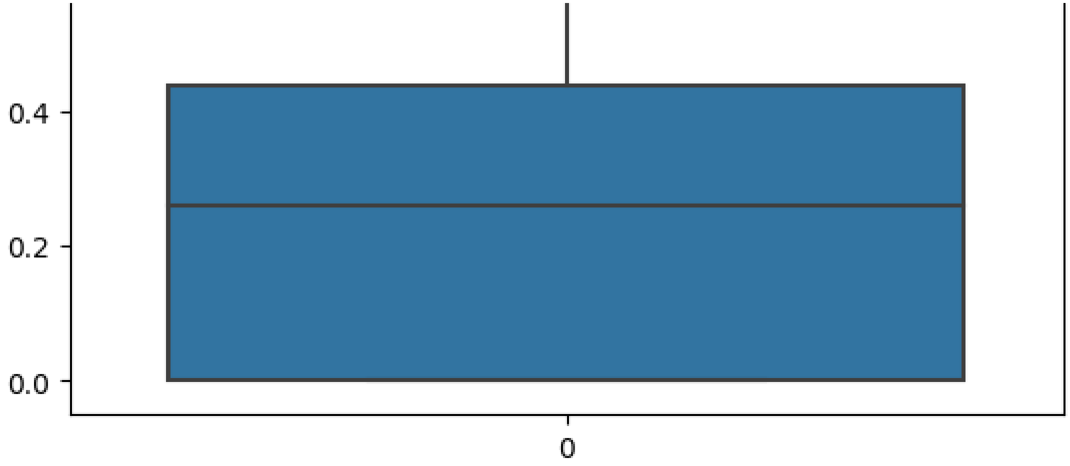

0.8
0.6
0.4
0.2
0.0
0

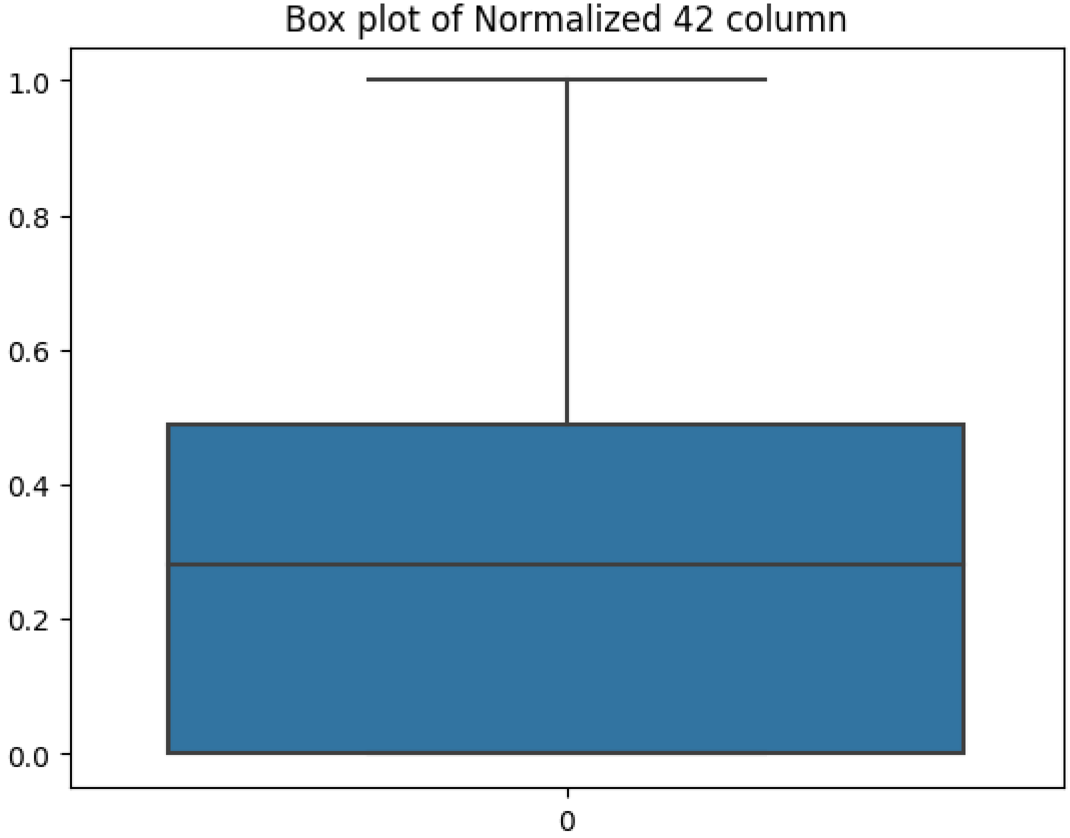

Box plot of Normalized 39 column
1.0
0.8
0.6
0.4
0.2


0.0
0

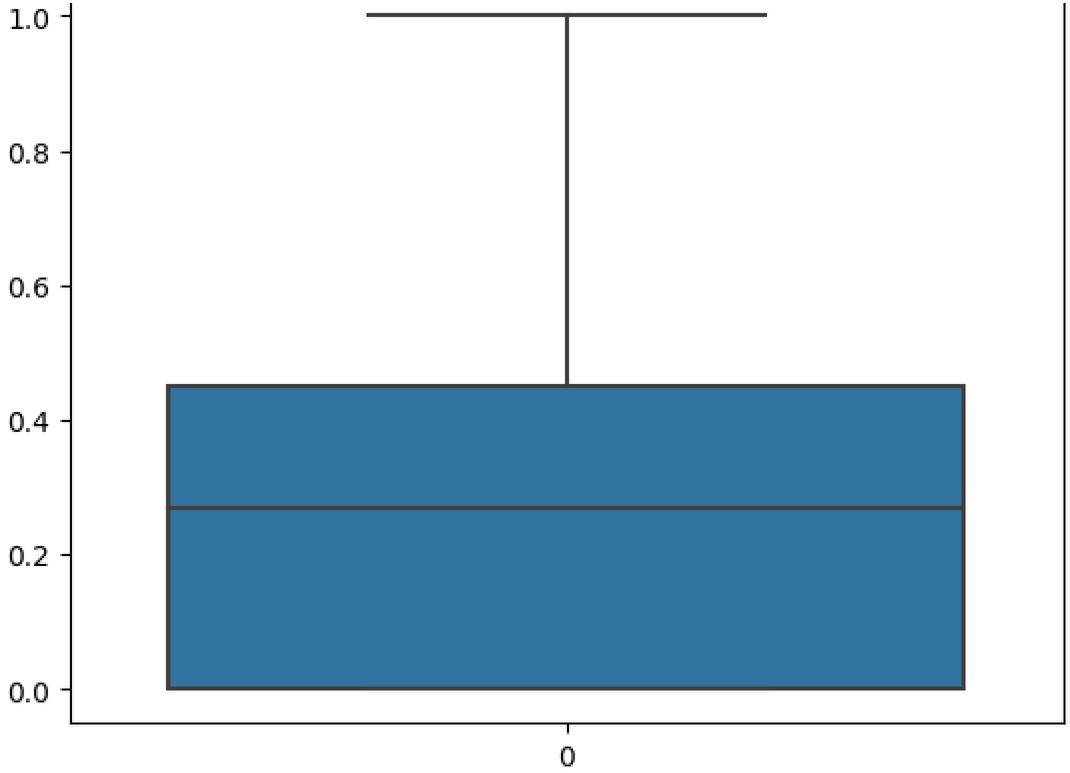

Box plot of Normalized 40 column
1.0
0.8
0.6
0.4
0.2
0.0
0

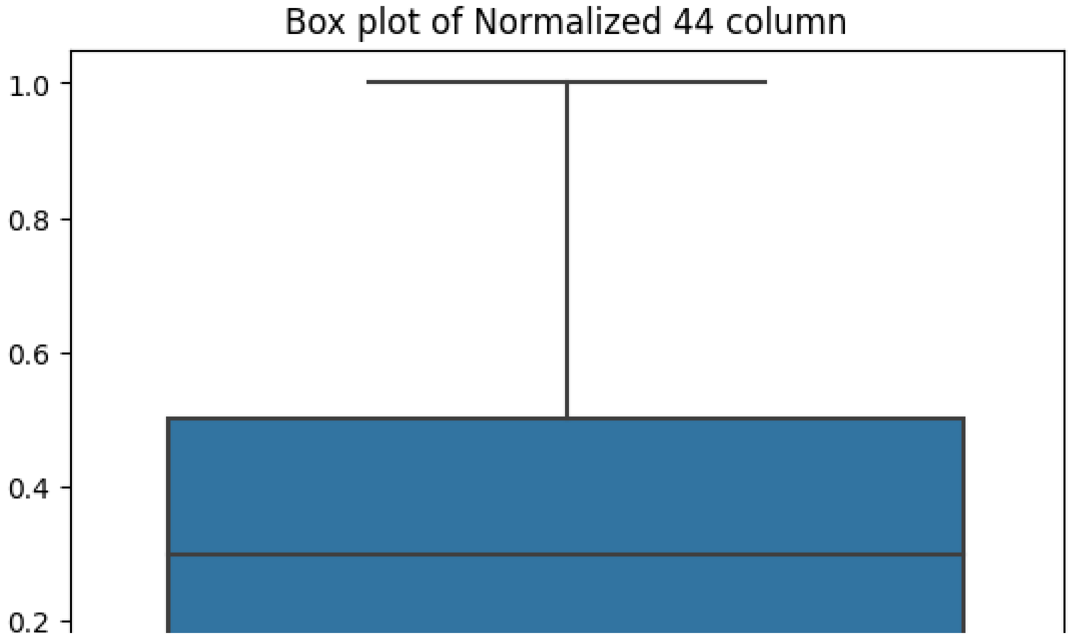

Box plot of Normalized 41 column
1.0
0.8
0.6

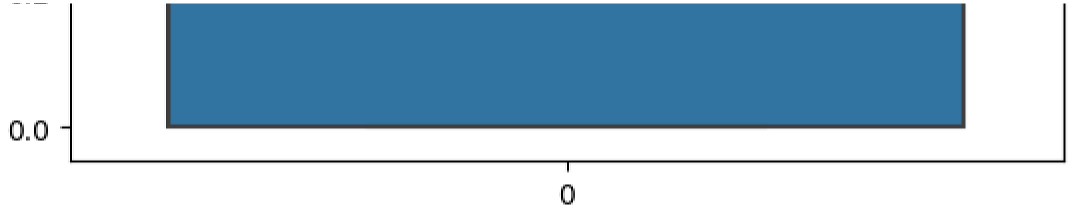

0.4
0.2
0.0
0

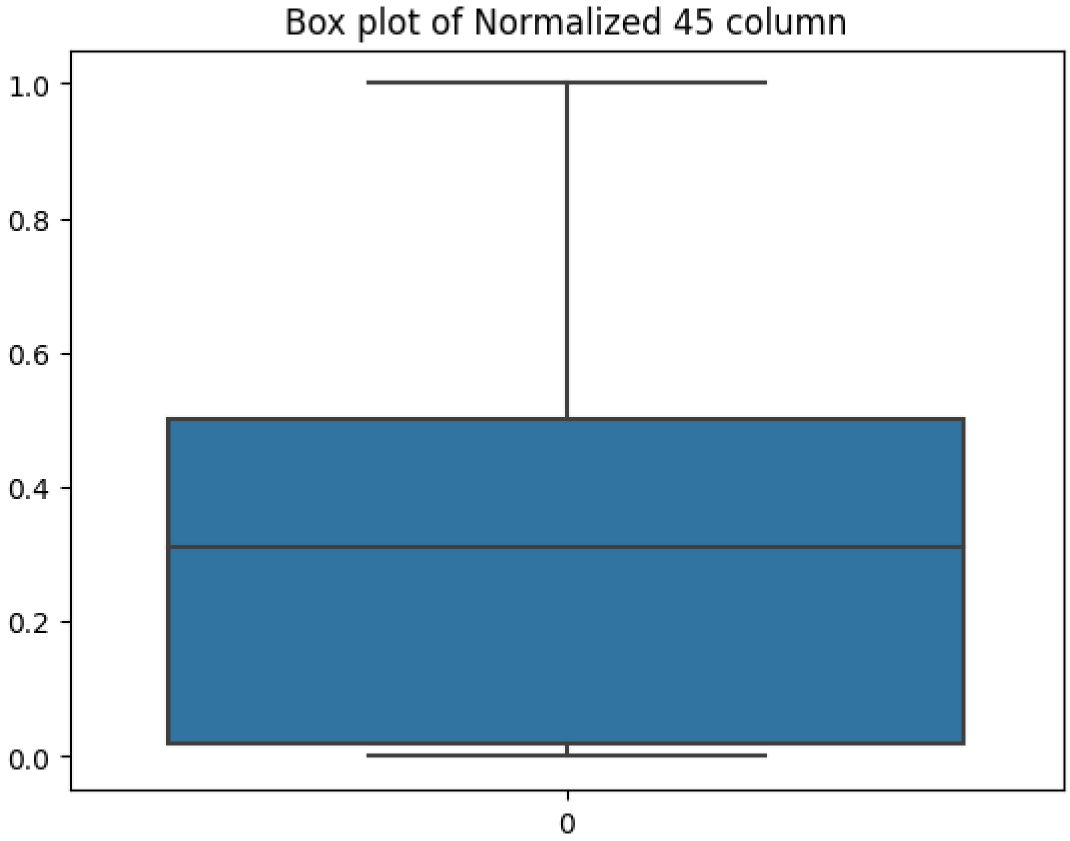

Box plot of Normalized 42 column
1.0
0.8
0.6
0.4
0.2
0.0
0

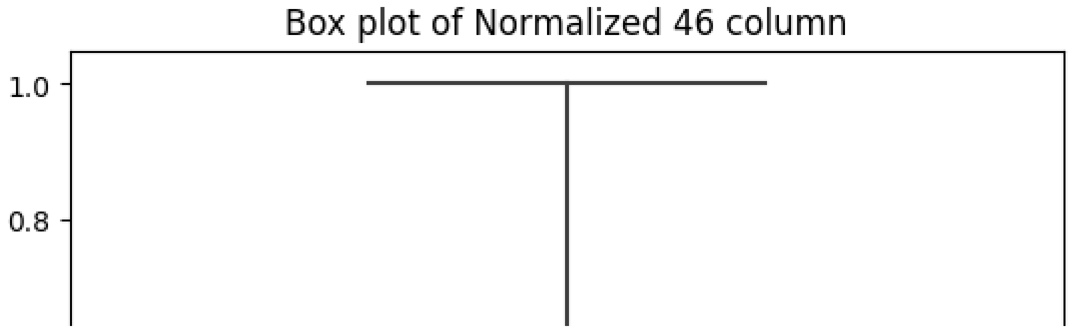
Box plot of Normalized 43 column

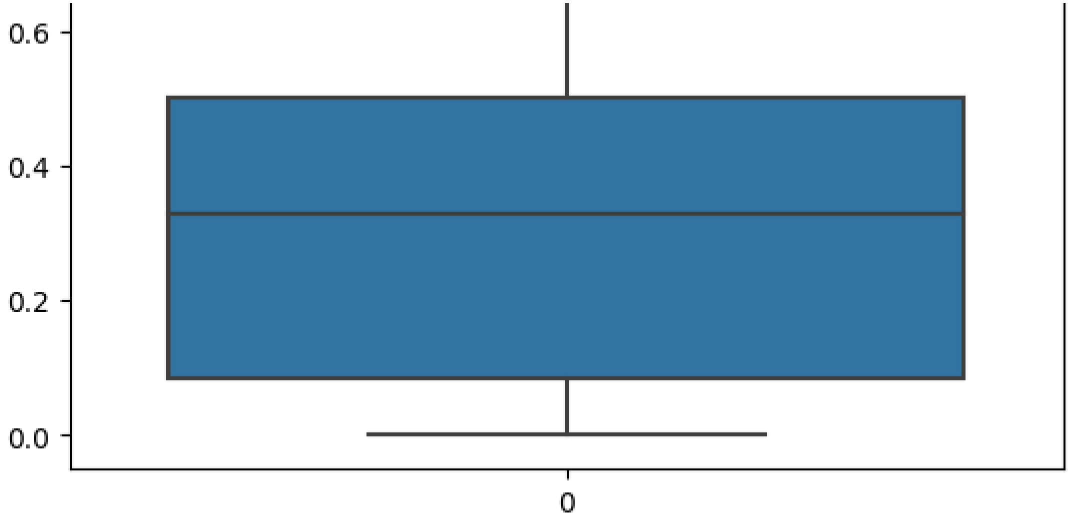

1.0
0.8
0.6
0.4
0.2
0.0
0


Box plot of Normalized 44 column
1.0
0.8
0.6
0.4
0.2

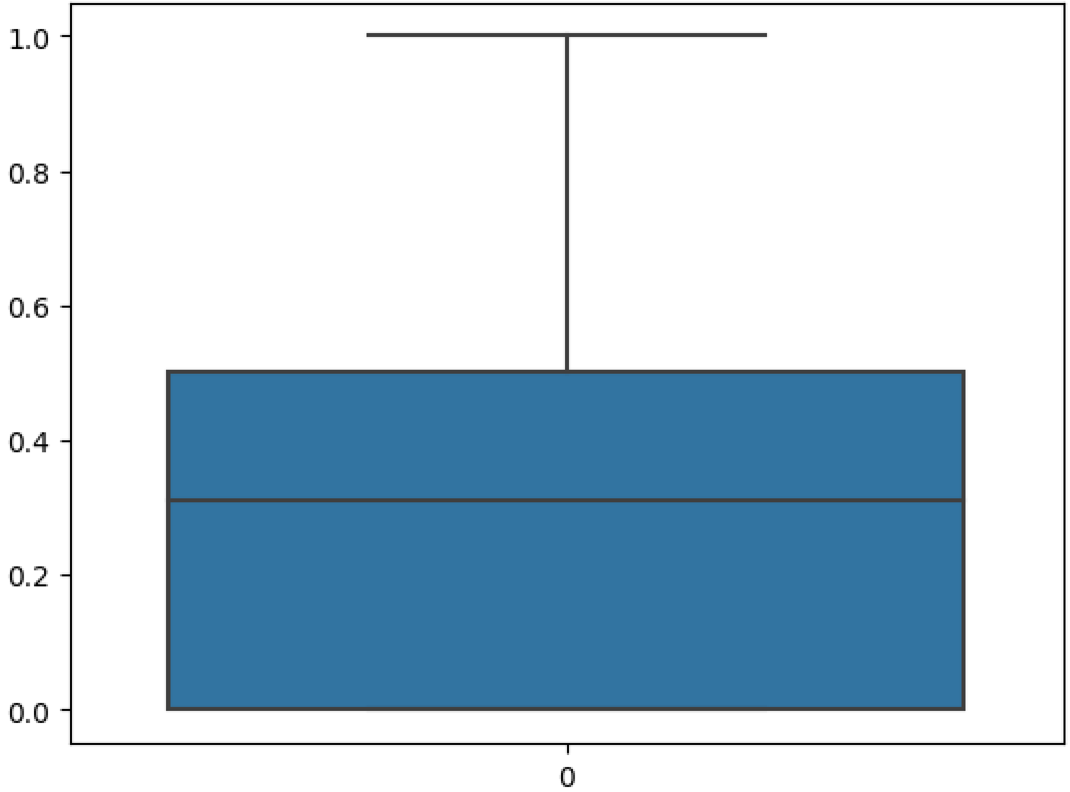

0.0
0


Box plot of Normalized 45 column
1.0
0.8
0.6
0.4
0.2
0.0
0

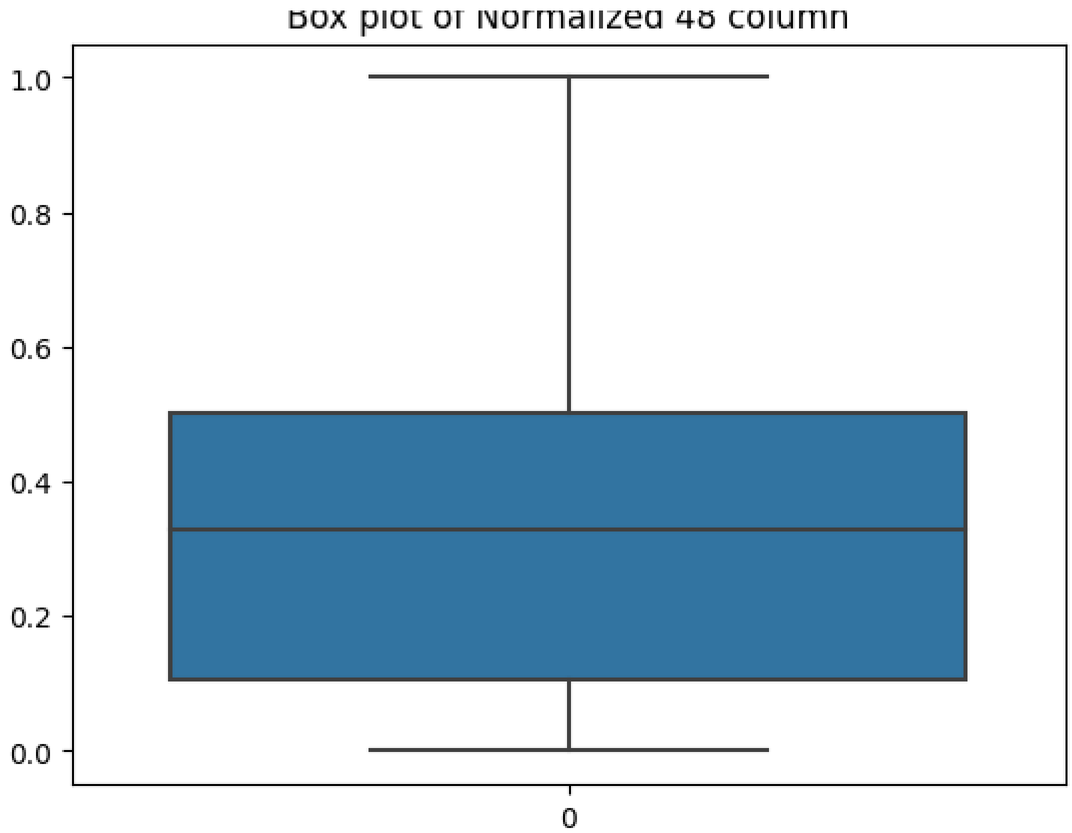

Box plot of Normalized 46 column
1.0
0.8

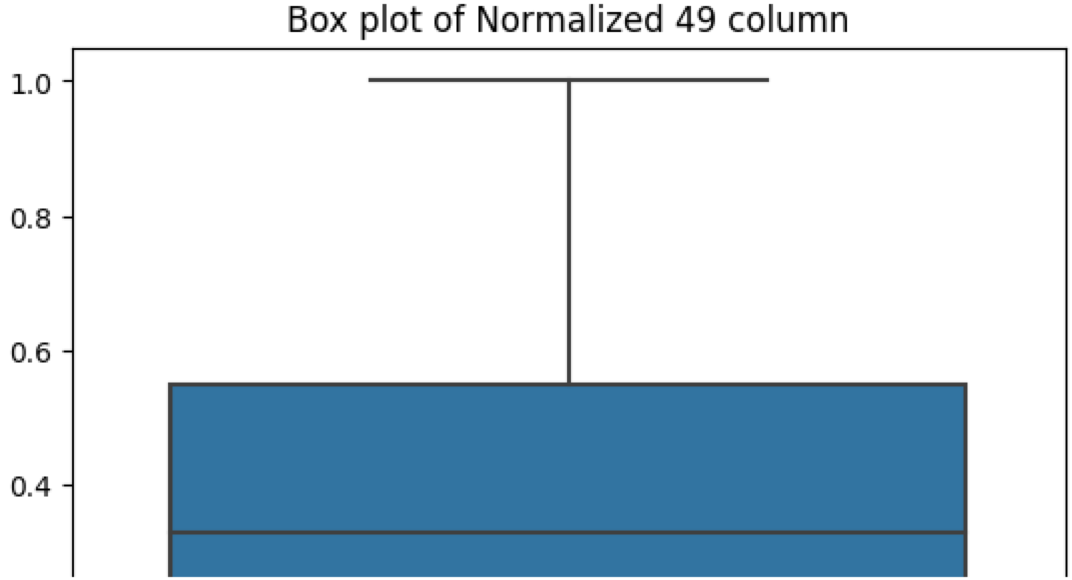


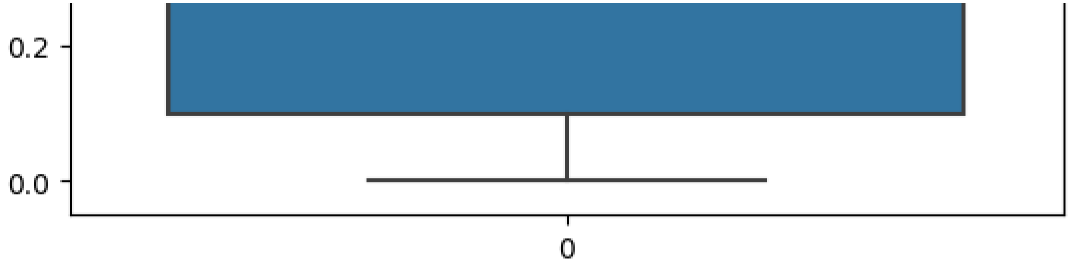
Box plot of Normalized 47 column

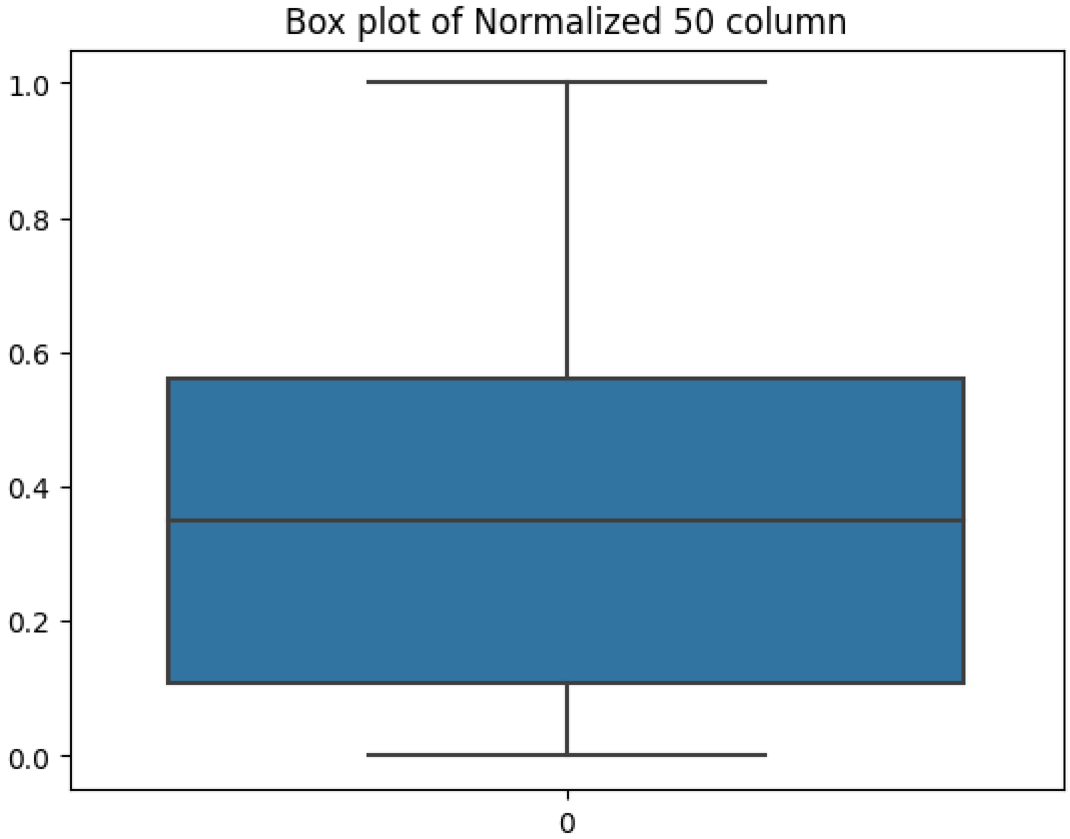


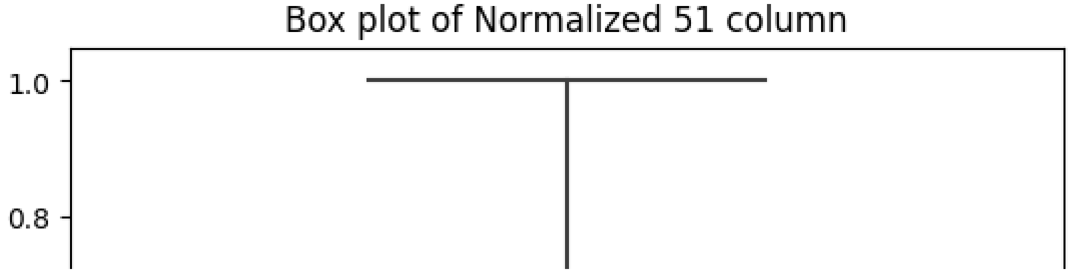
Box plot of Normalized 48 column

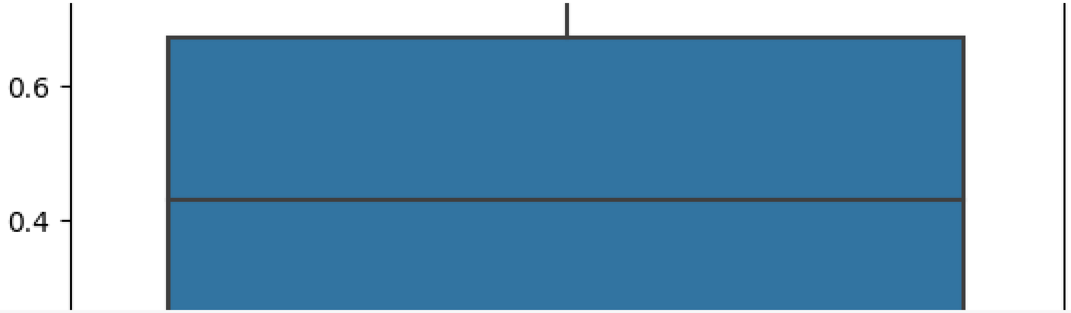

Box plot of Normalized 48 column
1.0
0.8
0.6
0.4
0.2
0.0
0


Box plot of Normalized 49 column
1.0
0.8
0.6
0.4

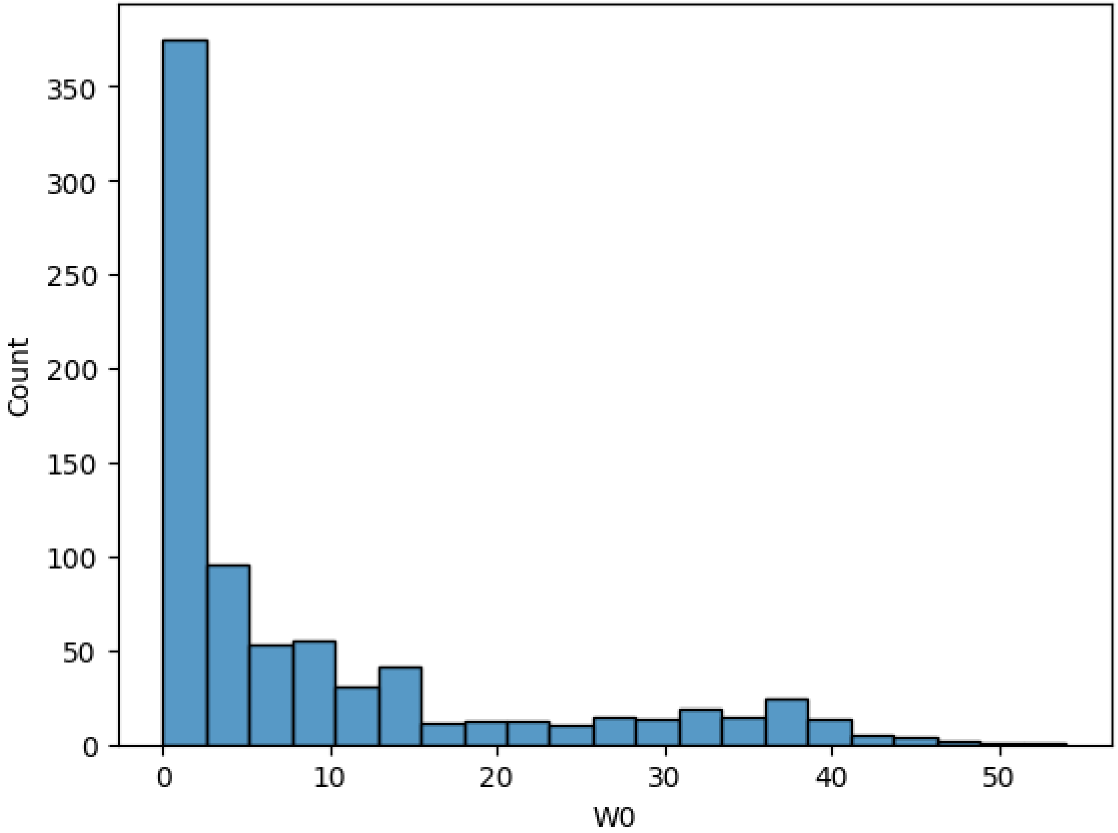

0.2
0.0
0


Box plot of Normalized 50 column
1.0
0.8
0.6
0.4
0.2
0.0
0

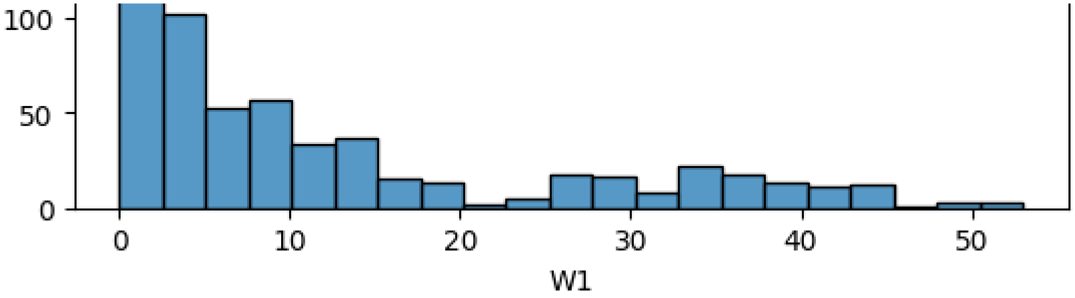

Box plot of Normalized 51 column
1.0
0.8

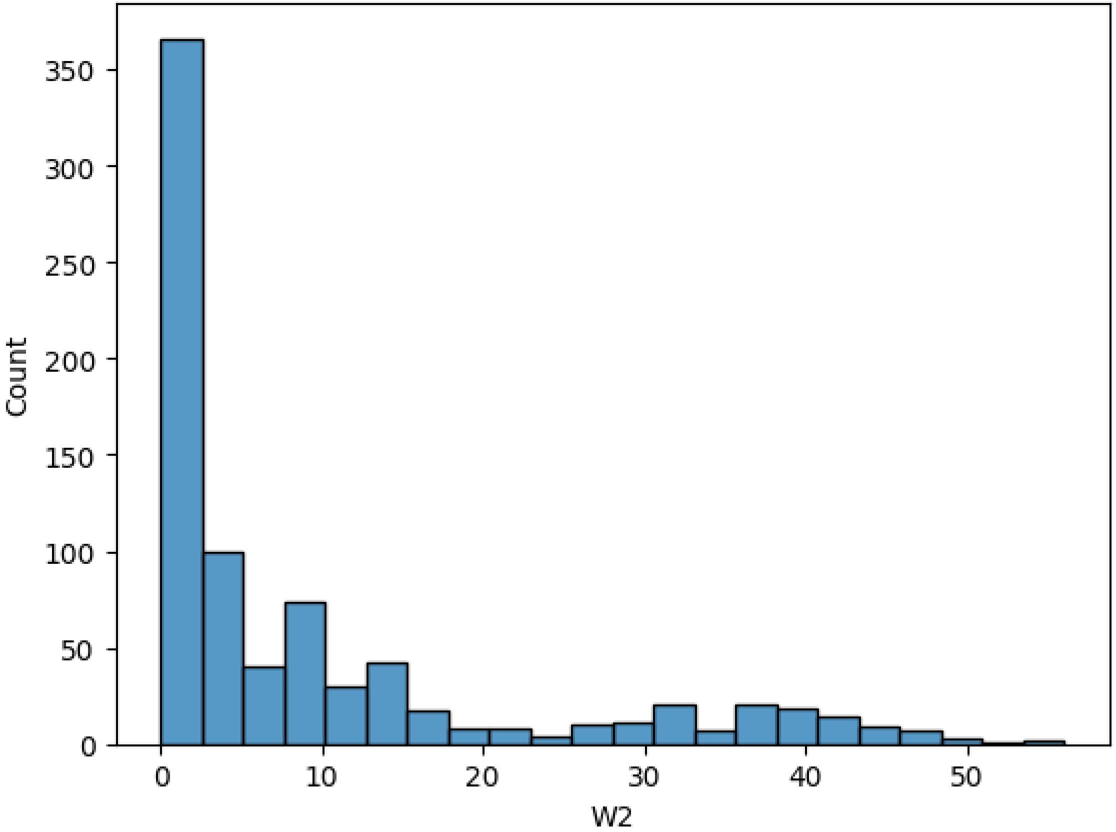


```
for i in file.select_dtypes(exclude=['object']).columns:
    sbs.histplot(file[i])
    plt.show()
```

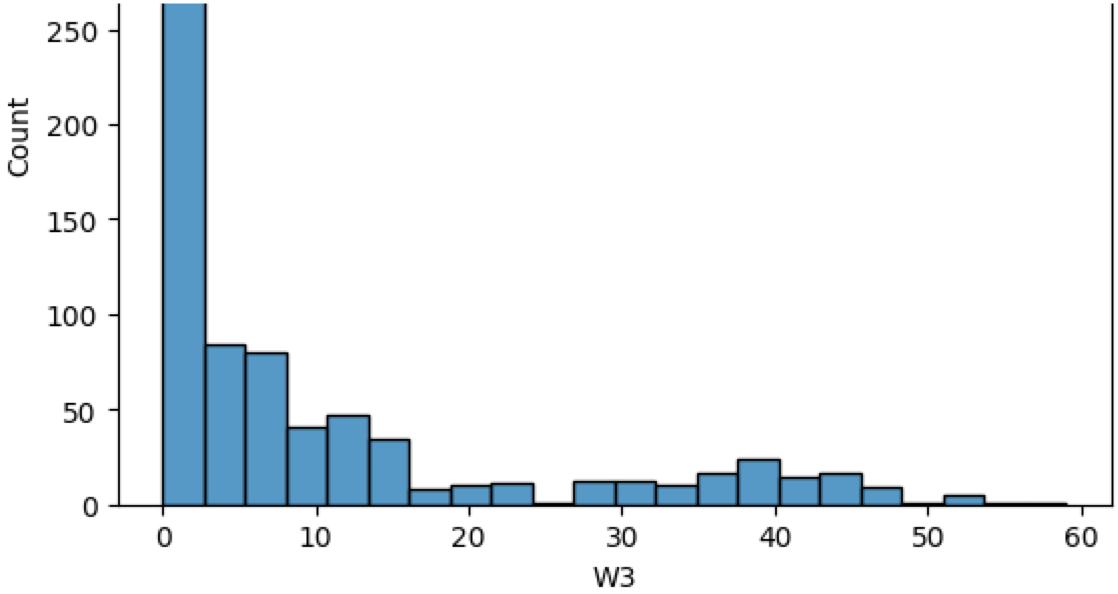

Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50
W0

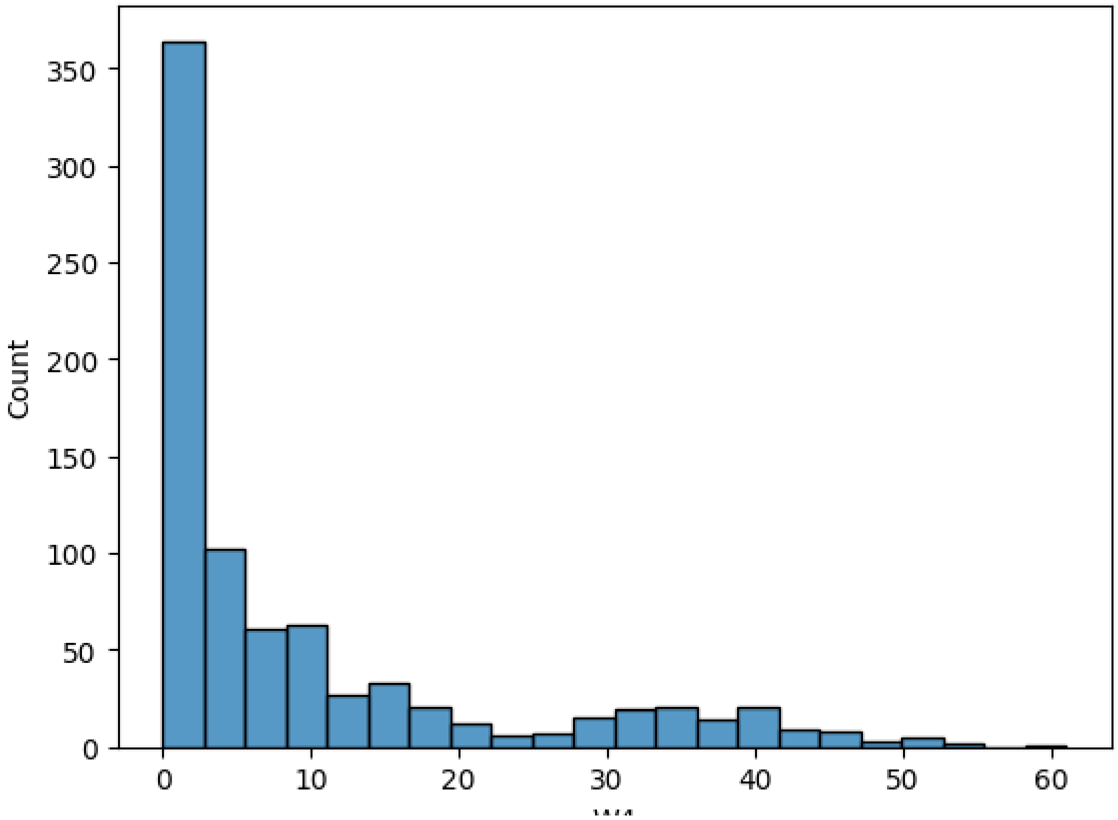
Count
150
200
250
300
350

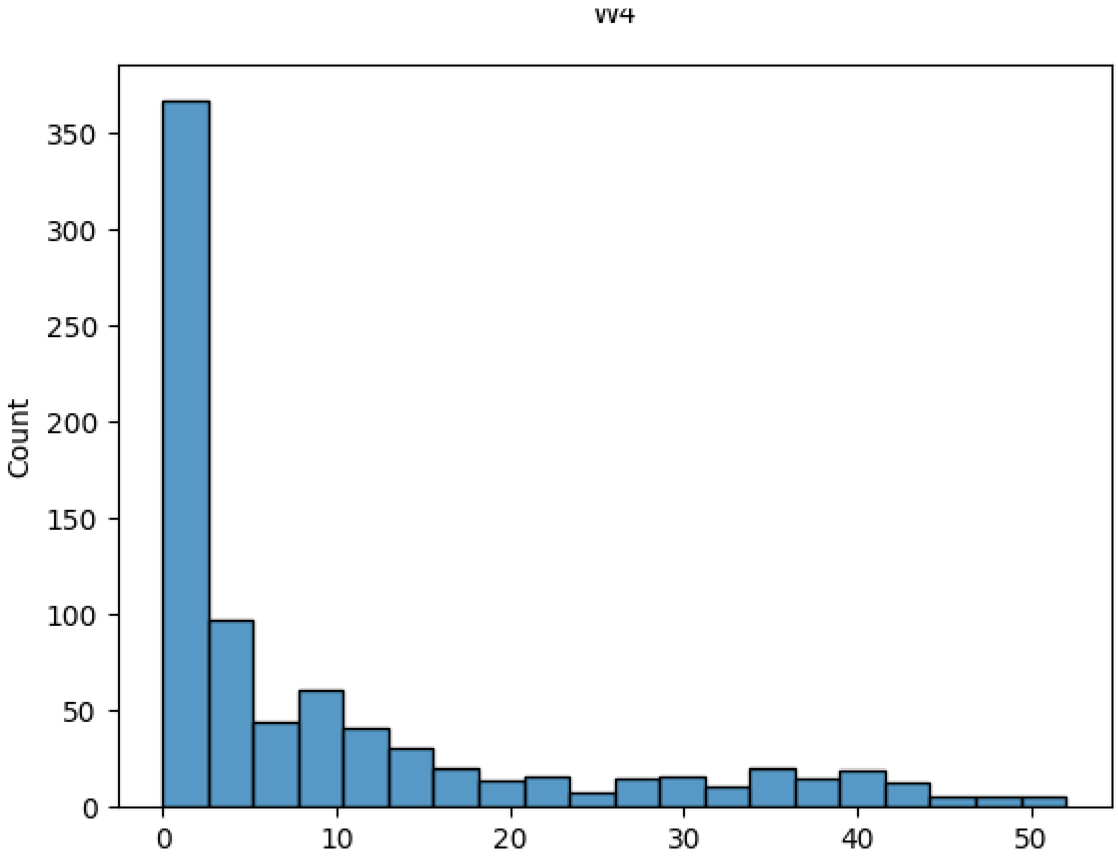

100
50
0
0
10
20
30
40
50
W1

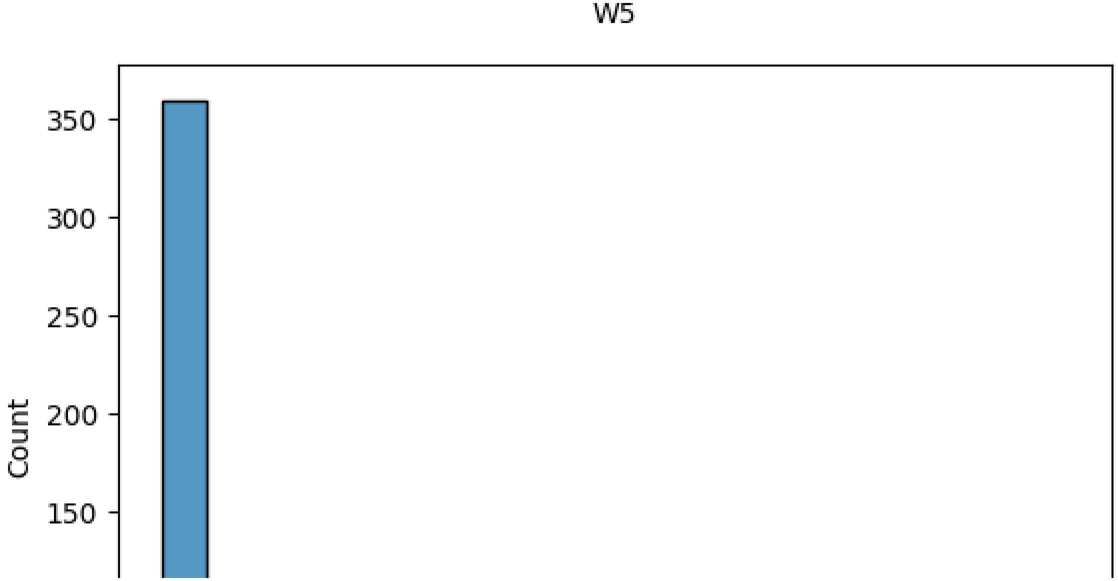

350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
W2

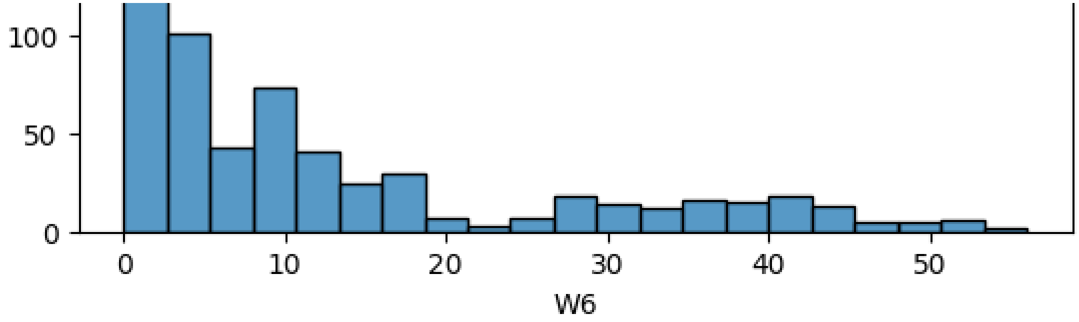
350
300

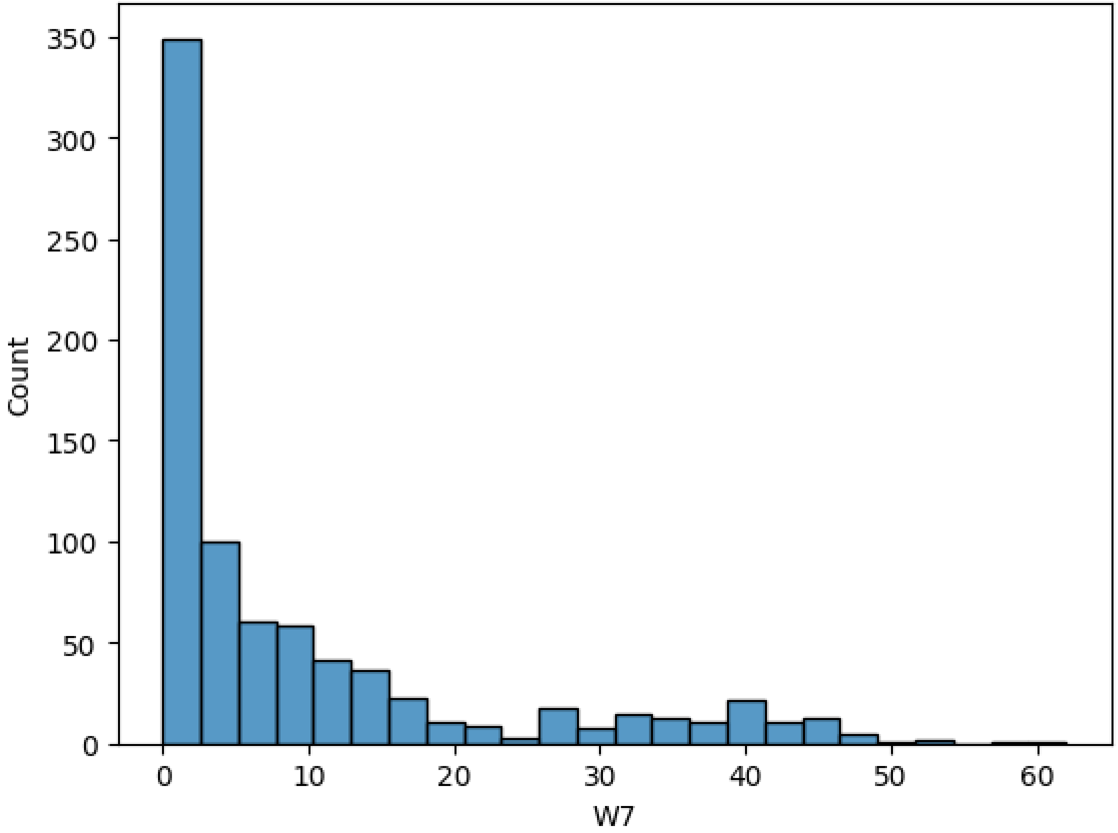

Count
0
50
100
150
200
250
0
10
20
30
40
50
60
W3


Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50
60
W4

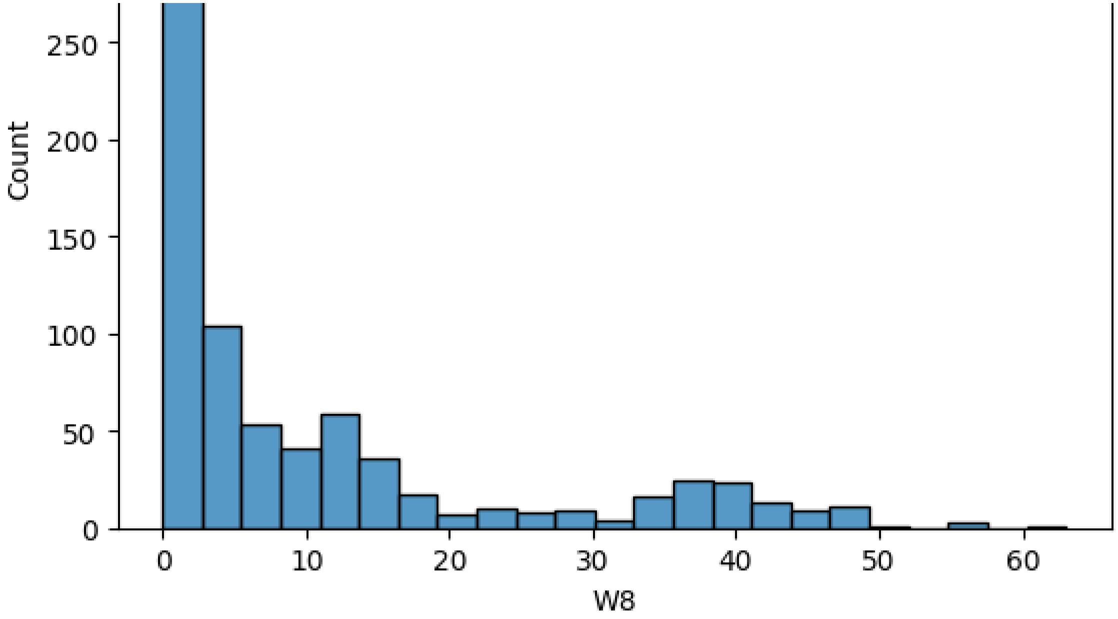

W4
Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50

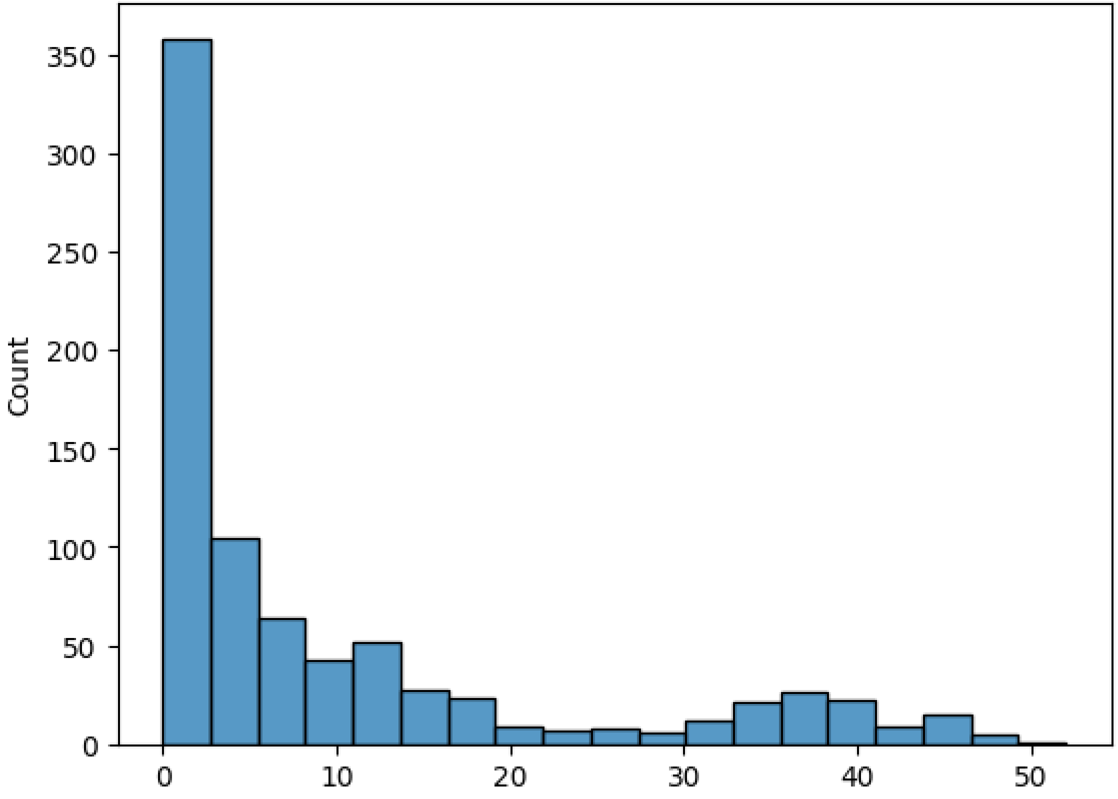

W5
Count
150
200
250
300
350


100
50
0
0
10
20
30
40
50
W6

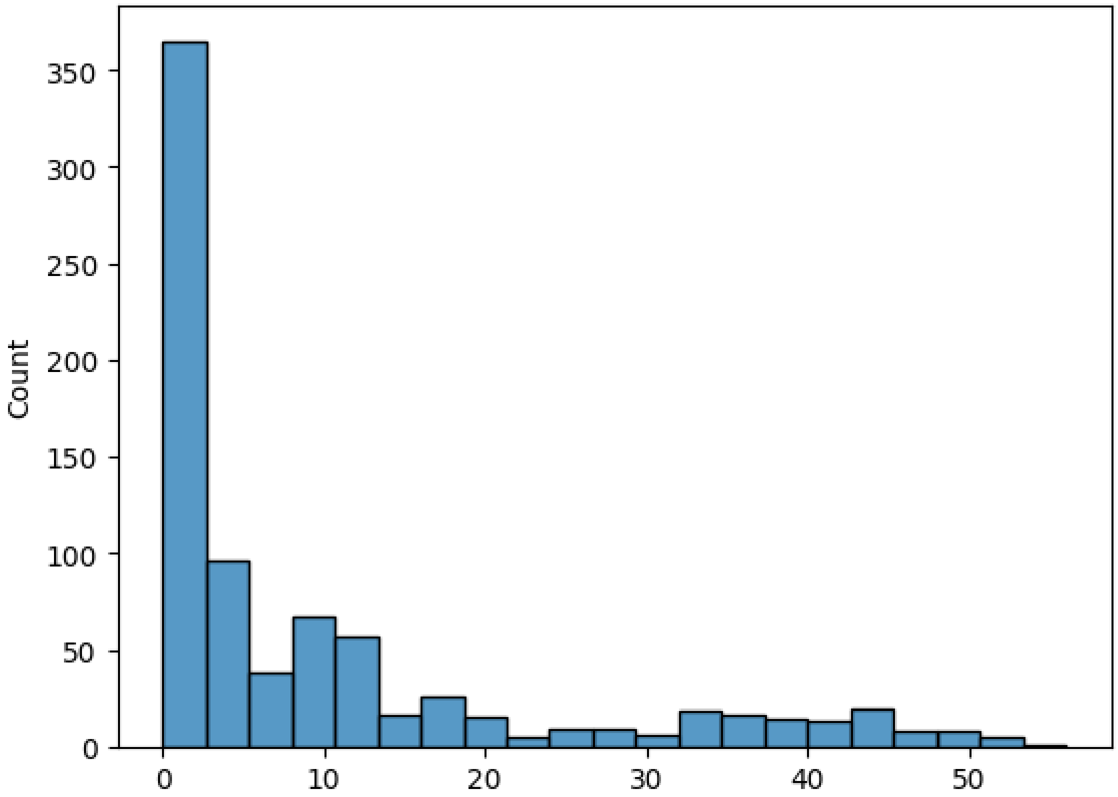

350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
60
W7

350
300


Count
250
200
150
100
50
0
0
10
20
30
40
50
60
W8

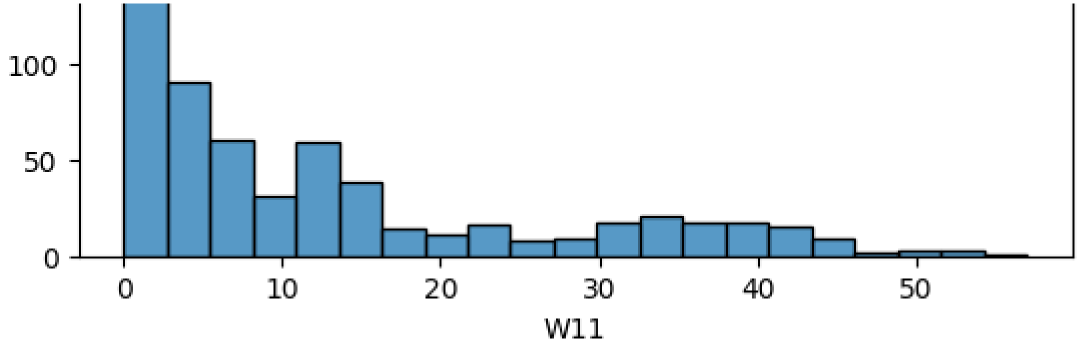

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50

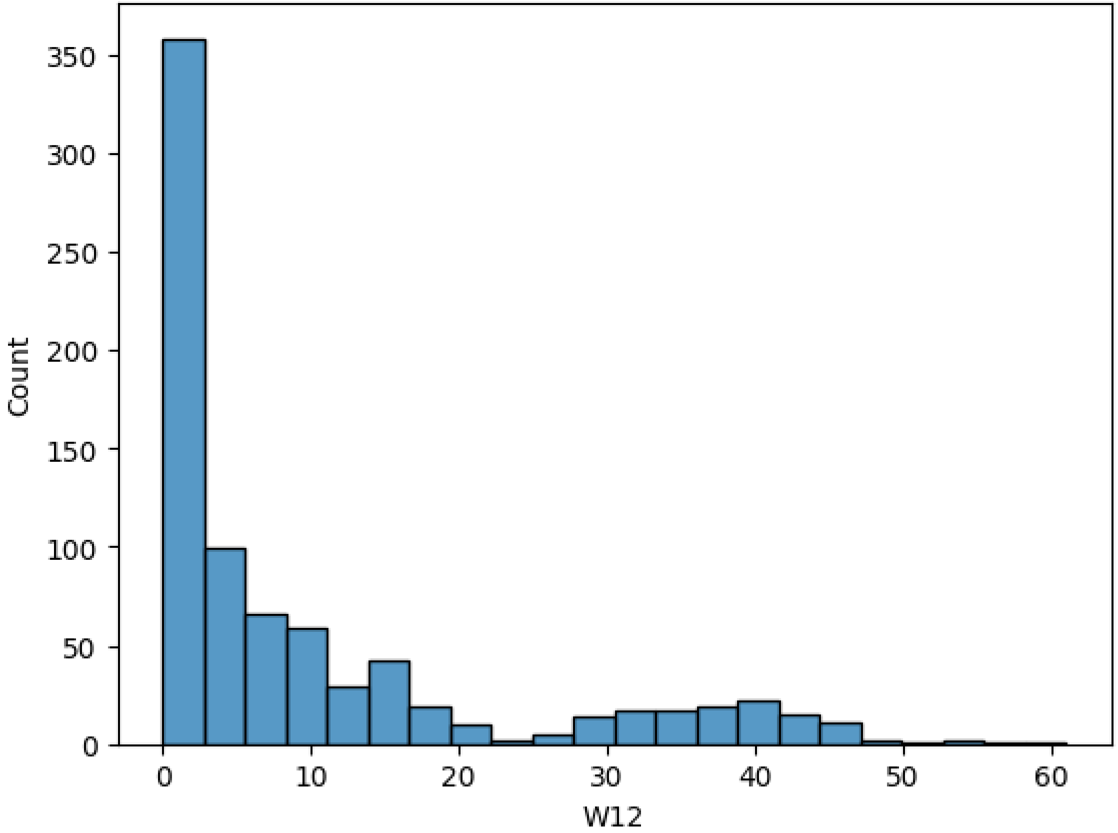
W9

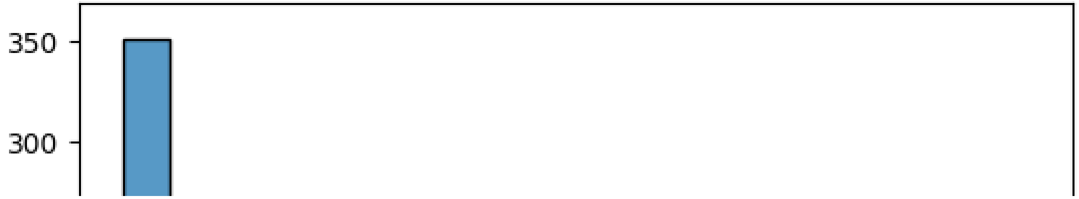

Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50

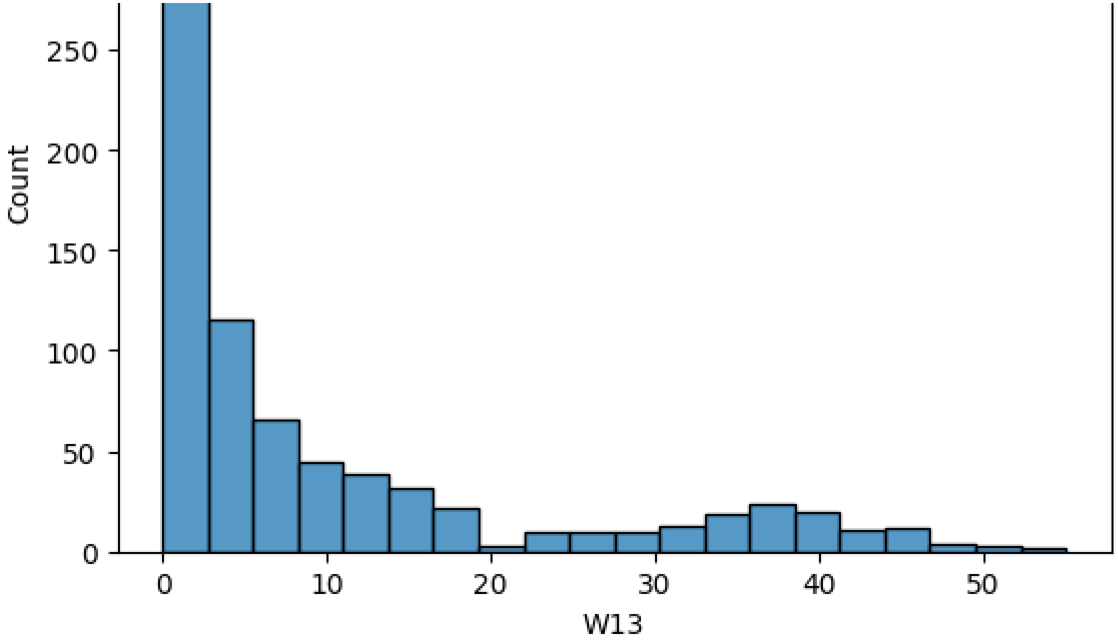
W10

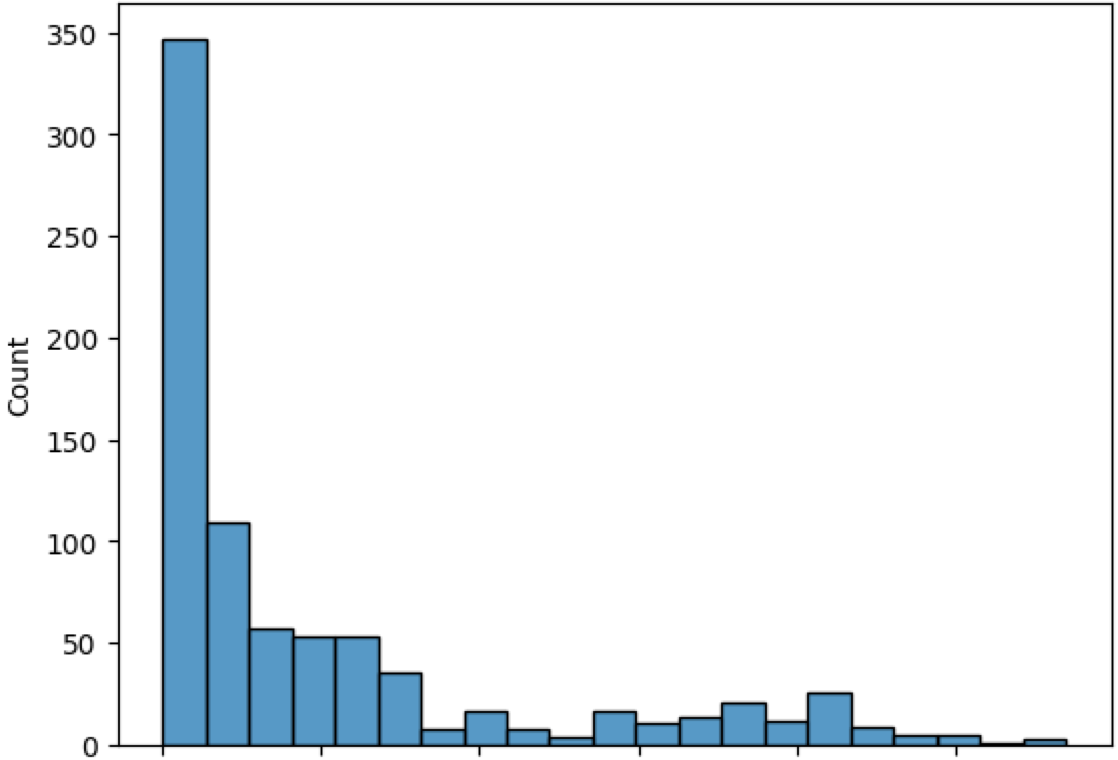
Count
150
200
250
300
350


100
50
0
0
10
20
30
40
50
W11

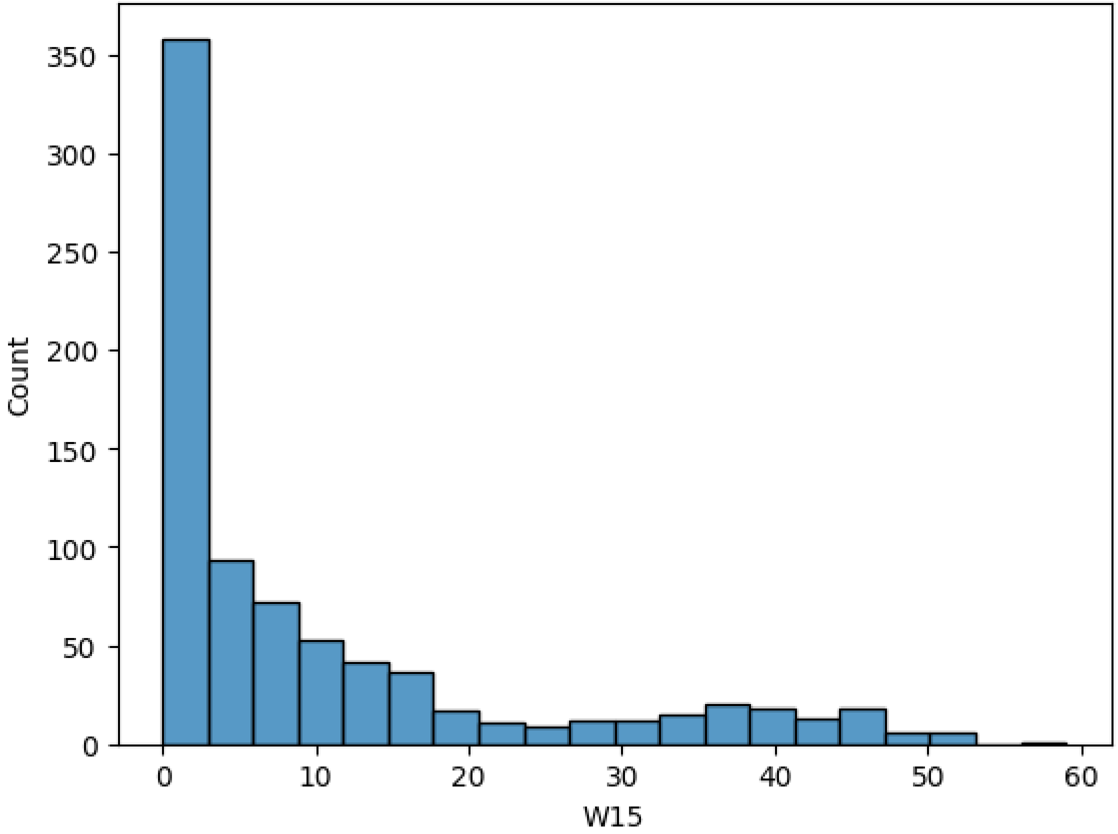

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
60
W12


350
300

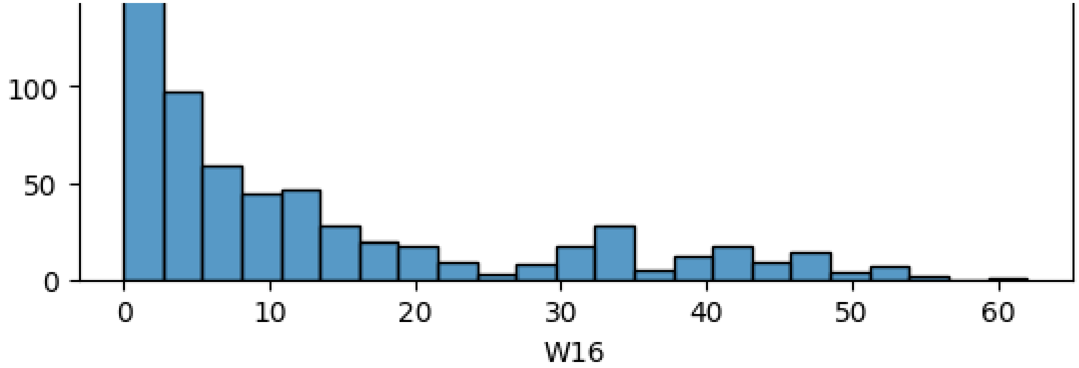

Count
0
50
100
150
200
250
0
10
20
30
40
50
W13

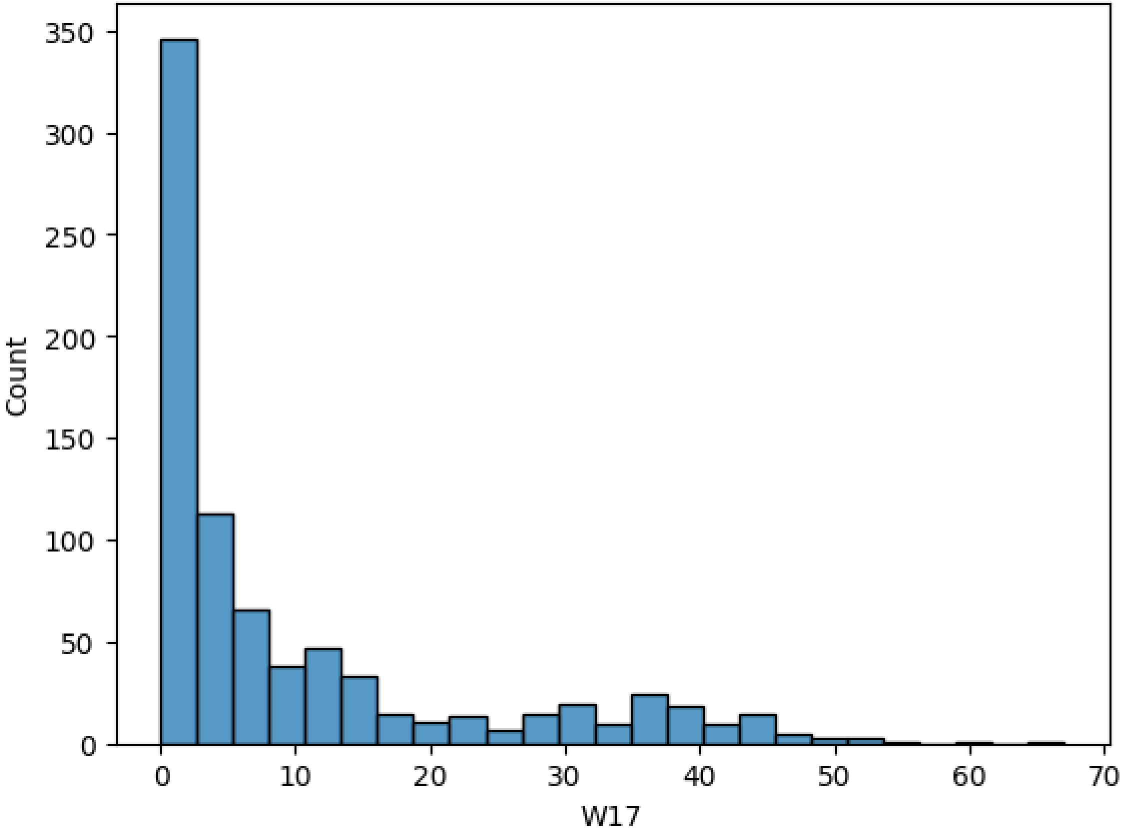

Count
0
50
100
150
200
250
300
350

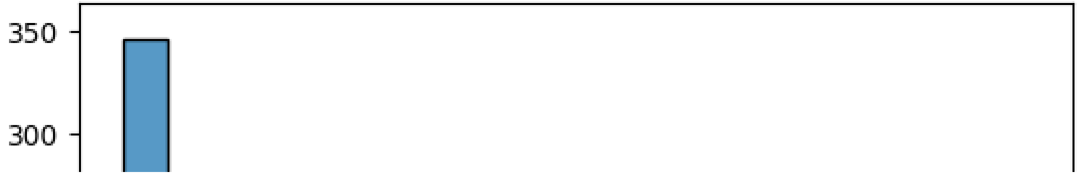
0
10
20
30
40
50
W14

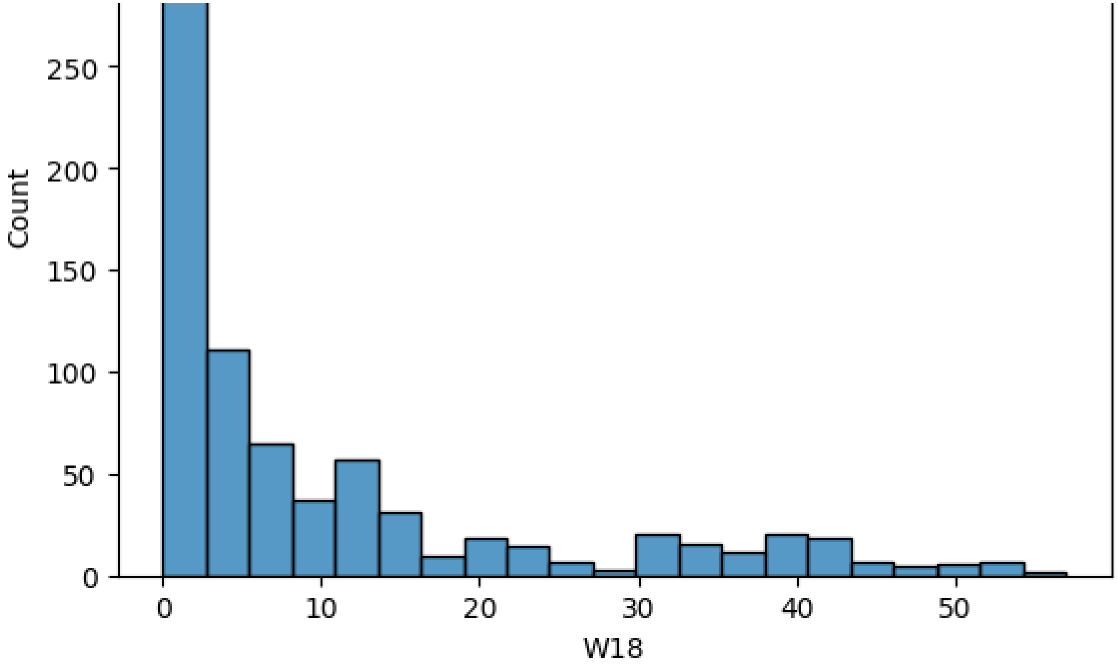

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
60
W15

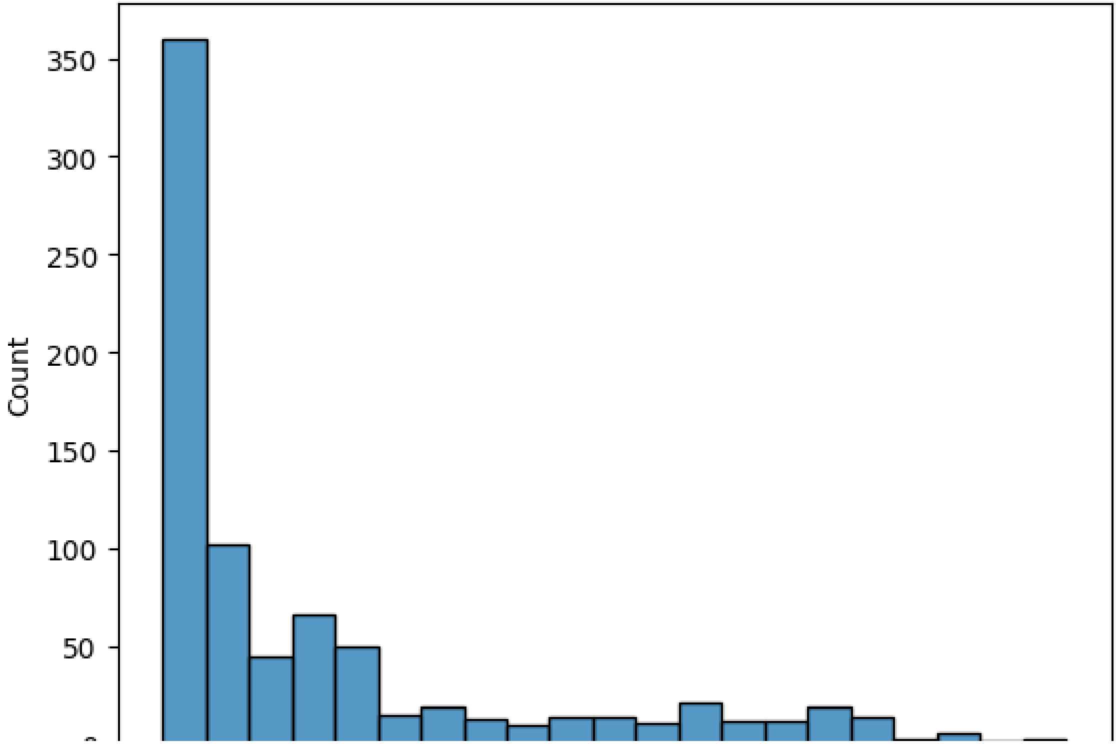
Count
350
300
250
200
150

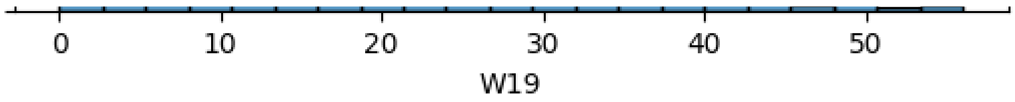

100
50
0
0
10
20
30
40
50
60
W16

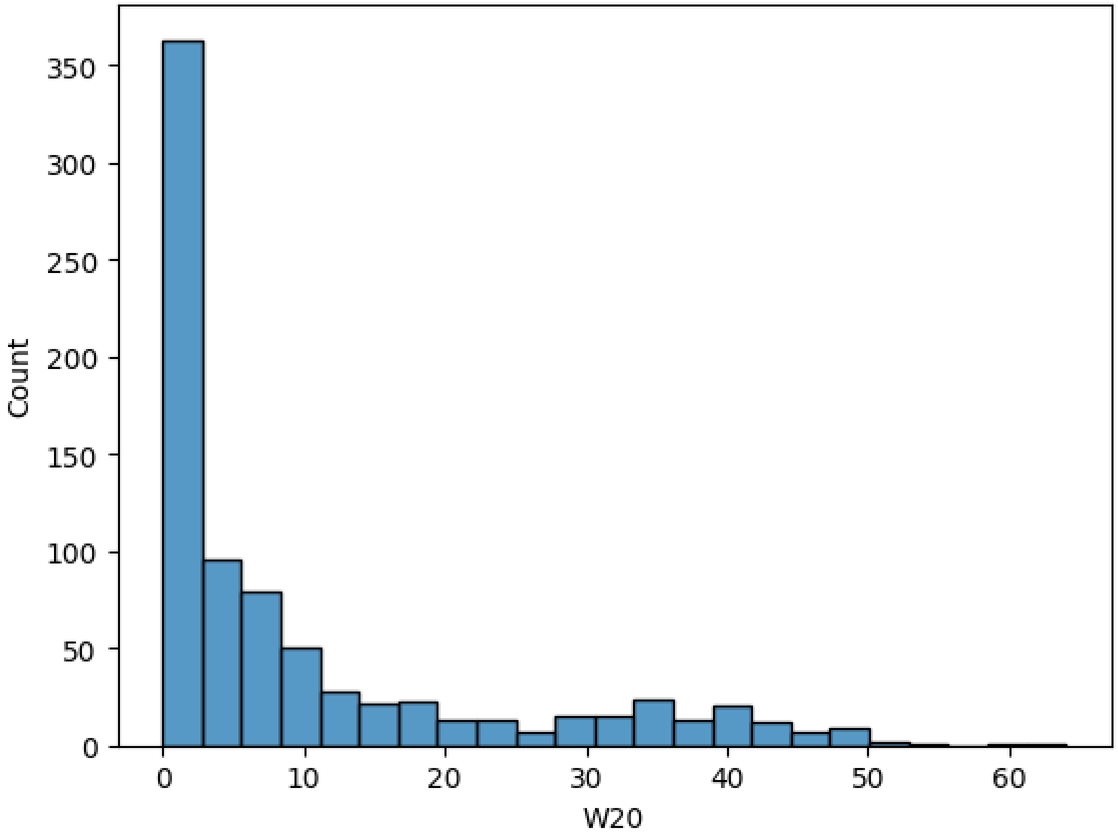

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
60
70
W17

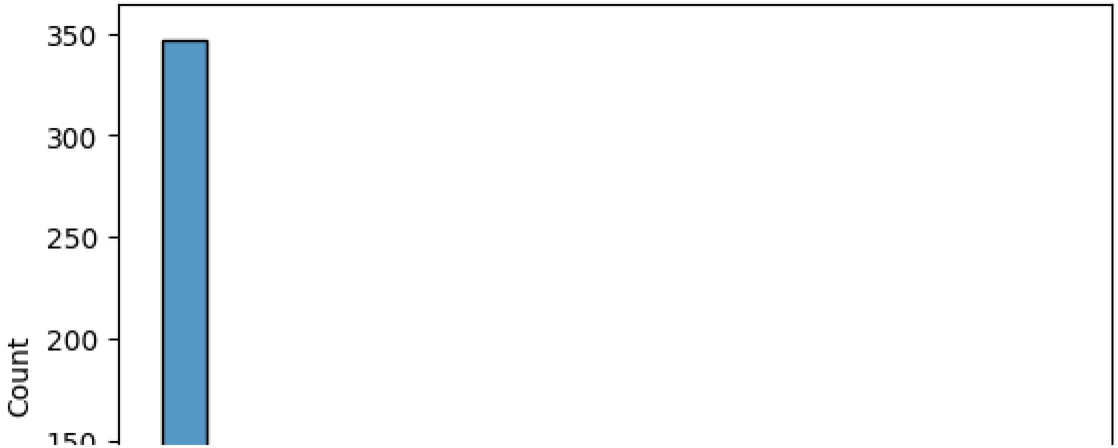

350
300

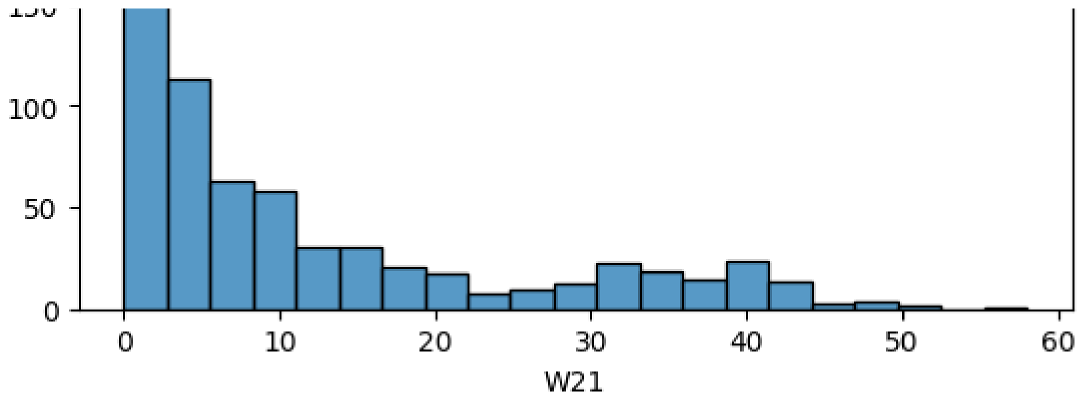

Count
250
200
150
100
50
0
0
10
20
30
40
50
W18

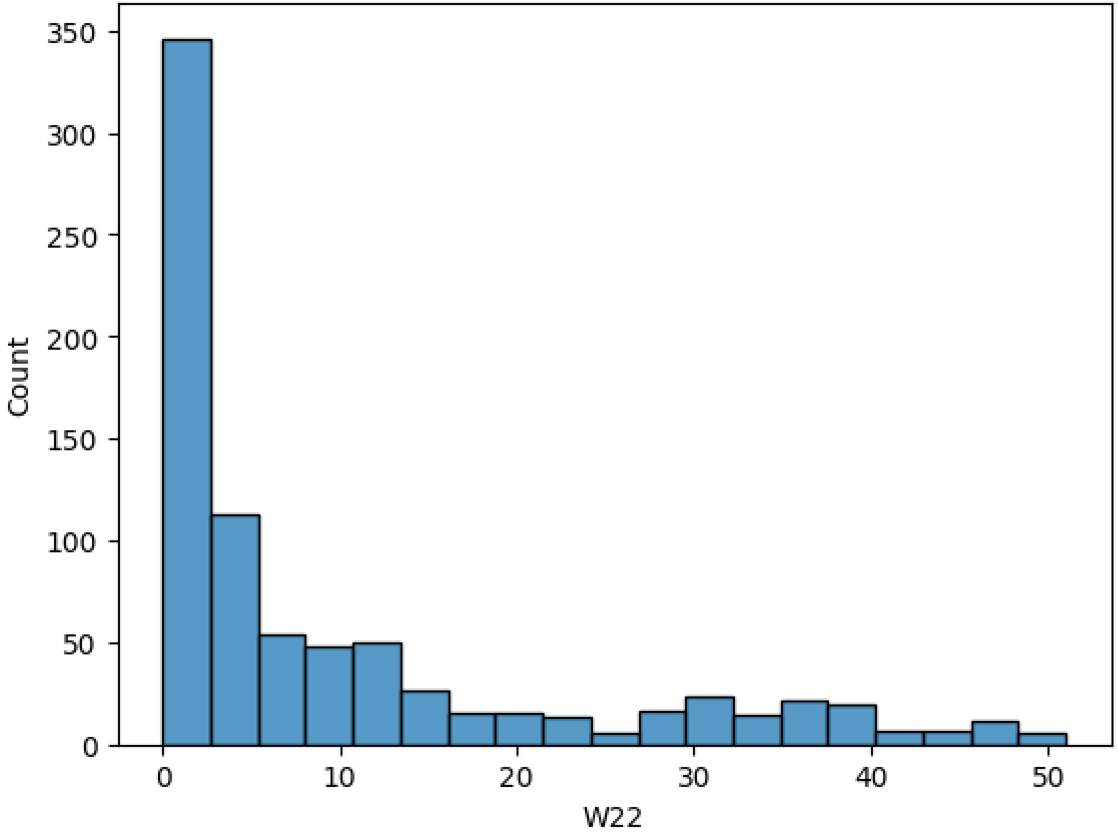

Count
350
300
250
200
150
100
50

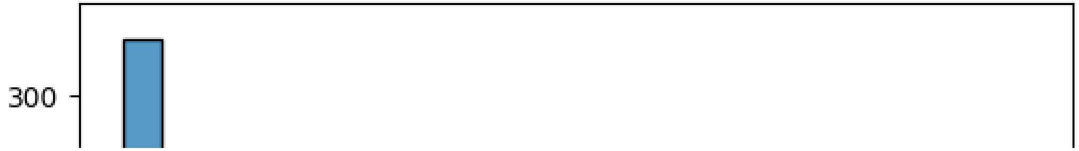

0
0
10
20
30
40
50
W19

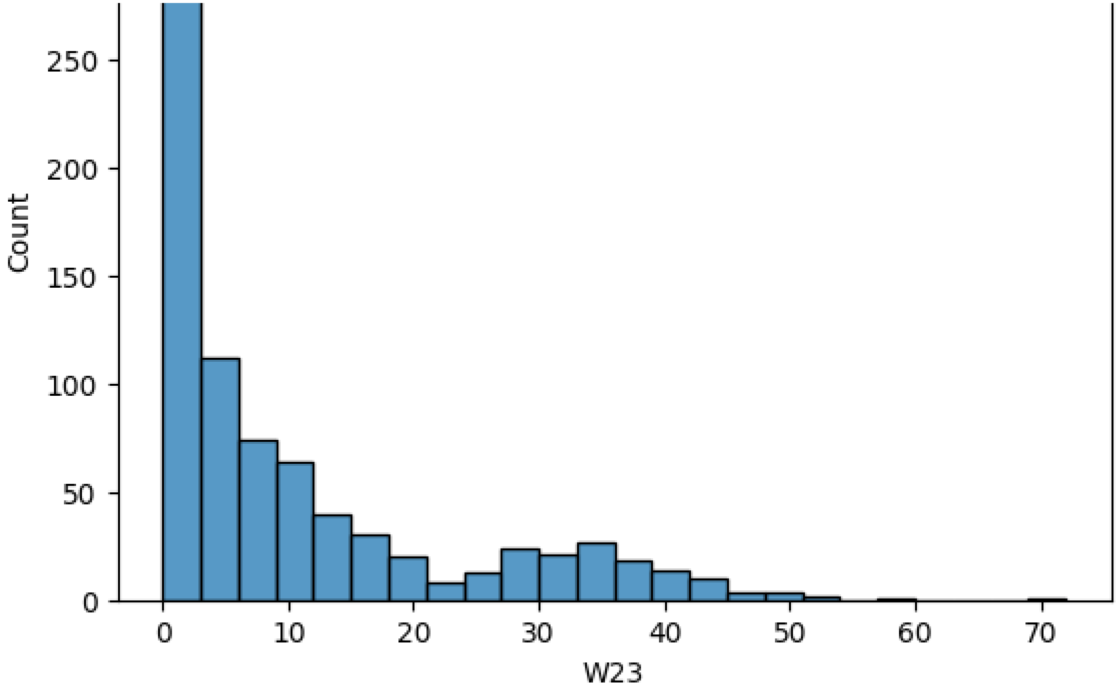

350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
60
W20

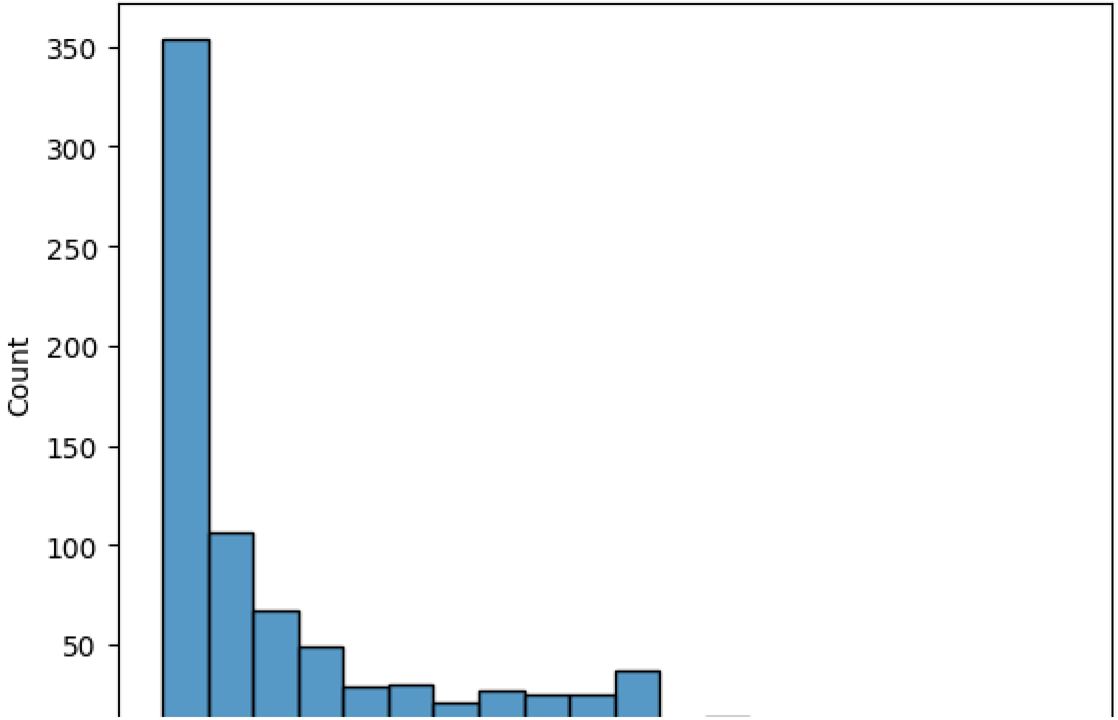

350
300
250
200
150
Count

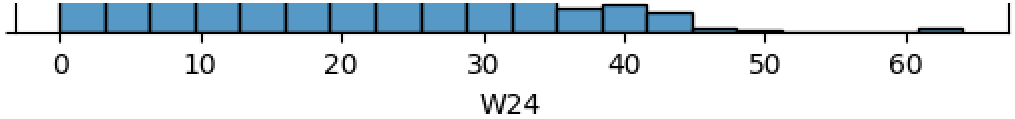

150
100
50
0
0
10
20
30
40
50
60
W21

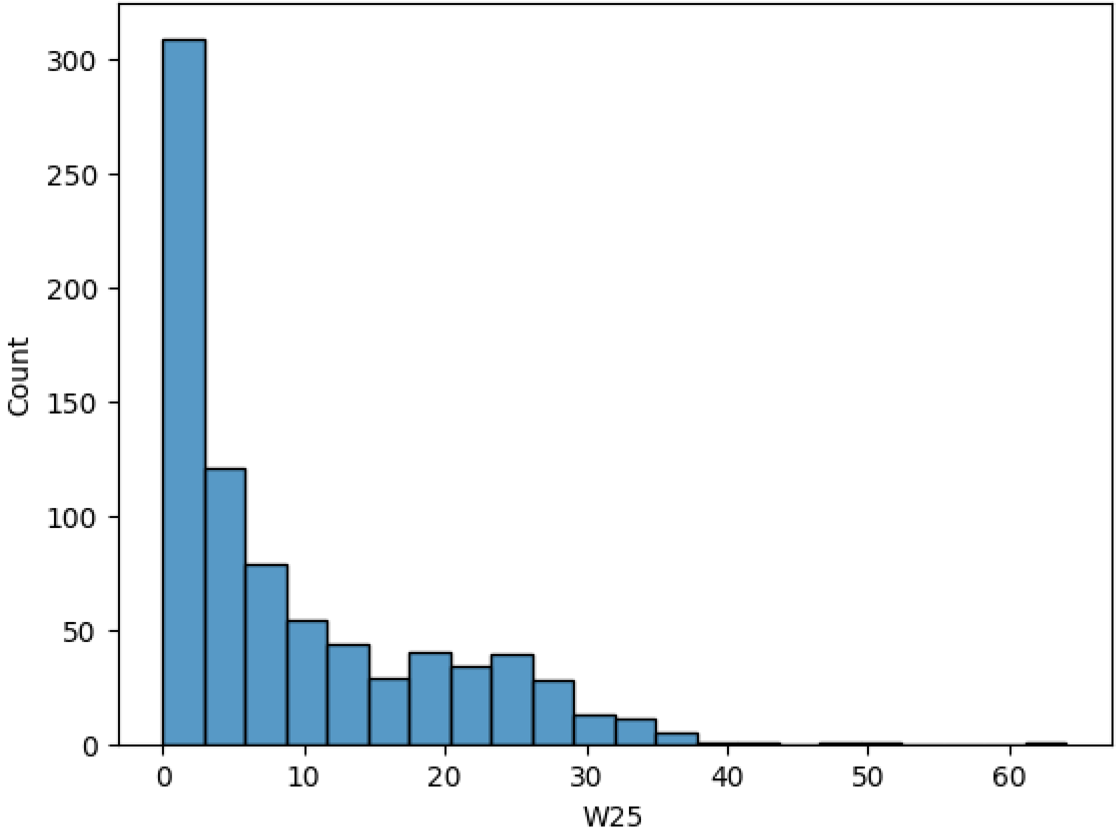

350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
W22


300

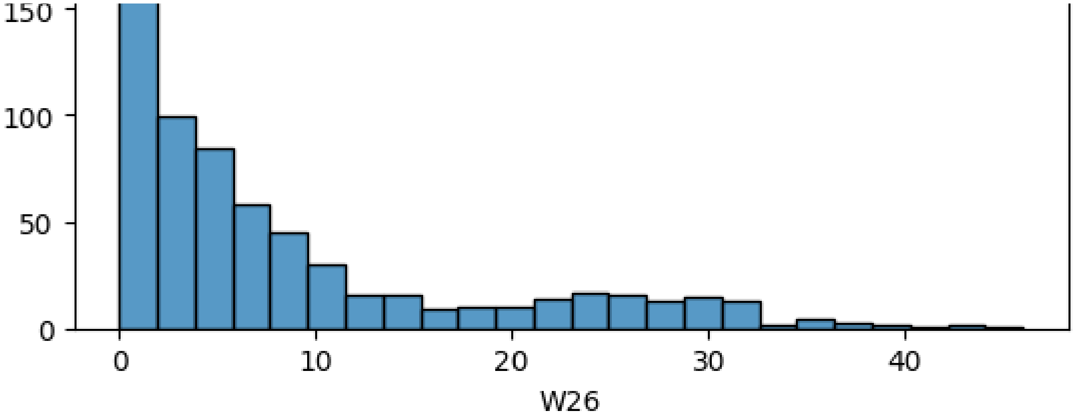

Count
250
200
150
100
50
0
0
10
20
30
40
50
60
70
W23

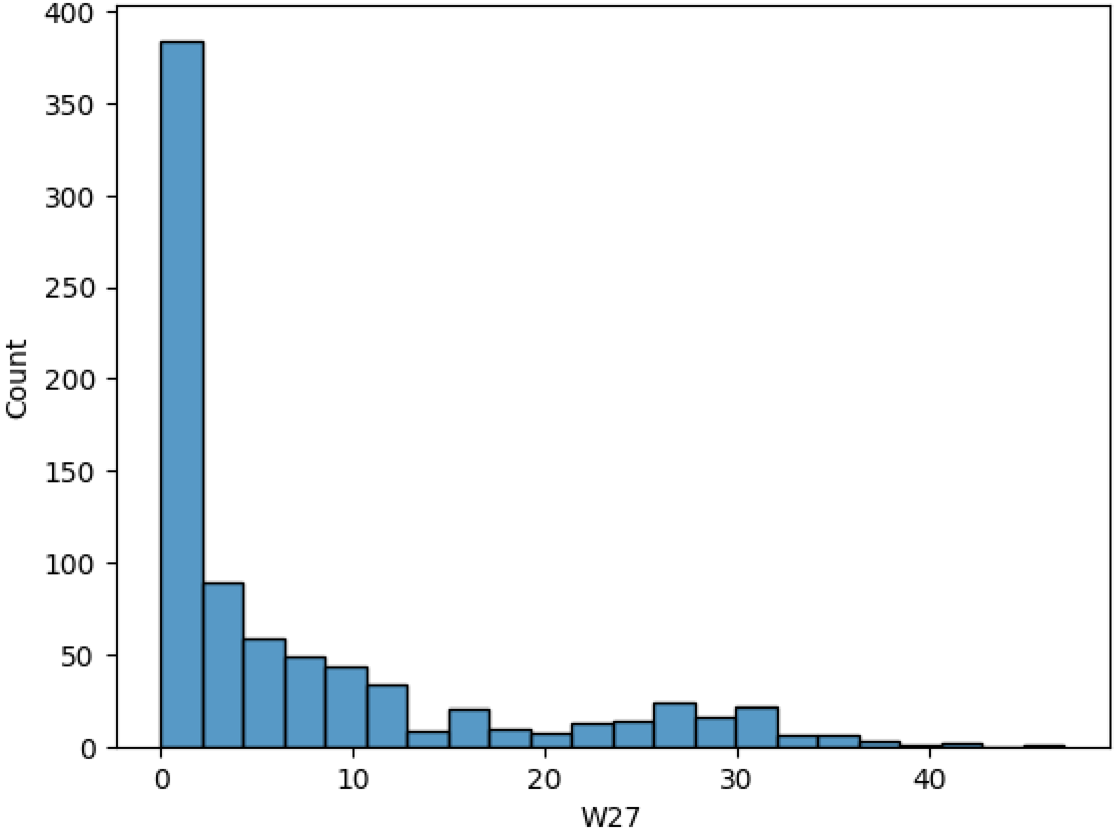

Count
350
300
250
200
150
100
50


0
0
10
20
30
40
50
60
W24

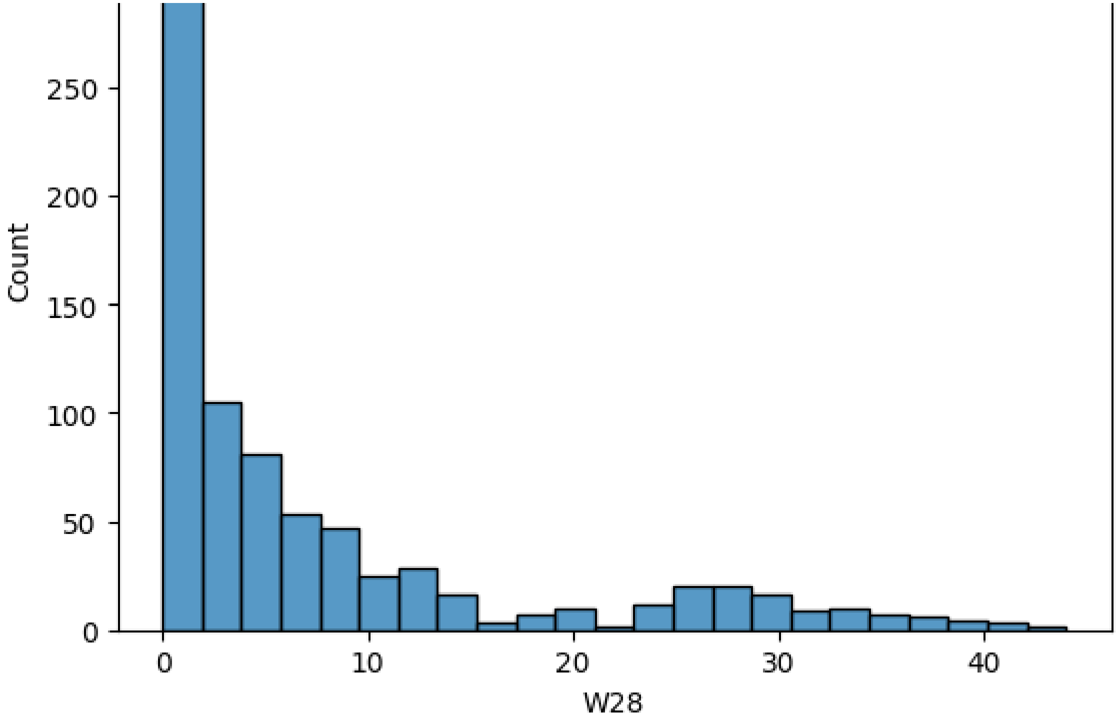

300
250
200
150
100
50
0
Count
0
10
20
30
40
50
60
W25

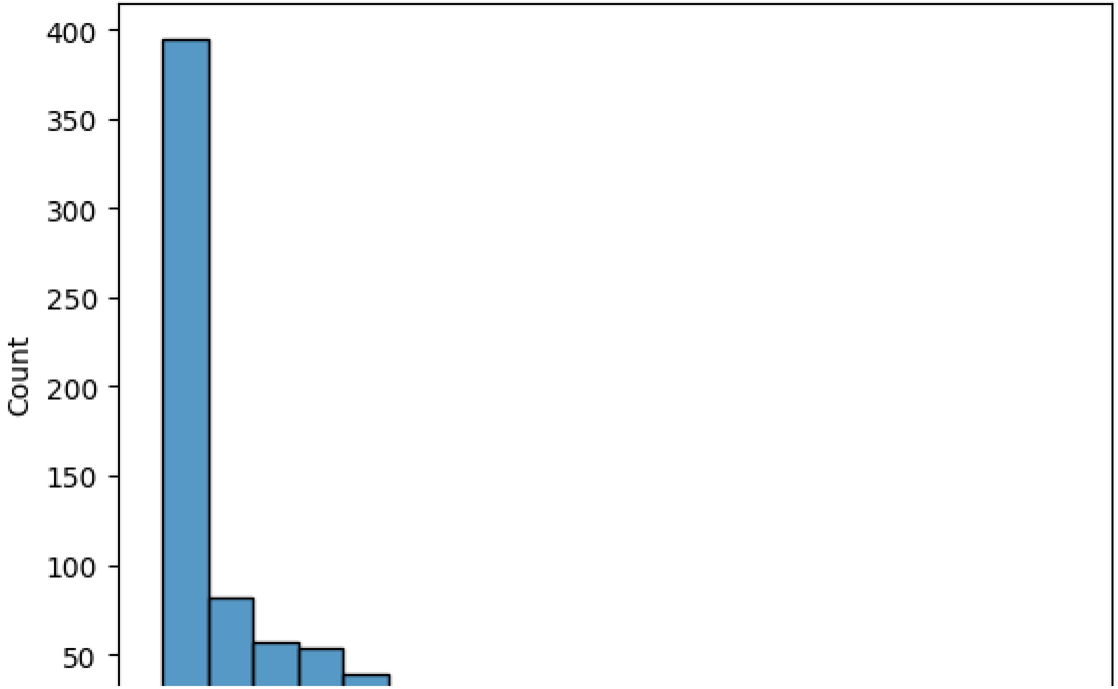
300
250
200
Count

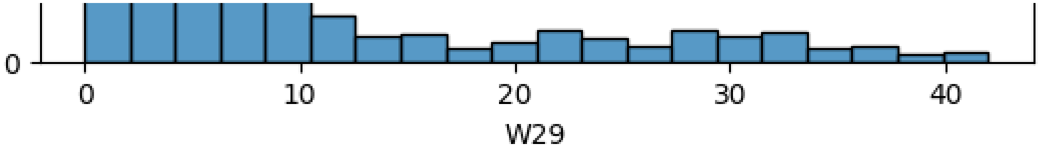

150
100
50
0
0
10
20
30
40
W26

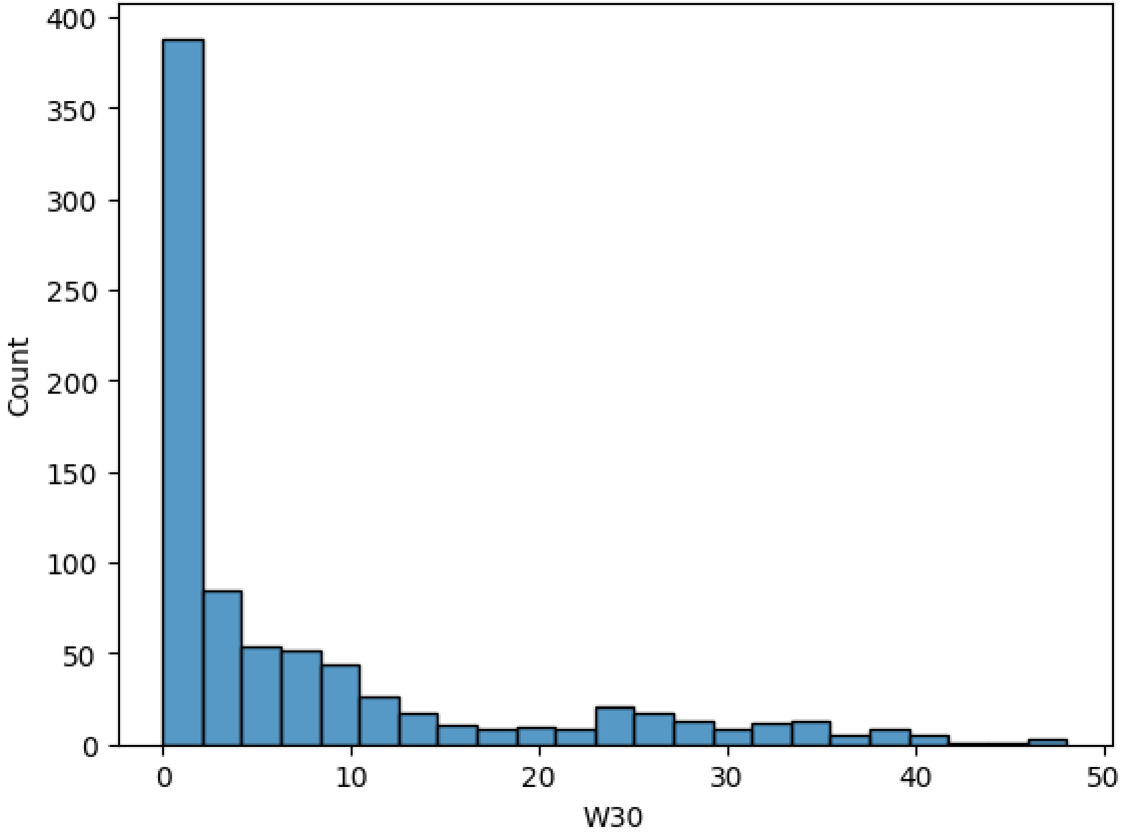

400
350
300
250
200
150
100
50
0
Count
0
10
20
30
40
W27

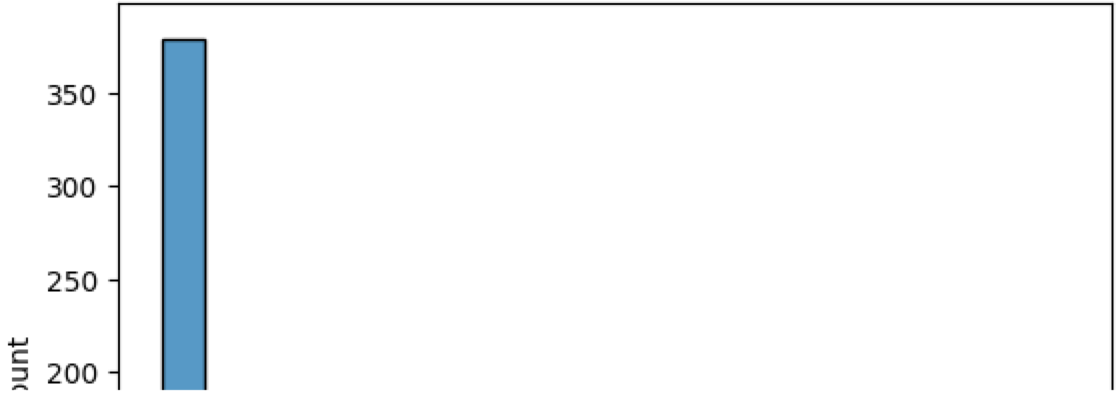
300

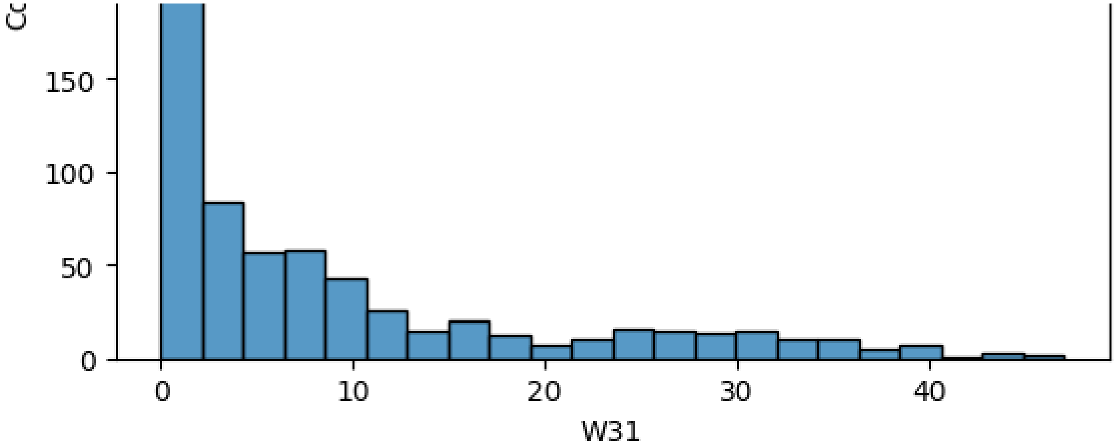

Count
250
200
150
100
50
0
0
10
20
30
40
W28

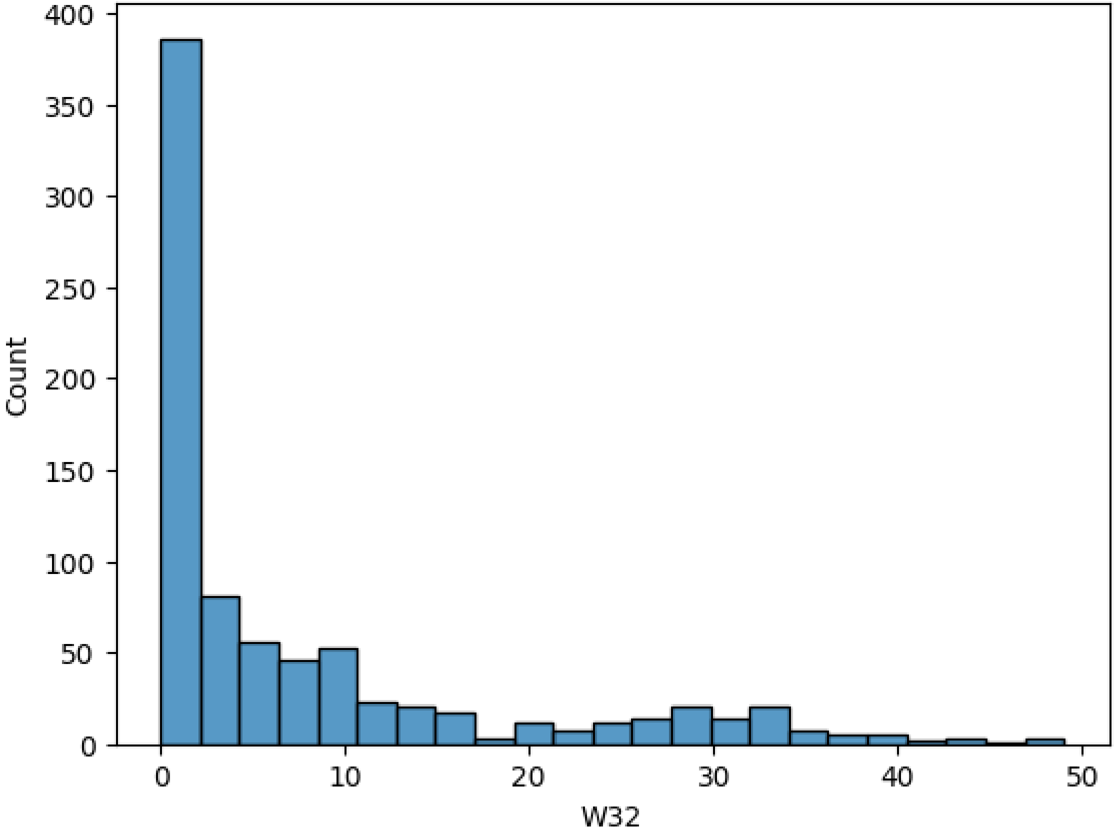

Count
400
350
300
250
200
150
100
50


0
0
10
20
30
40
W29

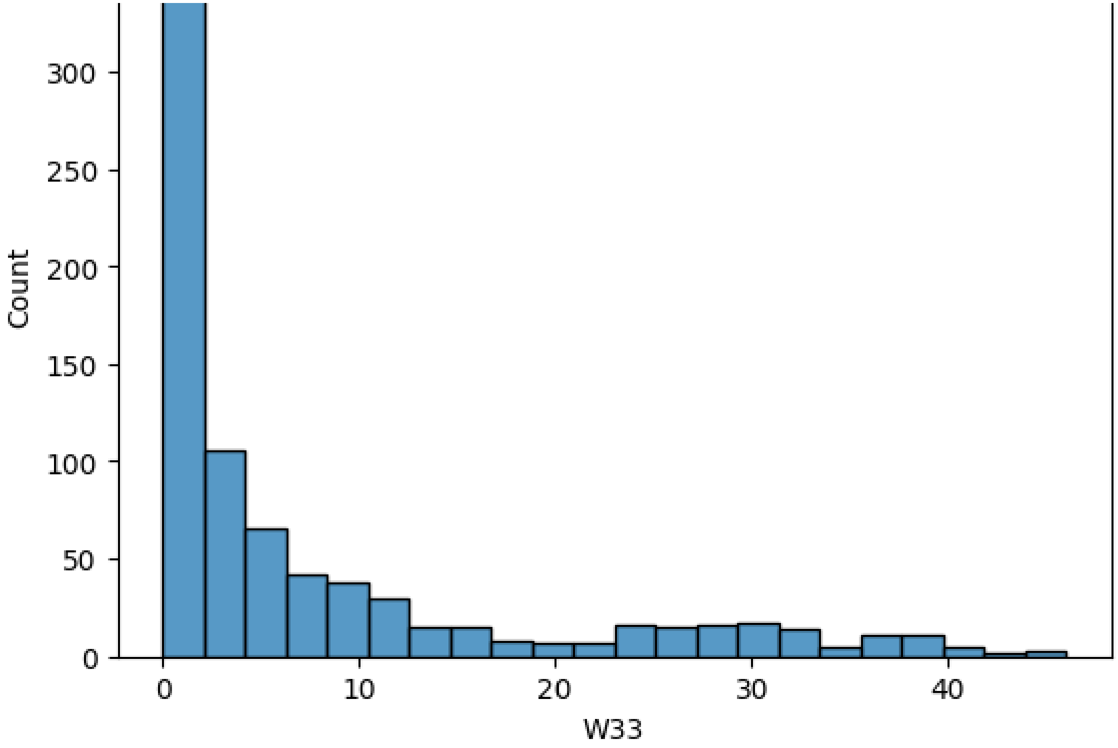

400
350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
W30

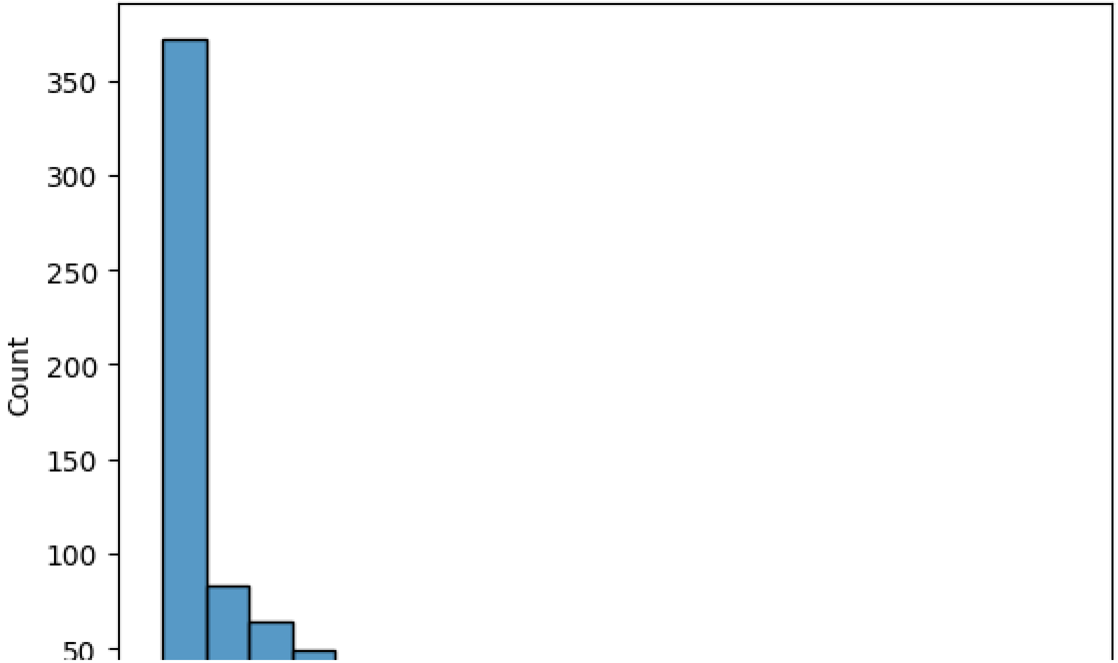

350
300
250
200
ount

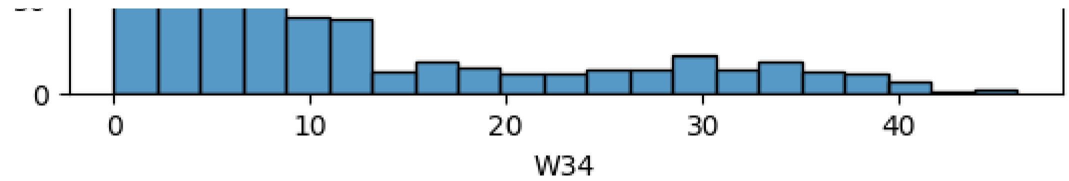

Co
150
100
50
0
0
10
20
30
40
W31

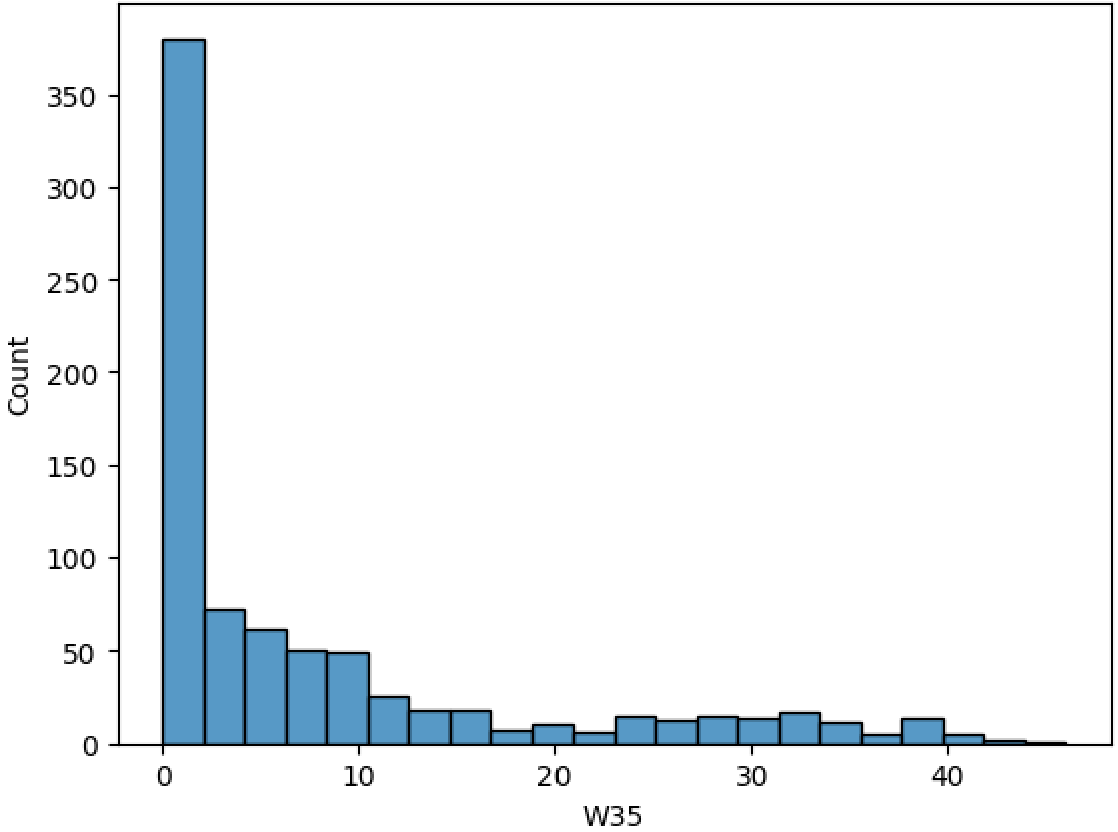

400
350
300
250
200
150
100
50
0
Count
0
10
20
30
40
50
W32

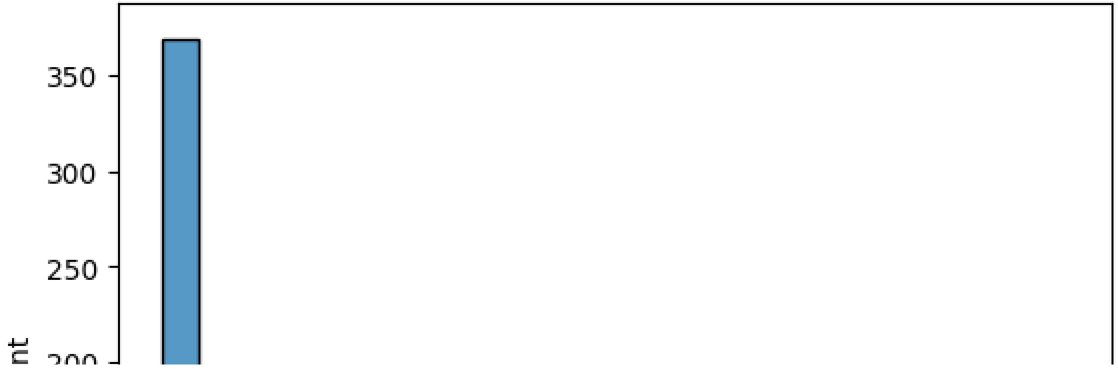
350

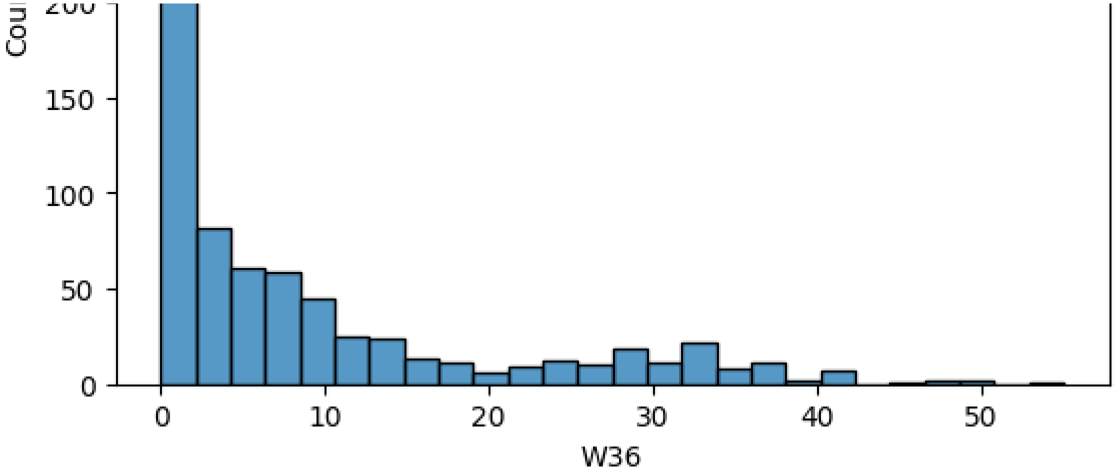

Count
300
250
200
150
100
50
0
0
10
20
30
40
W33

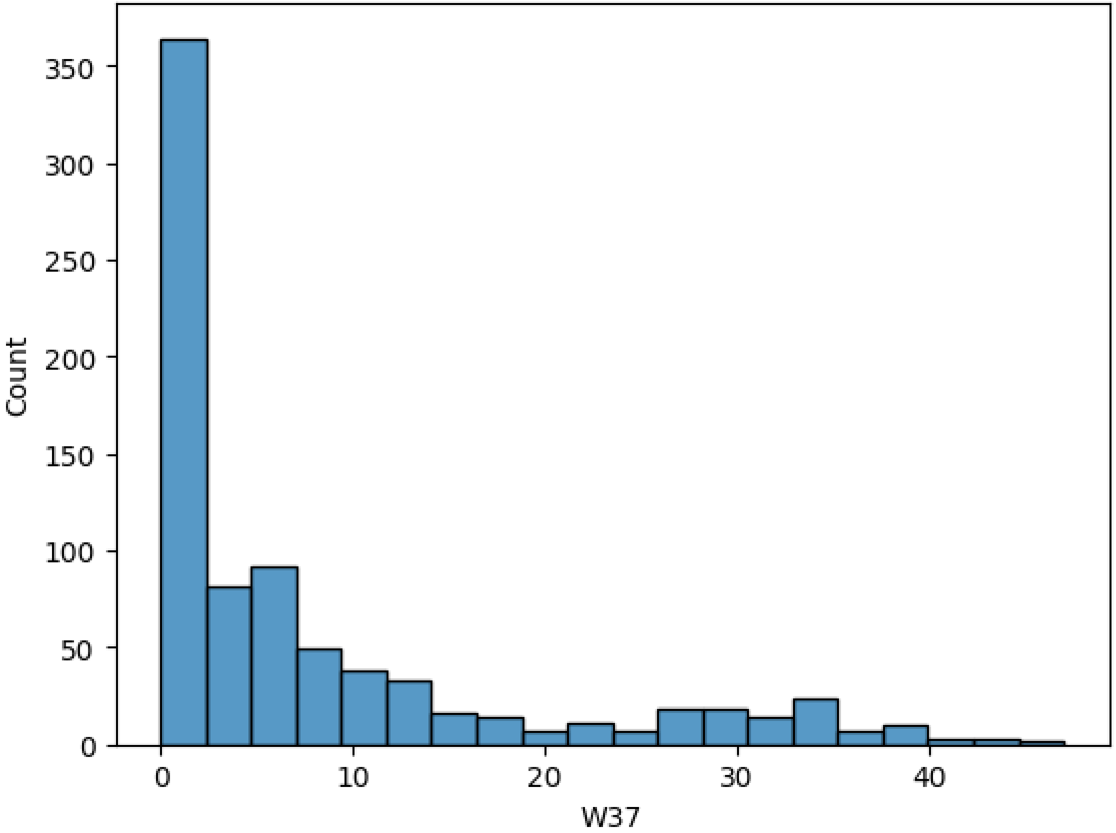

Count
350
300
250
200
150
100
50

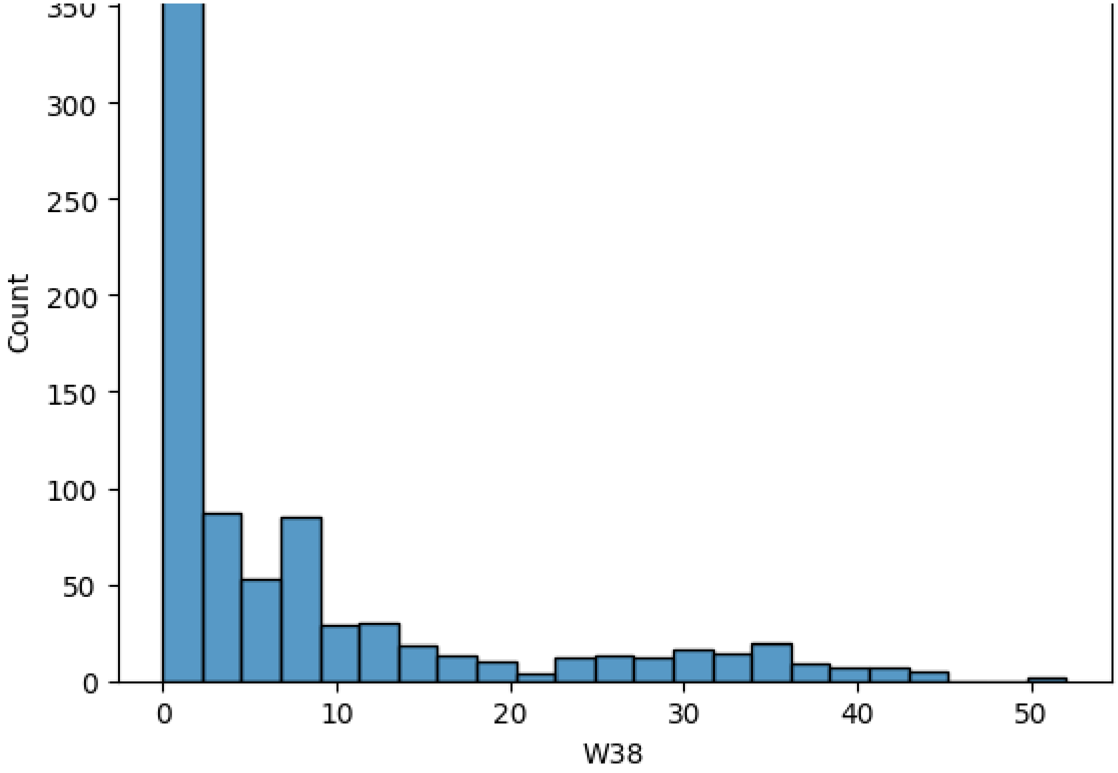

0
0
10
20
30
40
W34

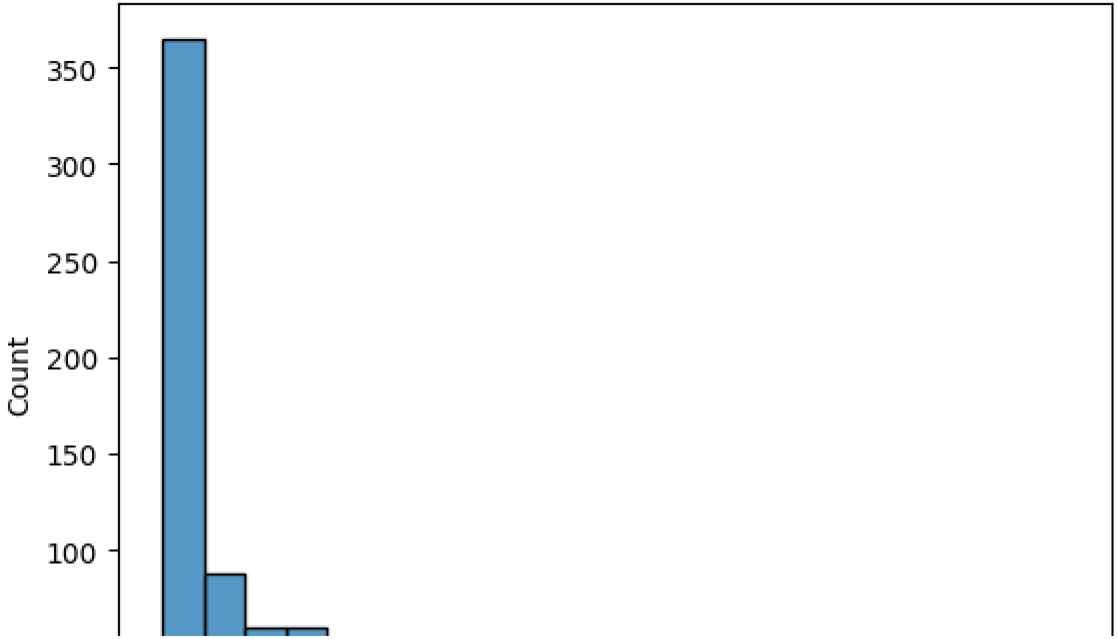

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
W35

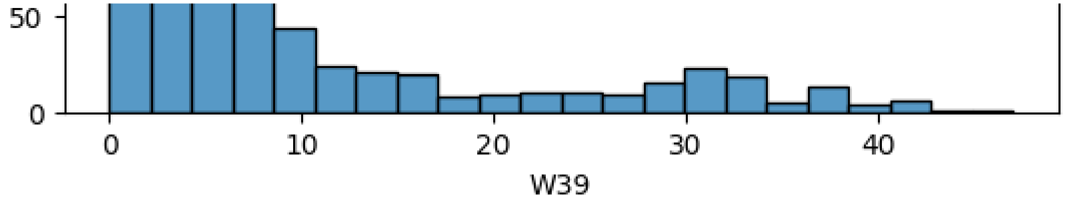

350
300
250
200

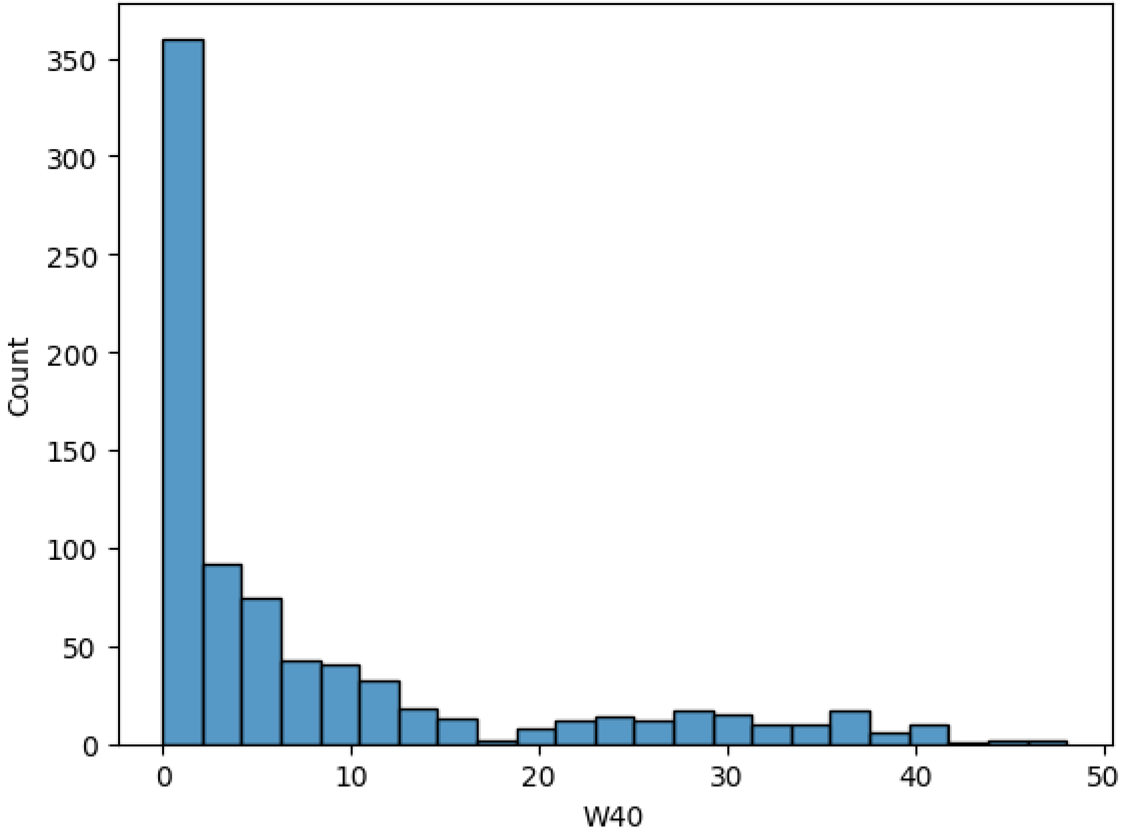

Cou
200
150
100
50
0
0
10
20
30
40
50
W36


Count
350
300
250
200
150
100
50
0
0
10
20
30
40
W37

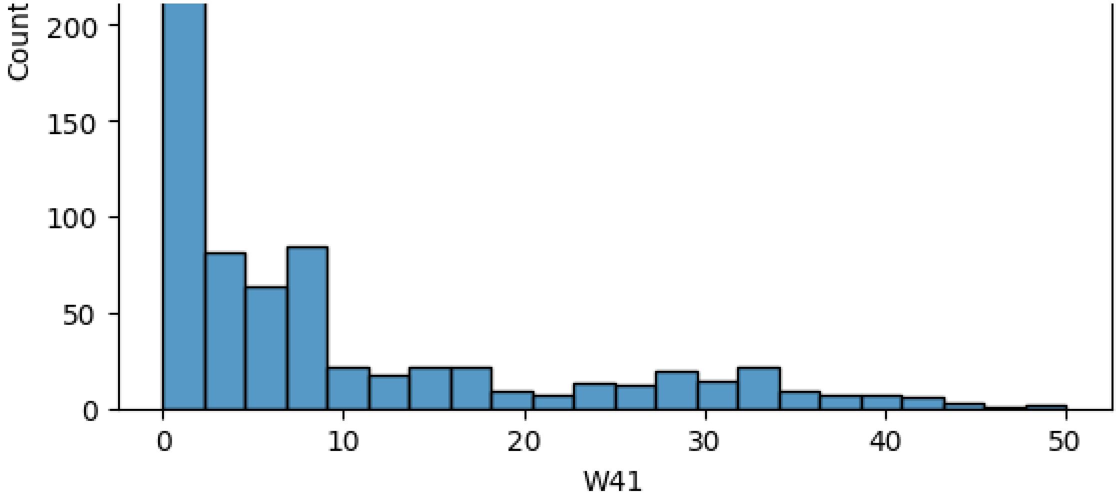

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
W38

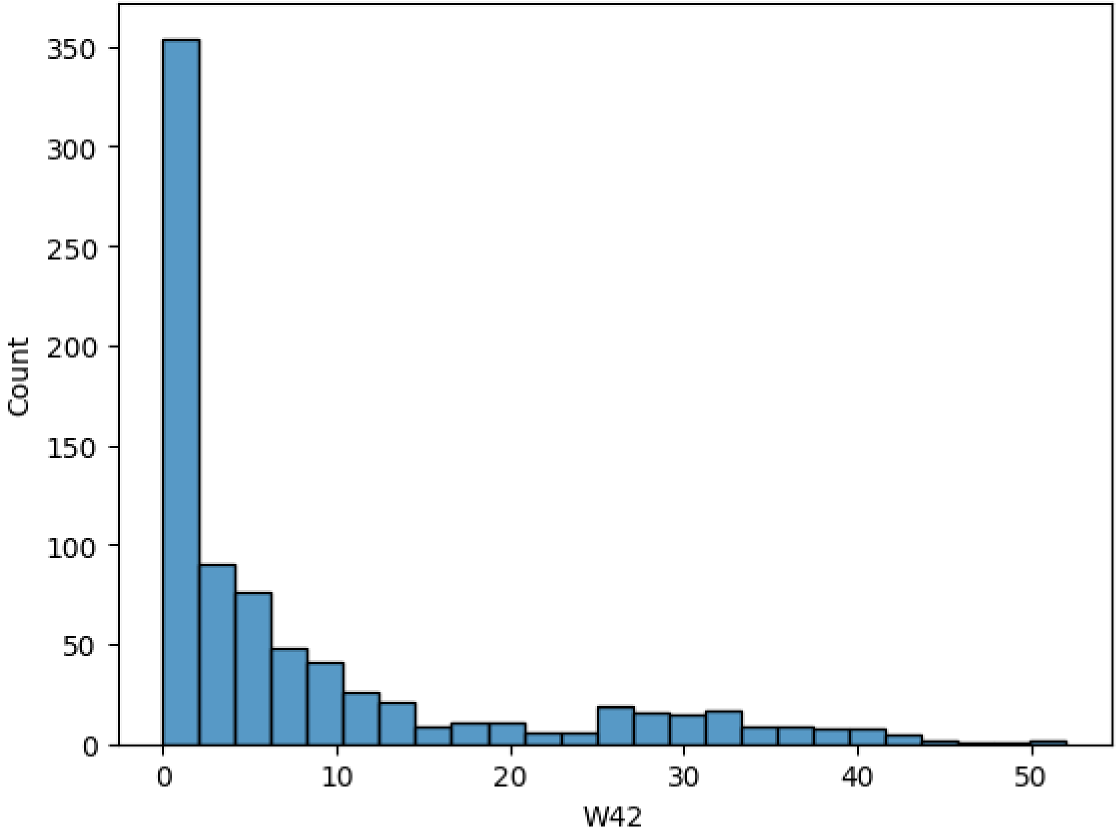

Count
350
300
250
200
150
100

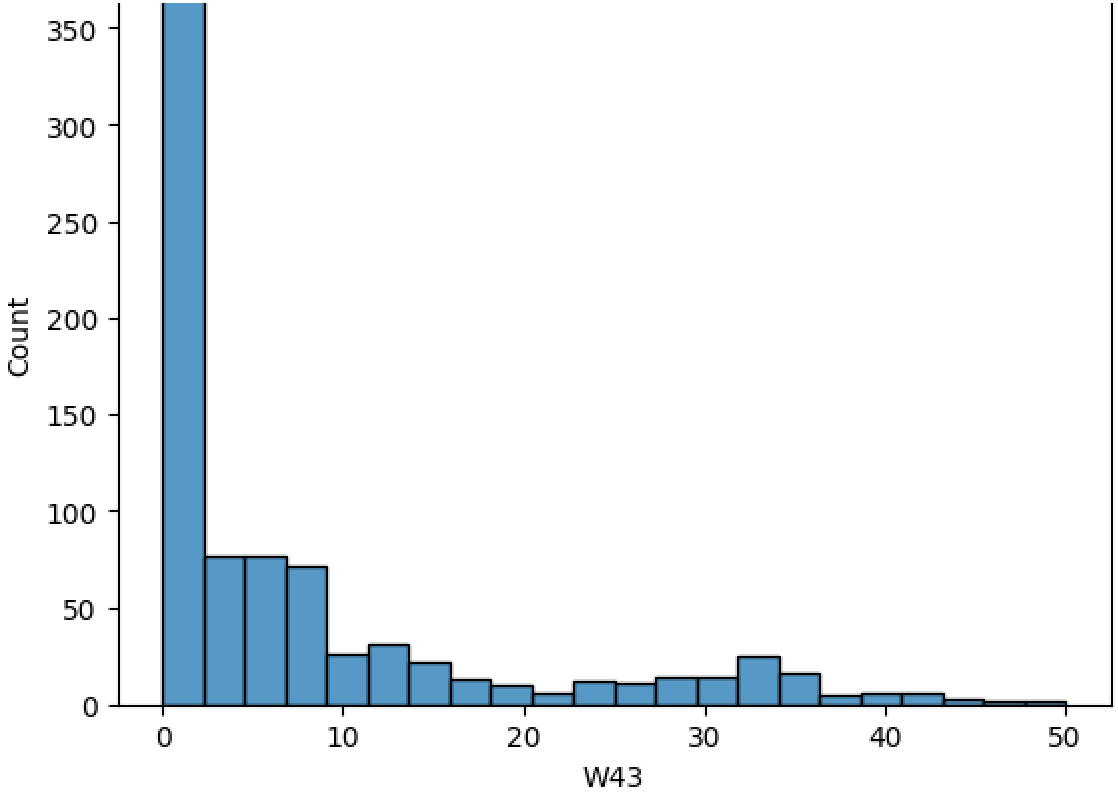

50
0
0
10
20
30
40
W39

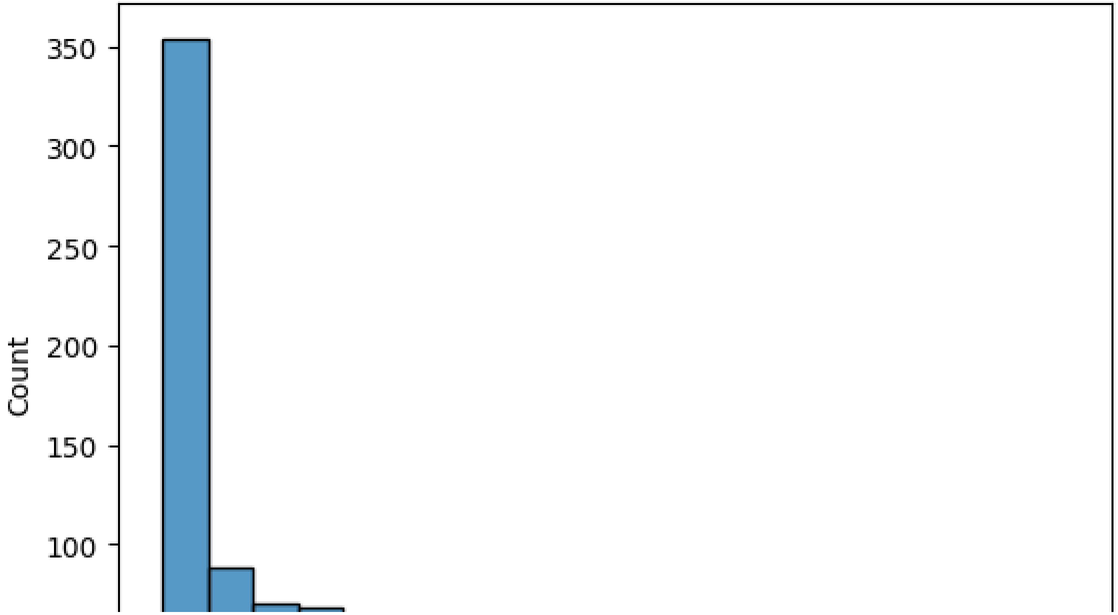

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
W40

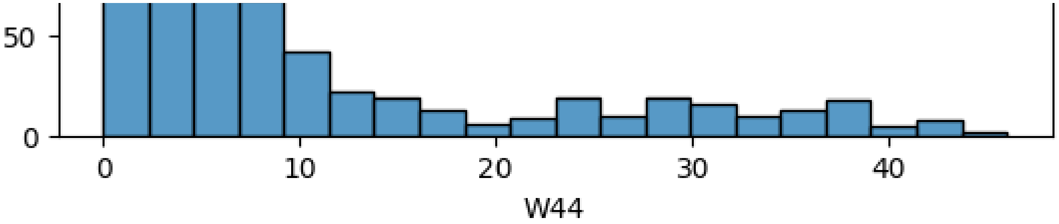
350
300
250

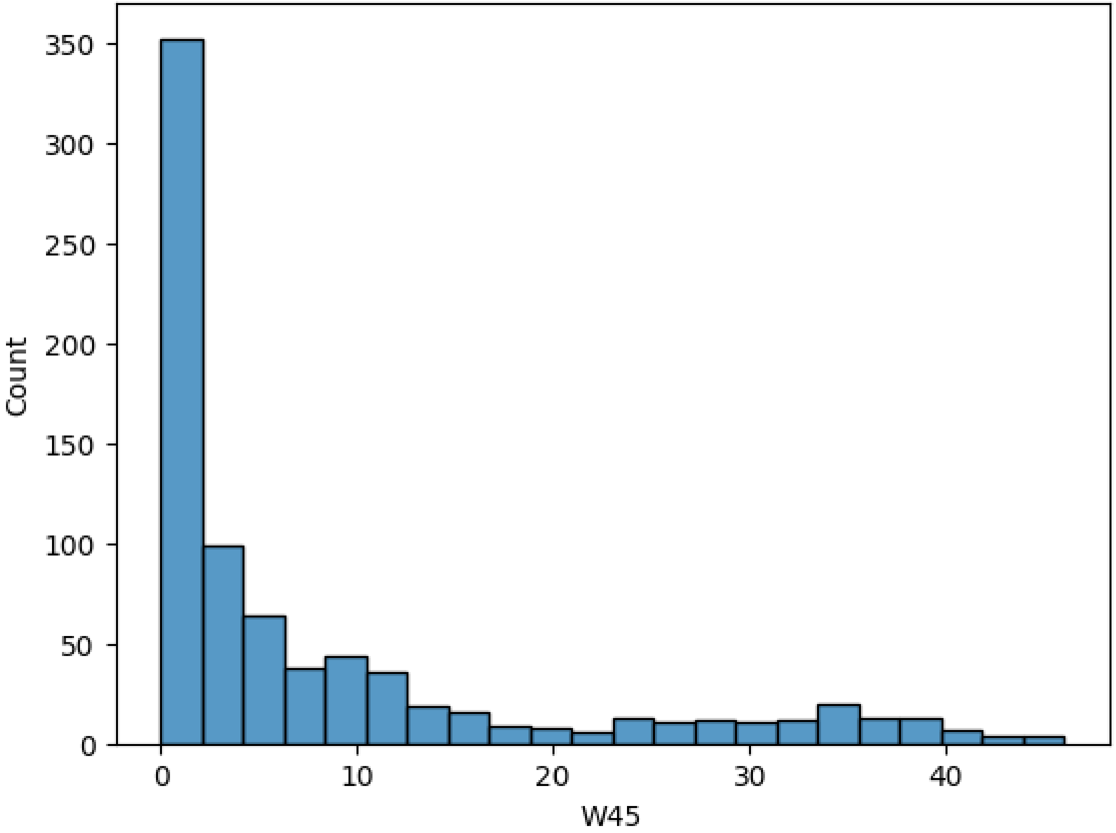

Count
200
150
100
50
0
0
10
20
30
40
50
W41


Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
W42

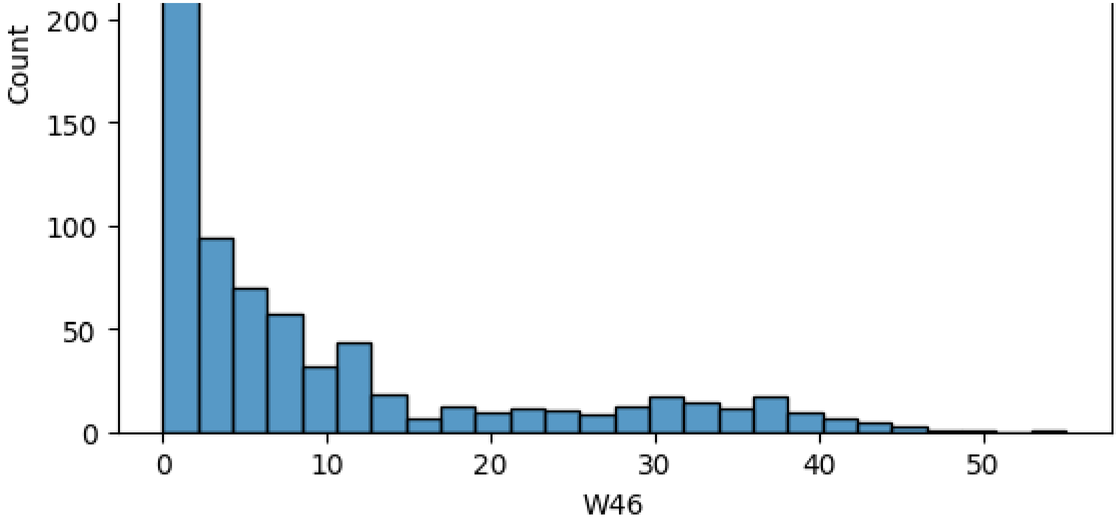

Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50
W43

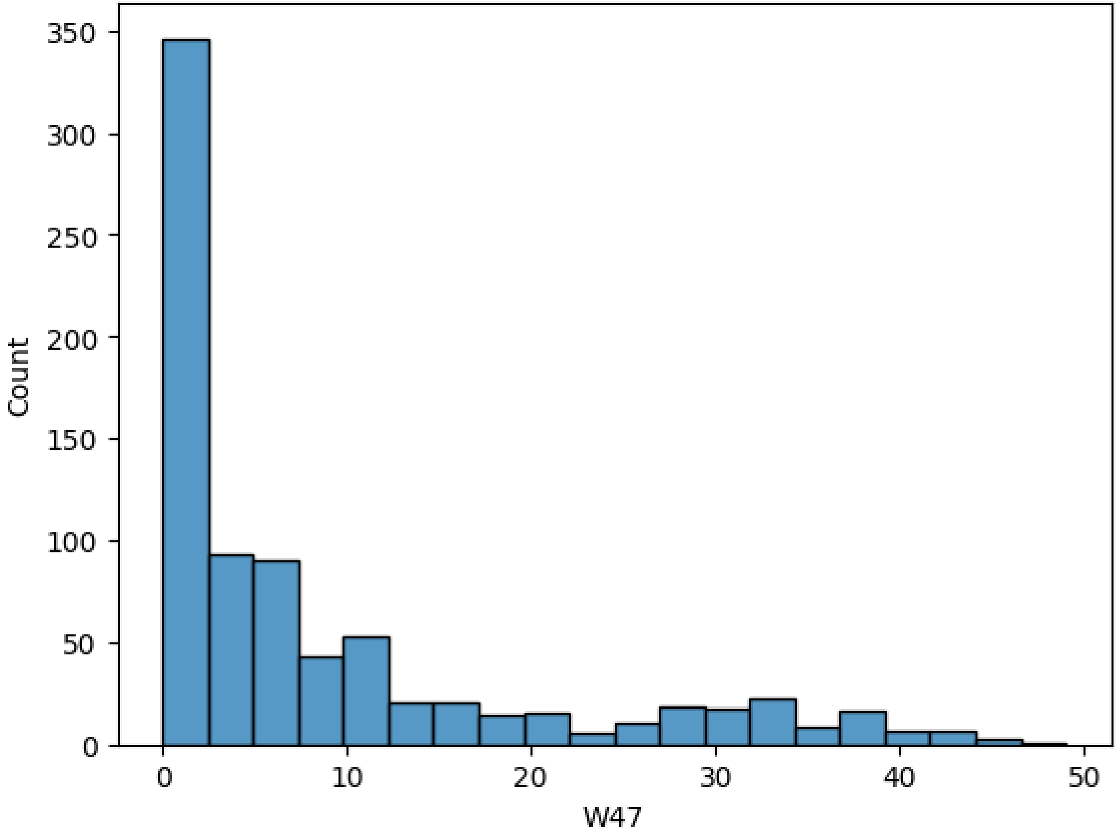

Count
100
150
200
250
300
350

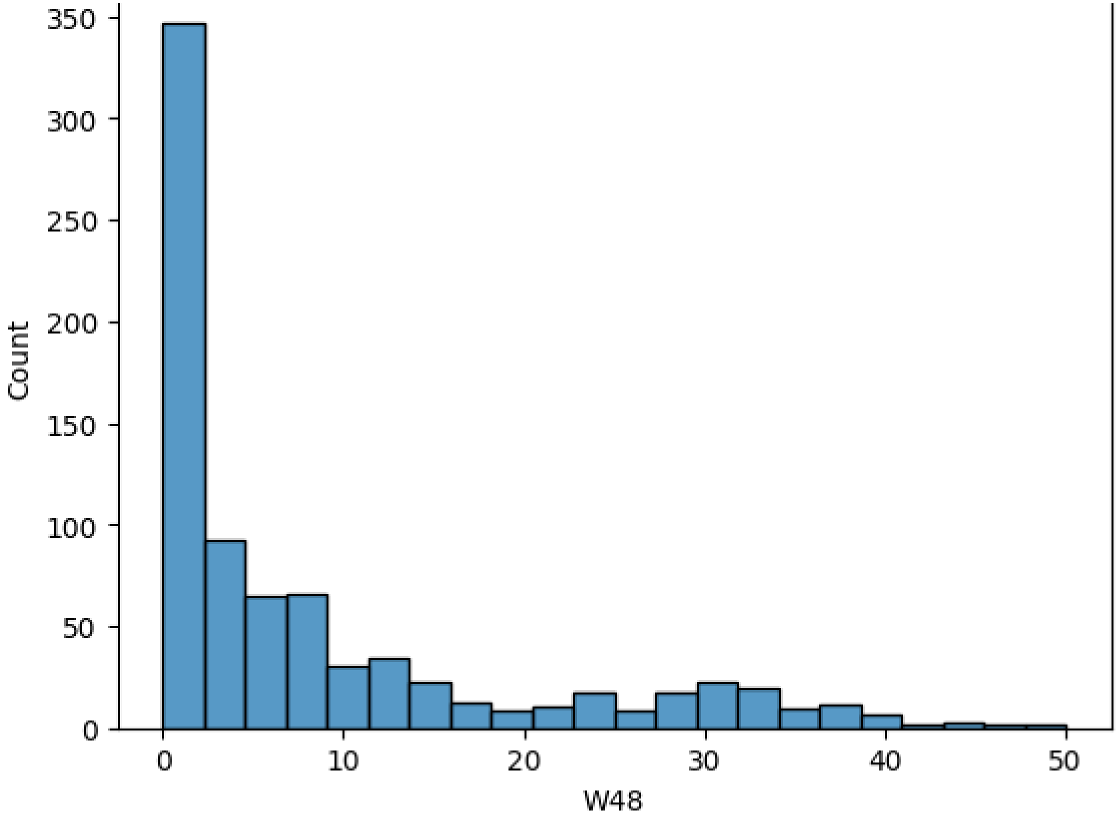

50
0
0
10
20
30
40
W44

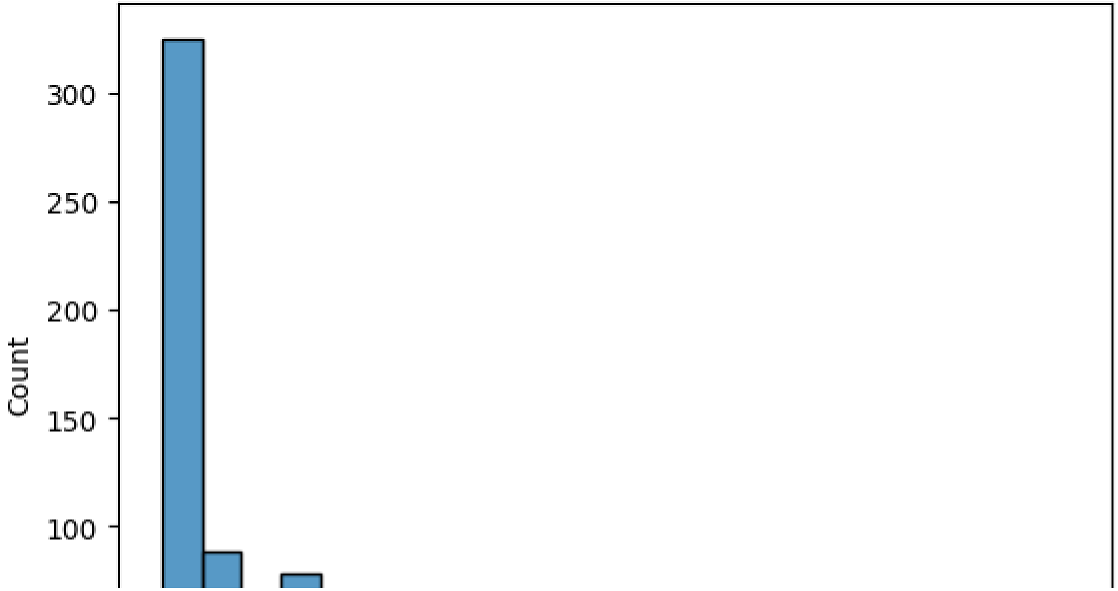

Count
350
300
250
200
150
100
50
0
0
10
20
30
40
W45

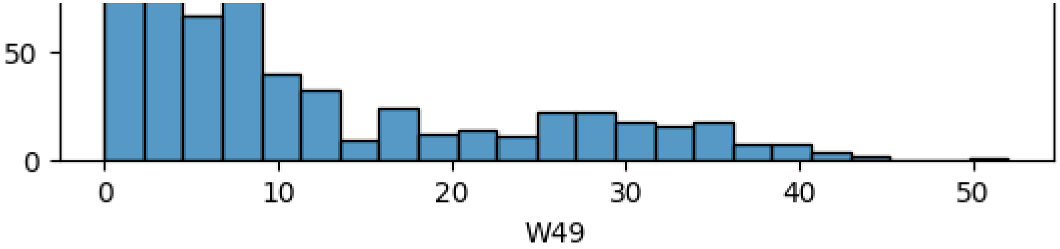
350
300
250

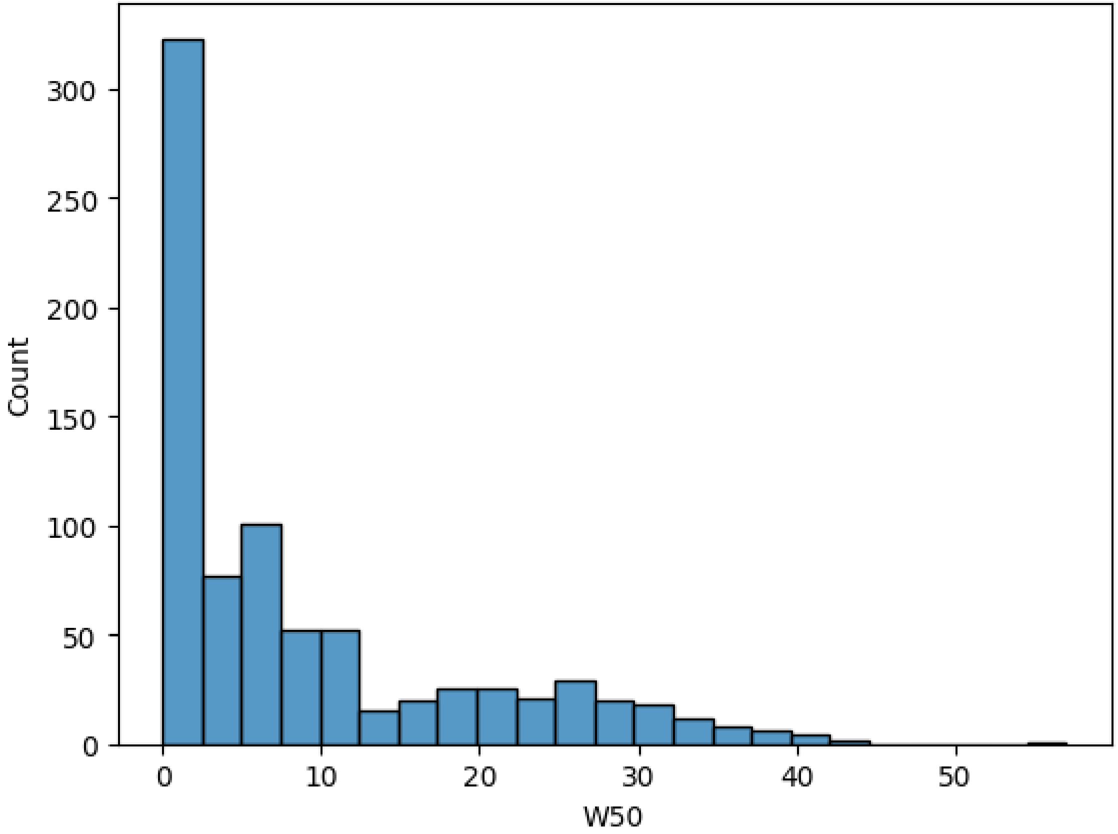

Count
0
50
100
150
200
0
10
20
30
40
50
W46


Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50
W47

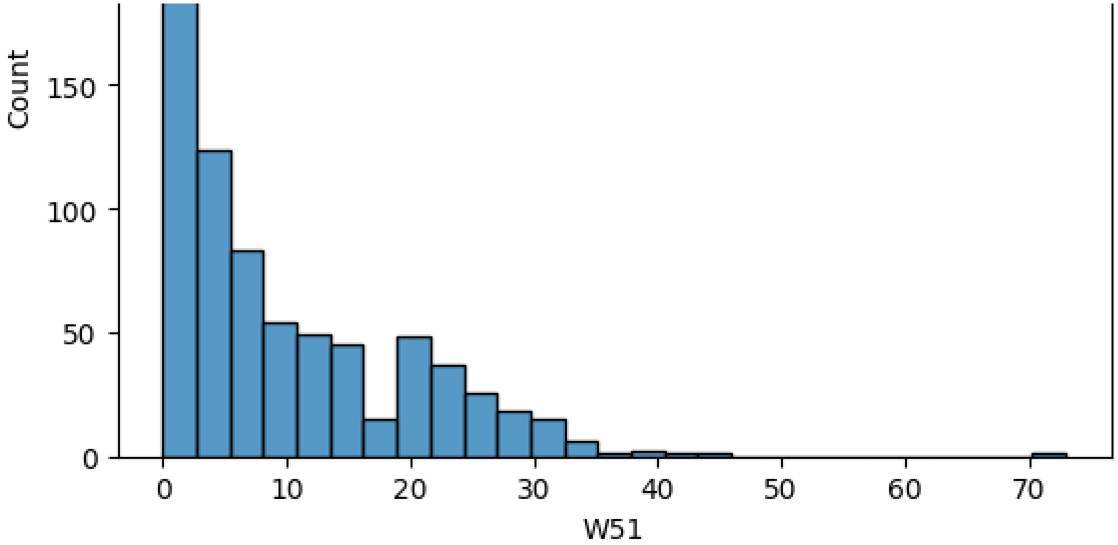

Count
0
50
100
150
200
250
300
350
0
10
20
30
40
50
W48

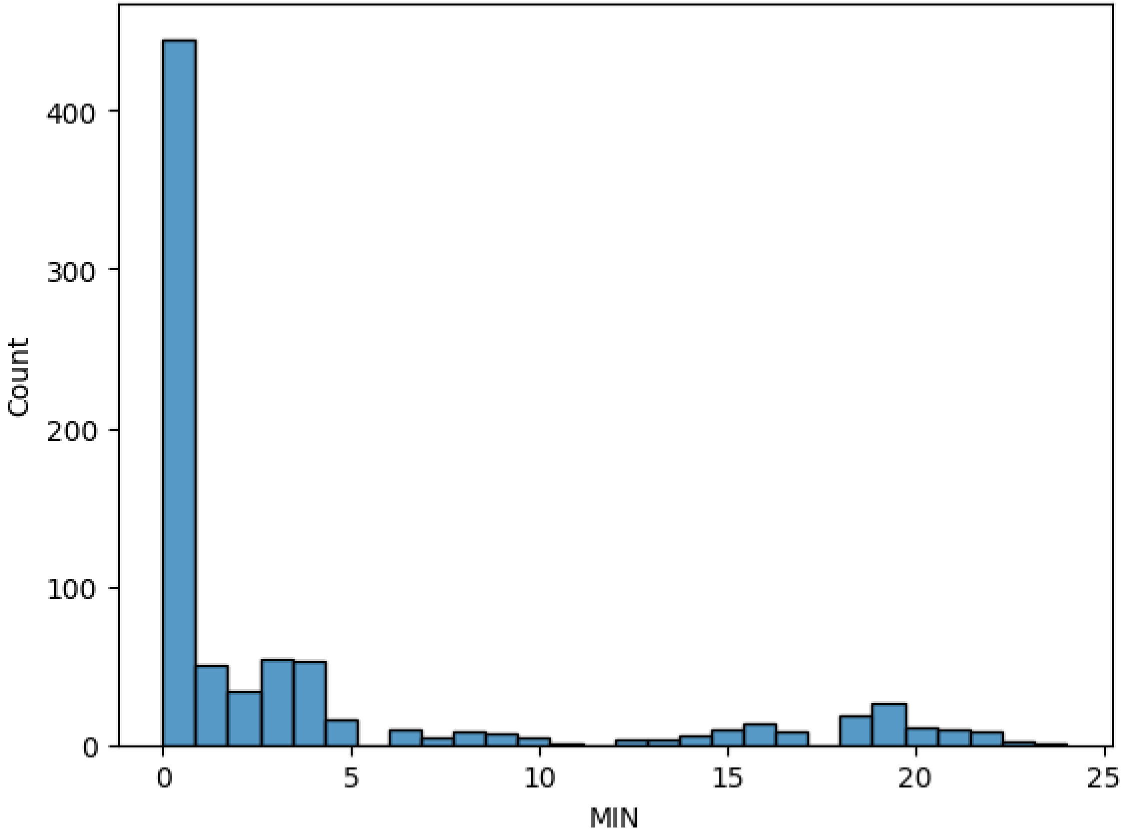

Count
100
150
200
250
300

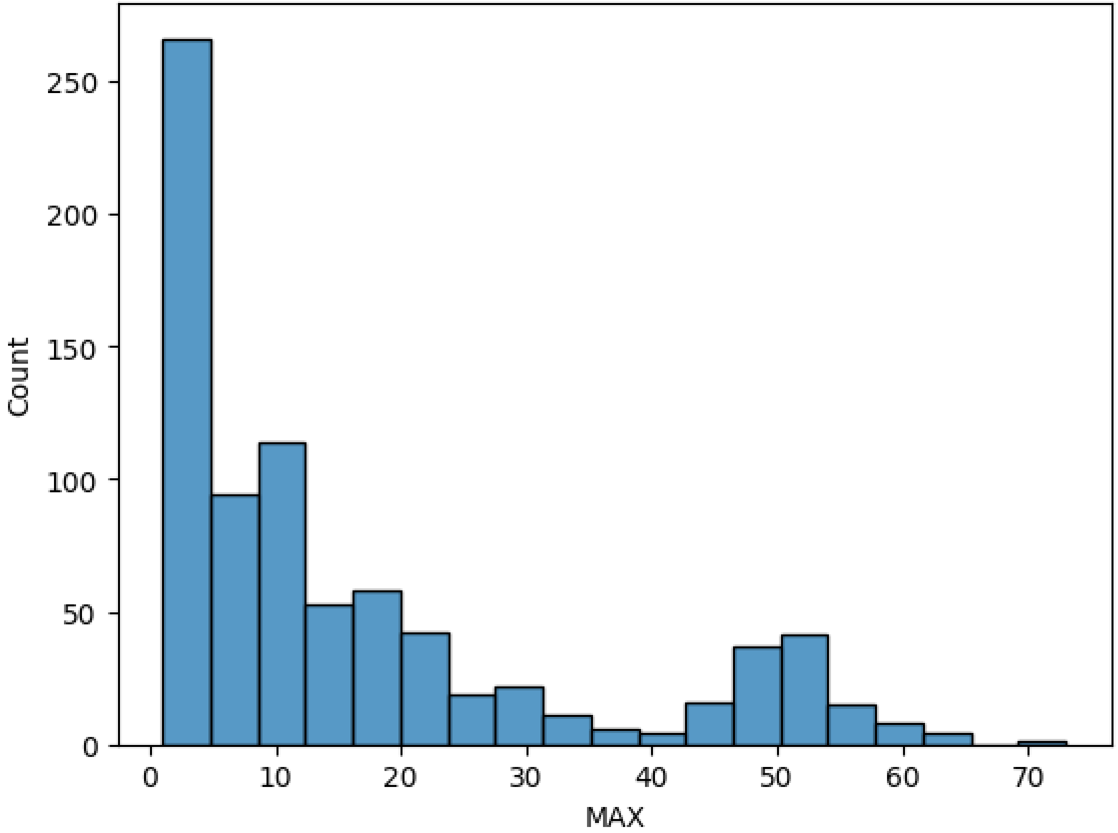

50
0
0
10
20
30
40
50
W49

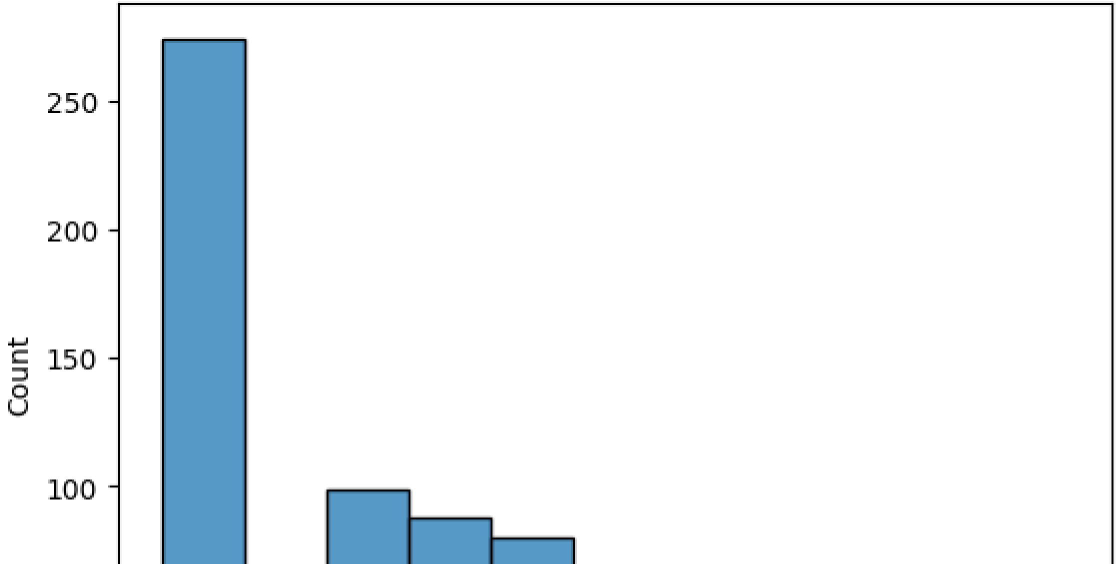

Count
300
250
200
150
100
50
0
0
10
20
30
40
50
W50

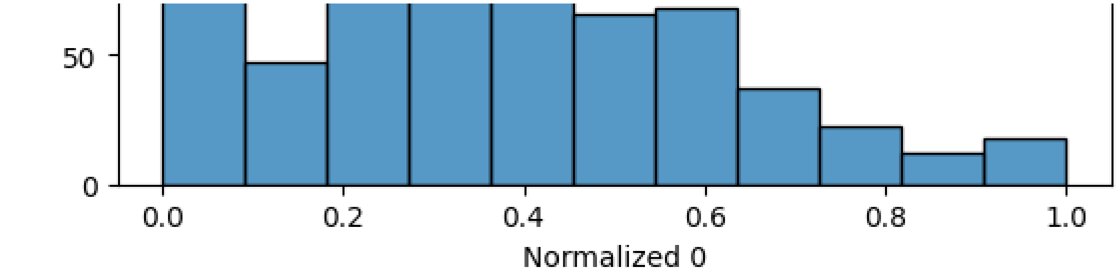
250
200

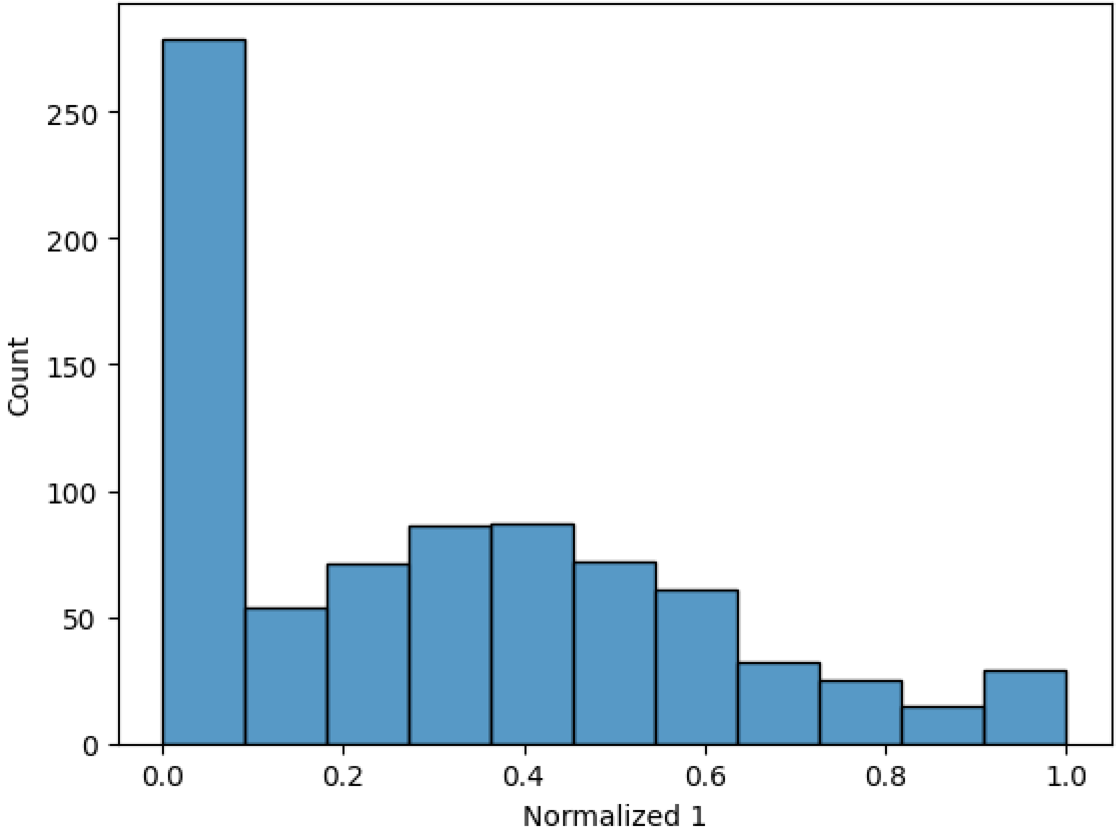

Count
150
100
50
0
0
10
20
30
40
50
60
70
W51


Count
400
300
200
100
0
0
5
10
15
20
25
MIN

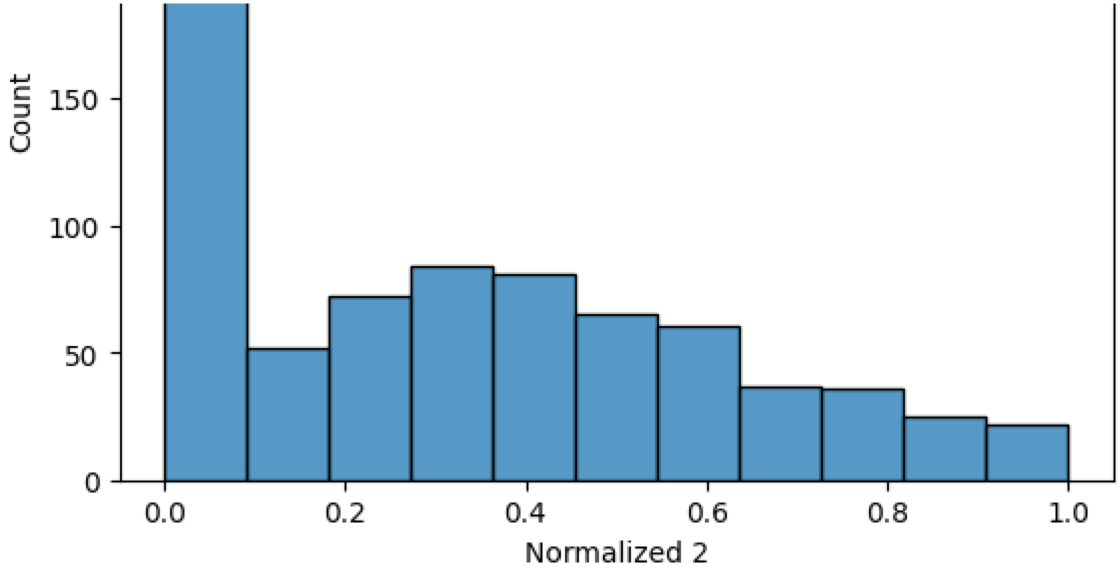

250
200
150
100
50
0
Count
0
10
20
30
40
50
60
70
MAX

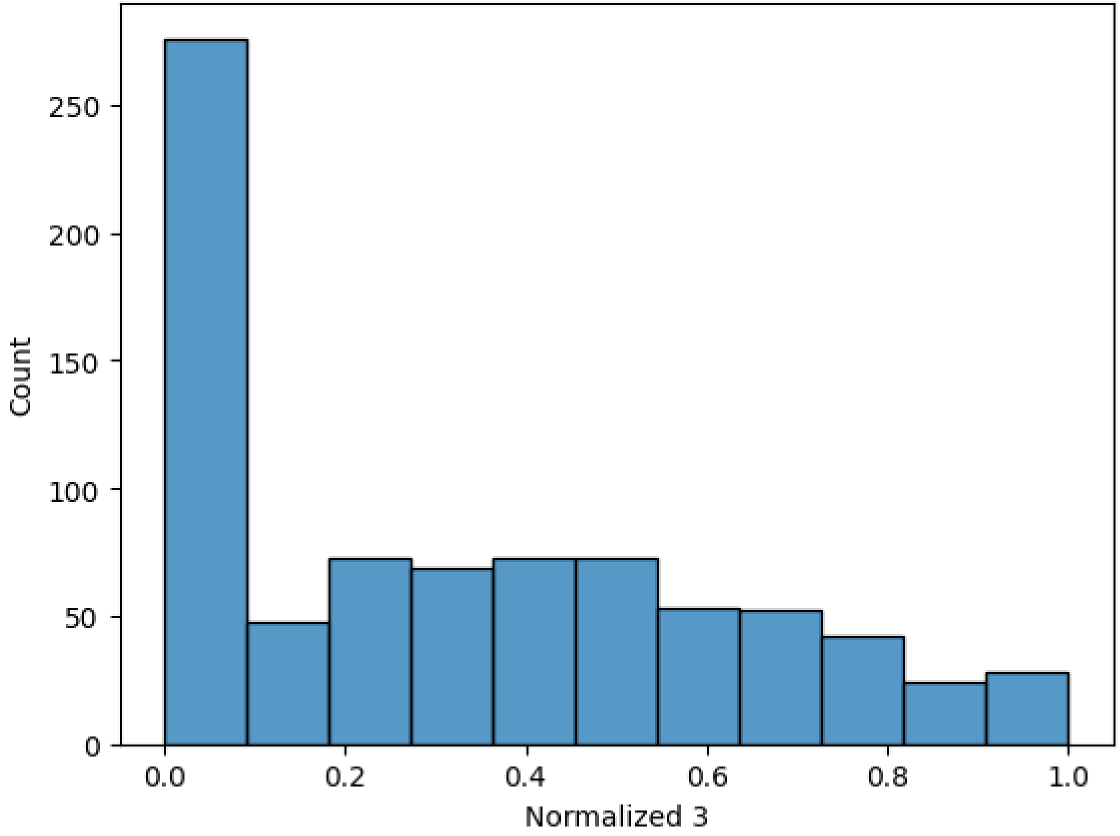

250
200
150
100
Count

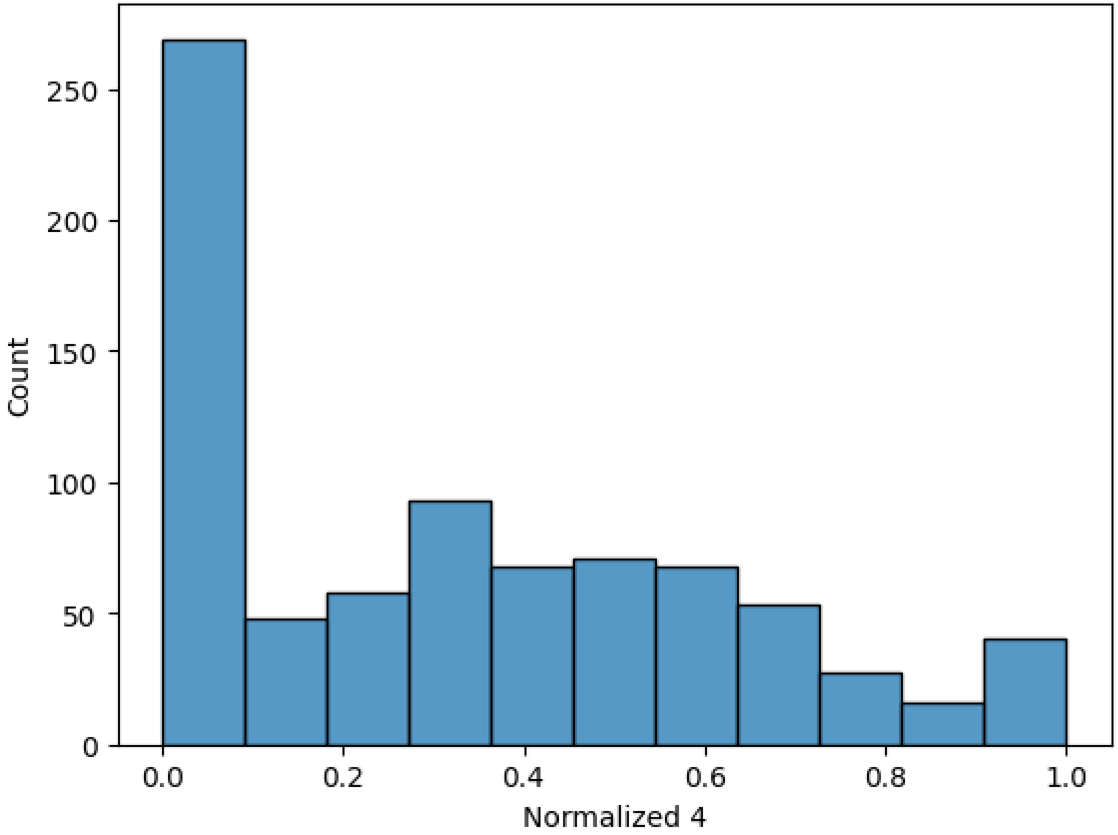

50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 0

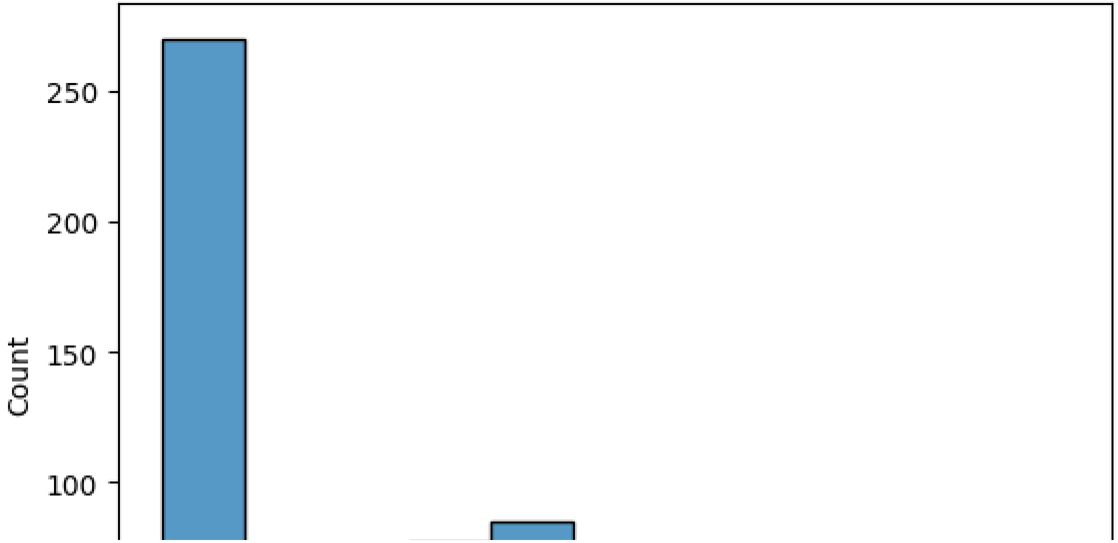

250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 1

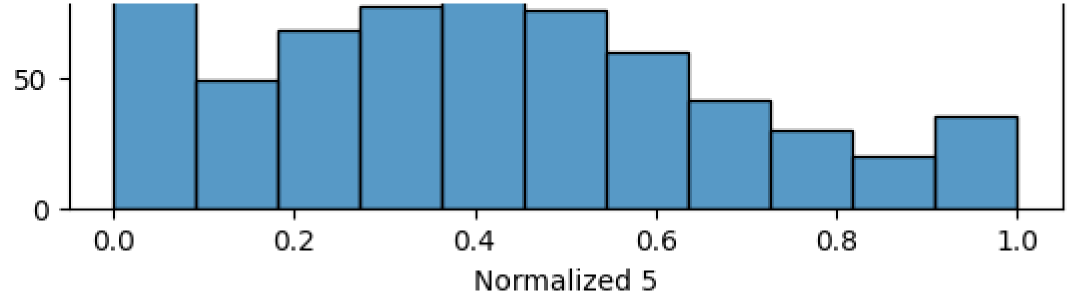
250
200

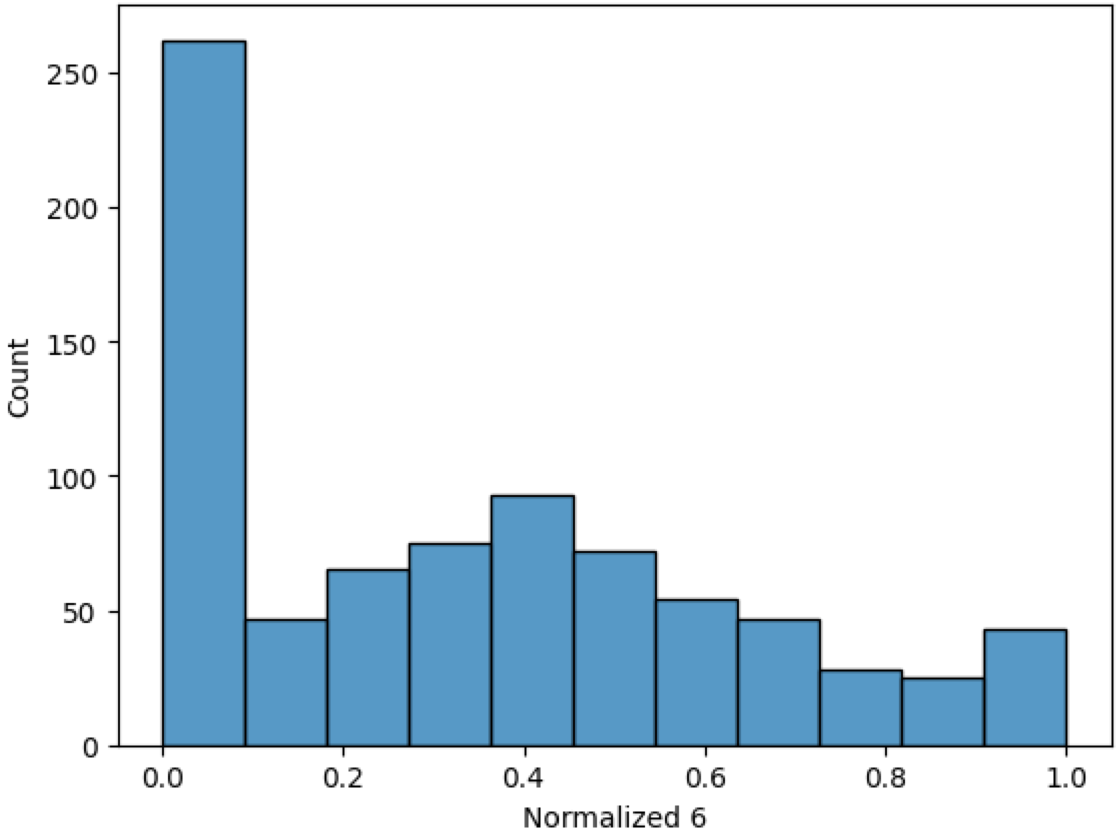

Count
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 2


Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 3

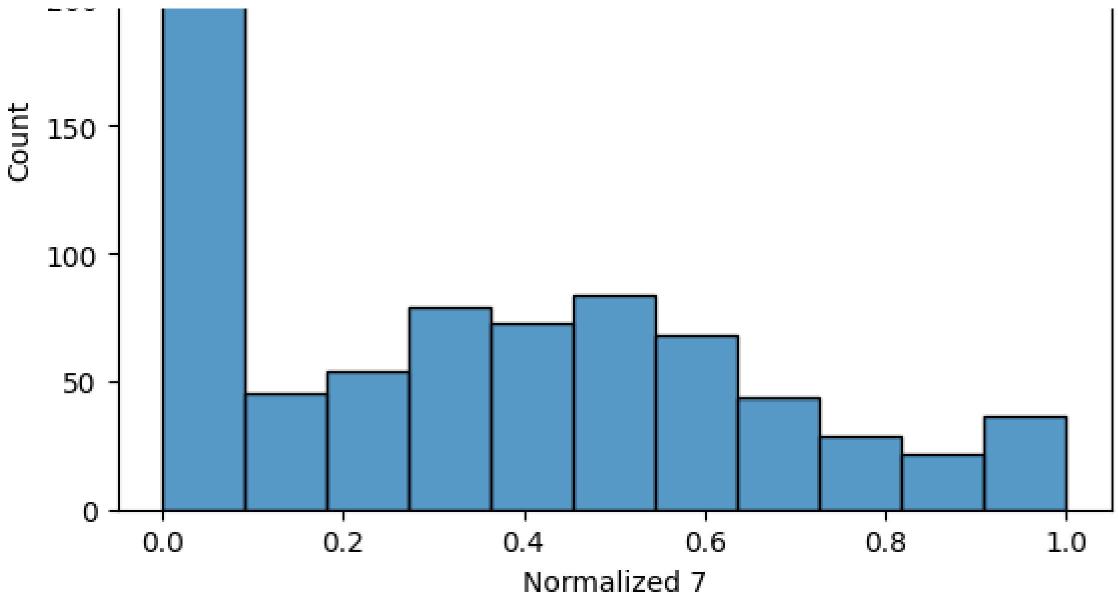

Count
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 4


Count
100
150
200
250

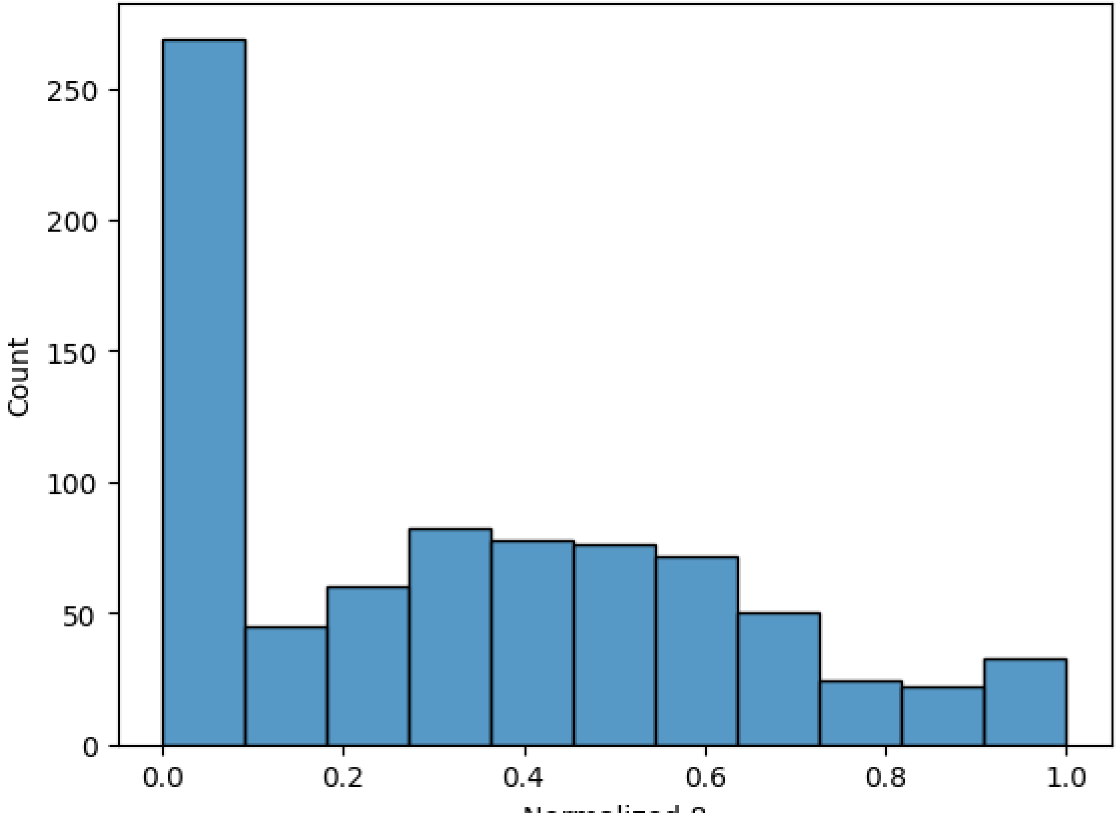

50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 5


Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 6

250
200

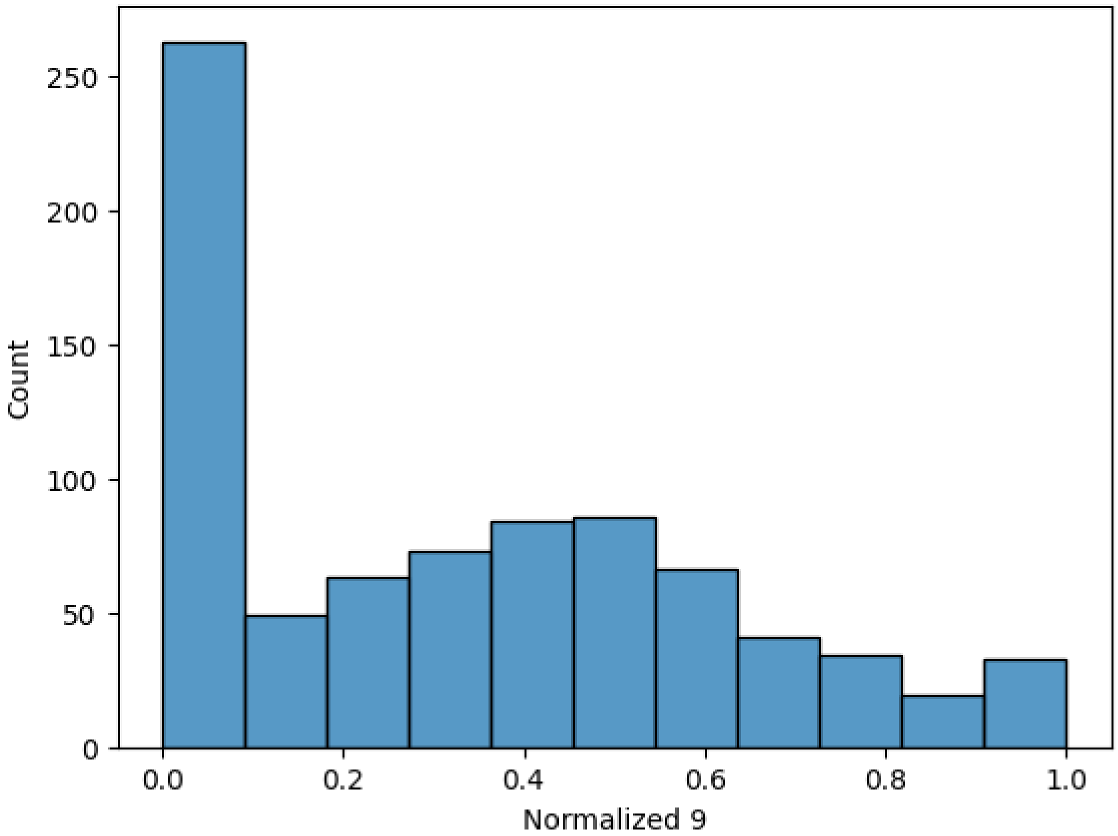


Count
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 7

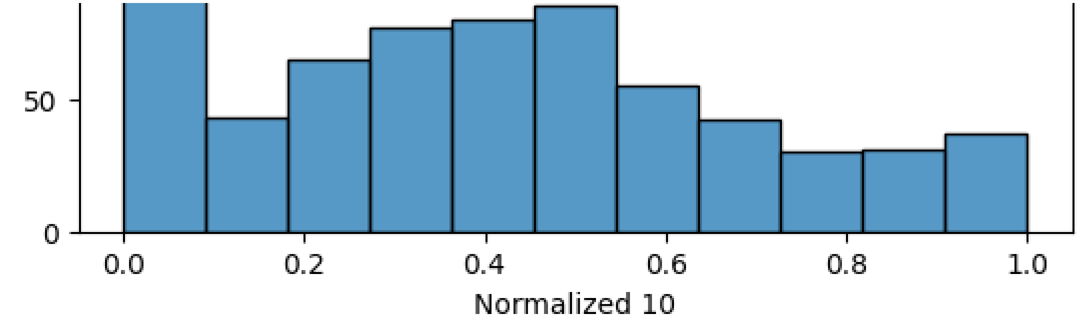


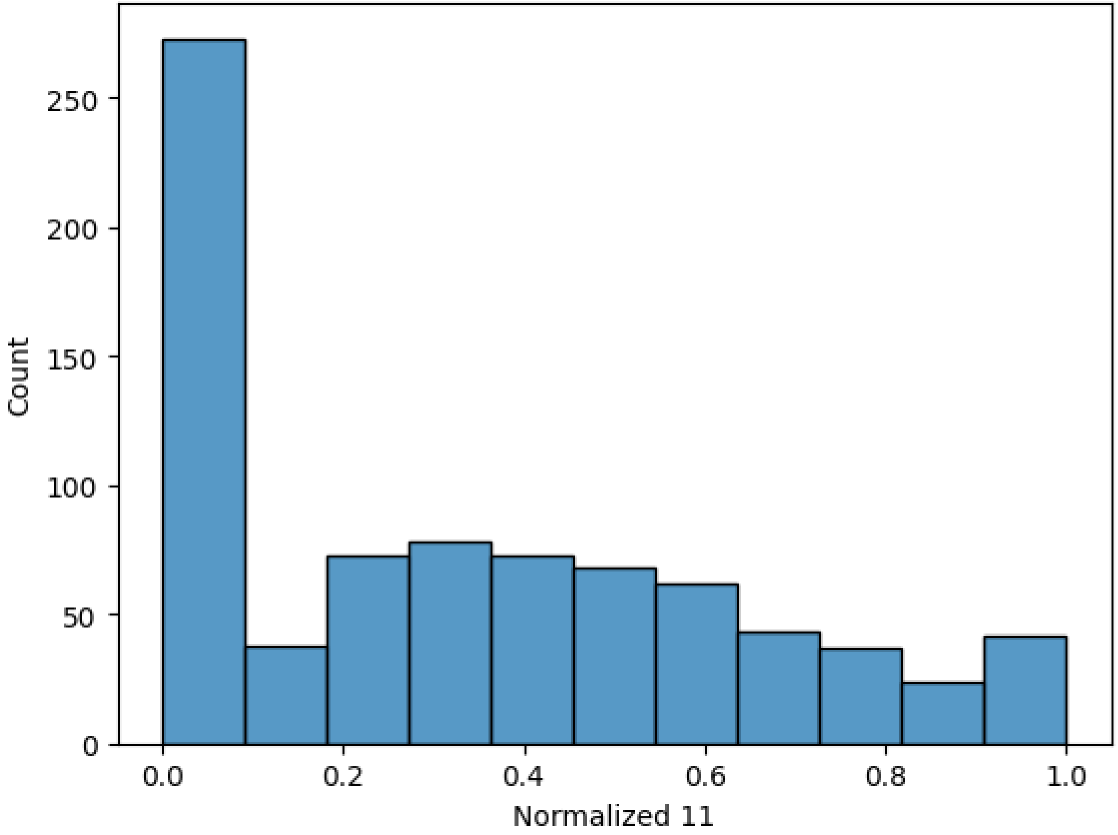
Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0

Normalized 8

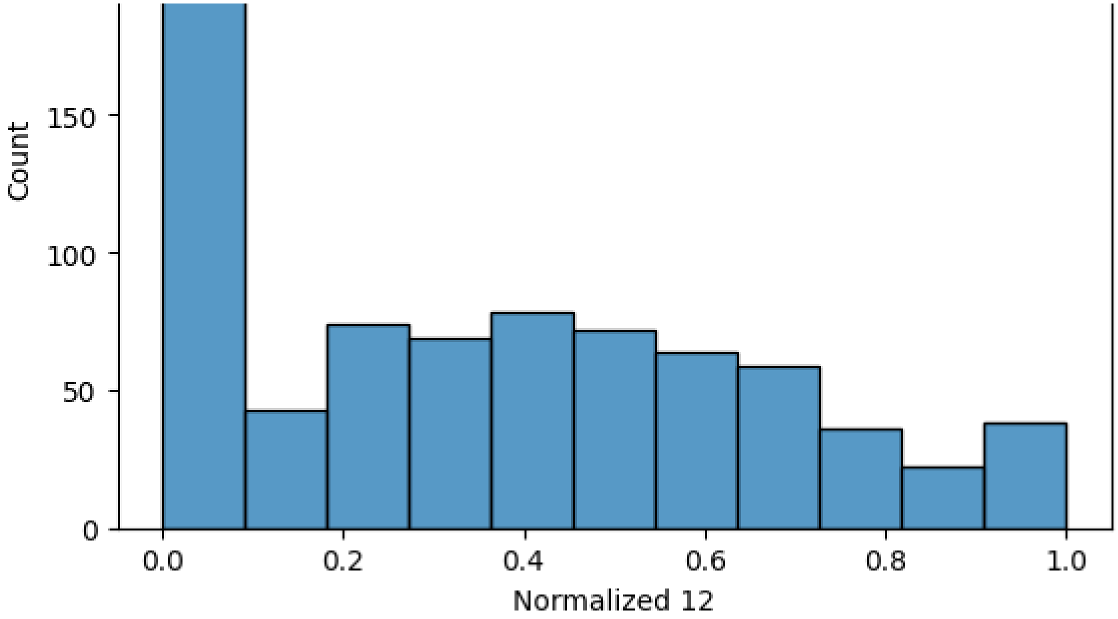


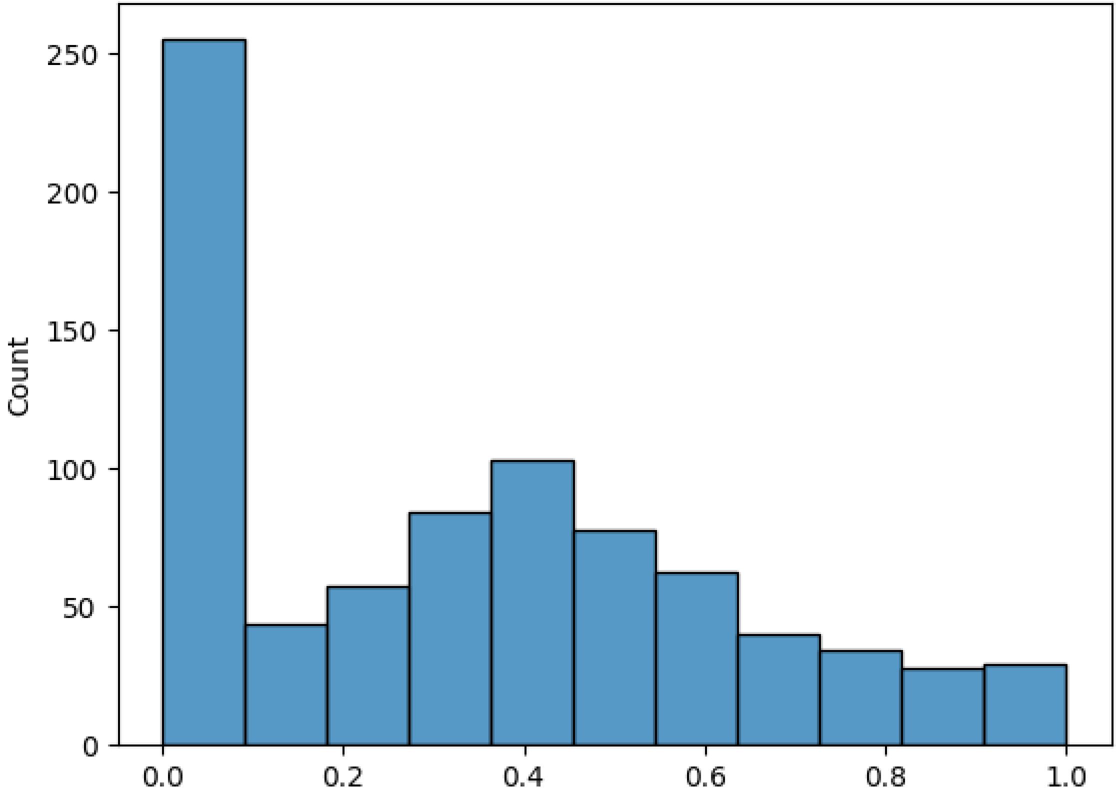
Count
250
200
150
100


50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 10

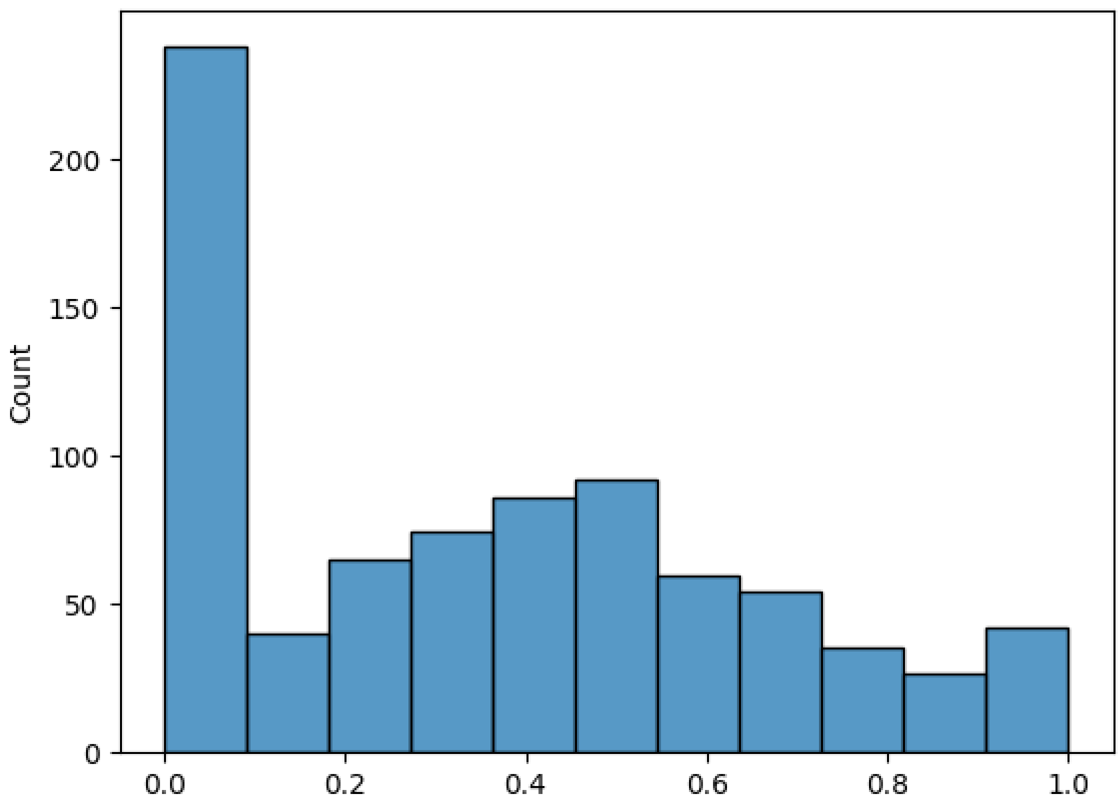

250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 11

250
200


Count
0
50
100
150
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 12

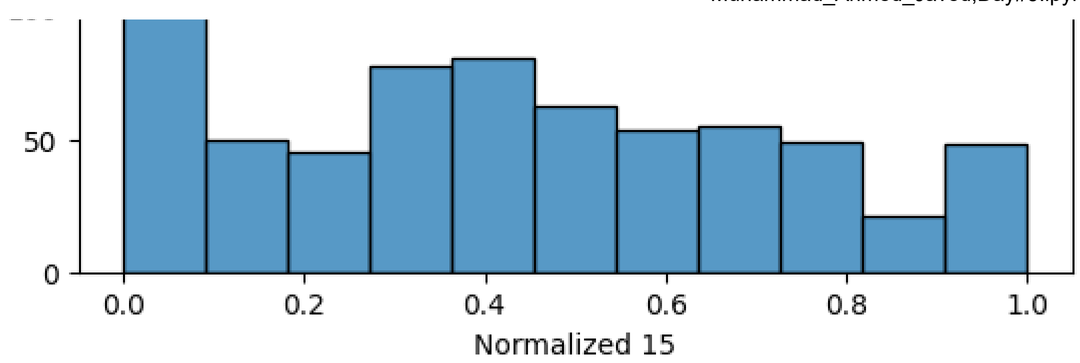

Count
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0

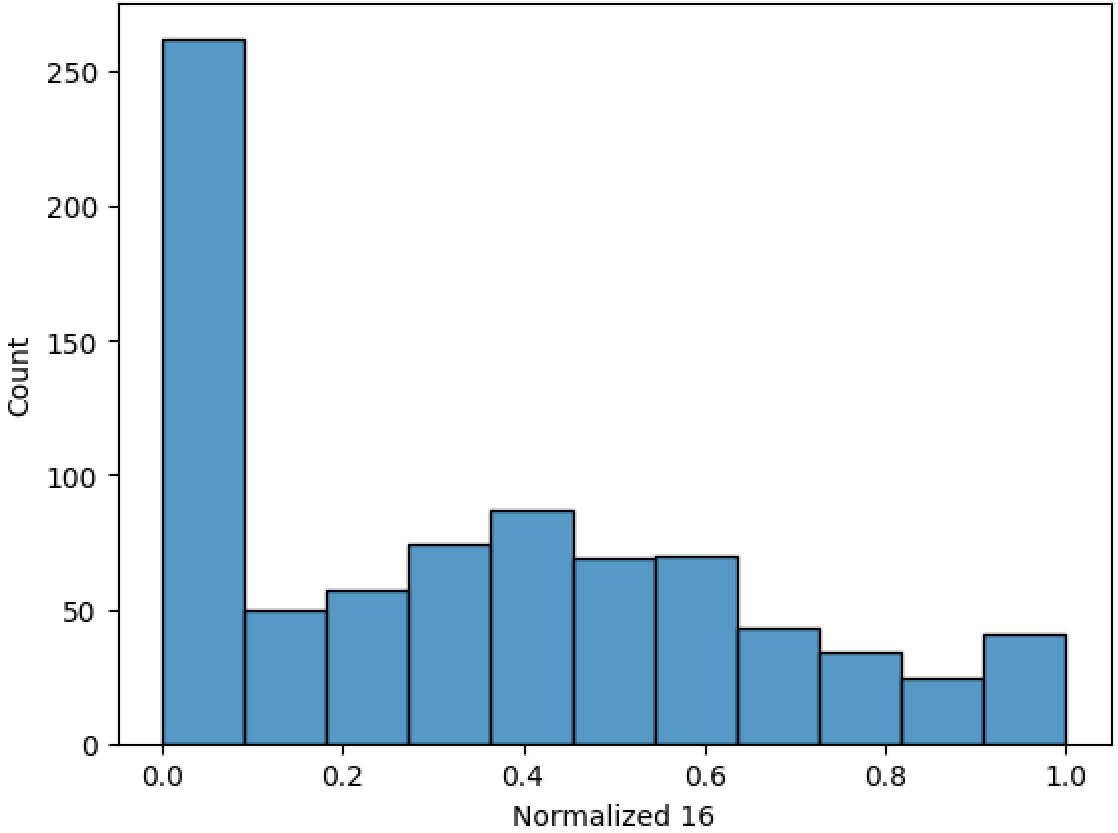
Normalized 13

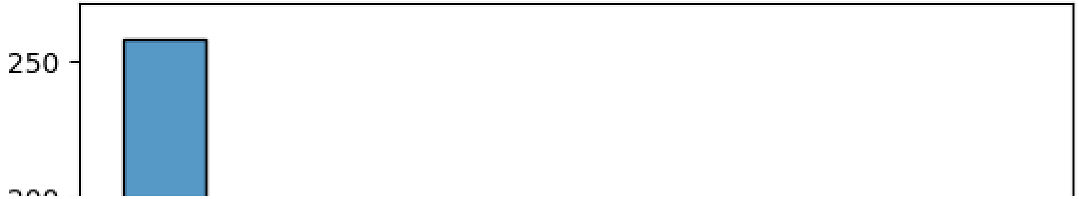


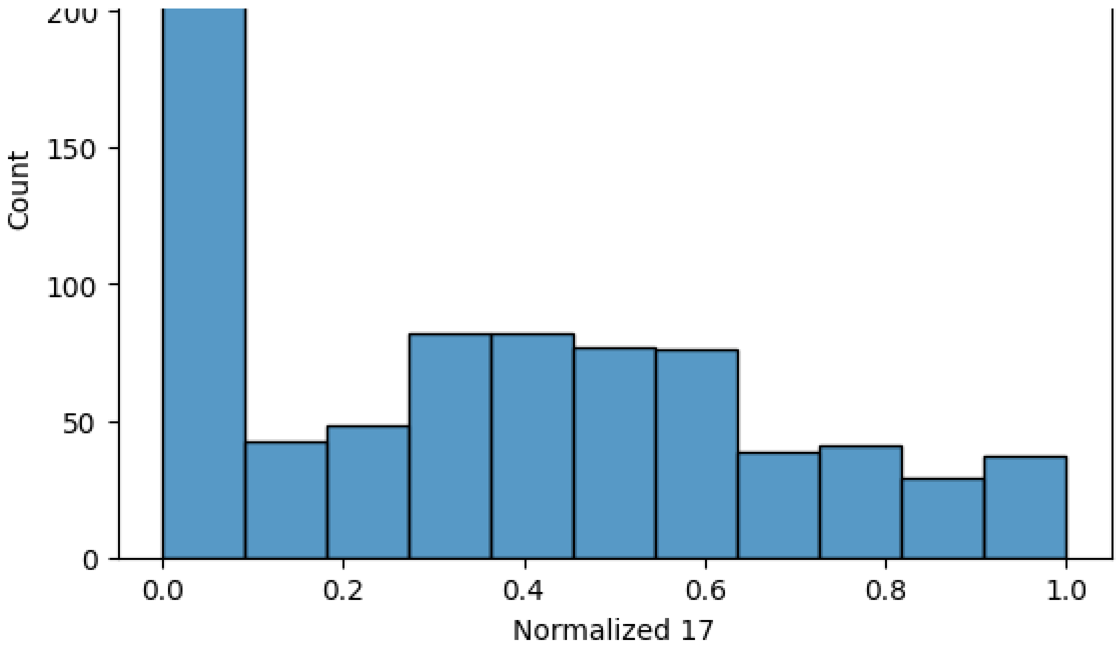
Normalized 14

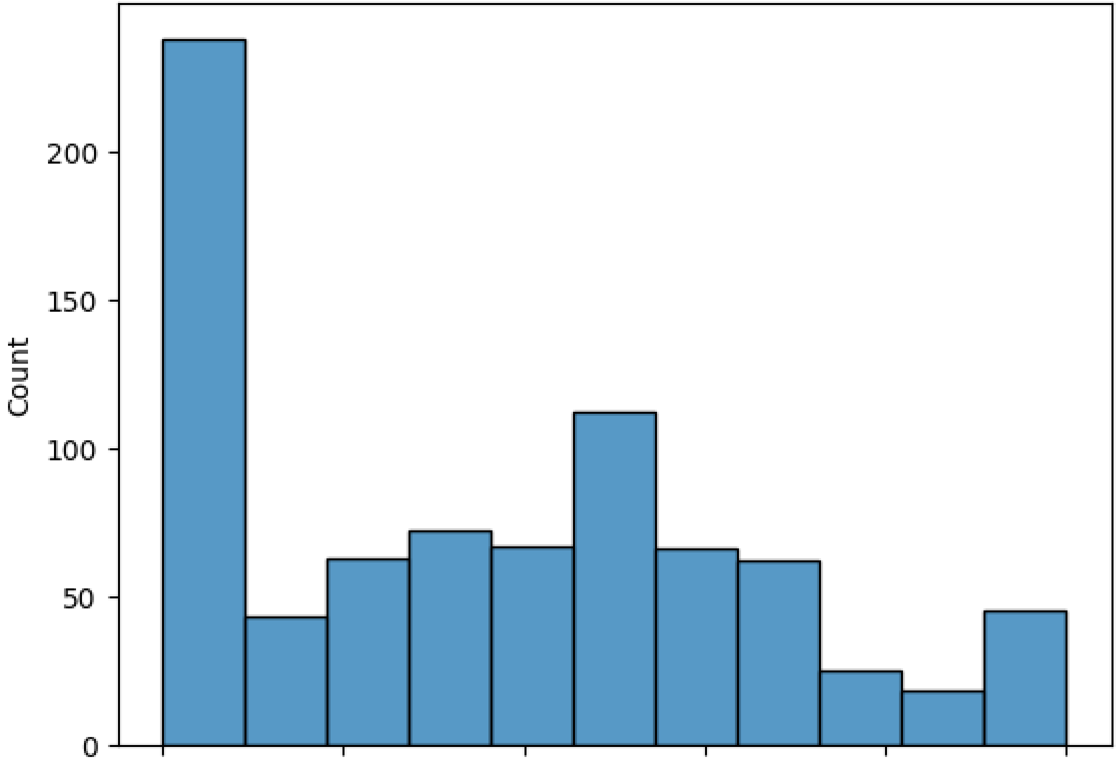
Count
100
150
200
250


50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 15

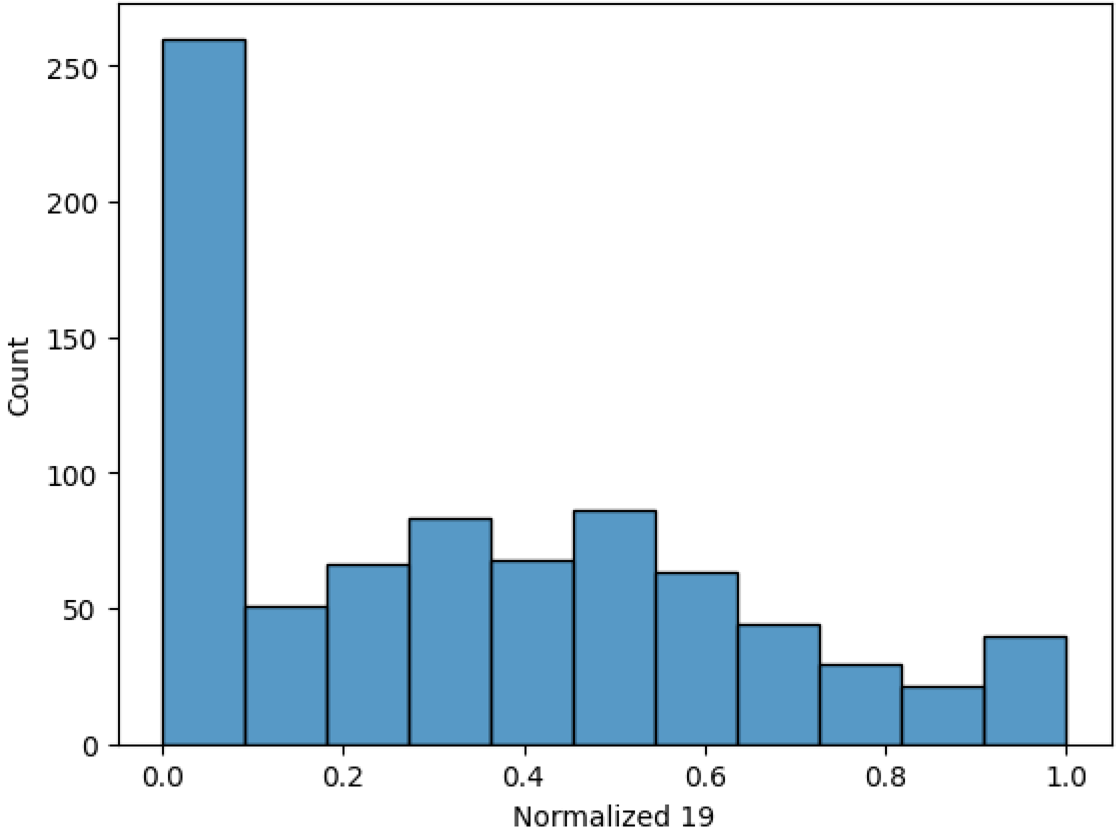

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 16


250

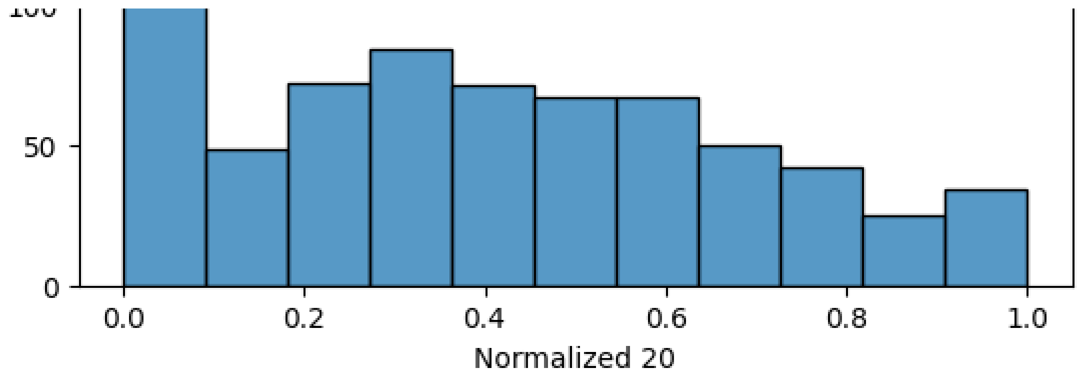

200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 17

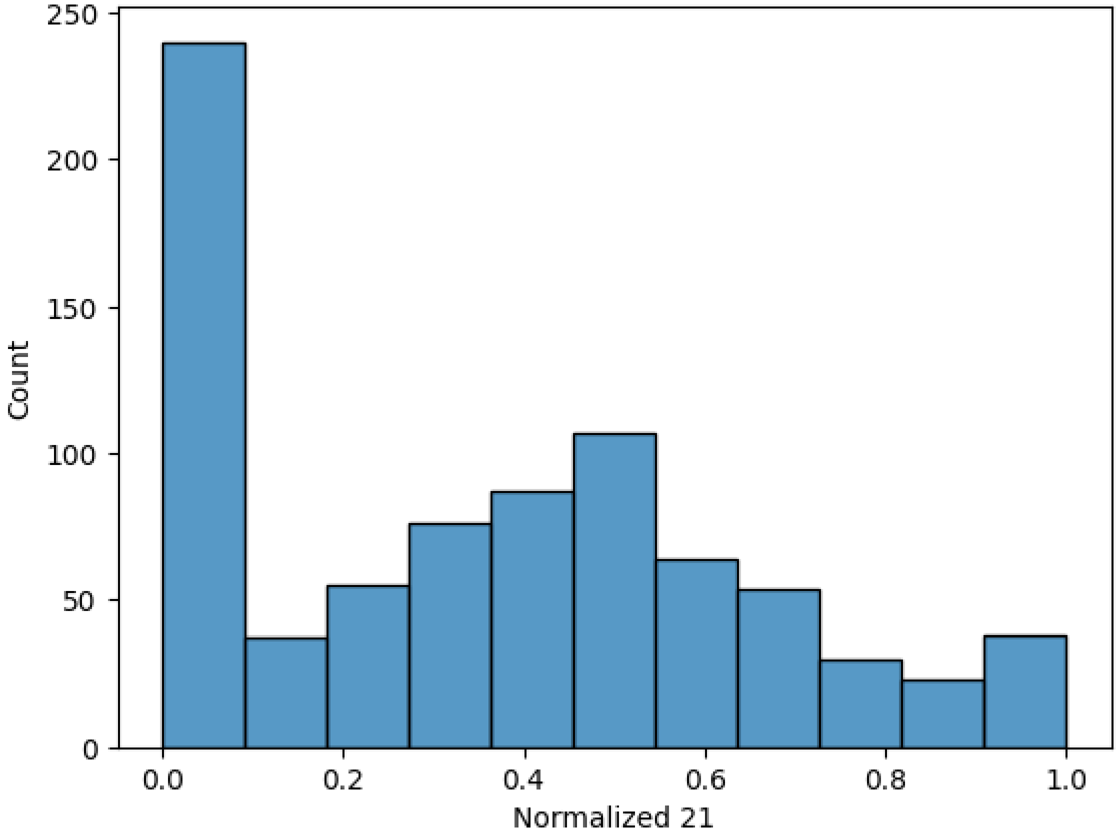

200
150
100
50
0
Count

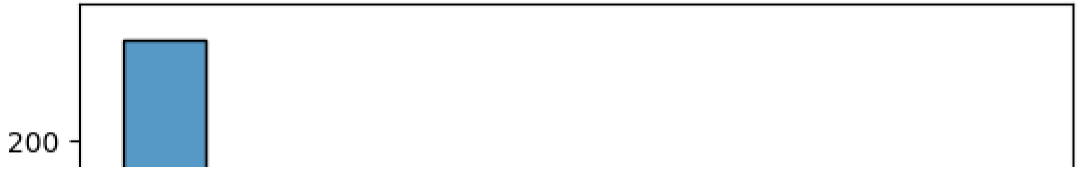
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 18

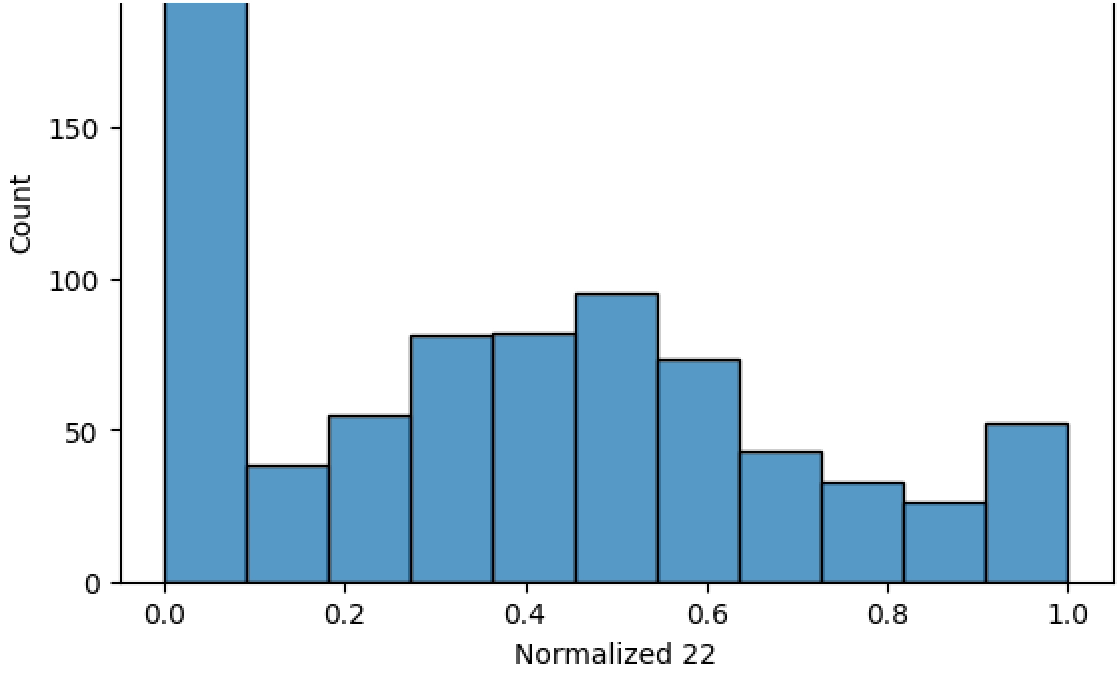

Count
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 19

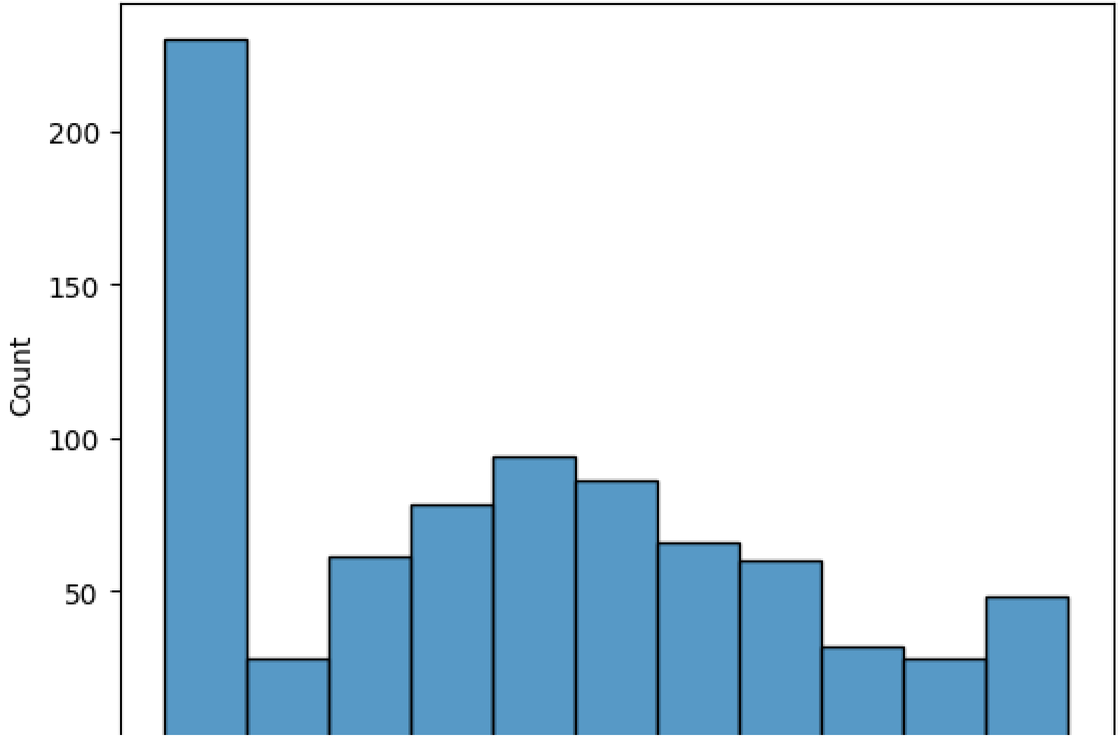
Count
100
150
200
250


100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 20

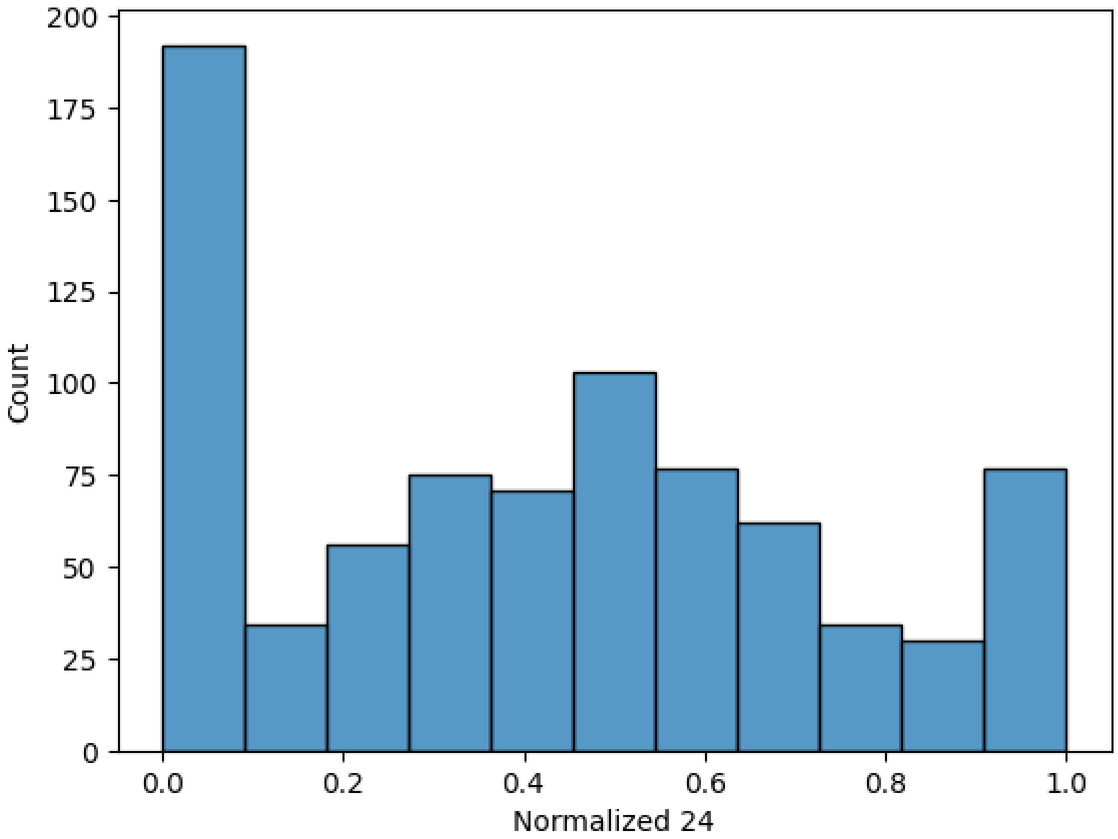

250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 21


200

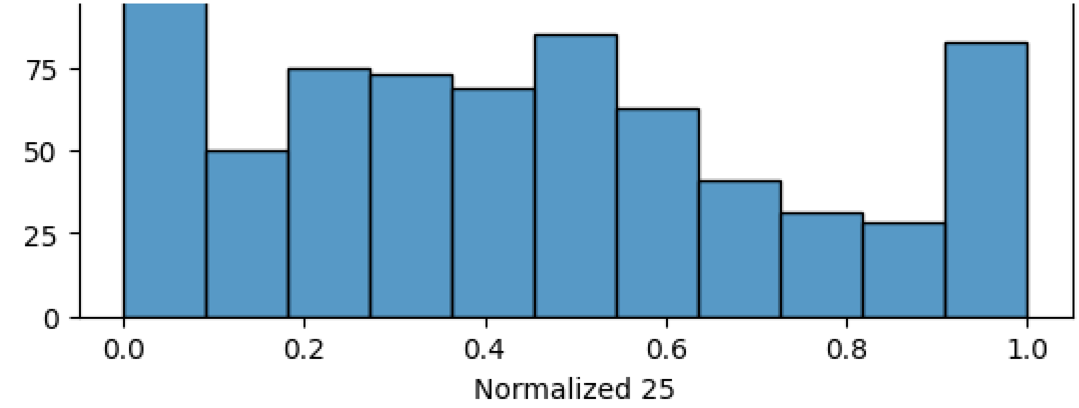

Count
0
50
100
150
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 22

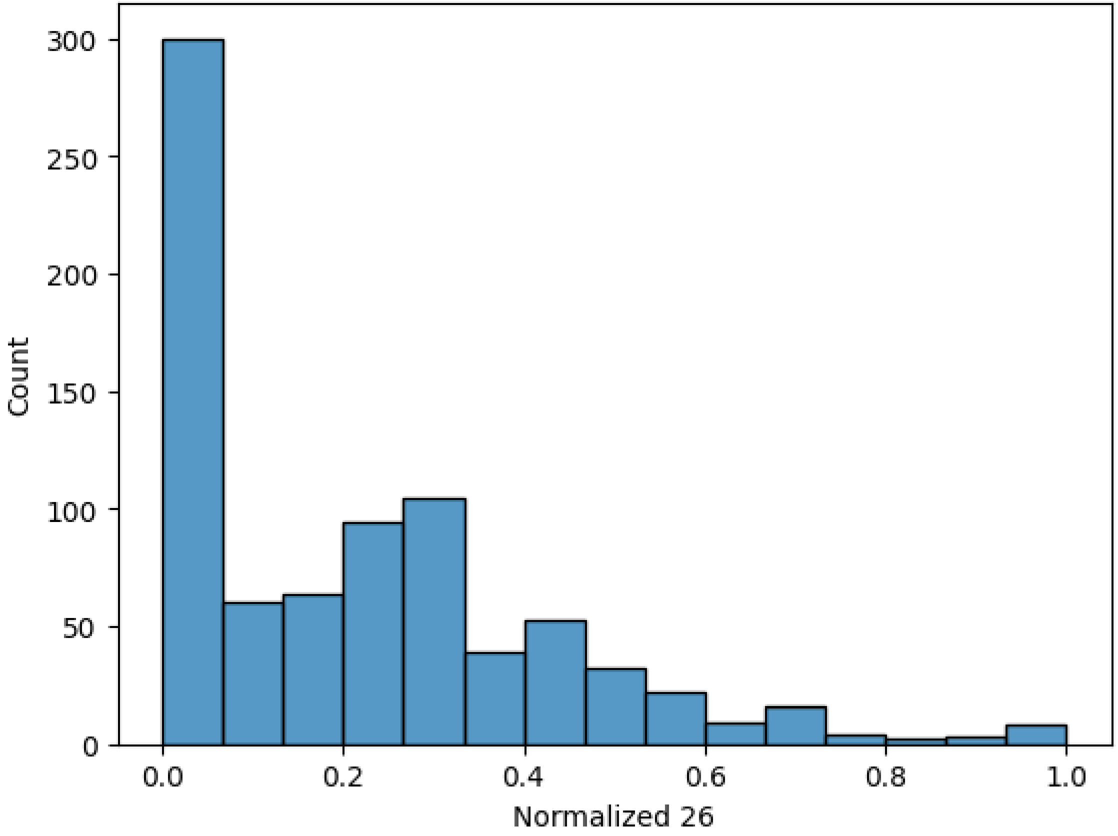

Count
50
100
150
200

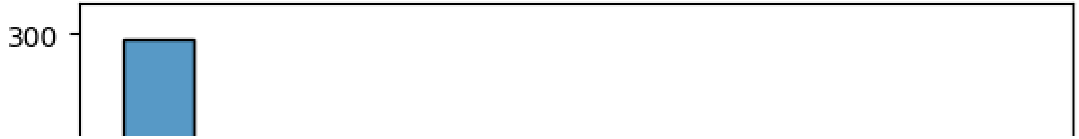
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 23

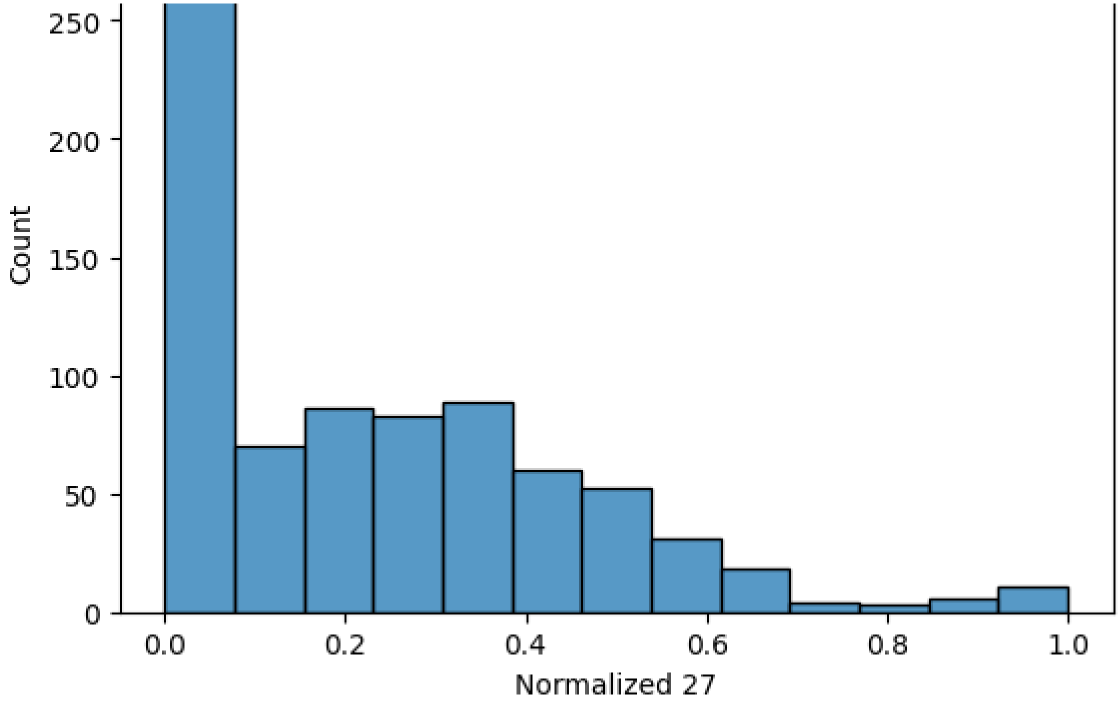

Count
200
175
150
125
100
75
50
25
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 24

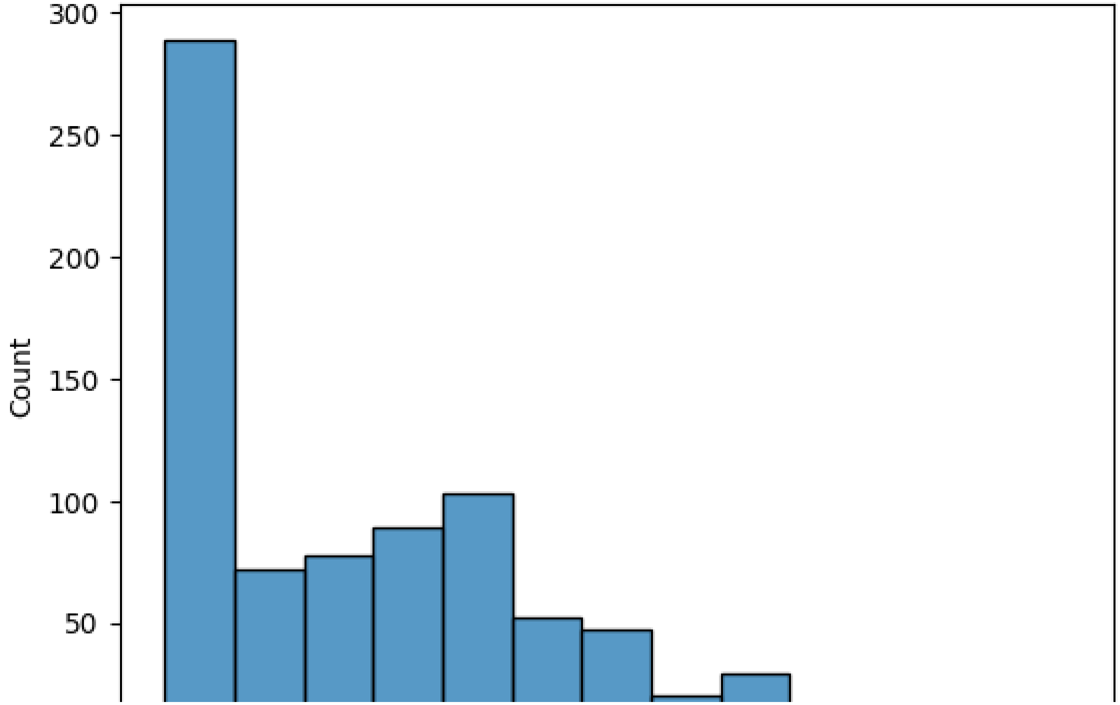
Count
200
175
150
125
100

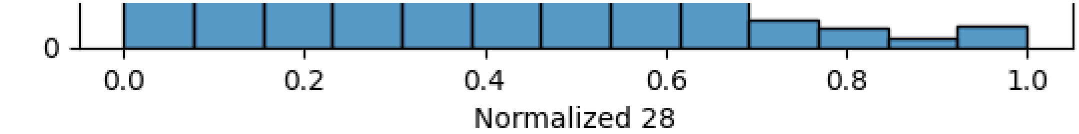

75
50
25
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 25

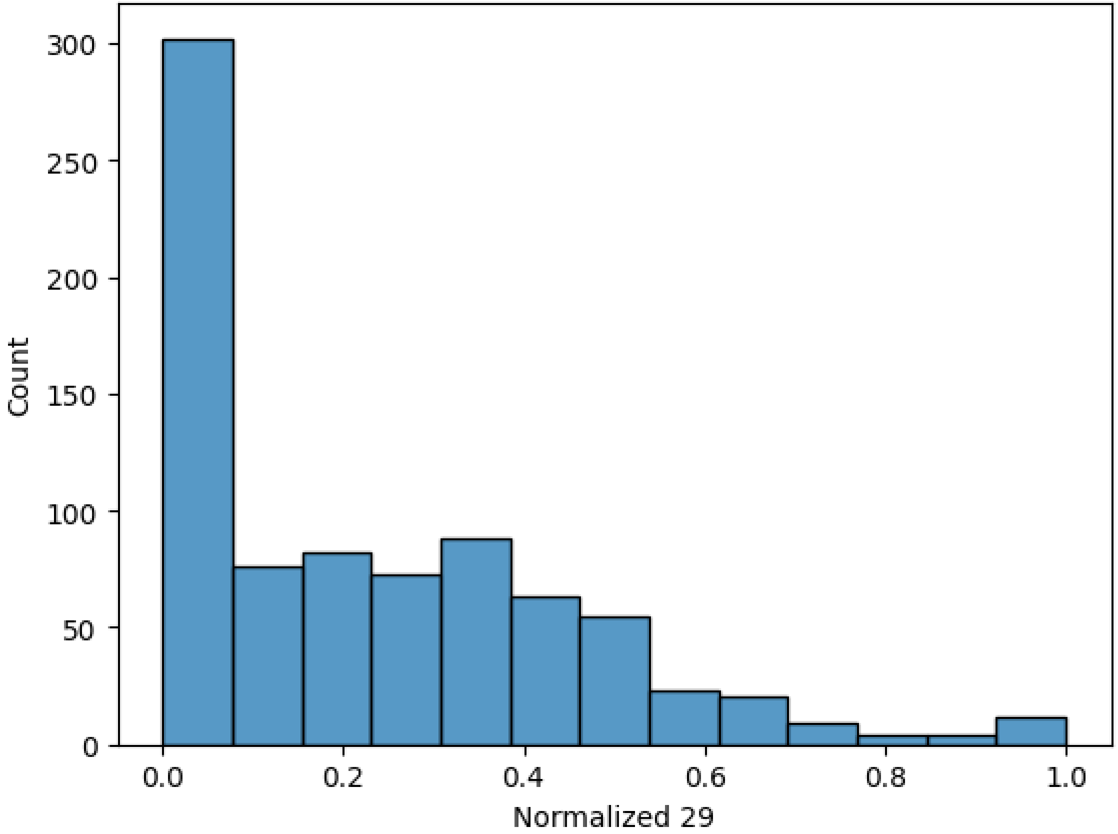

300
250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 26

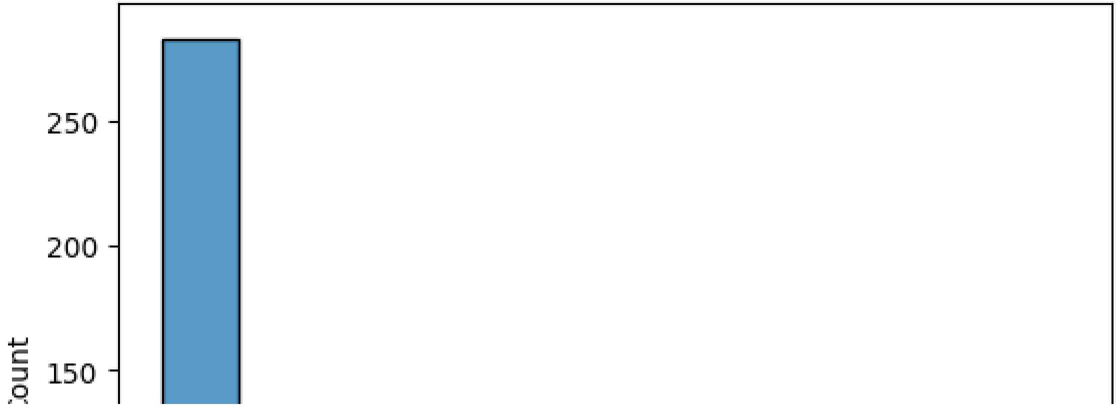

300

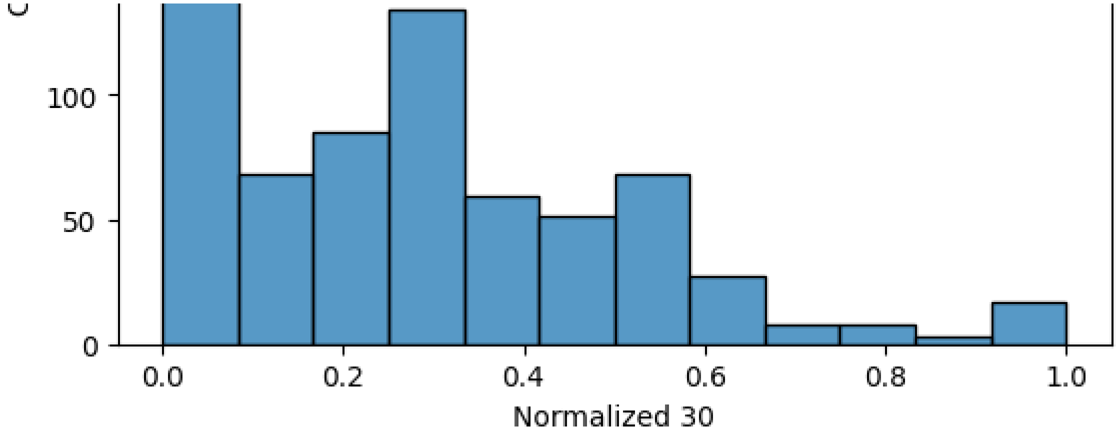

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 27

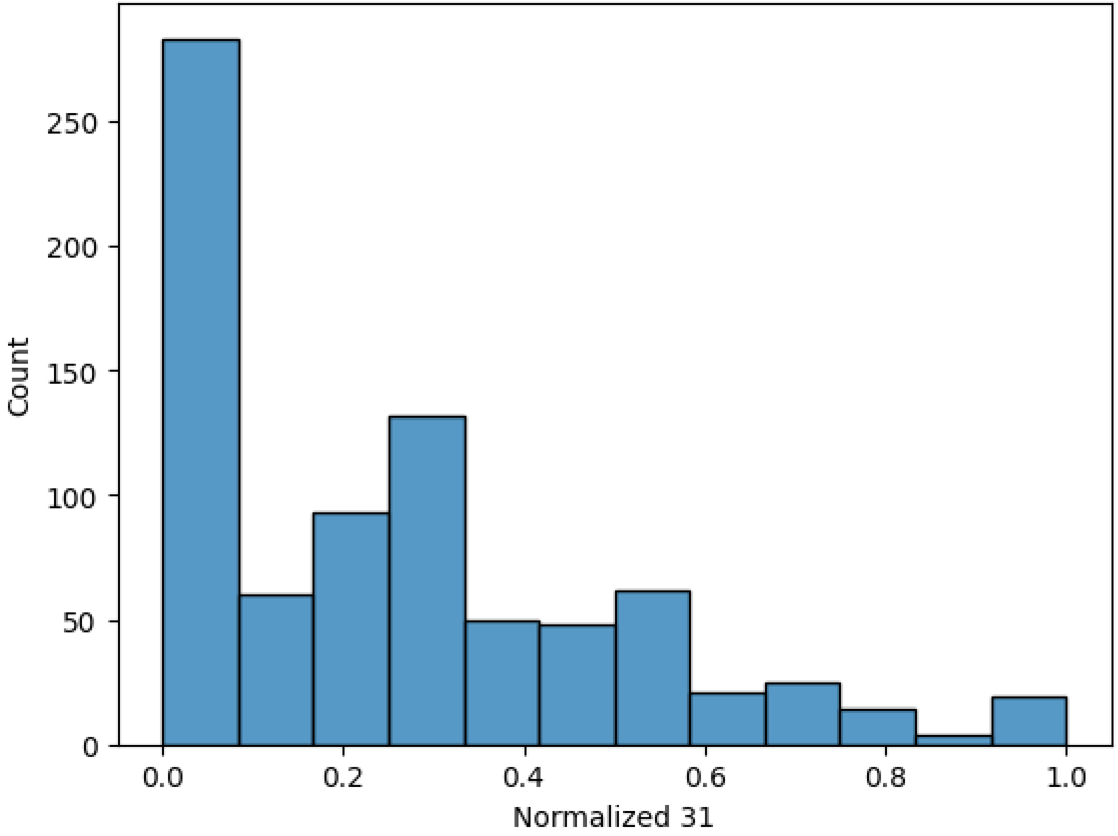

Count
300
250
200
150
100
50


0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 28

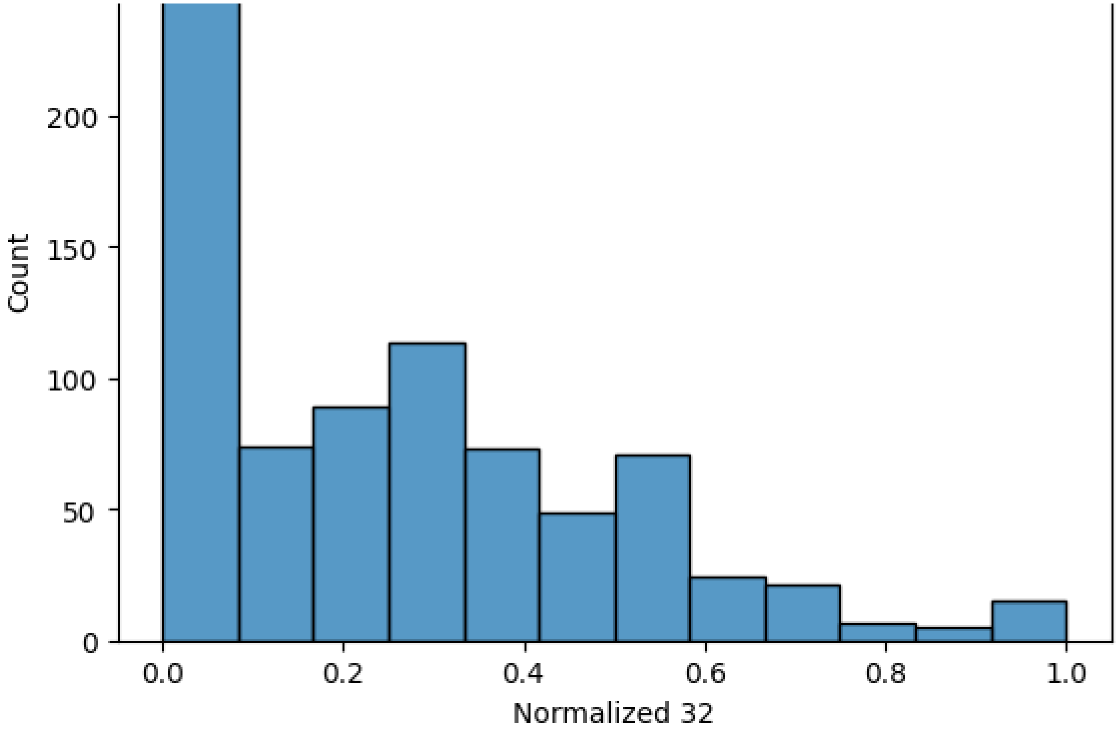

300
250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 29

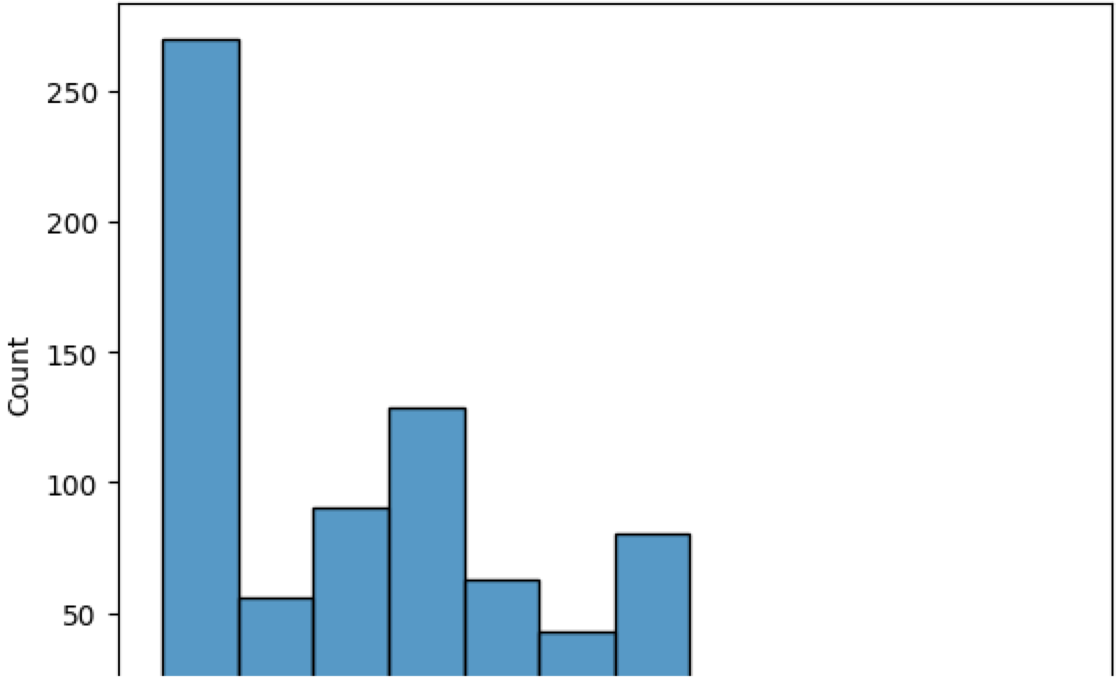

250
200
150
Count

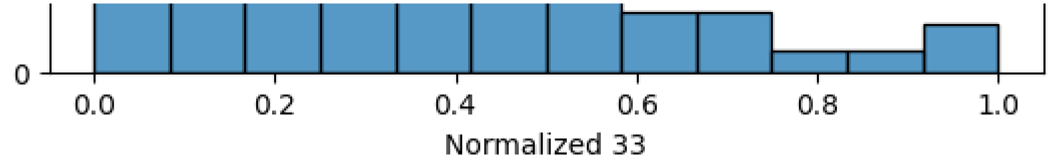

C
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 30

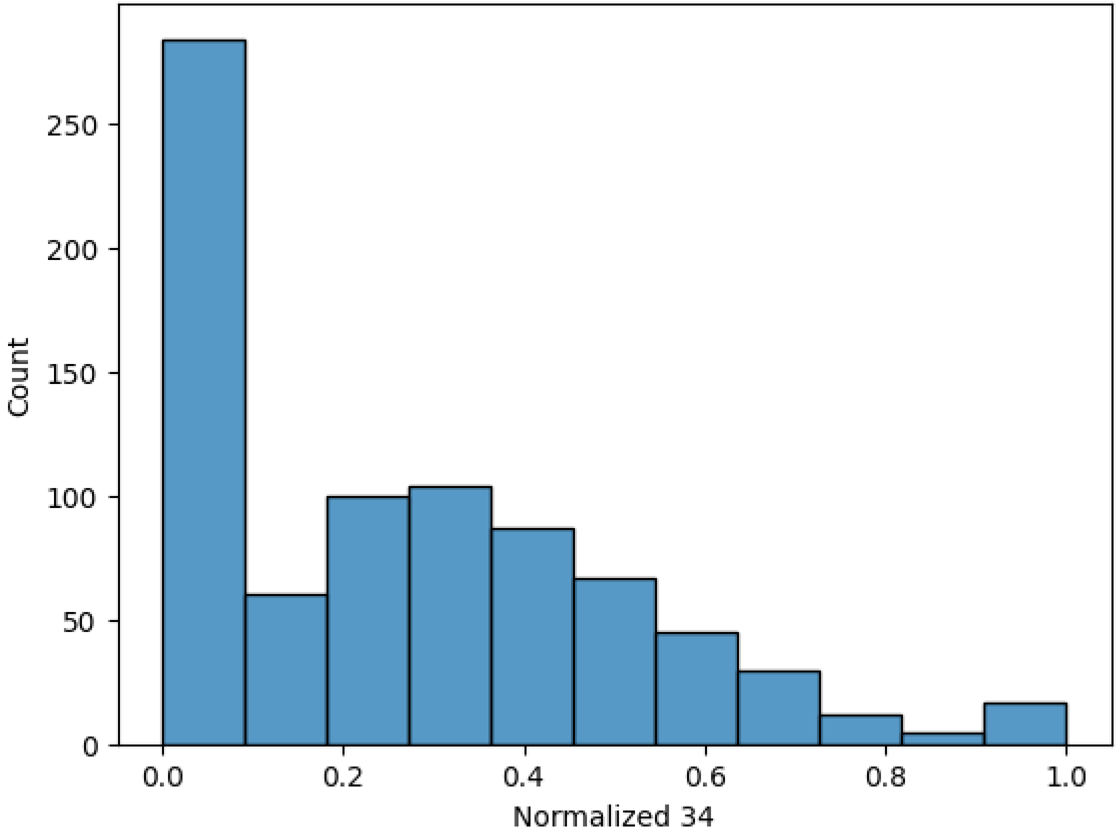

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 31

250

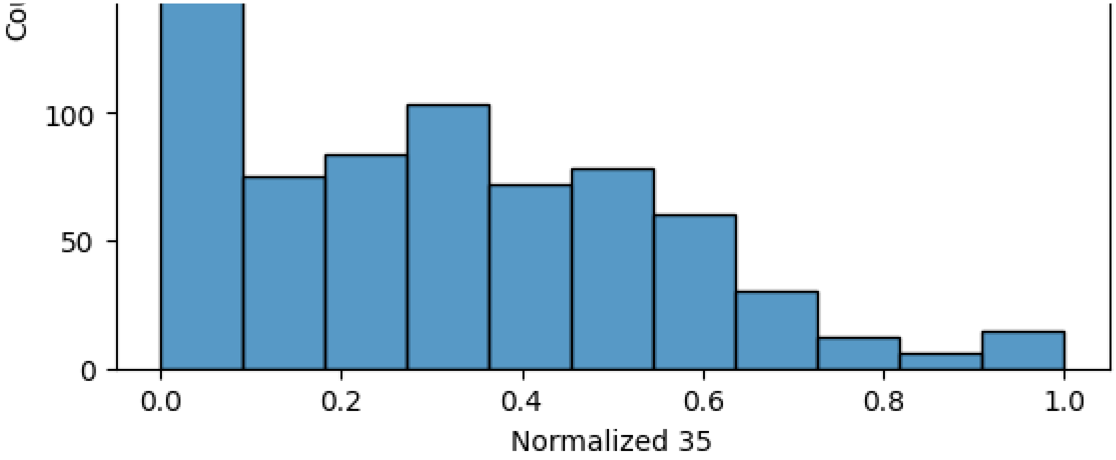

Count
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 32

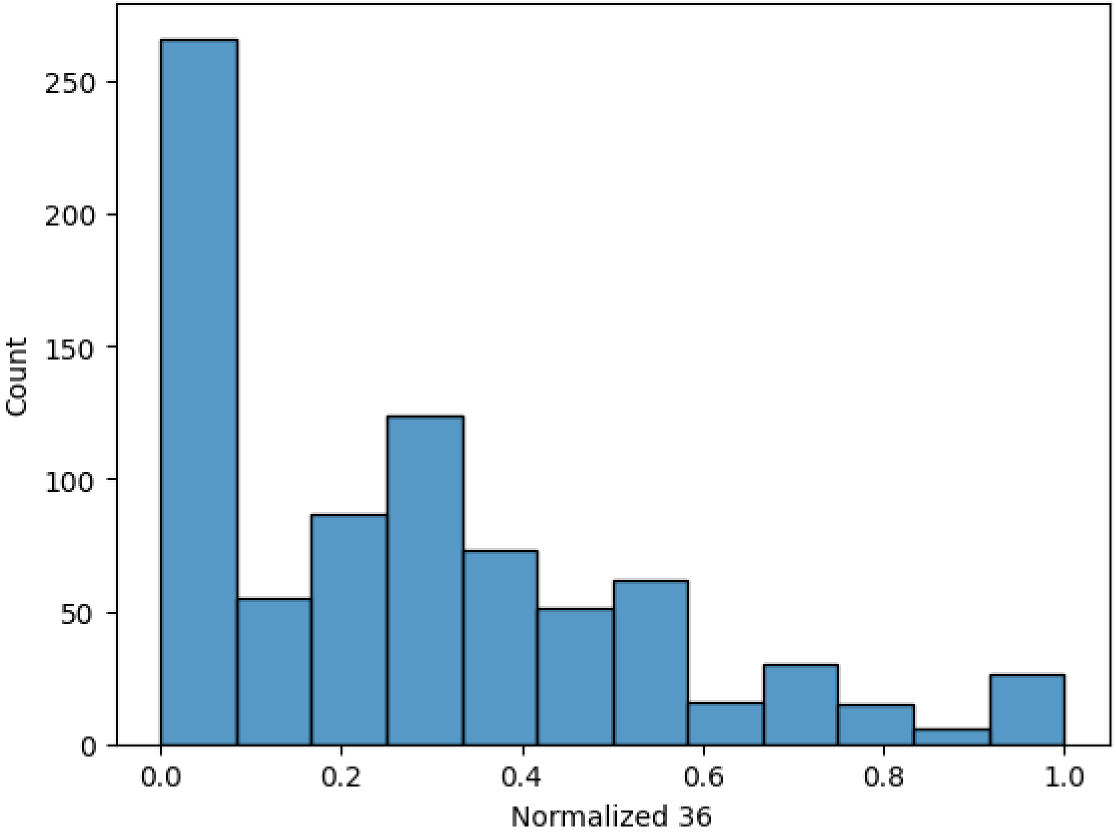

Count
50
100
150
200
250


0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 33

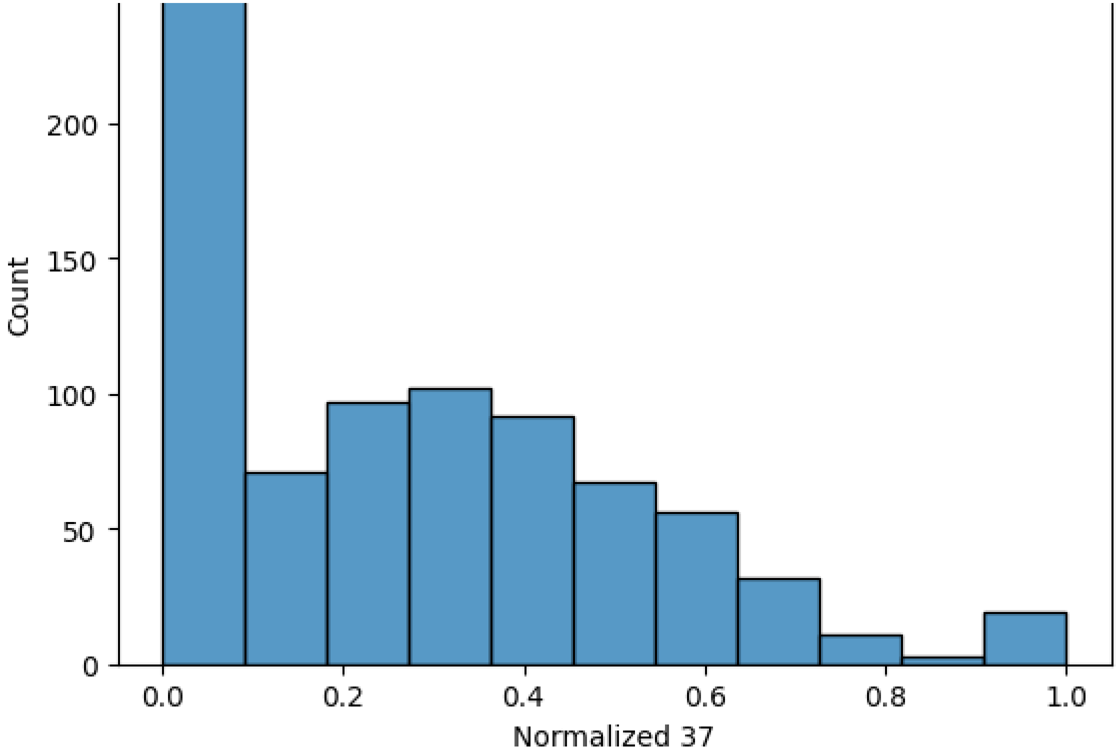

Count
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 34

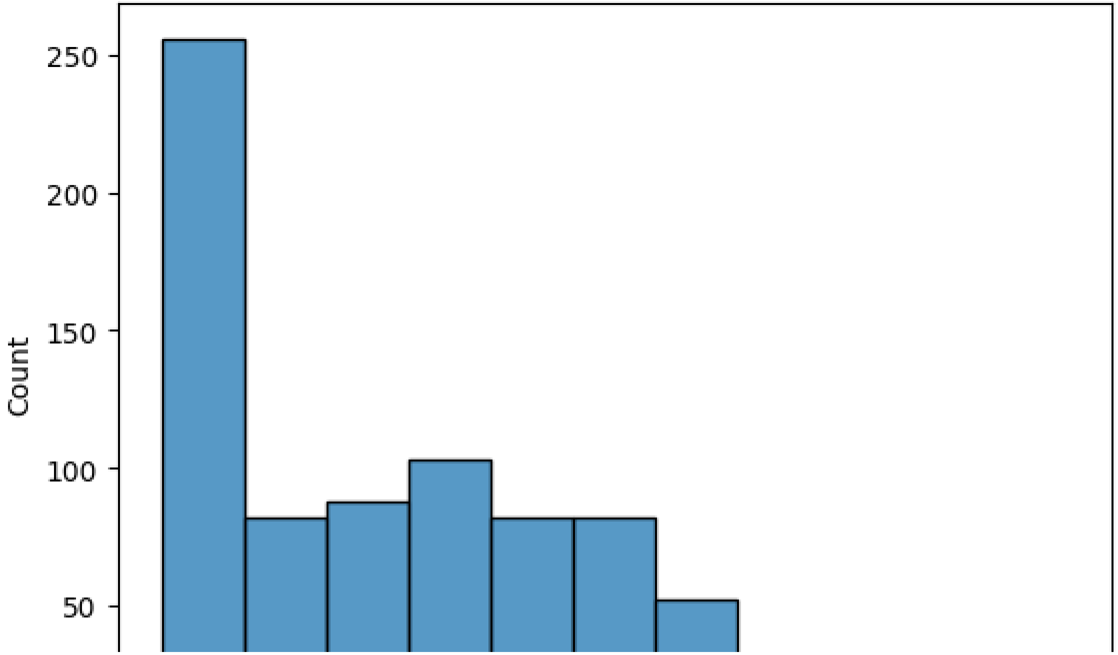
unt
150
200
250

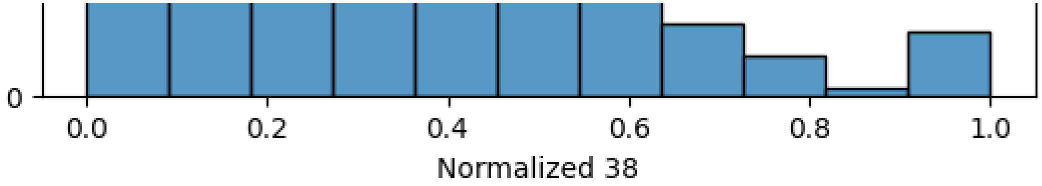

Co
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 35

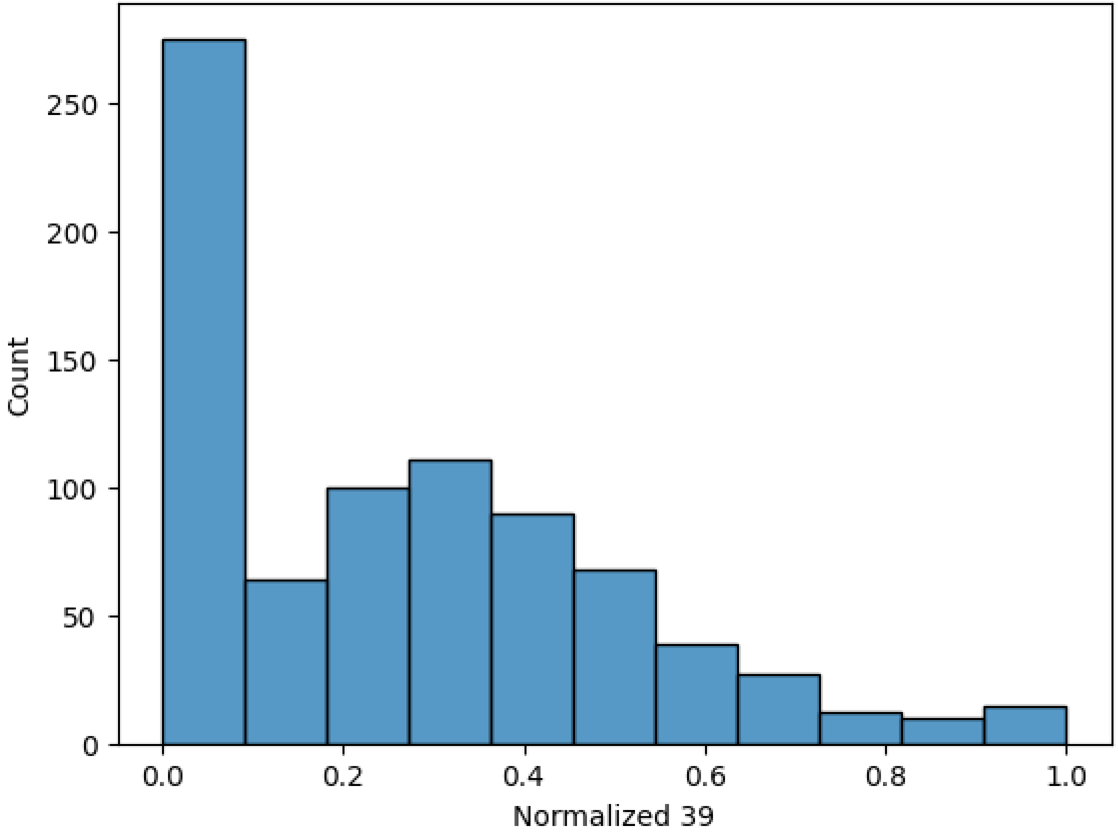

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 36

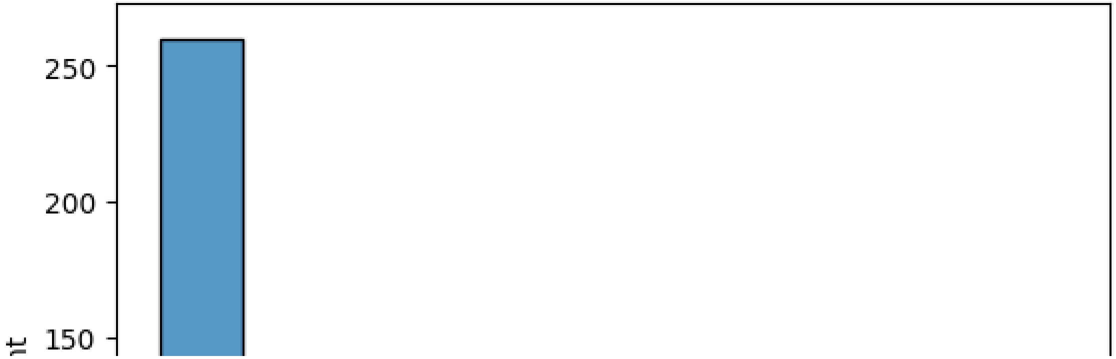
250

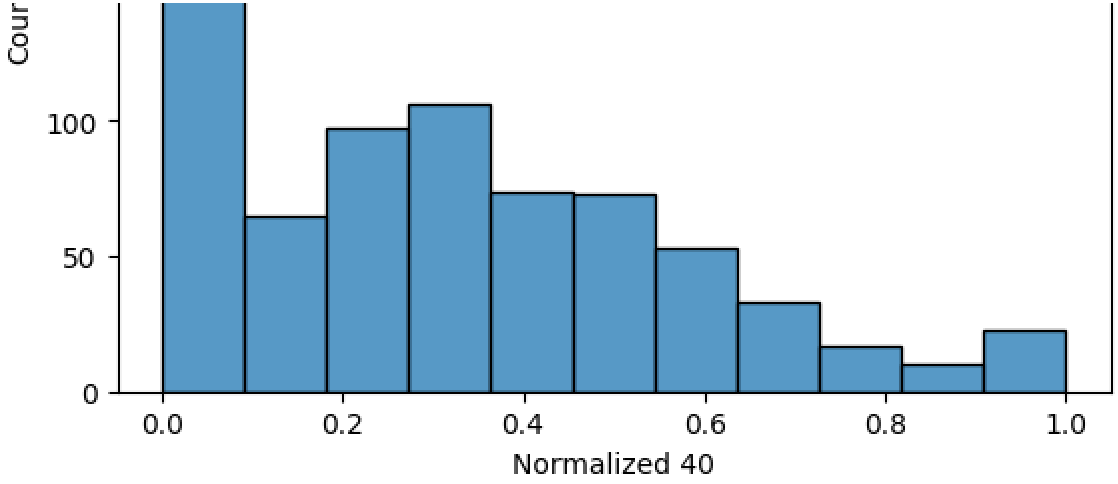

Count
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 37

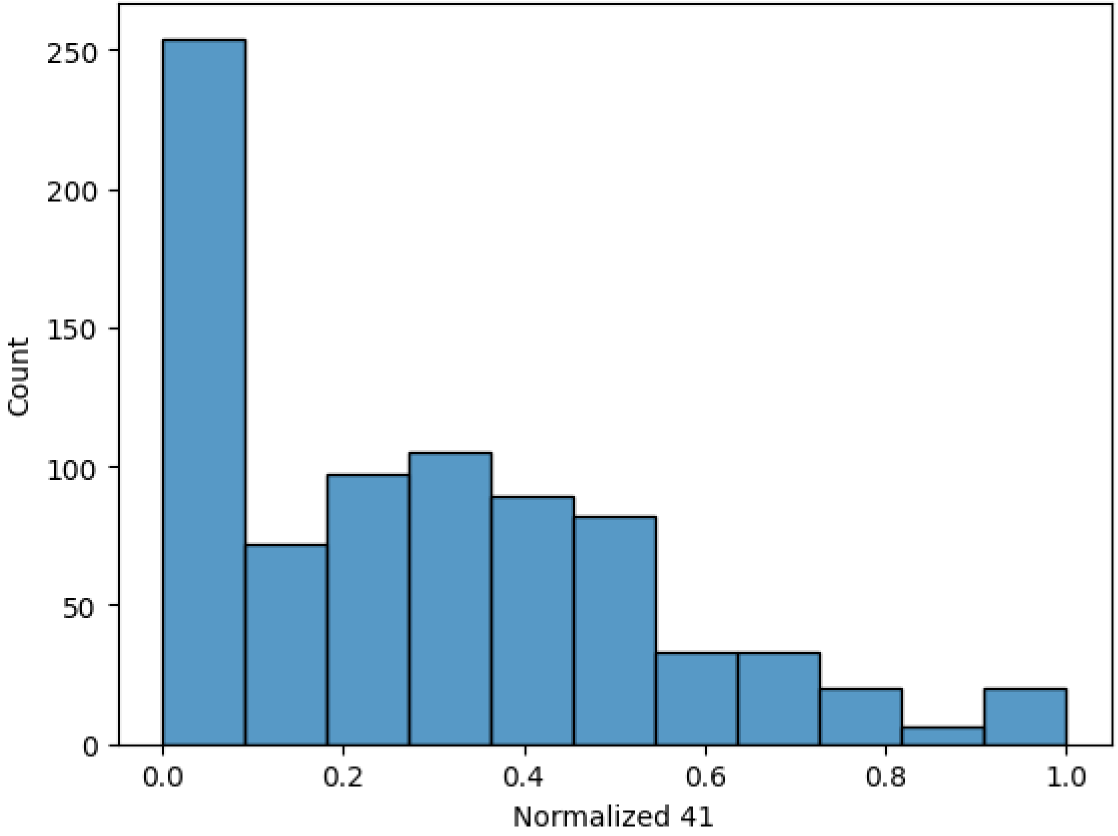

Count
250
200
150
100
50


0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 38

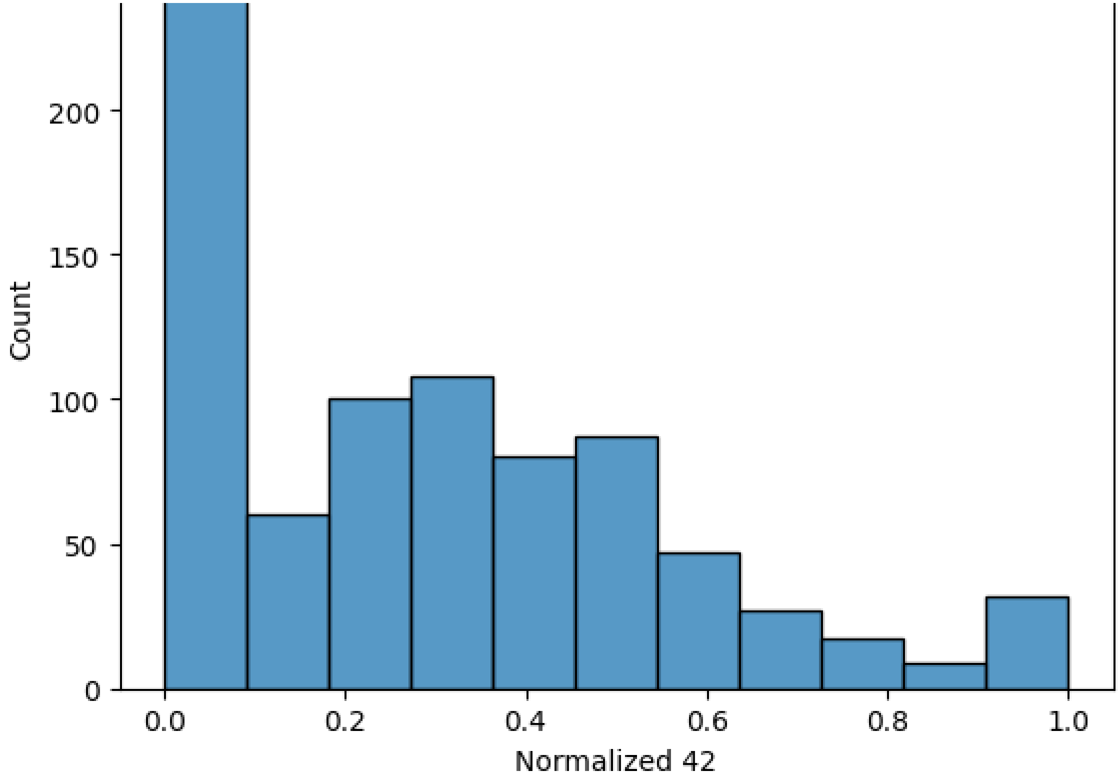

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 39

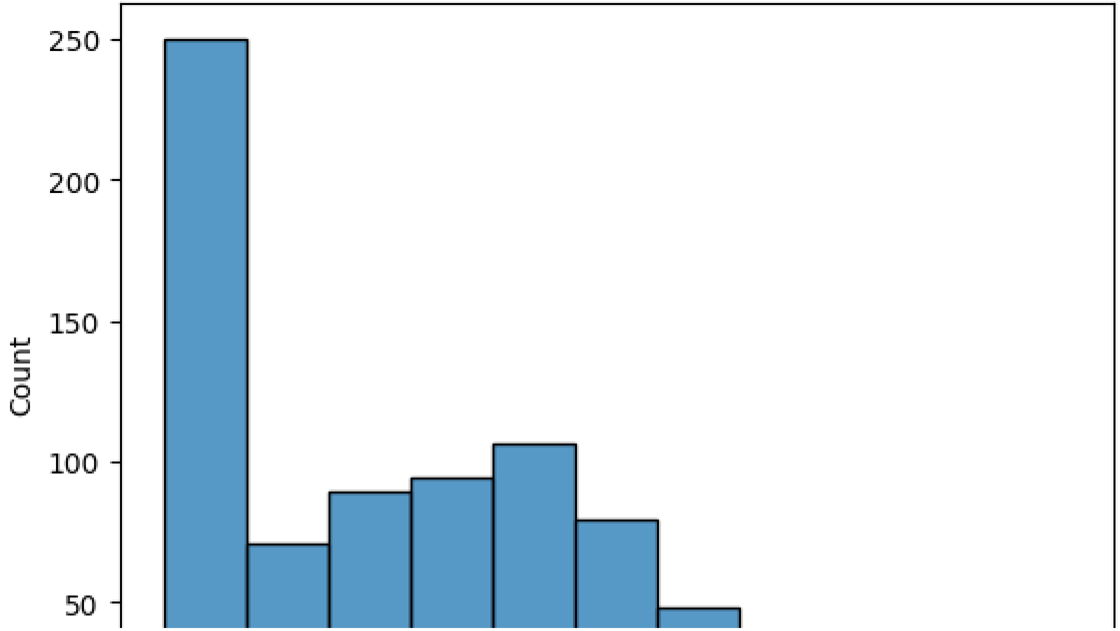

250
200
150
nt

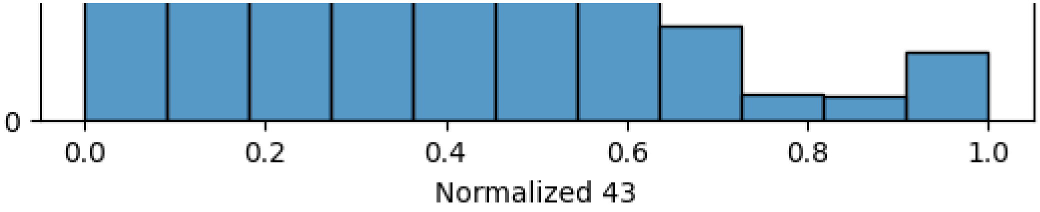

Cour
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 40

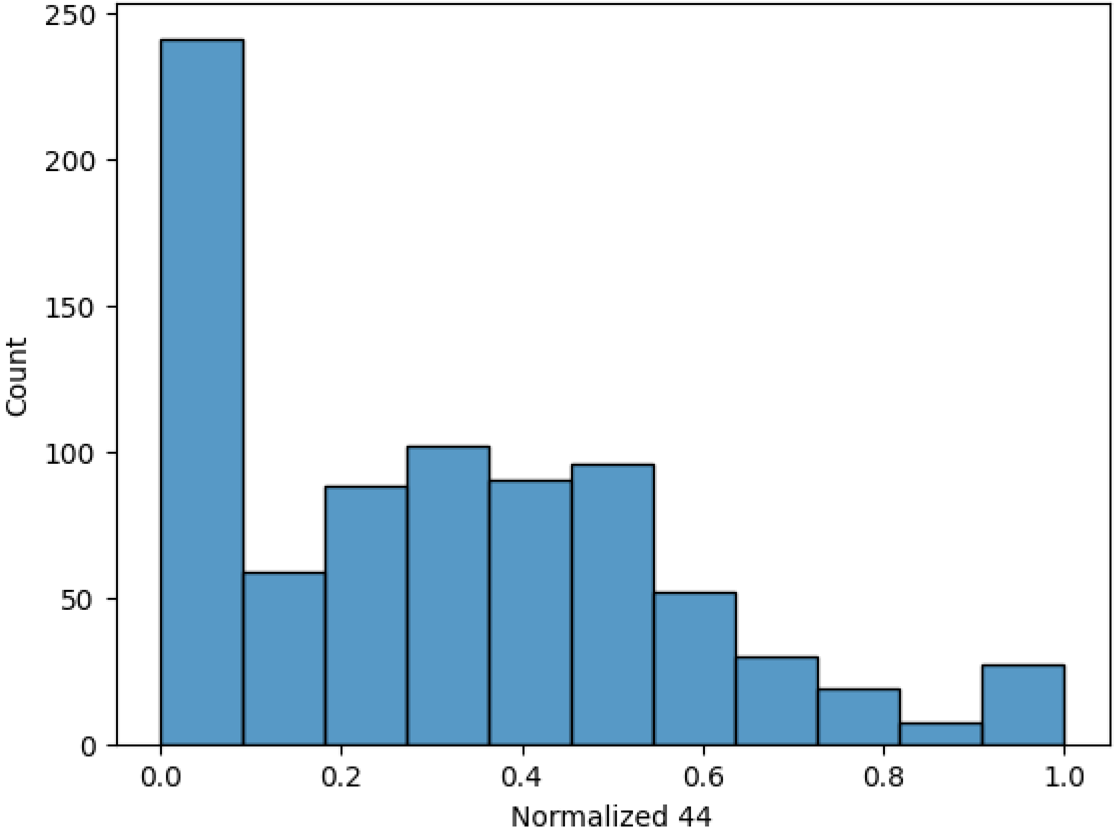

250
200
150
100
50
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 41

250

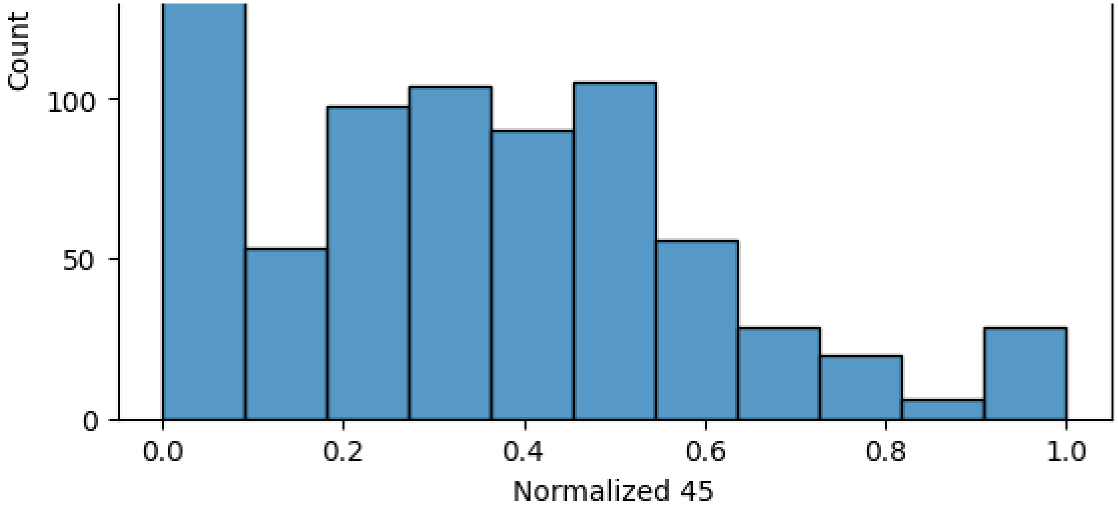

Count
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 42

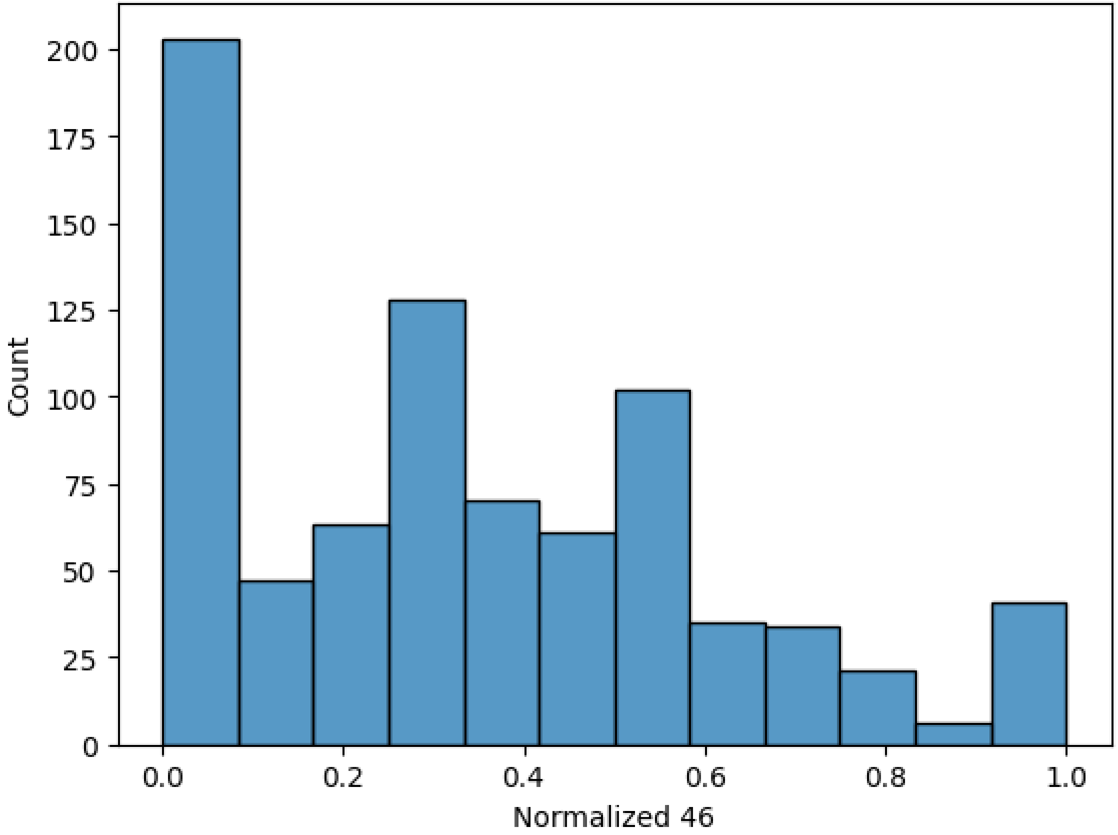

Count
250
200
150
100
50

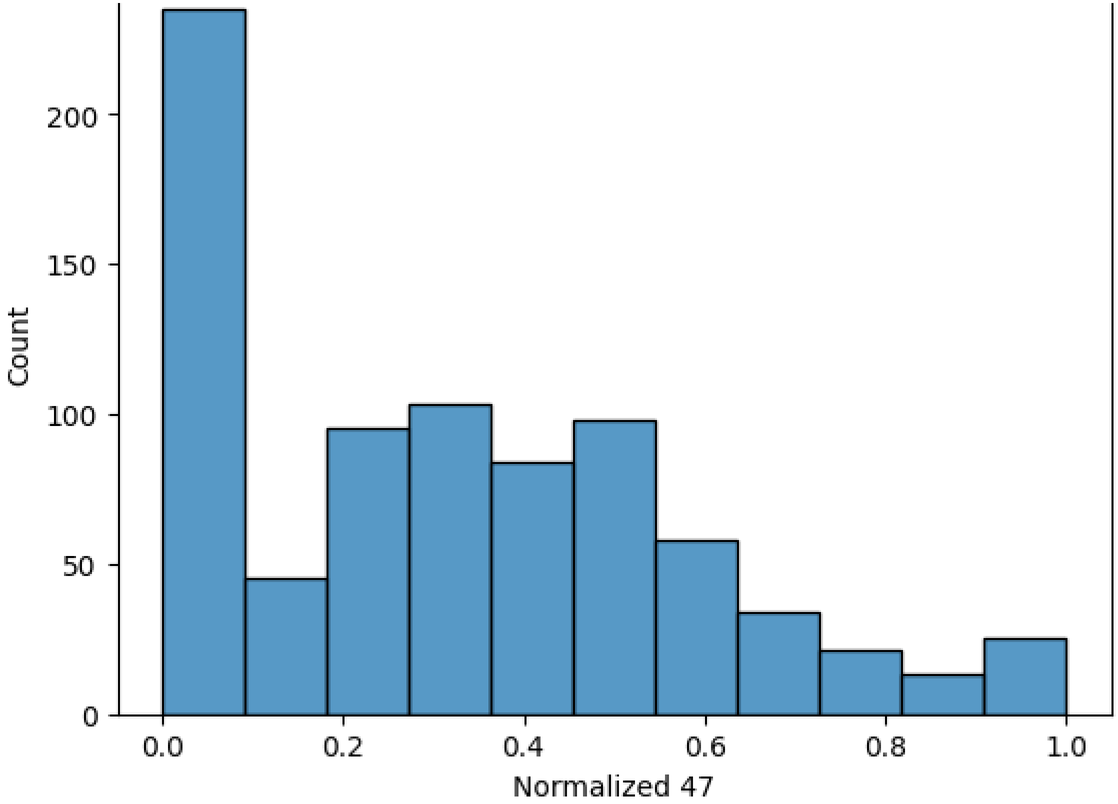

0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 43

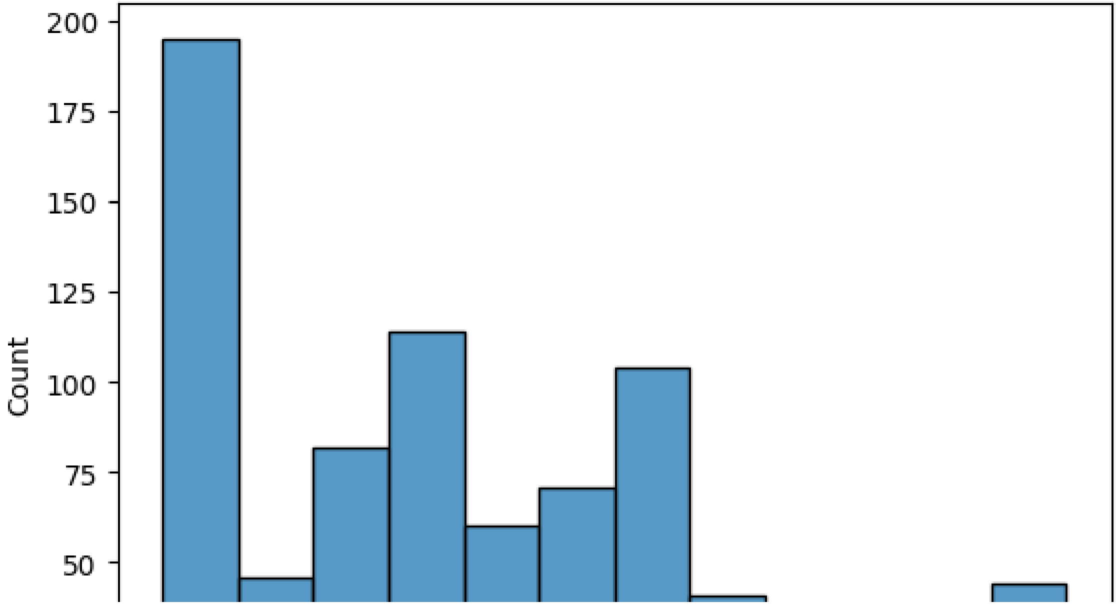

Count
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 44

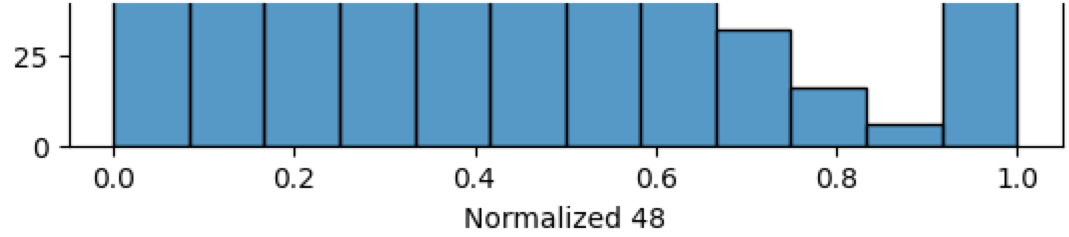
200
150

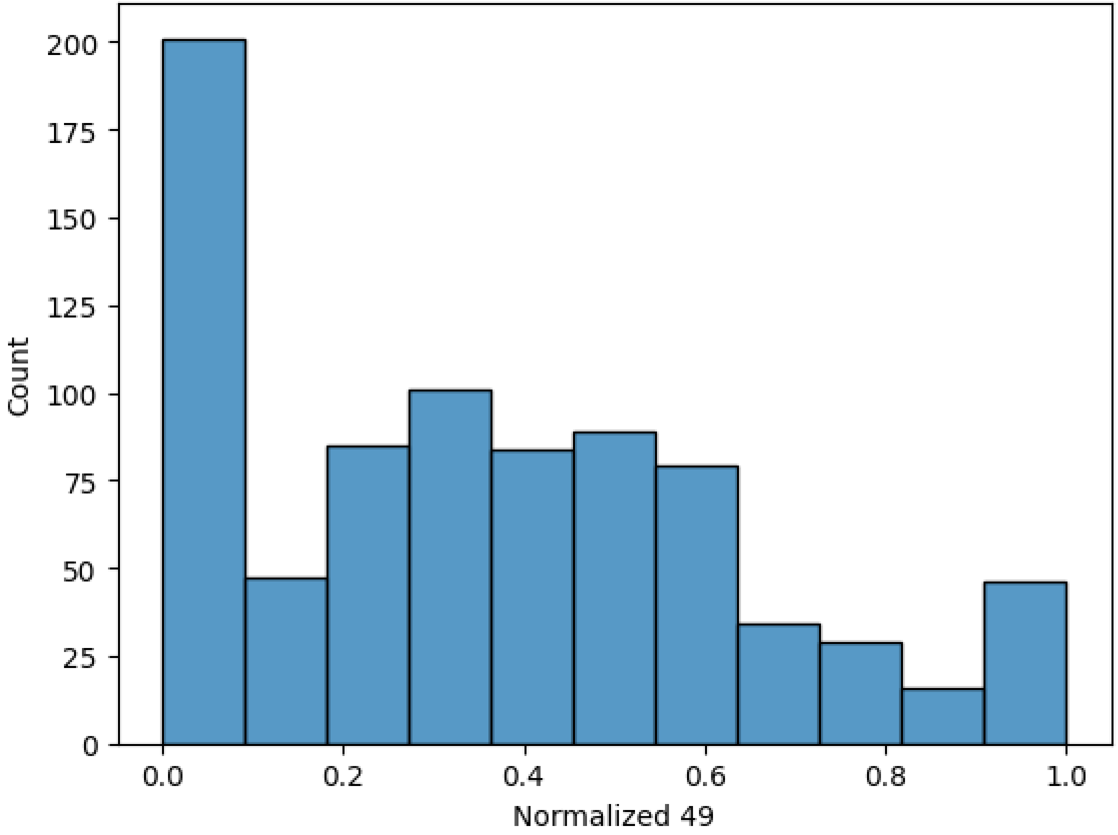

Count
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 45


200
175
150
125
100
75
50
25
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 46

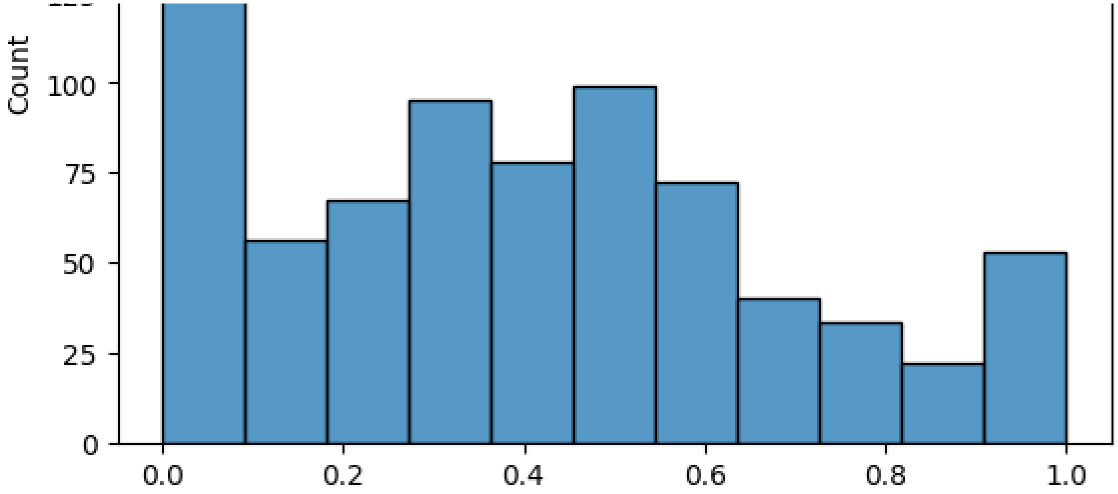

Count
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 47


Count
50
75
100
125
150
175
200


25
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 48


Count
200
175
150
125
100
75
50
25
0
0.0
0.2
0.4
0.6
0.8
1.0
Normalized 49

200
175
150
125



Removing Outliers:

```
def remove_outliers(data, threshold=3):
 z_scores=zscore(data)
 outlier= np.abs(z_scores) > threshold
 cleaned_data = data[~outlier]
 return cleaned_data
```

```
print("Columns after removing Outliers:")
remove_outliers(file.select_dtypes(include=['int','float']))
```

```
Columns after removing Outliers:
```

	W0	W1	W2	W3	W4	W5	W6	W7	W8	W9	...	Normalized 42	Normalized 43	Normalized 44	Normalized 45	Normalized 46	Norr
0	11.0	12.0	10.0	8.0	13.0	12.0	14.0	21.0	6.0	14.0	...	0.06	0.22	0.28	0.39	0.50	
1	7.0	6.0	3.0	2.0	7.0	1.0	6.0	3.0	3.0	3.0	...	0.20	0.40	0.50	0.10	0.10	
2	7.0	11.0	8.0	9.0	10.0	8.0	7.0	13.0	12.0	6.0	...	0.27	1.00	0.18	0.18	0.36	
3	12.0	8.0	13.0	5.0	9.0	6.0	9.0	13.0	13.0	11.0	...	0.41	0.47	0.06	0.12	0.24	
4	8.0	5.0	13.0	11.0	6.0	7.0	9.0	14.0	9.0	9.0	...	0.27	0.53	0.27	0.60	0.20	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
806	0.0	0.0	1.0	0.0	0.0	2.0	1.0	0.0	0.0	1.0	...	0.00	0.33	0.33	0.00	0.00	
807	0.0	1.0	0.0	0.0	1.0	2.0	2.0	6.0	0.0	1.0	...	0.43	0.43	0.57	0.29	0.57	

```
file.describe()
```

	W0	W1	W2	W3	W4	W5	W6	W7	W8	W9	...	N
count	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	811.000000	...	8

Scatter Plot:

I have choosen Week5 & Week 18 randomly:

```
plt.scatter(file['W5'],file['W18'],color='grey')
plt.xlabel('WEEK_5')
plt.ylabel('WEEK_18 ')
plt.title('SALES_TRANSACTION_DATA')
```

```
Text(0.5, 1.0, 'SALES_TRANSACTION_DATA')
```

K-Means Clustering:

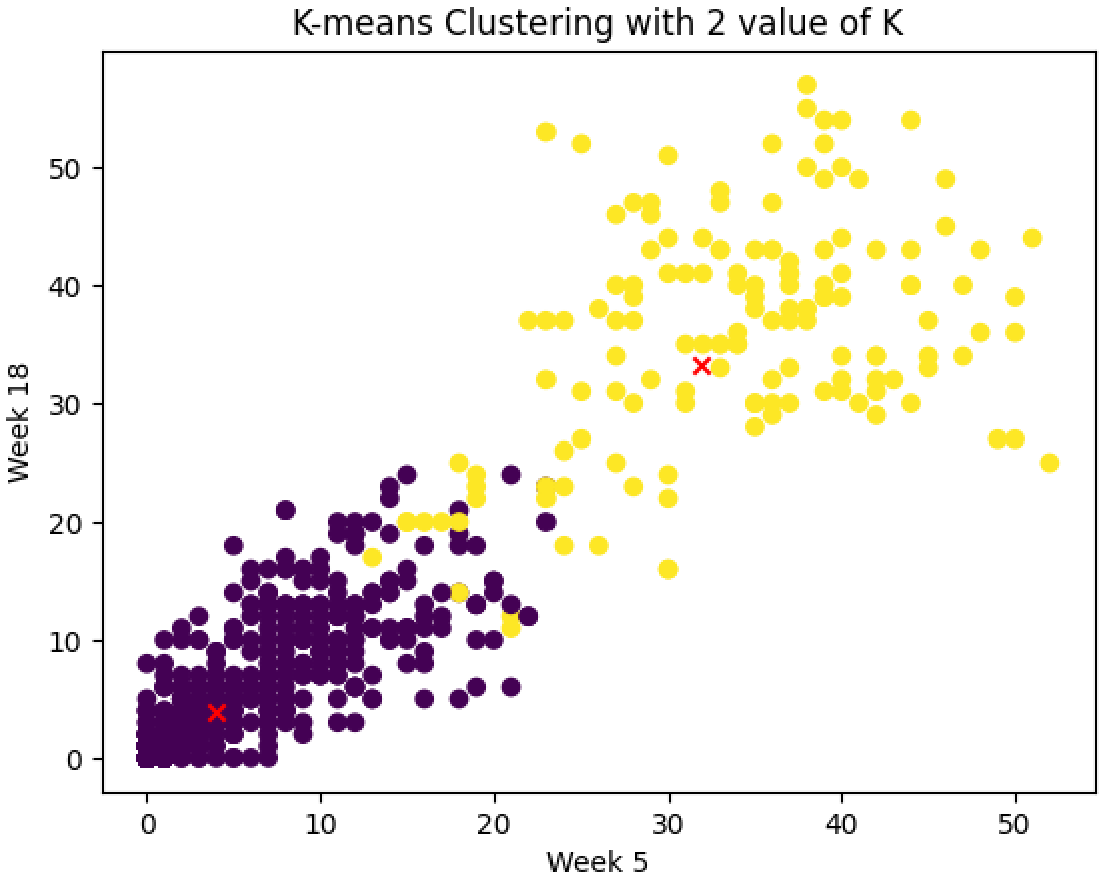
```
k = 3  # Number of clusters
kmeans = KMeans(n_clusters=k, random_state=42)
kmeans.fit(file.select_dtypes(exclude=['object']))

# Get the cluster labels
labels = kmeans.labels_

# Get the cluster centers
centers = kmeans.cluster_centers_

# Print the cluster labels and centers
for i in range(k):
    cluster_data = file[labels == i]
    cluster_center = centers[i]
    print(f"Cluster {i+1} - Center: {cluster_center}, Size: {len(cluster_data)}")
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will cha
  warnings.warn(
Cluster 1 - Center: [11.93401015 11.72588832 11.88324873 12.02030457 12.35532995 12.49746193
 12.23350254 12.08629442 12.54822335 12.37055838 12.36548223 12.97461929
 12.5177665  12.26395939 12.44670051 12.82233503 12.90862944 13.06598985
 12.6751269  12.6142132  12.75126904 12.83756345 13.37055838 13.43654822
 15.3857868  14.39086294  8.71573604  9.13705584  9.03045685  8.97461929
  9.35532995  9.5177665   9.66497462  9.71573604 10.14213198 10.21319797
  9.69543147 10.2284264  10.13705584  9.80203046  9.89847716 10.25380711
 10.1928934  10.29441624 10.47715736 10.42639594 10.57360406 10.99492386
 11.30964467 11.83756345 12.93908629 14.28426396  4.02030457 21.94416244
  0.45126904  0.43862944  0.45654822  0.45670051  0.47573604  0.48771574
  0.47756345  0.46725888  0.48532995  0.47538071  0.48385787  0.51451777
  0.48756345  0.46700508  0.48446701  0.50177665  0.50411168  0.51172589
  0.49040609  0.48020305  0.48700508  0.49279188  0.51649746  0.51791878
  0.61324873  0.56548223  0.27817259  0.30263959  0.29989848  0.28472081
  0.31142132  0.32005076  0.33187817  0.32649746  0.35441624  0.36238579
  0.32563452  0.3571066   0.35299492  0.3335533   0.33604061  0.35279188
  0.34812183  0.35807107  0.36335025  0.36304569  0.37213198  0.39598985
  0.40416244  0.43548223  0.48604061  0.55568528], Size: 197
Cluster 2 - Center: [33.70967742 35.19354839 36.74193548 38.37096774 36.70967742 35.94354839
 37.74193548 36.81451613 38.01612903 37.32258065 38.27419355 36.43548387
 37.87096774 37.41935484 38.02419355 38.80645161 38.96774194 37.34677419
 38.28225806 36.64516129 37.75806452 36.04032258 36.32258065 35.04032258
 33.         25.80645161 26.43548387 27.36290323 28.78225806 28.50806452
 29.66129032 28.95967742 30.17741935 30.44354839 30.44354839 30.29032258
 30.79032258 30.58064516 31.80645161 30.46774194 31.36290323 31.41129032
 32.09677419 31.73387097 31.78225806 32.5        33.00806452 32.35483871
 31.50806452 30.87903226 28.44354839 24.18548387 17.91935484 51.41129032
  0.47790323  0.52024194  0.56959677  0.61548387  0.56806452  0.54483871
  0.6016129   0.56774194  0.60379032  0.58637097  0.61290323  0.56508065
  0.59830645  0.58596774  0.605       0.62830645  0.63282258  0.58048387
  0.61024194  0.56322581  0.59209677  0.5433871   0.55169355  0.51169355
  0.44774194  0.23354839  0.25056452  0.28701613  0.32717742  0.31919355
  0.35483871  0.33064516  0.36685484  0.3766129   0.37895161  0.37709677
  0.38798387  0.38169355  0.41596774  0.37669355  0.40370968  0.40354839
  0.42516129  0.41556452  0.41790323  0.43903226  0.4508871   0.43008065
  0.40532258  0.38806452  0.31274194  0.18475806], Size: 124
Cluster 3 - Center: [1.40612245 1.48979592 1.46530612 1.54081633 1.58979592 1.54693878
 1.61836735 1.68979592 1.52857143 1.60612245 1.54693878 1.58163265
 1.64897959 1.63265306 1.77142857 1.65102041 1.55510204 1.68979592
 1.77959184 1.61836735 1.62040816 1.78979592 1.82244898 2.05306122
 2.28979592 2.40204082 1.31020408 1.30816327 1.30612245 1.3244898
 1.3244898  1.44081633 1.32040816 1.4755102  1.44693878 1.46530612
 1.5755102  1.5122449  1.53265306 1.53265306 1.62653061 1.56734694
 1.67346939 1.59795918 1.70408163 1.74489796 1.82857143 1.74285714
 1.83673469 2.14897959 2.26734694 2.84897959 0.10612245 5.15918367
 0.17661224 0.18704082 0.17997959 0.19002041 0.20604082 0.19471429
 0.20665306 0.20897959 0.18889796 0.20144898 0.19842857 0.19873469
 0.21344898 0.21277551 0.24095918 0.20997959 0.19316327 0.21879592
 0.23595918 0.21093878 0.21755102 0.24953061 0.25853061 0.27571429
 0.33146939 0.35840816 0.16834694 0.17485714 0.17665306 0.17610204
 0.18304082 0.20208163 0.18473469 0.20271429 0.19567347 0.20393878
 0.21806122 0.21191837 0.21210204 0.20916327 0.23671429 0.21846939
 0.24757143 0.22683673 0.25271429 0.26597959 0.29034694 0.25271429
 0.29461224 0.32073469 0.34281633 0.43812245], Size: 490
```

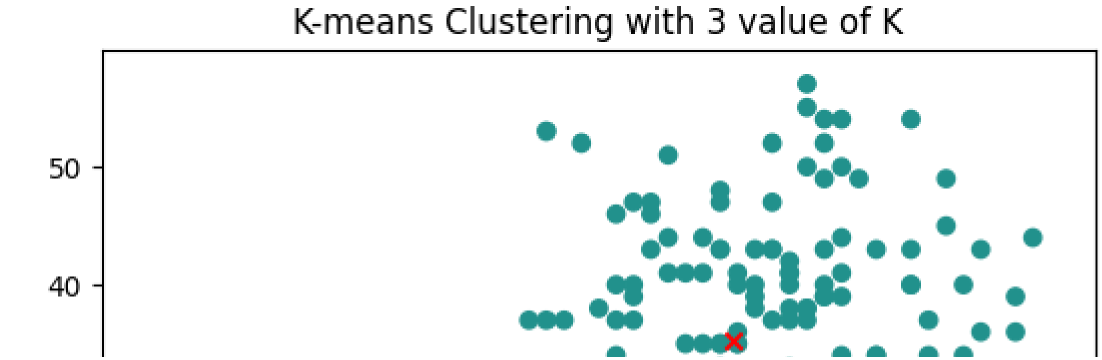
Visualization of K-Mean:

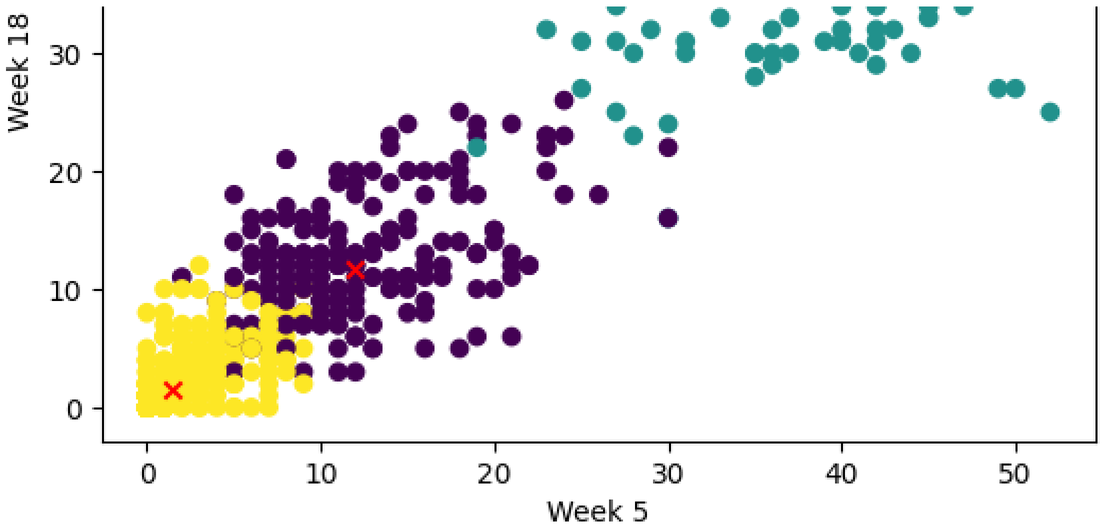
```
K = [2,3,4,5,6,7,8,9,10,20]  # Number of clusters
for i in K:
    kmeans = KMeans(n_clusters=i, random_state=42)
    kmeans.fit(file.select_dtypes(exclude=['object']))
    labels = kmeans.labels_
    centers = kmeans.cluster_centers_
    plt.scatter(file['W5'], file['W18'], c=labels, cmap='viridis')
    plt.scatter(centers[:, 0], centers[:, 1], c='red', marker='x')
    plt.xlabel('Week 5')
    plt.ylabel('Week 18')
    plt.title('K-means Clustering with '+ str(i)+ ' value of K')
    plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

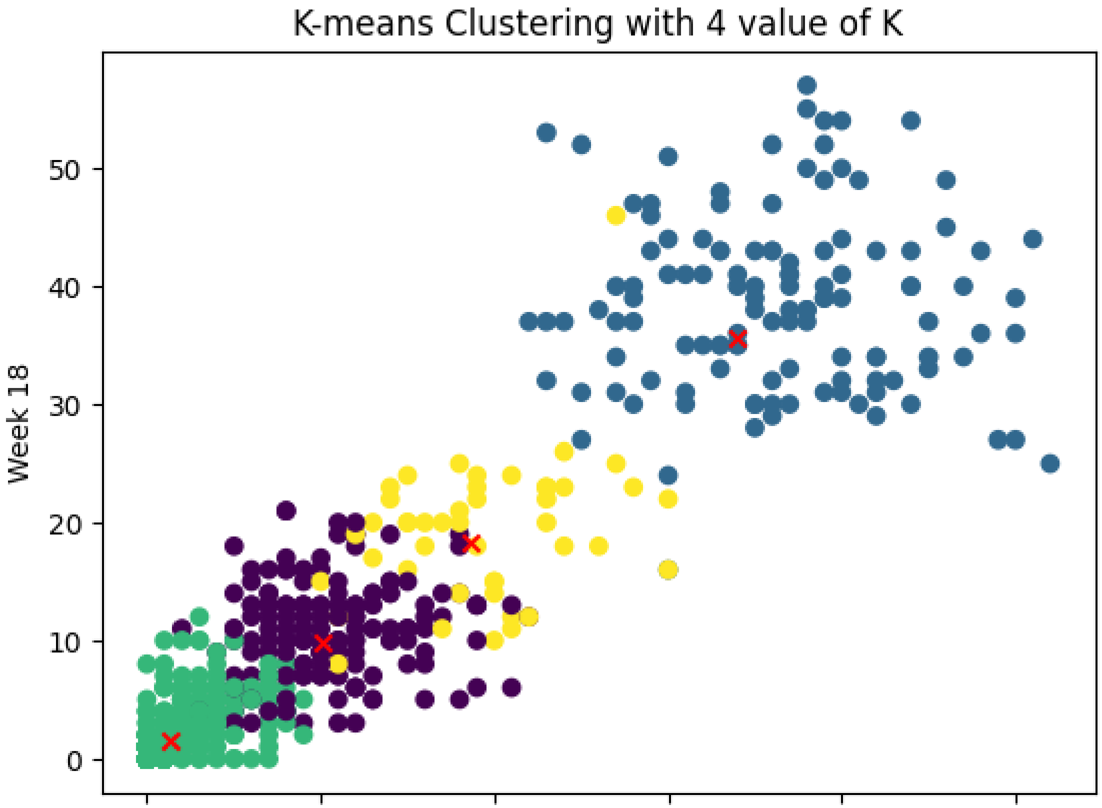


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

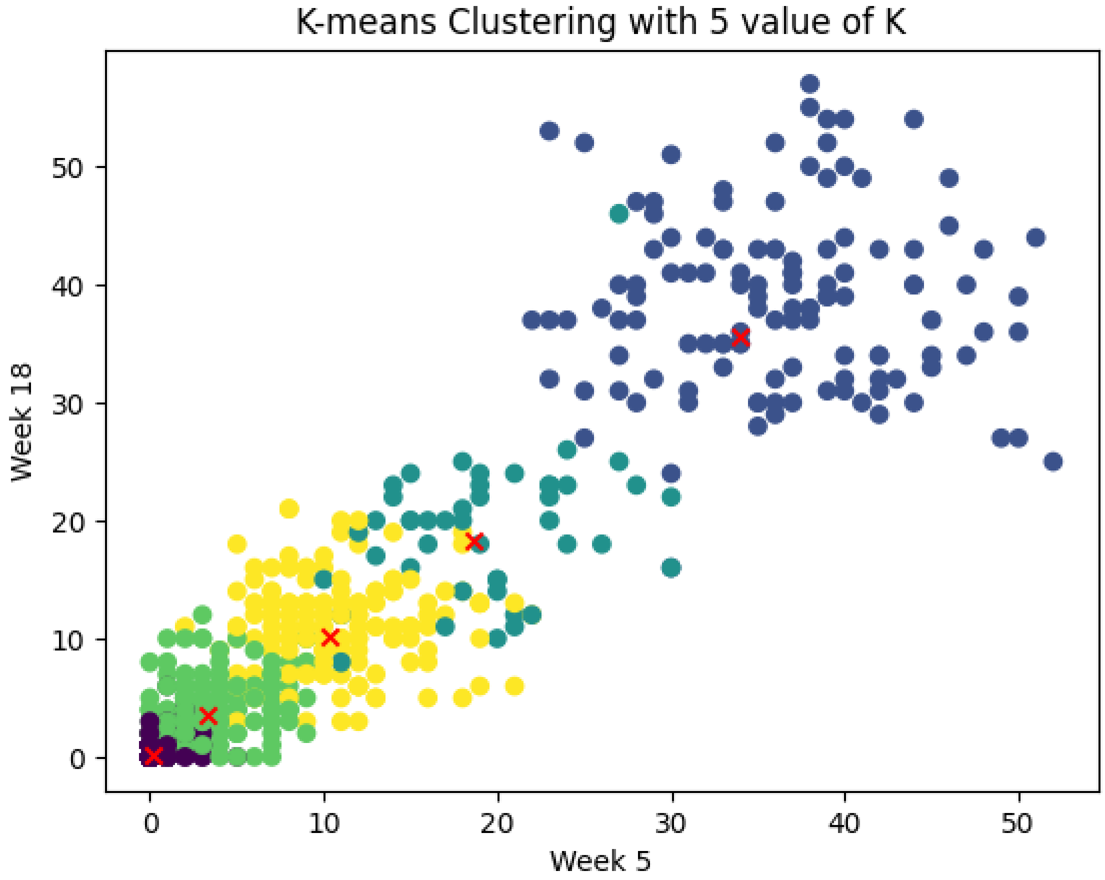




```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

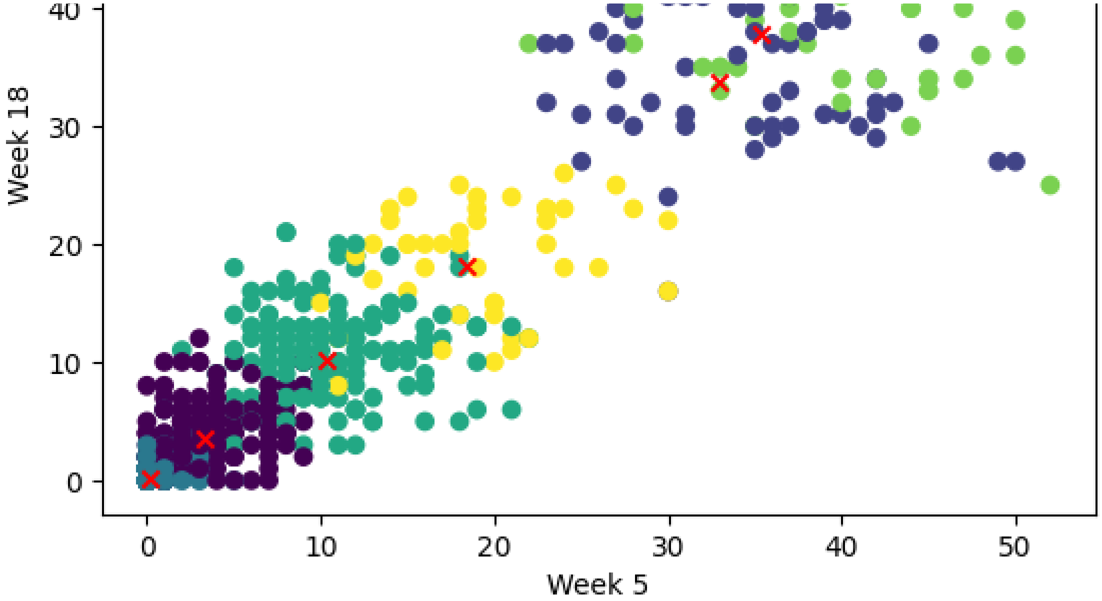


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

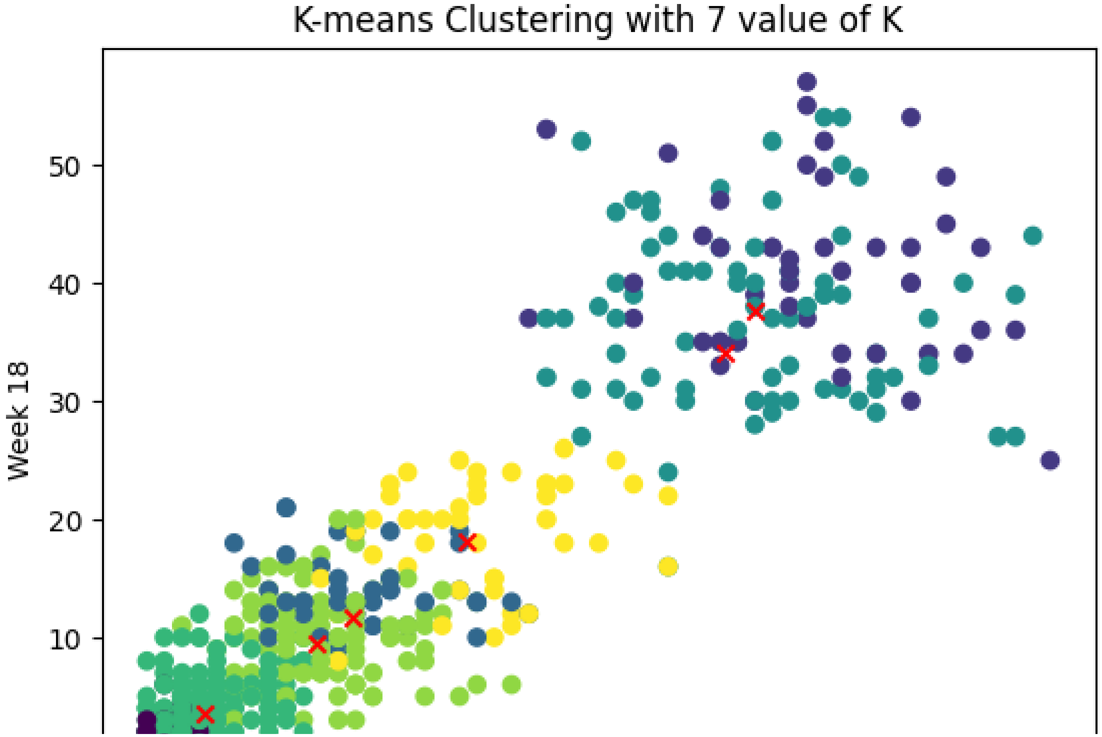


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

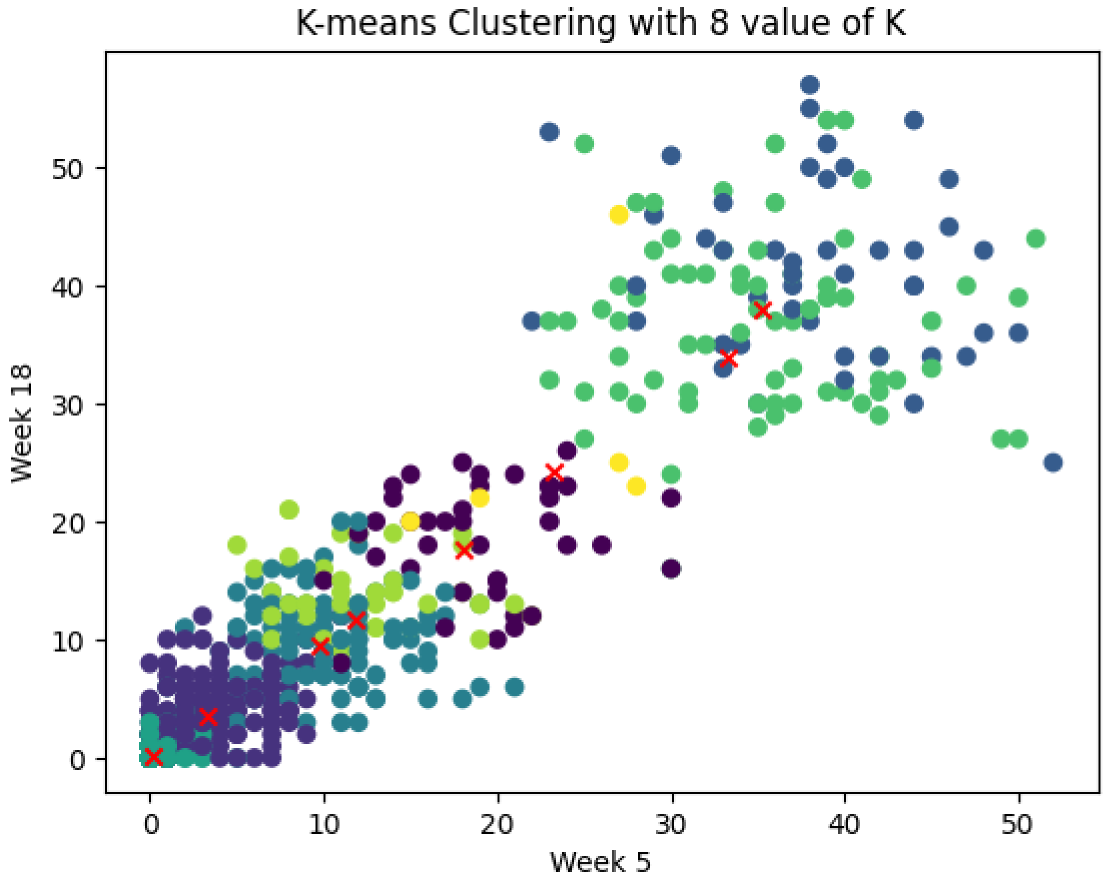
K-means Clustering with 6 value of K
50



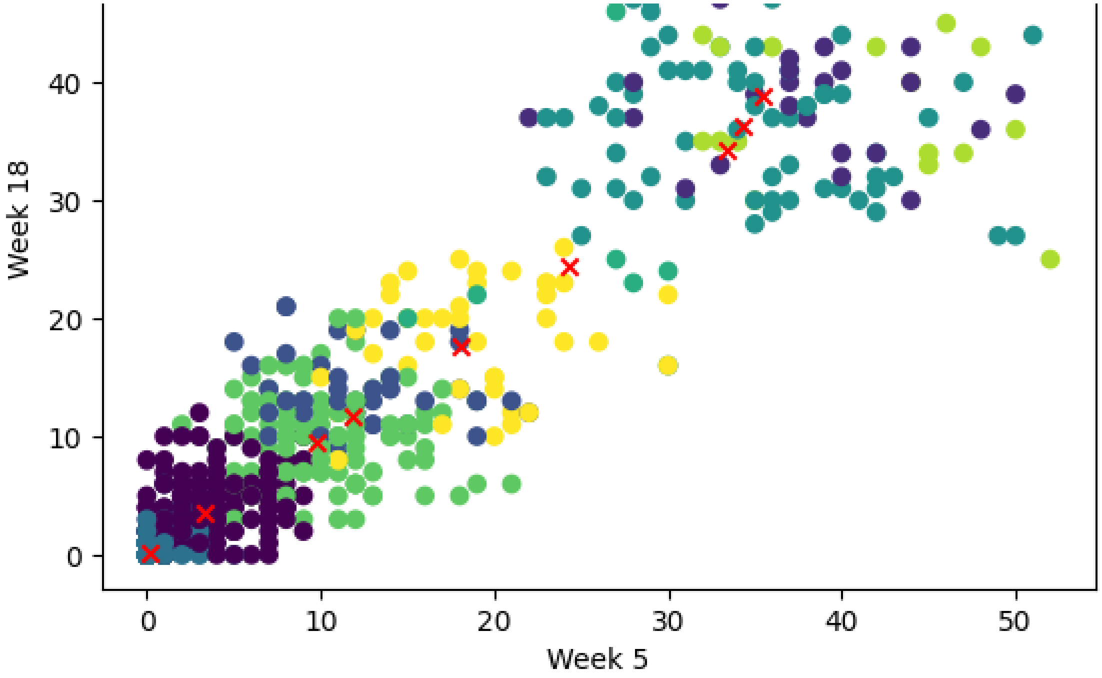
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



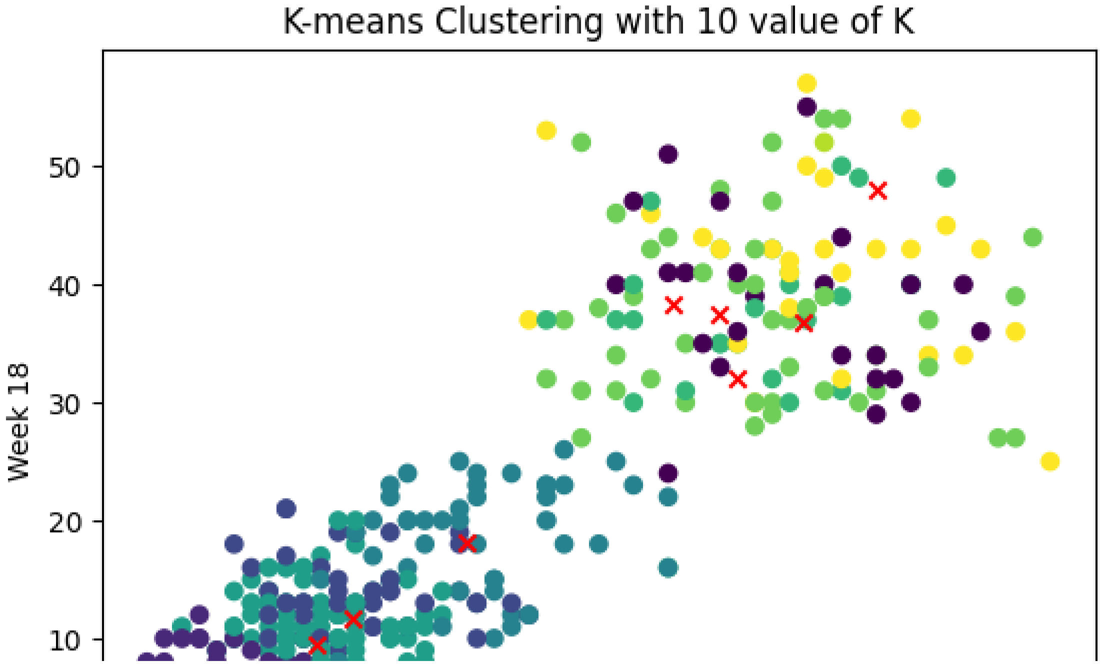
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

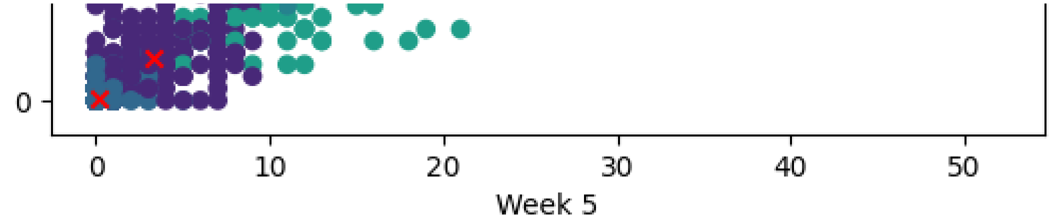


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

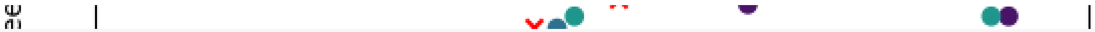




```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

K-means Clustering with 20 value of K

Liner Regression:

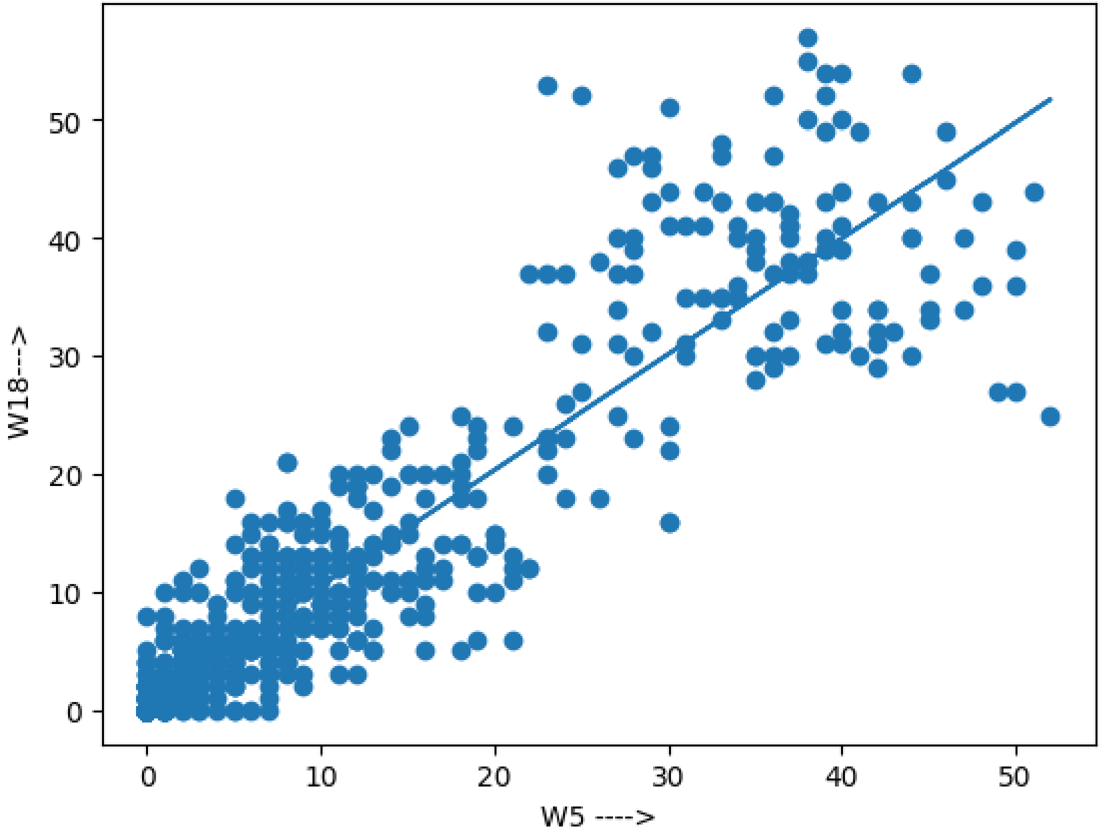


```
xn= np.array(file.W5)
yn= np.array(file.W18)
n=len(xn)
xn_mean = xn.mean()
yn_mean = yn.mean()
sum_xn = (xn).sum()
sum_yn = (yn).sum()
sum_xnyn = ((xn*yn).sum())*n
print(sum_xnyn)
sum_xn_sum_yn = sum_xn*sum_yn
sum_xnxn = ((xn*xn).sum())*n
print(sum_xnxn)
sum_xn_Square = sum_xn*sum_xn
slp = (sum_xnyn- sum_xn_sum_yn) / (sum_xnxn-sum_xn_Square)
print("Slope=" , slp)
```

```
Intcp = yn_mean - slp * xn_mean
print("Intercept=" ,Intcp)
```

```
168155984
166954893
Slope= 0.97991908131643
Intercept= 0.7313948369097005
```

```
plt.xlabel("W5 ---->")
plt.scatter(file.W5,file.W18)
plt.ylabel("W18---> ")
plt.plot(xn,slp*xn+Intcp)
plt.show()
```



Linear Regression General Form For Only One Independent Feature/Variable:

Y=mx+c.

where x=independent feature,Y=Predicted Y

m=slop , c=intercept

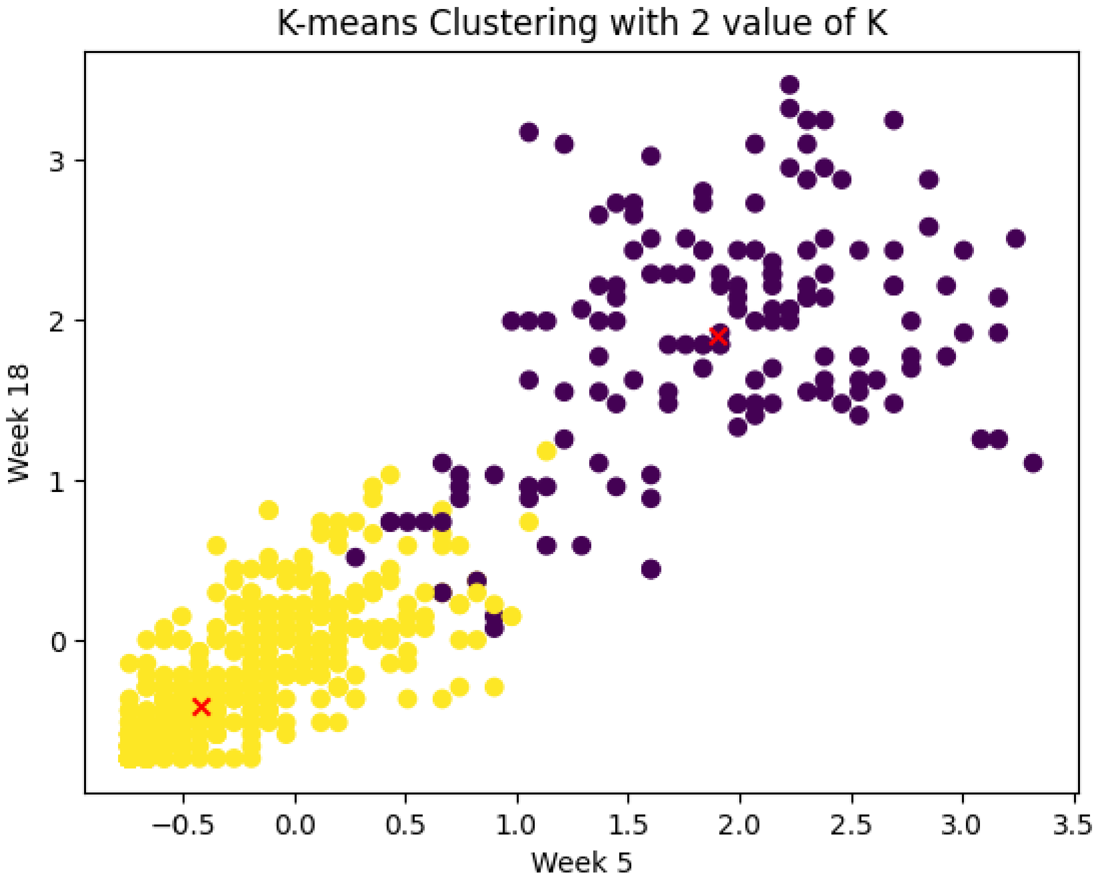
K-Means CLustering With Features Scaling:

```
scaler = StandardScaler()
scaled = pd.DataFrame(scaler.fit_transform(file.select_dtypes(exclude=['object'])))
scaled.columns=file.select_dtypes(exclude=['object']).columns
scaled.columns
```

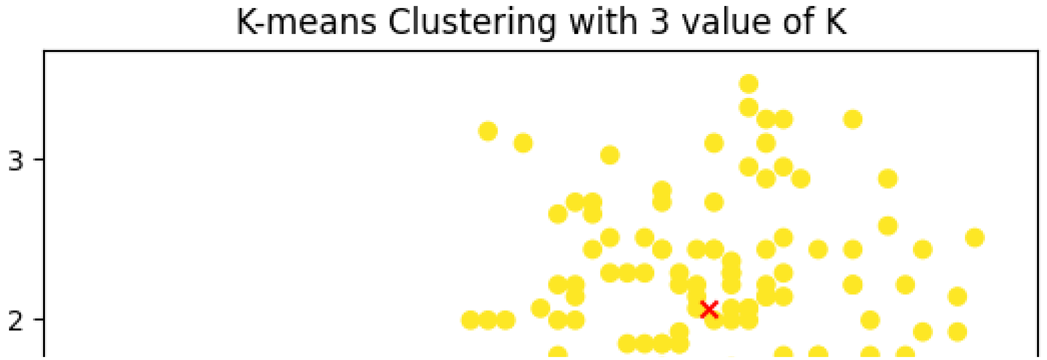
```
     Index(['W0', 'W1', 'W2', 'W3', 'W4', 'W5', 'W6', 'W7', 'W8', 'W9',
            ...
            'Normalized 42', 'Normalized 43', 'Normalized 44', 'Normalized 45',
            'Normalized 46', 'Normalized 47', 'Normalized 48', 'Normalized 49',
            'Normalized 50', 'Normalized 51'],
           dtype='object', length=106)
```

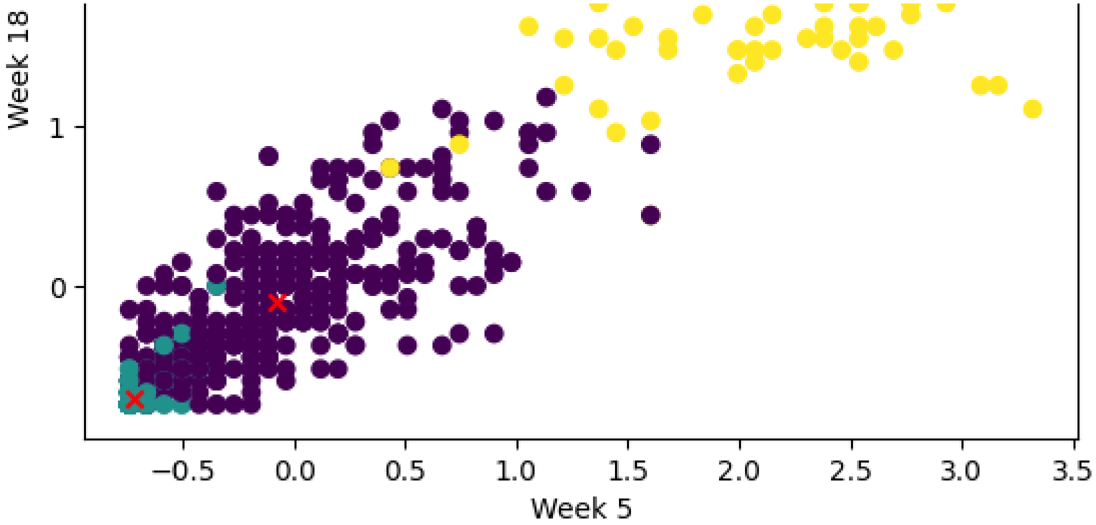
```
K = [2,3,4,5,6,7,8,9,10,20]  # Number of clusters
for i in K:
    kmeans = KMeans(n_clusters=i, random_state=42)
    kmeans.fit(scaled)
    labels = kmeans.labels_
    centers = kmeans.cluster_centers_
    plt.scatter(scaled['W5'], scaled['W18'], c=labels, cmap='viridis')
    plt.scatter(centers[:, 0], centers[:, 1], c='red', marker='x')
    plt.xlabel('Week 5')
    plt.ylabel('Week 18')
    plt.title('K-means Clustering with '+ str(i)+ ' value of K')
    plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



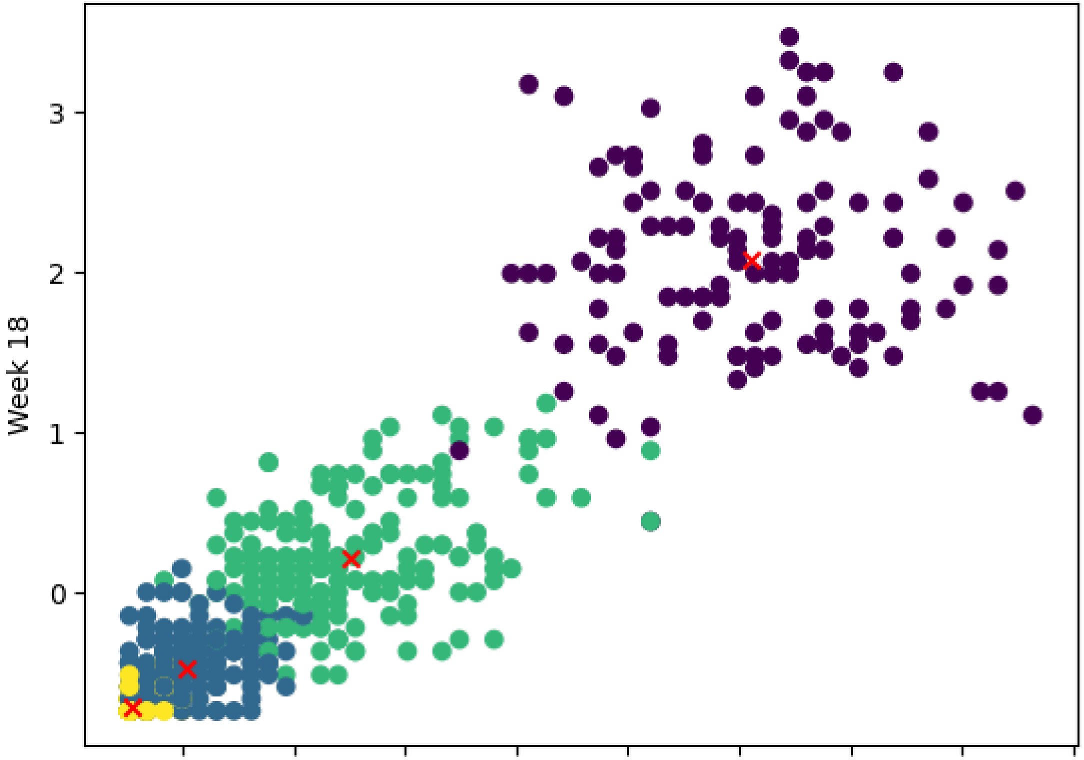
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



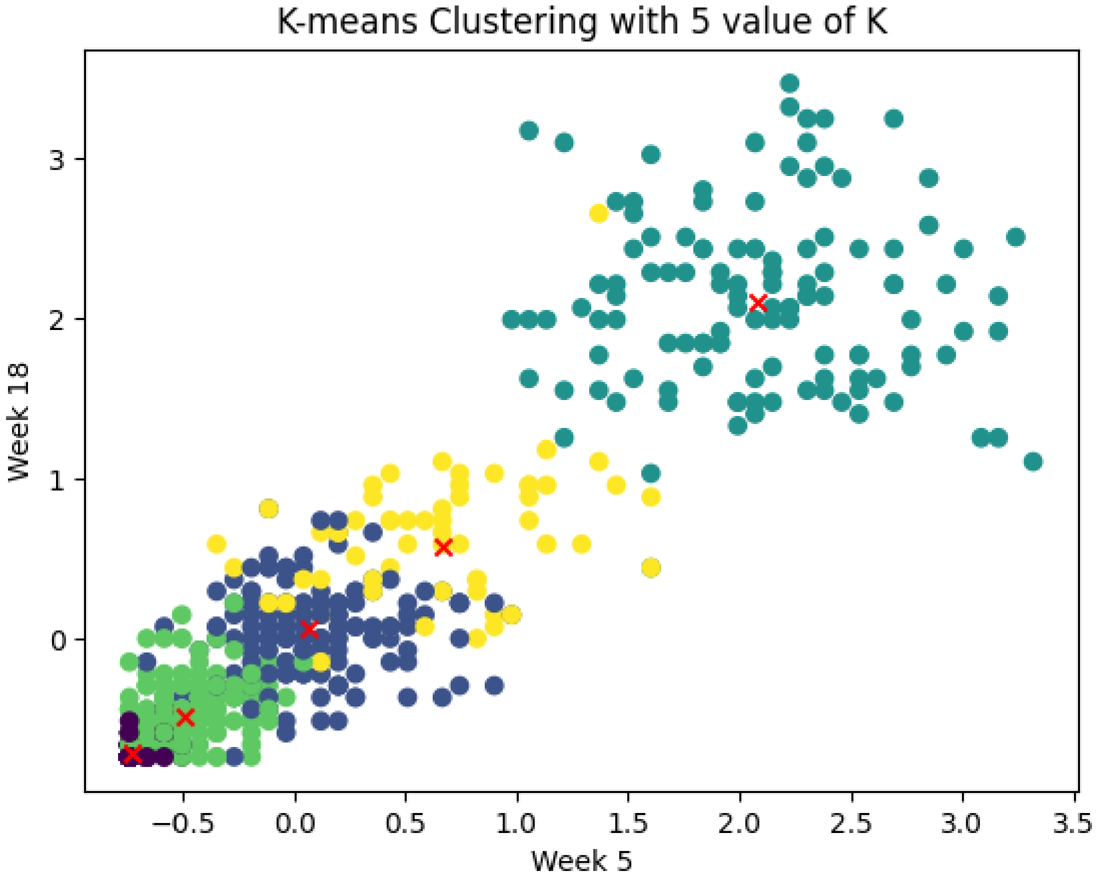


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

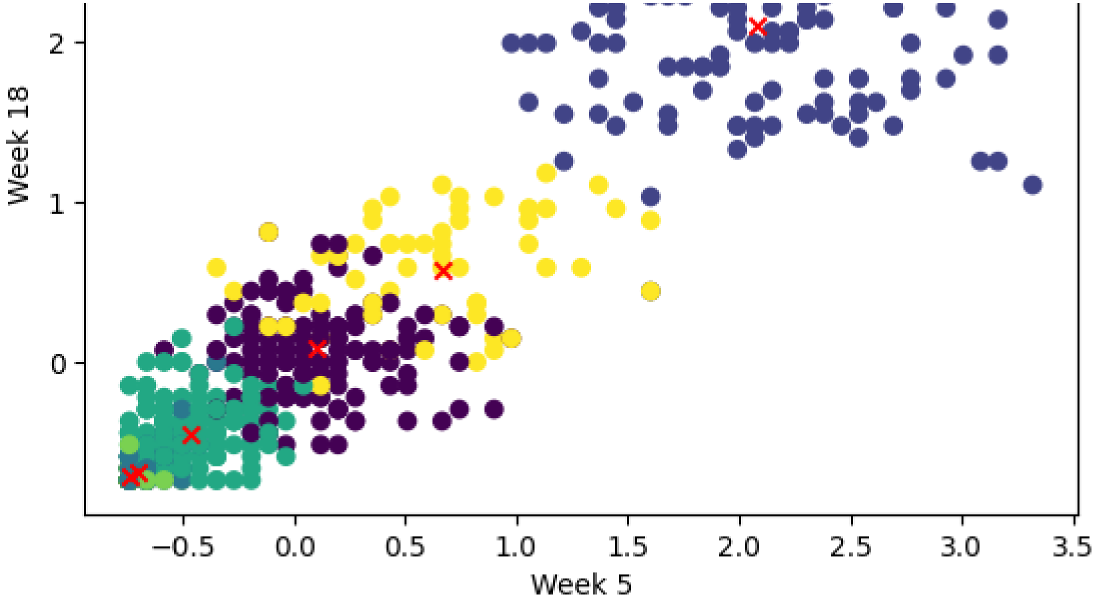
K-means Clustering with 4 value of K



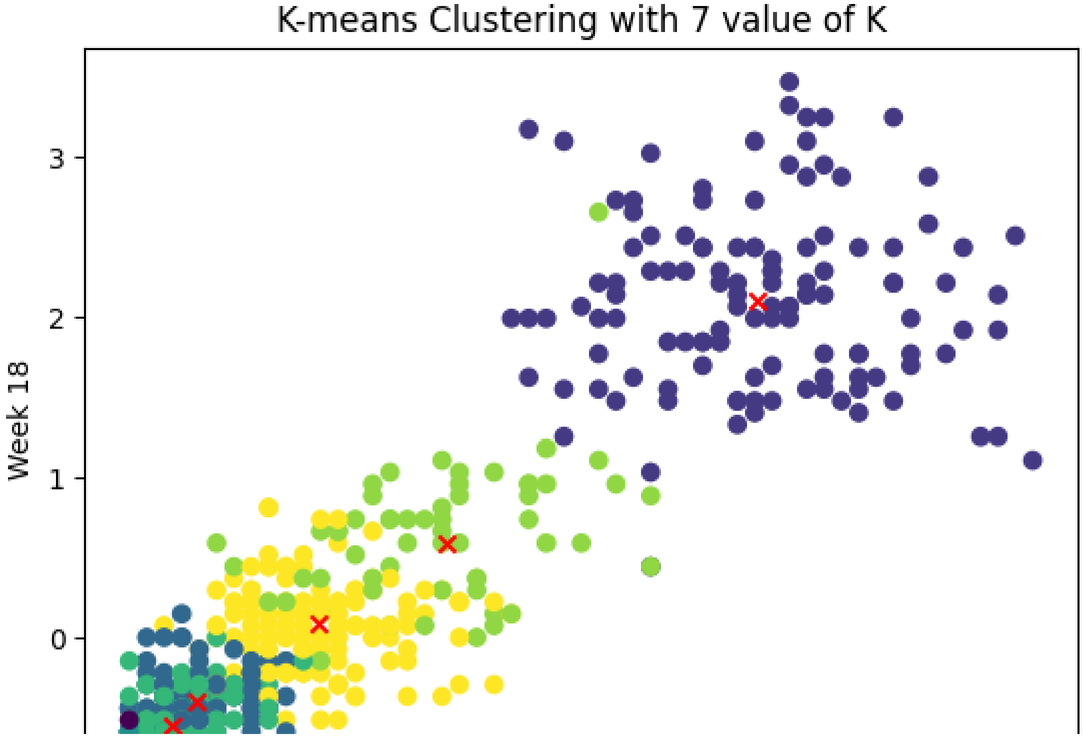
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

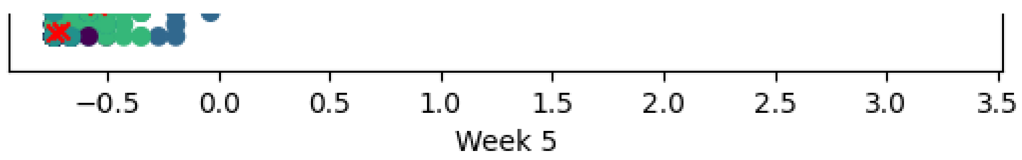


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

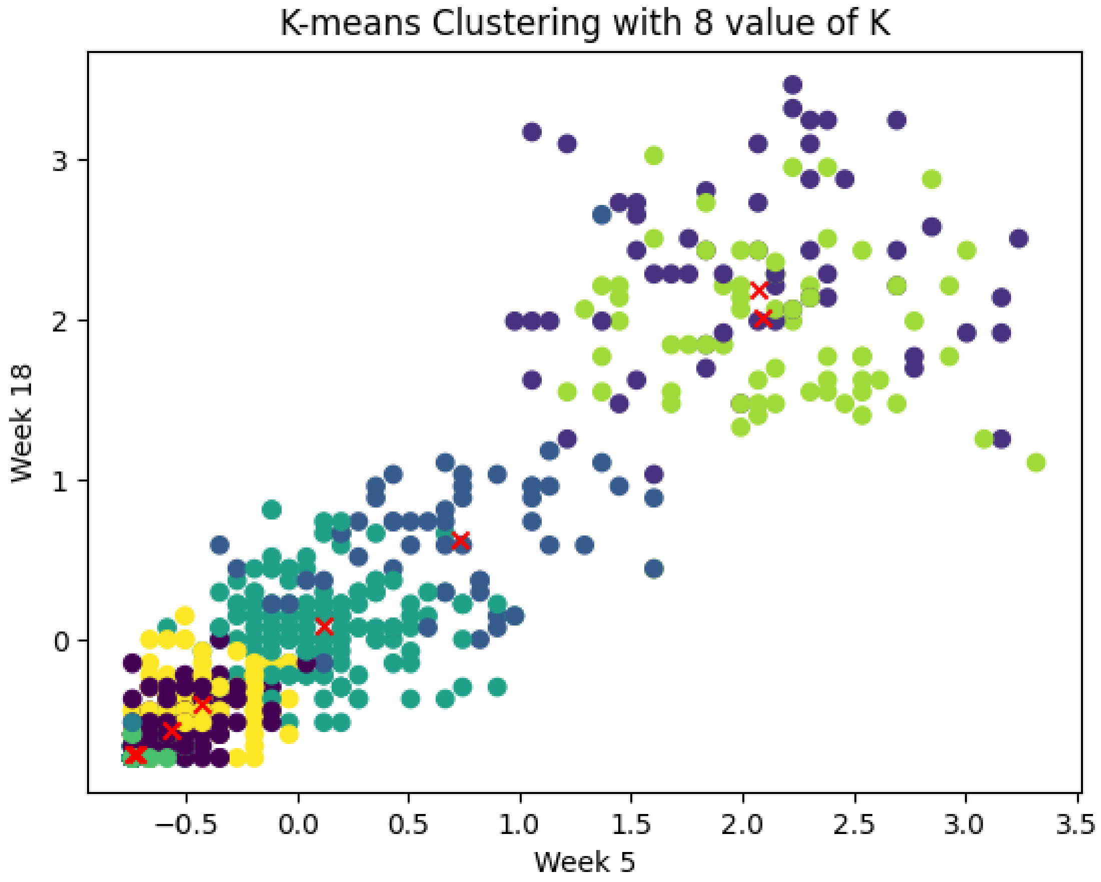


```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

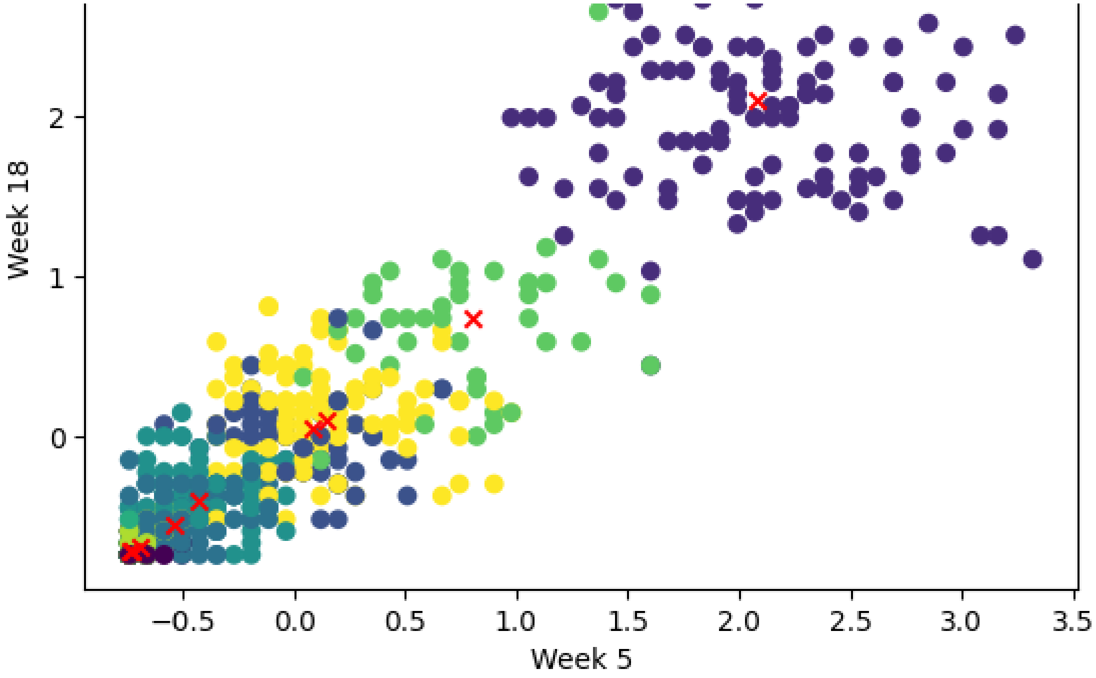




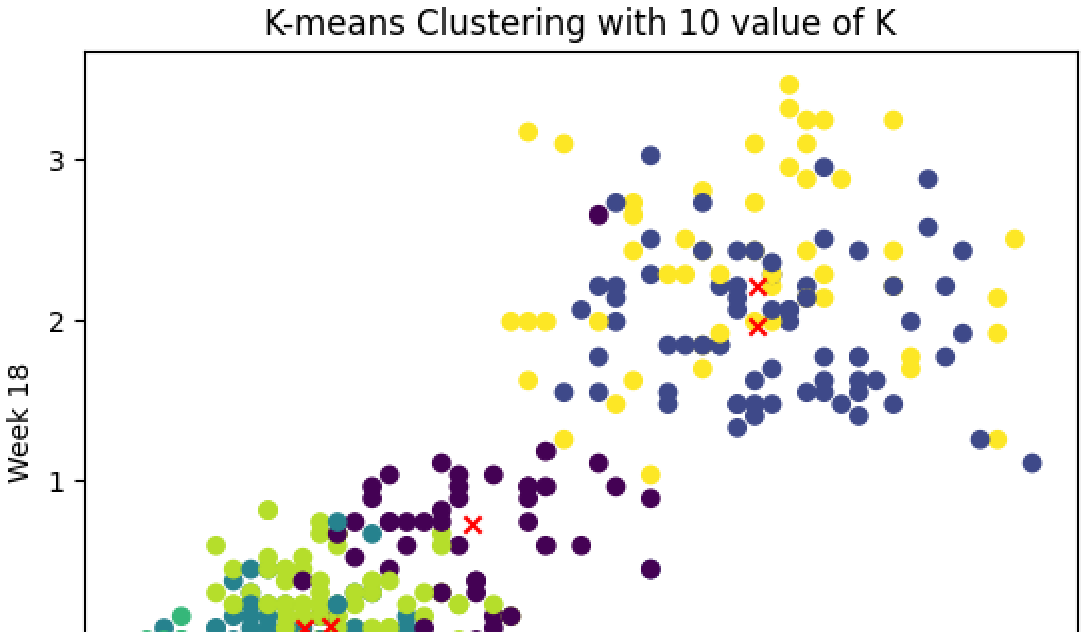
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```



```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will chang
  warnings.warn(
```

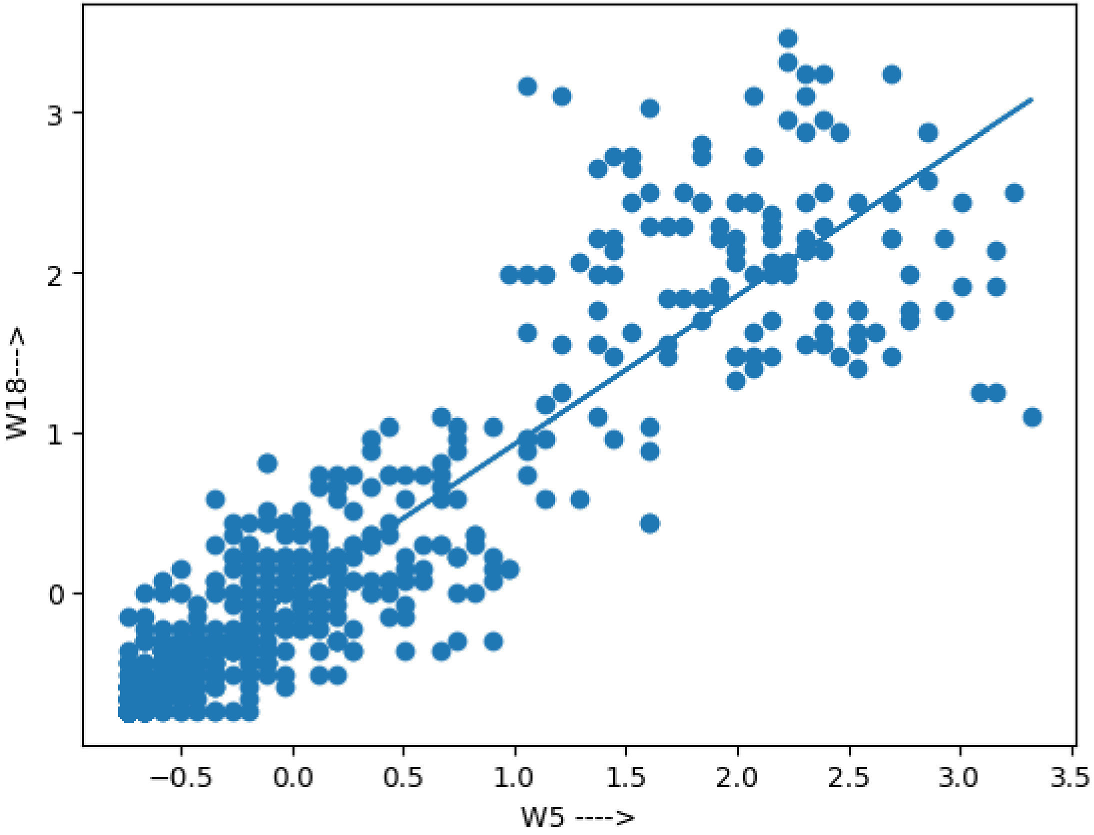


Liner Regression With Feature Scaling:

```
xn= np.array(scaled.W5)
yn= np.array(scaled.W18)
n=len(xn)
xn_mean = xn.mean()
yn_mean = yn.mean()
sum_xn = (xn).sum()
sum_yn = (yn).sum()
sum_xnyn = ((xn*yn).sum())*n
print(sum_xnyn)
sum_xn_sum_yn = sum_xn*sum_yn
sum_xnxn = ((xn*xn).sum())*n
print(sum_xnxn)
sum_xn_Square = sum_xn*sum_xn
slp = (sum_xnyn- sum_xn_sum_yn) / (sum_xnxn-sum_xn_Square)
print("Slope=" , slp)
Intcp = yn_mean - slp * xn_mean
print("Intercept=" ,Intcp)
```

```
609491.9885452681
657721.0
Slope= 0.9266725382727146
Intercept= -3.504526440247103e-17
```

```
plt.xlabel("W5 ---->")
plt.scatter(scaled.W5,scaled.W18)
plt.ylabel("W18---> ")
plt.plot(xn,slp*xn+Intcp)
plt.show()
```


3
2
1
0
W18--->
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
W5 ---->>

✓ 0s completed at 3:19 AM ● ×