In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [4]:
```
# Load data
client_data = pd.read_csv('client_data (1).csv')
price_data = pd.read_csv('price_data (1).csv')
```

In [7]:
```
client_data.head()
```

Out[7]:

| | id | channel_sales | cons_12m | cons_gas_12m | cons_last_month | date_activ |
|---|---|---|---|---|---|---|
| 0 | 24011ae4ebbe3035111d65fa7c15bc57 | foosdfpfkusacimwkcsosbicdxkicaua | 0 | 54946 | 0 | 2013-06-15 |
| 1 | d29c2c54acc38ff3c0614d0a653813dd | MISSING | 4660 | 0 | 0 | 2009-08-21 |
| 2 | 764c75f661154dac3a6c254cd082ea7d | foosdfpfkusacimwkcsosbicdxkicaua | 544 | 0 | 0 | 2010-04-16 |
| 3 | bba03439a292a1e166f80264c16191cb | lmkebamcaaclubfxadlmueccxoimlema | 1584 | 0 | 0 | 2010-03-30 |
| 4 | 149d57cf92fc41cf94415803a877cb4b | MISSING | 4425 | 0 | 526 | 2010-01-13 |

5 rows × 26 columns

In [8]:
```
price_data.head()
```

Out[8]:

| | id | price_date | price_off_peak_var | price_peak_var | price_mid_peak_var | price_off_peak_fix | price |
|---|---|---|---|---|---|---|---|
| 0 | 038af19179925da21a25619c5a24b745 | 2015-01-01 | 0.151367 | 0.0 | 0.0 | 44.266931 | |
| 1 | 038af19179925da21a25619c5a24b745 | 2015-02-01 | 0.151367 | 0.0 | 0.0 | 44.266931 | |
| 2 | 038af19179925da21a25619c5a24b745 | 2015-03-01 | 0.151367 | 0.0 | 0.0 | 44.266931 | |
| 3 | 038af19179925da21a25619c5a24b745 | 2015-04-01 | 0.149626 | 0.0 | 0.0 | 44.266931 | |
| 4 | 038af19179925da21a25619c5a24b745 | 2015-05-01 | 0.149626 | 0.0 | 0.0 | 44.266931 | |

In [9]:
```
client_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14606 entries, 0 to 14605
Data columns (total 26 columns):
 #   Column                           Non-Null Count  Dtype  
---  ------                           --------------  -----  
 0   id                               14606 non-null  object 
 1   channel_sales                    14606 non-null  object 
 2   cons_12m                         14606 non-null  int64  
 3   cons_gas_12m                     14606 non-null  int64  
 4   cons_last_month                  14606 non-null  int64  
 5   date_activ                       14606 non-null  object 
 6   date_end                         14606 non-null  object 
 7   date_modif_prod                  14606 non-null  object 
 8   date_renewal                     14606 non-null  object 
 9   forecast_cons_12m                14606 non-null  float64
 10  forecast_cons_year               14606 non-null  int64  
 11  forecast_discount_energy         14606 non-null  float64
 12  forecast_meter_rent_12m          14606 non-null  float64
 13  forecast_price_energy_off_peak   14606 non-null  float64
 14  forecast_price_energy_peak       14606 non-null  float64
 15  forecast_price_pow_off_peak      14606 non-null  float64
 16  has_gas                          14606 non-null  object 
 17  imp_cons                         14606 non-null  float64
 18  margin_gross_pow_ele             14606 non-null  float64
 19  margin_net_pow_ele               14606 non-null  float64
 20  nb_prod_act                      14606 non-null  int64  
 21  net_margin                       14606 non-null  float64
 22  num_years_antig                  14606 non-null  int64  
 23  origin_up                        14606 non-null  object 
 24  pow_max                          14606 non-null  float64
 25  churn                            14606 non-null  int64  
dtypes: float64(11), int64(7), object(8)
memory usage: 2.9+ MB
```

In [10]:
```
client_data.dtypes
```

Out[10]:
```
id                                object
channel_sales                     object
cons_12m                           int64
cons_gas_12m                       int64
cons_last_month                    int64
date_activ                        object
date_end                          object
date_modif_prod                   object
date_renewal                      object
forecast_cons_12m                float64
forecast_cons_year                 int64
forecast_discount_energy         float64
forecast_meter_rent_12m          float64
forecast_price_energy_off_peak   float64
forecast_price_energy_peak       float64
forecast_price_pow_off_peak      float64
has_gas                           object
imp_cons                         float64
margin_gross_pow_ele             float64
margin_net_pow_ele               float64
nb_prod_act                        int64
net_margin                       float64
num_years_antig                    int64
origin_up                         object
pow_max                          float64
churn                              int64
dtype: object
```

In [12]:
```
print(client_data.describe())
```

```
           cons_12m  cons_gas_12m  cons_last_month  forecast_cons_12m  \
count  1.460600e+04  1.460600e+04     14606.000000       14606.000000
mean   1.592203e+05  2.809238e+04     16090.269752        1868.614880
std    5.734653e+05  1.629731e+05     64364.196422        2387.571531
min    0.000000e+00  0.000000e+00         0.000000           0.000000
25%    5.674750e+03  0.000000e+00         0.000000         494.995000
50%    1.411550e+04  0.000000e+00       792.500000        1112.875000
75%    4.076375e+04  0.000000e+00      3383.000000        2401.790000
max    6.207104e+06  4.154590e+06    771203.000000       82902.830000

       forecast_cons_year  forecast_discount_energy  forecast_meter_rent_12m  \
count        14606.000000              14606.000000             14606.000000
mean          1399.762906                  0.966726                63.086871
std           3247.786255                  5.108289                66.165783
min              0.000000                  0.000000                 0.000000
25%              0.000000                  0.000000                16.180000
50%            314.000000                  0.000000                18.795000
75%           1745.750000                  0.000000               131.030000
max         175375.000000                 30.000000               599.310000

       forecast_price_energy_off_peak  forecast_price_energy_peak  \
count                    14606.000000                14606.000000
mean                         0.137283                    0.050491
std                          0.024623                    0.049037
min                          0.000000                    0.000000
25%                          0.116340                    0.000000
50%                          0.143166                    0.084138
75%                          0.146348                    0.098837
max                          0.273963                    0.195975

       forecast_price_pow_off_peak      imp_cons  margin_gross_pow_ele  \
count                 14606.000000  14606.000000          14606.000000
mean                     43.130056    152.786896             24.565121
std                       4.485988    341.369366             20.231172
min                       0.000000      0.000000              0.000000
25%                      40.606701      0.000000             14.280000
50%                      44.311378     37.395000             21.640000
75%                      44.311378    193.980000             29.880000
max                      59.266378  15042.790000            374.640000

       margin_net_pow_ele   nb_prod_act    net_margin  num_years_antig  \
count        14606.000000  14606.000000  14606.000000     14606.000000
mean            24.562517      1.292346    189.264522         4.997809
std             20.230280      0.709774    311.798130         1.611749
min              0.000000      1.000000      0.000000         1.000000
25%             14.280000      1.000000     50.712500         4.000000
50%             21.640000      1.000000    112.530000         5.000000
75%             29.880000      1.000000    243.097500         6.000000
max            374.640000     32.000000  24570.650000        13.000000

            pow_max         churn
count  14606.000000  14606.000000
mean      18.135136      0.097152
std       13.534743      0.296175
min        3.300000      0.000000
25%       12.500000      0.000000
50%       13.856000      0.000000
75%       19.172500      0.000000
max      320.000000      1.000000
```

In [13]:
```python
# Initial look
print("Client Data Info:\n")
client_data.info()
print("\nClient Data Preview:\n", client_data.head())

print("\nPrice Data Info:\n")
price_data.info()
print("\nPrice Data Preview:\n", price_data.head())

# Descriptive statistics
print("\nClient Data Description:\n", client_data.describe(include='all'))
print("\nPrice Data Description:\n", price_data.describe(include='all'))

# Unique value counts
print("\nUnique values in client_data:\n", client_data.nunique())
print("\nUnique values in price_data:\n", price_data.nunique())
```

```
Client Data Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14606 entries, 0 to 14605
Data columns (total 26 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
```

```
 0   id                              14606 non-null  object
 1   channel_sales                   14606 non-null  object
 2   cons_12m                        14606 non-null  int64
 3   cons_gas_12m                    14606 non-null  int64
 4   cons_last_month                 14606 non-null  int64
 5   date_activ                      14606 non-null  object
 6   date_end                        14606 non-null  object
 7   date_modif_prod                 14606 non-null  object
 8   date_renewal                    14606 non-null  object
 9   forecast_cons_12m               14606 non-null  float64
 10  forecast_cons_year              14606 non-null  int64
 11  forecast_discount_energy        14606 non-null  float64
 12  forecast_meter_rent_12m         14606 non-null  float64
 13  forecast_price_energy_off_peak  14606 non-null  float64
 14  forecast_price_energy_peak      14606 non-null  float64
 15  forecast_price_pow_off_peak     14606 non-null  float64
 16  has_gas                         14606 non-null  object
 17  imp_cons                        14606 non-null  float64
 18  margin_gross_pow_ele            14606 non-null  float64
 19  margin_net_pow_ele              14606 non-null  float64
 20  nb_prod_act                     14606 non-null  int64
 21  net_margin                      14606 non-null  float64
 22  num_years_antig                 14606 non-null  int64
 23  origin_up                       14606 non-null  object
 24  pow_max                         14606 non-null  float64
 25  churn                           14606 non-null  int64
dtypes: float64(11), int64(7), object(8)
memory usage: 2.9+ MB

Client Data Preview:
                                 id                    channel_sales  \
0  24011ae4ebbe3035111d65fa7c15bc57  foosdfpfkusacimwkcsosbicdxkicaua
1  d29c2c54acc38ff3c0614d0a653813dd                           MISSING
2  764c75f661154dac3a6c254cd082ea7d  foosdfpfkusacimwkcsosbicdxkicaua
3  bba03439a292a1e166f80264c16191cb  lmkebamcaaclubfxadlmueccxoimlema
4  149d57cf92fc41cf94415803a877cb4b                           MISSING

   cons_12m  cons_gas_12m  cons_last_month  date_activ    date_end  \
0         0         54946                0  2013-06-15  2016-06-15
1      4660             0                0  2009-08-21  2016-08-30
2       544             0                0  2010-04-16  2016-04-16
3      1584             0                0  2010-03-30  2016-03-30
4      4425             0              526  2010-01-13  2016-03-07

  date_modif_prod date_renewal  forecast_cons_12m  ...  has_gas  imp_cons  \
0      2015-11-01   2015-06-23               0.00  ...        t      0.00
1      2009-08-21   2015-08-31             189.95  ...        f      0.00
2      2010-04-16   2015-04-17              47.96  ...        f      0.00
3      2010-03-30   2015-03-31             240.04  ...        f      0.00
4      2010-01-13   2015-03-09             445.75  ...        f     52.32

   margin_gross_pow_ele  margin_net_pow_ele  nb_prod_act  net_margin  \
0                 25.44               25.44            2      678.99
1                 16.38               16.38            1       18.89
2                 28.60               28.60            1        6.60
3                 30.22               30.22            1       25.46
4                 44.91               44.91            1       47.98

  num_years_antig                        origin_up  pow_max  churn
0               3  lxidpiddsbxsbosboudacockeimpuepw   43.648      1
1               6  kamkkxfxxuwbdslkwifmmcsiusiuosws   13.800      0
2               6  kamkkxfxxuwbdslkwifmmcsiusiuosws   13.856      0
3               6  kamkkxfxxuwbdslkwifmmcsiusiuosws   13.200      0
4               6  kamkkxfxxuwbdslkwifmmcsiusiuosws   19.800      0

[5 rows x 26 columns]

Price Data Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 193002 entries, 0 to 193001
Data columns (total 8 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   id                  193002 non-null  object
 1   price_date          193002 non-null  object
 2   price_off_peak_var  193002 non-null  float64
 3   price_peak_var      193002 non-null  float64
 4   price_mid_peak_var  193002 non-null  float64
 5   price_off_peak_fix  193002 non-null  float64
 6   price_peak_fix      193002 non-null  float64
 7   price_mid_peak_fix  193002 non-null  float64
dtypes: float64(6), object(2)
```

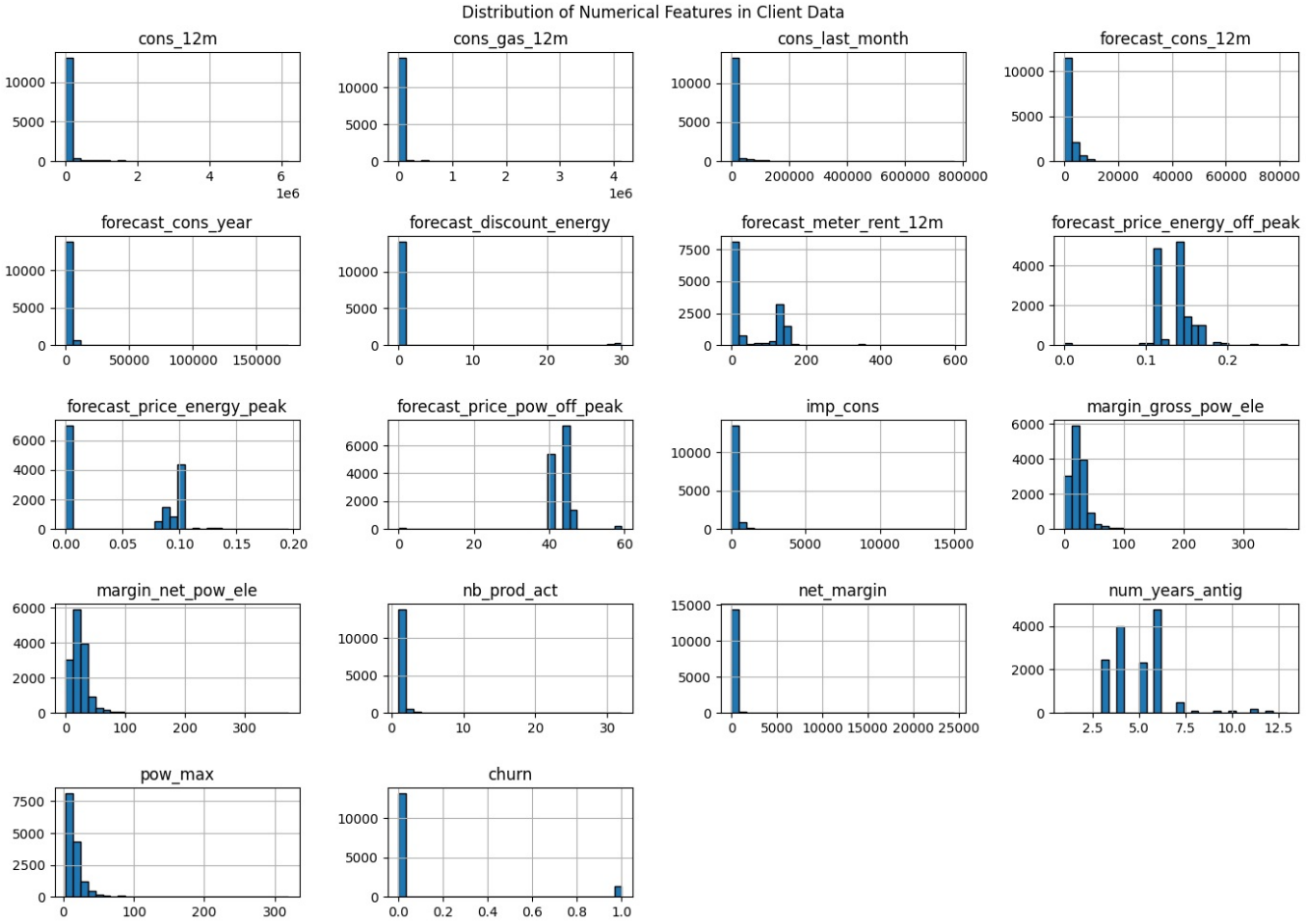
```
memory usage: 11.8+ MB

Price Data Preview:
                                id  price_date  price_off_peak_var  \
0  038af19179925da21a25619c5a24b745  2015-01-01            0.151367
1  038af19179925da21a25619c5a24b745  2015-02-01            0.151367
2  038af19179925da21a25619c5a24b745  2015-03-01            0.151367
3  038af19179925da21a25619c5a24b745  2015-04-01            0.149626
4  038af19179925da21a25619c5a24b745  2015-05-01            0.149626

   price_peak_var  price_mid_peak_var  price_off_peak_fix  price_peak_fix  \
0             0.0                 0.0           44.266931             0.0
1             0.0                 0.0           44.266931             0.0
2             0.0                 0.0           44.266931             0.0
3             0.0                 0.0           44.266931             0.0
4             0.0                 0.0           44.266931             0.0

   price_mid_peak_fix
0                 0.0
1                 0.0
2                 0.0
3                 0.0
4                 0.0

Client Data Description:
                                     id                    channel_sales  \
count                             14606                            14606
unique                            14606                                8
top     563dde550fd624d7352f3de77c0cdfcd  foosdfpfkusacimwkcsosbicdxkicaua
freq                                  1                             6754
mean                                NaN                              NaN
std                                 NaN                              NaN
min                                 NaN                              NaN
25%                                 NaN                              NaN
50%                                 NaN                              NaN
75%                                 NaN                              NaN
max                                 NaN                              NaN

            cons_12m  cons_gas_12m  cons_last_month  date_activ    date_end  \
count   1.460600e+04  1.460600e+04     14606.000000       14606       14606
unique           NaN           NaN              NaN        1796         368
top              NaN           NaN              NaN  2009-08-01  2016-02-01
freq             NaN           NaN              NaN          95         145
mean    1.592203e+05  2.809238e+04     16090.269752         NaN         NaN
std     5.734653e+05  1.629731e+05     64364.196422         NaN         NaN
min     0.000000e+00  0.000000e+00         0.000000         NaN         NaN
25%     5.674750e+03  0.000000e+00         0.000000         NaN         NaN
50%     1.411550e+04  0.000000e+00       792.500000         NaN         NaN
75%     4.076375e+04  0.000000e+00      3383.000000         NaN         NaN
max     6.207104e+06  4.154590e+06    771203.000000         NaN         NaN

       date_modif_prod date_renewal  forecast_cons_12m  ...  has_gas  \
count            14606        14606       14606.000000  ...    14606
unique            2129          386                NaN  ...        2
top         2015-11-01   2015-06-23                NaN  ...        f
freq               721          587                NaN  ...    11955
mean               NaN          NaN        1868.614880  ...      NaN
std                NaN          NaN        2387.571531  ...      NaN
min                NaN          NaN           0.000000  ...      NaN
25%                NaN          NaN         494.995000  ...      NaN
50%                NaN          NaN        1112.875000  ...      NaN
75%                NaN          NaN        2401.790000  ...      NaN
max                NaN          NaN       82902.830000  ...      NaN

            imp_cons  margin_gross_pow_ele  margin_net_pow_ele   nb_prod_act  \
count   14606.000000          14606.000000        14606.000000  14606.000000
unique           NaN                   NaN                 NaN           NaN
top              NaN                   NaN                 NaN           NaN
freq             NaN                   NaN                 NaN           NaN
mean      152.786896             24.565121           24.562517      1.292346
std       341.369366             20.231172           20.230280      0.709774
min         0.000000              0.000000            0.000000      1.000000
25%         0.000000             14.280000           14.280000      1.000000
50%        37.395000             21.640000           21.640000      1.000000
75%       193.980000             29.880000           29.880000      1.000000
max     15042.790000            374.640000          374.640000     32.000000

          net_margin num_years_antig                          origin_up  \
count   14606.000000    14606.000000                              14606
unique           NaN             NaN                                  6
top              NaN             NaN  lxidpiddsbxsbosboudacockeimpuepw
freq             NaN             NaN                               7097
mean      189.264522        4.997809                                NaN
```

```
std       311.798130        1.611749                            NaN
min         0.000000        1.000000                            NaN
25%        50.712500        4.000000                            NaN
50%       112.530000        5.000000                            NaN
75%       243.097500        6.000000                            NaN
max     24570.650000       13.000000                            NaN

            pow_max         churn
count   14606.000000  14606.000000
unique           NaN           NaN
top              NaN           NaN
freq             NaN           NaN
mean       18.135136      0.097152
std        13.534743      0.296175
min         3.300000      0.000000
25%        12.500000      0.000000
50%        13.856000      0.000000
75%        19.172500      0.000000
max       320.000000      1.000000

[11 rows x 26 columns]

Price Data Description:
                                     id  price_date  price_off_peak_var  \
count                            193002      193002       193002.000000
unique                            16096          12                 NaN
top     c18b6305122e4976739b8420d5b54ec5  2015-12-01                 NaN
freq                                 12       16094                 NaN
mean                                NaN         NaN            0.141027
std                                 NaN         NaN            0.025032
min                                 NaN         NaN            0.000000
25%                                 NaN         NaN            0.125976
50%                                 NaN         NaN            0.146033
75%                                 NaN         NaN            0.151635
max                                 NaN         NaN            0.280700

        price_peak_var  price_mid_peak_var  price_off_peak_fix  \
count    193002.000000       193002.000000       193002.000000
unique             NaN                 NaN                 NaN
top                NaN                 NaN                 NaN
freq               NaN                 NaN                 NaN
mean          0.054630            0.030496           43.334477
std           0.049924            0.036298            5.410297
min           0.000000            0.000000            0.000000
25%           0.000000            0.000000           40.728885
50%           0.085483            0.000000           44.266930
75%           0.101673            0.072558           44.444710
max           0.229788            0.114102           59.444710

        price_peak_fix  price_mid_peak_fix
count    193002.000000       193002.000000
unique             NaN                 NaN
top                NaN                 NaN
freq               NaN                 NaN
mean         10.622875            6.409984
std          12.841895            7.773592
min           0.000000            0.000000
25%           0.000000            0.000000
50%           0.000000            0.000000
75%          24.339581           16.226389
max          36.490692           17.458221

Unique values in client_data:
 id                               14606
channel_sales                        8
cons_12m                         11065
cons_gas_12m                      2112
cons_last_month                   4751
date_activ                        1796
date_end                           368
date_modif_prod                   2129
date_renewal                       386
forecast_cons_12m                13993
forecast_cons_year                4218
forecast_discount_energy            12
forecast_meter_rent_12m           3528
forecast_price_energy_off_peak     516
forecast_price_energy_peak         329
forecast_price_pow_off_peak         41
has_gas                              2
imp_cons                          7752
margin_gross_pow_ele              2391
margin_net_pow_ele                2391
```

```
nb_prod_act                       10
net_margin                     11965
num_years_antig                   13
origin_up                          6
pow_max                          698
churn                              2
dtype: int64

Unique values in price_data:
 id                    16096
price_date               12
price_off_peak_var     1853
price_peak_var         1189
price_mid_peak_var      711
price_off_peak_fix       66
price_peak_fix           31
price_mid_peak_fix       28
dtype: int64
```

In [14]:

```python
# Convert date columns if needed
if 'purchase_date' in price_data.columns:
    price_data['purchase_date'] = pd.to_datetime(price_data['purchase_date'])

# Plot numerical distributions from client_data
num_cols = client_data.select_dtypes(include=['int64', 'float64']).columns
client_data[num_cols].hist(bins=30, figsize=(14, 10), edgecolor='black')
plt.suptitle('Distribution of Numerical Features in Client Data')
plt.tight_layout()
plt.show()

# Plot churn distribution if exists
if 'churned' in client_data.columns:
    sns.countplot(x='churned', data=client_data)
    plt.title("Churn Distribution")
    plt.show()
```

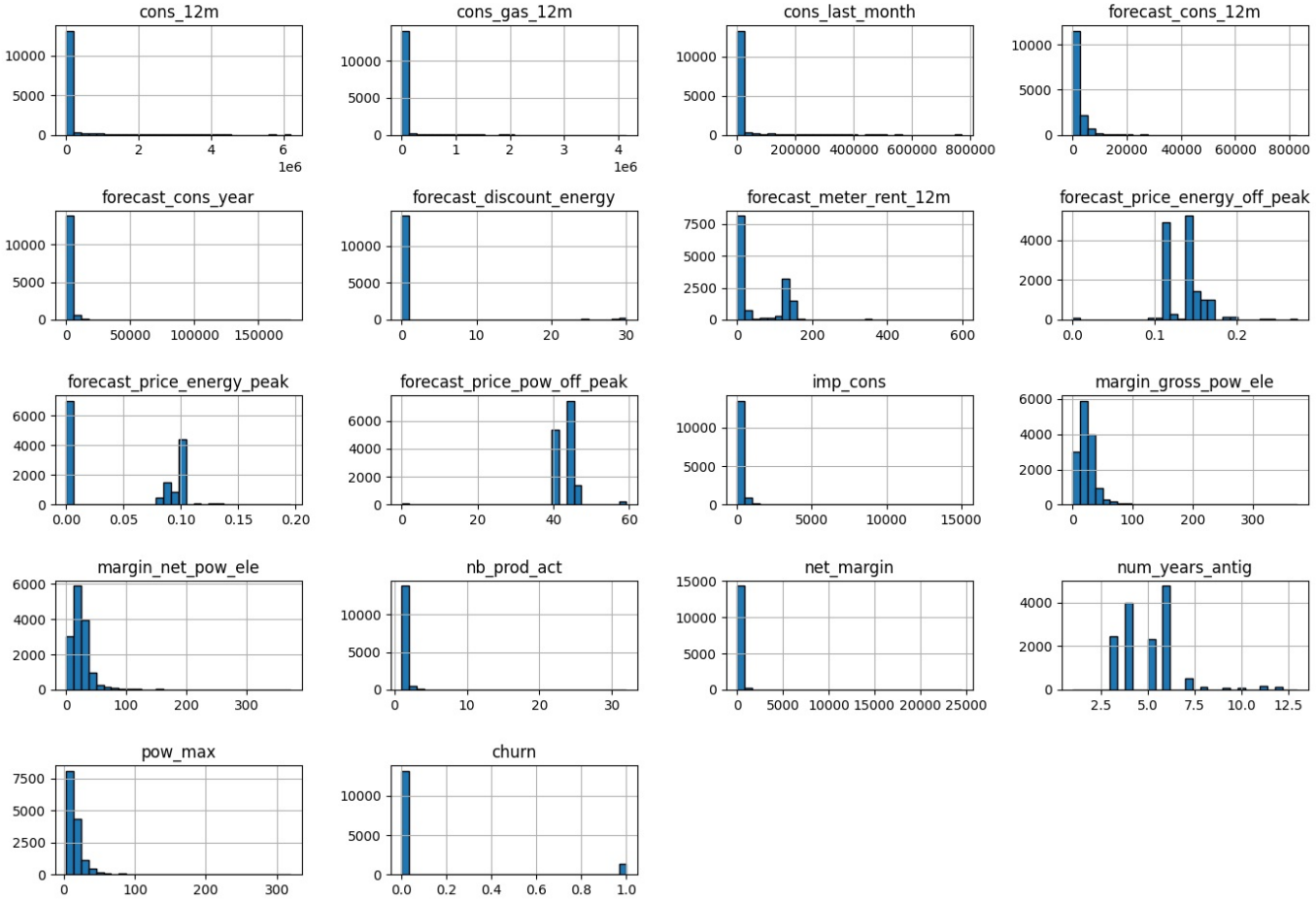



In [22]:

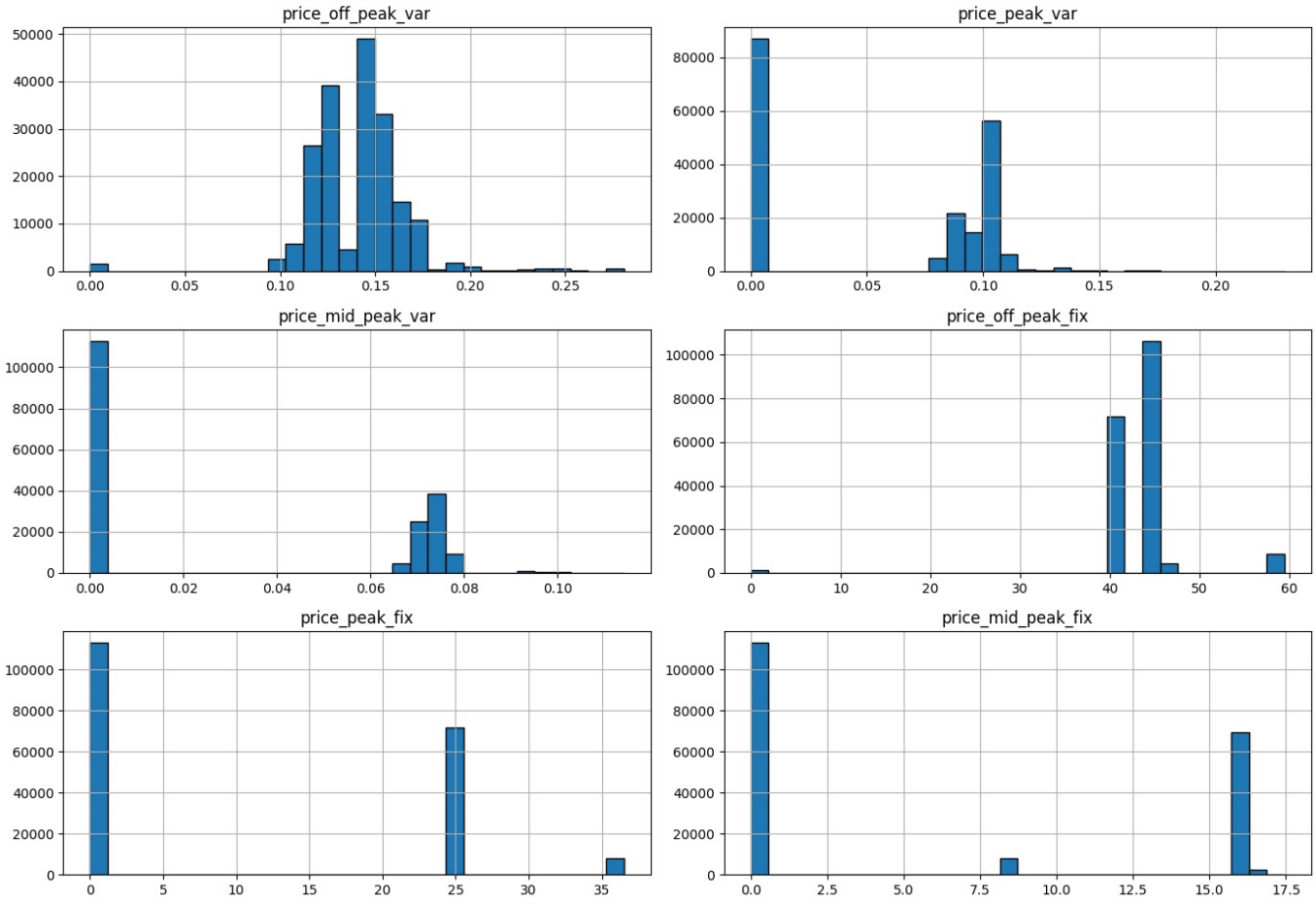
```python
def plot_distributions(df, title):
    num_cols = df.select_dtypes(include=['int64', 'float64']).columns
    df[num_cols].hist(bins=30, figsize=(14, 10), edgecolor='black')
    plt.suptitle(f'Distributions of Numeric Columns: {title}', fontsize=16)
    plt.tight_layout()
    plt.show()

plot_distributions(client_data, "client_data (1).csv")
plot_distributions(price_data, "price_data (1).csv")
```

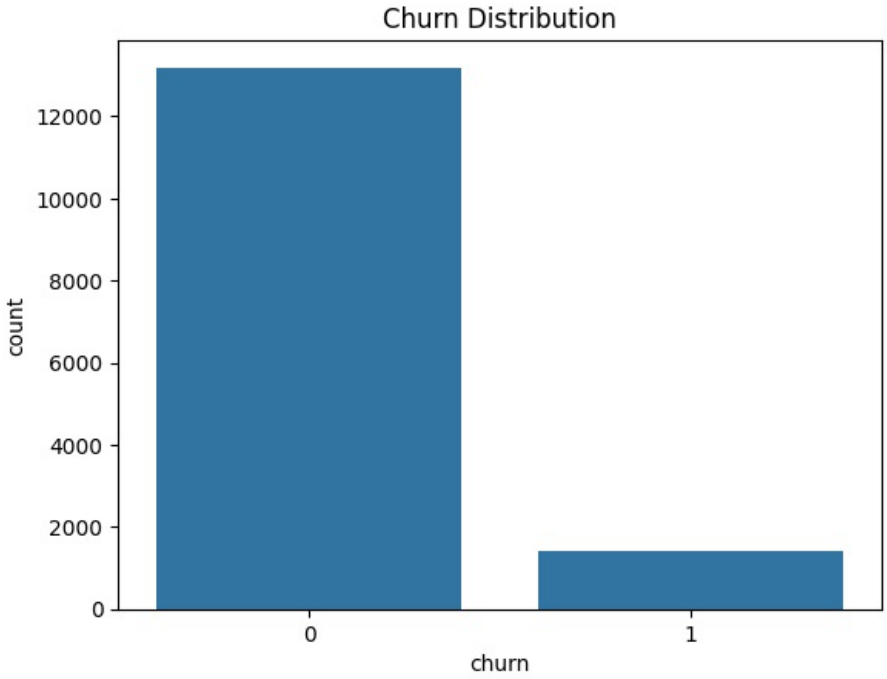
```python
#  8. Churn visualization (if it's inside client_data)
if 'churn' in client_data.columns or 'churned' in client_data.columns:
    churn_col = 'churn' if 'churn' in client_data.columns else 'churned'
    sns.countplot(x=churn_col, data=client_data)
    plt.title("Churn Distribution")
    plt.show()
```

Distributions of Numeric Columns: client_data (1).csv




Distributions of Numeric Columns: price_data (1).csv




Churn Distribution
12000
10000
8000
6000
4000
2000
0
count
0
1
churn

In [ ]: