Lab 4.2

Outlier identification IQR fILLING

Objective

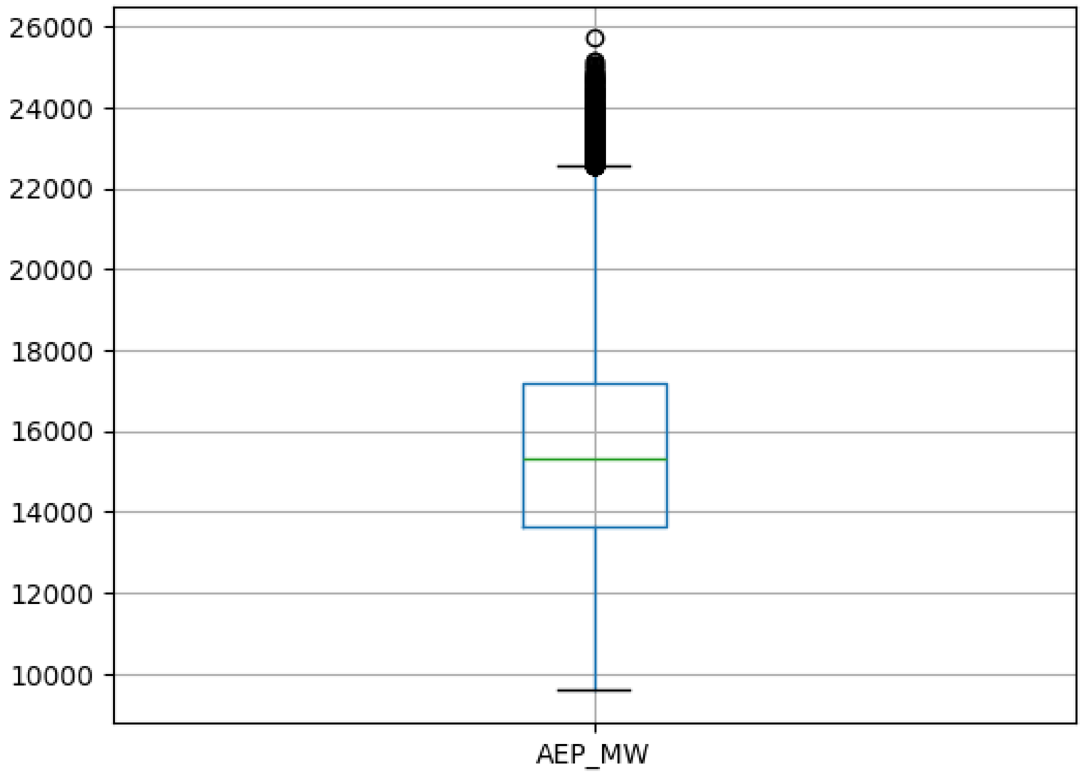
To identify outliers in a dataset using the Interquartile Range (IQR) method and apply appropriate techniques to handle (fill) the outliers for improved data quality.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#import seaborn as sns
%matplotlib inline

df = pd.read_csv(r'C:\Users\PMLS\labreports\
LAB4\2_Missing_Values_Filled.csv', parse_dates=['Datetime'],
index_col='Datetime')

df.boxplot('AEP_MW')

<Axes: >
```



```
def detect_outliers_iqr(data):
    outliers = []
    #data = sorted(data)
    q1 = data['AEP_MW'].quantile(0.25)
    q3 = data['AEP_MW'].quantile(0.75)
    # print(q1, q3)
    IQR = q3-q1
    lwr_bound = q1-(1.5*IQR)
    upr_bound = q3+(1.5*IQR)
    # print(lwr_bound, upr_bound)
    idx = 2
    idx_list = []
    for i in list(data['AEP_MW']):
        if (i<lwr_bound or i>upr_bound):
            outliers.append(i)
            idx_list.append(idx)
        idx = idx + 1
    data.loc[data['AEP_MW'] < lwr_bound, 'AEP_MW'] = np.nan
    data.loc[data['AEP_MW'] > upr_bound, 'AEP_MW'] = np.nan
    return outliers, data, idx_list# Driver code
```

It calculates the IQR (interquartile range) of the 'AEP_MW' column.

Finds outliers as values below Q1 - 1.5×IQR or above Q3 + 1.5×IQR.

Collects these outlier values and their indices.

Replaces outliers in the DataFrame with NaN.

Returns the outlier values list, the modified DataFrame, and the outlier indices list.

```
sample_outliers, p_data, index_of_out = detect_outliers_iqr(df)
print("Outliers from IQR method: ", sample_outliers)
```

```
Outliers from IQR method:  [22577.0, 22940.0, 22886.0, 22652.0,
22639.0, 22579.0, 22558.0, 22791.0, 23013.0, 22970.0, 22728.0,
23022.0, 23558.0, 23741.0, 22942.0, 22922.0, 23470.0, 23860.0,
24015.0, 23708.0, 23543.0, 22910.0, 22766.0, 22658.0, 23089.0,
23359.0, 23479.0, 23308.0, 22985.0, 22836.0, 23325.0, 23435.0,
23614.0, 23418.0, 23092.0, 22605.0, 22671.0, 22717.0, 22725.0,
22610.0, 22909.0, 23371.0, 23243.0, 23276.0, 23082.0, 22573.0,
22597.0, 23194.0, 23650.0, 23845.0, 23970.0, 23865.0, 23458.0,
22983.0, 22808.0, 22922.0, 23008.0, 22958.0, 23299.0, 23476.0,
23156.0, 22779.0, 23458.0, 24020.0, 24342.0, 24270.0, 24311.0,
24273.0, 24113.0, 23649.0, 23158.0, 22971.0, 23271.0, 23975.0,
24435.0, 24711.0, 24642.0, 24712.0, 24632.0, 24316.0, 23750.0,
23176.0, 23364.0, 23266.0, 23975.0, 24413.0, 24602.0, 24741.0,
24811.0, 24842.0, 24538.0, 23963.0, 23504.0, 23403.0, 22986.0,
23369.0, 23426.0, 23271.0, 22894.0, 22920.0, 22566.0, 22578.0,
23199.0, 23257.0, 23395.0, 23188.0, 22857.0, 22591.0, 23719.0,
24065.0, 23822.0, 23465.0, 22843.0, 22581.0, 23692.0, 24420.0,
24274.0, 23965.0, 23518.0, 22937.0, 22846.0, 23547.0, 23369.0,
22753.0, 23356.0, 23290.0, 22665.0, 22603.0, 22712.0, 23689.0,
23553.0, 22886.0, 23326.0, 23228.0, 22684.0, 22646.0, 22601.0,
22802.0, 23680.0, 23442.0, 22878.0, 23195.0, 23561.0, 23574.0,
23657.0, 23591.0, 23080.0, 22855.0, 23213.0, 23334.0, 23314.0,
23236.0, 22972.0, 22943.0, 23098.0, 22966.0, 23071.0, 22890.0,
22789.0, 23504.0, 23818.0, 23940.0, 24038.0, 24004.0, 23532.0,
22940.0, 22774.0, 23058.0, 23064.0, 22687.0, 23027.0, 23421.0,
23585.0, 23747.0, 23683.0, 23308.0, 22751.0, 22801.0, 23468.0,
23777.0, 23643.0, 23623.0, 23653.0, 23358.0, 22666.0, 22611.0,
23211.0, 23434.0, 23650.0, 23565.0, 23286.0, 22932.0, 22867.0,
23398.0, 23724.0, 23827.0, 23663.0, 23511.0, 22963.0, 22809.0,
23080.0, 23884.0, 24409.0, 24586.0, 24611.0, 24490.0, 24423.0,
23839.0, 23403.0, 23265.0, 23410.0, 24356.0, 24941.0, 25164.0,
25056.0, 25140.0, 24878.0, 24480.0, 23860.0, 23550.0, 23347.0,
23154.0, 24184.0, 24736.0, 24911.0, 24978.0, 25035.0, 24696.0,
24045.0, 23077.0, 22643.0, 22622.0, 23229.0, 23526.0, 23683.0,
23637.0, 23444.0, 22905.0, 22624.0, 22672.0, 22637.0, 23033.0,
23661.0, 23844.0, 23832.0, 23327.0, 22602.0, 22622.0, 23082.0,
23595.0, 23732.0, 23443.0, 22779.0, 22565.0, 23201.0, 24017.0,
24639.0, 24828.0, 24862.0, 24629.0, 23870.0, 23198.0, 23091.0,
22644.0, 23277.0, 24204.0, 24555.0, 24475.0, 24352.0, 24044.0,
23328.0, 22892.0, 23240.0, 23545.0, 23524.0, 23227.0, 23422.0,
24082.0, 24291.0, 24505.0, 24330.0, 23862.0, 23142.0, 22976.0,
22688.0, 22661.0, 22706.0, 23221.0, 23551.0, 23605.0, 23061.0,
22825.0, 23189.0, 23335.0, 23093.0, 22676.0, 23151.0, 23156.0,
22827.0, 22565.0, 22583.0, 22943.0, 22714.0, 22593.0, 23299.0,
23381.0, 23041.0, 22912.0, 23069.0, 23137.0, 22678.0, 23531.0,
24465.0, 24242.0, 23726.0, 22784.0, 22976.0, 22788.0, 22565.0,
22766.0, 22995.0, 23070.0, 23059.0, 22639.0, 22733.0, 23444.0,
23833.0, 23938.0, 23828.0, 23900.0, 23624.0, 22971.0, 22674.0,
22649.0, 22600.0, 22572.0, 22606.0, 22600.0, 22925.0, 23134.0,
23286.0, 23224.0, 22878.0, 22618.0, 22793.0, 22766.0, 22762.0,
23230.0, 23356.0, 23177.0, 22851.0, 22814.0, 23083.0, 23310.0,
23266.0, 23109.0, 22841.0, 22950.0, 22857.0, 22758.0, 22728.0,
22913.0, 23079.0, 23007.0, 25695.0, 22727.0, 22930.0, 22811.0,
22613.0, 22666.0, 23741.0, 23843.0, 23688.0, 23426.0, 22903.0,
22653.0, 23797.0, 24703.0, 24679.0, 24462.0, 24134.0, 23516.0,
22919.0, 22565.0, 23561.0, 23565.0, 23548.0, 23408.0, 22864.0,
22950.0, 23155.0, 22925.0, 22580.0, 22917.0, 22931.0, 22816.0,
22779.0, 22710.0, 22579.0, 22804.0, 23766.0, 23650.0, 23228.0,
22644.0, 22817.0, 22820.0, 23108.0, 23945.0, 23588.0, 22965.0,
22707.0, 22633.0, 22690.0, 22744.0, 22813.0, 22781.0, 22630.0,
22873.0, 22839.0, 22600.0, 22672.0, 22689.0, 22676.0, 22600.0,
22565.0, 23257.0, 23624.0, 23736.0, 23687.0, 23628.0, 23475.0,
23130.0, 22557.0, 22848.0, 22888.0, 22697.0, 22586.0, 22742.0,
22699.0, 22995.0, 23324.0, 23079.0, 23097.0, 23045.0, 22776.0,
22678.0, 22914.0, 23126.0, 23214.0, 23081.0, 22689.0, 22696.0,
23042.0, 23158.0, 23202.0, 23013.0, 22777.0, 22958.0, 22698.0,
22839.0, 23023.0, 22904.0, 22824.0, 23158.0, 23045.0, 22812.0,
22997.0, 22952.0, 22604.0, 22634.0, 23064.0, 22984.0, 22881.0,
23053.0, 22840.0, 22612.0, 22662.0, 22784.0, 22651.0, 22612.0,
22995.0, 23188.0, 23203.0, 23016.0, 22689.0, 22577.0, 22828.0,
22885.0, 22758.0, 23233.0, 23527.0, 23668.0, 23682.0, 23390.0,
23284.0, 22838.0, 22756.0, 23703.0, 24367.0, 24409.0, 24426.0,
24597.0, 24235.0, 24104.0, 23776.0, 23097.0, 23118.0, 23354.0,
23855.0, 24235.0, 23964.0, 23476.0, 22849.0, 22707.0, 22756.0,
22752.0, 22601.0, 22953.0, 23346.0, 23517.0, 23481.0, 23476.0,
23149.0, 22635.0, 22783.0, 23102.0, 23136.0, 22960.0, 22650.0,
22694.0, 22751.0, 22639.0, 22594.0, 22974.0, 23148.0, 23214.0,
22954.0, 22731.0, 22666.0, 22826.0, 22893.0, 22558.0, 22791.0,
23109.0, 23289.0, 23134.0, 22858.0, 22619.0, 23252.0, 23320.0,
22660.0, 22850.0, 22744.0, 22675.0, 22844.0, 22909.0, 22609.0,
22732.0, 22671.0, 22807.0, 22597.0, 22839.0, 23212.0, 23073.0,
22601.0, 22656.0, 23001.0, 22997.0, 23029.0, 22650.0, 22571.0,
22569.0, 22642.0, 22839.0, 22705.0, 22708.0, 22793.0, 22790.0,
22723.0, 22568.0, 22776.0, 22858.0, 22810.0, 22823.0, 22687.0,
22631.0, 22781.0, 22722.0, 23473.0, 23418.0, 23286.0, 22918.0,
23076.0, 23590.0, 23447.0, 23342.0, 23195.0, 22824.0, 22984.0,
22924.0, 22584.0, 23347.0, 23066.0, 22611.0, 22814.0, 23071.0,
23062.0, 22828.0, 23517.0, 24231.0, 24087.0, 23692.0, 23175.0,
22877.0, 23040.0, 22983.0, 22966.0, 23264.0, 23230.0, 22957.0,
22567.0, 23361.0, 24123.0, 24030.0, 23441.0, 22656.0, 22597.0,
22604.0, 23702.0, 24421.0, 24092.0, 23085.0, 22778.0, 22816.0,
22903.0, 22565.0, 23056.0, 23657.0, 23578.0, 23211.0, 22815.0,
23139.0, 23562.0, 23507.0, 23256.0, 22951.0, 22572.0, 23338.0,
23692.0, 23693.0, 23518.0, 23075.0, 22584.0, 22693.0, 23412.0,
24210.0, 24739.0, 24348.0, 23678.0, 23026.0, 22609.0, 22685.0,
22744.0, 22759.0, 22624.0, 22567.0]
```

```python
pd.isnull(p_data['AEP_MW']).sum()
```

```
np.int64(666)
```

```python
index_of_out
```

```
[1930,
 2625,
 2626,
 2769,
 2770,
 6473,
 6474,
 7048,
 7049,
 7050,
 7051,
 7142,
 7143,
 7144,
 7145,
 7166,
 7167,
 7168,
 7169,
 7170,
 7171,
 7172,
 7314,
 7315,
 7336,
 7337,
 7338,
 7339,
 7340,
 7359,
 7360,
 7361,
 7362,
 7363,
 7364,
 7383,
 7384,
 7385,
 7386,
 7387,
 7575,
 7576,
 7577,
 7578,
 7579,
 10689,
 15710,
 15711,
 15712,
 15713,
 15714,
 15715,
 15716,
 15717,
 15736,
 15737,
 15738,
 15760,
 15761,
 15762,
 15763,
 15764,
 16046,
 16047,
 16048,
 16049,
 16050,
 16051,
 16052,
 16053,
 16054,
 16055,
 16069,
 16070,
 16071,
 16072,
 16073,
 16074,
 16075,
 16076,
 16077,
 16078,
 16079,
 16093,
 16094,
 16095,
 16096,
 16097,
 16098,
 16099,
 16100,
 16101,
 16102,
 16103,
 16118,
 16119,
 16120,
 16121,
 16215,
 16216,
 16217,
 20457,
 20577,
 20578,
 20579,
 20580,
 20581,
 20582,
 20588,
 20589,
 20590,
 20591,
 20592,
 20599,
 20600,
 20601,
 20602,
 20603,
 20604,
 20605,
 20611,
 20612,
 20613,
 20614,
 20649,
 20650,
 20651,
 20662,
 20672,
 20673,
 20674,
 20675,
 20817,
 20818,
 20819,
 20829,
 20830,
 20840,
 20841,
 20842,
 20843,
 23775,
 23776,
 23777,
 23778,
 23779,
 23780,
 23967,
 23968,
 23969,
 23970,
 23971,
 23972,
 23991,
 23992,
 23993,
 23994,
 23995,
 24278,
 24279,
 24280,
 24281,
 24282,
 24283,
 24284,
 24285,
 24809,
 24810,
 24811,
 24812,
 24831,
 24832,
 24833,
 24834,
 24835,
 24836,
 24837,
 24854,
 24855,
 24856,
 24857,
 24858,
 24859,
 24860,
 24861,
 24878,
 24879,
 24880,
 24881,
 24882,
 24883,
 24884,
 24951,
 24952,
 24953,
 24954,
 24955,
 24956,
 24957,
 24958,
 24974,
 24975,
 24976,
 24977,
 24978,
 24979,
 24980,
 24981,
 24982,
 24983,
 24997,
 24998,
 24999,
 25000,
 25001,
 25002,
 25003,
 25004,
 25005,
 25006,
 25007,
 25021,
 25022,
 25023,
 25024,
 25025,
 25026,
 25027,
 25028,
 25029,
 25030,
 25046,
 25047,
 25048,
 25049,
 25050,
 25051,
 25052,
 25120,
 25121,
 25122,
 25191,
 25192,
 25193,
 25194,
 25195,
 25196,
 25336,
 25337,
 25338,
 25339,
 25340,
 25341,
 25342,
 25358,
 25359,
 25360,
 25361,
 25362,
 25363,
 25364,
 25365,
 25366,
 25367,
 25382,
 25383,
 25384,
 25385,
 25386,
 25387,
 25388,
 25480,
 25481,
 25482,
 25483,
 25484,
 25503,
 25504,
 25505,
 25506,
 25507,
 25508,
 25509,
 25510,
 25650,
 25651,
 25672,
 25673,
 25674,
 25675,
 25676,
 25696,
 25697,
 25698,
 25699,
 25700,
 28532,
 28533,
 28534,
 28545,
 28557,
 28569,
 28570,
 28976,
 28977,
 28978,
 28979,
 29060,
 29061,
 29062,
 29063,
 29072,
 29073,
 29074,
 29075,
 29076,
 29481,
 29482,
 29721,
 32271,
 32272,
 32273,
 32274,
 32275,
 32341,
 32342,
 32343,
 32344,
 32345,
 32346,
 32347,
 32348,
 32417,
 32418,
 32440,
 33234,
 33235,
 33255,
 33256,
 33257,
 33258,
 33259,
 33260,
 33281,
 33282,
 33283,
 33352,
 33353,
 33354,
 33355,
 33356,
 33544,
 33545,
 33546,
 33547,
 33548,
 33616,
 33617,
 33618,
 34386,
 34387,
 34409,
 34410,
 34411,
 35535,
 37041,
 37042,
 37043,
 37044,
 37627,
 37628,
 37629,
 37630,
 37631,
 37632,
 37639,
 37640,
 37641,
 37642,
 37643,
 37644,
 37645,
 37646,
 37651,
 37652,
 37653,
 37654,
 37655,
 37656,
 37665,
 37666,
 37667,
 37668,
 37737,
 37738,
 37739,
 37748,
 37749,
 37750,
 37760,
 37761,
 37762,
 37763,
 38108,
 38109,
 38110,
 38120,
 38121,
 38122,
 38123,
 46281,
 50514,
 50536,
 50537,
 50538,
 50539,
 50559,
 50560,
 50561,
 50562,
 50728,
 50729,
 50730,
 50897,
 50898,
 50918,
 50919,
 50920,
 50921,
 50922,
 50923,
 50924,
 50925,
 51040,
 51041,
 51042,
 51329,
 51330,
 51331,
 51351,
 51352,
 51353,
 51354,
 51355,
 51356,
 51399,
 51400,
 51401,
 51402,
 51403,
 51404,
 51423,
 51424,
 51425,
 51426,
 51427,
 51833,
 51834,
 51835,
 51857,
 51858,
 51859,
 54355,
 54356,
 54357,
 54358,
 54369,
 54370,
 54371,
 54379,
 54380,
 54381,
 54382,
 54393,
 54394,
 58600,
 58601,
 58602,
 58603,
 59559,
 59560,
 59561,
 59562,
 59563,
 59564,
 59585,
 59586,
 59587,
 59588,
 59607,
 59608,
 59609,
 59610,
 59611,
 59612,
 59613,
 59629,
 59630,
 59631,
 59632,
 59633,
 59634,
 59635,
 59636,
 59637,
 59638,
 59639,
 59653,
 59654,
 59655,
 59656,
 59657,
 59658,
 59753,
 59754,
 59755,
 59778,
 59799,
 59800,
 59801,
 59802,
 59803,
 59804,
 59805,
 59823,
 59824,
 59825,
 59826,
 59827,
 59895,
 59896,
 59897,
 59919,
 59920,
 59921,
 59922,
 59923,
 59924,
 60664,
 60665,
 60666,
 60667,
 67865,
 67866,
 67867,
 67868,
 67869,
 67886,
 67887,
 67888,
 68057,
 68058,
 68059,
 68081,
 68082,
 68083,
 68084,
 68296,
 68297,
 68298,
 68299,
 68318,
 68319,
 68320,
 68321,
 68534,
 68535,
 68536,
 68537,
 68706,
 68707,
 77056,
 77057,
 77058,
 77059,
 77079,
 77080,
 77081,
 77082,
 77083,
 77103,
 77104,
 77105,
 77106,
 77107,
 78400,
 78401,
 78402,
 81236,
 81237,
 81238,
 81239,
 81248,
 81249,
 81250,
 81251,
 81252,
 81253,
 81260,
 81261,
 81262,
 81609,
 81610,
 81611,
 81644,
 81645,
 81646,
 81647,
 81656,
 81657,
 81658,
 81659,
 81660,
 81753,
 81754,
 81755,
 81756,
 81764,
 81765,
 81766,
 81767,
 81776,
 81777,
 81778,
 81779,
 81780,
 81789,
 81790,
 81800,
 81801,
 81802,
 81803,
 90020,
 90021,
 90022,
 90023,
 90032,
 90033,
 90034,
 90035,
 90036,
 91040,
 91041,
 91042,
 91043,
 91044,
 91045,
 91052,
 91053,
 91054,
 91055,
 91056,
 91057,
 91062,
 91063,
 91064,
 91065,
 91066,
 91067,
 91068,
 91161,
 116193,
 116194,
 116217,
 116218,
 116577]
```

```
df['AEP_MW'] = df['AEP_MW'].interpolate(method='time')
```

```
pd.isnull(p_data['AEP_MW']).sum()
```

```
np.int64(0)
```

```
p_data.reset_index(inplace=True)
```

```
p_data.to_csv(r'C:\Users\PMLS\ML\
LAB4\3_Outlier_Identified.csv',index=False)
```

```
print(pd.__version__)
```

```
2.2.3
```