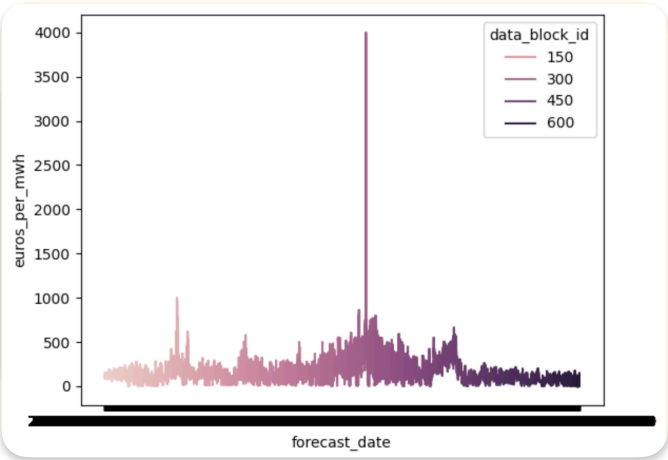


gas_prices.csv

| | forecast_date | lowest_price_per_mwh | highest_price_per_mwh | origin_date | data_block_id |
|---|---------------|----------------------|-----------------------|-------------|---------------|
| 0 | 2021-09-01 | 45.23 | 46.32 | 2021-08-31 | 1 |
| 1 | 2021-09-02 | 45.62 | 46.29 | 2021-09-01 | 2 |
| 2 | 2021-09-03 | 45.85 | 46.40 | 2021-09-02 | 3 |
| 3 | 2021-09-04 | 46.30 | 46.80 | 2021-09-03 | 4 |
| 4 | 2021-09-05 | 46.30 | 46.58 | 2021-09-04 | 5 |

electricity_prices.csv

| | forecast_date | euros_per_mwh | origin_date | data_block_id |
|---|---------------------|---------------|---------------------|---------------|
| 0 | 2021-09-01 00:00:00 | 92.51 | 2021-08-31 00:00:00 | 1 |
| 1 | 2021-09-01 01:00:00 | 88.90 | 2021-08-31 01:00:00 | 1 |
| 2 | 2021-09-01 02:00:00 | 87.35 | 2021-08-31 02:00:00 | 1 |
| 3 | 2021-09-01 03:00:00 | 86.88 | 2021-08-31 03:00:00 | 1 |
| 4 | 2021-09-01 04:00:00 | 88.43 | 2021-08-31 04:00:00 | 1 |
| 5 | 2021-09-01 05:00:00 | 93.58 | 2021-08-31 05:00:00 | 1 |
| 6 | 2021-09-01 06:00:00 | 118.70 | 2021-08-31 06:00:00 | 1 |
| 7 | 2021-09-01 07:00:00 | 135.44 | 2021-08-31 07:00:00 | 1 |
| 8 | 2021-09-01 08:00:00 | 140.72 | 2021-08-31 08:00:00 | 1 |
| 9 | 2021-09-01 09:00:00 | 130.00 | 2021-08-31 09:00:00 | 1 |



Merge on:
data_block_id
forecast_date

Merge on:
data_block_id
forecast_date

client.csv

| | product_type | county | elc_count | installed_capacity | is_business | date | data_block_id |
|---|--------------|--------|-----------|--------------------|-------------|------------|---------------|
| 0 | 1 | 0 | 108 | 952.89 | 0 | 2021-09-01 | 2 |
| 1 | 2 | 0 | 17 | 166.40 | 0 | 2021-09-01 | 2 |
| 2 | 3 | 0 | 688 | 7207.88 | 0 | 2021-09-01 | 2 |
| 3 | 0 | 0 | 5 | 400.00 | 1 | 2021-09-01 | 2 |
| 4 | 1 | 0 | 43 | 1411.00 | 1 | 2021-09-01 | 2 |
| 5 | 3 | 0 | 265 | 13417.79 | 1 | 2021-09-01 | 2 |
| 6 | 1 | 1 | 7 | 80.00 | 0 | 2021-09-01 | 2 |
| 7 | 3 | 1 | 42 | 472.60 | 0 | 2021-09-01 | 2 |
| 8 | 3 | 1 | 12 | 295.50 | 1 | 2021-09-01 | 2 |
| 9 | 3 | 2 | 33 | 414.50 | 0 | 2021-09-01 | 2 |

Length: 293,433
-> this is already
aggregated

Merge on:
County
product_type
is_business

data_block_id

train.csv

| | county | is_business | product_type | target | is_consumption | datetime | data_block_id | row_id | prediction_unit_id |
|---|--------|-------------|--------------|---------|----------------|---------------------|---------------|--------|--------------------|
| 0 | 0 | 0 | 1 | 0.713 | 0 | 2021-09-01 00:00:00 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 96.590 | 1 | 2021-09-01 00:00:00 | 0 | 1 | 0 |
| 2 | 0 | 0 | 2 | 0.000 | 0 | 2021-09-01 00:00:00 | 0 | 2 | 1 |
| 3 | 0 | 0 | 2 | 17.314 | 1 | 2021-09-01 00:00:00 | 0 | 3 | 1 |
| 4 | 0 | 0 | 3 | 2.904 | 0 | 2021-09-01 00:00:00 | 0 | 4 | 2 |
| 5 | 0 | 0 | 3 | 656.859 | 1 | 2021-09-01 00:00:00 | 0 | 5 | 2 |
| 6 | 0 | 1 | 0 | 0.000 | 0 | 2021-09-01 00:00:00 | 0 | 6 | 3 |
| 7 | 0 | 1 | 0 | 59.000 | 1 | 2021-09-01 00:00:00 | 0 | 7 | 3 |
| 8 | 0 | 1 | 1 | 0.000 | 0 | 2021-09-01 00:00:00 | 0 | 8 | 4 |
| 9 | 0 | 1 | 1 | 501.760 | 1 | 2021-09-01 00:00:00 | 0 | 9 | 4 |

Length: 18,165,168

Merge on:
County
forecast_date
data_block_id

Merge on:
County
forecast_date
data_block_id

forecast_weather.csv

| | latitude | longitude | origin_datetime | hours_ahead | temperature | dewpoint | cloudcover_high | cloudcover_low | cloudcover_mid | cloudcover_total | 10_metre_u_wind_component | 10_metre_v_wind_component | data_block_id | forecast_datetime |
|---|----------|-----------|---------------------------|-------------|-------------|-----------|-----------------|----------------|----------------|------------------|---------------------------|---------------------------|---------------|---------------------------|
| 0 | 57.6 | 21.7 | 2021-09-01 00:00:00+00:00 | 1 | 15.655786 | 11.553613 | 0.904816 | 0.019714 | 0.000000 | 0.905899 | -0.411328 | -9.106137 | 1 | 2021-09-01 01:00:00+00:01 |
| 1 | 57.6 | 22.2 | 2021-09-01 00:00:00+00:00 | 1 | 13.003931 | 10.689844 | 0.886322 | 0.004456 | 0.000000 | 0.886658 | 0.206347 | -5.355405 | 1 | 2021-09-01 01:00:00+00:01 |
| 2 | 57.6 | 22.7 | 2021-09-01 00:00:00+00:00 | 1 | 14.206567 | 11.671777 | 0.729034 | 0.005615 | 0.000000 | 0.730499 | 1.451587 | -7.417905 | 1 | 2021-09-01 01:00:00+00:01 |
| 3 | 57.6 | 23.2 | 2021-09-01 00:00:00+00:00 | 1 | 14.844507 | 12.264917 | 0.336304 | 0.074341 | 0.000626 | 0.385468 | 1.090869 | -9.163999 | 1 | 2021-09-01 01:00:00+00:01 |
| 4 | 57.6 | 23.7 | 2021-09-01 00:00:00+00:00 | 1 | 15.293848 | 12.458887 | 0.102875 | 0.088074 | 0.000015 | 0.176590 | 1.268481 | -8.975766 | 1 | 2021-09-01 01:00:00+00:01 |
| 5 | 57.6 | 24.2 | 2021-09-01 00:00:00+00:00 | 1 | 15.448999 | 12.878687 | 0.000000 | 0.232635 | 0.053558 | 0.237198 | 1.114184 | -7.702085 | 1 | 2021-09-01 01:00:00+00:01 |
| 6 | 57.6 | 24.7 | 2021-09-01 00:00:00+00:00 | 1 | 10.723535 | 10.032739 | 0.095764 | 0.222443 | 0.034836 | 0.302612 | 1.374927 | -2.276303 | 1 | 2021-09-01 01:00:00+00:01 |
| 7 | 57.6 | 25.2 | 2021-09-01 00:00:00+00:00 | 1 | 10.384893 | 8.832788 | 0.469421 | 0.105103 | 0.055603 | 0.528717 | 1.847339 | -2.719419 | 1 | 2021-09-01 01:00:00+00:01 |
| 8 | 57.6 | 25.7 | 2021-09-01 00:00:00+00:00 | 1 | 9.783228 | 8.720483 | 0.370117 | 0.091003 | 0.035339 | 0.430389 | 1.459521 | -2.114682 | 1 | 2021-09-01 01:00:00+00:01 |
| 9 | 57.6 | 26.2 | 2021-09-01 00:00:00+00:00 | 1 | 9.965112 | 9.094995 | 0.004456 | 0.247803 | 0.015030 | 0.252289 | 1.481128 | -2.359311 | 1 | 2021-09-01 01:00:00+00:01 |

Length: 61,641,216

historical_weather.csv

| | datetime | temperature | dewpoint | rain | snowfall | surface_pressure | cloudcover_total | cloudcover_low | cloudcover_mid | cloudcover_high | windspeed_10m | winddirection_10m | shortwave_radiation | direct_solar_radiation | diffuse_radiation | latitude | longitude | data_block_id |
|---|---------------------|-------------|----------|------|----------|------------------|------------------|----------------|----------------|-----------------|---------------|-------------------|---------------------|------------------------|-------------------|----------|-----------|---------------|
| 0 | 2021-09-01 00:00:00 | 14.4 | 12.0 | 0.0 | 0.0 | 1015.8 | 4 | 4 | 0 | 0 | 6.694444 | 3 | 0.0 | 0.0 | 0.0 | 57.6 | 21.7 | 1 |
| 1 | 2021-09-01 00:00:00 | 14.0 | 12.0 | 0.0 | 0.0 | 1010.6 | 7 | 8 | 0 | 0 | 4.944444 | 353 | 0.0 | 0.0 | 0.0 | 57.6 | 22.2 | 1 |
| 2 | 2021-09-01 00:00:00 | 14.4 | 12.8 | 0.0 | 0.0 | 1014.9 | 6 | 7 | 0 | 0 | 5.833333 | 348 | 0.0 | 0.0 | 0.0 | 57.6 | 22.7 | 1 |
| 3 | 2021-09-01 00:00:00 | 15.4 | 13.0 | 0.0 | 0.0 | 1014.4 | 4 | 2 | 4 | 0 | 7.111111 | 349 | 0.0 | 0.0 | 0.0 | 57.6 | 23.2 | 1 |
| 4 | 2021-09-01 00:00:00 | 15.9 | 12.6 | 0.0 | 0.0 | 1013.8 | 12 | 7 | 0 | 20 | 8.388889 | 360 | 0.0 | 0.0 | 0.0 | 57.6 | 23.7 | 1 |
| 5 | 2021-09-01 00:00:00 | 13.1 | 10.6 | 0.0 | 0.0 | 1013.4 | 1 | 1 | 0 | 0 | 4.500000 | 3 | 0.0 | 0.0 | 0.0 | 57.6 | 24.2 | 1 |
| 6 | 2021-09-01 00:00:00 | 13.0 | 10.7 | 0.0 | 0.0 | 1005.0 | 31 | 10 | 0 | 74 | 3.916667 | 357 | 0.0 | 0.0 | 0.0 | 57.6 | 24.7 | 1 |
| 7 | 2021-09-01 00:00:00 | 13.2 | 10.9 | 0.0 | 0.0 | 1005.8 | 41 | 16 | 0 | 80 | 3.722222 | 354 | 0.0 | 0.0 | 0.0 | 57.6 | 25.2 | 1 |
| 8 | 2021-09-01 00:00:00 | 13.3 | 10.9 | 0.0 | 0.0 | 1003.5 | 61 | 36 | 0 | 94 | 3.833333 | 354 | 0.0 | 0.0 | 0.0 | 57.6 | 25.7 | 1 |
| 9 | 2021-09-01 00:00:00 | 13.6 | 11.3 | 0.0 | 0.0 | 1003.5 | 97 | 68 | 11 | 97 | 3.722222 | 354 | 0.0 | 0.0 | 0.0 | 57.6 | 26.2 | 1 |

Merge on
lat and long

Get county column from lat and long in
weather_station_to_county_mapping.csv

weather_station_to_county_mapping.csv

| | county_name | longitude | latitude | county |
|---|-------------|-----------|----------|--------|
| 0 | NaN | 21.7 | 57.6 | NaN |
| 1 | NaN | 21.7 | 57.9 | NaN |
| 2 | NaN | 21.7 | 58.2 | NaN |
| 3 | NaN | 21.7 | 58.5 | NaN |
| 4 | NaN | 21.7 | 58.8 | NaN |
| 5 | NaN | 21.7 | 59.1 | NaN |
| 6 | NaN | 21.7 | 59.4 | NaN |
| 7 | NaN | 21.7 | 59.7 | NaN |
| 8 | NaN | 22.2 | 57.6 | NaN |
| 9 | NaN | 22.2 | 57.9 | NaN |