NYC 2025 Mayoral Election Forecast Model: Two-Compartment Multinomial Ridge with Offsets

Peter Driscoll

November 4, 2025

Purpose

This document is the authoritative specification for the NYC 2025 mayoral forecast. It defines the target, model form, priors, features, turnout assumptions, estimation, calibration, and outputs. It is the single source of truth for implementation and communication.

1 Objective and Unit of Analysis

Target: Election District (ED)–level *vote shares* for three candidates $\{k\} = \{\text{Mamdani}, \text{Cuomo}, \text{Sliwa}\}$, aggregatable to AD/borough/city. We forecast *shares* (not totals). Turnout is used to shape composition and aggregation weights, not to directly predict counts.

2 Model Overview

We model two disjoint "compartments" per ED i: EV (banked early voters) and R (remaining/Election Day voters). Within each compartment we use a multinomial logistic model with an offset prior, then mix the two by their ED-specific turnout weights.

2.1 Compartment models (per candidate k)

$$\eta_{ik}^{(c)} = o_{ik} \; + \; \gamma_k^{(c)} \; + \; \beta_k^\top \mathbf{x}_i^{(c)}, \qquad c \in \{\text{EV}, \text{R}\}, \tag{1}$$

$$p_{ik}^{(c)} = \frac{\exp\left(\eta_{ik}^{(c)}\right)}{\sum_\ell \exp\left(\eta_{i\ell}^{(c)}\right)}, \qquad \sum_k p_{ik}^{(c)} = 1. \tag{2}$$

2.2 Two-compartment mixture to ED-level share

Let T_i be the projected total turnout for ED i under a turnout scenario (Sec. 4). Define weights

$$w_i^{\text{EV}} = \frac{\text{EV}_{25,i}}{T_i}, \qquad w_i^{\text{R}} = 1 - w_i^{\text{EV}}.$$

The predicted ED share is

$$p_{ik} = w_i^{\text{EV}} \, p_{ik}^{(\text{EV})} \; + \; w_i^{\text{R}} \, p_{ik}^{(\text{R})}, \qquad \sum_k p_{ik} = 1. \tag{3}$$

3   Offsets (Pattern Priors)

Offsets o_{ik} anchor spatial patterns using the freshest and most relevant signals.

- **Mamdani:** $o_{iM} = \text{logit}\big(\text{Dem-primary Mamdani share in ED } i\big)$.

- **Cuomo:** $o_{iC} = 0$ (no direct Democratic-primary analog; learns from features and calibration).

- **Sliwa (weak stabilizer):** $o_{iS} = \alpha_S \, \text{logit}\big(\text{Sliwa 2021 share in ED } i\big)$ with small α_S to avoid over-weighting history while providing a plausible GOP spatial pattern.

All non-offset effects are shrunk via ridge (L2) to promote stability with small, noisy tabular data.

4   Turnout Scenarios and Weights

We do not directly model totals. Instead, turnout shapes composition and aggregation via T_i and the compartment weights.

- **Total turnout scenarios:** $T \in \{1.80\text{M}, 1.94\text{M}\}$.

- **Cannibalization of E-Day by EV:** $\rho \in \{0.4, 0.6\}$ is the fraction of 2021 Election Day volume cannibalized by 2025 EV.

- **Per-ED projection:** Let $\text{EDay21}_i = \text{Total21}_i - \text{EV}_{21,i}$. Define

$$\widehat{\text{EDay25}}_i \; = \; (1 - \rho)\, s \cdot \text{EDay21}_i,$$

where s scales so that $\sum_i \big(\text{EV}_{25,i} + \widehat{\text{EDay25}}_i\big) = T$. Then $T_i = \text{EV}_{25,i} + \widehat{\text{EDay25}}_i$ and $w_i^{\text{EV}} = \text{EV}_{25,i}/T_i$, $w_i^{\text{R}} = 1 - w_i^{\text{EV}}$.

5   Feature Sets (MECE by compartment)

All features are standardized (z-scores). Use a compact set to avoid collinearity; ridge will shrink noisy terms.

EV compartment (banked voters): $\mathbf{x}_i^{(\text{EV})}$

- youth_prop$_i$: ACS share age 18–29.

- student_share$_i$ *or* near_campus$_i$ (choose one).

- %EV_newly_registered$_i$ (best) *or* new_reg_rate$_i = \frac{\text{new regs 2024–25}}{\text{registrants}}$.

- income_above_median$_i$ (or tract income percentile).

- Optional access: ev_site_near$_i$ or distance.

Remaining / E-Day compartment: $\mathbf{x}_i^{(\mathbf{R})}$

- income_below_median$_i$ (complement to EV).

- remaining_gap$_i$ = target_bin_EV_rate − ED_EV_rate$_i$ (who has not voted yet).

- renter_share$_i$ *or* recent_mover_share$_i$ (pick one).

- Retain youth_prop$_i$ (late youth may vote on E-Day).

Candidate-specific stabilizer

- **Sliwa only:** gop_base$_i$ = Sliwa2021_share$_i$ with strong shrinkage (tiny ridge effect).

6 Estimation and Regularization

We fit ridge-penalized multinomial logits in each compartment (or a single model with EV/E-Day interactions if preferred operationally). Choose the ridge parameter λ via *spatial cross-validation* (folds grouped by AD or borough) to prevent leakage from neighboring EDs.

7 Poll Anchoring (Citywide Calibration)

Let $\bar{p}_k$ be the turnout-weighted citywide mean:

$$\bar{p}_k = \frac{\sum_i T_i\, p_{ik}}{\sum_i T_i}.$$

We apply small intercept shifts to $\gamma_k^{(\mathrm{EV})}$ and/or $\gamma_k^{(\mathrm{R})}$ (or a shared γ_k) so that $\bar{p}_k$ matches the latest polling averages.

Soft Sliwa prior (count-to-share): historical $\approx$ 316k votes implies $\sim 16\%{-}18\%$ share when $T \in [1.80, 1.94]$M; we encode this as a weak citywide penalty on $\bar{p}_S$ (polls remain the primary anchor).

8 Uncertainty and Simulation

We quantify uncertainty via:

- Bootstrap over EDs (resampling rows).

- A shared citywide shock added to linear predictors to induce correlation across EDs.

We run simulations across the grid $(T, \rho) \in \{1.80, 1.94\} \times \{0.4, 0.6\}$ and report 50% and 90% intervals for ED/AD/borough/city aggregates.

9 Data Inputs (Outstanding)

1. Democratic primary by ED (votes and shares).

2. EV 2025 by site/day (allocated to ED) and EV 2021.

3. Registration: current totals; new registrations 2024–25; party mix if available.

4. ACS: age 18–29; student share or near-campus flag; income variable.

5. Sliwa 2021 share by ED (or GOP registration share).

6. Numeric polling averages for Mamdani, Cuomo, Sliwa (used in calibration).

10   Implementation Plan (Order of Operations)

1. Build the ED design matrix: offsets o_{ik} and features $\mathbf{x}_i^{(\mathrm{EV})}, \mathbf{x}_i^{(\mathrm{R})}$; standardize features.

2. Compute $T_i, w_i^{\mathrm{EV}}, w_i^{\mathrm{R}}$ for each (T, ρ) scenario.

3. Fit compartment models with ridge; select λ via spatial CV.

4. Poll-align intercepts using turnout-weighted citywide means; include weak Sliwa share prior.

5. Simulate uncertainty across (T, ρ); export ED/AD/borough/city shares with 50/90% bands; render maps and a short methods note.

11   Summary

This specification combines (i) fresh *pattern priors* (Dem primary and a weak GOP map), (ii) composition-aware compartment weights derived from observed early voting and conservative E-Day scaling, and (iii) citywide calibration to current polling. The result is an interpretable, stable forecast of *vote shares* that updates coherently as early-vote and registration information evolves.

Update: Integration Status & Next Steps (November 4, 2025)

What We Have Accomplished

- **Pipeline built and validated end-to-end.** `build_base.py`, `build_turnout.py`, `build_features.py`, and `fit_compartment.py` run cleanly with key parity.

- **Turnout grid ready.** Long-form scenarios $(T, \rho) \in \{1.80, 1.94\} \times \{0.4, 0.6\}$ with sane T_i, $w^{\mathrm{EV}} \in [0, 1]$, $w^{\mathrm{R}} = 1 - w^{\mathrm{EV}}$; city totals match targets.

- **Offsets finite everywhere.** `pri_mam_share` and `sliwa21_share_eb` are clipped before logit; no $\pm\infty$.

- **Forecast artifacts written.** ED/AD/borough/city CSVs per scenario with probabilities summing to 1. (`forecast_ed_by_scenario.csv`, `forecast_ad_by_scenario.csv`, `forecast_borough_by_scenario.csv`, `forecast_city_by_scenario.csv`)

- **Diagnostics pass.** `debug_forecast.py` confirms weights, bounds, and aggregation sanity; citywide calibration equals polling means as designed.

Final Remaining Piece (Visualization & Publish)

Goal: Pipe ED-level results into an interactive map like the *early voting* example and publish to Substack.

Map spec.

- **Color channels (per-candidate layers):** Mamdani $\Rightarrow$ MamdaniBlue, Cuomo $\Rightarrow$ CuomoOrange, Sliwa $\Rightarrow$ SliwaRed.

- **Shade by share** p_{ik} with perceptual lightness (0%$\rightarrow$very light; 50%$\rightarrow$mid; 100%$\rightarrow$full).

- **Per-scenario toggle** (`scenario_tag`); hover tooltip shows `borough, AD, ED`, $p_M, p_C, p_S, T_i, w^{\text{EV}}, w^{\text{R}}$.

- **Static export for Substack:** PNG (2x retina) per candidate + a stacked triptych; captions include scenario label and citywide shares.

Minimal implementation sketch.

1. Join `forecast_ed_by_scenario.csv` to the ED GeoJSON (key: `ed_uid`).

2. Build three style functions (one per candidate) mapping p_{ik} $\rightarrow$fill color using the palette above.

3. Add a scenario dropdown; re-style the layer on change (no re-fetch).

4. Export static PNGs (Leaflet/Maplibre screenshot or browser print-to-PNG) for Substack.
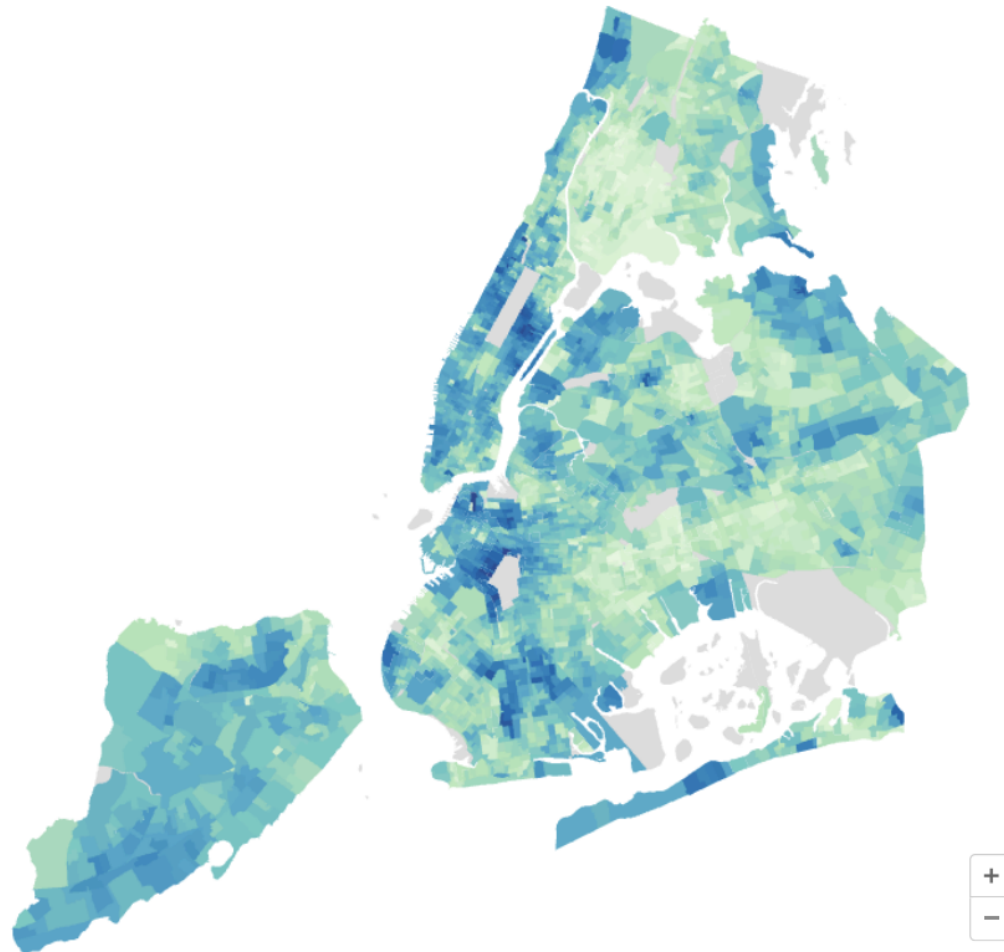
Where New Yorkers Voted Early

More than 732,000 New Yorkers cast their vote over the nine-day early voting period that ended Sunday.

See Neighborhood Summary →

Pct. of Registered Voters Who Voted Early

0%   42%

Map shows only election districts where more than 5 people voted early.
Map: Haidee Chu / THE CITY · Source: NYC Board of Elections, Department of City Planning

THE CITY

Figure 1: Early-vote example map style. Our forecast map will mirror this look-and-feel, with candidate layers colored blue (Mamdani), orange (Cuomo), and red (Sliwa).

Short Checklist to Ship the Map

- [ ] Load ED GeoJSON and left-join `forecast_ed_by_scenario.csv` by `ed_uid`.

- [ ] Implement color scales: MamdaniBlue, CuomoOrange, SliwaRed; legend with 0–42% style ticks to echo the reference.

- [ ] Add scenario selector and hover tooltips (ED keys + $p_M, p_C, p_S, T_i, w^{\text{EV}}$).

- [ ] Export static PNGs for Substack (one per candidate + triptych).