

Syracuse Urban Compliance Debt and Risk Forecasting

Track 3 - Urban Data Analysis (CuseHacks Datathon 2026)

Thesis. We quantify Syracuse's compliance debt (overdue open code-violation obligations), show where it concentrates and how city pressure signals co-move (crime, parking), and forecast which parcels/neighborhoods are most at risk next month.

Key result	What it means
Compliance debt is widespread	93.9% of open cases with a comply-by date are overdue (median 367 days).
Debt is concentrated	Top 10 neighborhoods account for 60.5% of total overdue-days.
Prediction is actionable	Top 1% highest-risk parcels achieve precision 0.146 (~4.1x lift over baseline).

Datasets used. Syracuse Parcel Map (Q1 2024), Code Violations (2017-present), Crime (2023-2025; trends use 2023-2024), Parking Violations (2023-present).

Unit of analysis. Parcels (SBL join) and neighborhoods (parcel-dissolved polygons).

Reproducibility. End-to-end notebook produces tables (CSV) and figures (PNG) under outputs/. No external datasets or LLMs used.

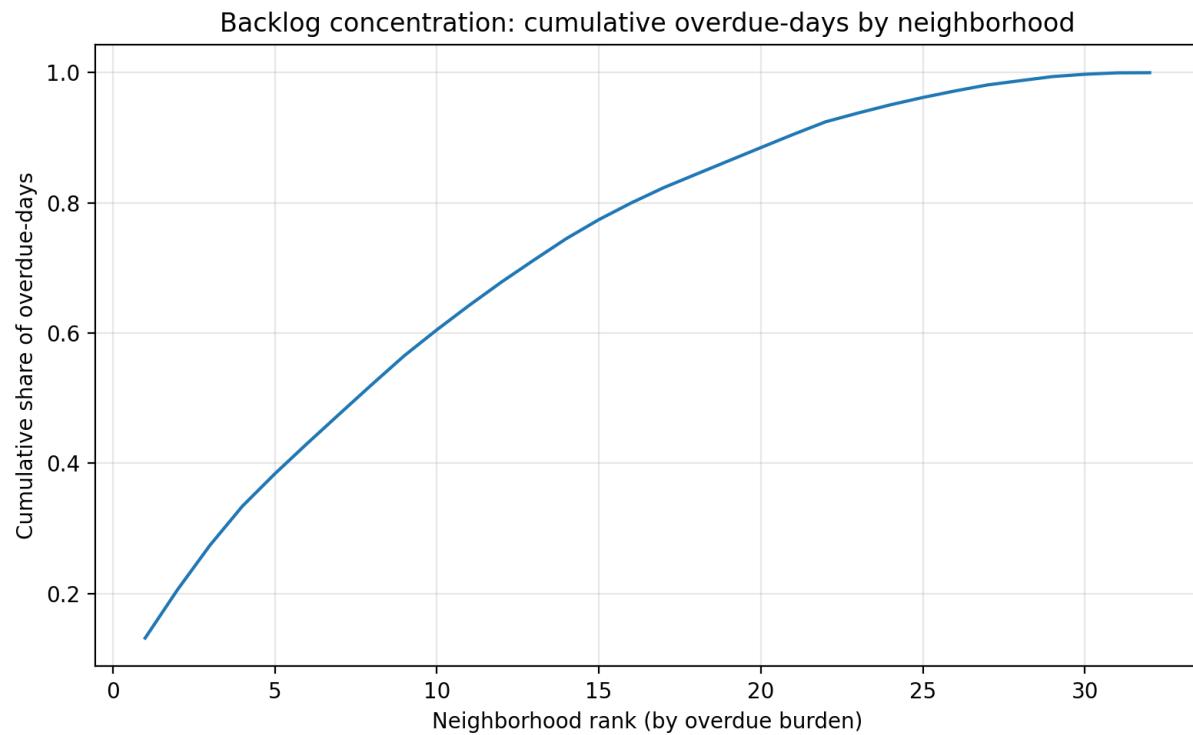
Best Insight: Measuring Compliance Debt

As of 2026-02-19 16:41:18.820, Syracuse has **17,985** open code-violation records. Among open records with a valid comply-by date (**17,969**), **93.9%** are overdue. Median overdue time is **367 days**; 95th percentile is **1848 days**.

Top neighborhoods by overdue-days burden:

Neighborhood	Overdue-days	Share of total
Northside	1,403,959	13.2%
Brighton	793,516	7.4%
Near Westside	728,028	6.8%
Eastwood	639,981	6.0%
Southwest	530,680	5.0%

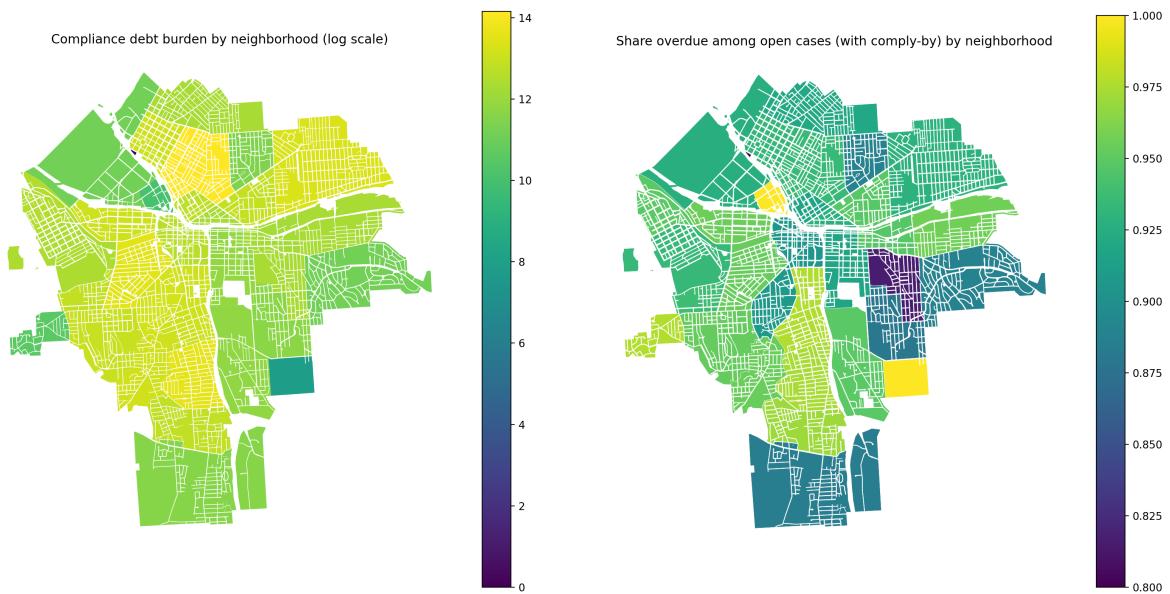
Backlog concentration curve



Interpretation: the first 10 neighborhoods account for about 60% of all overdue-days.

Best Visualization: Where debt concentrates

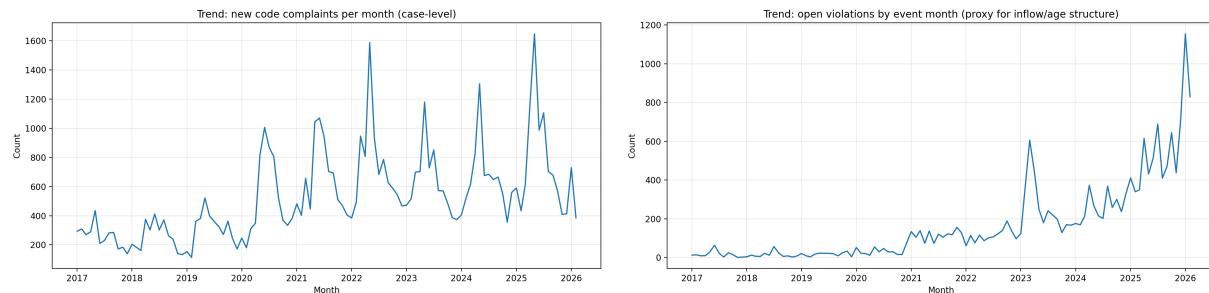
Neighborhood choropleths are built by dissolving parcel polygons and aggregating parcel-linked code records.



Left: compliance debt burden (log overdue-days). **Right:** overdue share among open cases (clipped to 0.80-1.00).

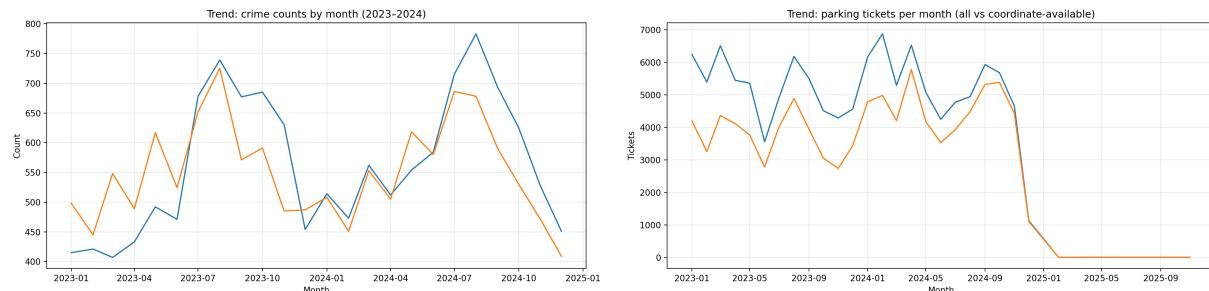
Best Trend: Enforcement volume over time

We report (1) new complaint/case inflow over time (case-level), and (2) event-month distribution of currently-open violations (proxy for backlog age structure; closure dates not available).



Best Trend: City pressure signals (2023-2024)

Crime trends use 2023-2024 (avoids partial 2025). Parking is shown as all tickets vs coordinate-available subset used for spatial features.



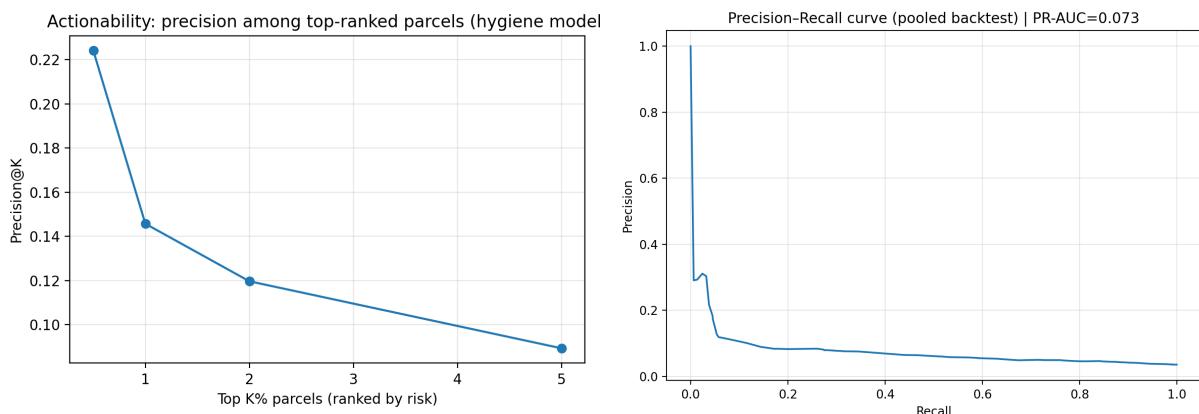
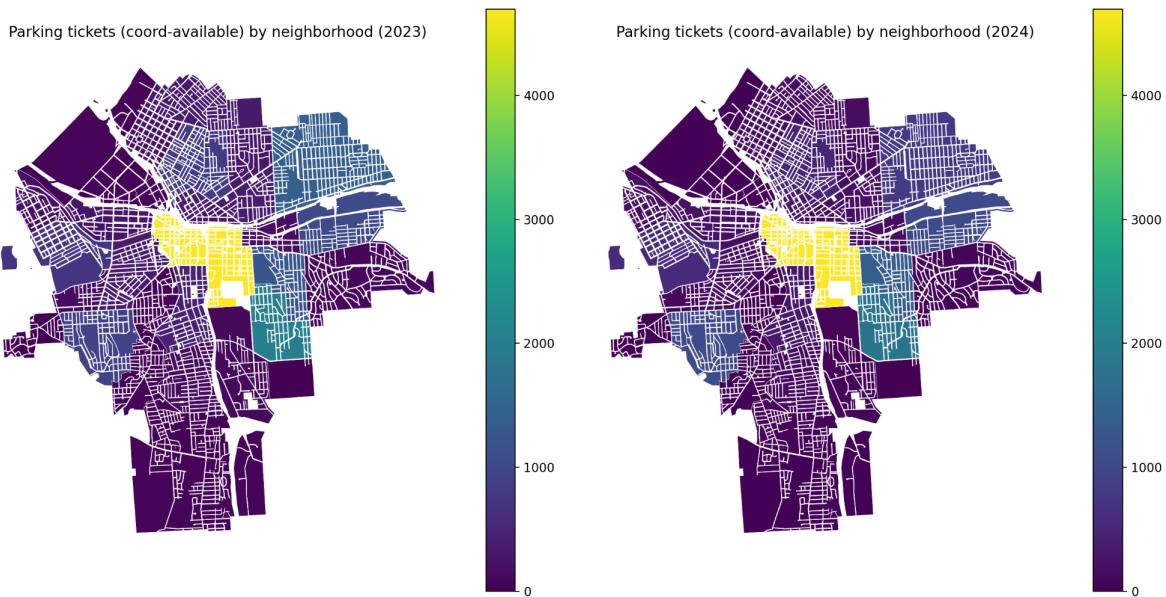
Best Visualization: Shared-scale spatial comparisons (2023 vs 2024)

Maps use shared color scales (vmax capped at the 95th percentile across both years) for comparability.



Best Prediction: Parcel-level next-month risk

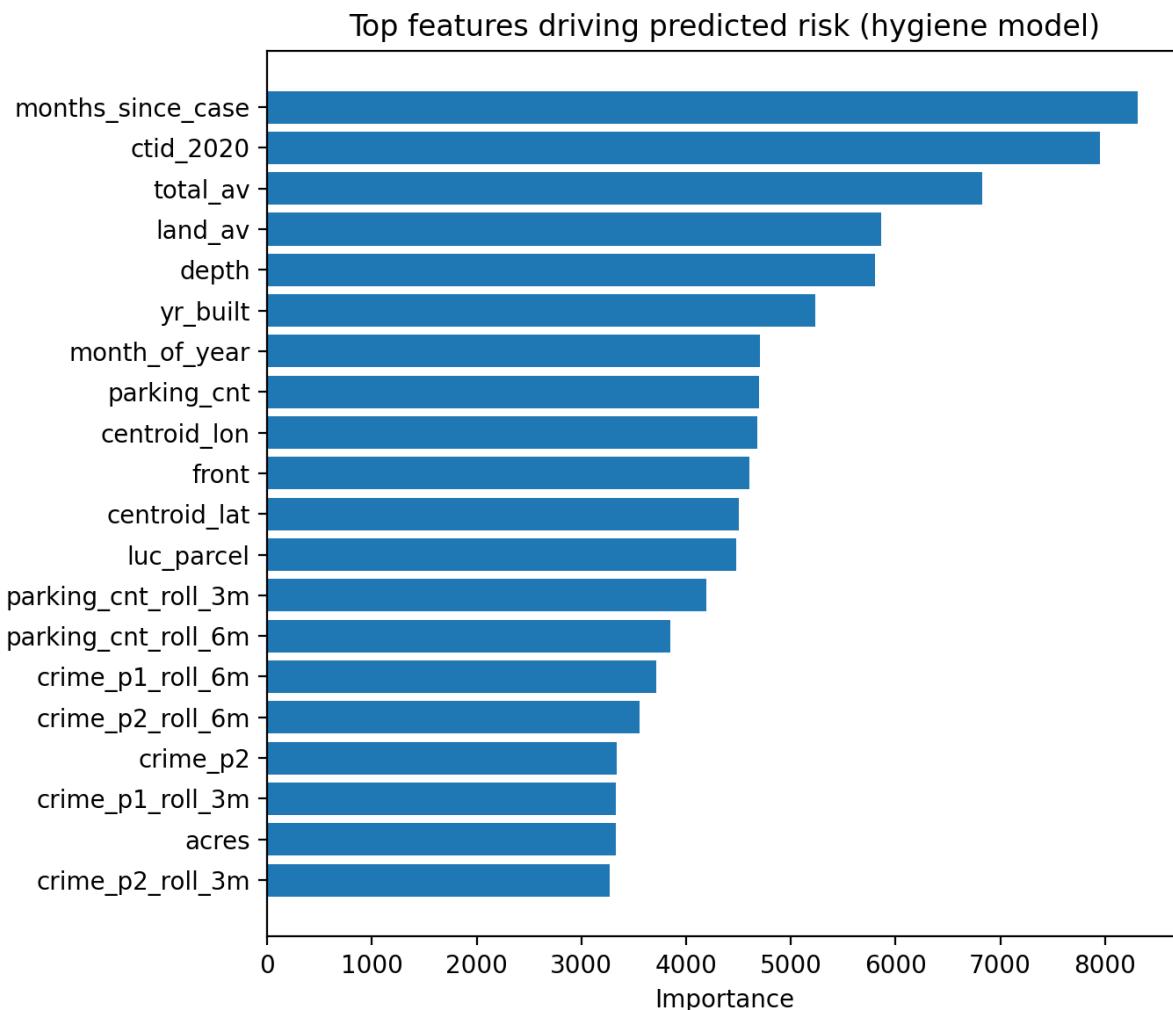
Target: whether a parcel receives at least one new code complaint next month ($y_{\text{next_1m}}$). Model: LightGBM with time-based backtesting. Hygiene feature set removes ID-like fields.



Actionability: Precision@1% = 0.146 (~4.1x lift over baseline 3.57%).

Interpretability and validated risk surface

Top drivers combine enforcement recency, parcel characteristics, seasonality, and neighborhood pressure (parking, crime).



Validated risk surface (2024-12 fold): top neighborhoods by mean predicted risk and observed next-month case rates.

Neighborhood	Mean risk	Actual rate	n parcels
Downtown	0.055	0.082	158
Hawley-Green	0.053	0.043	211
Prospect Hill	0.052	0.014	210
Park Ave.	0.051	0.074	475
Franklin Square	0.051	0.150	20
Washington Square	0.050	0.041	788
Near Eastside	0.050	0.019	258
Northside	0.050	0.031	2135
University Hill	0.050	0.037	191
Skunk City	0.049	0.032	404

Recommendations, limitations, reproducibility

Recommended use. If inspection capacity is limited, prioritize the top 0.5-2% highest-risk parcels each month, then allocate additional resources to high-burden neighborhoods highlighted by compliance-debt maps.

Limitations. Closure timestamps for closed cases are missing, so backlog stock cannot be reconstructed exactly. Parking coordinates are missing for some tickets, so spatial features use the coordinate-available subset. Crime 2025 data is partial; modeling window uses 2023-2024.

Reproducibility checklist. Run notebook top-to-bottom; outputs written to outputs/tables and outputs/visualizations. No external datasets and no LLMs were used.