In [24]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [25]:
```
d0=pd.read_csv('D:/Projects/Seqwa Assignment/The_One-_to_Four-Family_Home_Performance_with_ENERGY_STAR__Program__Beginning_2010.csv')
d0.head()
```

Out[25]:

	Reporting Period	Home Performance Project ID	Home Performance Site ID	Project County	Project City	Project Zip	Gas Utility	Electric Utility
0	01/31/2018	P00000780513	S00000167844	Jefferson	Three Mile Bay	13693	NaN	National Grid
1	01/31/2018	P00000896731	S00000220979	Orleans	Kent	14477	NaN	National Grid
2	01/31/2018	P00000900996	S00000246366	Saint Lawrence	Potsdam	13676	NaN	National Grid
3	01/31/2018	P00000935291	S00000265719	Oneida	Rome	13440	NaN	National Grid
4	01/31/2018	P00000936199	S00000267582	Oneida	Rome	13440	NaN	National Grid

5 rows × 21 columns

In [26]:
```
d0.shape
```

Out[26]: (48871, 21)

In [27]:
```
d0.describe()
```

Out[27]:

	Project Zip	Total Project Cost (USD)	Total Incentives (USD)	Amount Financed Through Program (USD)	Estimated Annual kWh Savings	Esti A M Sa
count	48871.000000	48871.000000	48871.000000	48871.000000	48871.000000	48871.0
mean	13211.800434	8476.919605	1675.083894	2886.845471	458.446850	31.6244
std	1319.040080	6051.524131	1612.013314	5323.094549	2097.060386	28.7212
min	10001.000000	152.000000	0.000000	0.000000	-27585.000000	-362.00
25%	12010.000000	4427.500000	491.500000	0.000000	0.000000	14.0000
50%	13501.000000	6879.000000	1003.000000	0.000000	1.000000	26.0000
75%	14464.000000	10766.500000	2460.000000	4334.500000	400.000000	42.0000
max	14905.000000	94772.000000	12000.000000	25000.000000	174258.000000	455.000

In [28]:
```
d0.shape
print(d0.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48871 entries, 0 to 48870
Data columns (total 21 columns):
Reporting Period                                                    48871
non-null object
Home Performance Project ID                                         48871
non-null object
Home Performance Site ID                                            48871
non-null object
Project County                                                      48871
non-null object
Project City                                                        48871
non-null object
Project Zip                                                         48871
non-null int64
Gas Utility                                                         34308
non-null object
Electric Utility                                                    48871
non-null object
Project Completion Date                                             48871
non-null object
Customer Type                                                       48871
non-null object
Low-Rise or Home Performance Indicator                              48871
non-null object
Total Project Cost (USD)                                            48871
non-null int64
Total Incentives (USD)                                              48871
non-null int64
Type of Program Financing                                           15381
non-null object
Amount Financed Through Program (USD)                               48871
non-null int64
Pre-Retrofit Home Heating Fuel Type                                 48871
non-null object
Estimated Annual kWh Savings                                        48871
non-null int64
Estimated Annual MMBtu Savings                                      48871
non-null int64
First Year Energy Savings $ Estimate (USD)                          48871
non-null int64
Homeowner Received Green Jobs-Green NY Free/Reduced Cost Audit (Y/N)    48871
non-null object
Location                                                            48870
non-null object
dtypes: int64(7), object(14)
memory usage: 7.8+ MB
None
```

In [29]:
```
#column names
print(d0.columns)
```

```
Index(['Reporting Period', 'Home Performance Project ID',
       'Home Performance Site ID', 'Project County', 'Project City',
       'Project Zip', 'Gas Utility', 'Electric Utility',
       'Project Completion Date', 'Customer Type',
       'Low-Rise or Home Performance Indicator', 'Total Project Cost (USD)',
       'Total Incentives (USD)', 'Type of Program Financing',
       'Amount Financed Through Program (USD)',
       'Pre-Retrofit Home Heating Fuel Type', 'Estimated Annual kWh Savings',
       'Estimated Annual MMBtu Savings',
       'First Year Energy Savings $ Estimate (USD)',
       'Homeowner Received Green Jobs-Green NY Free/Reduced Cost Audit (Y/
N)',
       'Location'],
      dtype='object')
```

In [30]:
```
print(d0['Estimated Annual MMBtu Savings'].nunique())
print(d0.nunique())
```

```
360
Reporting Period                                                      1
Home Performance Project ID                                       48871
Home Performance Site ID                                          47367
Project County                                                       62
Project City                                                       1439
Project Zip                                                        1486
Gas Utility                                                          15
Electric Utility                                                      9
Project Completion Date                                            1978
Customer Type                                                         2
Low-Rise or Home Performance Indicator                                2
Total Project Cost (USD)                                          15296
Total Incentives (USD)                                             4677
Type of Program Financing                                             4
Amount Financed Through Program (USD)                              8839
Pre-Retrofit Home Heating Fuel Type                                  10
Estimated Annual kWh Savings                                       4643
Estimated Annual MMBtu Savings                                      360
First Year Energy Savings $ Estimate (USD)                         3602
Homeowner Received Green Jobs-Green NY Free/Reduced Cost Audit (Y/N)    2
Location                                                           1675
dtype: int64
```

In [31]:
```
print(d0.dtypes)
```

```
Reporting Period                                                  objec
t
Home Performance Project ID                                       objec
t
Home Performance Site ID                                          objec
t
Project County                                                    objec
t
Project City                                                      objec
t
Project Zip                                                        int6
4
Gas Utility                                                       objec
t
Electric Utility                                                  objec
t
Project Completion Date                                           objec
t
Customer Type                                                     objec
t
Low-Rise or Home Performance Indicator                            objec
t
Total Project Cost (USD)                                           int6
4
Total Incentives (USD)                                             int6
4
Type of Program Financing                                         objec
t
Amount Financed Through Program (USD)                              int6
4
Pre-Retrofit Home Heating Fuel Type                               objec
t
Estimated Annual kWh Savings                                       int6
4
Estimated Annual MMBtu Savings                                     int6
4
First Year Energy Savings $ Estimate (USD)                         int6
4
Homeowner Received Green Jobs-Green NY Free/Reduced Cost Audit (Y/N)    objec
t
Location                                                          objec
t
dtype: object
```

Answer 1

In [33]:
```
d0.groupby(['Pre-Retrofit Home Heating Fuel Type'])['Estimated Annual MMBtu Savings'].sum()
```

Out[33]:
```
Pre-Retrofit Home Heating Fuel Type
Anthracite Coal         9
Coal                 1427
Electricity         -3916
Kerosene             2905
Natural Gas        483834
Natural gas        497717
Oil                482771
Propane             62516
Wood                15339
Wood Pellets         2915
Name: Estimated Annual MMBtu Savings, dtype: int64
```

Answer 2

In [37]:
```
d0.groupby(['Project County','Pre-Retrofit Home Heating Fuel Type'])['First Ye
ar Energy Savings $ Estimate (USD)'].max()
```

Out[37]:

Project County	Pre-Retrofit Home Heating Fuel Type	
Albany	Electricity	2114
	Kerosene	5981
	Natural Gas	2201
	Natural gas	2889
	Oil	4934
	Propane	1778
	Wood	1863
	Wood Pellets	1142
Allegany	Electricity	180
	Natural Gas	1273
	Natural gas	589
	Oil	2251
	Propane	2439
	Wood	730
Bronx	Electricity	2140
	Natural Gas	1850
	Natural gas	1831
	Oil	4788
Broome	Coal	1051
	Electricity	2066
	Kerosene	600
	Natural Gas	2523
	Natural gas	2287
	Oil	4700
	Propane	2171
	Wood	986
	Wood Pellets	1465
Cattaraugus	Electricity	1556
	Natural Gas	1743
	Natural gas	1720
		...
Washington	Wood Pellets	489
Wayne	Coal	596
	Electricity	3409
	Kerosene	1298
	Natural Gas	2802
	Natural gas	2034
	Oil	5170
	Propane	4625
	Wood	1238
	Wood Pellets	1128
Westchester	Electricity	2925
	Kerosene	47
	Natural Gas	8320
	Natural gas	3310
	Oil	9173
	Propane	3065
	Wood	1462
Wyoming	Coal	111
	Electricity	1667
	Natural Gas	1390
	Natural gas	1198
	Oil	4373
	Propane	2781
Yates	Coal	85
	Electricity	1608

```
Natural Gas                                1011
Natural gas                                1735
Oil                                        2638
Propane                                    2544
Wood                                       1390
Name: First Year Energy Savings $ Estimate (USD), Length: 443, dtype: int64
```

answer 3

In [58]:
```
df=d0
```

In [59]:
```
print((df.groupby(['Project County'])['First Year Energy Savings $ Estimate (USD)'].mean().sort_values() > 800).sum())
```

```
33
```

answer 4

In [40]:
```
df = d0.groupby(['Project City'])['Estimated Annual kWh Savings'].mean().sort_
values()
df
```

Out[40]:

Project City	
Vermontville	-16408.000000
Fultonville	-12543.500000
Jefferson	-9569.500000
Granville	-7846.000000
Central Sq	-7428.000000
East Berne	-6394.500000
Rock City Falls	-6235.000000
Cuba	-5228.000000
Ellicottville	-5028.000000
Grafton	-4700.500000
Middle Grove	-4565.500000
Ferndale	-4489.000000
Modena	-4362.500000
North Chatham	-4227.600000
Sardinia	-4110.000000
Craryville	-4015.750000
Cambridge	-3775.166667
Darien Center	-3758.400000
Schodack Lndg	-3544.000000
Bolton Landing	-3537.400000
Bakers Mills	-3411.000000
Whitehall	-3278.000000
New Woodstock	-3273.625000
Salamanca	-2634.666667
Farmersville Station	-2397.000000
St Johnsville	-2345.333333
Mountain Dale	-2281.000000
Salem	-2274.000000
Delanson	-2211.500000
Franklin	-2118.000000
...	...
Burke	4914.000000
Westfield	5116.666667
Great Valley	5124.666667
Sinclairville	5406.800000
Saranac Lake	5566.184615
Somers	5652.040816
Malden Bridge	5678.000000
Brantingham	5683.000000
Turin	5687.500000
Lily Dale	5709.500000
Altmar	5890.000000
Smallwood, Ny	6149.000000
Dannemora	6322.000000
Stephentown	6709.500000
Fort Ann	6787.714286
Verona Beach	6979.000000
Tupper Lake	7078.000000
Caneadea	7608.000000
Knox	8213.000000
Saranac	9014.000000
Westbrookville	9492.000000
Mc Donough	10181.000000
Smallwood	11159.000000
Oswegatchie	11387.000000
Buskirk	13181.000000

```
Hoffmeister             14340.000000
Swain                   16410.000000
Lake Placid             21616.500000
Copake Falls            22382.000000
Fort Johnson            31063.250000
Name: Estimated Annual kWh Savings, Length: 1439, dtype: float64
```

answer 5

In [12]:
```
print(d0['Pre-Retrofit Home Heating Fuel Type'].value_counts())
```

```
Natural gas        17044
Natural Gas        15657
Oil                12117
Propane             2189
Electricity         1236
Wood                 351
Wood Pellets         125
Kerosene             101
Coal                  50
Anthracite Coal        1
Name: Pre-Retrofit Home Heating Fuel Type, dtype: int64
```

In [35]:
```
data=d0
print(data.head())
```

```
  Reporting Period Home Performance Project ID Home Performance Site ID  \
0       01/31/2018                P00000780513              S00000167844
1       01/31/2018                P00000896731              S00000220979
2       01/31/2018                P00000900996              S00000246366
3       01/31/2018                P00000935291              S00000265719
4       01/31/2018                P00000936199              S00000267582

   Project County    Project City  Project Zip Gas Utility Electric Utility
  \
0       Jefferson  Three Mile Bay        13693         NaN    National Grid

1         Orleans            Kent        14477         NaN    National Grid

2  Saint Lawrence         Potsdam        13676         NaN    National Grid

3          Oneida            Rome        13440         NaN    National Grid

4          Oneida            Rome        13440         NaN    National Grid


  Project Completion Date Customer Type  \
0              07/06/2012      Assisted
1              07/14/2014      Assisted
2              09/16/2014        Market
3              01/23/2015      Assisted
4              02/09/2015        Market

                     ...                     Total Project Cost (USD)  \
0                    ...                                        10000
1                    ...                                         2980
2                    ...                                        13500
3                    ...                                         6000
4                    ...                                         8915

   Total Incentives (USD)  Type of Program Financing  \
0                    5000                        NaN
1                    1490                        NaN
2                    1350                        NaN
3                    3000                        NaN
4                     892                        NaN

  Amount Financed Through Program (USD)  Pre-Retrofit Home Heating Fuel Typ
e  \
0                                     0                                  Na
N
1                                     0                                  Na
N
2                                     0                                  Na
N
3                                     0                                  Na
N
4                                     0                                  Na
N

  Estimated Annual kWh Savings  Estimated Annual MMBtu Savings  \
0                            0                             19
1                          280                             31
```

```
2                            0                                26
3                          181                                14
4                            0                                38

   First Year Energy Savings $ Estimate (USD)  \
0                                         689
1                                         571
2                                         702
3                                         388
4                                        1001

   Homeowner Received Green Jobs-Green NY Free/Reduced Cost Audit (Y/N)  \
0                                                 Yes
1                                                 Yes
2                                                  No
3                                                 Yes
4                                                 Yes

                                    Location
0                Three Mile\nBay, NY 13693\n
1     Kent, NY 14477\n(43.346379, -78.13651)
2  Potsdam, NY 13676\n(44.65525, -74.932432)
3    Rome, NY 13440\n(43.214717, -75.454665)
4    Rome, NY 13440\n(43.214717, -75.454665)

[5 rows x 21 columns]
```

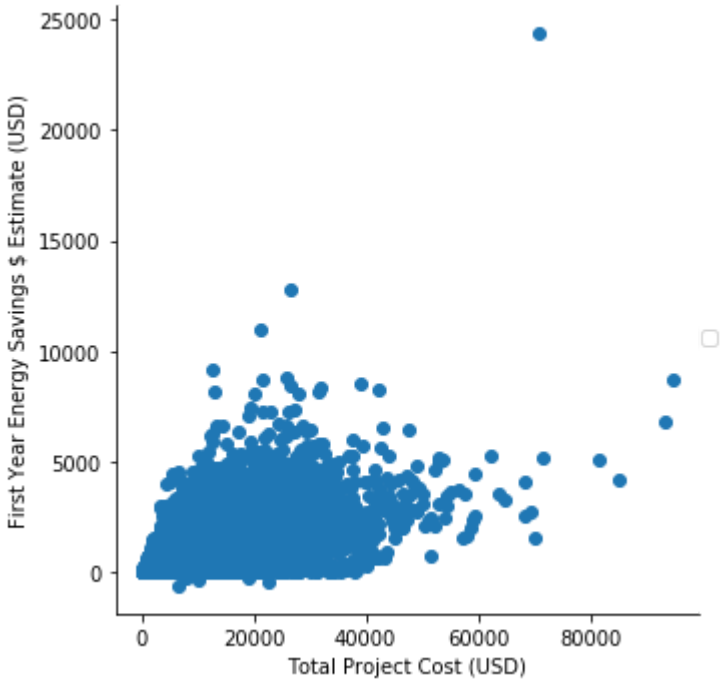
In [25]:
```
d0['Project City'].value_counts()
```

Out[25]:

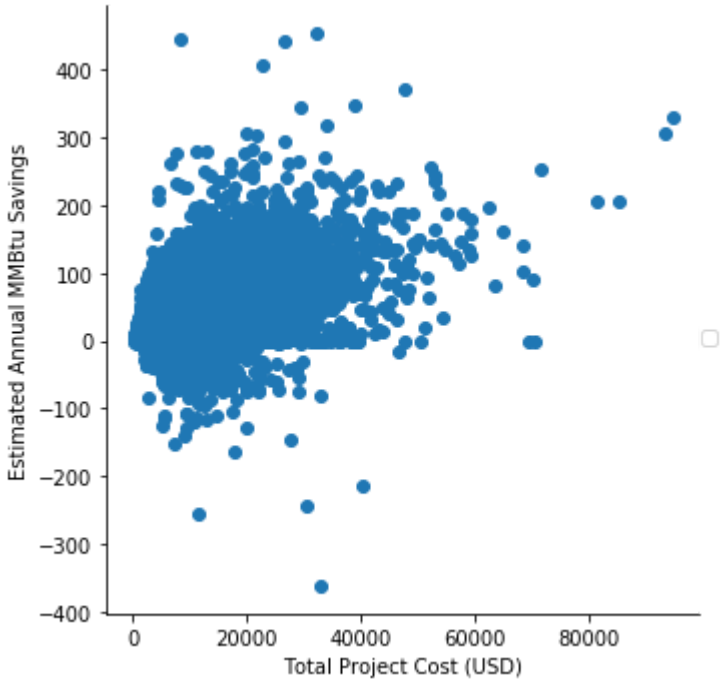
Rochester	5421
Buffalo	2268
Watertown	1949
Syracuse	1576
Fort Drum	686
Webster	667
Ithaca	614
Pittsford	614
Utica	596
Tonawanda	530
Liverpool	467
Canandaigua	394
Penfield	361
Rome	347
Niagara Falls	341
Albany	326
North Chili	323
Fairport	320
West Babylon	318
White Plains	315
Schenectady	305
Brockport	302
Poughkeepsie	292
Binghamton	284
North Tonawanda	284
Baldwinsville	273
Geneva	257
Brooklyn	250
Scarsdale	218
Victor	213
...	
Piseco	1
Alton	1
Breezy Point	1
Champlain	1
Keene Valley	1
Brant	1
Belle Harbor	1
Hinsdale	1
Croton Falls	1
East Marion	1
Elmhurst	1
Portchester	1
Glendale	1
Clarence Ctr	1
South Butler	1
Oswegatchie	1
Lawtons	1
Alexander	1
Pierrepont Manor	1
Java Village	1
Grand Gorge	1
Godeffroy	1
Tioga Center	1
Rhinecliff	1
Clifton Spgs	1
Ferndale	1

```
Fort Hunter            1
Kings Point            1
Mt Tremper             1
Nineveh                1
Name: Project City, Length: 1439, dtype: int64
```

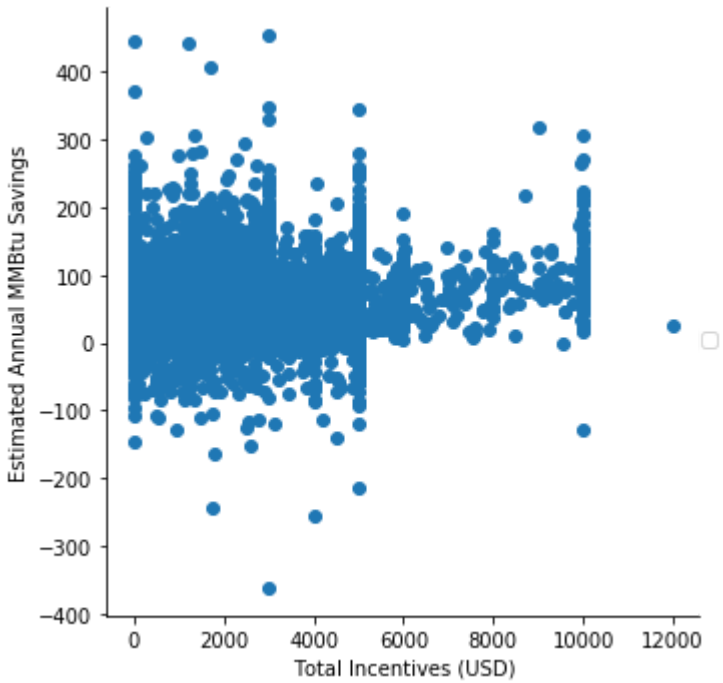
In [43]:
```
fig=sns.FacetGrid(d0,size=5)
fig.map(plt.scatter,'Total Project Cost (USD)','First Year Energy Savings $ Es
timate (USD)')
fig.add_legend()
plt.show()
```



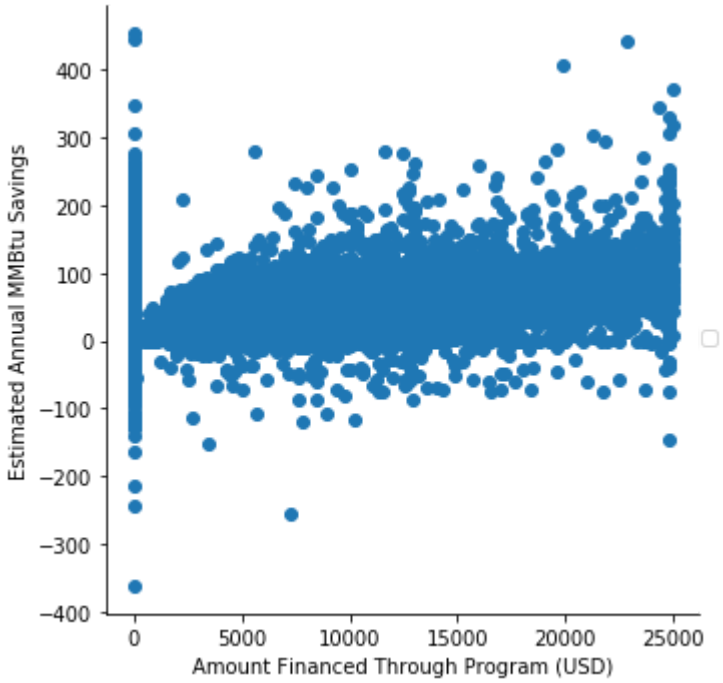
In [44]:
```
fig=sns.FacetGrid(d0,size=5)
fig.map(plt.scatter,'Total Project Cost (USD)','Estimated Annual MMBtu Saving
s')
fig.add_legend()
plt.show()
```



In [45]:
```
fig=sns.FacetGrid(d0,size=5)
fig.map(plt.scatter,'Total Incentives (USD)','Estimated Annual MMBtu Savings')
fig.add_legend()
plt.show()
```



In [47]:
```
fig=sns.FacetGrid(d0,size=5)
fig.map(plt.scatter,'Amount Financed Through Program (USD)','Estimated Annual
 MMBtu Savings')
fig.add_legend()
plt.show()
```



In [53]:
```
print(d0['Estimated Annual MMBtu Savings'].corr(d0['Total Project Cost (USD)'
]))
```

```
0.5819275741054862
```

In [51]:
```
print(d0['Estimated Annual MMBtu Savings'].corr(d0['Amount Financed Through Pr
ogram (USD)']))
```

```
0.42988907300114676
```

In [52]:
```
print(d0['Estimated Annual MMBtu Savings'].corr(d0['Total Incentives (USD)']))
```

```
0.27638852209516734
```

In [29]:
```

```

```
Natural gas     15657
Name: Pre-Retrofit Home Heating Fuel Type, dtype: int64
```

Observation

After analysing the above graph and correlation values we can say that 'Total Project Cost (USD)' has more effect on the'TEstimated Annual MMBtu Savings'.