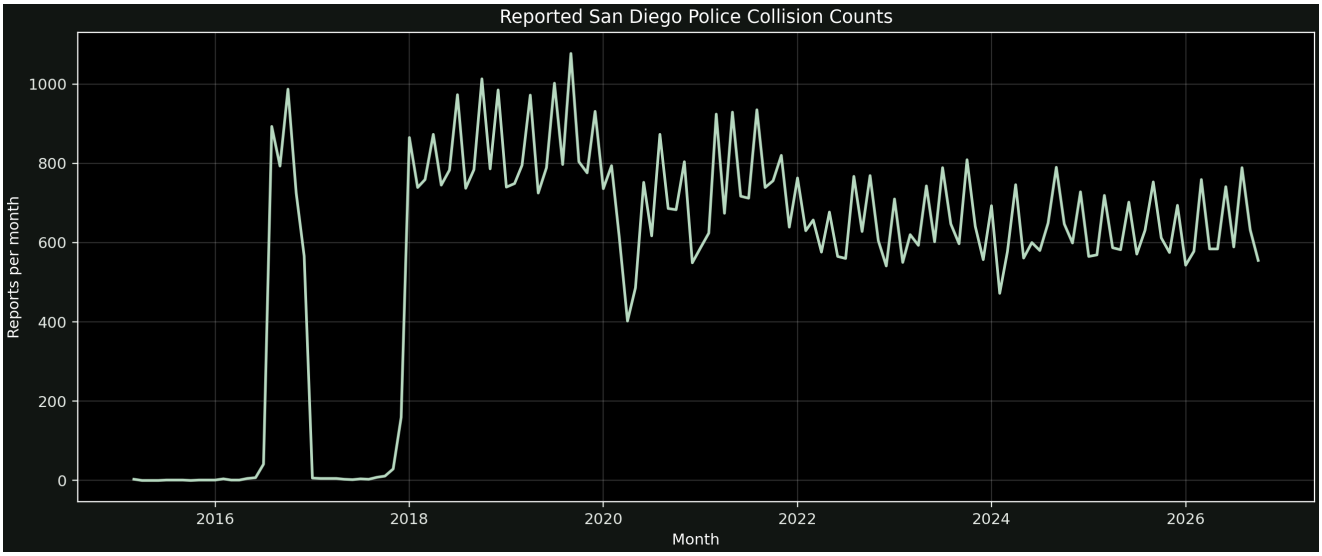
CRASH TEST

Spatiotemporal Traffic Collision Risk Forecasting

APPLIED URBAN SAFETY ANALYTICS / RESEARCH RELEASE 1.0

01 / Research overview

CRASH TEST is a reproducible platform for weekly collision-risk forecasting across San Diego Police beats. The source snapshot contains 78,715 unique reports across 148 beats. It studies aggregate risk ranking, not individual crashes or exact crash locations.



02 / Problem definition

The unit is a police beat and Monday-starting week. Primary target: next-week report count. Secondary target: next-week occurrence. The ranking objective identifies comparatively elevated aggregate risk, not causal danger or an operational allocation recommendation.

03 / Data sources and preprocessing

The City of San Diego Police basic-report source supplies report time, police beat, injury, fatality, and hit-and-run fields. It is updated daily, excludes many property-damage-only events, and does not cover freeway crashes handled by CHP. Duplicate IDs and invalid dates/beats are removed; checksum, schema, coverage, and missingness are saved on each run.

04 / Feature engineering

Lagged counts at 1, 2, 4, 8, 12, and 52 weeks, shifted trailing count and severity summaries, beat identity, and cyclical seasonality form the feature set. Every historical feature is shifted before the prediction week, preventing target leakage.

05 / Forecasting methodology

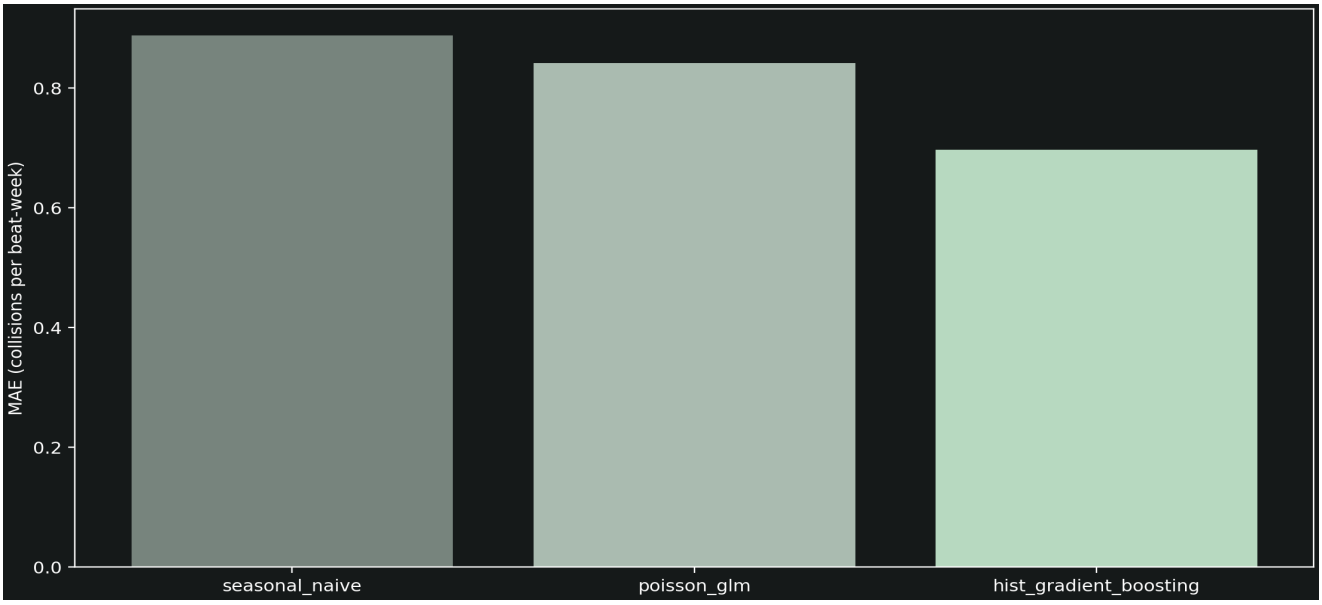
Seasonal-naive and Poisson regression are compared with Poisson histogram gradient boosting. A distinct occurrence classifier is calibrated with logistic calibration. No deep-learning method is used because the available aggregate evidence does not justify its additional complexity.

06 / Model training and evaluation

Training ends in 2023, validation is 2024, and 2025 is untouched test data. Two rolling origins evaluate 2023 and 2024 after their own training cutoffs. A deterministic disjoint police-beat test estimates transfer to unseen administrative areas.

07 / Results

On 7,696 held-out 2025 beat-weeks, gradient boosting achieved MAE 0.697, compared with 0.842 for Poisson regression and 0.888 for seasonal naive. Top-decile ranking lift was 3.587.



08 / Prediction uncertainty

The occurrence classifier is calibrated on 2024 only and evaluated in 2025: average precision 0.814, Brier score 0.164. Bootstrap MAE improvement over seasonal naive is 0.190; it does not guarantee behavior under distribution shift.

09 / Model interpretation

Held-out permutation importance measures the count-error change when a feature is shuffled. It is predictive diagnostics, not causal attribution; correlated time-history variables can share or conceal apparent importance.

10 / Limitations

A disjoint evaluation of 30 held-out beats yielded MAE 0.639. Reporting bias, property-damage-only omissions, CHP freeway coverage, administrative aggregation, absent verified point coordinates, uncertain exposure, and temporal/geographic shift limit generalization and operational use.

11 / Conclusions

Leakage-safe historical features improve aggregate 2025 beat-week count forecasting relative to the evaluated seasonal baseline. The study does not demonstrate point-level prediction, causal drivers, or readiness for decisions about people or communities.

12 / Reproducibility

Download official source CSVs, install dependencies, run the baseline and expanded scripts, then serve the read-only API and dashboard. The pipeline emits provenance checksums, processed panels, saved predictions, metrics, and figures. Raw and row-level data remain excluded from Git.

References

City of San Diego, Traffic collisions - basic reports: https://data.sandiego.gov/datasets/traffic-collisions/
City of San Diego, Traffic collision details: https://data.sandiego.gov/datasets/traffic-collision-details/
City of San Diego, Traffic ADT counts: https://data.sandiego.gov/datasets/traffic-adt-counts/