Model Stacking

In questo notebook sono usati i dataset su cui è stata operata la fase preliminare di **preparazione dei dati**, la **divisione** in base alla contea e la **selezione delle feature**.

È costruito un **metamodello** che fa utilizzo di un dataset le cui feature sono le predizioni dei tre modelli costruiti, rispettivamente un **DecisionTreeRegressor**, un **AdaBoostRegressor** e un **RandomForestRegressor**.

Il metamodello è allenato è allenato con un **LinearRegressor** sulle stesse instanze di train usate per generare le predizioni del nuovo dataset.

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

from sklearn.metrics      import mean_squared_error
from sklearn.ensemble     import RandomForestRegressor
from sklearn.ensemble     import AdaBoostRegressor
from sklearn.tree         import DecisionTreeRegressor
from sklearn.linear_model import LinearRegression
## Lettura dei dati

warnings.filterwarnings('ignore')
```

Lettura dei dati

In [3]:
```
# local file paths

dir_name = 'selezione'
region_names = np.array(['A', 'B', 'C'])

fp_Xtrain = []
fp_Xval   = []
fp_Xtest  = []
fp_ytrain = []
fp_yval   = []
fp_ytest  = []

for i in range(3):
    fp_Xtrain.append(dir_name + f'/X_train{region_names[i]}.csv')
    fp_Xval  .append(dir_name + f'/X_val{  region_names[i]}.csv')
    fp_Xtest .append(dir_name + f'/X_test{ region_names[i]}.csv')
    fp_ytrain.append(dir_name + f'/y_train{region_names[i]}.csv')
    fp_yval  .append(dir_name + f'/y_val{  region_names[i]}.csv')
    fp_ytest .append(dir_name + f'/y_test{ region_names[i]}.csv')
```

In [4]:
```
# Lettura dei dati

X_train = []
X_val   = []
X_test  = []
y_train = []
y_val   = []
y_test  = []

for i in range(3):
    X_train.append(pd.read_csv(fp_Xtrain[i], low_memry=False))
    X_val  .append(pd.read_csv(fp_Xval  [i], low_memry=False))
    X_test .append(pd.read_csv(fp_Xtest [i], low_memry=False))
    y_train.append(pd.read_csv(fp_ytrain[i], low_memry=False))
    y_val  .append(pd.read_csv(fp_yval  [i], low_memry=False))
    y_test .append(pd.read_csv(fp_ytest [i], low_memry=False))

X_train = np.array(X_train, dtype=object)
X_val   = np.array(X_val,   dtype=object)
X_test  = np.array(X_test,  dtype=object)
y_train = np.array(y_train, dtype=object)
y_val   = np.array(y_val,   dtype=object)
y_test  = np.array(y_test,  dtype=object)
```

In [5]:
```
def dimensionality(y=False):
    for i in range(3):
        print(f'X_train{region_names[i]}: {X_train[i].shape}')
        print(f'X_val{region_names[i]}:   {X_val  [i].shape}')
        print(f'X_test{region_names[i]}:  {X_test [i].shape}')
        if y:
            print(f'y_train{region_names[i]}: {y_train[i].shape}')
            print(f'y_val{region_names[i]}:   {y_val  [i].shape}')
            print(f'y_test{region_names[i]}:  {y_test [i].shape}')
        print()
```

In [6]:
```
dimensionality(y=True)
```

```
X_trainA: (26819, 55)
X_valA:   (9006, 55)
X_testA:  (9085, 55)
y_trainA: (26819, 1)
y_valA:   (9006, 1)
y_testA:  (9085, 1)

X_trainB: (8119, 32)
X_valB:   (2658, 32)
X_testB:  (2606, 32)
y_trainB: (8119, 1)
y_valB:   (2658, 1)
y_testB:  (2606, 1)

X_trainC: (64771, 33)
X_valC:   (21908, 33)
X_testC:  (21876, 33)
y_trainC: (64771, 1)
y_valC:   (21908, 1)
y_testC:  (21876, 1)
```

Modelli precedentemente individuati

Albero di decisione

In [7]:
```
dt_model = np.array([
    DecisionTreeRegressor(max_leaf_nodes = 10),
    DecisionTreeRegressor(max_leaf_nodes = 10),
    DecisionTreeRegressor(max_leaf_nodes = 10)
])
```

Boosting

In [8]:
```
boost_model = np.array([
    AdaBoostRegressor(dt_model[0], n_estimators = 20),
    AdaBoostRegressor(dt_model[1], n_estimators = 20),
    AdaBoostRegressor(dt_model[2], n_estimators = 20)
])
```

Foresta

In [9]:
```
rf_model = np.array([
    RandomForestRegressor(n_estimators = 850, n_jobs = -1),
    RandomForestRegressor(n_estimators = 700, n_jobs = -1),
    RandomForestRegressor(n_estimators = 700, n_jobs = -1)
])
```

In [10]:
```
models = np.array([
    dt_model,
    boost_model,
    rf_model
])
```

Costruzione del Metamodello

Costruzione di un dataset **Y** cui ciascuna colonna contiene le **predizioni dei modelli selezionati per le istanze di Train**. Le colonne hanno il nome del tipo di modello usato. L'operazione è usata per ottenere un dataset Y sia per il **Train** che per il **Test**.

In [11]:
```
def gen_Y(X, y, models, col_names, index):
    Y = pd.DataFrame()
    for i in range(len(col_names)):
        models[i][index].fit(X[index], y[index])
        Y[col_names[i]] = models[i][index].predict(X[index])
    return Y
```

In [12]:
```
def gen_arrY(X, y, models, col_names):
    Y = []
    for i in range(len(X)):
        Y.append(gen_Y(X, y, models, col_names, i))
    Y = np.array(Y, dtype=object)
    return Y
```

In [13]:
```
col_names = ['DecisionTree', 'Boosting', 'RandomForest']
```

In [14]:
```
Y_train = gen_arrY(X_train, y_train, models, col_names)
Y_test  = gen_arrY(X_test,  y_test,  models, col_names)
```

In [15]:
```
Y_train[0].head(20)
```

	DecisionTree	Boosting	RandomForest
0	0.050485	0.266712	0.055412
1	0.007182	0.113340	0.025957
2	0.007182	0.056837	-0.006884
3	0.007182	0.032949	-0.019451
4	0.017906	0.095386	0.021818
5	0.007182	0.091944	0.023328
6	0.017906	0.201193	0.123257
7	0.007182	0.182836	-0.028503
8	0.017906	0.182836	0.026966
9	0.007182	0.071312	-0.002090
10	0.007182	0.188450	0.006948
11	0.007182	0.032391	0.015780
12	0.007182	0.201981	-0.004285
13	0.007182	0.198270	-0.027130
14	0.007182	0.103031	-0.000350
15	0.017906	0.061509	0.035279
16	0.017906	-0.023329	-0.004317
17	0.017906	0.058858	0.003308
18	0.007182	0.023251	-0.000682
19	0.017906	0.112282	0.097573

In [16]:
```
Y_test[0].head(20)
```

	DecisionTree	Boosting	RandomForest
0	0.007520	0.034881	0.019503
1	0.040975	0.659259	0.013394
2	0.007520	-0.889670	0.017264
3	0.007520	-0.025480	-0.052040
4	0.040975	0.861412	0.002068
5	0.040975	0.921076	-0.021154
6	0.007520	-1.097865	-0.000320
7	0.007520	0.007690	-0.010995
8	0.011041	0.039852	-0.033108
9	0.007520	0.007690	-0.044424
10	0.011041	0.137659	0.010438
11	0.007520	-0.505845	0.010550
12	0.011041	0.006302	0.023889
13	0.007520	0.592034	0.029549
14	0.007520	0.051201	0.028119
15	0.007520	0.005953	-0.008425
16	0.011041	0.032760	0.070837
17	0.007520	0.673133	0.065244
18	0.011041	0.710624	0.019638
19	0.007520	0.420446	0.014377

In [17]:
```
def dimensionalityY():
    for i in range(3):
        print(region_names[i])
        print(f'X_train{region_names[i]}: {X_train[i].shape}')
        print(f'Y_train{region_names[i]}: {Y_train[i].shape}')
        print(f'y_train{region_names[i]}: {y_train[i].shape}')
        print()
        print(f'X_test{region_names[i]}:  {X_test[i].shape}')
        print(f'Y_test{region_names[i]}:  {Y_test[i].shape}')
        print(f'y_test{region_names[i]}:  {y_test[i].shape}')
        print()
```

In [18]:
```
dimensionalityY()
```

```
A
X_trainA: (26819, 55)
Y_trainA: (26819, 3)
y_trainA: (26819, 1)

X_testA: (9085, 55)
Y_testA: (9085, 3)
y_testA: (9085, 1)

B
X_trainB: (8119, 32)
Y_trainB: (8119, 3)
y_trainB: (8119, 1)

X_testB: (2606, 32)
Y_testB: (2606, 3)
y_testB: (2606, 1)

C
X_trainC: (64771, 33)
Y_trainC: (64771, 3)
y_trainC: (64771, 1)

X_testC: (21876, 33)
Y_testC: (21876, 3)
y_testC: (21876, 1)
```

Regressione Lineare

Modello di regressione lineare allenato sul nuovo dataset costruito con le predizioni per le **Y di train** e le stesse **y di train** su cui sono stati allenati i modelli. Le predizioni sono calcolate sulla base del nuovo dataset costruito con le predizioni per le **Y di test**.

In [19]:
```
def linear_regression_preds(Y_train, y_train, Y_test):
    preds = []
    for Y_trn, y_trn, Y_tst in zip(Y_train, y_train, Y_test):
        preds.append(
            LinearRegression(fit_intercept=True).fit(Y_trn, y_trn).predict(Y_tst)
        )
    preds = np.array(preds, dtype = object)
    return preds
```

In [20]:
```
y_preds = linear_regression_preds(Y_train, y_train, Y_test)
```

Sulla base del confronto tra predizioni e valori attesi calcolo l'errore.

In [21]:
```
def get_mse(y_trues, y_preds):
    mse = []
    for y_true, y_pred in zip(y_trues, y_preds):
        mse.append(mean_squared_error(y_true, y_pred))
    mse = np.array(mse, dtype = object)
    return mse
```

In [22]:
```
mse = get_mse(y_test, y_preds)
```

In [23]:
```
for i in range(3):
    print(region_names[i])
    print(mse[i])
    print()
```

```
A
0.0016989797078586966

B
0.0023132091661643032

C
0.0022809873002163383
```

Confronto con l'errore del modello banale

In [24]:
```
def mse_trivial(y_train, y):
    y_pred = np.repeat(y_train.mean(), len(y_train))
    return ((y_pred - y.reshape(-1,1))**2).mean()
```

In [25]:
```
for i in range(3):
    print(region_names[i])
    print(mse_trivial(y_test[i].values.ravel(), y_preds[i]))
    print()
```

```
A
0.02757366856545814

B
0.01923225064440397

C
0.02754626652860815
```

Analisi dei risultati

Rispetto al modello banale gli errori del metamodello sono circa dieci volte più precise: seppur il modello non sia molto preciso, combinare le informazioni di più modelli sembra essere risultato vantaggioso.

Migliori Predizioni

In [26]:
```
def best_prediction(X, y_trues, y_preds, index, nrow=10):
    err  = abs(y_trues[index] - y_preds[index])
    err  = err.values.ravel()
    args = err.argsort()
    for i in range(nrow):
        print(
            "Row : %d, logerror : %2.7f error: %2.7f" % (
                args[i],
                (y_trues[index].iloc[args[i]])['logerror'],
                err[args[i]]
            )
        )
    return X[index].iloc[args[:nrow]].transpose()
```

Prima Contea 1286

In [47]:
```
best_prediction(X_test, y_test, y_preds, 0, nrow=5)
```

```
Row : 4204, logerror : 0.0469000 error: 0.0000018
Row : 7126, logerror : -0.0182000 error: 0.0000019
Row : 7055, logerror : -0.0050000 error: 0.0000040
Row : 7353, logerror : 0.0198000 error: 0.0000044
Row : 4263, logerror : 0.0046090 error: 0.0000059
```

	4204	7126	7055	7353	4263
bathroomcnt	2.000000e+00	2.000000e+00	3.000000e+00	2.500000e+00	1.000000e+00
bedroomcnt	2.000000e+00	3.000000e+00	4.000000e+00	3.000000e+00	1.000000e+00
buildingqualitytypeid	7.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00
calculatedbathnbr	2.000000e+00	2.000000e+00	3.000000e+00	2.500000e+00	1.000000e+00
calculatedfinishedsquarefeet	1.274000e+03	1.264000e+03	2.127000e+03	1.345000e+03	8.220000e+02
finishedsquarefeet12	1.274000e+03	1.264000e+03	2.127000e+03	1.345000e+03	8.220000e+02
fireplacecnt	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
latitude	3.366740e+07	3.372443e+07	3.361572e+07	3.358339e+07	3.377900e+07
longitude	1.178505e+08	1.180366e+08	1.176718e+08	1.177196e+08	1.179970e+08
lotsizesquarefeet	7.205000e+03	6.030000e+03	8.710000e+03	2.550000e+03	7.205000e+03
rawcensustractandblock	6.059063e+07	6.059100e+07	6.059032e+07	6.059063e+07	6.059088e+07
regionidcity	5.265000e+04	2.521800e+04	1.277300e+04	3.708600e+04	4.296700e+04
regionidzip	9.694700e+04	9.696700e+04	9.699500e+04	9.697400e+04	9.620300e+04
roomcnt	0.000000e+00	6.000000e+00	5.000000e+00	0.000000e+00	0.000000e+00
unitcnt	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	2.006000e+03	1.960000e+03	1.968000e+03	1.990000e+03	1.983000e+03
structuretaxvaluedollarcnt	2.500940e+05	5.604800e+04	1.442240e+05	1.265110e+05	5.993700e+04
taxvaluedollarcnt	4.617750e+05	4.641480e+05	3.221080e+05	3.187550e+05	2.071750e+05
landtaxvaluedollarcnt	2.116810e+05	4.081000e+05	1.778840e+05	1.922440e+05	1.472380e+05
taxamount	5.697100e+03	5.336560e+03	3.316140e+03	3.246380e+03	2.674880e+03
int_transactiondate	2.630000e+02	2.030000e+02	2.630000e+02	1.810000e+02	8.600000e+01
period_mean_price	1.852401e-02	8.281449e-03	1.852401e-02	8.717589e-03	8.947002e-03
neighborhood_mean_price	1.916152e-02	1.168283e-02	1.227379e-02	1.051390e-02	1.159592e-02
living_area_prop	2.185004e-01	2.096186e-01	2.442021e-01	5.274510e-01	2.185004e-01
tax_ratio	8.105440e+01	8.697513e+01	9.713341e+01	9.818783e+01	7.745208e+01
tax_prop	1.181466e+00	1.373389e-01	8.107756e-01	6.580752e-01	4.070756e-01
buildingqualitytypeid_na_flag	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
unitcnt_na_flag	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
assessmentyear_2015.0	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00
heatingorsystemtypeid_1.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_6.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_7.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_10.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_11.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_12.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_13.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_18.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_24.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
poolcnt_1.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_0100	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_122	0.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_34	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00
propertycountylandusecode_rare	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_31.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_247.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_248.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_260.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_261.0	0.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_263.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_264.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_265.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_266.0	1.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00
propertylandusetypeid_267.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_269.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00

Condiderazioni:

- il valore assoluto dei `logerror` è relativamente piccolo.
- le grandezze di 4 delle 5 case è simile come indicano `bedroomcnt`, `bathroomcnt` e `lotsizesquarefeet`. Una casa ha un numero contenuto di stanze e una superficie conseguentemente piccola rispetto alle altre.
- tutte le case hanno lo stesso `buildingqualitytypeid` pari a 7
- tutte le case sono di città differenti ma della stessa regione.
- tutte le case hanno lo stesso `rawcensustractandblock`.
- la variabilità del `neighborhood_mean_price` è contenuta.
- tutte e cinque le case sono sprovviste di impianto anticendio (`fireplacecnt` = 0) e sono sprovviste di piscina (`poolcnt` = 0).
- quattro case su cinque sono state assegnate nel 2015 (flag `assessmentyear_2015` acceso).
- per tutte e cinque le case il flag `buildingqualitytypeid` e `unitcnt` acceso.
- per tutte le case i flag `heatingorsystemtypeid` sono sempre spenti (valori non presenti nei train oppure altre feature eliminate dalla selezione).

Seconda Contea 2061

In [48]:
```
best_prediction(X_test, y_test, y_preds, 1, nrow=5)
```

```
Row : 1698, logerror : 0.0609842 error: 0.0000032
Row : 773, logerror : 0.1458000 error: 0.0000075
Row : 989, logerror : -0.0051408 error: 0.0000372
Row : 794, logerror : 0.0411000 error: 0.0000412
Row : 1794, logerror : -0.0243876 error: 0.0000658
```

	1698	773	989	794	1794
bathroomcnt	3.500000e+00	2.000000e+00	2.000000e+00	3.000000e+00	2.000000e+00
bedroomcnt	4.000000e+00	3.000000e+00	4.000000e+00	3.000000e+00	4.000000e+00
calculatedbathnbr	3.500000e+00	2.000000e+00	2.000000e+00	3.000000e+00	2.000000e+00
calculatedfinishedsquarefeet	3.695000e+03	1.858000e+03	1.716000e+03	2.030000e+03	2.030000e+03
finishedsquarefeet12	3.695000e+03	1.858000e+03	1.716000e+03	2.030000e+03	2.030000e+03
fireplacecnt	3.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
latitude	3.418754e+07	3.417726e+07	3.417802e+07	3.425167e+07	3.421825e+07
longitude	1.192310e+08	1.192079e+08	1.189490e+08	1.187672e+08	1.188812e+08
lotsizesquarefeet	6.006000e+03	5.227000e+03	9.342000e+03	1.065300e+04	8.323000e+03
rawcensustractandblock	6.111004e+07	6.111004e+07	6.111006e+07	6.111008e+07	6.111006e+07
regionidcity	1.315000e+04	4.719800e+04	3.427800e+04	2.711000e+04	3.427800e+04
regionidzip	9.710600e+04	9.710900e+04	9.635100e+04	9.711800e+04	9.638300e+04
roomcnt	6.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00
unitcnt	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	2.004000e+03	1.977000e+03	1.965000e+03	1.965000e+03	1.978000e+03
structuretaxvaluedollarcnt	4.660000e+05	1.300000e+05	2.960000e+05	1.275330e+05	2.864790e+05
taxvaluedollarcnt	9.300000e+05	3.700000e+05	5.960000e+05	1.961930e+05	5.729580e+05
landtaxvaluedollarcnt	4.640000e+05	2.400000e+05	3.000000e+05	6.866000e+04	2.864790e+05
taxamount	1.597284e+04	4.311700e+03	6.295180e+03	2.531680e+03	6.480120e+03
int_transactiondate	1.970000e+02	1.750000e+02	2.020000e+02	3.000000e+01	2.560000e+02
period_mean_price	1.313502e-02	8.967911e-03	1.313502e-02	9.852440e-03	1.567975e-02
neighborhood_mean_price	1.015588e-02	6.582700e-03	8.166625e-03	1.740458e-02	1.846930e-02
living_area_prop	6.152181e-01	3.554620e-01	1.836866e-01	1.905566e-01	2.439024e-01
tax_ratio	5.822384e+01	8.581302e+01	9.467561e+01	7.749519e+01	8.841781e+01
tax_prop	1.004310e+00	5.416667e-01	9.866667e-01	1.857457e+00	1.000000e+00
assessmentyear_2015.0	0.000000e+00	1.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00
poolcnt_1.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_261.0	1.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
propertylandusetypeid_265.0	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_266.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_275.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00

Condiderazioni:

- il valore assoluto dei `logerror` varia in un range relativamente contenuto.
- il numero di stanze è molto simile: `bedroomcnt` è pari a 3 o 4, `bathroomcnt` a 2 o 3, infatti `lotsizesquarefeet` è simile per quattro case, mentre per una quinta casa ha una superficie decisamente maggiore (infatti `living_area_prop` per questa casa è circa il doppio delle altre).
- la casa con molta superficie ha inoltre ben 3 impianti anticendio (`fireplacecnt`); le altre quattro ne hanno solo uno.
- tutte le case hanno lo stesso `rawcensustractandblock`.
- tutte le case sono di città e regioni differenti.
- tutte e cinque le case sono sprovviste di piscina (`poolcnt` = 0).
- quattro case su cinque sono state assegnate nel 2015 (flag `assessmentyear_2015` accesso).
- quattro delle cinque case hanno `propertylandusetypeid` pari a 261.

Terza Contea 3101

In [50]:
```
best_prediction(X_test, y_test, y_preds, 2, nrow=5)
```

```
Row : 3725, logerror : -0.1831224 error: 0.0000005
Row : 15384, logerror : -0.0273226 error: 0.0000042
Row : 17022, logerror : 0.0357668 error: 0.0000042
Row : 8248, logerror : 0.0070000 error: 0.0000094
Row : 8436, logerror : -0.0790000 error: 0.0000123
```

	3725	15384	17022	8248	8436
bathroomcnt	2.000000e+00	1.000000e+00	1.000000e+00	3.000000e+00	2.000000e+00
bedroomcnt	3.000000e+00	2.000000e+00	2.000000e+00	3.000000e+00	2.000000e+00
buildingqualitytypeid	6.000000e+00	5.000000e+00	4.000000e+00	4.000000e+00	7.000000e+00
calculatedbathnbr	2.000000e+00	1.000000e+00	1.000000e+00	3.000000e+00	2.000000e+00
calculatedfinishedsquarefeet	1.690000e+03	9.640000e+02	1.113000e+03	2.443000e+03	9.800000e+02
finishedsquarefeet12	1.690000e+03	9.640000e+02	1.113000e+03	2.443000e+03	9.800000e+02
latitude	3.415907e+07	3.418092e+07	3.403501e+07	3.457394e+07	3.440290e+07
longitude	1.183154e+08	1.185585e+08	1.181530e+08	1.180077e+08	1.184640e+08
lotsizesquarefeet	7.418000e+03	2.177600e+04	5.450000e+03	7.138000e+03	7.802700e+04
rawcensustractandblock	6.037312e+07	6.037139e+07	6.037530e+07	6.037911e+07	6.037920e+07
regionidcity	3.960540e+05	1.244700e+04	1.788200e+04	4.022700e+04	5.431100e+04
regionidzip	9.643800e+04	9.637900e+04	9.600300e+04	9.733000e+04	9.635200e+04
unitcnt	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	1.939000e+03	1.926000e+03	1.949000e+03	2.006000e+03	1.990000e+03
structuretaxvaluedollarcnt	1.065380e+05	1.780400e+05	7.647700e+04	1.919000e+05	6.919200e+04
taxvaluedollarcnt	2.127370e+05	2.967320e+05	2.733250e+05	2.399000e+05	1.810780e+05
landtaxvaluedollarcnt	1.061990e+05	1.186920e+05	1.968480e+05	4.800000e+04	1.118860e+05
taxamount	2.374790e+03	3.740090e+03	3.816970e+03	5.051830e+03	2.715390e+03
int_transactiondate	7.300000e+01	8.100000e+01	1.100000e+02	1.670000e+02	1.440000e+02
period_mean_price	1.198506e-02	1.198506e-02	7.500441e-03	8.206334e-03	5.133757e-03
neighborhood_mean_price	7.247531e-03	1.671850e-02	1.348337e-02	7.631008e-03	5.674360e-03
living_area_prop	2.278242e-01	4.426892e-02	2.042202e-01	3.422527e-01	1.255976e-02
tax_ratio	8.958139e+01	7.933820e+01	7.160785e+01	4.748774e+01	6.668582e+01
tax_prop	1.003192e+00	1.500017e+00	3.885079e-01	3.997917e+00	6.184152e-01
assessmentyear_2015.0	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00
heatingorsystemtypeid_2.0	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00
heatingorsystemtypeid_7.0	1.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_0100	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00
propertycountylandusecode_0101	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_rare	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertyzoningdesc_LAR1	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertyzoningdesc_rare	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00

Condiderazioni:

- il valore assoluto dei `logerror` varia in un range relativamente molto piccolo.
- ci sono case un po' più piccole e un po' più grandi: il `bedroomcnt` e `bathroomcnt` variano tra 1 e 4, ciò si riflette in `finishedsquarefeet12`.
- `calculatedfinishedsquarefeet` e `finishedsquarefeet12` per tutte e cinque le case hanno lo stesso valore.
- tre case sono state costruite prima degli anni '50 come evidenzia `yearbuilt`.
- tutte le case sono di città e regioni differenti.
- tre case hanno `heatingorsystemtypeid` pari a 2, le altre due a 7.
- quattro case su cinque hanno `propertycountylandusecode` pari a 100.
- tutte e cinque le case hanno un `propertyzoningdesc` non frequente.

Peggiori Predizoini

In [52]:
```
def worst_prediction(X, y_trues, y_preds, index, nrow=10):
    err  = abs(y_trues[index] - y_preds[index])
    err  = err.values.ravel()
    args = err.argsort()[::-1]
    for i in range(nrow):
        print(
            "Row : %d, logerror : %2.7f error: %2.7f" % (
                args[i],
                (y_trues[index].iloc[args[i]])['logerror'],
                err[args[i]]
            )
        )
    return X[index].iloc[args[:nrow]].transpose()
```

Prima Contea 1286

In [53]:
```
worst_prediction(X_test, y_test, y_preds, 0, nrow=5)
```

```
Row : 1481, logerror : 2.4656840 error: 0.7085885
Row : 6773, logerror : 2.0550000 error: 0.6617327
Row : 7731, logerror : -0.1984006 error: 0.6601608
Row : 7107, logerror : -0.0694000 error: 0.5384818
Row : 6451, logerror : 2.3224013 error: 0.5320029
```

	1481	6773	7731	7107	6451
bathroomcnt	2.500000e+00	0.000000e+00	0.000000e+00	2.500000e+00	1.000000e+00
bedroomcnt	4.000000e+00	0.000000e+00	0.000000e+00	3.000000e+00	4.000000e+00
buildingqualitytypeid	7.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00	7.000000e+00
calculatedbathnbr	2.500000e+00	2.000000e+00	2.000000e+00	2.500000e+00	1.000000e+00
calculatedfinishedsquarefeet	2.186000e+03	1.547000e+03	1.547000e+03	1.454000e+03	1.764000e+03
finishedsquarefeet12	2.186000e+03	1.524000e+03	1.524000e+03	1.454000e+03	1.764000e+03
fireplacecnt	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
latitude	3.369593e+07	3.367318e+07	3.363002e+07	3.374811e+07	3.366206e+07
longitude	1.179011e+08	1.176577e+08	1.175828e+08	1.179631e+08	1.179998e+08
lotsizesquarefeet	6.750000e+03	7.205000e+03	8.014800e+04	2.000000e+03	5.551000e+03
rawcensustractandblock	6.059064e+07	6.059052e+07	6.059032e+07	6.059089e+07	6.059099e+07
regionidcity	3.803200e+04	2.521800e+04	1.759700e+04	3.478000e+04	2.521800e+04
regionidzip	9.695800e+04	9.696200e+04	9.698900e+04	9.699000e+04	9.696600e+04
roomcnt	8.000000e+00	0.000000e+00	0.000000e+00	6.000000e+00	7.000000e+00
unitcnt	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	1.969000e+03	1.969000e+03	1.969000e+03	2.012000e+03	1.905000e+03
structuretaxvaluedollarcnt	1.349660e+05	2.255544e+06	2.540461e+06	1.222010e+05	3.601750e+05
taxvaluedollarcnt	6.269830e+05	4.074029e+06	5.080924e+06	2.874010e+05	1.694922e+06
landtaxvaluedollarcnt	4.920170e+05	1.818485e+06	2.540463e+06	1.652000e+05	1.334747e+06
taxamount	7.232840e+03	4.314370e+04	5.188794e+04	4.926280e+03	1.878934e+04
int_transactiondate	8.000000e+01	2.720000e+02	3.300000e+01	3.640000e+02	2.570000e+02
period_mean_price	8.947002e-03	1.852401e-02	1.442575e-02	2.093990e-02	2.563148e-02
neighborhood_mean_price	2.303311e-02	9.226918e-03	1.729015e-02	1.354075e-02	6.784414e-03
living_area_prop	3.238518e-01	2.185004e-01	2.185004e-01	7.270000e-01	3.177806e-01
tax_ratio	8.668559e+01	9.442929e+01	9.792110e+01	5.834037e+01	9.020657e+01
tax_prop	2.743117e-01	1.240342e+00	9.999992e-01	7.397155e-01	2.698452e-01
buildingqualitytypeid_na_flag	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
unitcnt_na_flag	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
assessmentyear_2015.0	0.000000e+00	1.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00
heatingorsystemtypeid_1.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_6.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_7.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_10.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_11.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_12.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_13.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_18.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
heatingorsystemtypeid_24.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
poolcnt_1.0	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_0100	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_122	1.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
propertycountylandusecode_34	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_rare	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_31.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_247.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_248.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_260.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_261.0	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
propertylandusetypeid_263.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_264.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_265.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_266.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_267.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_269.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00

Condiderazioni:

- il valore assoluto dei `logerror` è positivo e piccolo per tre predizioni, negativo e piccolo per le altre due.
- ci sono due case il cui `bedroomcnt` e `bathroomcnt` è pari a 0, il che è alquanto improbabile. Può essere sintomo di un dataset mal compilato o di un tipo di informazione mancante non interita come Nan.
- nessuna casa ha un impianto anticendio (`fireplacecnt` pari a 0)
- fatta eccezione per una casa venduta l'8 Gennaio, `int_transactiondate` suggerisce che le altre quattro siano state vendute nei mesi finali dell'anno.
- una casa ha una piscina (bit di `poolcnt` acceso).
- tutte le case sono di città e regioni differenti.
- tutte e cinque le case hanno il missing flag per buildingqualitytypeid e unitcnt acceso.
- per nessuna casa ho informazione su `heatingorsystemtypeid` (non presente in train o eliminato dalla feature selection)
- quattro case su cinque hanno `propertylandusetypeid` pari a 261 e quattro su cinque `propertycountylandusecode` pari a 122.

Seconda Contea 2061

In [56]:
```
worst_prediction(X_test, y_test, y_preds, 1, nrow=5)
```

```
Row : 1085, logerror : -0.0336000 error: 0.6240503
Row : 56, logerror : 2.3300000 error: 0.5950939
Row : 1393, logerror : 2.1773627 error: 0.5317995
Row : 1845, logerror : -2.2381122 error: 0.4723513
Row : 2238, logerror : 1.3631378 error: 0.4209428
```

Out[56]:

	1085	56	1393	1845	2238
bathroomcnt	2.000000e+00	2.000000e+00	2.000000e+00	1.000000e+00	2.500000e+00
bedroomcnt	3.000000e+00	3.000000e+00	5.000000e+00	3.000000e+00	2.000000e+00
calculatedbathnbr	2.000000e+00	2.000000e+00	2.000000e+00	1.000000e+00	2.500000e+00
calculatedfinishedsquarefeet	1.160000e+03	1.696000e+03	1.860000e+03	1.759000e+03	1.240000e+03
finishedsquarefeet12	1.160000e+03	1.696000e+03	1.860000e+03	1.759000e+03	1.240000e+03
fireplacecnt	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
latitude	3.425852e+07	3.426906e+07	3.422859e+07	3.418204e+07	3.417732e+07
longitude	1.187694e+08	1.188007e+08	1.191919e+08	1.191742e+08	1.187859e+08
lotsizesquarefeet	7.000000e+03	8.257000e+03	6.534000e+03	6.000000e+03	7.205000e+03
rawcensustractandblock	6.111008e+07	6.111008e+07	6.111003e+07	6.111004e+07	6.111008e+07
regionidcity	2.711000e+04	2.711000e+04	1.315000e+04	1.315000e+04	5.365500e+04
regionidzip	9.711800e+04	9.711800e+04	9.710700e+04	9.710400e+04	9.639300e+04
roomcnt	5.000000e+00	6.000000e+00	9.000000e+00	7.000000e+00	5.000000e+00
unitcnt	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	1.963000e+03	1.976000e+03	1.977000e+03	1.951000e+03	1.990000e+03
structuretaxvaluedollarcnt	2.016460e+05	1.871250e+05	2.430000e+05	1.227120e+05	2.537050e+05
taxvaluedollarcnt	4.032920e+05	4.775830e+05	4.880000e+05	2.197130e+05	4.224960e+05
landtaxvaluedollarcnt	2.016460e+05	2.904580e+05	2.450000e+05	9.700100e+04	1.687910e+05
taxamount	4.741940e+03	5.467440e+03	5.589700e+03	2.643160e+03	5.680340e+03
int_transactiondate	7.000000e+01	7.000000e+01	1.010000e+02	2.410000e+02	3.300000e+01
period_mean_price	4.897452e-03	4.897452e-03	1.617989e-02	1.080729e-02	1.993256e-02
neighborhood_mean_price	1.740458e-02	1.063207e-02	8.821141e-03	6.311392e-03	1.281764e-02
living_area_prop	1.657143e-01	2.054015e-01	2.846648e-01	2.931667e-01	2.185004e-01
tax_ratio	8.504789e+01	8.735039e+01	8.730343e+01	8.312512e+01	7.437865e+01
tax_prop	1.000000e+00	6.442412e-01	9.918367e-01	1.265059e+00	1.503072e+00
assessmentyear_2015.0	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
poolcnt_1.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_261.0	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00
propertylandusetypeid_265.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_266.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00
propertylandusetypeid_275.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00

Condiderazioni:

- per tre istante il valore assoluto del `logerror` è molto grande.
- il numero di stanze è molto simile: `bedroomcnt` è pari a 2 o 3 (una casa ha ben 5 camere da letto) , `bathroomcnt` a 1 o 2, infatti `lotsizesquarefeet` è simile.
- quattro case su cinque hanno un impianto anticendio (`fireplacecnt`).
- ci sono due coppie di case della stessa città.
- quattro case su cinque sono state costruite tra gli anni '60 e '70.
- due case sono state vendute il 7 gennaio.
- tutte e cinque le case sono sprovviste di piscina (`poolcnt` = 0).
- quattro delle cinque case hanno `propertylandusetypeid` pari a 261.

Terza Contea 3101

In [55]:

```
worst_prediction(X_test, y_test, y_preds, 2, nrow=5)
```

```
Row : 5889, logerror : 2.2370694 error: 1.0158254
Row : 8793, logerror : 2.6198761 error: 1.0031500
Row : 12840, logerror : 0.6217000 error: 0.9122410
Row : 18706, logerror : -2.1756961 error: 0.7333424
Row : 2702, logerror : -1.3240000 error: 0.6130974
```

Out[55]:

	5889	8793	12840	18706	2702
bathroomcnt	7.000000e+00	6.000000e+00	2.000000e+00	2.000000e+00	2.000000e+00
bedroomcnt	1.100000e+01	1.200000e+01	3.000000e+00	1.000000e+00	4.000000e+00
buildingqualitytypeid	8.000000e+00	5.000000e+00	7.000000e+00	8.000000e+00	4.000000e+00
calculatedbathnbr	7.000000e+00	6.000000e+00	2.000000e+00	2.000000e+00	2.000000e+00
calculatedfinishedsquarefeet	6.111000e+03	8.469000e+03	1.202000e+03	1.678000e+03	1.232000e+03
finishedsquarefeet12	6.111000e+03	8.469000e+03	1.202000e+03	1.678000e+03	1.232000e+03
latitude	3.416221e+07	3.407233e+07	3.402572e+07	3.403857e+07	3.404255e+07
longitude	1.185005e+08	1.182988e+08	1.183835e+08	1.186553e+08	1.189082e+08
lotsizesquarefeet	7.205000e+03	7.205000e+03	5.402000e+03	2.296800e+04	8.660700e+04
rawcensustractandblock	6.037140e+07	6.037211e+07	6.037702e+07	6.037800e+07	6.037800e+07
regionidcity	1.244700e+04	1.244700e+04	5.161700e+04	1.252000e+04	1.252000e+04
regionidzip	9.634900e+04	9.598500e+04	9.609700e+04	9.611600e+04	9.611600e+04
unitcnt	4.000000e+00	2.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
yearbuilt	1.949000e+03	1.909000e+03	1.940000e+03	2.002000e+03	1.953000e+03
structuretaxvaluedollarcnt	3.450000e+06	2.252280e+05	1.669360e+05	7.590160e+05	4.934100e+04
taxvaluedollarcnt	5.373376e+06	2.485282e+06	4.103250e+05	9.120827e+06	4.410090e+05
landtaxvaluedollarcnt	1.923376e+06	2.260054e+06	2.433890e+05	8.361811e+06	3.916680e+05
taxamount	6.439638e+04	3.034287e+04	5.483480e+03	1.094690e+05	5.804210e+03
int_transactiondate	2.430000e+02	1.570000e+02	3.530000e+02	1.220000e+02	1.440000e+02
period_mean_price	1.900508e-02	9.990173e-03	1.486509e-02	5.965922e-03	5.133757e-03
neighborhood_mean_price	6.635053e-03	1.426688e-02	-2.698349e-04	-5.838539e-03	-5.371798e-02
living_area_prop	2.185004e-01	2.185004e-01	2.225102e-01	7.305817e-02	1.422518e-02
tax_ratio	8.344221e+01	8.190662e+01	7.482931e+01	8.331883e+01	7.598088e+01
tax_prop	1.793721e+00	9.965602e-02	6.858814e-01	9.077173e-02	1.259766e-01
assessmentyear_2015.0	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00	1.000000e+00
heatingorsystemtypeid_2.0	1.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00
heatingorsystemtypeid_7.0	0.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00	1.000000e+00
propertycountylandusecode_0100	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
propertycountylandusecode_0101	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertycountylandusecode_rare	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertylandusetypeid_246.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertyzoningdesc_LAR1	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
propertyzoningdesc_rare	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00

Condiderazioni:

- il valore assoluto dei `logerror` è per tre case molto alto.
- due case sono decisamente molto grandi: `bahroomcnt` e `bedroomcnt` indicano che una casa ha 7 bagni e 11 camere da letto, l'altra 6 bagni e 12 camere. Anche `calculatedfinishedsquarefeet` è di conseguenza molto grande.
- `calculatedfinishedsquarefeet` e `finishedsquarefeet12` per tutte e cinque le case hanno lo stesso valore.
- la prima casa è stata costruita nel '50, la seconda addirittura nel 1909.
- due coppie di case sono della stessa città.
- due case hanno `heatingorsystemtypeid` pari a 2, le altre tre a 7.
- quattro su cinque case hanno un `propertyzoningdesc` non frequente e `propertycountylandusecode` pari a 100.