



daniel hardesty lewis

ai for cities and real estate

I build tools that let investors, developers, and policymakers understand how land actually behaves — beyond zoning rules or spreadsheet assumptions, under real constraints and future uncertainty

my work combines open data, artificial intelligence, and political awareness to evaluate real estate at scale — parcel by parcel, market by market. I focus on explainability, reproducibility, and decision-ready insight, not black-box valuation

selected works and works-in-progress

openRange benchmark & trailMarkers pricing model

a 20-year Manhattan valuation dataset and commercial AI sales pricing model trained on public tax and sales records. achieves comparable accuracy to Zillow on far less data. open-source and built for reproducibility

oppsAlert

an explainable AI model that identifies where neighbors are likely to successfully oppose new housing. accuracy high enough to use before filing. know what drives opposition to redevelopment

dealMiner

a fully automated, broker-free single family portfolio sourcing pipeline. built from tens of millions of public records. in Houston, the model surfaces minimum 11% IRR, 1.65× EM deals in minutes at the fraction of the cost of a back-office team

work-in-progress: propalytics

a generalized valuation engine that identifies finance-first refined asset classes and true comparables of low-density buildings, ranks their probabilities of success under future scenarios, and produces interactive, investor-facing maps of alpha

about me

[NSF Big10](#) | [TOP500 supercomputing](#) | [Bagnold Medal contributor](#)

Columbia GSAPP (Urban Planning, 2026)

UT Austin B.S. [Pure Mathematics](#)

founder @ Summit Geospatial

built at the intersection of land, data, and development

[LinkedIn](#) • daniel@summitgeospatial.com



selected work

openRange benchmark & trailMarkers pricing model

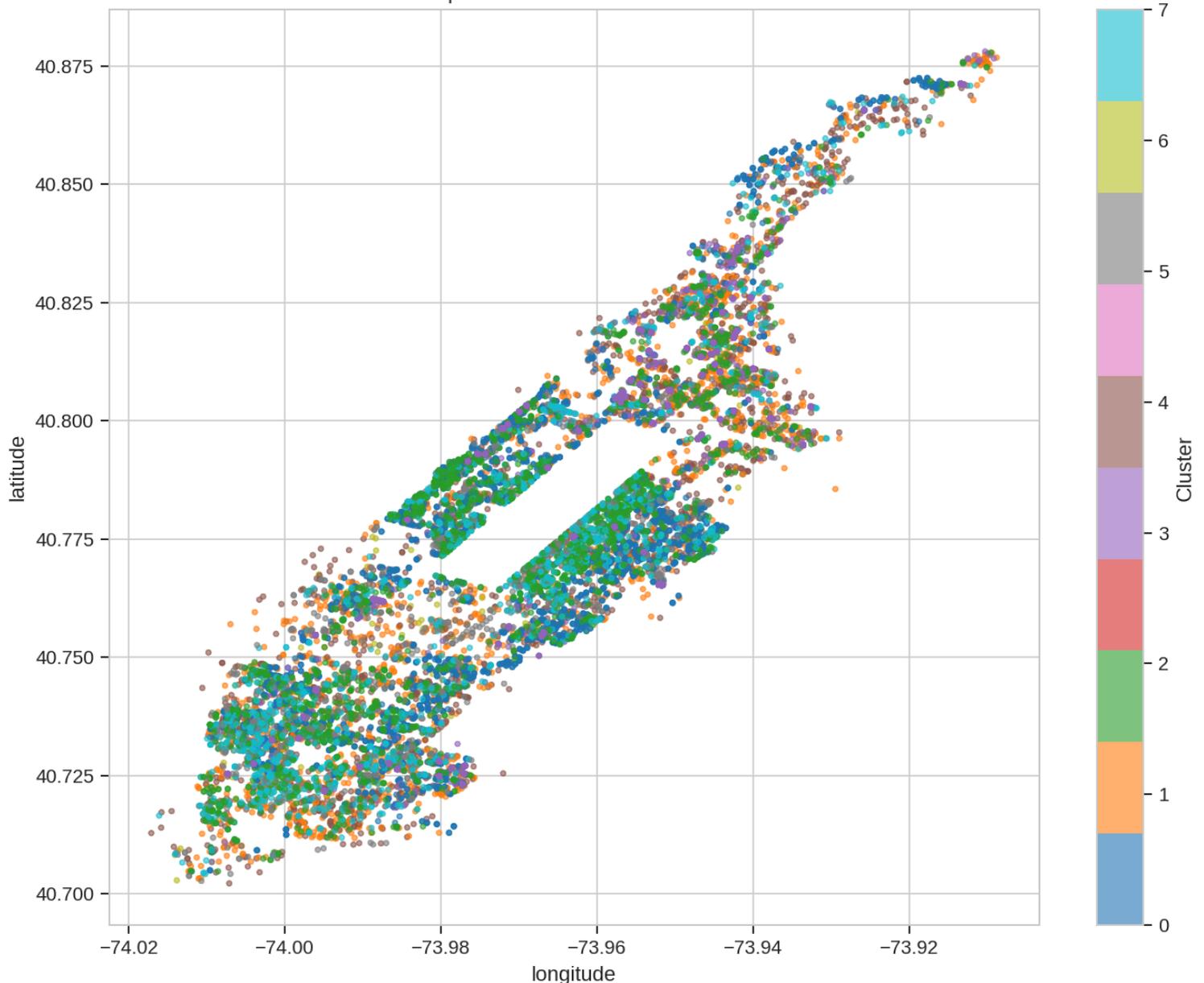
parcel-level valuation

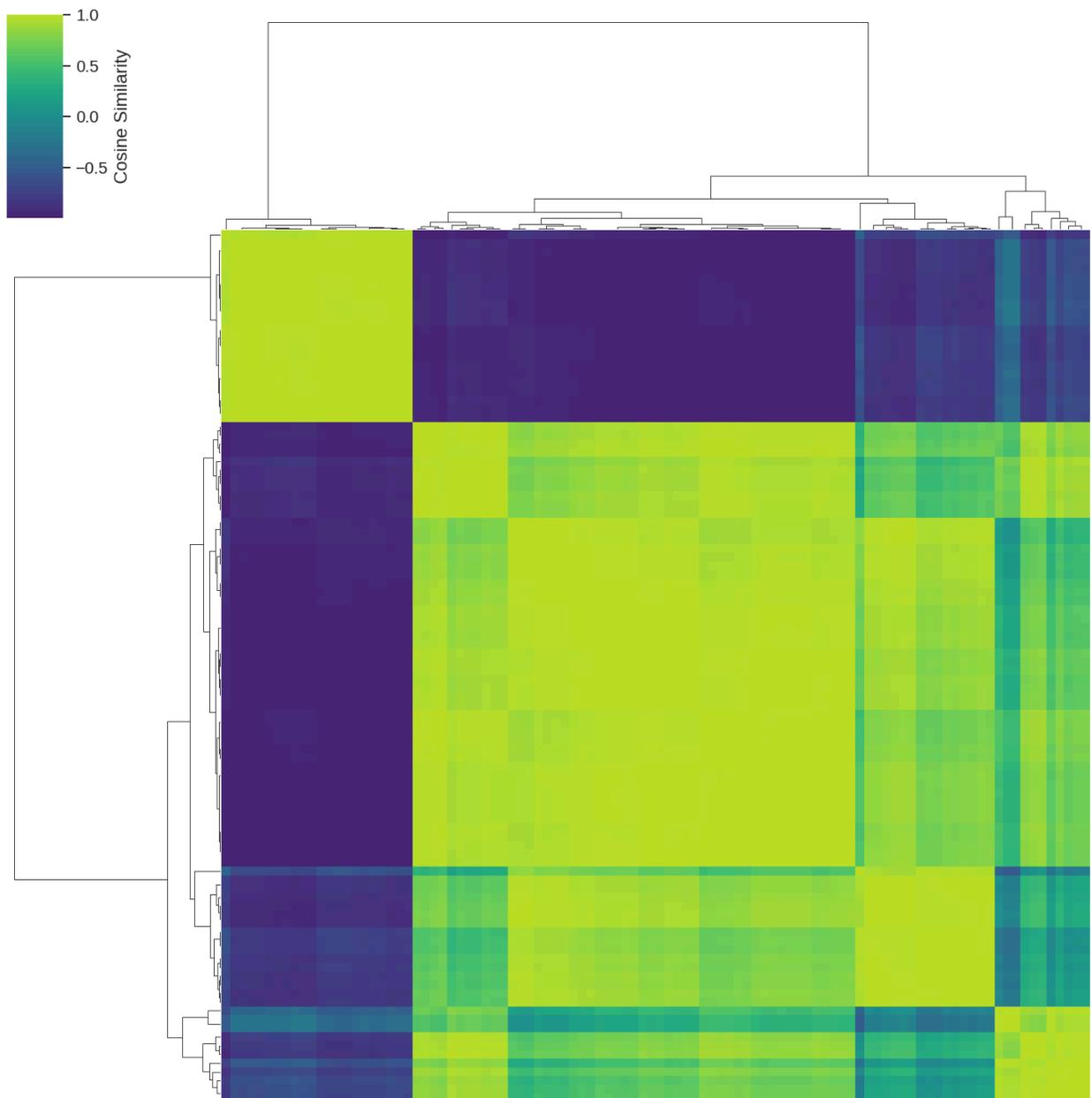
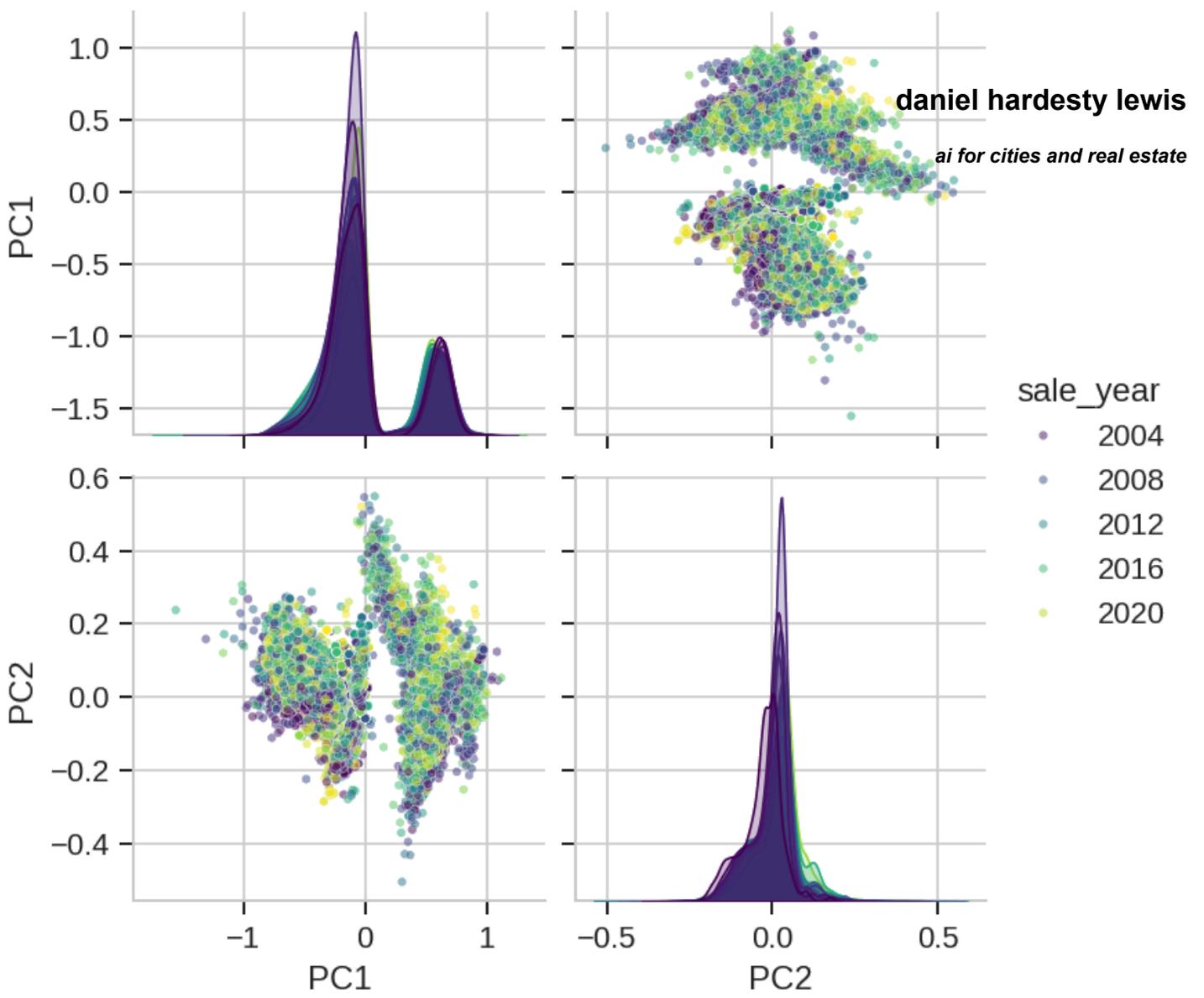
Manhattan, 2003-2022 vintage

- **what we built:** a comprehensive dataset covering 20 years of Manhattan parcel transactions, engineered from NYC's PLUTO tax and ACRIS sales records, with 850,000+ parcel-year observations
- **technical depth:** includes 100+ numeric and categorical variables such as zoning, floor area ratio (FAR), assessment histories, and temporally-lagged sales
- **baseline model:** semi-supervised latent factor model effectively handles missing data and sparse sales labels
- **insights:** reveals market dynamics through stable clusters, helping quantify urban trends like gentrification and investment rotation
- **open & reproducible:** fully documented and open-source, built for easy adoption and iteration
- **within sight of commercial competitors:** 12% MAPE vs Zillow's 8%

latent space of Manhattan

Spatial Distribution of Clusters







selected work

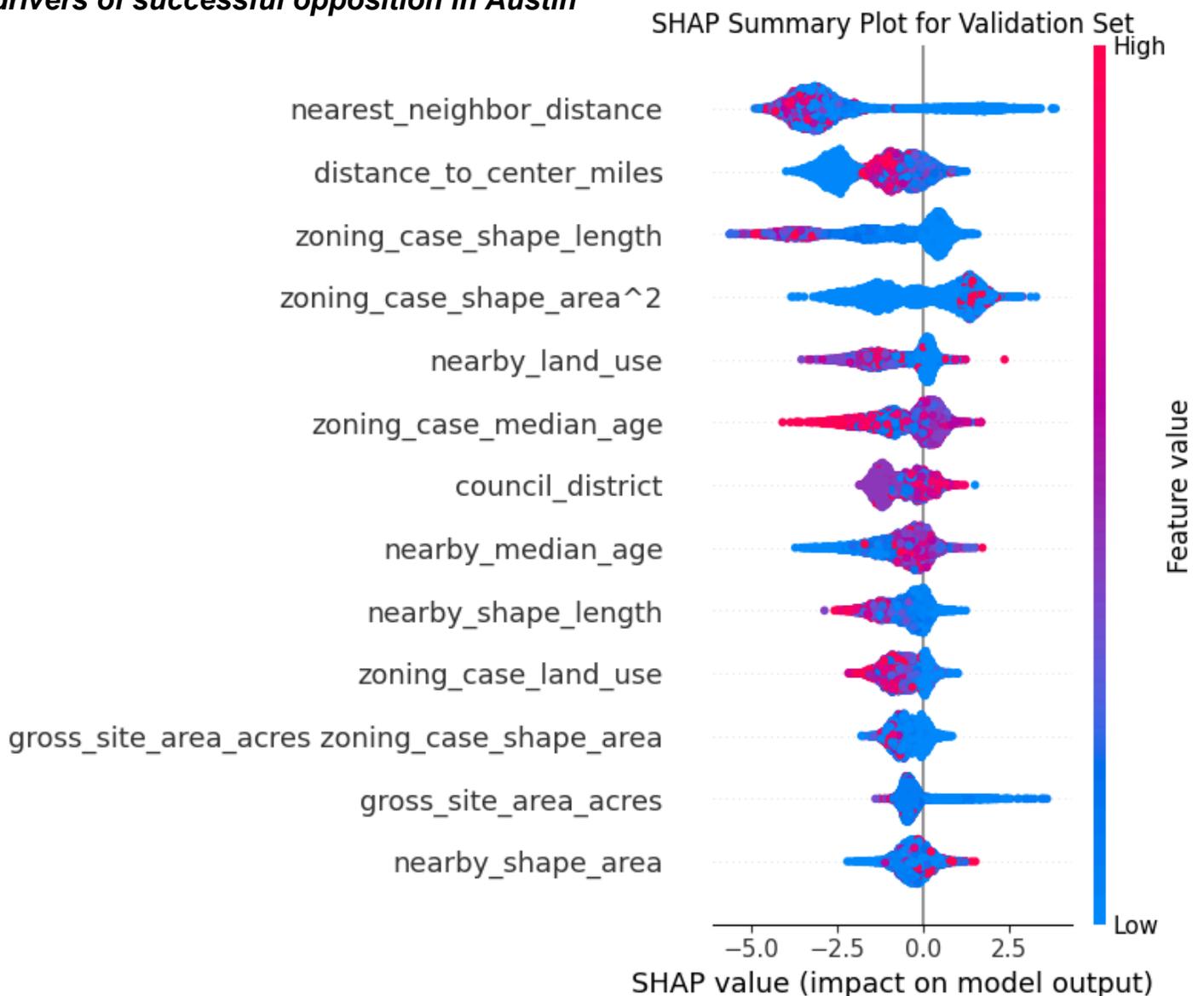
oppsAlert

predictive modeling of zoning opposition

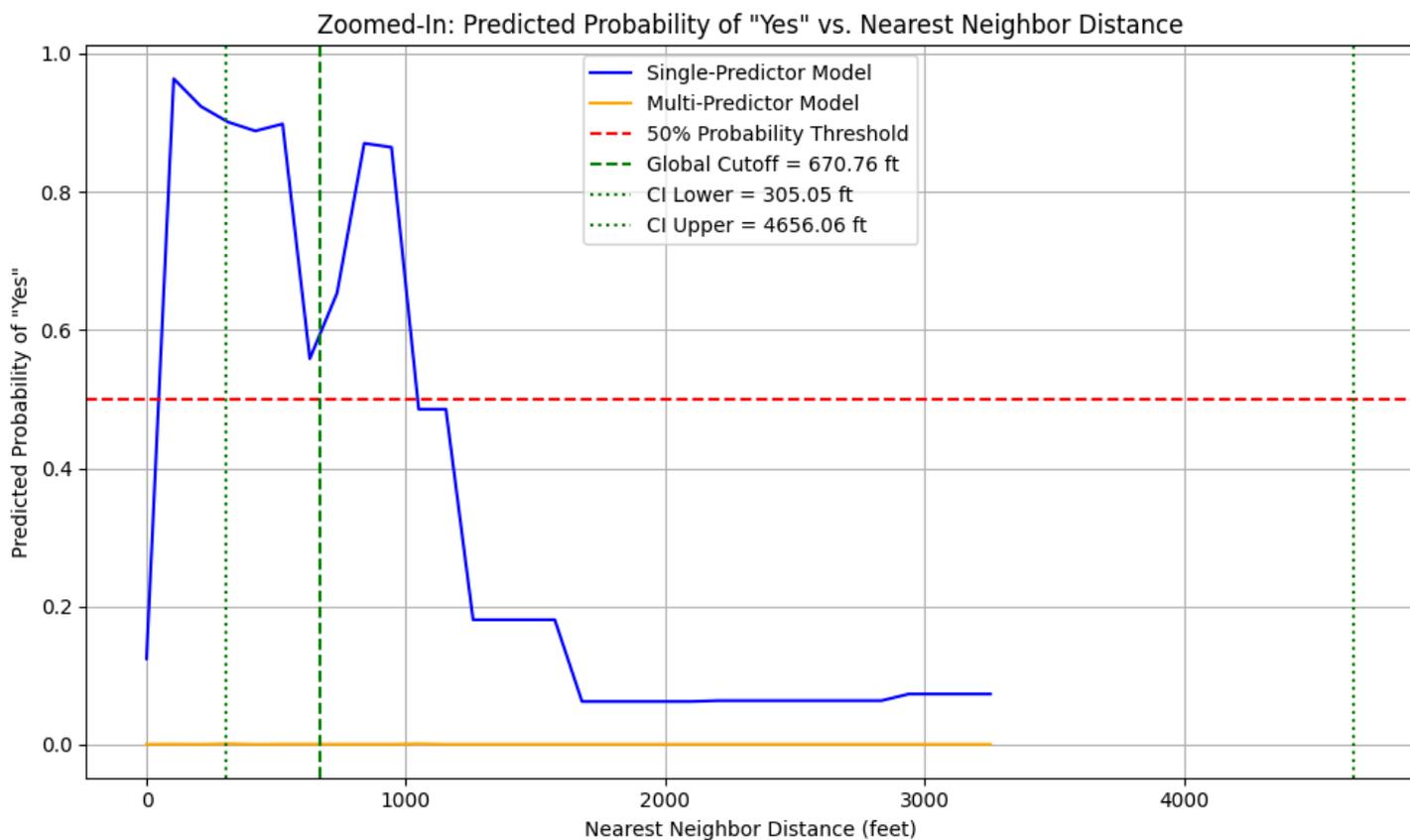
Austin, 2007-2024 vintage

- **why it matters:** identifies areas where a small group of residents could successfully oppose new housing developments, impacting urban growth and housing affordability
- **data-driven model:** utilizes zoning cases, demographic data, and historical opposition records
- **explainable AI:** SHAP-explainable ML model accurately predicts zoning opposition, facilitating proactive community engagement
- **use-case ready:** offers actionable insights for developers, city planners, and advocacy groups, significantly reducing investment risk and delays

drivers of successful opposition in Austin



probability of successful opposition by distance





selected work

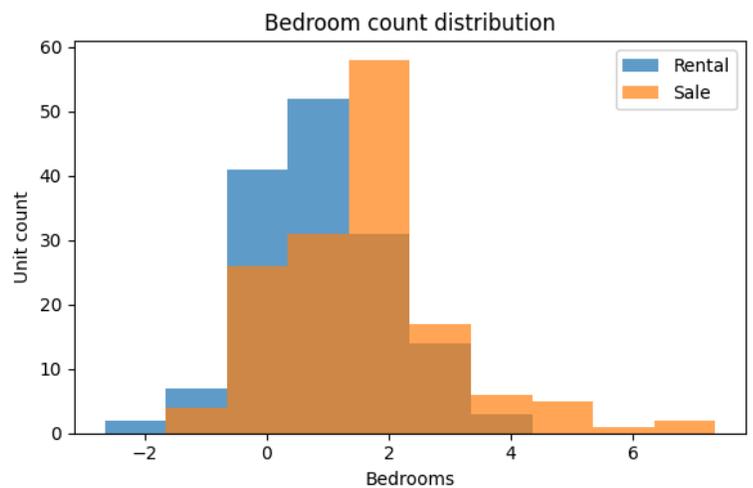
dealMiner

automated single-family rental portfolio sourcing

Houston, 2015-2025 vintage

- **instant deal discovery:** leveraging a backtested heuristic-driven approach, analyzes millions of public property records to instantly identify profitable rental portfolios
 - **workflow simplicity:** easily transferable methodology; ready for replication in new markets with minimal adjustments
 - **proven performance:** demonstrates high predictive accuracy and robust financial returns (IRR and Equity Multiple) in extensive backtesting
 - **broker-free automation:** removes manual effort and subjective bias, delivering consistent investment opportunities quickly
-

sample single-family portfolio



work-in-progress: propalytics

interactive real estate performance heat maps

Queens, 2024 vintage

- **vision:** provide investors and developers with interactive heat maps displaying real estate financial performance metrics (IRR, Equity Multiple) across entire markets
- **dynamic scenario modeling:** users can adjust future market scenarios—such as construction costs and land values — to instantly visualize impact on investment potential
- **advanced clustering:** combines advanced machine learning clustering techniques to accurately represent market asset classes and optimize property comparisons
- **scalable:** generalizable model framework suitable for various geographies, from local neighborhoods to major metropolitan areas

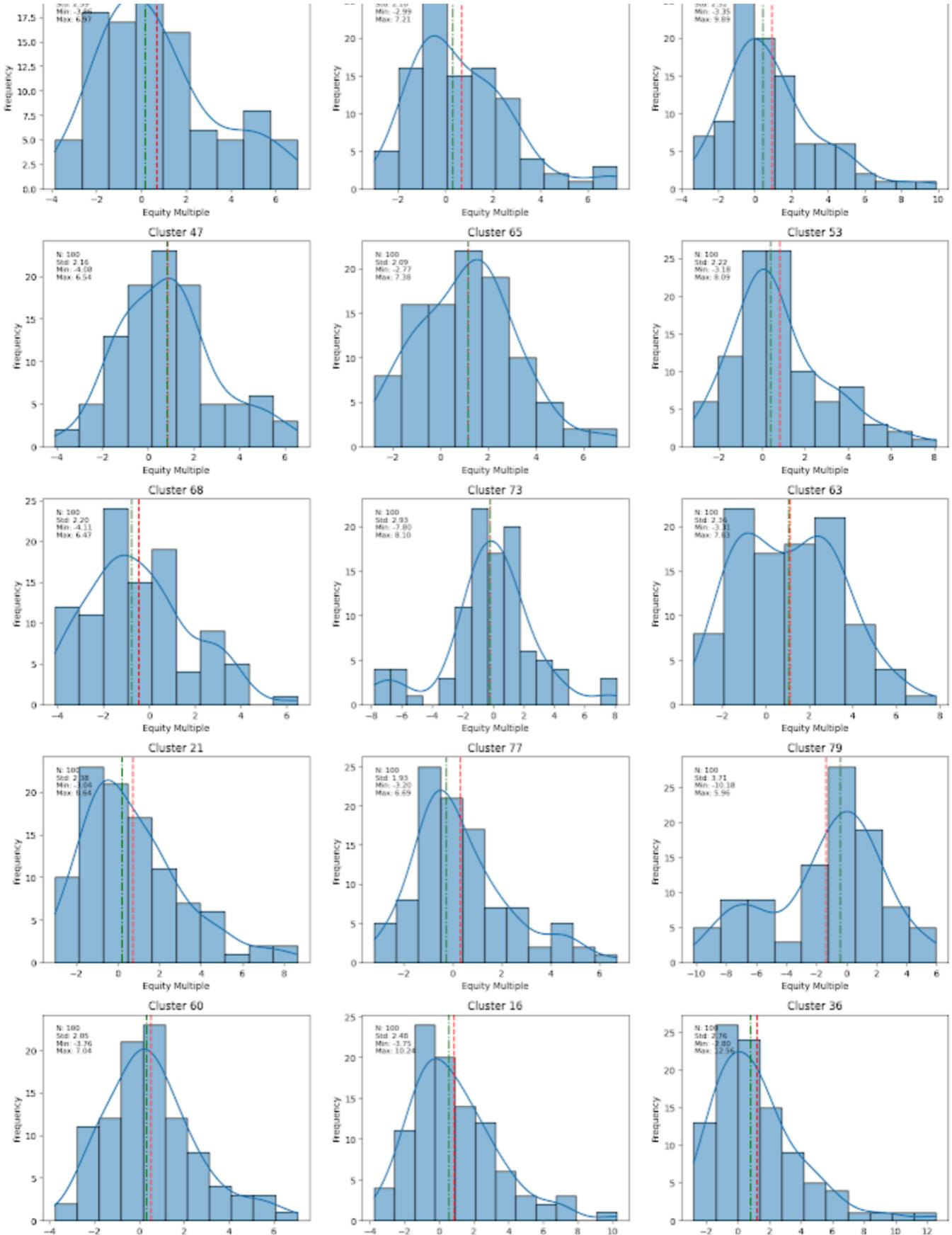
sample clustering in Queens, NYC





selected work

ranking probabilistic financial futures in Queens



contact me

[\(713\) 371-7875](tel:(713)371-7875)

[NSF Big10](#) | [TOP500 supercomputing](#) | [Bagnold Medal contributor](#)

Columbia GSAPP (Urban Planning, 2026)

UT Austin B.S. [Pure Mathematics](#)

founder @ Summit Geospatial

built at the intersection of land, data, and development

[LinkedIn](#) • daniel@summitgeospatial.com

selected work whitepapers follow on the next page



selected work

openRange & trailMarkers: open, parcel-level valuation corpus

Manhattan, 2003-2022 vintage

What we built

- **20-year panel, Manhattan.** Built to handle 1M+ tax-lot-years, engineered from NYC's public tax and sales records
- **Rich feature stack.** 100+ numeric & categorical variables spanning zoning, FAR, assessment histories, and temporally-lagged sales signals
- **Open & reproducible.** Full pipeline, code, and docs on GitHub; trains on a GPU in <8 min

Baseline model

- **Autoconfigured Missing-data Importance-Weighted Auto-Encoder (autoMIWAE)** handles the twin headaches of the domain: **structured missingness** and **extreme label sparsity**: in our test case less than 10% of rows have a sale price
- 2D latent space + unstable-variance aware price head
- **93% R2; 12% MAPE.** Within spitting distance of best-in-class Automated Valuation Models such as Zillow's 7-8% median error
- Generates calibrated uncertainty: within one-sigma of uncertainty for 42% of parcels, which is essential to automate underwriting

Market insights unlocked

- **Binary regime & 8 subsegments.** Latent-space clustering of 75% by silhouette score cleanly separates everyday vs. unique assets, then subdivides into eight stable clusters that we found shift roughly 2% each year, giving a live portrait of gentrification & capital rotation
- **Transparent performance context.** Asset-class by asset-class performance metrics; our benchmark makes open the denominators the industry usually doesn't

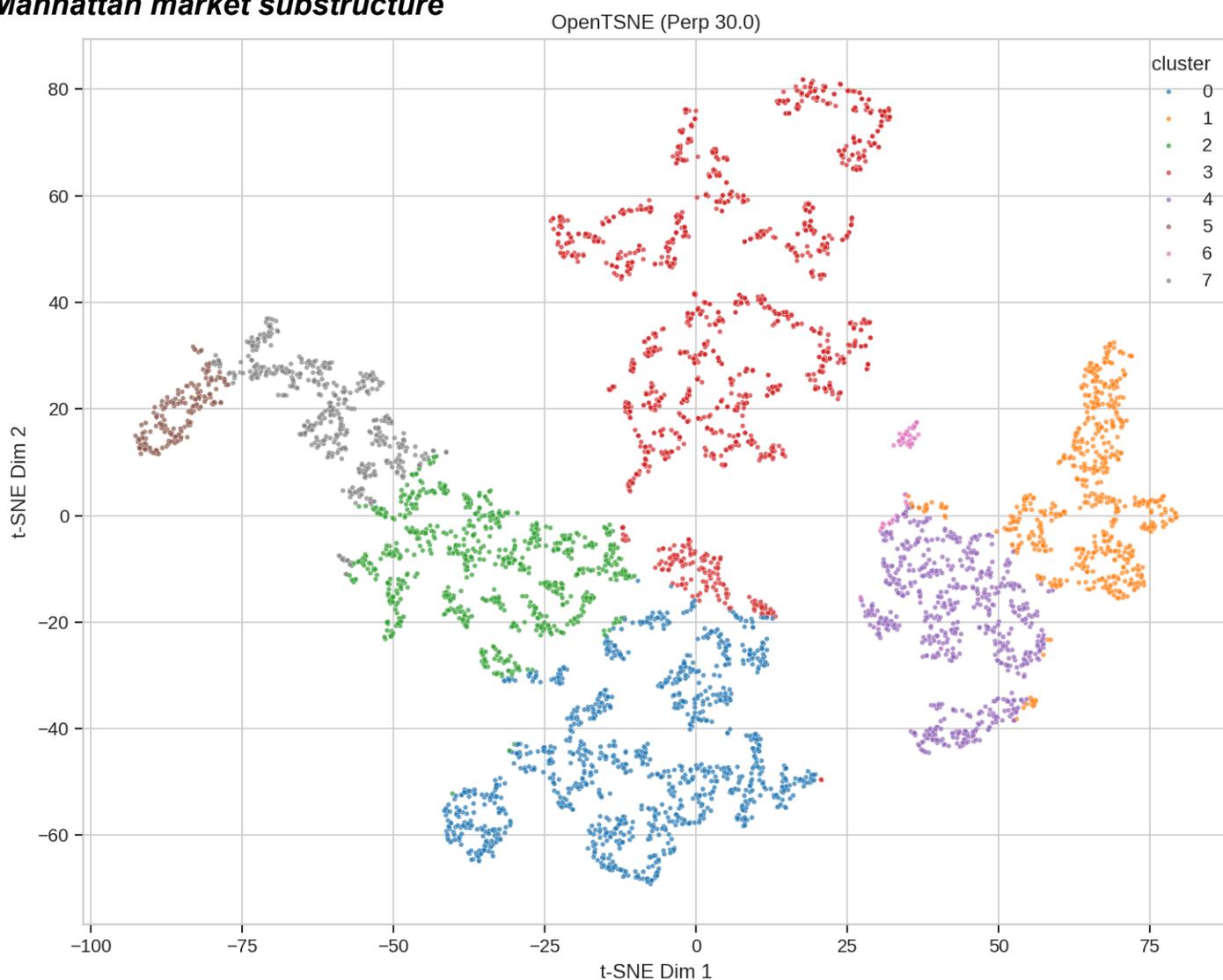
Why you should care

1. **Finally, an ImageNet-grade testbed for real estate.** Apples-to-apples comparisons to the rest of AI
2. **Bridges the data gap.** 100× the typical size of academic sets, big enough to stress-test deep models, yet fully open — no contracts or API tolls
3. **Built for rapid iteration.** Entire loop in the span of a meeting: perfect for live demos, hackathons, or proof-of-concept sprints
4. **Direct line to production value.** Uncertainty scores and cluster IDs can flow into risk screens, mass appraisal triage, or dynamic pricing engines from day one
5. **Road-map ready.** Swap in Brooklyn, Queens, etc, alternative data, and more — our open source code is fertile ground for industry follow-ons

Call to action

If your team is benchmarking new valuation architectures — or needs transparent data to de-risk an investment in AI talent — fork the repo, beat the 12% MAPE, and profit. Every point you shave off moves the needle toward better valuing the market.

Manhattan market substructure



The most important clusters of property features cohere in the above manner.

selected work

a baseline for every asset class on the market

Model	# Observations	Median sale price	RMSE	normRMSE	MAPE	R ²
<i>All single family</i>						
autoMIWAE-A (PLUTO-A)	4,188	\$1.30 M	\$4.00 M	3.08×	22.0%	0.73
Yu RF (NYC-MSA)	4,802	\$0.635 M	\$2.15 M	3.38×	30.9%	0.80
FHFA NYC HPI (NYC-MSA)	~10,000,000 ³	\$1.30 M	\$0.032 M	0.024×	2.4%	0.98
<i>Detached SF</i>						
autoMIWAE-A1 (PLUTO-A1)	230	\$0.80 M	\$0.21 M	0.26×	12.0%	0.93
FHFA All-transactions HPI (US->NYC-MSA-eq.)	>20,000,000 [†]	-	-	-	4.5%	-
Zillow SFH HPI (US->NYC-MSA-eq.)	-	\$2.05 M	\$0.049 M	0.024×	2.4%	-
<i>Detached SF (Zillow-eq.)</i>						
autoMIWAE-A149T (PLUTO-A1+A4+A9 top14%-trimmed)	3,605	\$1.30 M	\$3.25 M	2.50×	15.0%	0.75
autoMIWAE-A1 (PLUTO-A1)	230	\$0.80 M	\$0.21 M	0.26×	12.0%	0.93
Zillow Zestimate Off-market (NYC, lot-level)	-	\$1.50 M	\$0.124 M	0.082×	8.2%	-
Zillow Zestimate On-market (NYC, lot-level)	1,880	\$1.50 M	\$0.036 M	0.024×	2.4%	0.98
<i>Office</i>						
autoMIWAE-O4 (PLUTO-O4)	546	\$299.12 M	\$159.44 M	0.53×	32.0%	0.67
NPI CRE DNN (US->NYC-MSA-eq., 117k sqft)	400,370	\$16.18 M	\$3.69 M	0.23×	10.4%	0.96
NPI CRE DNN (US->NYC-MSA-eq., 2k sqft)	400,370	\$0.276 M	\$0.063 M	0.23×	10.4%	0.96
<i>Vacant land</i>						
autoMIWAE-V1 (PLUTO-V1)	3,531	\$16.77 M	\$8.05 M	0.48×	29.0%	0.76
Melbourne land XGBoost (MB->NYC-MSA-eq.)	26,700	\$11.65 M	\$3.29 M	0.28×	13.9%	0.86

oopsAlert: AI radar for rezoning friction

Austin, 2007-2024 vintage

0 ▶ Why it matters

Austin’s 20%-within-200 ft rezoning protest-petition rule turns a handful of landowners into a veto bloc, stalling housing capacity and pushing costs. 90% of rezonings sail through; the rest blow up budgets and timelines. A parcel-level AI screen lets both the city and developers see trouble coming months in advance and engage early — protecting the public’s equity goals while simultaneously de-risking capital.

1 ▶ Data & feature stack

Layer	Source	Signal captured
Petition files	FOIA pull of every Austin protest from 2007 to 2024	<i>dependent variable</i> : “Yes, I’m opposed / No, I’m not opposed”
Zoning cases	city rezoning files & geometry	proposed vs. existing zoning use
Demographics	census block-group ACS	income, race, tenure gaps between neighbor and owner
Spatial	GIS & TCAD parcels	distance to neighbor, 200-ft buffer from rezoning, miles to city center

Engineering highlights

- 43 raw columns → 51 model features
- Key transforms: demographic deltas, log-distances, zoning × demographic interaction terms
- Rare-event target (2.4 % “Yes, I’m opposed”) with advanced rare-class balancing

2 ▶ Model in production

Variant	Hold-out ROC-AUC	Acc.	F1 (Opposed)	Notes
<i>XGBoost tuned</i>	0.968	98%	0.58	<i>selected</i>
RF / GBoost	0.95–0.96	94-97%	0.38-0.53	a close second
Logit	0.82	63%	0.10	the baseline

10-fold CV + 20 % time-split hold-out; training code in Colab notebook.



selected work

Top SHAP drivers of successful opposition

1. *distance to neighbor* ↓
interpretation: opposition is local
 2. *distance to city center* ↓
interpretation: opposition *increases* closer to city center
 3. *owner more likely hispanic than neighbor* ↑, *owner higher-earning than neighbor* ↑
interpretation: differences drive opposition
 4. *neighbor more likely a renter than owner, neighbor commutes more than owner*
interpretation: decreased housing stress increases opposition
-

3 ▶ Rules of thumb for planners

- **Zone of influence ≈ 700 ft:** inside this ring, odds of a valid petition double
 - **Equity lens:** less opposition doesn't mean less attention: tailor outreach & benefits to majority-renter blocks with large Hispanic share
 - **Council district map:** in Austin, districts 3, 5, 9 drive 60 % of historical opposition: focus staff time there
 - **What-if policy tests:** lowering the threshold for a valid petition to 10% cuts predicted oppositions by 41% & expanding the valid buffer to 300 ft adds only 4% new opposants
-

4 ▶ Use-cases

Stakeholder	How to act on the score
City planning	Queue flyers & small-area meetings for parcels which have a ≥60% probability of successful protest
Developers	Bake 6-week cushion & design concessions into budgets where probability is >40%; otherwise stick to the standard timeline
Housing advocates	Spot silent-exclusion zones – where there is low opposition but high need – and lobby for inclusionary engagement overlays
Legal reform	Simulation shows where a super-majority hurdle will occur, which drives many dead deals. The evidence we provide helps legislators tweak the statute, as they are considering doing right now in 2025

5 ▶ Big take-aways

1. **Opposition is predictable**
2. **Location & demographic deltas explain more than 97% of variance**
3. **Transparent, SHAP-explainable model makes the politics actionable**
4. **Scalable:** this same approach is transferable to any protest-statute state; next-up will be NYC

Bottom line: A 30-second API call can now warn: “This rezoning will need a supermajority council vote and has a 68% probability of protest delay — budget extra outreach or adjust the site plan.” That’s planning with eyes wide open.

selected work

dealMiner: vibe-driven portfolio picker

Instant deal discovery from 40 M+ public records

Houston, 2015-2025 vintage

0 ▶ Why read this?

In the time it takes to skim this page, a rules-of-thumb picker plus a six-factor linear regression can vacuum 40 million U.S. tax-lot rows, rank them, and surface an **institutional-grade single-family-rental (SFR) acquisition**. Everything here is machine-selected; dropping the workflow into another metro is switching the source, not a complete re-model.

1 ▶ Workflow as applied to Houston, in 90 seconds

Step	Code block does...	Key numbers
Clean snapshot	one row per lot → cluster-median impute → repair common recording errors	take sufficient sample of detached homes, for Houston: 25k
GPT where location matters	vibe-code the location, location, location → in Houston: GPT easily tells you it's the Inner Loop, Outer Loop, and Outside the Loop	location..., an absolute necessity
Have ChatGPT score	vibe-code what matters → in Houston, that's: (1) replacement discount +30%, (2) cheap \$/SF -20%, (3) amenities +15%, (4) bedrooms +15%, (5) quality +10%, (6) permit freshness +10%	which will yield your final_scores
Calculate yield	rent PSF = $\max(0.75, \text{base} - 0.0001 \times \text{SF})$ with base 1.25 / 1.05 / 0.85 by band	use GPT to mirror this against your local MLS's leases
Pick	the top rentals with $\geq 8.5\%$ yield & the top flips	produce a portfolio
Roll-up	aggregates to Gross Potential Rent	then feed the pro-forma

We backtested this approach against Houston for 2015-25, using 10 temporal folds, and we achieve a **99.5% R², median abs. error \$9.4k (4.6%)**. — In markets like Houston, SFR values behave like stationary bonds, so plain linear regression was sufficient.

2 ▶ Sample single-family portfolio

Metric	Result
Units / SF	300 homes · 817 k SF
Purchase price	\$98.5 M (\$121/SF)
Yr-1 NOI / cap	\$5.42 M · 5% cap
Capital stack	75% LTC @ 6%; GP/LP = 10%/90%
Yr-5 exit	\$121.6 M @ 5% cap
Returns	11% project IRR · 1.65× EM · 15.7% GP IRR
Sensitivities	+100 bp rent → 14.3% IRR · +100 bp cap → 5.3% IRR

Vibe the proforma based on the recent context for Houston: record 44.6 k SFR leases (+3.3 % YoY); REIT exits 3.8–4.2% caps. We underwrite at 5% to absorb Harris Co. 1.77% tax drag

3 ▶ Upcoming build-out

Tier	Status	Next steps
Data lake	ingested 20 years across 3 metros	refresh them quarterly
Extract the vibes	99% R ² in Houston	validate vibed factors for other metros
	model upgrade	now port to NYC real-estate data
Pick properties	minimum established for Houston	plug-and-play to other metros
Fund wrapper	scoping	PE / REIT / retail sleeves

Think Fama-French for housing: metro cycle, scale, discount, yield spread, permit momentum—cleaner data means higher explanatory power

4 ▶ Take-away for AI-first investors

dealMiner demonstrates a **data-and-vibes-driven, minutes-long** pipeline that delivers a minimum diversified, levered IRR of 11% — with no need to whisper to brokers or fold out hundreds of Excel tables. The same script can light up any metro, feed Argus, and provide a baseline from which to select the best deals. Everything is fully open so your team can replicate *and beat* these numbers in your next target city.



selected work

work-in-progress: propalytics — market-wide alpha, one glance at a time

Queens, 2024 vintage

What it is

An **interactive web heat-map** that colors every parcel in a metro by its projected development returns (IRR / equity multiple). Zoom out for a metro-wide capital radar, zoom in to a single lot for a deal teaser — all continuously recomputed under the future you believe.

How it works

Pipeline stage	Core method	Why it matters
1. Refined asset classes	Pareto-tuned clustering on FAR, height, existing SF, latitude/longitude	yields 20k–50k low-density parcels into 80 statistically tight building typologies — which are drop-in property comps at scale
2. Scenario generator	market sliders or pre-loaded data feeds for cost, revenue, debt, hold, discount & growth	let users impose <i>their</i> macro/micro view instead making the assumption for them
3. Monte-Carlo sampler	100–1 000 draws per cluster per scenario	yields rank-ordered financial-returns curves with confidence intervals
4. Map renderer	web tile service (fork of Summit Geospatial stack)	instantly zoom in from the MSA level down to the parcel

The user experience

1. **Pick a market** (e.g. Houston and the NYC outer boroughs)
 2. **Adjust sliders** — construction cost \uparrow 10%, cap-rate drift +75 bp, etc
 3. **Heat map re-colors** in <2 s; hover any parcel for IRR curve, click for CSV/Argus export
 4. **Bookmark top decile clusters**; download lead list or invite propalytics to help out
-

Why existing tools fall short

Incumbent focus	The gap we fill
CoStar / CompStak — descriptive rents & sales	No forward-looking return math.
Internal mega-models (e.g., Greystar 400-sheet workbook)	Over-parameterized; can't sweep an entire metro in minutes.
Zillow iBuying	Guarded, single-family only, kept proprietary.

propalytics is **purpose-built for first-pass alpha discovery**.

Business model

- **SaaS subscription** (tiered by number of metros & scenario runs) covers compute + data
 - **2-5% promote share** on deals sourced via the platform — aligns incentives with client upside
-

Traction & tech proof

- NYC pilot in Queens: 25,000 parcels, 80 refined asset clusters, <3 weeks dev time
- Bootstrap Monte-Carlo engine already benchmarks more than 50 \times as fast than manual desktop Argus loops

selected work

- Advisory conversations: MetaProp, CRE attorneys, SFR developers, institutional LP analysts
-

Next steps on the roadmap

1. **Scale to 10 top MSAs** (Houston, Phoenix, Atlanta, etc.) by Q4 2025
 2. **Portfolio API** — stream parcel scores straight into investment CRMs
 3. **Fair-value overlay** using satellite & permit deltas for mid-cycle repricing
 4. **Capital-partner marketplace**: match high-scoring sites with equity and debt
-

propalytics gives investors a head-start portfolio filter — and the confidence they began their search where the alpha lives.