In [2]:
```
import pandas as pd
```

In [6]:
```
data=pd.read_csv('https://raw.githubusercontent.com/WalePhenomenon/climate_change/master/fuel_ferc1.csv')
data.to_csv('hamoyeproject',index_label=False)
```

In [3]:
```
project=pd.read_csv('hamoyeproject.csv')
```

In [4]:
```
project.head()
```

Out[4]:

	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
0	f1_fuel_1994_12_1_0_7	1	1994	rockport	coal	ton	5377489.0	16.590	18.59	18.53	1.121
1	f1_fuel_1994_12_1_0_10	1	1994	rockport total plant	coal	ton	10486945.0	16.592	18.58	18.53	1.120
2	f1_fuel_1994_12_2_0_1	2	1994	gorgas	coal	ton	2978683.0	24.130	39.72	38.12	1.650
3	f1_fuel_1994_12_2_0_7	2	1994	barry	coal	ton	3739484.0	23.950	47.21	45.99	1.970
4	f1_fuel_1994_12_2_0_10	2	1994	chickasaw	gas	mcf	40533.0	1.000	2.77	2.77	2.570

In [5]:
```
A=[1,2,3,4,5,6]
B=[13,21,34]
```

In [6]:
```
A.append(B)
```

In [7]:
```
A.extend(B)
```

In [8]:
```
A.pop(A.index([13,21,34]))
```

Out[8]:
```
[13, 21, 34]
```

In [9]:
```
A
```

Out[9]:
```
[1, 2, 3, 4, 5, 6, 13, 21, 34]
```

In [10]:
```
import numpy as np
```

In [11]:
```
np.eye(3)
```

Out[11]:
```
array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])
```

In [12]:
```
np.identity(3)
```

Out[12]:
```
array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])
```

In [13]:
```
project.head()
```

Out[13]:

	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
0	f1_fuel_1994_12_1_0_7	1	1994	rockport	coal	ton	5377489.0	16.590	18.59	18.53	1.121
1	f1_fuel_1994_12_1_0_10	1	1994	rockport total plant	coal	ton	10486945.0	16.592	18.58	18.53	1.120
2	f1_fuel_1994_12_2_0_1	2	1994	gorgas	coal	ton	2978683.0	24.130	39.72	38.12	1.650
3	f1_fuel_1994_12_2_0_7	2	1994	barry	coal	ton	3739484.0	23.950	47.21	45.99	1.970
4	f1_fuel_1994_12_2_0_10	2	1994	chickasaw	gas	mcf	40533.0	1.000	2.77	2.77	2.570

In [14]:
```
pd.DataFrame(project.groupby('fuel_type_code_pudl').mean()['fuel_cost_per_unit_burned'])
#MIN IS GAS
```

Out[14]:

	fuel_cost_per_unit_burned
fuel_type_code_pudl	
coal	67.421830
gas	13.659397
nuclear	4955.157002
oil	168.877086
other	18.253856
waste	19.518122

In [15]:
```
project.columns
```

Out[15]:
```
Index(['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'],
      dtype='object')
```

In [16]:
```
np.std(project['fuel_mmbtu_per_unit'])
```

Out[16]:
```
10.600040781504145
```

In [17]:
```
project.describe()
```

Out[17]:

	utility_id_ferc1	report_year	fuel_qty_burned	fuel_mmbtu_per_unit	fuel_cost_per_unit_burned	fuel_cost_per_unit_delivered	fuel_cost_per_mmbtu
count	29523.000000	29523.000000	2.952300e+04	29523.000000	29523.000000	2.952300e+04	29523.000000
mean	118.601836	2005.806050	2.622119e+06	8.492111	208.649031	9.175704e+02	19.304354
std	74.178353	7.025483	9.118004e+06	10.600220	2854.490090	6.877593e+04	2091.540939
min	1.000000	1994.000000	1.000000e+00	0.000001	-276.080000	-8.749370e+02	-41.501000
25%	55.000000	2000.000000	1.391700e+04	1.024000	5.207000	3.778500e+00	1.940000
50%	122.000000	2006.000000	2.533220e+05	5.762694	26.000000	1.737100e+01	4.127000
75%	176.000000	2012.000000	1.424034e+06	17.006000	47.113000	4.213700e+01	7.745000
max	514.000000	2018.000000	5.558942e+08	341.260000	139356.000000	7.964521e+06	359278.000000

In [18]:
```
#standard deviation is 10.06
#75 percentile is 17.01
```

In [ ]:
```

```

In [ ]:
```

```

In [19]:
```
import scipy as sp
```

In [21]:
```
from scipy.stats import skew,kurtosis
```

In [22]:
```
skw=sp.stats.skew(project.fuel_qty_burned)
kur=sp.stats.kurtosis(project.fuel_qty_burned)
```

In [23]:
```
print('The skewness is {} and  the kurtosis is {} of fuel quantity burned'.format(skw,kur))
```
```
The skewness is 15.850699077422112 and  the kurtosis is 651.2589365474387 of fuel quantity burned
```

In [24]:
```
missing=[col for col in project.columns if project[col].isnull().any()]
```

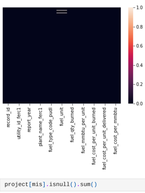
In [25]:
```
miss=project.dropna(axis=1)
```

In [26]:
```
mis=[]
for col in project.columns:
    if col not in project.dropna(axis=1).columns:
        mis.append(col)
        print(col)
```
```
fuel_unit
```

In [27]:
```
import seaborn as sns
```

In [28]:
```
sns.heatmap(project.isnull(),yticklabels=False)
```

Out[28]:
```
<AxesSubplot:>
```



In [29]:
```
project[mis].isnull().sum()
```

Out[29]:
```
fuel_unit    180
dtype: int64
```

In [30]:
```
project[mis].isnull().count()
```

Out[30]:
```
fuel_unit    29523
dtype: int64
```

In [31]:
```
feauture='fuel_price'
total=29523

percentage=(180/total)*100
```

In [32]:
```
percentage
```

Out[32]:
```
0.6096944136774718
```

In [33]:
```
# CATEGORICAL AND MEAN IMPUTATION
pd.get_dummies(project[mis])
```

Out[33]:

	fuel_unit_btu	fuel_unit_gal	fuel_unit_gramsU	fuel_unit_kgU	fuel_unit_mcf	fuel_unit_mmbtu	fuel_unit_mwdth	fuel_unit_mwhth	fuel_unit_ton
0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	1
4	0	0	0	0	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...
29518	0	0	0	0	1	0	0	0	0
29519	0	0	0	0	1	0	0	0	0
29520	0	0	0	0	1	0	0	0	0
29521	0	0	0	0	0	0	0	0	1
29522	0	0	0	0	1	0	0	0	0

29523 rows × 9 columns

In [34]:
```
project.corr()
```

Out[34]:

	utility_id_ferc1	report_year	fuel_qty_burned	fuel_mmbtu_per_unit	fuel_cost_per_unit_burned	fuel_cost_per_unit_delivered	fuel_cost_per_mmbtu
utility_id_ferc1	1.000000	0.093323	-0.057447	-0.066946	-0.037863	-0.016414	0.006122
report_year	0.093323	1.000000	0.012952	-0.110853	0.013599	-0.014043	0.010261
fuel_qty_burned	-0.057447	0.012952	1.000000	-0.080946	-0.018535	-0.003551	-0.001896
fuel_mmbtu_per_unit	-0.066946	-0.110853	-0.080946	1.000000	-0.010034	-0.009039	-0.005884
fuel_cost_per_unit_burned	-0.037863	0.013599	-0.018535	-0.010034	1.000000	0.011007	-0.000437
fuel_cost_per_unit_delivered	-0.016414	-0.014043	-0.003551	-0.009039	0.011007	1.000000	-0.000109
fuel_cost_per_mmbtu	0.006122	0.010261	-0.001896	-0.005884	-0.000437	-0.000109	1.000000

In [35]:
```
projectct.head()
```

Out[35]:

	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
0	f1_fuel_1994_12_1_0_7	1	1994	rockport	coal	ton	5377489.0	16.590	18.59	18.53	1.121
1	f1_fuel_1994_12_1_0_10	1	1994	rockport total plant	coal	ton	10486945.0	16.592	18.58	18.53	1.120
2	f1_fuel_1994_12_2_0_1	2	1994	gorgas	coal	ton	2978683.0	24.130	39.72	38.12	1.650
3	f1_fuel_1994_12_2_0_7	2	1994	barry	coal	ton	3739484.0	23.950	47.21	45.99	1.970
4	f1_fuel_1994_12_2_0_10	2	1994	chickasaw	gas	mcf	40533.0	1.000	2.77	2.77	2.570

In [37]:
```
coal=project[project['fuel_type_code_pudl']=='coal']
```

In [38]:
```
coal.head()
```

Out[38]:

	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
0	f1_fuel_1994_12_1_0_7	1	1994	rockport	coal	ton	5377489.0	16.590	18.59	18.53	1.121
1	f1_fuel_1994_12_1_0_10	1	1994	rockport total plant	coal	ton	10486945.0	16.592	18.58	18.53	1.120
2	f1_fuel_1994_12_2_0_1	2	1994	gorgas	coal	ton	2978683.0	24.130	39.72	38.12	1.650
3	f1_fuel_1994_12_2_0_7	2	1994	barry	coal	ton	3739484.0	23.950	47.21	45.99	1.970
5	f1_fuel_1994_12_2_0_13	2	1994	e. c. gaston unit 5	coal	ton	2124933.0	23.922	44.24	43.25	1.850

In [39]:
```
coal_94=coal[coal['report_year']==1994]
```

In [40]:
```
coal_94
```

Out[40]:

	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
0	f1_fuel_1994_12_1_0_7	1	1994	rockport	coal	ton	5377489.0	16.590	18.59	18.53	1.121
1	f1_fuel_1994_12_1_0_10	1	1994	rockport total plant	coal	ton	10486945.0	16.592	18.58	18.53	1.120
2	f1_fuel_1994_12_2_0_1	2	1994	gorgas	coal	ton	2978683.0	24.130	39.72	38.12	1.650
3	f1_fuel_1994_12_2_0_7	2	1994	barry	coal	ton	3739484.0	23.950	47.21	45.99	1.970
5	f1_fuel_1994_12_2_0_13	2	1994	e. c. gaston unit 5	coal	ton	2124933.0	23.922	44.24	43.25	1.850
...	...	...	...	...	...	...	...	...	...	...	...
1224	f1_fuel_1994_12_45_0_10	45	1994	lee	coal	ton	228247.0	25.204	37.78	45.76	1.499
1227	f1_fuel_1994_12_45_1_1	45	1994	dan river	coal	ton	165067.0	24.704	38.23	38.56	1.548
1229	f1_fuel_1994_12_45_1_7	45	1994	cliffside	coal	ton	862674.0	25.078	40.97	41.00	1.634
1230	f1_fuel_1994_12_45_1_10	45	1994	riverbend	coal	ton	396121.0	24.952	39.95	43.70	1.601
1232	f1_fuel_1994_12_45_2_1	45	1994	buck	coal	ton	205382.0	24.706	39.19	39.37	1.586

475 rows × 11 columns

In [41]:
```
coal_98=coal[coal['report_year']==1998]
```

In [42]:
```
coal_98
```

Out[42]:

	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
4618	f1_fuel_1998_12_1_0_7	1	1998	rockport total avg	coal	ton	4965881.0	17.198	19.282	19.338	1.126
4619	f1_fuel_1998_12_1_0_10	1	1998	rockport total plant	coal	ton	10000938.0	17.196	19.280	19.340	1.122
4620	f1_fuel_1998_12_4_0_1	4	1998	units 1 - 3	coal	ton	1974682.0	22.266	26.550	26.540	1.190
4622	f1_fuel_1998_12_4_0_4	4	1998	unit 4	coal	ton	385512.0	22.330	26.550	26.540	1.190
4624	f1_fuel_1998_12_6_0_1	6	1998	clinch river	coal	ton	1834553.0	24.534	30.936	30.013	1.261
...	...	...	...	...	...	...	...	...	...	...	...
5708	f1_fuel_1998_12_145_0_10	145	1998	comanche	coal	ton	2732360.0	17.134	15.151	15.035	0.884
5710	f1_fuel_1998_12_145_0_13	145	1998	craig	coal	ton	3246889.0	20.058	21.055	21.386	1.050
5712	f1_fuel_1998_12_145_1_1	145	1998	hayden	coal	ton	802459.0	21.092	22.596	22.691	1.071
5714	f1_fuel_1998_12_145_1_4	145	1998	pawnee	coal	ton	2104342.0	16.684	14.035	13.596	0.842
5716	f1_fuel_1998_12_145_1_7	145	1998	valmont 5	coal	ton	592834.0	21.842	24.875	25.128	1.139

431 rows × 11 columns

In [46]:
```
project[project['fuel_cost_per_unit_delivered']==project['fuel_cost_per_unit_delivered'].max()]
```

Out[46]:

	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
3564	f1_fuel_1997_12_9_0_8	9	1997	peach bt 2&3	nuclear	gramsU	210474.0	0.000065	37.847	7964521.0	0.578

In [48]:
```
#1997
```

In [ ]:
```

```