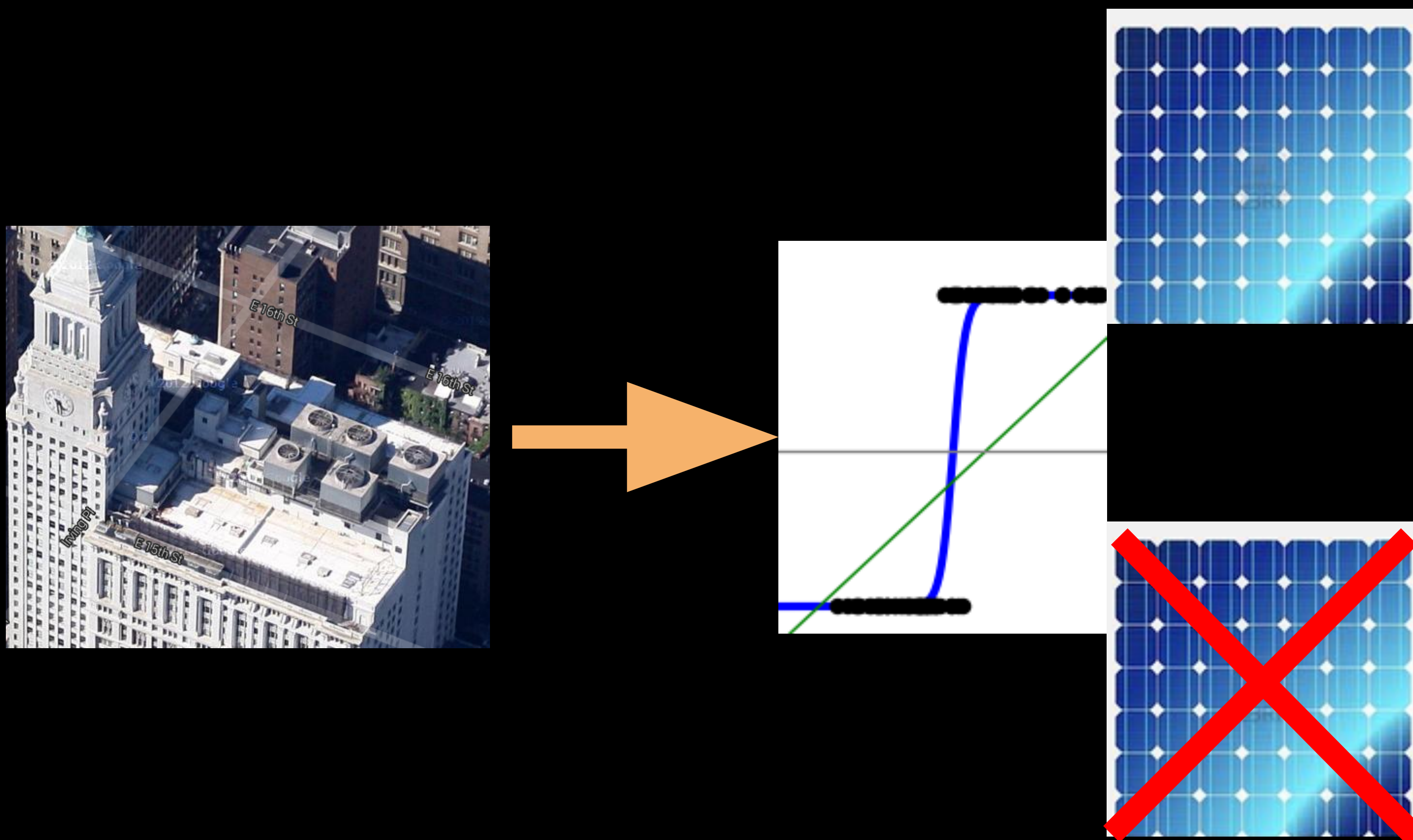
Distributed Generation Adoption Predictor

David Logsdon

Same page:

- Solar Photovoltaics (PV) have been rapidly adopted in NYC and Westchester
- Declining panel prices, healthy incentives, and high electric rates
- My company (ConEd) oversees grid interconnection and is interested in policy/rate issues and customer adoption

Conceptual Idea



Most time consuming...

```
######DATA MUNGING...
##Acount numbers are either 15 or 14 digits in either database, going to 14 digits to standardize for later joining
PV_MASTER['Acct_No'] = PV_MASTER['Acct_No'].map( lambda x : str(x)[0:14])
ROOFTOP_POTENTIALS['Acct_No'] = ROOFTOP_POTENTIALS['Acct_No'].map(lambda x : str(x)[0:14])

##Boro refers to the NYC Borrough (or WS for Westchester)
ROOFTOP_POTENTIALS['Boro_QN']= ROOFTOP_POTENTIALS['Boro'].map(lambda x: 1 if (x=='QN') else 0)
ROOFTOP_POTENTIALS['Boro_BX']= ROOFTOP_POTENTIALS['Boro'].map(lambda x: 1 if (x=='BX') else 0)
ROOFTOP_POTENTIALS['Boro_MN']= ROOFTOP_POTENTIALS['Boro'].map(lambda x: 1 if (x=='MN') else 0)
ROOFTOP_POTENTIALS['Boro_SI']= ROOFTOP_POTENTIALS['Boro'].map(lambda x: 1 if (x=='SI') else 0)
ROOFTOP_POTENTIALS['Boro_WS']= ROOFTOP_POTENTIALS['Boro'].map(lambda x: 1 if (x=='WS') else 0)

##Zonal Data for NYISO electricity market zone id.
ROOFTOP_POTENTIALS['Zone_H'] = 0
ROOFTOP_POTENTIALS['Zone_H'] = ROOFTOP_POTENTIALS['Zone'].map(lambda x: 1 if (x=='H') else 0)
ROOFTOP_POTENTIALS['Zone_H'] = ROOFTOP_POTENTIALS['Zone_H'].map(lambda x: int(x))
ROOFTOP_POTENTIALS['Zone_I'] = 0
ROOFTOP_POTENTIALS['Zone_I'] = ROOFTOP_POTENTIALS['Zone'].map(lambda x: 1 if (x=='I') else 0)
ROOFTOP_POTENTIALS['Zone_I'] = ROOFTOP_POTENTIALS['Zone_I'].map(lambda x: int(x))
ROOFTOP_POTENTIALS['Zone_J'] = 0
ROOFTOP_POTENTIALS['Zone_J'] = ROOFTOP_POTENTIALS['Zone'].map(lambda x: 1 if (x=='J') else 0)
ROOFTOP_POTENTIALS['Zone_J'] = ROOFTOP_POTENTIALS['Zone_J'].map(lambda x: int(x))

##Some fomatting on the entries to smooth them out for the logit model.
ROOFTOP_POTENTIALS['Max_kW'] = ROOFTOP_POTENTIALS['Max_kW'].replace(['#DIV/0!', np.nan], 0)
ROOFTOP_POTENTIALS['Max_kW'] = ROOFTOP_POTENTIALS['Max_kW'].map(lambda x : str(x).replace(',',''))
ROOFTOP_POTENTIALS['Max_kW'] = ROOFTOP_POTENTIALS['Max_kW'].map(lambda x : float(x))
ROOFTOP_POTENTIALS['min load'] = ROOFTOP_POTENTIALS['min load'].replace(['#DIV/0!', np.nan], 0)
```

Dep. Variable:	Has_PV	No. Observations:	211
Model:	Logit	Df Residuals:	187
Method:	MLE	Df Model:	23
Date:	Tue, 13 Aug 2013	Pseudo R-squ.:	0.5311
Time:	20:11:45	Log-Likelihood:	-68.584
converged:	False	LL-Null:	-146.25
		LLR p-value:	1.269e-21

	coef	std err	z	P>|z|	[95.0% Conf. Int.]	
Yr_kWh	-4.527e-06	2.72e-06	-1.665	0.096	-9.86e-06	8.01e-07
Max_kW	-0.4352	3.618	-0.120	0.904	-7.526	6.655
LF	4.4792	2.516	1.780	0.075	-0.452	9.411
min load	1.8446	14.470	0.127	0.899	-26.516	30.205
kw export	0.2043	2.03e+04	1.01e-05	1.000	-3.98e+04	3.98e+04
NumBldgs	-0.0952	0.143	-0.665	0.506	-0.376	0.186
BldgArea	8.472e-05	2.58e-05	3.288	0.001	3.42e-05	0.000
NumFloors	-1.0041	0.287	-3.494	0.000	-1.567	-0.441
BldgFront	-0.0067	0.005	-1.353	0.176	-0.016	0.003
BldgDepth	-0.0169	0.006	-2.830	0.005	-0.029	-0.005
AreaPerFloor	0.0035	0.007	0.481	0.630	-0.011	0.018
roof capacity	-0.3878	0.811	-0.478	0.632	-1.977	1.201
Zone_H	16.3462	4.75e+07	3.44e-07	1.000	-9.3e+07	9.3e+07
Zone_I	-0.2857	2.197	-0.130	0.897	-4.592	4.021
Zone_J	-1.5453	1.815	-0.851	0.395	-5.103	2.012
Boro_QN	-0.4704	1.000	-0.470	0.638	-2.430	1.489
Boro_BX	-1.4673	1.669	-0.879	0.379	-4.738	1.803
Boro_MN	-1.8803	1.773	-1.061	0.289	-5.355	1.595
Boro_SI	-0.5728	1.676	-0.342	0.732	-3.857	2.711
Boro_WS	14.7596	1174.483	0.013	0.990	-2287.184	2316.703
Cis_Lat	1.5726	6.362	0.247	0.805	-10.897	14.042
Cis_Long	0.8557	3.498	0.245	0.807	-5.999	7.711
Block	3.315e-05	8.41e-05	0.394	0.693	-0.000	0.000
Lot	-3.425e-05	0.000	-0.161	0.872	-0.000	0.000

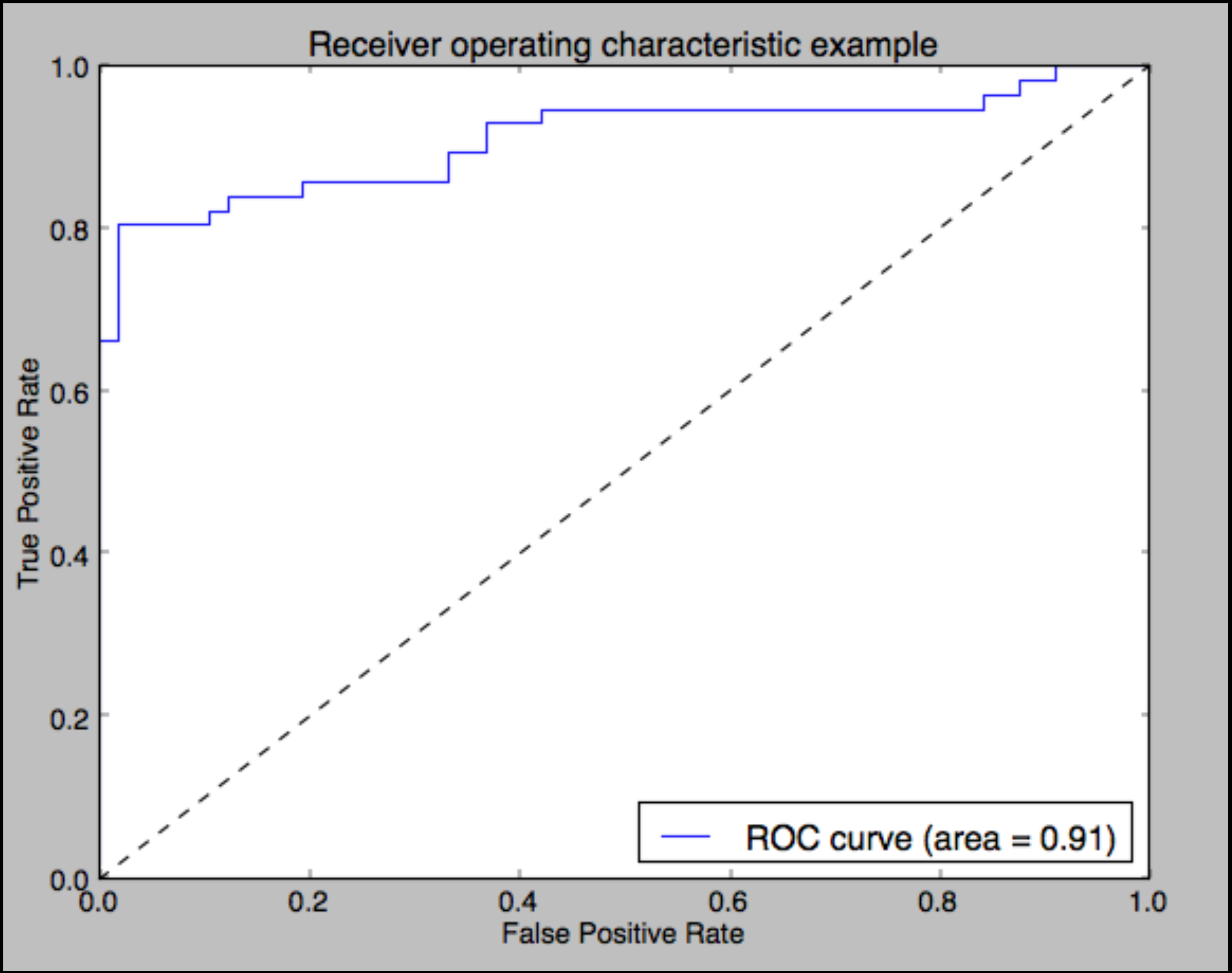
Dep. Variable:	Has_PV	No. Observations:	211
Model:	Logit	Df Residuals:	187
Method:	MLE	Df Model:	23
Date:	Tue, 13 Aug 2013	Pseudo R-squ.:	0.5311
Time:	20:11:45	Log-Likelihood:	-68.584
converged:	False	LL-Null:	-146.25
		LLR p-value:	1.269e-21

	coef	std err	z	P>|z|	[95.0% Conf. Int.]	
Yr_kWh	-4.527e-06	2.72e-06	-1.665	0.096	-9.86e-06	8.01e-07
Max_kW	-0.4352	3.618	-0.120	0.904	-7.526	6.655
LF	4.4792	2.516	1.780	0.075	-0.452	9.411
min load	1.8446	14.470	0.127	0.899	-26.516	30.205
kw export	0.2043	2.03e+04	1.01e-05	1.000	-3.98e+04	3.98e+04
NumBldgs	-0.0952	0.143	-0.665	0.506	-0.376	0.186
BldgArea	8.472e-05	2.58e-05	3.288	0.001	3.42e-05	0.000
NumFloors	-1.0041	0.287	-3.494	0.000	-1.567	-0.441
BldgFront	-0.0067	0.005	-1.353	0.176	-0.016	0.003
BldgDepth	-0.0169	0.006	-2.830	0.005	-0.029	-0.005
AreaPerFloor	0.0035	0.007	0.481	0.630	-0.011	0.018
roof capacity	-0.3878	0.811	-0.478	0.632	-1.977	1.201
Zone_H	16.3462	4.75e+07	3.44e-07	1.000	-9.3e+07	9.3e+07
Zone_I	-0.2857	2.197	-0.130	0.897	-4.592	4.021
Zone_J	-1.5453	1.815	-0.851	0.395	-5.103	2.012
Boro_QN	-0.4704	1.000	-0.470	0.638	-2.430	1.489
Boro_BX	-1.4673	1.669	-0.879	0.379	-4.738	1.803
Boro_MN	-1.8803	1.773	-1.061	0.289	-5.355	1.595
Boro_SI	-0.5728	1.676	-0.342	0.732	-3.857	2.711
Boro_WS	14.7596	1174.483	0.013	0.990	-2287.184	2316.703
~~Cis_Lat~~	~~1.5726~~	~~6.362~~	~~0.247~~	~~0.805~~	~~-10.897~~	~~14.042~~
~~Cis_Long~~	~~0.8557~~	~~3.498~~	~~0.245~~	~~0.807~~	~~-5.999~~	~~7.711~~
~~Block~~	~~3.315e-05~~	~~8.41e-05~~	~~0.394~~	~~0.693~~	~~-0.000~~	~~0.000~~
~~Lot~~	~~-3.425e-05~~	~~0.000~~	~~-0.161~~	~~0.872~~	~~-0.000~~	~~0.000~~

Logit Regression Results

Dep. Variable:	Has_PV	No. Observations:	211
Model:	Logit	Df Residuals:	206
Method:	MLE	Df Model:	4
Date:	Tue, 13 Aug 2013	Pseudo R-squ.:	0.3483
Time:	22:27:27	Log-Likelihood:	-95.307
converged:	True	LL-Null:	-146.25
		LLR p-value:	3.896e-21

	coef	std err	z	P>|z|	[95.0% Conf. Int.]	
Yr_kWh	3.514e-07	2.91e-07	1.207	0.228	-2.19e-07	9.22e-07
LF	1.8311	0.762	2.403	0.016	0.338	3.324
BldgArea	0.0001	2.36e-05	4.486	0.000	5.97e-05	0.000
NumFloors	-1.3127	0.277	-4.746	0.000	-1.855	-0.771
BldgDepth	-0.0117	0.003	-4.089	0.000	-0.017	-0.006


Receiver operating characteristic example
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.91)

Further things to try

- Explore: predictor or detector?
- Naive Bayes model, regularization
- Use weights
- Locational Data → dummy variables, bins
- Build a more complete training set (156 out of 1,400, mostly commercial customers)
- Bring in other databases
- New technology types (small CHP)