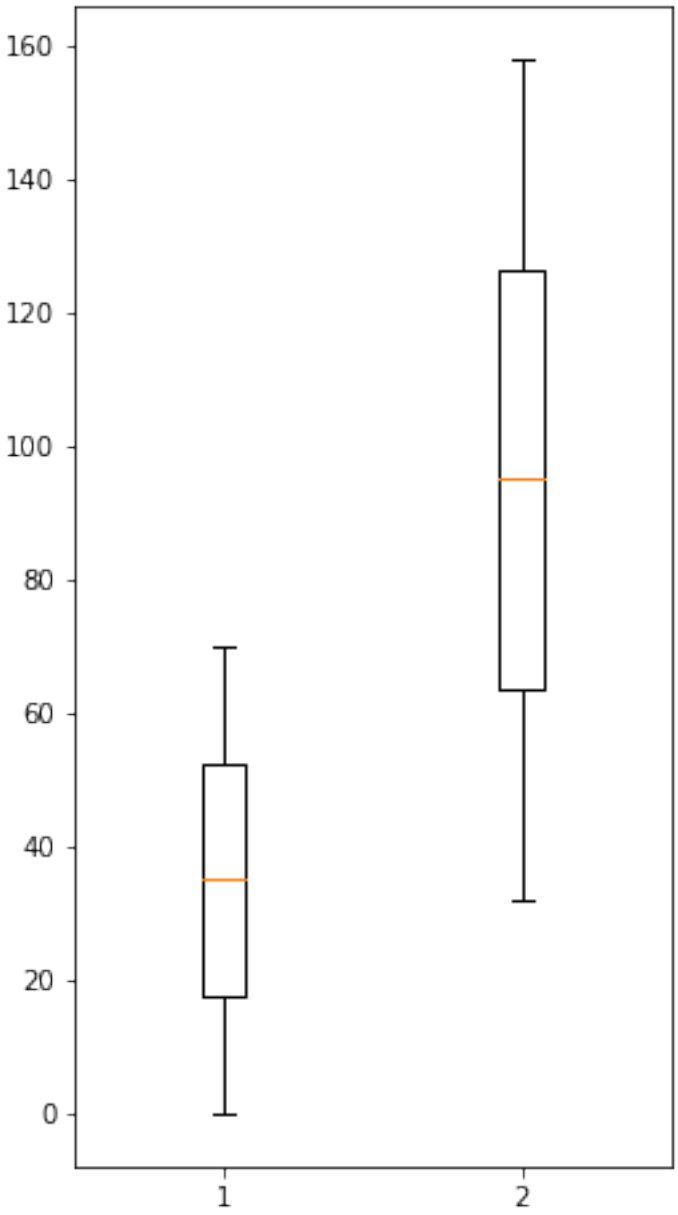
Data scale 1

July 12, 2022

[ ]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

data_ct = [0, 10, 20, 30, 40, 50, 60, 70]
data_ft = [32, 50, 68, 86, 104, 122, 140, 158]
data = [data_ct, data_ft]
data
```

[ ]:
```
[[0, 10, 20, 30, 40, 50, 60, 70], [32, 50, 68, 86, 104, 122, 140, 158]]
```

[ ]:
```
fig = plt.figure(figsize = (4, 8))
plt.boxplot(data)
plt.show()
```



[ ]:
```
temps = {"ct": [0, 10, 20, 30, 40, 50, 60, 70], "ft": [32, 50, 68, 86, 104,␣
↪122, 140, 158]}
df = pd.DataFrame(temps)
df
```

[ ]:

	ct	ft
0	0	32
1	10	50
2	20	68

```
3  30   86
4  40  104
5  50  122
6  60  140
7  70  158
```

1 Standard Scaler

[ ]:
```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(df)
df_scaled = scaler.transform(df)
pd.DataFrame(df_scaled, columns = ["ct_std", "ft_std"])
```

[ ]:
```
     ct_std    ft_std
0 -1.527525 -1.527525
1 -1.091089 -1.091089
2 -0.654654 -0.654654
3 -0.218218 -0.218218
4  0.218218  0.218218
5  0.654654  0.654654
6  1.091089  1.091089
7  1.527525  1.527525
```

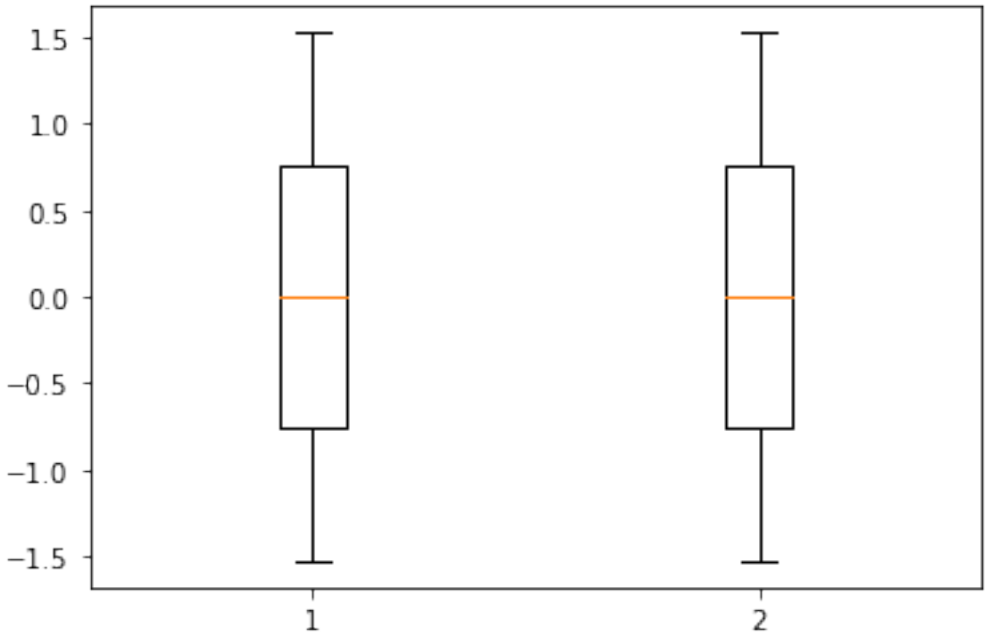
[ ]:
```
print(df.mean())
```

```
ct    35.0
ft    95.0
dtype: float64
```

[ ]:
```
print(df_scaled.mean())
print(df_scaled.var())
```

```
0.0
1.0
```

[ ]:
```
plt.boxplot(df_scaled)
plt.show()
```



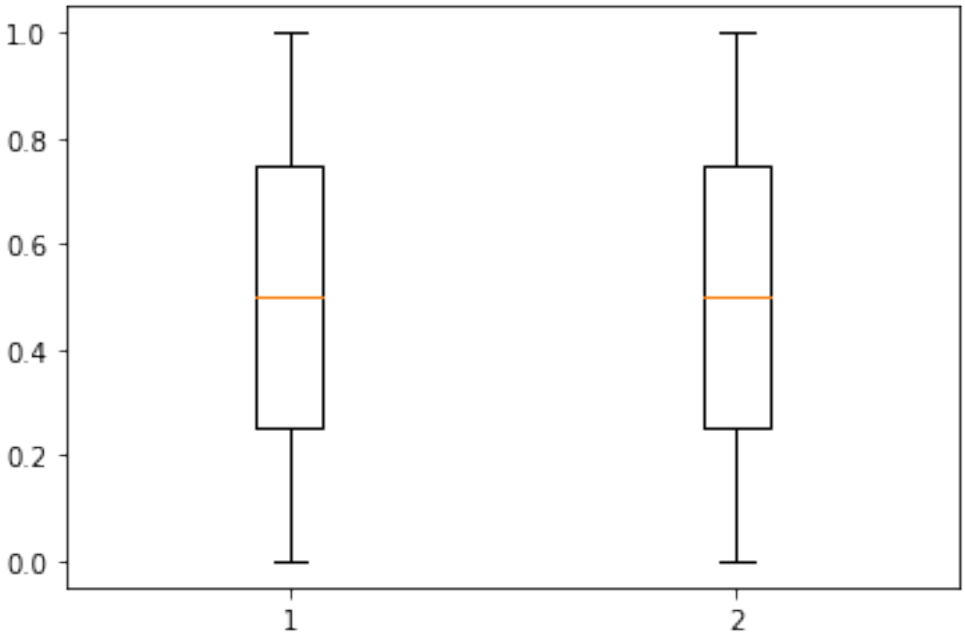
2 MinMax Scaler

[ ]:
```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
df_minmax = scaler.fit_transform(df)

pd.DataFrame(df_minmax, columns = ['ct_minmax', 'ft_minmax'])
```

[ ]:

	ct_minmax	ft_minmax
0	0.000000	0.000000
1	0.142857	0.142857
2	0.285714	0.285714
3	0.428571	0.428571
4	0.571429	0.571429
5	0.714286	0.714286
6	0.857143	0.857143
7	1.000000	1.000000

[ ]:
```
plt.boxplot(df_minmax)
plt.show()
```



3 Outlier

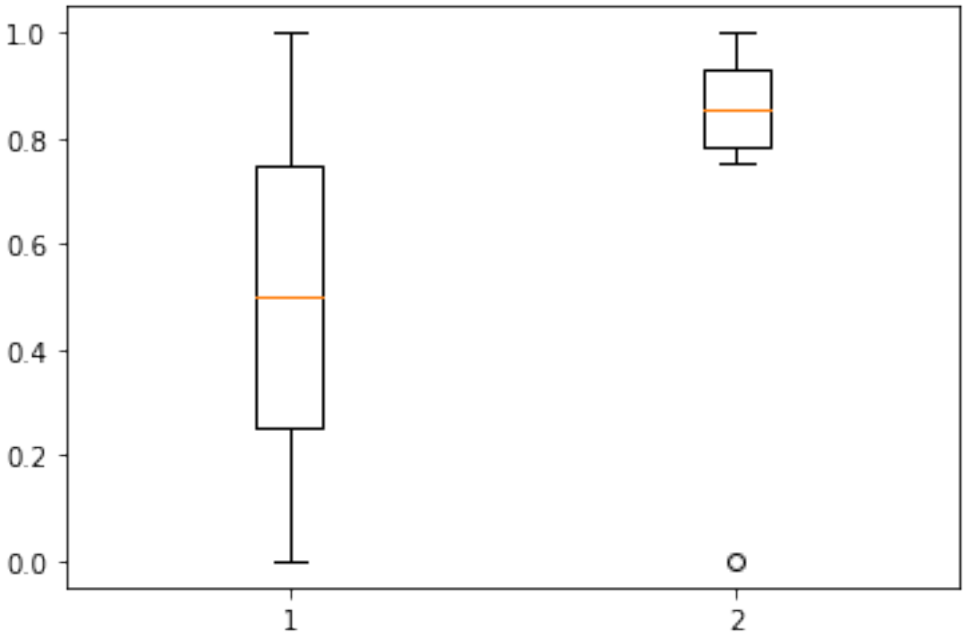
[ ]:
```
temps = {"ct": [0, 10, 20, 30 , 40 , 50, 60, 70], "ft": [32, 50, 68, 86, 104,␣
↪122, 140, -300]}
df_outlier = pd.DataFrame(temps)
df_outlier
```

[ ]:

	ct	ft
0	0	32
1	10	50
2	20	68
3	30	86
4	40	104
5	50	122
6	60	140
7	70	-300

4 MinMax Scaler with Outlier

[ ]:
```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
df_minmax_outlier = scaler.fit_transform(df_outlier)
plt.boxplot(df_minmax_outlier)
plt.show()
```



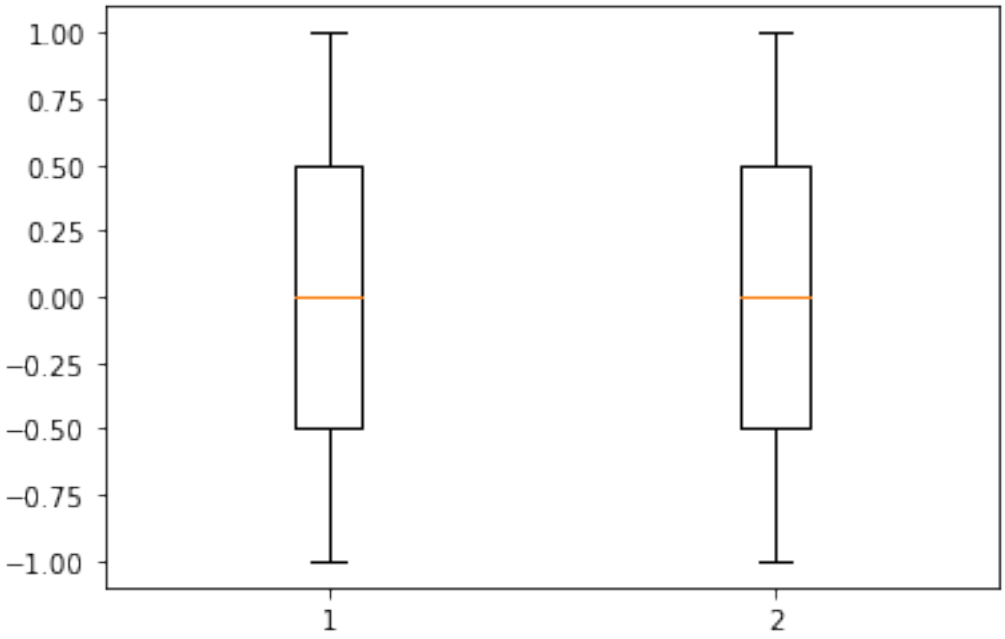
5 Robust Scaler without Outlier

[ ]:
```
from sklearn.preprocessing import RobustScaler
scaler = RobustScaler()
df_robust = scaler.fit_transform(df)

pd.DataFrame(df_robust, columns = ['ct_robust', 'ft_robust'])
```

[ ]:

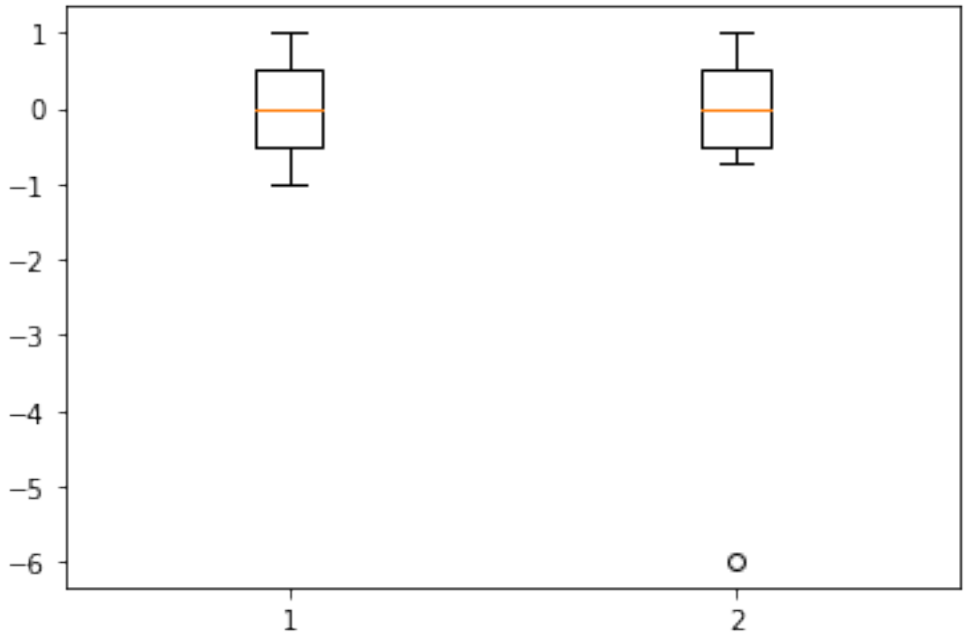
	ct_robust	ft_robust
0	-1.000000	-1.000000
1	-0.714286	-0.714286
2	-0.428571	-0.428571
3	-0.142857	-0.142857
4	0.142857	0.142857
5	0.428571	0.428571
6	0.714286	0.714286
7	1.000000	1.000000

[ ]:
```
plt.boxplot(df_robust)
plt.show()
```



6 Robust Scaler with Outlier

[ ]:
```
temps = {"ct": [0, 10, 20, 30, 40, 50, 60, 70], "ft": [32, 50, 68, 86, 104,␣
 ↪122, 140, -300]}
df_outlier = pd.DataFrame(temps)

from sklearn.preprocessing import RobustScaler
scaler = RobustScaler()
df_minmax_outlier =scaler.fit_transform(df_outlier)
plt.boxplot(df_minmax_outlier)
plt.show()
```


1
0
−1
−2
−3
−4
−5
−6
1
2

[ ]: