```
fuel_data = pd.read_csv('fuel_ferc1.csv')
```

```
fuel_data
```

```
                     record_id  utility_id_ferc1  report_year  \
0        f1_fuel_1994_12_1_0_7                 1         1994   
1       f1_fuel_1994_12_1_0_10                 1         1994   
2        f1_fuel_1994_12_2_0_1                 2         1994   
3        f1_fuel_1994_12_2_0_7                 2         1994   
4       f1_fuel_1994_12_2_0_10                 2         1994   
...                        ...               ...          ...   
29518  f1_fuel_2018_12_12_0_13                12         2018   
29519   f1_fuel_2018_12_12_1_1                12         2018   
29520  f1_fuel_2018_12_12_1_10                12         2018   
29521  f1_fuel_2018_12_12_1_13                12         2018   
29522  f1_fuel_2018_12_12_1_14                12         2018   

           plant_name_ferc1 fuel_type_code_pudl fuel_unit 
fuel_qty_burned  \
0                  rockport                coal       ton 
5377489.0   
1      rockport total plant                coal       ton 
10486945.0   
2                    gorgas                coal       ton 
2978683.0   
3                     barry                coal       ton 
3739484.0   
4                 chickasaw                 gas       mcf 
40533.0   
...                     ...                 ...       ... 
...   
29518    neil simpson ct #1                 gas       mcf 
18799.0   
29519  cheyenne prairie 58%                 gas       mcf 
806730.0   
29520     lange ct facility                 gas       mcf 
104554.0   
29521       wygen 3 bhp 52%                coal       ton 
315945.0   
29522       wygen 3 bhp 52%                 gas       mcf 
17853.0   

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                   16.590                      18.59   
1                   16.592                      18.58   
2                   24.130                      39.72   
3                   23.950                      47.21   
4                    1.000                       2.77   
...                    ...                        ...   
29518                1.059                       4.78   
```

```
29519                1.050                      3.65
29520                1.060                      4.77
29521               16.108                      3.06
29522                1.059                      0.00

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                             18.53                1.121
1                             18.53                1.120
2                             38.12                1.650
3                             45.99                1.970
4                              2.77                2.570
...                             ...                  ...
29518                          4.78                9.030
29519                          3.65                6.950
29520                          4.77                8.990
29521                         14.76                1.110
29522                          0.00               11.680

[29523 rows x 11 columns]
```

```
import pandas as pd
```

```
fuel_data.describe(include= 'all' )
```

```
                      record_id  utility_id_ferc1   report_year  \
count                     29523      29523.000000  29523.000000   
unique                    29523               NaN           NaN   
top     f1_fuel_2003_12_193_2_6               NaN           NaN   
freq                          1               NaN           NaN   
mean                        NaN        118.601836   2005.806050   
std                         NaN         74.178353      7.025483   
min                         NaN          1.000000   1994.000000   
25%                         NaN         55.000000   2000.000000   
50%                         NaN        122.000000   2006.000000   
75%                         NaN        176.000000   2012.000000   
max                         NaN        514.000000   2018.000000   

       plant_name_ferc1 fuel_type_code_pudl fuel_unit  fuel_qty_burned
\
count             29523               29523     29343     2.952300e+04

unique             2315                   6         9              NaN

top           big stone                 gas       mcf              NaN

freq                156               11486     11354              NaN

mean                NaN                 NaN       NaN     2.622119e+06

std                 NaN                 NaN       NaN     9.118004e+06
```

```
min                 NaN                 NaN       NaN     1.000000e+00

25%                 NaN                 NaN       NaN     1.381700e+04

50%                 NaN                 NaN       NaN     2.533220e+05

75%                 NaN                 NaN       NaN     1.424034e+06

max                 NaN                 NaN       NaN     5.558942e+08
```

```
        fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
count          29523.000000               29523.000000
unique                  NaN                        NaN
top                     NaN                        NaN
freq                    NaN                        NaN
mean               8.492111                 208.649031
std               10.600220                2854.490090
min                0.000001                -276.080000
25%                1.024000                   5.207000
50%                5.762694                  26.000000
75%               17.006000                  47.113000
max              341.260000             139358.000000

        fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
count                   2.952300e+04         29523.000000
unique                           NaN                  NaN
top                              NaN                  NaN
freq                             NaN                  NaN
mean                    9.175704e+02            19.304354
std                     6.877593e+04          2091.540939
min                    -8.749370e+02           -41.501000
25%                     3.778500e+00             1.940000
50%                     1.737100e+01             4.127000
75%                     4.213700e+01             7.745000
max                     7.964521e+06        359278.000000
```

```
fuel_data.isnull().sum()
```

```
record_id                         0
utility_id_ferc1                  0
report_year                       0
plant_name_ferc1                  0
fuel_type_code_pudl               0
fuel_unit                       180
fuel_qty_burned                   0
fuel_mmbtu_per_unit               0
fuel_cost_per_unit_burned         0
fuel_cost_per_unit_delivered      0
```

```
fuel_cost_per_mmbtu               0
dtype: int64
```

```
#use groupby to count the sum of each unique value in the fuel unit
column
fuel_data.groupby( 'fuel_unit' )[ 'fuel_unit' ].count()
fuel_data[[ 'fuel_unit' ]] = fuel_data[[ 'fuel_unit' ]].fillna(value=
'mcf' )
```

```
fuel_data.isnull().sum()
```

```
record_id                       0
utility_id_ferc1                0
report_year                     0
plant_name_ferc1                0
fuel_type_code_pudl             0
fuel_unit                       0
fuel_qty_burned                 0
fuel_mmbtu_per_unit             0
fuel_cost_per_unit_burned       0
fuel_cost_per_unit_delivered    0
fuel_cost_per_mmbtu             0
dtype: int64
```

```
fuel_data.groupby( 'report_year' )[ 'report_year' ].count()
#group by the fuel type code year and print the first entries in all
the groups formed
fuel_data.groupby( 'fuel_type_code_pudl' ).first()
```

```
                                  record_id  utility_id_ferc1
report_year  \
fuel_type_code_pudl

coal                    f1_fuel_1994_12_1_0_7                 1
1994
gas                    f1_fuel_1994_12_2_0_10                 2
1994
nuclear                 f1_fuel_1994_12_2_1_1                 2
1994
oil                     f1_fuel_1994_12_6_0_2                 6
1994
other                  f1_fuel_1994_12_11_0_6                11
1994
waste                   f1_fuel_1994_12_9_0_3                 9
1994

                     plant_name_ferc1 fuel_unit  fuel_qty_burned  \
fuel_type_code_pudl
coal                         rockport       ton        5377489.0
gas                         chickasaw       mcf          40533.0
nuclear              joseph m. farley       kgU           2260.0
```

```
oil                      clinch river       bbl           6510.0
other                      w.f. wyman       bbl          55652.0
waste                    b.l. england       ton           2438.0

                     fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
fuel_type_code_pudl
coal                           16.590000                     18.590
gas                             1.000000                      2.770
nuclear                         0.064094                     28.770
oil                             5.875338                     32.130
other                           0.149719                     14.685
waste                           0.015939                     34.180

                     fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu

fuel_type_code_pudl

coal                                       18.530                1.121

gas                                         2.770                2.570

nuclear                                     0.000                0.450

oil                                        23.444                5.469

other                                      15.090                2.335

waste                                      34.180                1.072
```

```
fuel_df1 = fuel_data.iloc[ 0 : 19000 ].reset_index(drop= True )

fuel_df1
```

```
                      record_id  utility_id_ferc1  report_year  \
0         f1_fuel_1994_12_1_0_7                 1         1994
1        f1_fuel_1994_12_1_0_10                 1         1994
2         f1_fuel_1994_12_2_0_1                 2         1994
3         f1_fuel_1994_12_2_0_7                 2         1994
4        f1_fuel_1994_12_2_0_10                 2         1994
...                         ...               ...          ...
18995   f1_fuel_2009_12_182_1_9               182         2009
18996  f1_fuel_2009_12_182_1_10               182         2009
18997  f1_fuel_2009_12_182_1_13               182         2009
18998  f1_fuel_2009_12_182_1_14               182         2009
18999    f1_fuel_2009_12_79_0_1                79         2009

           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
0                  rockport                coal       ton
5377489.0
```

```
1      rockport total plant                coal       ton
10486945.0
2                     gorgas                coal       ton
2978683.0
3                      barry                coal       ton
3739484.0
4                  chickasaw                 gas       mcf
40533.0
...                      ...                 ...       ...
...
18995              lake road                 gas       mcf
340857.0
18996              lake road                 oil       mcf
771.0
18997            iatan (18%)                coal       ton
414142.0
18998            iatan (18%)                 oil       bbl
5761.0
18999               montrose                coal       ton
2050919.0

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                16.590000                     18.590
1                16.592000                     18.580
2                24.130000                     39.720
3                23.950000                     47.210
4                 1.000000                      2.770
...                    ...                        ...
18995             1.000000                      4.711
18996             5.801544                     84.899
18997            16.718000                     18.509
18998             5.537910                     83.636
18999            17.160000                     29.629

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                            18.530                1.121
1                            18.530                1.120
2                            38.120                1.650
3                            45.990                1.970
4                             2.770                2.570
...                             ...                  ...
18995                         4.711                4.711
18996                        84.899               14.634
18997                        17.570                1.107
18998                        72.280               15.102
18999                        28.330                1.727

[19000 rows x 11 columns]
```

```
fuel_df2 = fuel_data.iloc[ 19000 :].reset_index(drop= True )
```

```
fuel_df2

                     record_id  utility_id_ferc1  report_year  \
0        f1_fuel_2009_12_79_0_2                79         2009   
1        f1_fuel_2009_12_79_0_4                79         2009   
2        f1_fuel_2009_12_79_0_5                79         2009   
3        f1_fuel_2009_12_79_0_7                79         2009   
4       f1_fuel_2009_12_79_0_10                79         2009   
...                         ...               ...          ...   
10518   f1_fuel_2018_12_12_0_13                12         2018   
10519    f1_fuel_2018_12_12_1_1                12         2018   
10520   f1_fuel_2018_12_12_1_10                12         2018   
10521   f1_fuel_2018_12_12_1_13                12         2018   
10522   f1_fuel_2018_12_12_1_14                12         2018   

           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
0                  montrose                 oil       bbl
22912.0   
1                hawthorn 5                coal       ton
2408123.0   
2                hawthorn 5                 gas       mcf
82141.0   
3            hawthorn 6 & 9                 gas       mcf
1701680.0   
4            hawthorn 7 & 8                 gas       mcf
82601.0   
...                     ...                 ...       ...
...   
10518    neil simpson ct #1                 gas       mcf
18799.0   
10519  cheyenne prairie 58%                 gas       mcf
806730.0   
10520     lange ct facility                 gas       mcf
104554.0   
10521       wygen 3 bhp 52%                coal       ton
315945.0   
10522       wygen 3 bhp 52%                 gas       mcf
17853.0   

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                 5.770422                     65.443   
1                16.782000                     18.598   
2                 1.000000                      6.238   
3                 1.000000                      4.885   
4                 1.000000                      5.383   
...                    ...                        ...   
10518             1.059000                      4.780   
10519             1.050000                      3.650   
10520             1.060000                      4.770   
```

```
10521            16.108000                      3.060
10522             1.059000                      0.000

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                            67.540               11.341
1                            18.310                1.108
2                             6.238                6.238
3                             4.885                4.885
4                             5.383                5.383
...                             ...                  ...
10518                         4.780                9.030
10519                         3.650                6.950
10520                         4.770                8.990
10521                        14.760                1.110
10522                         0.000               11.680

[10523 rows x 11 columns]
```

```
fuel_data.iloc[ 19000 :]
```

```
                     record_id  utility_id_ferc1  report_year  \
19000   f1_fuel_2009_12_79_0_2                79         2009
19001   f1_fuel_2009_12_79_0_4                79         2009
19002   f1_fuel_2009_12_79_0_5                79         2009
19003   f1_fuel_2009_12_79_0_7                79         2009
19004  f1_fuel_2009_12_79_0_10                79         2009
...                        ...               ...          ...
29518  f1_fuel_2018_12_12_0_13                12         2018
29519   f1_fuel_2018_12_12_1_1                12         2018
29520  f1_fuel_2018_12_12_1_10                12         2018
29521  f1_fuel_2018_12_12_1_13                12         2018
29522  f1_fuel_2018_12_12_1_14                12         2018

           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
19000              montrose                 oil       bbl
22912.0
19001            hawthorn 5                coal       ton
2408123.0
19002            hawthorn 5                 gas       mcf
82141.0
19003        hawthorn 6 & 9                 gas       mcf
1701680.0
19004        hawthorn 7 & 8                 gas       mcf
82601.0
...                     ...                 ...       ...
...
29518    neil simpson ct #1                 gas       mcf
18799.0
29519  cheyenne prairie 58%                 gas       mcf
```

```
806730.0
29520     lange ct facility                gas       mcf
104554.0
29521       wygen 3 bhp 52%               coal       ton
315945.0
29522       wygen 3 bhp 52%                gas       mcf
17853.0

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
19000             5.770422                     65.443
19001            16.782000                     18.598
19002             1.000000                      6.238
19003             1.000000                      4.885
19004             1.000000                      5.383
...                    ...                        ...
29518             1.059000                      4.780
29519             1.050000                      3.650
29520             1.060000                      4.770
29521            16.108000                      3.060
29522             1.059000                      0.000

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
19000                        67.540               11.341
19001                        18.310                1.108
19002                         6.238                6.238
19003                         4.885                4.885
19004                         5.383                5.383
...                             ...                  ...
29518                         4.780                9.030
29519                         3.650                6.950
29520                         4.770                8.990
29521                        14.760                1.110
29522                         0.000               11.680

[10523 rows x 11 columns]
```

```
fuel_data.iloc[ 0 : 19000 ]
```

```
                      record_id  utility_id_ferc1  report_year  \
0         f1_fuel_1994_12_1_0_7                 1         1994
1        f1_fuel_1994_12_1_0_10                 1         1994
2         f1_fuel_1994_12_2_0_1                 2         1994
3         f1_fuel_1994_12_2_0_7                 2         1994
4        f1_fuel_1994_12_2_0_10                 2         1994
...                         ...               ...          ...
18995   f1_fuel_2009_12_182_1_9               182         2009
18996  f1_fuel_2009_12_182_1_10               182         2009
18997  f1_fuel_2009_12_182_1_13               182         2009
18998  f1_fuel_2009_12_182_1_14               182         2009
18999    f1_fuel_2009_12_79_0_1                79         2009
```

```
           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
0                  rockport                coal       ton
5377489.0
1       rockport total plant                coal       ton
10486945.0
2                    gorgas                coal       ton
2978683.0
3                     barry                coal       ton
3739484.0
4                 chickasaw                 gas       mcf
40533.0
...                     ...                 ...       ...
...
18995             lake road                 gas       mcf
340857.0
18996             lake road                 oil       mcf
771.0
18997           iatan (18%)                coal       ton
414142.0
18998           iatan (18%)                 oil       bbl
5761.0
18999              montrose                coal       ton
2050919.0

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                16.590000                     18.590
1                16.592000                     18.580
2                24.130000                     39.720
3                23.950000                     47.210
4                 1.000000                      2.770
...                    ...                        ...
18995             1.000000                      4.711
18996             5.801544                     84.899
18997            16.718000                     18.509
18998             5.537910                     83.636
18999            17.160000                     29.629

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                            18.530                1.121
1                            18.530                1.120
2                            38.120                1.650
3                            45.990                1.970
4                             2.770                2.570
...                             ...                  ...
18995                         4.711                4.711
18996                        84.899               14.634
18997                        17.570                1.107
18998                        72.280               15.102
```

```
18999                        28.330                1.727

[19000 rows x 11 columns]
```

```
assert len(fuel_data) == (len(fuel_df1) + len(fuel_df2))
```

```
#an inner merge will lose rows that do not match in both dataframes
pd.merge(fuel_df1, fuel_df2, how= "inner" )
```

```
Empty DataFrame
Columns: [record_id, utility_id_ferc1, report_year, plant_name_ferc1,
fuel_type_code_pudl, fuel_unit, fuel_qty_burned, fuel_mmbtu_per_unit,
fuel_cost_per_unit_burned, fuel_cost_per_unit_delivered,
fuel_cost_per_mmbtu]
Index: []
```

```
#outer merge returns all rows in both dataframes
pd.merge(fuel_df1, fuel_df2, how= "outer" )
```

```
                     record_id  utility_id_ferc1  report_year  \
0        f1_fuel_1994_12_1_0_7                 1         1994
1       f1_fuel_1994_12_1_0_10                 1         1994
2        f1_fuel_1994_12_2_0_1                 2         1994
3        f1_fuel_1994_12_2_0_7                 2         1994
4       f1_fuel_1994_12_2_0_10                 2         1994
...                        ...               ...          ...
29518  f1_fuel_2018_12_12_0_13                12         2018
29519   f1_fuel_2018_12_12_1_1                12         2018
29520  f1_fuel_2018_12_12_1_10                12         2018
29521  f1_fuel_2018_12_12_1_13                12         2018
29522  f1_fuel_2018_12_12_1_14                12         2018

           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
0                  rockport                coal       ton
5377489.0
1      rockport total plant                coal       ton
10486945.0
2                    gorgas                coal       ton
2978683.0
3                     barry                coal       ton
3739484.0
4                 chickasaw                 gas       mcf
40533.0
...                     ...                 ...       ...
...
29518    neil simpson ct #1                 gas       mcf
18799.0
29519  cheyenne prairie 58%                 gas       mcf
806730.0
29520     lange ct facility                 gas       mcf
```

```
104554.0
29521       wygen 3 bhp 52%                coal       ton
315945.0
29522       wygen 3 bhp 52%                 gas       mcf
17853.0

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                   16.590                      18.59
1                   16.592                      18.58
2                   24.130                      39.72
3                   23.950                      47.21
4                    1.000                       2.77
...                    ...                        ...
29518                1.059                       4.78
29519                1.050                       3.65
29520                1.060                       4.77
29521               16.108                       3.06
29522                1.059                       0.00

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                             18.53                1.121
1                             18.53                1.120
2                             38.12                1.650
3                             45.99                1.970
4                              2.77                2.570
...                             ...                  ...
29518                          4.78                9.030
29519                          3.65                6.950
29520                          4.77                8.990
29521                         14.76                1.110
29522                          0.00               11.680

[29523 rows x 11 columns]
```

```
#removes rows from the right dataframe that do not have a match with
the left
#and keeps all rows from the left
pd.merge(fuel_df1, fuel_df2, how= "left" )
```

```
                      record_id  utility_id_ferc1  report_year  \
0         f1_fuel_1994_12_1_0_7                 1         1994
1        f1_fuel_1994_12_1_0_10                 1         1994
2         f1_fuel_1994_12_2_0_1                 2         1994
3         f1_fuel_1994_12_2_0_7                 2         1994
4        f1_fuel_1994_12_2_0_10                 2         1994
...                         ...               ...          ...
18995   f1_fuel_2009_12_182_1_9               182         2009
18996  f1_fuel_2009_12_182_1_10               182         2009
18997  f1_fuel_2009_12_182_1_13               182         2009
18998  f1_fuel_2009_12_182_1_14               182         2009
```

```
18999    f1_fuel_2009_12_79_0_1                79         2009

           plant_name_ferc1 fuel_type_code_pudl fuel_unit
fuel_qty_burned  \
0                  rockport                coal       ton
5377489.0
1       rockport total plant                coal       ton
10486945.0
2                    gorgas                coal       ton
2978683.0
3                     barry                coal       ton
3739484.0
4                 chickasaw                 gas       mcf
40533.0
...                     ...                 ...       ...
...
18995             lake road                 gas       mcf
340857.0
18996             lake road                 oil       mcf
771.0
18997           iatan (18%)                coal       ton
414142.0
18998           iatan (18%)                 oil       bbl
5761.0
18999              montrose                coal       ton
2050919.0

       fuel_mmbtu_per_unit  fuel_cost_per_unit_burned  \
0                16.590000                     18.590
1                16.592000                     18.580
2                24.130000                     39.720
3                23.950000                     47.210
4                 1.000000                      2.770
...                    ...                        ...
18995             1.000000                      4.711
18996             5.801544                     84.899
18997            16.718000                     18.509
18998             5.537910                     83.636
18999            17.160000                     29.629

       fuel_cost_per_unit_delivered  fuel_cost_per_mmbtu
0                            18.530                1.121
1                            18.530                1.120
2                            38.120                1.650
3                            45.990                1.970
4                             2.770                2.570
...                             ...                  ...
18995                         4.711                4.711
18996                        84.899               14.634
18997                        17.570                1.107
```

```
18998                        72.280               15.102
18999                        28.330                1.727

[19000 rows x 11 columns]
```

```
pd.concat([fuel_data, data_to_concat]).reset_index(drop= True )
```

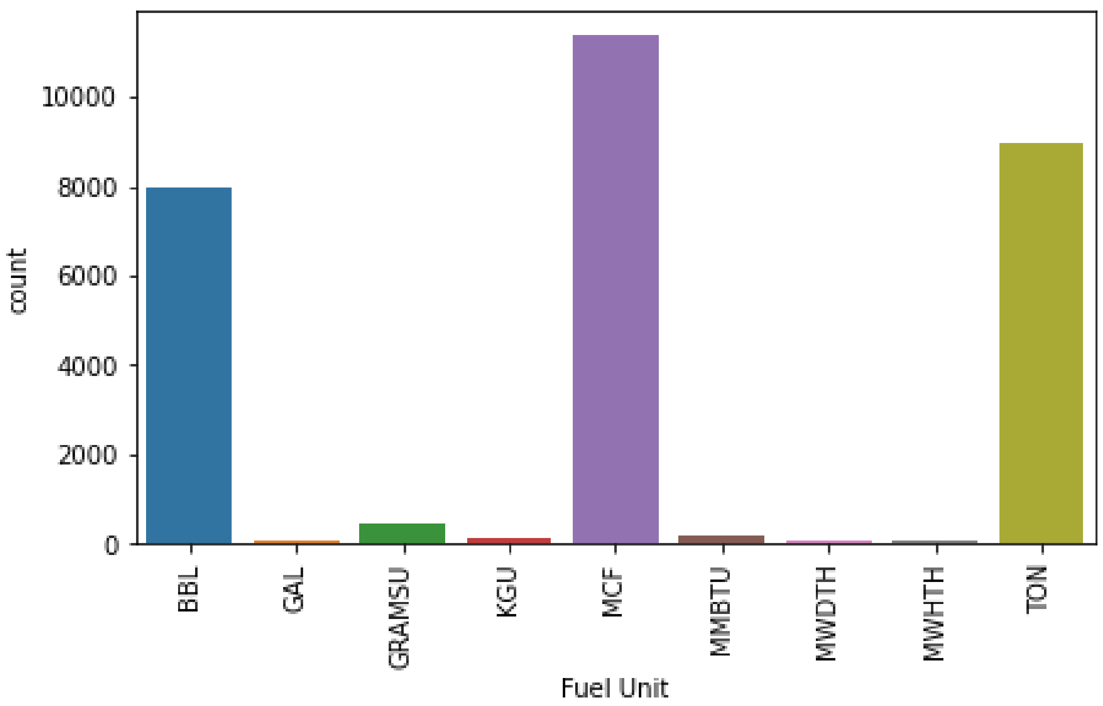
```
----------------------------------------------------------------------
-----
NameError                                 Traceback (most recent call
last)
<ipython-input-37-48aad84ddbc5> in <module>
----> 1 pd.concat([fuel_data, data_to_concat]).reset_index(drop=
True )

NameError: name 'data_to_concat' is not defined
```

```
#check for duplicate rows
fuel_data.duplicated().any()
```

```
False
```

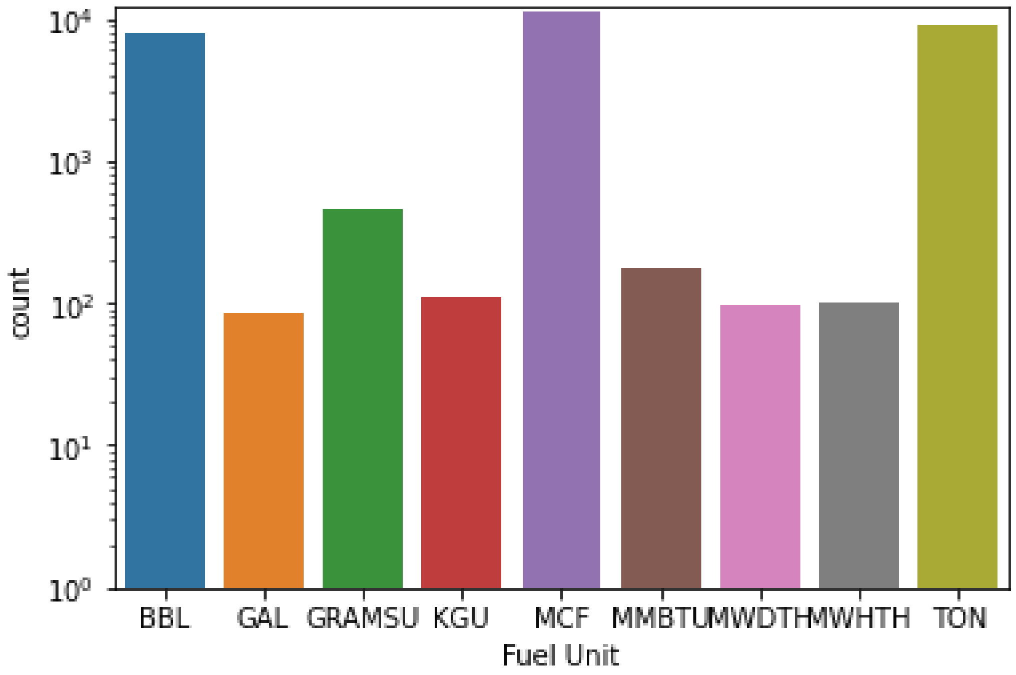
```
# Import plotting library
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(7,4))
plt.xticks(rotation=90)
fuel_unit = pd.DataFrame({'unit':['BBL', 'GAL', 'GRAMSU', 'KGU',
'MCF', 'MMBTU',
'MWDTH', 'MWHTH', 'TON'],
'count':[7998, 84, 464, 110, 11354, 180, 95, 100, 8958]})
sns.barplot(data=fuel_unit, x='unit', y='count')
plt.xlabel('Fuel Unit')
```

```
Text(0.5, 0, 'Fuel Unit')
```



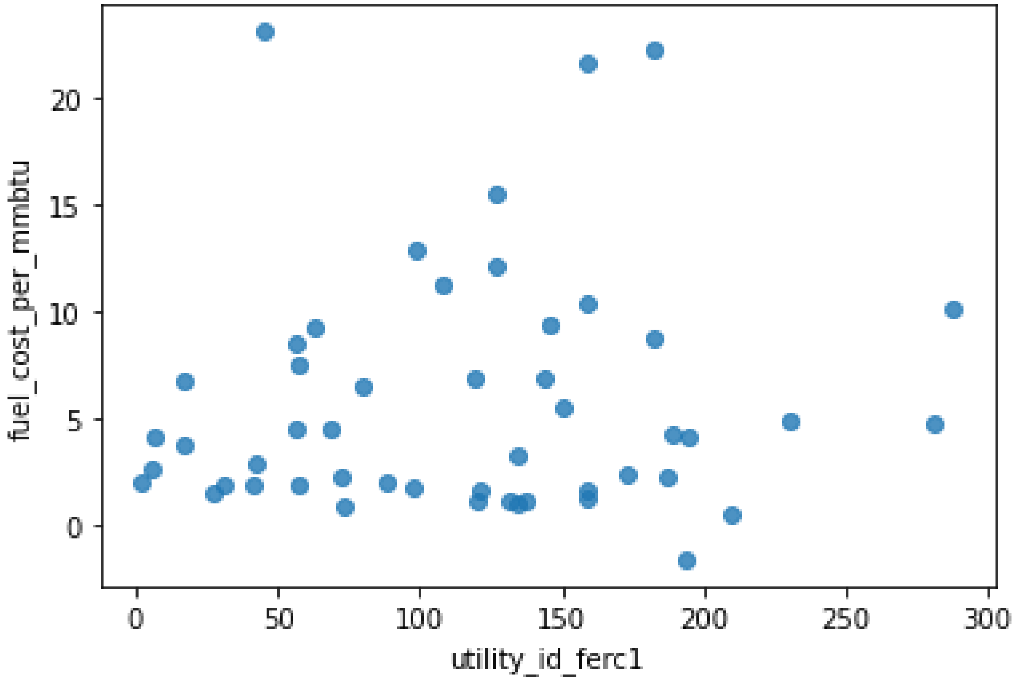
```
g = sns.barplot(data=fuel_unit, x='unit', y='count')
g.set_yscale("log")
g.set_ylim(1, 12000)
plt.xlabel('Fuel Unit')
```

```
Text(0.5, 0, 'Fuel Unit')
```



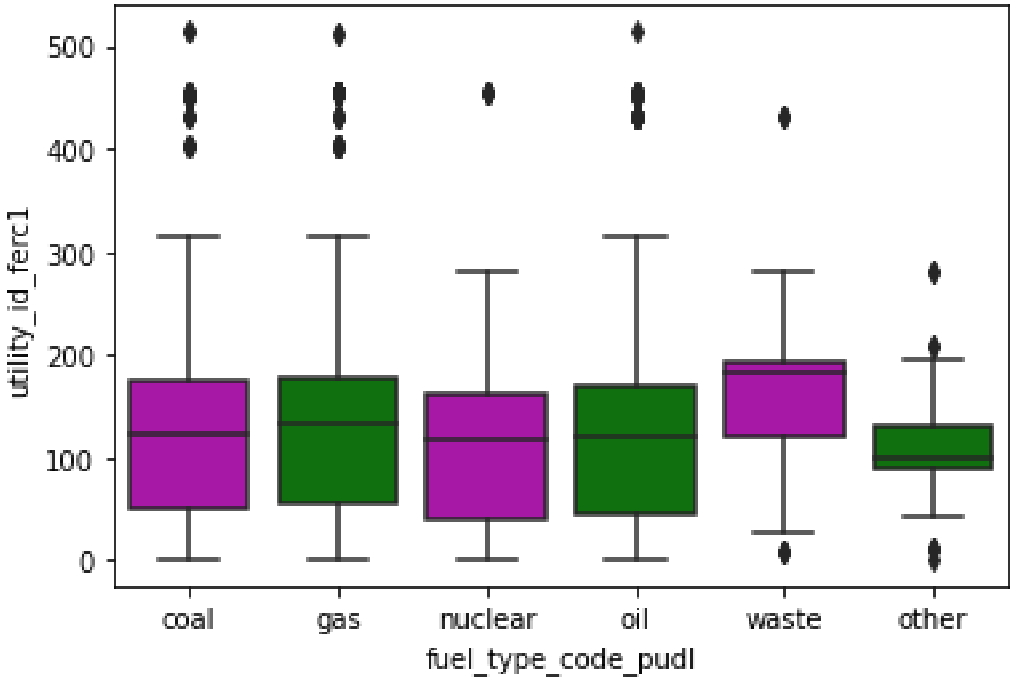
```
# Select a sample of the dataset
sample_df = fuel_data.sample(n=50, random_state=4)
sns.regplot(x=sample_df["utility_id_ferc1"],
y=sample_df["fuel_cost_per_mmbtu"],
fit_reg=False)
```

```
<AxesSubplot:xlabel='utility_id_ferc1', ylabel='fuel_cost_per_mmbtu'>
```



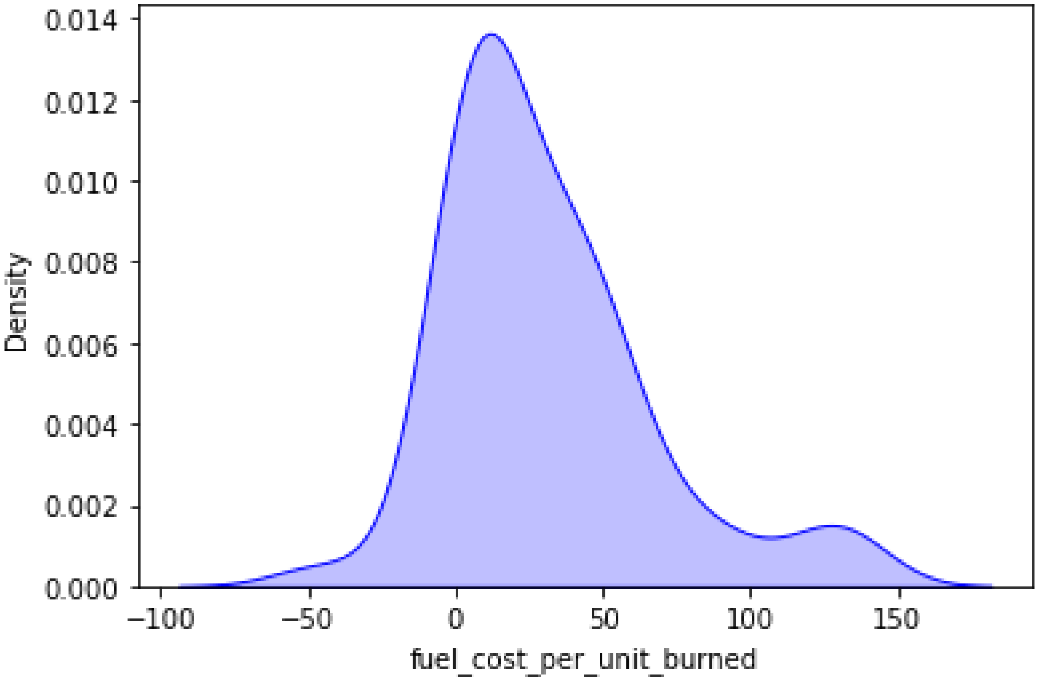
```
sns.boxplot(x="fuel_type_code_pudl", y="utility_id_ferc1",
palette=["m", "g"], data=fuel_data)
```

```
<AxesSubplot:xlabel='fuel_type_code_pudl', ylabel='utility_id_ferc1'>
```



```
sns.kdeplot(sample_df['fuel_cost_per_unit_burned'], shade=True,
color="b")
```

```
<AxesSubplot:xlabel='fuel_cost_per_unit_burned', ylabel='Density'>
```


Density
fuel_cost_per_unit_burned