KEENE MASS APPRAISAL CONSULTING
SENSIBLE SOLUTIONS.
Keenemac.com (267)307-6071

Size Codes

KEVIN KEENE

What Are Size Codes?

Size codes are an example of feature engineering that creates categorical variables from a range of building size – building square footage or building square meters

Originating in regression models for Philadelphia, they are used in addition to a continuous building size curve or linear rate to capture market preferences or penalties associated with the 'right size' or 'wrong size' building. These preferences may be present in some markets and can explain unique or peculiar market dynamics

Size Codes

KEENE MASS APPRAISAL CONSULTING
SENSIBLE SOLUTIONS.

Size codes were developed and used in *all* regression models for Philadelphia – residential and non-residential alike. Typically, between 1 and 3 sizes are significant in every model (except for the office building sales model), with different combinations of preferences and penalties across the entire model catalog. They are not co-linear with other attributes in any of the models, and greatly improve the explanatory power and performance of most models. It is fair to say that the market could not be understood without them!

In Philadelphia 5 categories were created – smallest, small, average, large and largest, but you may create as many or as few as needed to reflect the preferences and nuances of your market. You can base your categories on a bell curve or an equal distribution, such as quintiles.

Use Case Examples

An aging population may place a premium on smaller homes. 'Empty Nesters' will sell larger houses to downsize into more manageable space. If the supply of smaller homes is limited, you might see significantly higher prices per square foot – possibly even higher raw prices– for smaller homes within the same neighborhoods as larger ones

In prestigious neighborhoods there may be a premium on larger houses that are more suitable for entertaining

When energy costs are high, buyers may avoid large houses that are more difficult to heat or cool in favor of smaller houses

Shifts in the mortgage industry may create demand for larger or smaller houses. When the industry shifted from supporting mostly single wage earner households to those with multiple wage earners, larger houses became more desirable in some areas

These are just some examples, but all of them would require adjustments that would not be reflected in a continuous curve based on building size to capture market dynamics

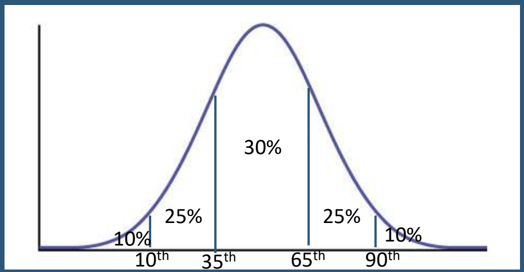
Method for Creating Size Codes

The steps in the process

Determine the distinct classes of property for which you want codes
(See sample table on next slide)

For each class, array the inventory by building size in ascending or descending order

Navigate to your desired breakpoints – we used the 10th 35th 65th and 90th percentiles - to create 5 groups centered on the median size



Look for 'Natural' breaks to determine limits – You don't want an 1835 sf house in a different group than an 1836 sf house!

Assign values to all properties within each size range

Transform data into categorical attributes by size code

Sample Size Code Table

Categories created for Philadelphia's Inventory

Some classes had to be broken into more discrete sub-classes.

If all industrial had been grouped together, most of the '5's would have been warehouses

If all apartment buildings had been grouped together, most of the '5's would have been high-rises

Row houses tend to be smaller than twins, which tend to be smaller than singles, so we used different breakpoints for each type.

For any given group or sub-group, the logic of smallest/small/average/large/largest holds up well.

In every model, the size code associated with each account was used, so there were only 4 binaries (smallest/small/large/largest) entered in each model. We didn't have a set of binaries for each of rows, twins and singles, as the number of observations required to support that many binaries would have been prohibitive in many cases

Size Codes

Accounts	Style	Flpn	Desc	1 min	1 max	2 min	2 max	3 min	3 max	4 min	4 ma›	5 min	5 max
	%		Percentile Ranges	0	10	11	35	36	65	66	90	91	100
86	1	*	Air Rights / Subterranian	0	0	1	1500	1501	10000	10001	70000	70001	1000000
3918	2	*	NonResidential Condo	0	0	0	0	0	0	0	0	0	0
30151	2	2	Residential Condo	1	600	601	900	901	1300	1301	2000	2001	600000
	4	*	Stacked PUD	1	600	601	900	901	1300	1301	2000	2001	600000
4125	A	*	Apartments	1	4	5	18	19	32	33	63	64	1000
51	A	H	Apartments (Hi Rise)	1	30	31	110	111	200	201	300	301	1000
5633	C	*	Retail(Other)	1	1199	1200	2200	2201	4999	5000	15000	15001	2000000
22	C	A/B	Anchor,Big Box or Ret Wai	1	80000	80001	110000	110001	140000	140001	200000	200001	500000
14674	D	*	Mixed Use	1	1300	1301	1700	1701	2500	2501	3600	3601	2000000
761	E	*	Schools	0	5000	5001	22000	22001	60000	60001	160000	160001	3000000
2099	G	*	Garages	1	325	326	670	671	1440	1441	4400	4401	1000000
153	H	*	Hotels	1	2149	2150	3000	3001	9400	9401	200000	200001	1500000
3241	I	*	Industrial Other	1	1300	1301	2800	2801	7900	7901	45000	45001	2000000
3000	I	W	Industrial Warehouses	1	1450	1451	3500	3501	11100	11101	48000	48001	3000000
	J	*	Recreation and Assembly	1	1100	1101	2000	2001	7200	7201	20000	20001	2000000
12264	L	*	Land (NonResidential)	1	600	601	1070	1071	2500	2501	23000	23001	200000000
33369	L	R	Land (Residential)	1	600	601	785	786	1200	1201	2640	2641	20000000
36936	M	*	Multi-Family	1	1300	1301	1600	1601	2200	2201	2800	2801	160000
786	O	*	Offices (Other)	1	2050	2051	4300	4301	11000	11001	36000	36001	2000000
74	O	A	Offices (HiRise)	1	60000	60001	270000	270001	700000	700001	1E+06	1E+06	9000000
108	O	B	Offices (MidRise)	1	5000	5001	28000	28001	120001	120001	400000	400001	2000000
1178	P	*	Public Use	1	1200	1201	3999	4000	22000	22001	72000	72001	2000000
330694	R	*	Row Dwellings	1	885	886	1058	1059	1300	1301	1599	1600	2000000
27021	S	*	Detached Dwellings	1	1176	1177	1470	1471	2176	2177	3300	3301	50000
65044	T	*	Twins Dwellings	1	1080	1081	1260	1261	1652	1653	2126	2127	1000000
1610	W	*	Churches	1	1899	1900	3499	3500	11599	11600	25000	25001	400000
265	Y	*	Health Facilities	1	1620	1621	3200	3201	40000	40001	230000	230001	1000000

Will It Work For You?

Experiment with the technique and test it in some of your models. If it provides additional explanatory power I hope you can use it in your practice!

THANK YOU!

KEENE MASS APPRAISAL CONSULTING

SENSIBLE SOLUTIONS.

Keenemac.com (267)307-6071