

Importing necessary packages

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import pickle
import datetime
```

Importing the Dataset

```
data = pd.read_csv("BCG_Final Dataset.csv")
```

```
data.isna().sum()
```

```
Unnamed: 0      0
id              0
cons_12m        0
cons_gas_12m    0
cons_last_month 0
date_activ      0
date_end        0
date_modif_prod 0
date_renewal     0
forecast_cons_12m 0
forecast_cons_year 0
forecast_discount_energy 0
forecast_meter_rent_12m 0
forecast_price_energy_p1 0
forecast_price_energy_p2 0
forecast_price_pow_p1 0
has_gas         0
imp_cons        0
margin_gross_pow_ele 0
margin_net_pow_ele 0
nb_prod_act     0
net_margin      0
num_years_antig 0
origin_up       0
pow_max         0
churn           0
dtype: int64
```

Dropping Id column

```
data1 = data.drop(['id'], axis = 1)
```

```
data1.isna().sum()
```

```

Unnamed: 0      0
cons_12m        0
cons_gas_12m    0
cons_last_month 0
date_activ      0
date_end        0
date_modif_prod 0
date_renewal    0
forecast_cons_12m 0
forecast_cons_year 0
forecast_discount_energy 0
forecast_meter_rent_12m 0
forecast_price_energy_p1 0
forecast_price_energy_p2 0
forecast_price_pow_p1 0
has_gas         0
imp_cons        0
margin_gross_pow_ele 0
margin_net_pow_ele 0
nb_prod_act     0
net_margin      0
num_years_antig 0
origin_up       0
pow_max         0
churn           0
dtype: int64

```

Creating feature matrix for independent and dependent variables

```
X = data1.iloc[:, 1:24]
```

```
Y = data1['churn']
```

```
data1 = data1.drop(['Unnamed: 0'], axis = 1)
```

```
data1
```

	cons_12m	cons_gas_12m	cons_last_month	date_activ	date_end	date_modif_prod
0	309275	0	10025	2012-11-07	2016-11-06	2012-11-07
1	4660	0	0	2009-08-21	2016-08-30	2009-08-21
2	544	0	0	2010-04-16	2016-04-16	2010-04-16
3	1584	0	0	2010-03-30	2016-03-30	2010-03-30
4	121335	0	12400	2010-04-08	2016-04-08	2010-04-08
...
15669	32270	47940	0	2012-05-24	2016-05-08	2015-05-08
					2016-05-08	

Converting Date variables into Datetime type

```

15674      4844      0      470  2012-02-08  2016-02-08  2012-02-08
from datetime import date

15672      131      0      0  2012-08-30  2016-08-30  2012-08-30
data1['date_activ'] = pd.to_datetime(data1.date_activ)

15673      8730      0      0  2009-12-18  2016-12-18  2009-12-18
data1['date_end'] = pd.to_datetime(data1.date_end)

data1['date_renewal'] = pd.to_datetime(data1.date_renewal)

```

Adding column to display number of months left before renewal

```
act_months_left'] = (data1['date_end'].dt.year - data1['date_activ'].dt.year)*12 + data1['dat
```

```
data1['contract_months_left']
```

```

0      48
1      84
2      72
3      72
4      72
..
15669  48
15670  48
15671  48
15672  48

```

15673 84

Name: contract months left, Length: 15674, dtype: int64

data1

	cons_12m	cons_gas_12m	cons_last_month	date_activ	date_end	date_modif_prod
0	309275	0	10025	2012-11-07	2016-11-06	2012-11-07
1	4660	0	0	2009-08-21	2016-08-30	2009-08-21
2	544	0	0	2010-04-16	2016-04-16	2010-04-16
3	1584	0	0	2010-03-30	2016-03-30	2010-03-30
4	121335	0	12400	2010-04-08	2016-04-08	2010-04-08
...
15669	32270	47940	0	2012-05-24	2016-05-08	2015-05-08
15670	7223	0	181	2012-08-27	2016-08-27	2012-08-27
15671	1844	0	179	2012-02-08	2016-02-07	2012-02-08
15672	131	0	0	2012-08-30	2016-08-30	2012-08-30
15673	8730	0	0	2009-12-18	2016-12-17	2009-12-18

15674 rows × 7 columns

cons_gas_12m', 'cons_last_month', 'date_activ', 'date_end', 'date_modif_prod', 'date_renewal

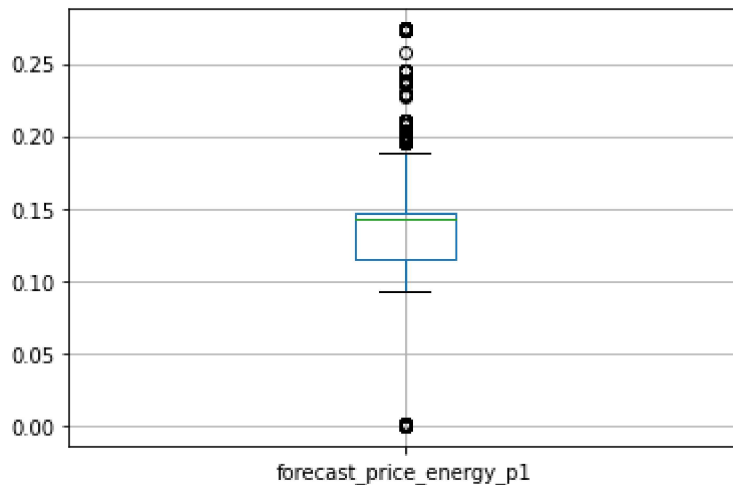
data1

energy_p2	forecast_price_pow_p1	has_gas	imp_cons	margin_gross_pow_ele	margin_net_po
0.088347	58.995952	No	831.80		41.76
0.000000	44.311378	No	0.00		16.38
0.087899	44.311378	No	0.00		28.60
0.000000	44.311378	No	0.00		30.22
0.093746	40.606701	No	1052.37		3.18
...
0.000000	44.311378	Yes	0.00		27.88
0.091892	58.995952	No	15.94		0.00

Identifying Outliers

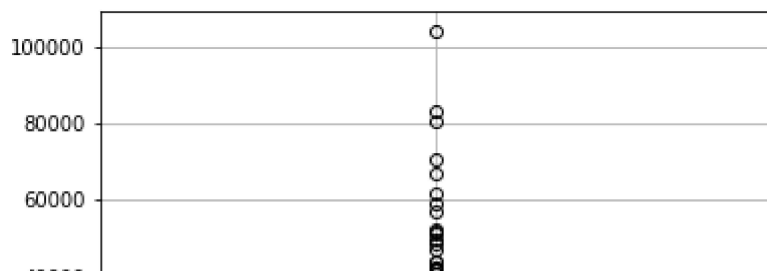
```
data.boxplot(column = 'forecast_price_energy_p1')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f705d7e3d50>



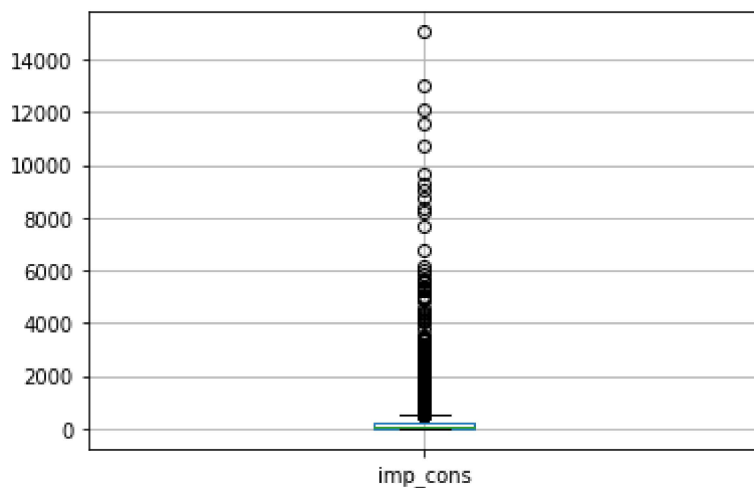
```
data.boxplot(column = 'forecast_cons_12m')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705c2adb90>
```



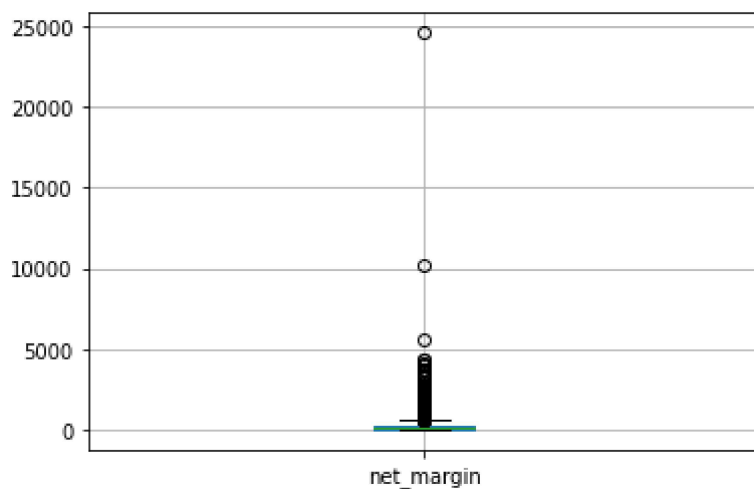
```
data.boxplot(column = 'imp_cons')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705c2b7050>
```



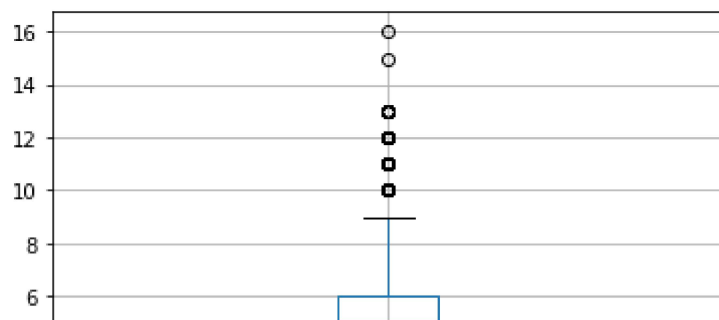
```
data.boxplot(column = 'net_margin')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705bd96690>
```



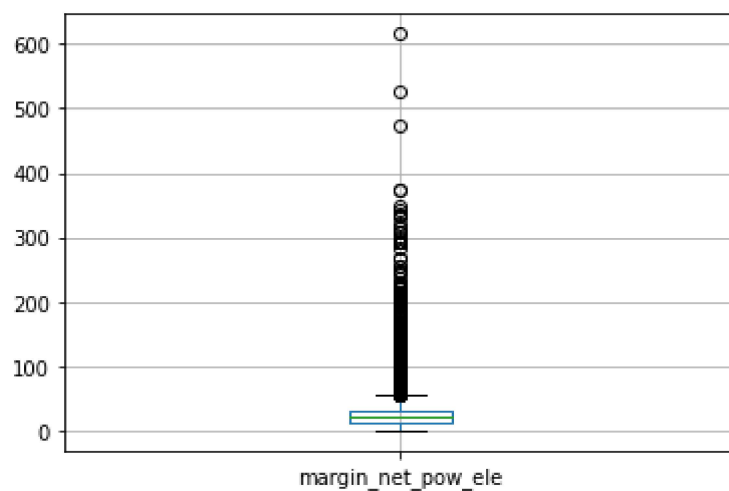
```
data.boxplot(column = 'num_years_antig')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705bcc8810>
```



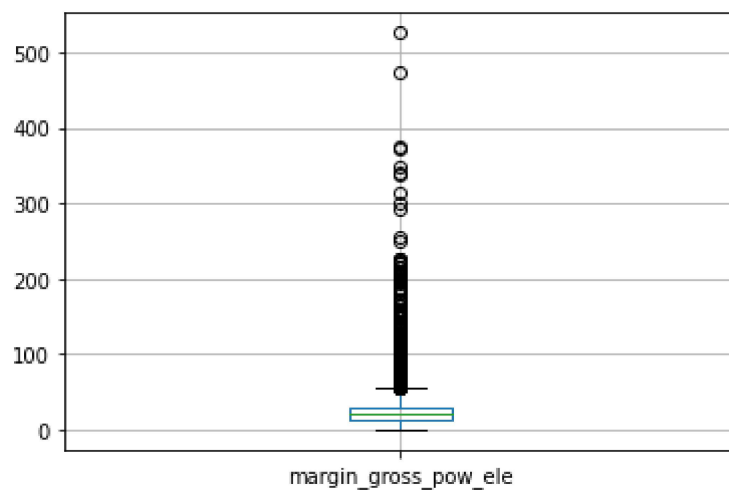
```
data.boxplot(column = 'margin_net_pow_ele')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705bc42e10>
```



```
data.boxplot(column = 'margin_gross_pow_ele')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f705bd5b890>
```



✓ 0s completed at 7:45 PM

● ✕