IAAO

URISA

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

Comparable Sales in the Age of Artificial Intelligence

Langyue Wang (Larry) and Joseph Wehrli (Joe)

Tyler Technologies, Inc.
Toronto, Ontario and Cincinnati, Ohio

{Larry.Wang, Joe.Wehrli}@tylertech.com

IAAO GIS/VALUATION TECHNOLOGIES CONFERENCE URISA

Motivation

- Explore the possibility to bring AI and machine learning techniques in the comparable sales approach
  - Simplify the procedures, reduce subjectiveness, improve efficiency and accuracy, and achieve high automation

- Empower people, not replace people

- Simplify usage of CAMA systems and broaden accessibility of AVMs to everyday users

Motivation

- Develop and propose practical feature importance based approaches for comparable sales selection

- Validate and test the proposed method through experiments

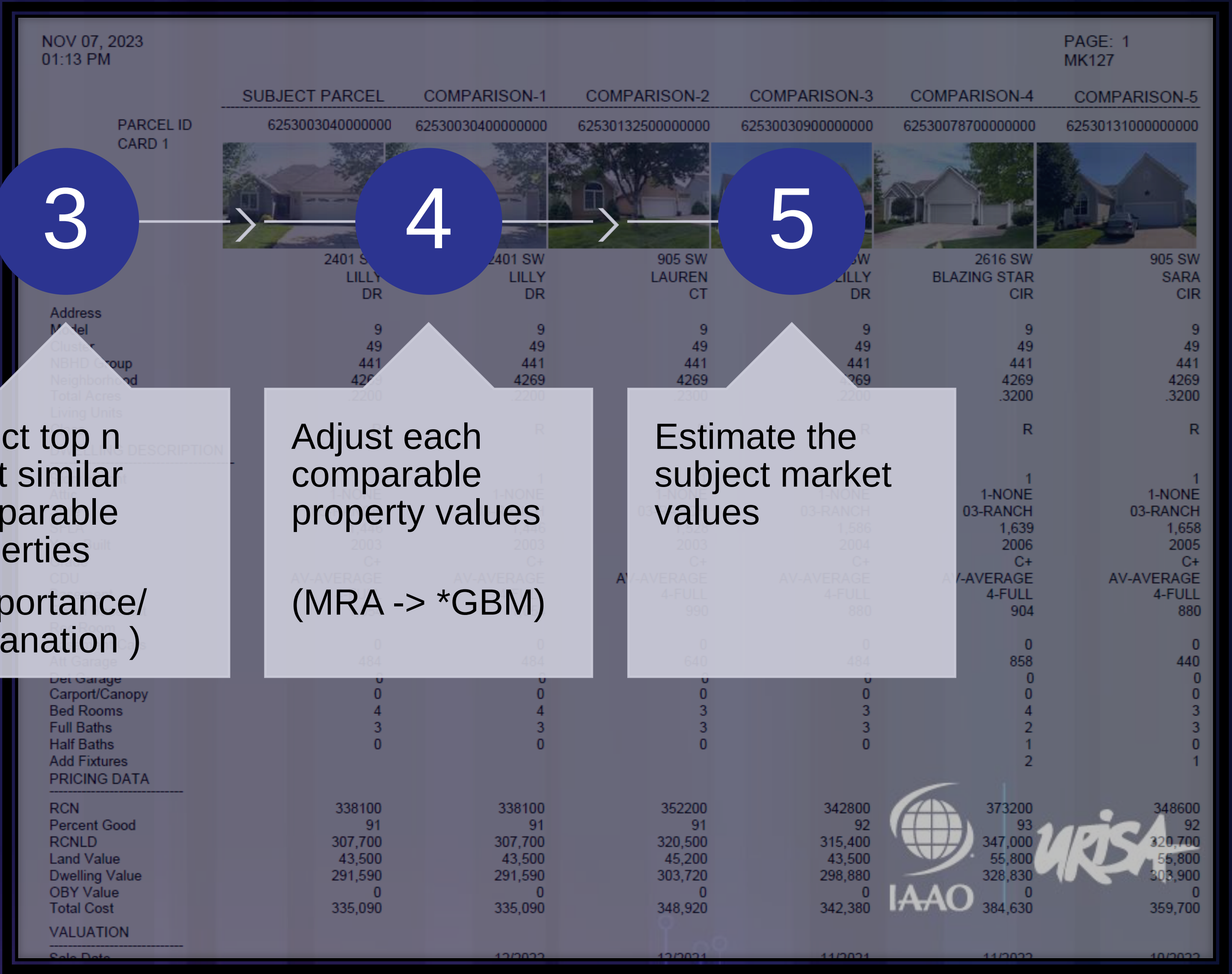
- Foster excitement around AI-enabled evolution of Sales Comparison Appraisal, inspire other variations of innovation

Agenda

- Comparable Appraisal procedure and limitations
- Gradient Boosting and Feature Importance
- Other two popular feature importances: Permutation Feature Importance and SHAP importance
- Proposed feature importance based comparable appraisal method
- Experiments, results and comparison analysis
- Concluding remarks

Comparable Appraisal

1. Select a pool of candidate properties from a database of recent sales
2. Rank the candidates based on similarity distance measures
3. Select top n most similar comparable properties (*Importance/ Explanation)
4. Adjust each comparable property values (MRA -> *GBM)
5. Estimate the subject market values



NOV 07, 2023
01:13 PM

PAGE: 1
MK127

	SUBJECT PARCEL	COMPARISON-1	COMPARISON-2	COMPARISON-3	COMPARISON-4	COMPARISON-5
PARCEL ID CARD 1	6253003040000000	6253003040000000	6253013250000000	6253003090000000	6253007870000000	6253013100000000
Address	2401 S[illegible] LILLY DR	[illegible]401 SW LILLY DR	905 SW LAUREN CT	[illegible]W [illegible]ILLY DR	2616 SW BLAZING STAR CIR	905 SW SARA CIR
Carport/Canopy	0	0	0	0	0	0
Bed Rooms	4	4	3	3	4	3
Full Baths	3	3	3	3	2	3
Half Baths	0	0	0	0	1	0
Add Fixtures					2	1
PRICING DATA						
RCN	338100	338100	352200	342800	373200	348600
Percent Good	91	91	91	92	93	92
RCNLD	307,700	307,700	320,500	315,400	347,000	320,700
Land Value	43,500	43,500	45,200	43,500	55,800	55,800
Dwelling Value	291,590	291,590	303,720	298,880	328,830	303,900
OBY Value	0	0	0	0	0	0
Total Cost	335,090	335,090	348,920	342,380	384,630	359,700
VALUATION						

Weights used in Traditional Comparable Approach

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

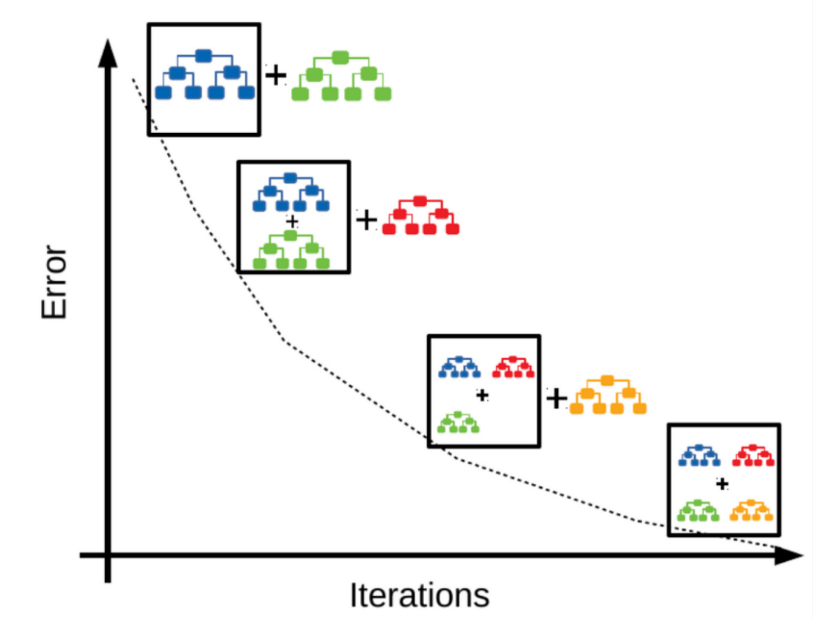
- MRA coefficients/ z-scores
- Empirical weights (manually tweaked subjective values)
- Coefficients depends on the scale of the input features
- Empirically picked weights needs a large amount of time, model specific, subjective

Variable Name	Weights-Variable	Weights-Constan	Subj Data	Comp Data	ABS Diff	[W*(diff)]^2
LANDVAL	0.0018		226,000	244000	18000	1050
SFLA	0.075		5,300	5000	300	506
AGEMAX	5		20	20	0	0
FIXTOT	10		22	22	0	0
GFACT	200		1.85	2.5	0.65	16900
SALEMON	5		0	24	24	14400
TOTGAR	0.09			0	0	0
FINBSMTOT	0.07				0	0
XCOORD	0.015		1250	3250	2000	900
YCOORD	0.015		1250	3250	2000	900
STORIES		50	1	1	0	0
NBHD		150	410	400	1	22500
NGROUP		100	1	1	0	0
STYLE		50	1	1	0	0
						57156
		Sum of Squares		57156.01		
		Distance Points		239		

IAAO URISA

Gradient Boosting in Machine Learning

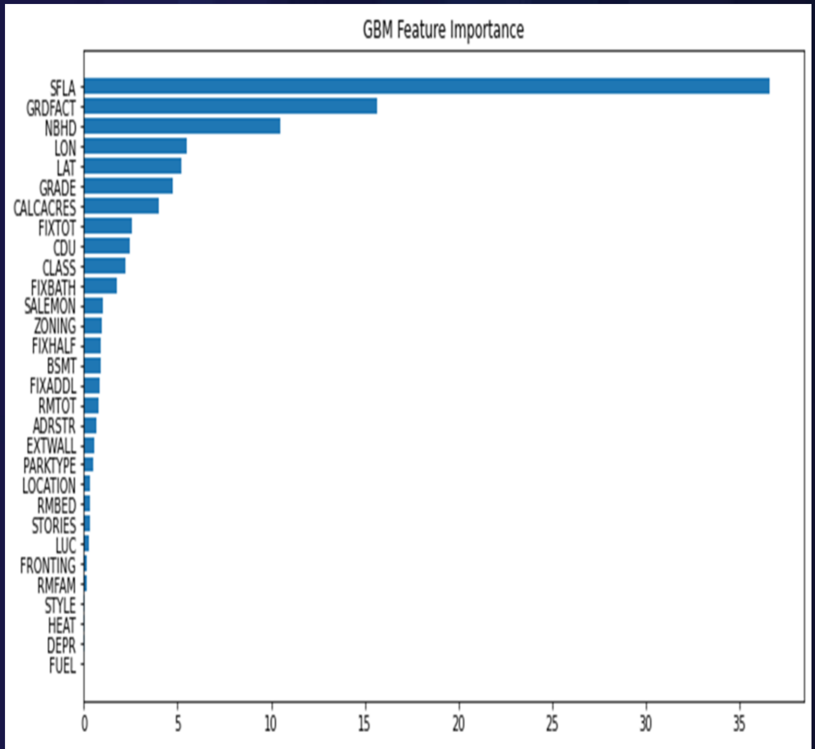
- Gradient Boosting Machines (GBM) is a powerful ensemble technique which combines the predictions of multiple weak learners sequentially to create a single more accurate strong learner
- The weak learners are usually tree based models
- GBMs are among the current state-of-the-art ML techniques on tabular data in a variety of tasks such as prediction and regression.
- Can handle both numerical and categorical data, which eliminates the need for data conversion or transformation



(Reference: https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764 /)

Feature Importance in Gradient Boosting Models

- GBMs provide a score, called feature importance, that indicates how useful or valuable each feature was in the construction of the boosted decision trees
- This importance is calculated explicitly for each attribute in the dataset, allowing attributes to be ranked and compared to each other.
- The more an attribute is used to make key decisions with decision trees, the higher its relative importance.



Popular Gradient Boosting Tools

- Besides scikit-learn implementations, the three most famous boosting algorithm implementations that have provided various recipes for winning ML competitions are:
  1. CatBoost
  2. XGBoost
  3. LightGBM

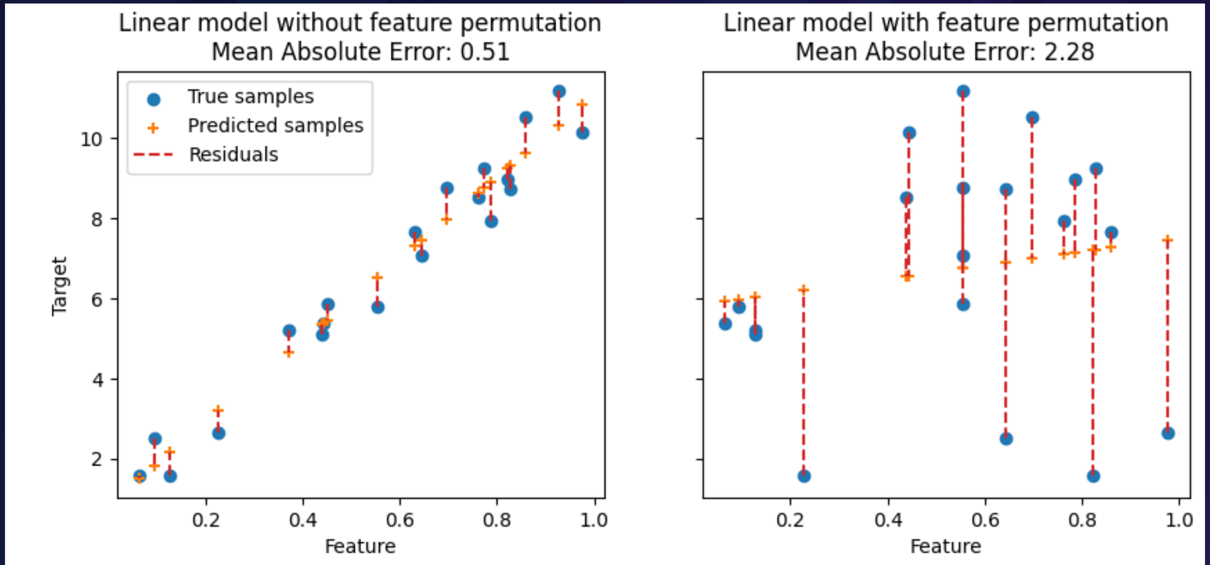
- CatBoost (coined from "Category" and "Boosting") was chosen for this research and experiments, because it
  - Best supports Categorical and Text data
  - Fastest prediction time and best performance

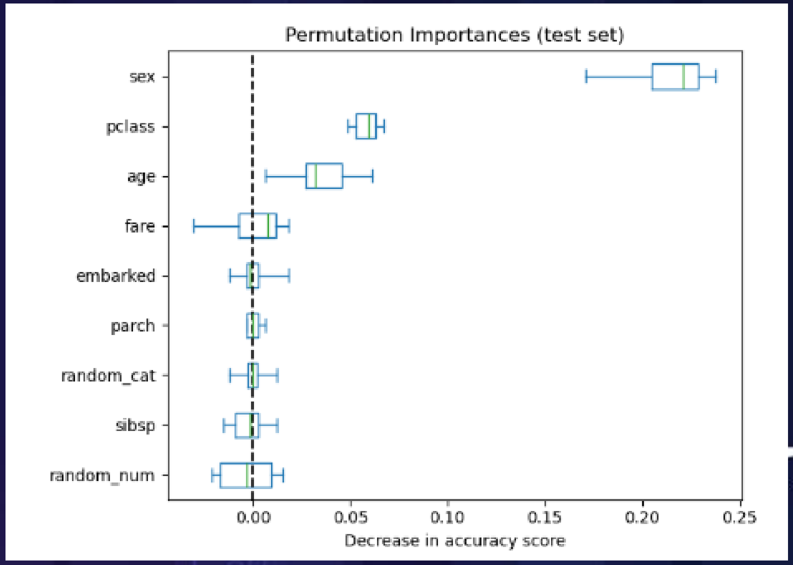
  (based on some benchmark comparison research)

Permutation Feature Importance

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

- Permutation feature importance measures the degradation of the model’s score after randomly shuffling the values of a single feature
- A feature is “important” if shuffling its values increases the model error, because the model relied on the feature for the prediction
- Permutation feature importance is **model-agnostic**

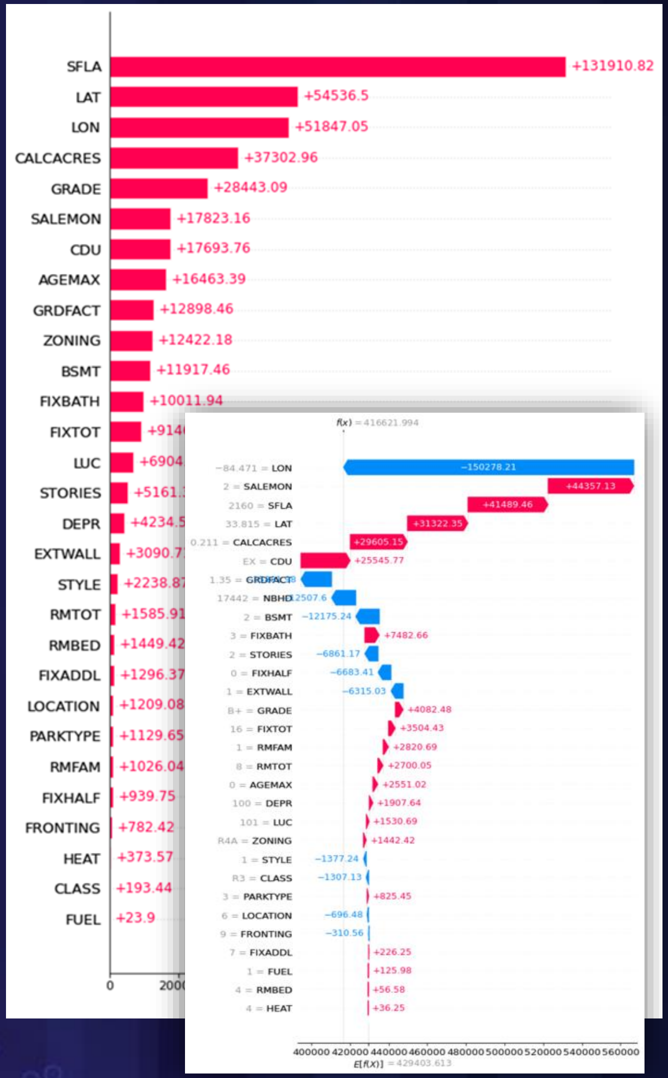




(Ref: https://scikit-learn.org/stable/modules/permutation_importance.html)

SHAP Importance

- SHAP (SHapley Additive exPlanations), based on cooperative game theory, calculates a value (shapely value) that represents the contribution of each feature (as a player) to the model (as a team) outcome
- Shapely value is the average marginal contribution of a feature across all the possible combinations of features.
- The value has both direction and magnitude, SHAP importance is represented in absolute value form
- Model-agnostic and consistent, also good for global and local model explanation



Feature Importances based Comps Approach -1

- Similarity Measure

Suppose there are N candidate properties and K attributes/features used for comps selection, the Euclidean distance between the i^{th} candidate property and the subject property:

$$D_i = \sqrt{\sum_{j=1}^{K} \frac{W_j}{\sum W_j} \begin{Bmatrix} \left(\frac{X_{ij}-X_{sj}}{s_j}\right)^2 & X_{ij}\ is\ numerical \\ \begin{matrix} 1 & X_{ij} == X_{sj} \\ 0 & X_{ij} <> X_{sj} \end{matrix} & X_{ij}\ is\ categorical \end{Bmatrix}}$$

i=1,2.3 …. N

j=1,2.3 …. K

D_i : Weighted Standardized Euclidean Distance

W_j : j^{th} attribute feature importance weight

X_{ij} : the value of the j^{th} attribute of the i^{th} property

X_{sj} : the value of the j^{th} attribute of the subject property

s_j : standard deviation of j^{th} attribute

- Feature importance values of each attribute are used as the weights

Feature Importances based Comps Approach -2

- Estimate subject market value

Based on the previous Similarity Measure, select top 5 candidate properties as comparable sales to estimate a subject market value

$$ESP = GBM_{subj} + \underbrace{w_1\ Resid_1 + w_2\ Resid_2 + \ldots + w_n\ Resid_n}_{\text{Weighted GBM adjustments}} \quad (n=5)$$

ESP : Estimated Subject Price

GBM_{subj} : GBM model prediction for the subject

$$Resid_n = SP_n - GBM_n$$

SP_n : sale price of the nth comparable property

GBM_n : GBM predicted price for the nth comparable property

w_n : inverse distance weight

$$\sum_{n=1}^{5} w_n = 1$$

[Reference: Improving Mass Appraisal Valuation Models Using Spatio-Temporal Methods by Richard A. Borst, PhD]

Experiments – Property Sales Data

- Area:

  Fulton County, Georgia

- Sales data:

  31,125 residential single-family sales

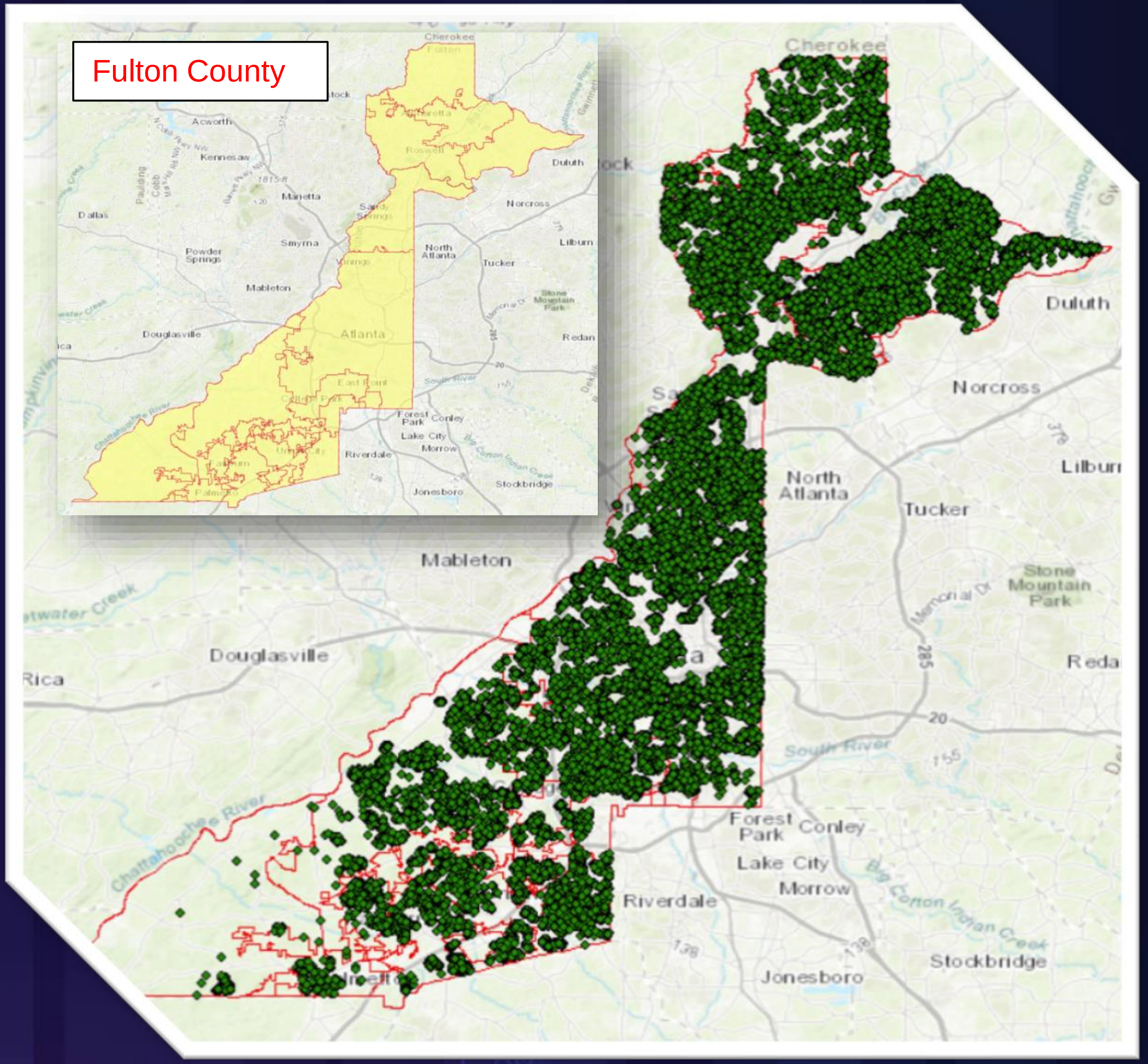
  from Jan 1,2017 to Dec 31,2019

- 17 numerical variables/features

  CALCACRES, FRONTING, STORIES, YRBLT, EXTWALL, RMTOT, RMBED, RMFAM, FIXBATH, FIXHALF, FIXADDL, FIXTOT, BSMT, HEAT, FUEL, SFLA, GRDFACT, DEPR ,LAT, LON, SALEMON

- 13 categorical variables/features

  NBHD, STYLE, ZONING, GRADE, CDU, LOCATION, ADRSTR, BSMT, HEAT, FUEL, FRONTING, EXTWALL, PARKTYPE



GBM Feature Importance Result

- Training CatBoostRegressor GBM model with randomly split the sales dataset to 80% for training and 20% for test
- Use Optuna for tuning the key hyperparameters
- Best hyperparameters:

{ 'iterations': 1722, 'learning_rate': 0.08528,

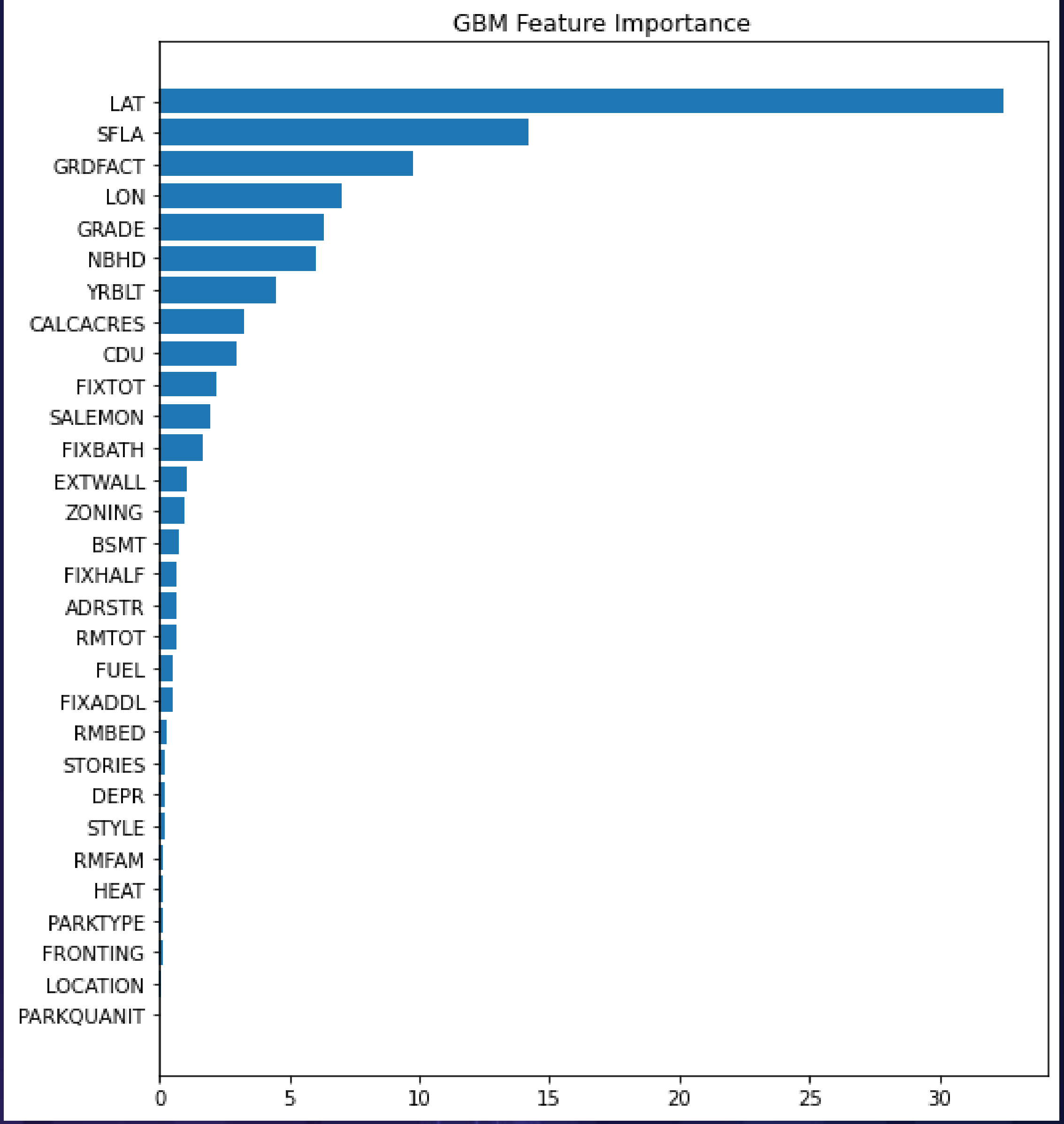
'depth': 7, 'subsample': 0.757490517169143,

'colsample_bylevel': 0.8673982334421928,

'min_data_in_leaf': 32, 'l2_leaf_reg': 1.5 }

RMSE - CatBoost (training): 65334.09250867287

RMSE - CatBoost (test): 117713.72364887314



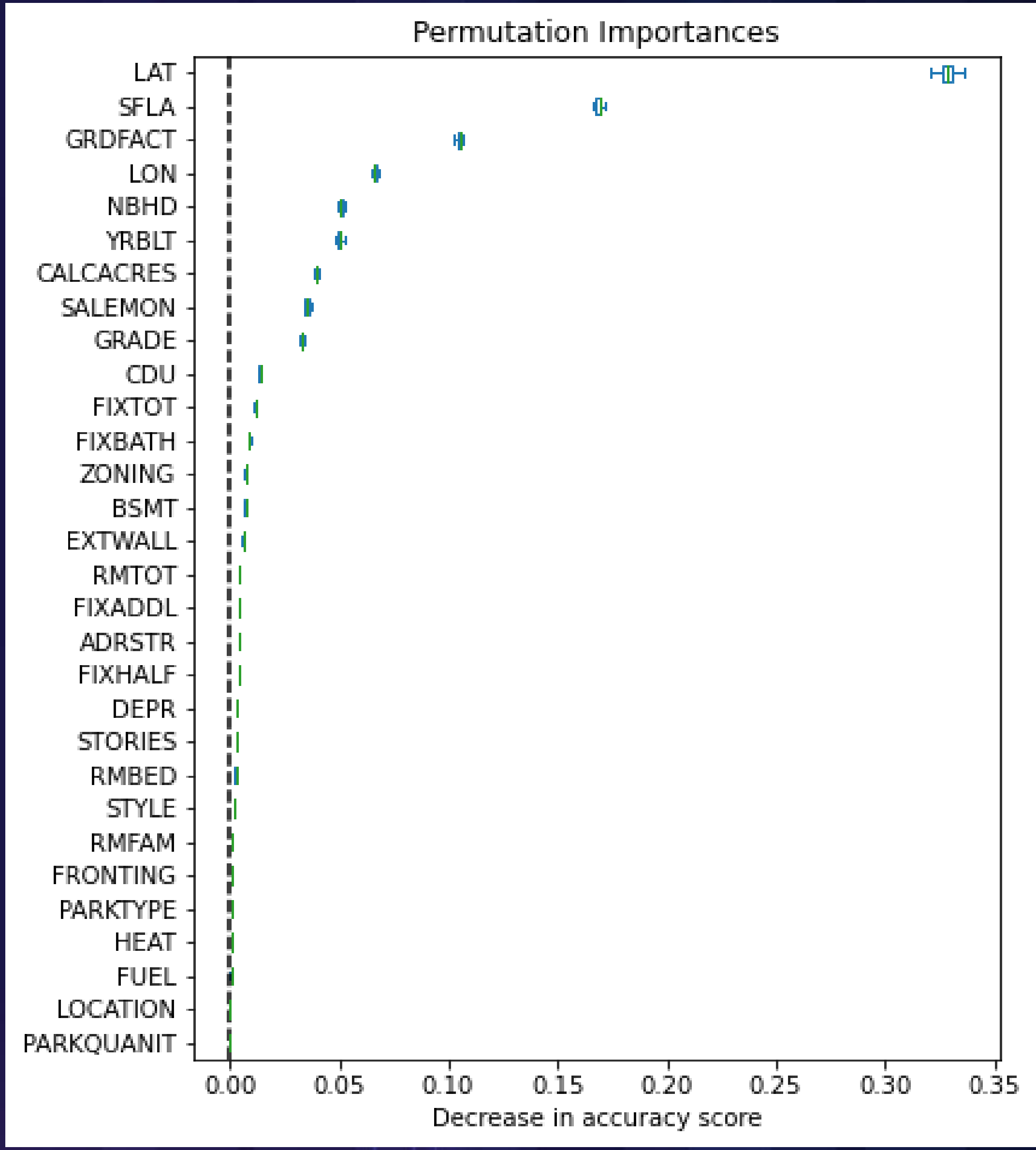
Permutation Feature Importance Result

- Use sklearn.inspection.permutation_importance function
- Use the same previous GBM model as the estimator
- Parameters:

  n_repeats: 10

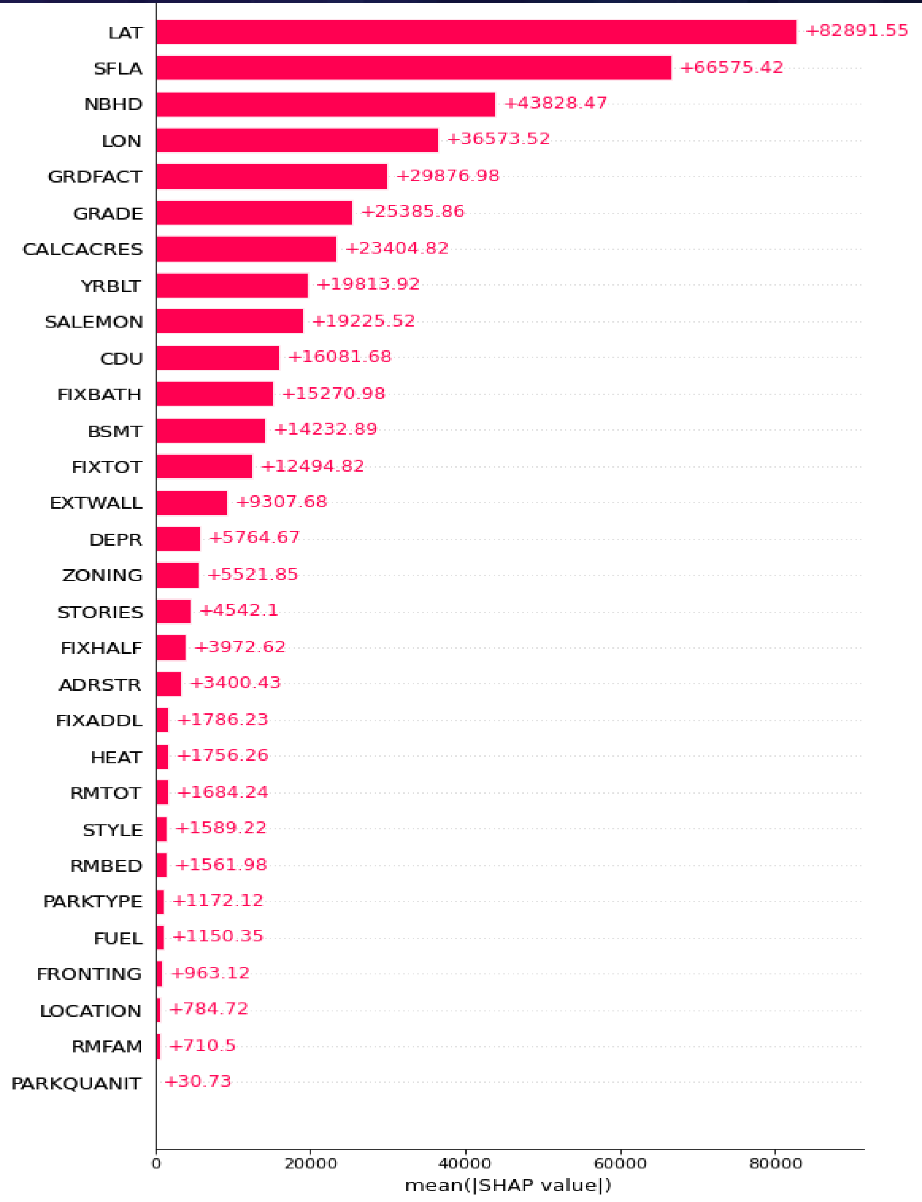
  (number of time to permute a feature)

  Others: default



SHAP Feature Importance Result

- Use Python SHAP, which can be installed from either PyPI or conda-forge
- Use the GBM model in the SHAP explainer



Comparison – Feature Importances

	GBM		Permutation		SHAP	
1	LAT	32.49	LAT	34.29	LAT	18.37
2	SFLA	14.23	SFLA	17.63	SFLA	14.75
3	GRDFACT	9.77	GRDFACT	10.96	**NBHD**	9.71
4	LON	7.02	LON	6.89	LON	8.10
5	**GRADE**	6.33	**NBHD**	5.30	GRDFACT	6.62
6	**NBHD**	6.01	YRBLT	5.22	**GRADE**	5.62
7	YRBLT	4.53	CALCACRES	4.09	CALCACRES	5.19
8	CALCACRES	3.25	SALEMON	3.66	YRBLT	4.39
9	**CDU**	2.99	**GRADE**	3.39	SALEMON	4.26
10	FIXTOT	2.23	**CDU**	1.42	**CDU**	3.56
11	SALEMON	2.01	FIXTOT	1.18	FIXBATH	3.38
12	FIXBATH	1.68	FIXBATH	0.92	**BSMT**	3.15
13	**EXTWALL**	1.08	**ZONING**	0.74	FIXTOT	2.77
14	**ZONING**	0.99	**BSMT**	0.71	**EXTWALL**	2.06
15	**BSMT**	0.77	**EXTWALL**	0.59	DEPR	1.28

Importance values are all converted to percentage values

Categorical features are in BOLD fonts

Comparison – Feature Importances (Cont'd)

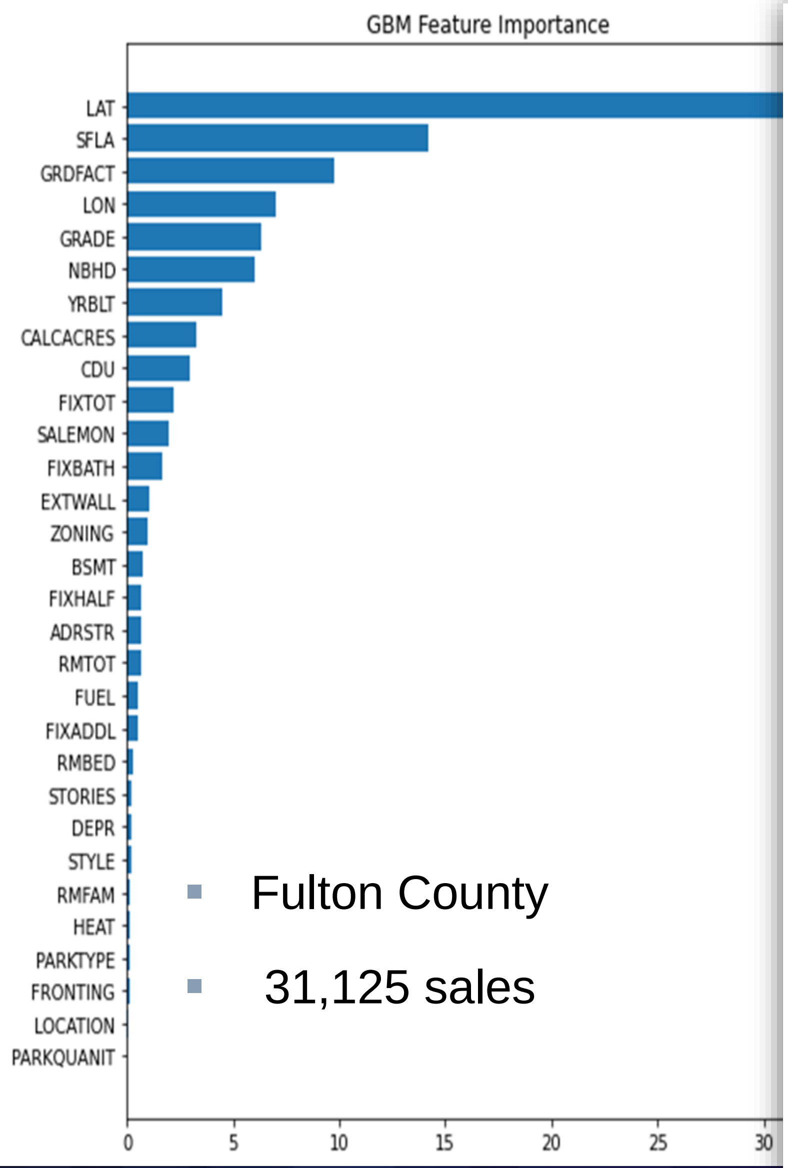
	GBM		Permutation		SHAP	
16	FIXHALF	0.68	RMTOT	0.44	**ZONING**	1.22
17	**ADRSTR**	0.66	FIXADDL	0.42	STORIES	1.00
18	RMTOT	0.64	**ADRSTR**	0.41	FIXHALF	0.88
19	**FUEL**	0.53	FIXHALF	0.40	**ADRSTR**	0.75
20	FIXADDL	0.49	DEPR	0.30	FIXADDL	0.40
21	RMBED	0.31	STORIES	0.28	**HEAT**	0.39
22	STORIES	0.25	RMBED	0.24	RMTOT	0.37
23	DEPR	0.23	**STYLE**	0.20	**STYLE**	0.35
24	**STYLE**	0.21	RMFAM	0.09	RMBED	0.35
25	RMFAM	0.15	**FRONTING**	0.08	**PARKTYPE**	0.26
26	**HEAT**	0.12	**PARKTYPE**	0.06	**FUEL**	0.25
27	**PARKTYPE**	0.12	**HEAT**	0.06	**FRONTING**	0.21
28	**FRONTING**	0.12	**FUEL**	0.01	**LOCATION**	0.17
29	**LOCATION**	0.08	**LOCATION**	0.007	RMFAM	0.16
30	PARKQUANIT	0.002	PARKQUANIT	0.0007	PARKQUANIT	0.007

Importance values are all converted to percentage values

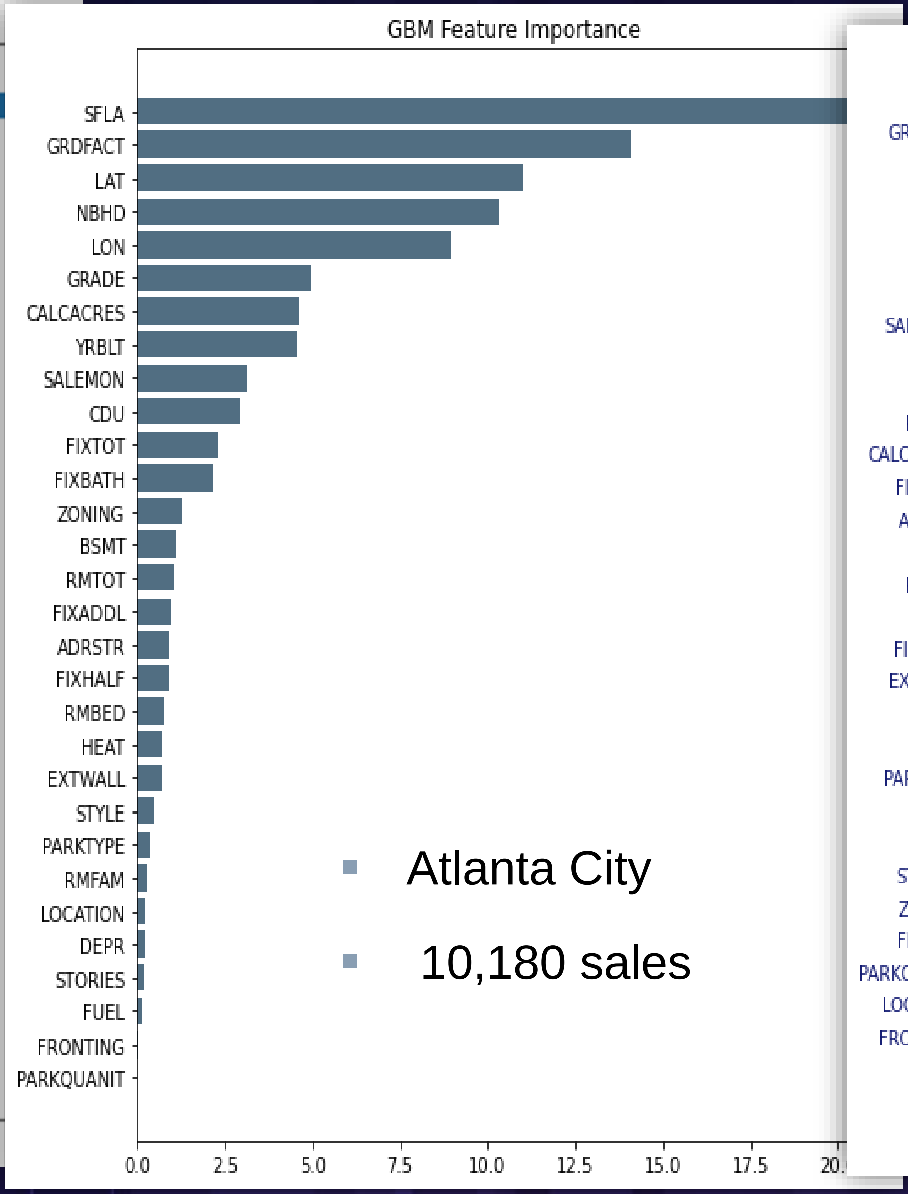
Categorical features are in BOLD fonts

GBM Feature Importance at Various Scales

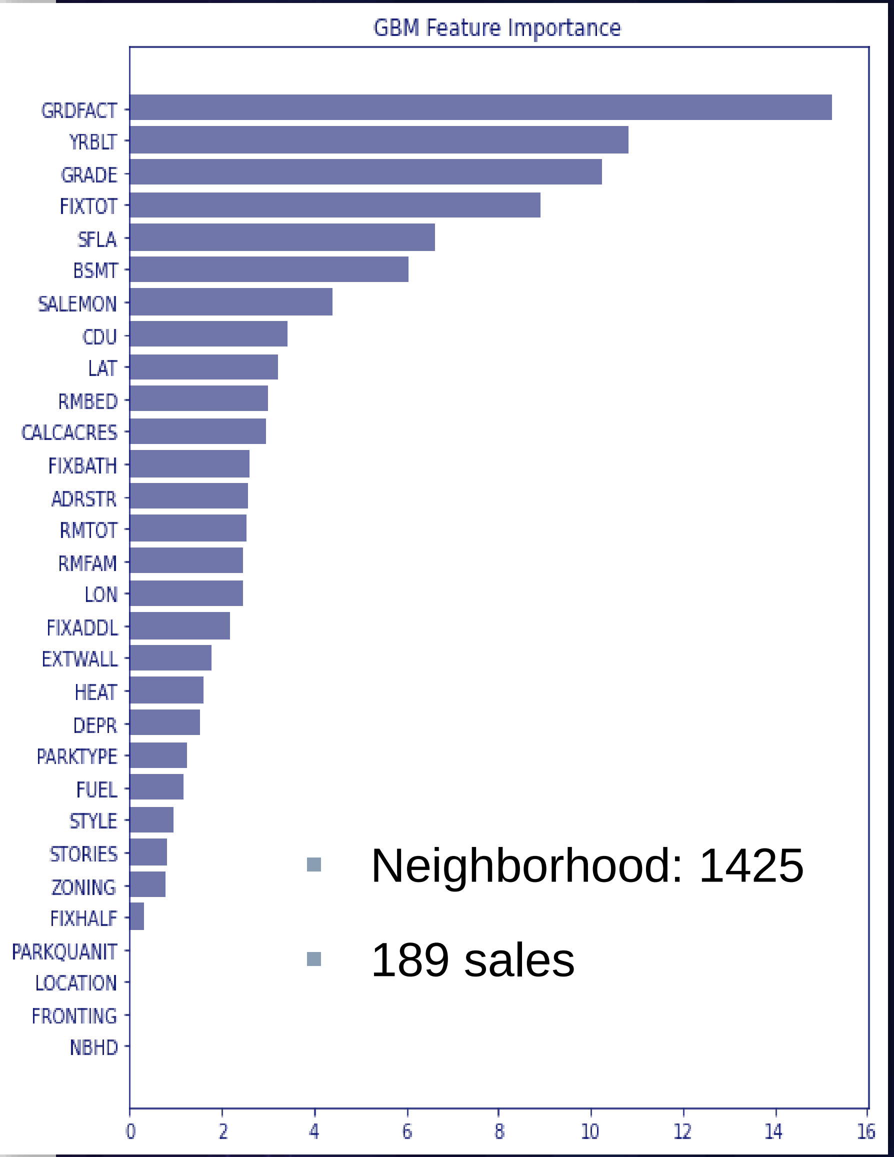
GIS/VALUATION TECHNOLOGIES CONFERENCE 2024



- Fulton County
- 31,125 sales



- Atlanta City
- 10,180 sales



- Neighborhood: 1425
- 189 sales

Experiments – Subject Data and Valuation

- Valuation Date:

  Jan 1, 2020

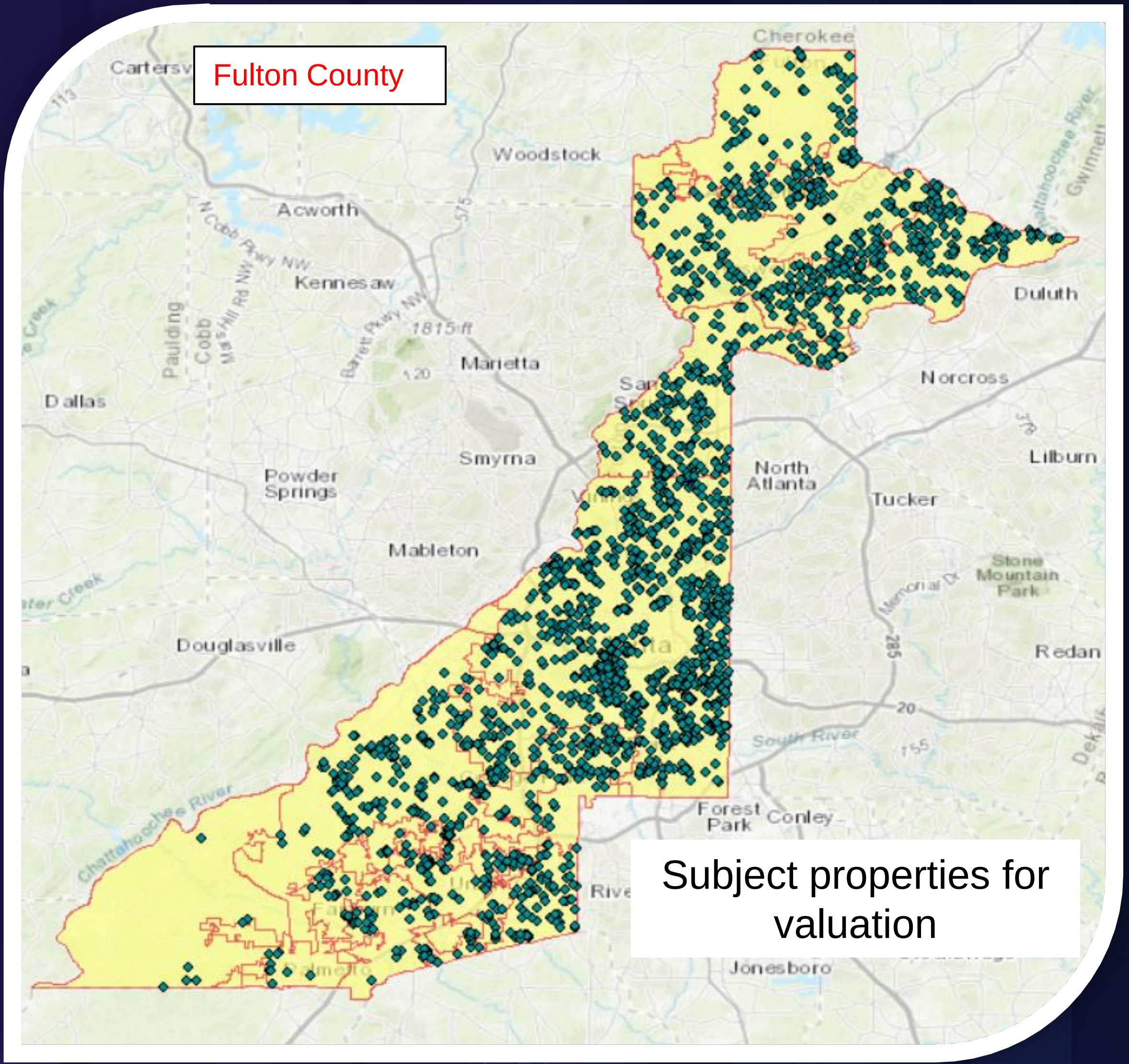
- Issue:

  No sales really sold on Jan 1, 2020, therefore no true sales prices for calculating valuation accuracy

- Solution:

  Randomly selected 2,272 residential single-family sales in Dec 2019, Jan 2020 and Feb 2020, use their sale prices as the true prices (approximately)

- Set SALEMON of all subjects to 0



Market Price Estimation Comparison - Method

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

Subject Dataset (N)
(Dec 1, 2019 - Feb 29,2020)

Sales Data
(2017-2019)

Existing CAMA Database

Repeat N

Choose one as Subject S

Calculate similarity distances between S and each sales data

Select 5 most similar properties as comparable sales (comps)

Estimate the market price of S using mean of the sale prices of the 5 comps

Estimate the market price of S using the weighted GBM adjustment method

Calculate residuals between the estimated price and the Subject real sale price

Calculate Error Metrics and Ratio Study

Pull the 5 Comps of each of the Subject S and the estimated market prices

Calculate Error Metrics and Ratio Study

Compare

Market Price Estimation – Mean Comps Price

- Area: Fulton county
- No. of Subjects: 2,273

- **Error Metrics**

Comps from	AVG Price	Median Price	R2	RMSE	MAE	MAPE	RRSE	RAE	COV
GBM	445,701.91	330,000	0.861	142,470.60	65,518.42	15.19	0.370	0.255	32.23
Permutation			0.863	142,665.52	65,069.70	15.07	0.371	0.254	32.27
SHAP			0.861	142,223.52	66,708.57	15.46	0.369	0.26	32.17
CAMA			0.866	139,708.04	65,707.71	15.09	0.362	0.256	31.60

- **Sales Ratio Study**

Comps from	AVG Price	Median Price	Median Sales Ratio	Mean Sales Ratio	COD	PRD
GBM	445,701.91	330,000	0.984	1.001	15.353	1.032
Permutation			0.99	1.006	15.193	1.031
SHAP			0.986	1.003	15.62	1.034
CAMA			0.973	0.989	15.256	1.049

Market Price Estimation - Weighted GBM Adjustment

- Area: Fulton County
- No. of Subjects: 2,273

- **Error Metrics**

Comps from	AVG Price	Median Price	R2	RMSE	MAE	MAPE	RRSE	RAE	COV
GBM	445,824.14	330,000	0.91	113,334.68	42,104.74	10.28	0.29	0.16	25.62
Permutation			0.91	113,092.36	41,734.66	10.30	0.29	0.16	25.57
SHAP			0.91	112,798.40	41,909.23	10.385	0.29	0.16	25.51
CAMA			0.85	148,156.06	66,467.73	15.36	0.38	0.26	33.49

- **Sales Ratio Study**

Comps from	AVG Price	Median Price	Median Sales Ratio	Mean Sales Ratio	COD	PRD
GBM	445,824.14	330,000	0.998	0.977	10.288	1.005
Permutation			0.998	0.978	10.315	1.004
SHAP			0.997	0.977	10.398	1.005
CAMA			0.953	0.945	15.179	1.034

Market Price Estimation – Comparison – Various Area

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

- Error Metrics (Method: Weighted GBM Adjustment)

Area	No. of Subjects / No. of Sales	AVG Price	Median Price	Comps Selected using	R2	RMSE	MAE	MAPE	RRSE	RAE	COV
Fulton County	2273 / 31,125	445,824.14	330,000	GBM	0.91	113,334.68	42,104.74	10.28	0.29	0.16	25.62
				Permutation	0.91	113,092.36	41,734.66	10.30	0.29	0.16	25.57
				SHAP	0.91	112,798.40	41,909.23	10.385	0.29	0.16	25.51
				CAMA	0.85	148,156.06	66,467.73	15.36	0.38	0.26	33.49
Atlanta City	900 / 10,180	495,426.47	333,000	GBM	0.93	12,4551.47	54,279.85	14.81	0.26	0.16	25.64
				Permutation	0.926	12,8996.90	55,045.23	14.69	0.27	0.17	26.56
				SHAP	0.925	129,198.51	55,427.47	14.96	0.27	0.17	26.60
				CAMA	0.85	181,253.96	86,241.71	19.82	0.38	0.26	37.32
NBHD 1425	18 / 189	458,598.33	409,250	GBM	0.80	71,754.14	45,847.49	9.76	0.44	0.35	NA
				Permutation	0.797	72,896.52	45,973.62	9.68	0.45	0.35	NA
				SHAP	0.81	71,131.37	44,669.75	9.41	0.44	0.34	NA
				CAMA	0.61	100,572.85	69,475.00	19.73	0.62	0.53	NA

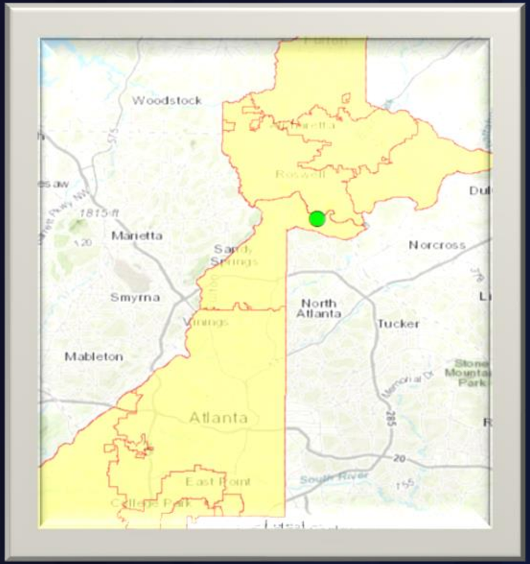
Market Price Estimation – Comparison – Various Area

- Ratio Study (Method: Weighted GBM Adjustment)

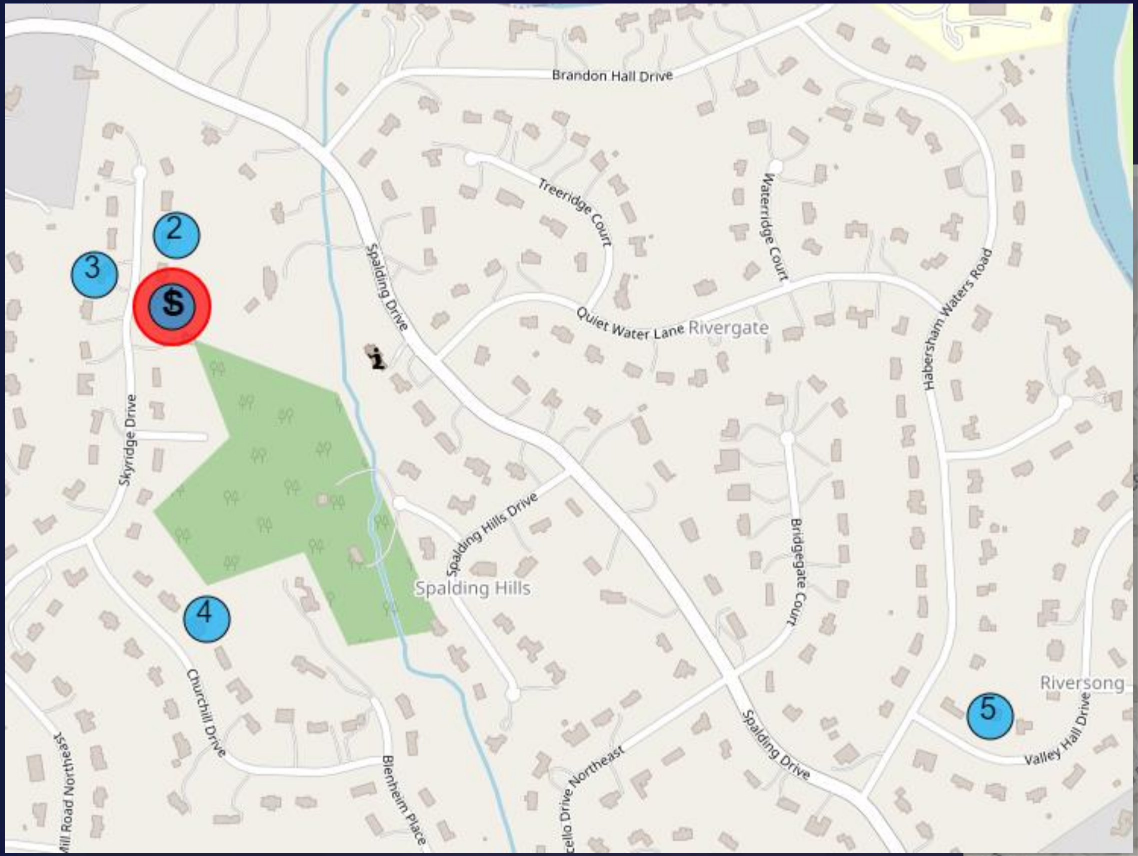
NBHD	No. of Subjects / No. of Sales	AVG Price	Median Price	Comps Selected using	Median Sales Ratio	Mean Sales Ratio	COD	PRD
Fulton County	2,273 / 31,125	445,824.14	330,000	GBM	0.998	0.977	10.288	1.005
				Permutation	0.998	0.978	10.315	1.004
				SHAP	0.997	0.977	10.398	1.005
				CAMA	0.953	0.945	15.179	1.034
Atlanta City	900 / 10180	495,426.48	330,000	GBM	0.996	0.961	14.838	0.991
				Permutation	0.996	0.961	14.725	0.991
				SHAP	0.996	0.963	14.993	0.991
				CAMA	0.937	0.941	20.074	1.049
NBHD 1425	18 / 189	458,598.33	409,250	GBM	0.998	0.978	9.763	1.013
				Permutation	0.998	0.98	9.664	1.014
				SHAP	0.996	0.975	9.427	1.012
				CAMA	0.924	0.995	18.355	1.05

Individual Subject and Comps Comparison - 1

Subject Location

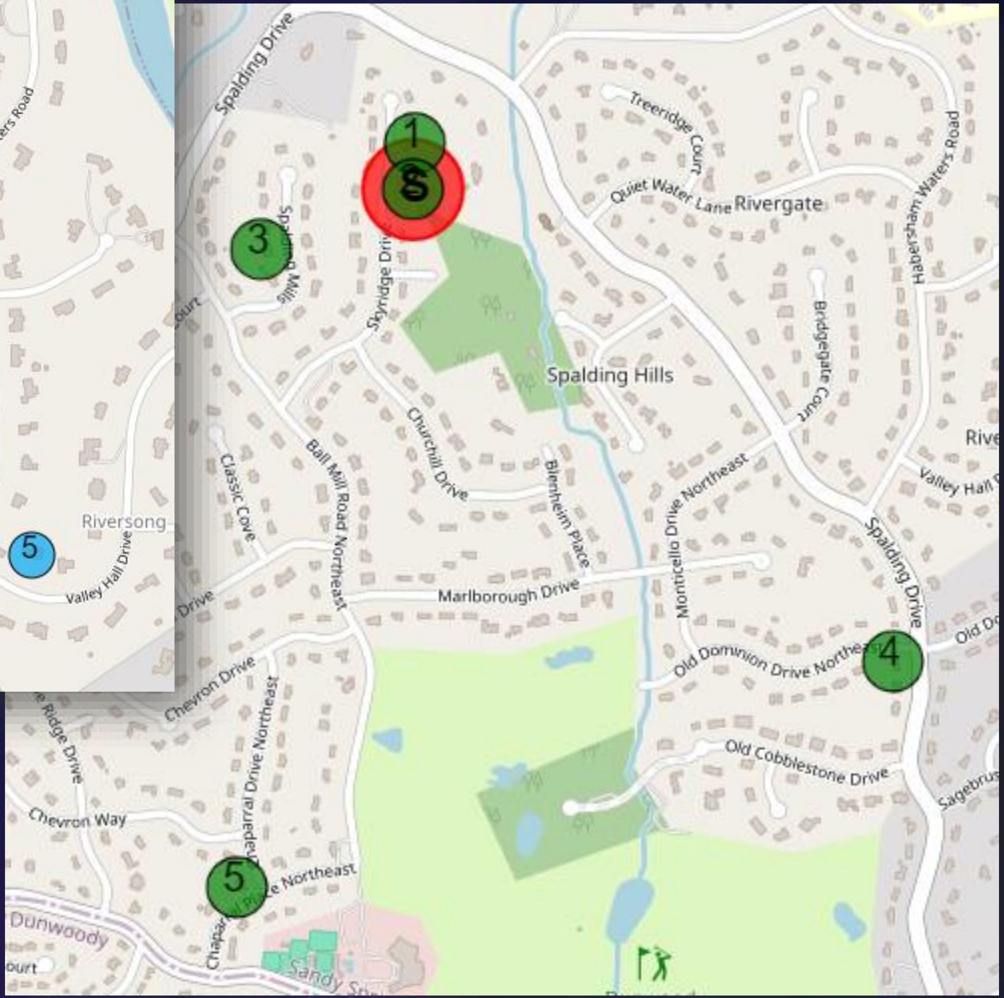


Subject sold on
2/17/2020
Price: **$503,575**



Comps selected using GBM feature importance and estimation calculated using weighted adjustment method

Comps selected in CAMA



Our Estimated Price: $513,244.99   Residual: -9,669.99

CAMA Estimated Price: $541,590   Residual: -38,015.00

Individual Subject and Comps comparison - 1

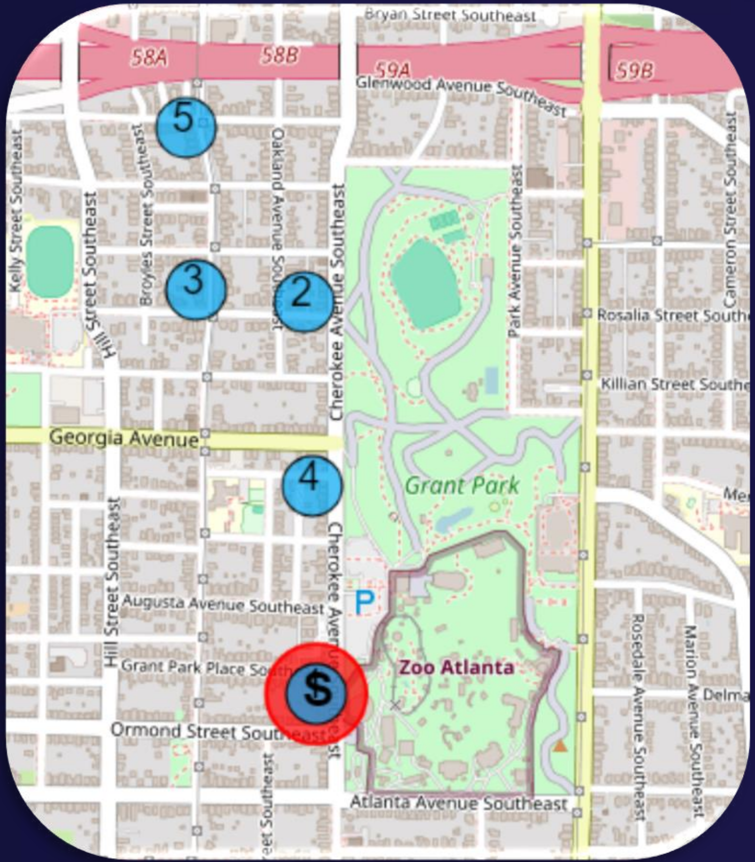
	SUBJECT	COMP1	COMP2	COMP3	COMP4	COMP5	CAMA_COMP1	CAMA_COMP2	CAMA_COMP3	CAMA_COMP4	CAMA_COMP5
PARID	06 035800010152	06 035800010152	06 035800010137	06 035800010061	06 0352 LL0572	06 035100020281	06 035800010137	06 035800010152	06 035700010245	06 0342 LL0237	06 035200020223
PRICE	503575	509000	437000	615000	700000	585000	437000	509000	660000	542500	537500
SALEDT	2/17/2020	2/11/2019	5/21/2019	5/17/2019	6/7/2019	3/27/2019	5/21/2019	2/11/2019	3/29/2019	6/24/2019	9/13/2019
ADRSTR	SKYRIDGE	SKYRIDGE	SKYRIDGE	SKYRIDGE	CHURCHILL	VALLEY HALL	SKYRIDGE	SKYRIDGE	SPALDING MILL	SPALDING	CHAPARRAL
NBHD	604	604	604	604	607	606	604	604	604	604	604
STYLE	1	1	1	1	1	1	1	1	1	8	1
ZONING	R2	R2	R2	R2	R2C	R2	R2	R2	R2	AG1	R2
GRADE	A	A	A	A+	A	A	A	A	B+	B+	B
CDU	VG	VG	VG	EX	EX	EX	VG	VG	VG	VG	VG
LOCATION	6	6	6	6	6	6	6	6	6	6	6
BSMT	4	4	4	4	4	2	4	4	4	4	4
HEAT	4	4	4	4	4	4	4	4	4	4	4
FUEL	1	1	1	1	1	1	1	1	1	1	1
FRONTING	9	9	9	9	9	9	9	9	9	9	9
EXTWALL	7	7	1	7	7	7	1	7	7	9	7
PARKTYPE	3	3	3	3	3	3	3	3	3	1	3
SFLA	3351	3351	3256	3424	3972	3764	3256	3351	3884	3181	3134
GRDFACT	1.55	1.55	1.55	1.67	1.55	1.55	1.55	1.55	1.35	1.35	1.26
CALCACRES	1.0645	1.0645	1.0021	0.9952	1.0207	0.9734	1.0021	1.0645	1.1548	1.0365	1.146
STORIES	2	2	2	2	2	1.5	2	2	2	2	2
YRBLT	1979	1979	1979	1975	1977	1976	1979	1979	1976	1973	1972
RMTOT	9	9	14	9	9	8	14	9	9	8	8
RMBED	4	4	5	4	4	4	5	4	4	3	4
RMFAM	0	0	0	0	0	1	0	0	0	0	0
FIXBATH	3	3	4	3	4	3	4	3	4	3	2
FIXHALF	0	0	0	1	0	1	0	0	1	1	1
FIXADDL	3	3	3	4	2	2	3	3	2	3	2
FIXTOT	12	12	15	15	14	13	15	12	16	14	10
DEPR	100	100	100	100	100	100	100	100	100	98	98
LAT	33.980051	33.980051	33.980886	33.980432	33.9764	33.97526	33.980886	33.980051	33.978989	33.971656	33.967645
LON	-84.31223	-84.31223	-84.312163	-84.313327	-84.311736	-84.300664	-84.312163	-84.31223	-84.315474	-84.301904	-84.315997
SALEMON	0	10	7	7	6	9	7	10	9	6	3
PARKQUANIT	2	2	2	2	2	2	2	2	2	2	2

Individual Subject and Comps Comparison - 2

GIS/VALUATION TECHNOLOGIES CONFERENCE 2024

Subject Location

Subject sold on
1/27/2020
Price: $713,000



Comps selected using SHAP feature importance

Comps selected in CAMA

Our Estimated Price: $760,164.47 Residual: - $47,164.46

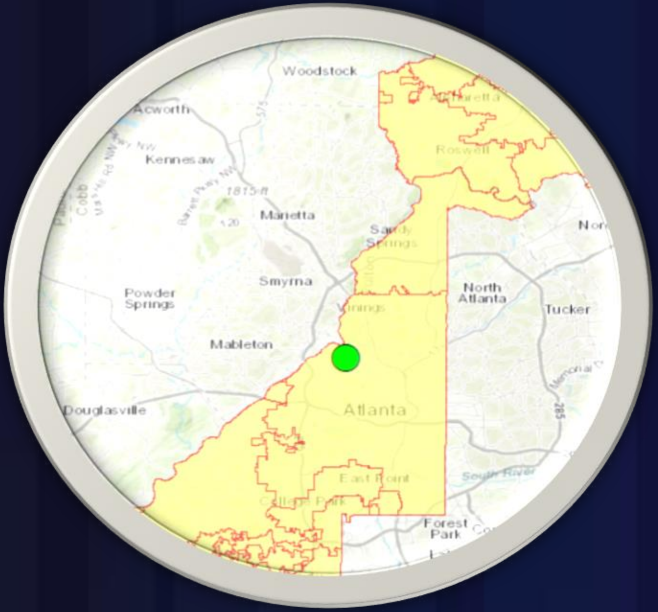
CAMA Estimated Price: $649,870.00 Residual: $63,130.00

Individual Subject and Comps Comparison - 2

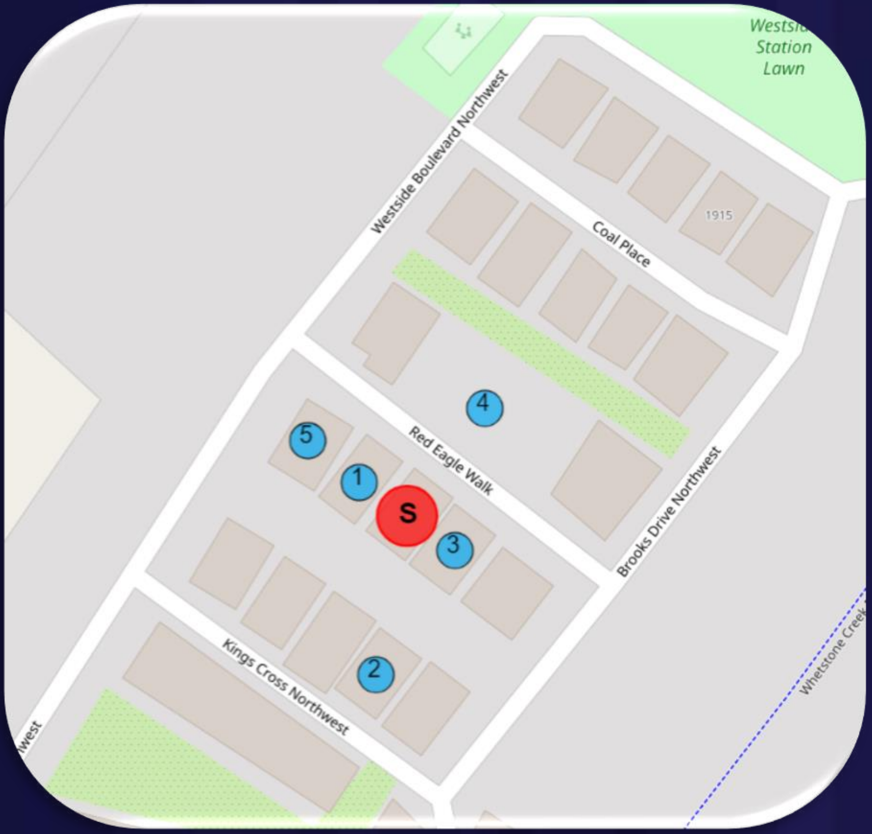
	SUBJECT	COMP1	COMP2	COMP3	COMP4	COMP5	CAMA_COMP1	CAMA_COMP2	CAMA_COMP3	CAMA_COMP4	CAMA_COMP5
PARID	14 004300060160	14 004300060160	14 004400071034	14 004400070697	14 004300020982	14 004400040864	14 004300030155	14 004400070697	14 004300030650	14 004300060277	14 004400040864
PRICE	713000	669000	625000	585000	515000	500000	590000	585000	550000	570000	500000
SALEDT	1/27/2020	1/12/2018	11/15/2019	11/22/2019	7/12/2019	10/19/2018	7/15/2019	11/22/2019	6/6/2019	8/8/2019	10/19/2018
ADRSTR	CHEROKEE	CHEROKEE	MILLEDGE	GRANT	PAVILION	GRANT	BASS	GRANT	GRANT PARK	ORMOND	GRANT
NBHD	14269	14269	14269	14269	14269	14269	14269	14269	14269	14269	14269
STYLE	1	1	1	1	1	1	8	1	8	8	1
ZONING	R5	R5	R5	R5	R5	R5	R5	R5	R5	R5	R5
GRADE	A-	A-	B+	C+	B+	B+	B+	C+	B+	B+	B+
CDU	EX	EX	VG	EX	VG	GD	EX	EX	VG	VG	GD
LOCATION	6	6	6	6	6	6	6	6	6	6	6
BSMT	3	3	2	3	2	3	2	3	2	2	3
HEAT	4	4	4	4	4	4	4	4	4	4	4
FUEL	1	1	1	1	1	1	1	1	1	1	1
FRONTING	9	9	9	9	9	9	9	9	9	9	9
EXTWALL	6	6	1	1	1	1	1	1	1	1	1
PARKTYPE	3	3	3	3	3	3	3	3	2	3	3
SFLA	1768	1768	1479	1794	1900	1747	1828	1794	1786	1894	1747
GRDFACT	1.45	1.45	1.35	1.08	1.35	1.35	1.35	1.08	1.35	1.35	1.35
CALCACRES	0.3122	0.3122	0.1221	0.1722	0.1339	0.1879	0.1531	0.1722	0.1066	0.1825	0.1879
STORIES	1	1	1	1	1	1	2	1	2	2	1
YRBLT	1920	1920	1920	1920	1920	1920	1997	1920	2003	1998	1920
RMTOT	7	7	6	6	5	7	4	6	8	8	7
RMBED	3	3	3	3	3	3	3	3	3	3	3
RMFAM	0	0	0	0	0	0	1	0	1	1	0
FIXBATH	2	2	2	2	2	2	2	2	2	2	2
FIXHALF	1	1	0	0	0	1	1	0	1	1	1
FIXADDL	2	2	2	4	2	3	4	4	5	4	3
FIXTOT	10	10	8	10	8	11	12	10	13	12	11
DEPR	100	100	98	100	90	96	100	100	98	98	96
LAT	33.732122	33.732122	33.739125	33.739325	33.735825	33.742222	33.735001	33.739325	33.732873	33.731723	33.742222
LON	-84.37406	-84.37406	-84.374322	-84.376634	-84.374146	-84.376834	-84.377473	-84.376634	-84.3773	-84.375305	-84.376834
SALEMON	0	23	1	1	5	14	5	1	6	4	14
PARKQUANIT	2	2	2	2	2	2	2	2	2	2	2

Individual Subject and Comps Comparison - 3

Subject Location

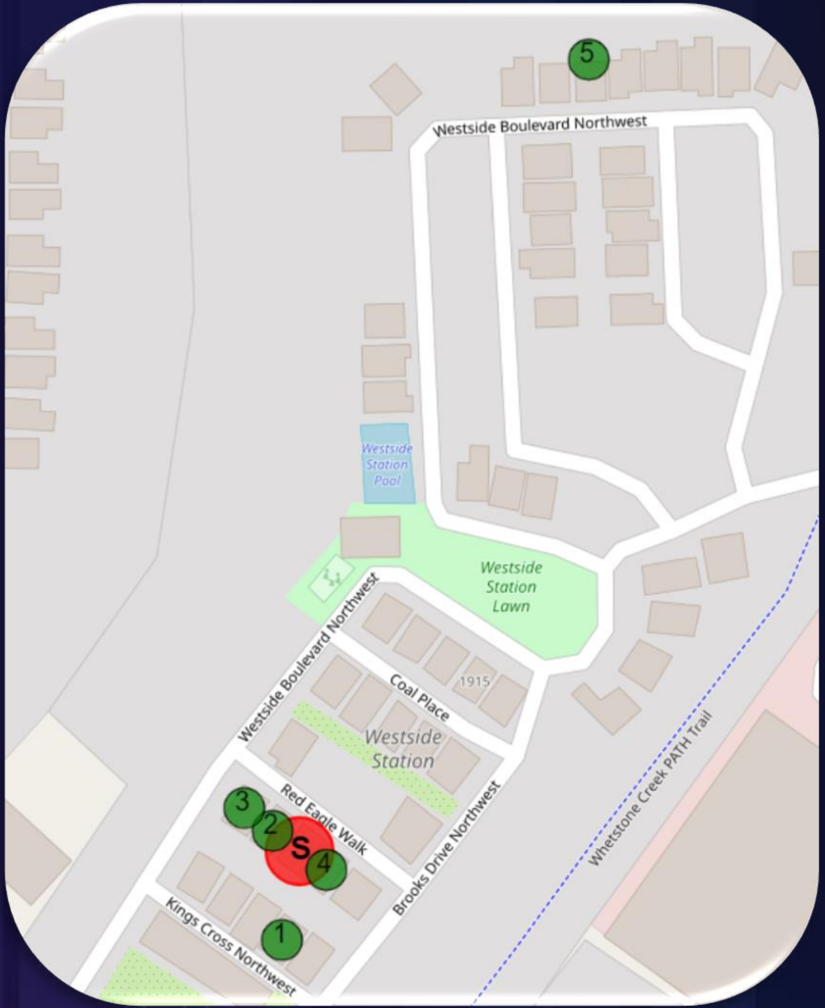


Subject sold on
2/21/2020
Price: $608,900



Comps selected using
Permutation feature importance

Comps selected in CAMA



Our Estimated Price: $654,676.63 Residual: - $45,776.63

CAMA Estimated Price: $570,190.00 Residual: $38,710.00

Individual Subject and Comps Comparison - 3

	SUBJECT	COMP1	COMP2	COMP3	COMP4	COMP5	CAMA_COMP1	CAMA_COMP2	CAMA_COMP3	CAMA_COMP4	CAMA_COMP5
PARID	17 0229 LL4341	17 0229 LL4358	17 0229 LL4309	17 0229 LL4333	17 0229 LL4390	17 0229 LL4366	17 0229 LL4309	17 0229 LL4358	17 0229 LL4366	17 0229 LL4333	17 0229 LL5116
PRICE	608900	609900	614900	609900	575000	568900	614900	609900	568900	609900	584900
SALEDT	2/21/2020	10/9/2019	8/19/2019	6/27/2019	4/10/2019	5/28/2019	8/19/2019	10/9/2019	5/28/2019	6/27/2019	10/8/2019
ADRSTR	RED EAGLE	RED EAGLE	KINGS CROSS	RED EAGLE	RED EAGLE	RED EAGLE	KINGS CROSS	RED EAGLE	RED EAGLE	RED EAGLE	WESTSIDE
NBHD	17365	17365	17365	17365	17365	17365	17365	17365	17365	17365	17365
STYLE	1	1	1	1	1	1	1	1	1	1	1
ZONING	I1	I1	I1	I1	I1	I1	I1	I1	I1	I1	I1
GRADE	B+	B+	B+	B+	B+	B+	B+	B+	B+	B+	A
CDU	EX	EX	EX	EX	EX	EX	EX	EX	EX	EX	GD
LOCATION	3	3	3	3	3	3	3	3	3	3	6
BSMT	4	4	4	4	1	1	4	4	1	4	1
HEAT	4	4	4	4	4	4	4	4	4	4	4
FUEL	1	1	1	1	1	1	1	1	1	1	1
FRONTING	4	4	4	4	4	4	4	4	4	4	9
EXTWALL	1	1	1	1	1	9	1	1	9	1	1
PARKTYPE	1	1	1	1	1	1	1	1	1	1	3
SFLA	$\boxed{2592}$	$\boxed{2408}$	2608	2390	2804	2670	$\boxed{2608}$	2408	2670	2390	2720
GRDFACT	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.55
CALCACRE	0.06	0.06	0.06	0.06	0.06	0.08	0.06	0.06	0.08	0.06	0.112
STORIES	2	2	2	2	2	2	2	2	2	2	2
YRBLT	2019	2019	2019	2019	2017	2019	2019	2019	2019	2019	2018
RMTOT	6	5	7	10	10	9	7	5	9	10	10
RMBED	4	3	3	3	4	4	3	3	4	3	4
RMFAM	2	2	4	1	1	5	4	2	5	1	1
FIXBATH	4	4	3	3	3	3	3	4	3	3	3
FIXHALF	0	0	1	1	1	1	1	0	1	1	
FIXADDL	8	8	8	8	7	8	8	8	8	8	6
FIXTOT	20	20	19	19	18	19	19	20	19	19	15
DEPR	100	100	100	100	100	100	100	100	100	100	96
LAT	33.811398	33.811462	33.811102	33.811333	33.811599	33.811538	33.811102	33.811462	33.811538	33.811333	33.814015
LON	-84.452703	-84.45281	-84.452772	-84.452597	-84.452528	-84.452926	-84.452772	-84.45281	-84.452926	-84.452597	-84.451548
SALEMON	0	2	4	6	8	7	4	2	7	6	2
PARKQUA	3	3	3	3	3	3	3	3	3	3	2

Observation; What we learned

- These are preliminary experiments, but they show very promising potential
- Importance-based comparable selection picks very reasonable comparable sales; similar, and in some cases, slightly better results were achieved in our preliminary experiments
- Comps sales appraisal is simplified using importance-based weights, strong potential to retire MRA and use GBM as the source of comps adjustments
- Comprehensive case study of its usage in practice needed, even so, “best comps” are highly subjective

Discussion and Future work

- Importance in similarity vs. Importance in prediction
- Shift away from empirical distance function (generative like controls)
- Extract hedonic prices from GBM (interpretable)
- Leverage importance and explanation metrics to score (and guide) assessor’s further adjustment of weights and addition of variables with lower predictive importance
- “AI always in the loop” to learn from comp overrides done by assessors; online learning procedures feed learnings forward to future comps selections

Concluding Remarks

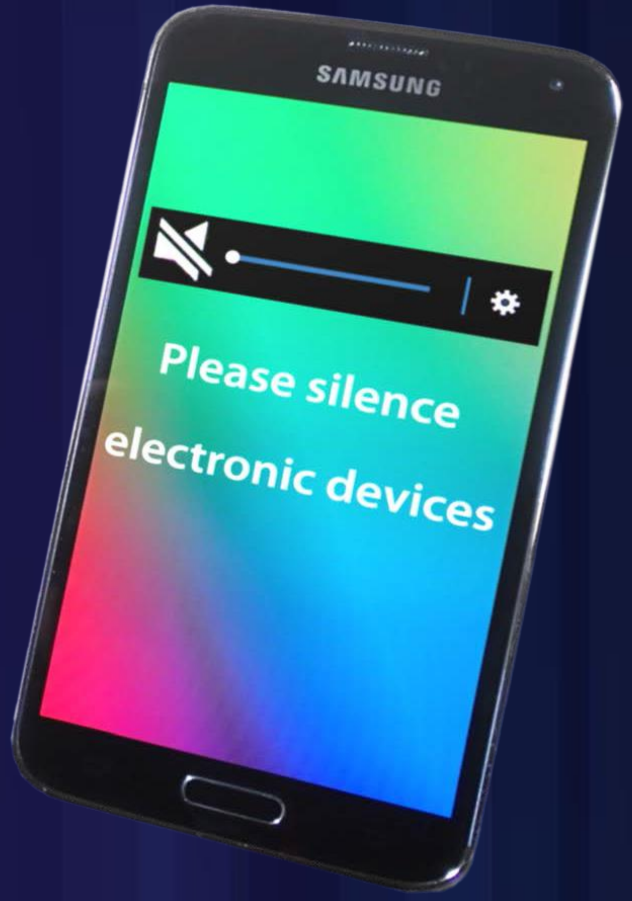
- Empower the assessor; similar performance, less effort
- Avoidance of complexity
  - Avoids preprocessing and MRA Calibration steps
  - Avoids initial comps weight determination step
  - Avoids nbhd segmentation as a pre-req for calibration
  - Avoids tedium; straightforward end-to-end automation
- Good tools for understanding (SHAP, marginal values)
- Intuitive assessor product controls available after the initial AI calibrations; full control without limitation
  - Assessor still has full power to add/remove other selection variables and further adjust or constraint weights

Continuing Education (CE) Credit

Recertification Credit forms for CE credit can be collected from the registration desk on Thursday

Housekeeping

- The conference proceedings will be available approximately 8 weeks after the conference
- Please silence your electronic devices
- Attendance at this conference counts toward GIS Professional (GISP) Certification and Renewal




GIS-Pro 2024
October 7-10, 2024 • Portland, Maine

TOGETHER: DARING TO CLIMB NEW HEIGHTS

IAAO

2024

ANNUAL CONFERENCE & EXHIBITION

August 25-28, 2024 • Denver, Colorado

IAAO | URISA