INTERNATIONAL RESEARCH SYMPOSIUM
Amsterdam, The Netherlands · December 4 - 5, 2024

ASSESSMENT INNOVATION & COLLABORATION WITH A FOCUS ON AI

Representation Matters

IAAO

IIFI
IPTI
International Property Tax Institute

Representation Matters

Measuring Model-Market Representation, a HI heuristic to simplify AI complexity

James Williams Office of Property Assessment
City of Philadelphia

Alex Raju Office of Property Assessment
City of Philadelphia

Kevin Keene Keene Mass Appraisal Consulting

Thanks For Joining Us

Philadelphia – Founded in 1650 by William Penn

- 587,000 parcels with a Total Assessed Value of $231 billion
- Not driven by any specific industry, or characterized by one identity
- Over 490,000 Residential parcels
- Home to Penn, LaSalle, Drexel, Temple, St. Joseph's and many smaller colleges and universities
- 24 large office buildings
- 4,500 industrial parcels
- 42,000 Vacant land parcels
- No farms. No oil or gas

Commercial Real Estate

Drexel University buys 'highly strategic' 7-Eleven site adjacent to campus

Drexel University acquired 3401 Lancaster Ave., where a 7-Eleven has operated for more than 30 years adjacent to the school's campus.

PAUL SCHWEDELSON

By Paul Schwedelson – Reporter, Philadelphia Business Journa
Aug 19, 2024

Drexel University bought a 7-Eleven convenience store site t
eventually allow for campus expansion.

Commercial Real Estate

Temple University acquires key North Broad Street property for future expansion

The recently acquired parcel at 2034 N. Broad St.

GOOGLE MAPS

By Ryan Mulligan – Reporter, Philadelphia Business Journal
Apr 7, 2023 **Updated** Apr 7, 2023 3:49pm EDT

The university is formulating a new campus master plan that will likely include development on the recently acquired parcel, said Temple COO Ken Kaiser.

Are They Similar?

A well-represented sales file is the foundation for a reliable and effective multiple regression model. It enhances data quality, feature engineering, and validation, ultimately leading to more accurate property value estimates.

The Number 1 Rule In Real Estate

Representation of Unsold Properties

Whenever we use ratio studies to analyze assessment performance, or use sales based methods to estimate property value, we are assuming:

That properties that sell are similar to properties that do not sell

- AND -

That representation of unsold properties is proportionate to sales activity

Is it safe or reasonable to make these assumptions?

How can we test the validity of these assumptions?

Two Methods for Measuring Representation

This session will discuss and demonstrate two methods for measuring and understanding representation of unsold properties and identifying properties that are not represented by sales

Keene Group Summary (KGS)

Assigns properties to groups using a common schema (Group ID)
Summarizes and presents data about each group

Heuristic Measure of Representation (HMR)

Assigns properties to groups defined by each regression model
Derive a score for each property that reflects the degree of generalization

Uses

Gain more precise insight into model and assessment performance

Facilitate review of valuation projections

Identify submarkets that are not well served by the valuation process

Representation of Unsold Properties

Whenever we cannot make direct observations or comparisons, we must generalize or make assumptions

With every degree of difference, estimates of value based on generalizations become less accurate and less reliable

Degrees of difference typically include

Neighborhoods

Construction quality

Condition

Age

Property types

Building or lot sizes

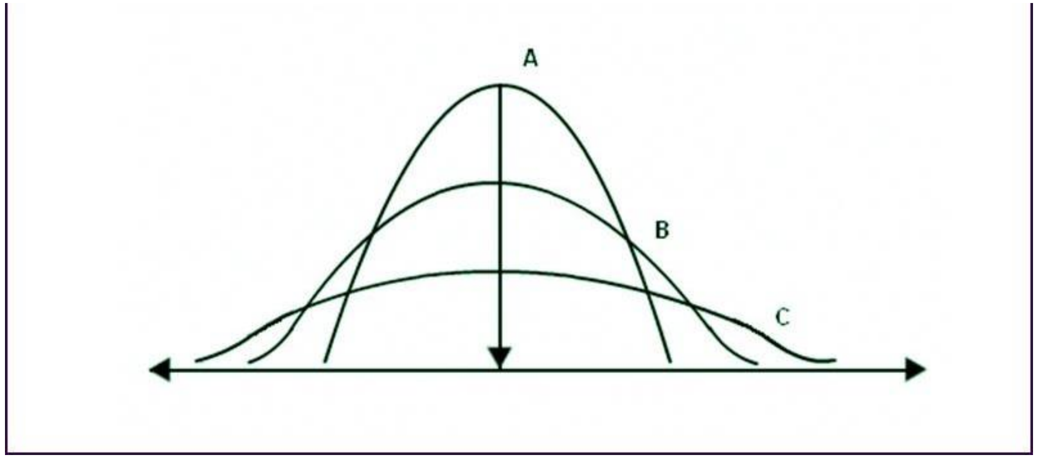
Price or value class

Model Performance

Performance metrics

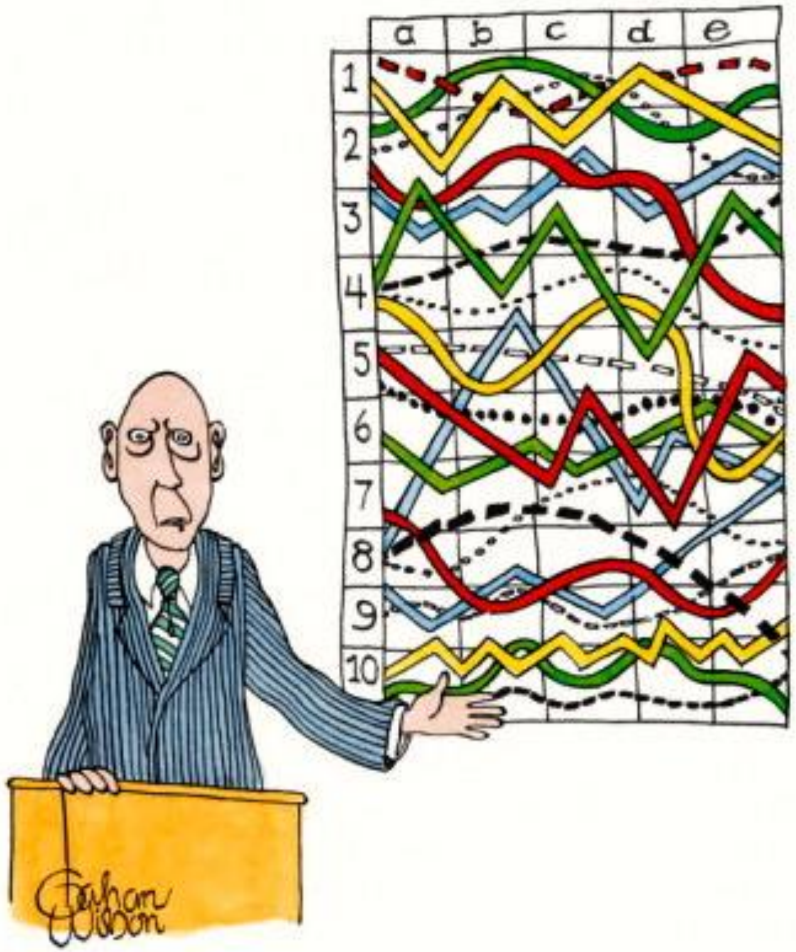
R^2, COD, PRD, PRB, MEDIAN, etc.

- Adjusted R Square: 0.925
- Median: 0.993
- Price Related Differential: 1.021
- Price Related Bias: 0.015
- Coefficient of Dispersion: 0.101



Performance can be measured as a whole, by independent variables and/or independent variable groups/subsets

We need to identify the cases or submarkets where our models are *not* working well



"I'll pause for a moment so you can let this information sink in."

Group Summary Method

Why Use Grouping IDs?

It is not particularly difficult to make good decisions in appraising properties. The hard part of mass appraisal is making sure that, when decisions are applied, they affect

- Every property that should be affected and
- No properties that should not be affected

Group IDs allow us to designate properties as members of groups and make decisions at the group level. This ensures that all properties in the group are affected equally. It also allows us to have different methods, adjustment coefficients and techniques for different groups of properties. We can also keep aggregate or summary data for all groups and easily publish that data to the world.

Group IDs make databases much more efficient, avoiding multi-key joins between tables and simplifying retrieval of data through queries.

The Power of Persistent Groups

Consistent means that the group is always the same, no matter who accesses the data.
Persistent means that the group, and data about the group, exists at all times.

Every account has an Group ID. There are around 59,000 distinct Group IDs in the Philadelphia data.

Every transaction has a SGroup ID. There are around 23,000 groups with 1 or more validated sales.

Because the groups are both Consistent and Persistent, we can maintain data about Group IDs and SGroup IDs.

Every account can be linked to the data about it's Group ID and SGroup ID groups, including the number of accounts, number of valid sales, average size, wtd SP/SF, average sale price, median ratio of assessment, and many others.

We can compare any given account to what is typical for the group, allowing us to find those that are at significant variance. How well does the sale price of a new transaction match what we know about what is typical for the group? How well does a specific value align with other properties in the group?

We can also compare attributes of a set of records to attributes of records that are NOT in the dataset. This is called 'what is' to 'what is not' comparison.

You can't do this if you don't group, or if you group 'on the fly'!

Aggregating Information

CAMA systems keep a lot of information about specific accounts. We also keep information about groups of accounts. What are some of the things that we know about groups?

Number of properties in the group

Number of sales in the group

Level of Sales Activity

Average Price

Price per SF

MV per SF

Range of MV per SF

Median Level of Assessment

Land to Building Ratio

Average MV

Number of groups on each block

Groups with few or no sales

All of these things can be related to other aggregates, or compared to specific accounts

Anatomy of Group ID

Group IDs provide a 'snapshot' of a property, bringing together the most important elements that describe a property in one place

B211-TJ5-34-0

Location

B211

Zone

Group

Area

GMA

Characteristics

TJ5 - 34

Property Type

Size

Era Built

Condition

Special Considerations

0

Sub Group

Summary Tables

GroupId	Accounts	TotalProjMV	AvgProjMV	MinProjMV	MaxProjMV	ProjMVRange	AvgProjMVSF	WtdAvgProjN	MinProjMVSF	MaxProjMVS	ProjMVSFRange	TotalBldSF	AvgBldSF	MinBldSF	MaxBldSF	BldSFRange	AvgLotSF	MinLotSF	MaxLotSF	LotSFRange	AvgYearBlt	MinYearBlt	MaxYearBlt	YearBltRange
A111LRSA540	2	$94,000	$47,000	$41,500	$52,500	$11,000	$23.73		$23.73	$27.93	$4.20						2002	1880	2125	245				
A111M2B0514	2	$411,100	$205,550	$200,400	$210,700	$10,300	$169.97	$169.81	$169.97	$173.51	$3.54	2421	1210	1155	1266	111	2382	2097	2666	569	1950	1949	1950	1

GroupId	Accounts	ValidSales	TotalProjMV	AvgProjMV	MinProjMV	MaxProjMV	ProjMVRange	AvgProjMVSF	WtdAvgProjN	MinProjMVSF	MaxProjMVS	ProjMVSFRar	TotalBldSF	AvgBldSF	MinBldSF	MaxBldSF	BldSFRange	AvgLotSF	MinLotSF	MaxLotSF	LotSFRange	AvgYearBlt
A111RPF2533	4	3	$887,100.00	$221,775.00	$218,100.00	$226,300.00	$8,200.00	$198.01	$198.01	$198.01	$202.05	$4.04	4480	1120	1120	1120	0	1475	1470	1484	14	1949
A111RPF2534	44	5	$6,835,000.00	$155,340.91	$151,900.00	$158,000.00	$6,100.00	$138.70	$138.70	$138.70	$141.07	$2.37	49280	1120	1120	1120	0	1480	1468	1493	25	1949
A111RPW2434	1		$152,300.00	$152,300.00	$152,300.00	$152,300.00	$0.00	$125.25	$125.25	$125.25	$125.25	$0.00	1216	1216	1216	1216	0	1474	1474	1474	0	1945
A111RPW2531	1		$225,400.00	$225,400.00	$225,400.00	$225,400.00	$0.00	$201.25	$201.25	$201.25	$201.25	$0.00	1120	1120	1120	1120	0	1755	1755	1755	0	1949
A111RPW2532	60	58	$18,097,500.00	$301,625.00	$275,500.00	$353,700.00	$78,200.00	$261.07	$260.61	$261.07	$315.80	$54.73	69444	1157	1120	1296	176	1791	1301	4262	2961	1949
A111RPW2533	170	73	$41,441,400.00	$243,772.94	$212,800.00	$306,500.00	$93,700.00	$213.50	$213.23	$213.50	$252.59	$39.09	194350	1143	1120	1296	176	1776	1292	9000	7708	1949
A111RPW2534	2029	118	$344,459,200.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1949
A111RPW2535	8	10	$794,900.00																			1949
A111RPW2536	2	1	$147,200.00																			1949
A111RPW2542	1	1	$318,600.00																			1950
A111RPW2543	16	7	$4,202,300.00																			1949
A111RPW2544	177	11	$32,449,700.00																			1949
A111RPW2552	1	1	$357,700.00																			1949
A111RPW2553	1		$293,500.00																			1949
A111RPW2554	3		$611,300.00																			1949
A111RPW3533	1	1	$252,800.00																			1949
A111RPW3534	4		$695,100.00																			1949
A123M2C0315	1		$105,100.00																			1920

GroupId	Parcel_ID	ParIdNum	OPA_ACCOUI	PROPID	SEC_FLD	BLOCK_ID	ADDRESS	C
A111RPW2534	1001483344	1001483344	344166100	7262007312	3257	7262007300	7312 SHERWOOD RD	0
A111RPW2534	1001675471	1001675471	343278700	8947001334	3257	8947001300	1334 N 75TH ST	0
A111RPW2534	1001675468	1001675468	343278400	8947001328	3257	8947001300	1328 N 75TH ST	0
A111RPW2534	1001675472	1001675472	343278800	8947001336	3257	8947001300	1336 N 75TH ST	0
A111RPW2534	1001675473	1001675473	343278900	8947001338	3257	8947001300	1338 N 75TH ST	0
A111RPW2534	1001675474	1001675474	343279000	8947001340	3257	8947001300	1340 N 75TH ST	0
A111RPW2534	1001675475	1001675475	343279100	8947001342	3257	8947001300	1342 N 75TH ST	0
A111RPW2534	1001675476	1001675476	343279200	8947001344	3257	8947001300	1344 N 75TH ST	0
A111RPW2534	1001675477	1001675477	343279300	8947001346	3257	8947001300	1346 N 75TH ST	0
A111RPW2534	1001675478	1001675478	343279400	8947001348	3257	8947001300	1348 N 75TH ST	0
A111RPW2534	1001106925	1001106925	343213100	1926007422	3257	1926007400	7422 BROOKHAVEN RD	0
A111RPW2534	1001106921	1001106921	343212800	1926007416	3257	1926007400	7416 BROOKHAVEN RD	0
A111RPW2534	1001099866	1001099866	343221700	1874007530	3257	1874007500	7530 BRENTWOOD RD	0

These tables store information about persistent groups. Every account can be linked to these tables by Group ID, so that any account can be compared to the summary data and all accounts in a group can be identified

Using Summaries to Understand Representation

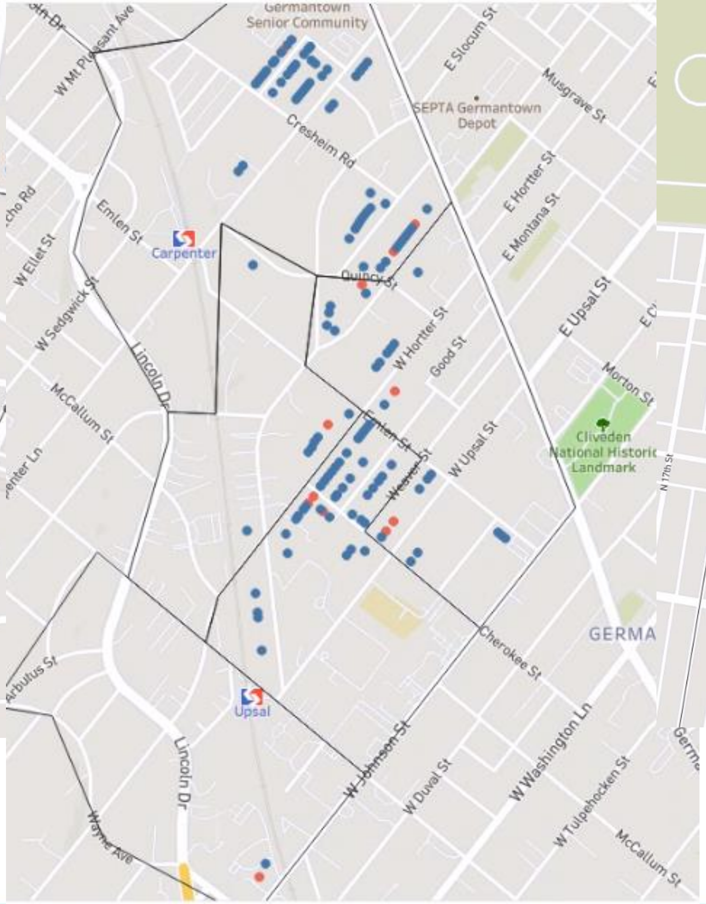
Properties will not be proportionately represented by sales as sales may represent different numbers of accounts

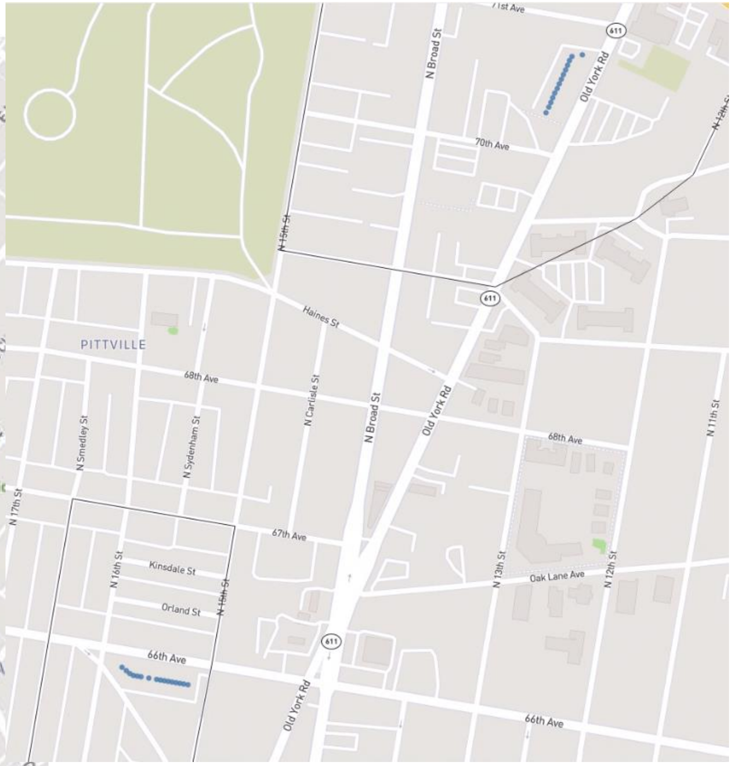
Some groups of properties may be represented by few or no sales

GroupId	Accounts	Agv MV	Min MV	Max MV	MVRange	WtdAvgMVSF	Min / Max MVSF		*Range*	AvgBldSF	AvgLotSF
	Valid Sales	Agv TASP	MinTASP	MaxTASP	TASPRange	WtdAvgTASPSF	Min / Max TASPSF			AvgSBldSF	AvgSLotSF
M111RPF2344	301	$165,418	$155,200	$185,900	$30,700	$117.22	$117.41	$130.34	$12.93	1,411	1,522
	17	$170,057	$139,064	$191,300	$52,236	$116.82	$96.36	$131.19	$34.83	1,456	1,529
M612TOS3353	43	$531,742	$440,300	$816,700	$376,400	$192.35	$196.39	$242.49	$46.10	2,764	4,072
	22	$500,055	$366,316	$658,954	$292,638	$192.46	$123.13	$272.45	$149.33	2,598	4,169
M621TOS3354	49	$473,682	$432,500	$603,400	$170,900	$200.34	$200.74	$216.00	$15.26	2,364	3,303
	2	$463,211	$453,694	$472,728	$19,034	$204.42	$202.18	$206.61	$4.43	2,266	3,225
M111RPW2444	16	$179,244	$178,400	$179,300	$900	$119.18	$119.18	$119.22	$0.04	1,504	2,400
	2	$176,882	$174,114	$179,650	$5,536	$117.61	$115.77	$119.45	$3.68	1,504	2,400
M111RPW2445	1	$111,100	$111,100	$111,100	$0	$73.87	$73.87	$73.87	$0.00	1,504	2,400
	1	$109,275	$109,275	$109,275	$0	$72.66	$72.66	$72.66	$0.00	1,504	2,400
M111RPW2454	30	$187,770	$181,800	$204,700	$22,900	$112.57	$112.68	$120.30	$7.61	1,668	2,707

Mapping Representation

Properties Sold and Unsold by GroupID





Group Summaries

Here's a sample breakdown of representation by groups of similar properties

	Groups	Accts	Sales	Pct of Groups	Pct of Accts	Pct of Sales	Representation Pct
SF Total	35,233	422,996	54,966				13%
1 account	15,708	15,708	4,094	45%	4%	7%	26%
Lt 5 Accounts	25,245	40,836	10,905	72%	10%	20%	27%
100 or more Accts	845	201,453	16,030	2%	48%	29%	8%
500 or more Accts	65	47,401	3,530	0.2%	11%	6%	7%
No Sales	20,951	59,763	-	59%	14%	0%	0%
At least 1 sale	14,282	363,233	54,966	41%	86%	100%	15%
At least 3 sales	5,002	291,239	43,126	14%	69%	78%	15%
Less than 3 sales	30,231	131,757	11,840	86%	31%	22%	9%
10 or more sales	1,141	172,065	24,895	3%	41%	45%	14%
Condition 7	1,337	3,715	745	4%	1%	1%	20%
Condition 6	1,088	2,094	954	3%	0%	2%	46%
Condition 5	3,234	11,889	4,044	9%	3%	7%	34%
Condition 4	18,109	352,270	26,092	51%	83%	47%	7%
Condition 3	6,902	31,406	10,861	20%	7%	20%	35%
Condition 2	3,625	14,007	8,552	10%	3%	16%	61%
Condition 1	898	7,572	3,718	3%	2%	7%	49%

There are 21,000 groups representing almost 60,000 properties that are not represented by any sales

We can identify every property that is in any of these groups

Distribution of Unrepresented Properties

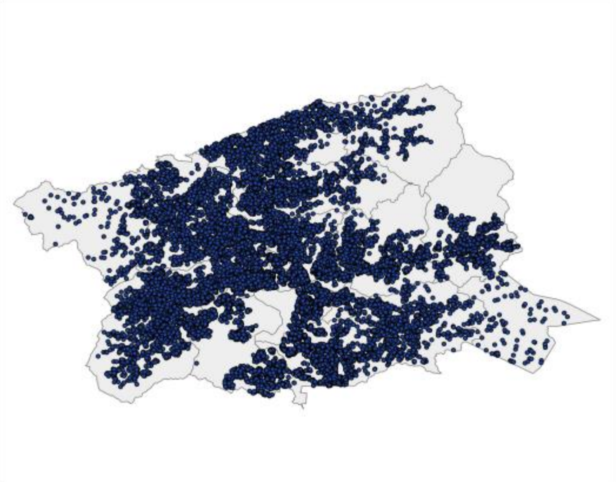
ValueClass * Represented2 Crosstabulation

			Represented2 No	Represented2 Yes	Total
ValueClass	Below 120k	Count	7288	2777	10065
		% within ValueClass	72.4%	27.6%	100.0%
	120k to 175k	Count	6123	3648	9771
		% within ValueClass	62.7%	37.3%	100.0%
	175k to 215k	Count	4963	5088	10051
		% within ValueClass	49.4%	50.6%	100.0%
	215k to 250k	Count	4651	5866	10517
		% within ValueClass	44.2%	55.8%	100.0%
	250k to 282k	Count	3883	5585	9468
		% within ValueClass	41.0%	59.0%	100.0%
	282k to 322k	Count	4363	5833	10196
		% within ValueClass	42.8%	57.2%	100.0%
	322k to 374k	Count	4643	5329	9972
		% within ValueClass	46.6%	53.4%	100.0%
	374k to 460k	Count	4791	5289	10080
		% within ValueClass	47.5%	52.5%	100.0%
	460k to 640k	Count	4820	5063	9883
		% within ValueClass	48.8%	51.2%	100.0%
	640k+	Count	4504	3670	8174
		% within ValueClass	55.1%	44.9%	100.0%
Total		Count	50029	48148	98177
		% within ValueClass	51.0%	49.0%	100.0%

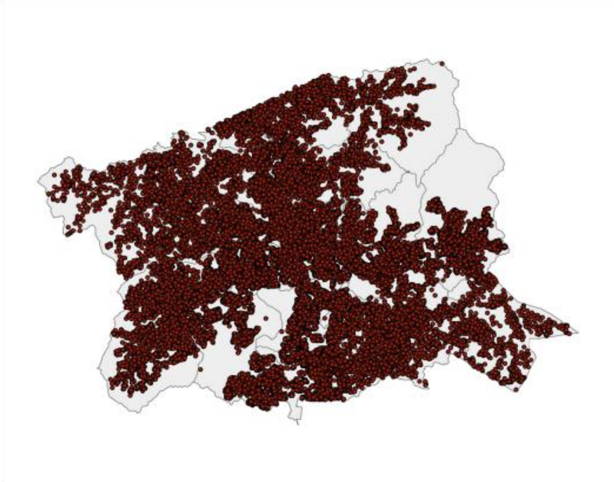
Condition * Represented2 Crosstabulation

		Represented2 No	Represented2 Yes	Total
Condition	Fair	2775	201	2976
		93.2%	6.8%	100.0%
	Good	4923	5871	10794
		45.6%	54.4%	100.0%
	Normal	36784	39281	76065
		48.4%	51.6%	100.0%
	Poor	1102	52	1154
		95.5%	4.5%	100.0%
	Renovated	3020	2737	5757
		52.5%	47.5%	100.0%
	Unsound	260	6	266
		97.7%	2.3%	100.0%
Total		50029	48148	98177
		51.0%	49.0%	100.0%

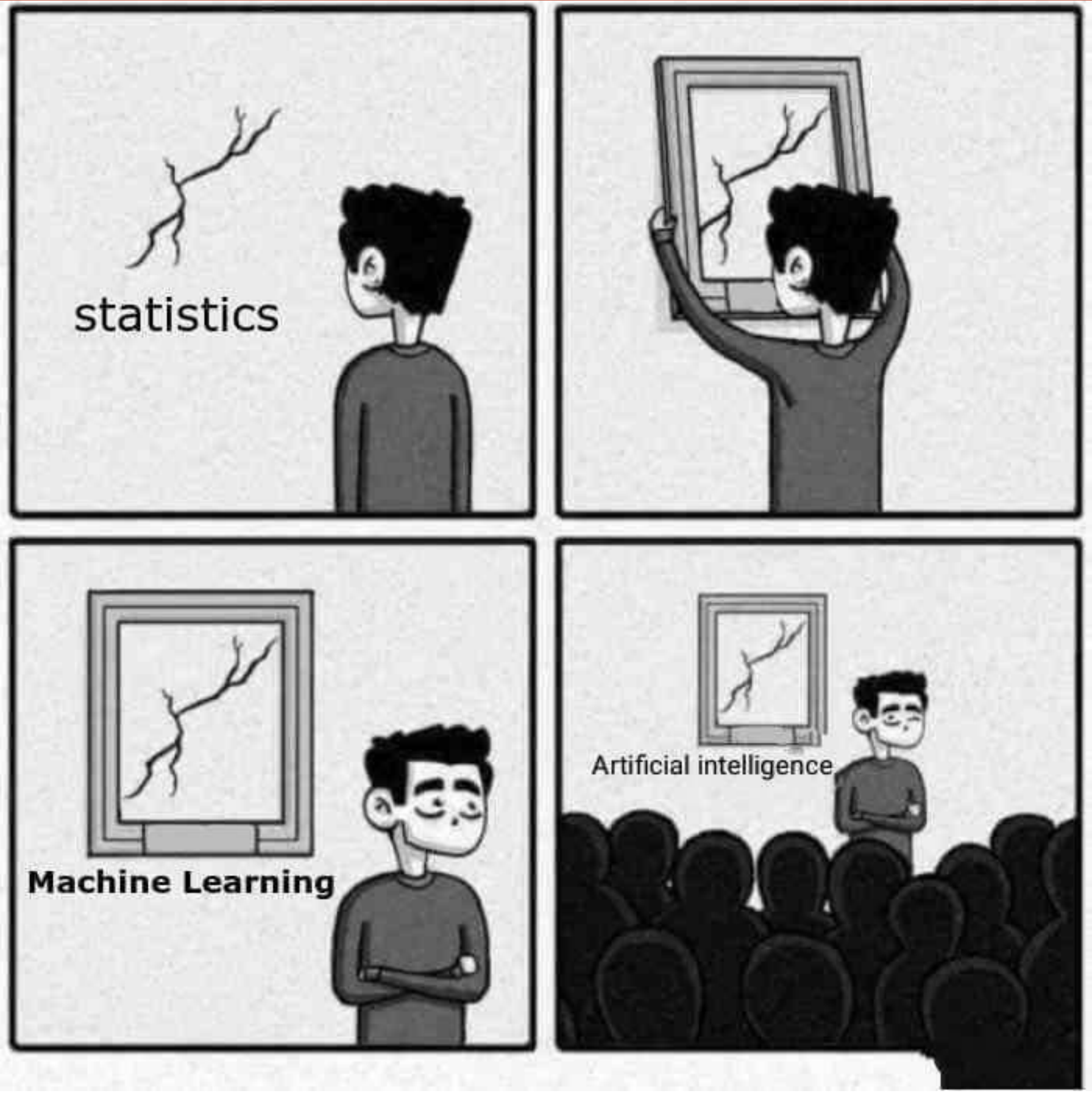
Buncombe County Represented vs Unrepresented Properties



Represented

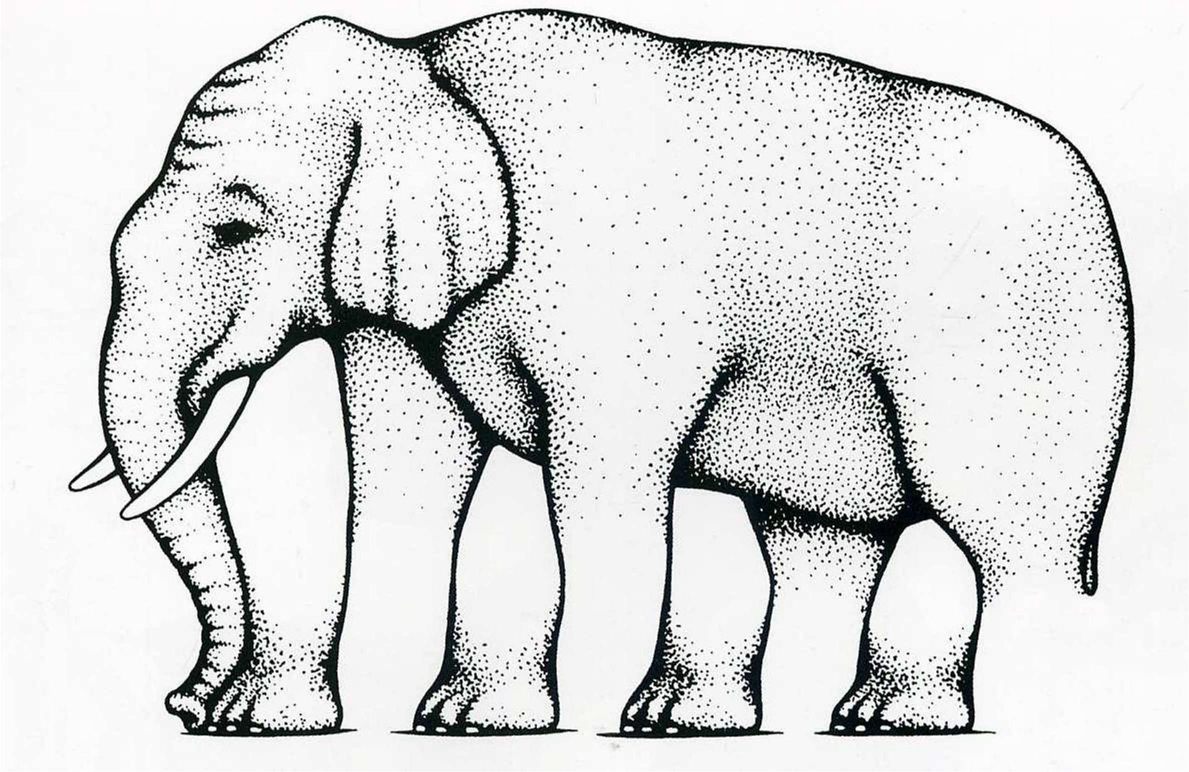


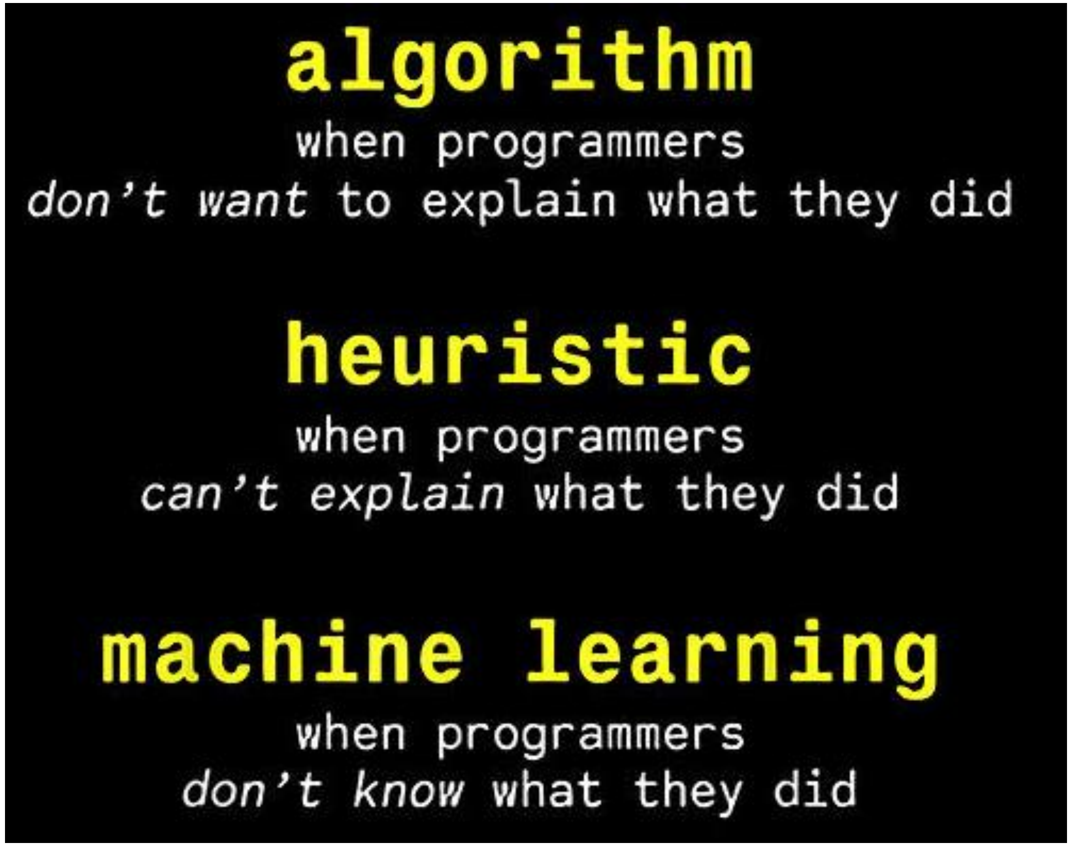
Unrepresented



original comic by sandserif

Heuristic Measure of Representation





Projections

What happens when a binary variable does not develop a co-efficient?

- It gets treated like the base
- Contributory value is Generalized for that Attribute

Why does it happen?

- Under-represented / Not represented in Sales, unable to generate coefficient with significance.

Are there ways to identify and quantify the degree of generalization in a projection / prediction?

HMR Scoring Algorithm

HMR requires the use of regression models

Identify the base binary for each categorical variable

Identify Binaries that developed coefficients.

Assign a score for each categorical variable that is

- NOT used as the binary base

AND

- DID NOT generate a coefficient.

Tally scores for each case

Heuristic Measure of Representation

Final Score = Degree of Generalization

Each case will get a Score

All cases in a group will have the same Score

Scores will range from 0 to n, where n is the number of unrepresented binary variables

Higher Scores indicate :

Higher Degree of Generalization

Lower Degrees of Representation

by the Model

Simple Example

- **Condition**
  - NEW
  - REHAB
  - ABOVE AVERAGE
  - *Average (base)*
  - BELOW AVERAGE
  - POOR
  - SEALED
- **Location**
  - LOC_M111
  - LOC_M112
  - *LOC_M123 (base)*
  - LOC_M131
  - LOC_M132
  - LOC_M133
  - LOC_M141
  - LOC_M143

Coefficients	
REHAB	0.355
ABOVE_AVG	0.210
BELOW_AVG	-0.478
POOR	-0.731
SEALED	-1.131

LOC_M111	-0.048
LOC_M131	-0.214
LOC_M133	-0.205
LOC_M141	-0.325

- **Style**
  - *Row (base)*
  - Single
  - Twin

SINGLE	0.214

NEW; Twin; LOC_M112; LOC_M132; LOC_M143 will generate scores of '1'.

All bases & binaries that have coefficients will generate scores of '**0**'

For this model the schema will be location/condition/style ***'M111S4'***

Each model will specify a different grouping schema

HMR Scoring Example

ID	ADDR	Locale	Style	Condition	PredMV	L_Score	C_Score	S_Score	TOT_SCORE
8386000300	322 WINONA ST	M111	R	3	$ 280,000.00	0	0	0	0
5136000600	607 LOCUST AVE	M112	T	5	$ 110,000.00	1	0	1	2
2390000100	126 E CLIVEDEN ST	M113	S	3	$ 229,900.00	1	0	0	1
1746007400	7426 BEVERLY RD	M131	S	1	$ 245,000.00	0	1	0	1
3108001800	1825 W ELEANOR ST	M132	R	3	$ 140,000.00	1	0	0	1
8196001400	1412 E WEAVER ST	M133	R	4	$ 115,000.00	0	0	0	0
2296005500	5518 CHEW AVE	M141	T	4	$ 131,500.00	0	0	1	1
3666006300	6375 GERMANTOWN AVE	M143	R	1	$ 242,500.00	1	1	0	2
2330000000	65 E CLAPIER ST	M143	R	4	$ 275,000.00	1	0	0	1
1650006400	6415 N BEECHWOOD ST	M132	T	4	$ 149,900.00	1	0	1	2
2366006200	6215 CLEARVIEW ST	M133	R	3	$ 70,000.00	0	0	0	0
5136000600	617 LOCUST AVE	M131	T	5	$ 104,000.00	0	0	1	0
2532000100	151 E COULTER ST	M132	S	3	$ 272,000.00	1	0	0	1
5180001500	1518 W LOUDON ST	M133	S	5	$ 50,000.00	0	0	0	0
8817005900	5986 N 20TH ST	M141	R	1	$ 235,000.00	0	1	0	1
8817007300	7347 N 20TH ST	M141	R	4	$ 179,900.00	0	0	0	0
7308004500	4514 N SMEDLEY ST	M123	T	4	$ 129,900.00	0	0	1	1

'M112T5' generates a score of '2'

Degrees of Generalization

HMR Score Allows you to :

- Recognize parcel groups in the population that are under-represented or not represented in the sales data used to create the Model
- Respecify / Recalibrate your variables & Improve your Models

Scoring can be scaled or weighted based on variable importance

Can be used in both feedback & feedforward pipeline

Bake in Intelligence within Big Data

Create a Group Field of Attributes in Model Specification

- Sales File
- Master File

Identify Representation based on the Group

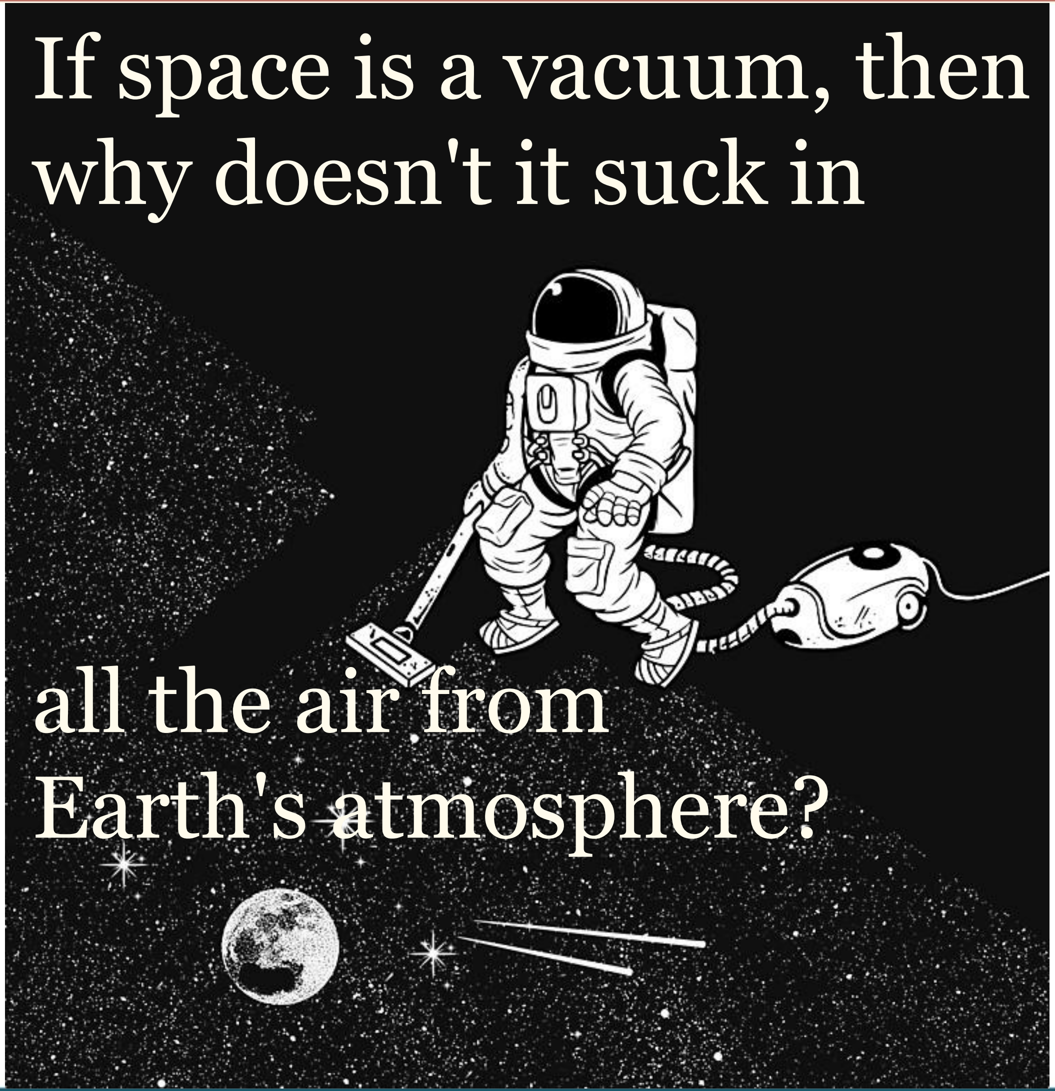
Implement a Scoring for the Group

Measure performance of those Groups

Transplantable Intelligence

- Comparables Engine
- Ratio Studies
- Neighborhood Definition
- Future Model Specifications

Asking the same question?




MY BEAUTIFUL REGRESSION MODEL
BAD SALES VALIDATION
OUTLIERS
UNREPRESENTED PARCELS
@JWTHEBLUEPRINT

Conclusions

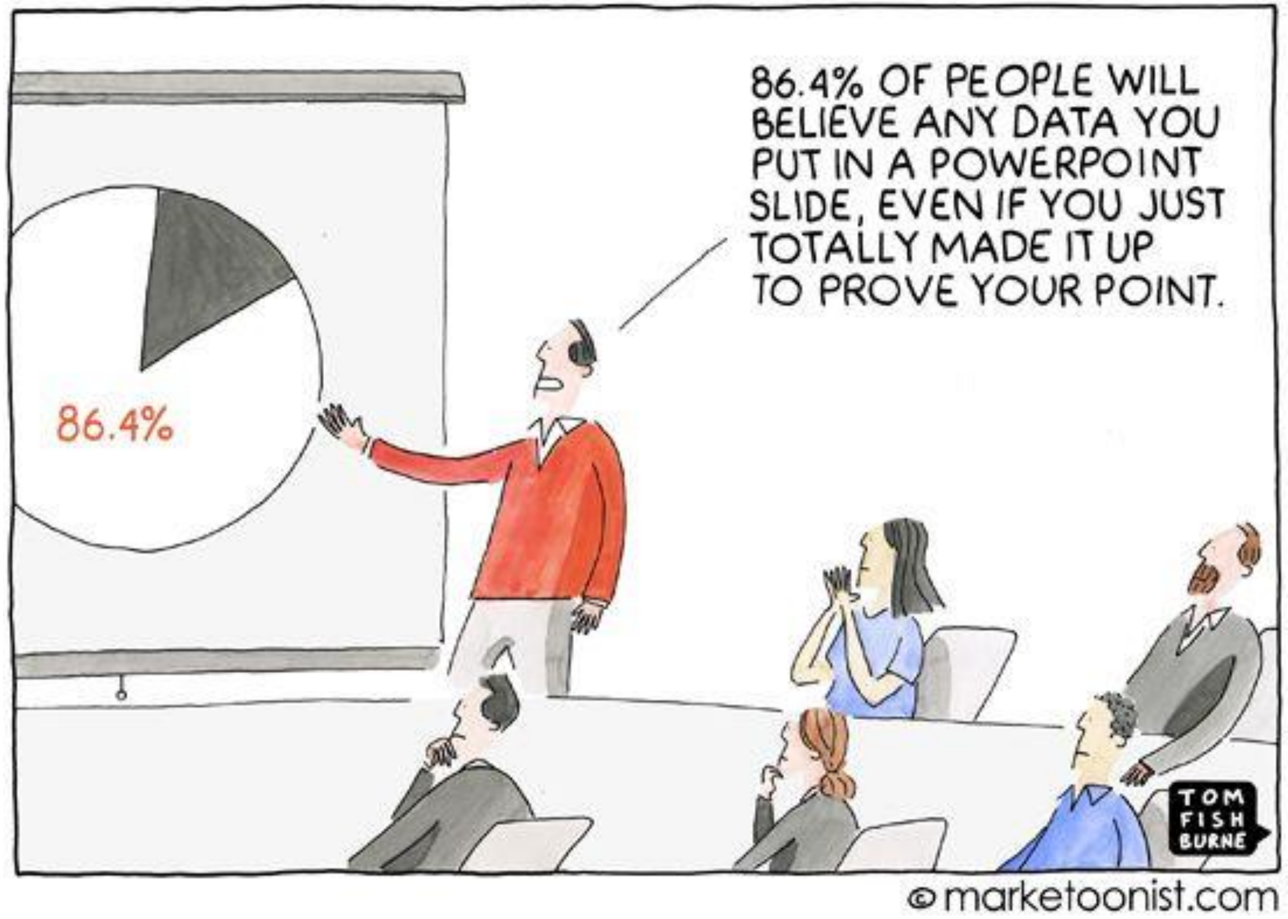
Both methods support analysis by either groups or specific cases

Methods can be combined to show summary data, degrees of generalization and confidence intervals

By creating groups of comparable properties and summarizing data about those groups, we can:

- Gain more precise insight into model and assessment performance
- Facilitate review of valuation projections
- Precisely identify properties that are un-represented or under-represented
- Identify submarkets that are not well served by your valuation process
- Better understand representation through maps and visualizations

Thank You!



Questions ?

Comments?

James.R.Williams@phila.gov
Kevin@keenemac.com
Alex.Raju@phila.gov