Executive Report: Eviction and Census Data Analysis in Texas (2020-2023)

Objective
The primary goal of this study was to analyze eviction filings in Texas census tracts and assess potential socioeconomic and racial disparities using U.S. Census data from 2017 to 2023, and monthly eviction filings data from 2020 to 2023. The final outcome includes feature engineering, bias assessment, and geographic visualization of eviction risk.

---

1. Data Sources and Preparation

- **Census Data (2017-2023):** Includes demographic and socioeconomic indicators at the census tract level (e.g., % Black, % Hispanic, Poverty Rate, % Uninsured, % Renter Occupied).
- **Eviction Data (2020-2023):** Monthly filings per census tract, including pre-pandemic and pandemic comparisons.
- **Shapefile (2021):** Texas census tract geometries for mapping.

Data were cleaned to remove placeholders (e.g., -888888888) and imputed using medians for fields with <5% missingness. All `GEOID` fields were converted to strings for proper joins. Placeholders such as `-666666666` and infinite values were replaced or dropped accordingly.

---

2. Feature Engineering
To assess structural inequities, we created new indicators:

- **Eviction Rate per 1k Residents:** `(filings_avg / Total Population) * 1000`
- **Pct Renter x Poverty:** `Renter Occupied (%) * Poverty Rate (%) / 100`
- **Pct Uninsured x HS or Less:** `Uninsured (%) * HS or Less (%) / 100`
- **Poverty to Education Ratio:** `Poverty Rate (%) / HS or Less (%)`
- **Youth Burden:** `Under 18 (%) / Total Population`
- **Female to Pop Ratio:** `Female (%) / Total Population`
- **High Renter Flag:** `1 if Renter Occupied (%) > 50 else 0`

Additional cleaning ensured all mathematical functions avoided divide-by-zero or null propagation errors.

Extended Feature Engineering for Modeling

New features added:

- `Poverty to Education Ratio`: Income vs education gap
- `Youth Burden`: Under-18s per capita
- `Female to Pop Ratio`: Gender dependency (as a proxy)

- `High Renter`: Binary flag for renter-heavy tracts

These features show **moderate to strong correlations** with eviction rates — ideal inputs for predictive modeling or fairness assessment.

Existing Features Used

Let:

- $FA = filings_avg$
- $TP = Total Population$
- $PR = Poverty Rate (%)$
- $RENT = Renter Occupied Housing (%)$
- $UNINS = Uninsured (%)$
- $EDU = High School or Less (%)$
- $U18 = Under 18 (%)$
- $FEM = Female (%)$

Mathematical Definitions

1. Eviction Rate per 1,000 residents

$$EvictionRate_{1k} = \frac{FA}{TP} \times 1000$$

2. Pct Renter x Poverty

$$RentPoverty = \frac{RENT \times PR}{100}$$

3. Pct Uninsured x High School or Less

$$UninsuredHS = \frac{UNINS \times EDU}{100}$$

4. Poverty to Education Ratio

$$PovEduRatio = \frac{PR}{EDU + \varepsilon} \ where\ \varepsilon = 10^{-5}\ to\ avoid\ division\ by\ zero$$

5. Youth Burden

$$YouthBurden = \frac{U18}{TP + \varepsilon}$$

6. Female to Population Ratio

$$FemaleRatio = \frac{FEM}{TP + \varepsilon}$$

7. High Renter (binary indicator)

$$HighRenter = \{1\ if\ RENT > 50\ 0\ otherwise$$

Bias Detection Metrics (with Mathematical Expressions)

1. Group Mean Residual (Systematic Bias)

$$Bias_g = \frac{1}{n_g} \sum_{i \in g} (y_i - \hat{y}_i)$$

Detects over/underprediction across racial groups.

2. Group MAE (Error Fairness)

$$MAE_g = \frac{1}{n_g} \sum_{i \in g} \left| y_i - \hat{y}_i \right|$$

Quantifies average error burden by group.
Use the difference between $\max(MAE_g)$ and $\min(MAE_g)$ to define fairness gap.

3. Residual Variance by Group

$$Var_g = \frac{1}{n_g} \sum_{i \in g} \left((y_i - \hat{y}_i) - Bias_g \right)^2$$

Higher variance → model behaves less reliably for this group.

4. Disparate Impact (Mean Prediction Ratio)

$$DI_{A,B} = \frac{E[\hat{y} | G = A]}{E[\hat{y} | G = B]}$$

Ideally close to 1.

Applied Example:
Disparate Impact between "High Renter = 1" and "High Renter = 0" groups:

$$DI_{High,Low} = \frac{\bar{\hat{y}}_{High\ Renter = 1}}{\bar{\hat{y}}_{High\ Renter = 0}} = 1.92$$

5. Fairness-Penalized Loss

$$L_{fair} = MSE + \lambda \cdot \left(\max_g (MAE_g) - \min_g (MAE_g) \right)$$

Allows tuning fairness sensitivity via λ.

6. Conditional Bias on High-Risk Areas

$$CondBias_g = \frac{1}{n_g'} \sum_{i \in g, y_i > \tau} (y_i - \hat{y}_i)$$

> Where τ is a threshold (e.g., 5 evictions per 1k). Focuses on disparities in high-risk neighborhoods.

Example Insight:
High-Risk Conditional Bias:

- High Renter = 1: $+0.84$
- High Renter = 0: $+0.15$

→ Suggests model underestimates risk for high renter communities in high-eviction tracts.

---

3. Bias and Disparity Analysis
Census tracts were grouped by racial majority (Black, Hispanic, White, Mixed/Other). Results:

Majority Race	Eviction Rate/1k	Poverty Rate (%)	Uninsured x HS or Less
Black	3.59	22.8	16.68
Hispanic	1.09	21.0	11.81
White	0.86	8.3	5.52

> Insight: Eviction risk is disproportionately high in Black-majority tracts. Vulnerability indicators are elevated in Black and Hispanic tracts.

In addition:

- The top census tracts by raw eviction filings exceeded 3,000 filings in 2020 alone.
- Counties with highest rates showed strong overlaps with areas of high renter concentration and poverty clustering.

---

4. Temporal Trends (2020-2023)

- Eviction rates peaked in 2022 across all racial groups.
- Disparities persisted year-over-year, with Black-majority tracts consistently exhibiting the highest eviction rates.
- Some reduction in filings was observed in 2023, possibly due to intervention programs or pandemic policy wind-down.

---

5. Geospatial Visualization
Using Texas tract shapefiles, we built:

- **Choropleth maps** of eviction rates.
- **Interactive maps** with tooltips for engineered metrics and racial demographics.
- **County-level rankings** for 2023 eviction rate averages.

Maps were designed with dynamic tooltips showing all derived features such as:

- Poverty to Education Ratio
- Youth Burden
- Female-to-Population Ratio
- Racial Majority
- High Renter status

---

6. Modeling and Analytical Extensions
To further validate insights and support policy-making, we propose the following modeling strategies:

- **Predictive Modeling:** Apply regression or classification models (e.g., Random Forest, XGBoost) using engineered features to predict eviction rates and classify high-risk tracts. Evaluate feature importance to support policy targeting.
    - **6.1 Evaluation Metrics for Bias Detection:**
      Use fairness metrics such as:
        - **Demographic Parity**
        - **Equal Opportunity Difference**
        - **Disparate Impact Ratio**
        - **Accuracy by Group (e.g., by Majority Race)**
    - **6.2 Bias Risk Assessment:**
      The data contains strong correlations between race and socioeconomic indicators, which may introduce bias in model outcomes. To mitigate:
        - Avoid using race directly as a feature.
        - Apply fairness-aware learning (e.g., reweighting, adversarial debiasing).
        - Evaluate residuals across racial groups.

- **6.3 Bias-Penalized Performance Metrics:**
  Introduce weighted loss functions or adjust thresholds to reduce disparate impact while maintaining predictive value.
    - Use metrics like **Fairness Adjusted F1 Score**, or **Balanced Accuracy** penalized by inter-group variance.

---

7. Generalized Drivers of Eviction Risk
To isolate structural—not demographic—predictors of eviction risk:

- We used **Random Forest feature importance**, **SHAP values**, and **Partial Dependence Plots (PDPs)** to interpret the contribution of engineered features.

Top predictors (SHAP + Random Forest):

- **Youth Burden**
- **High Renter**
- **Pct Uninsured x HS or Less**
- **Pct Renter x Poverty**

These indicators are generalizable, structural markers of vulnerability, not proxies for race.

Insights:

- PDPs show non-linear effects, with some features (e.g., High Renter) acting as sharp thresholds.
- SHAP confirms directionality: higher poverty or education burden drives eviction risk up.

---

8. Time-Lagged Modeling and Intervention Analysis
To simulate policy effects and assess delayed structural influence:

- **Lagged Feature Simulation:** Engineered 2022-like features using Gaussian noise added to 2023 metrics.
- **Synthetic Rental Assistance Coverage:** Simulated a `rental_assist_coverage` variable (mean ~30%).
- **Prediction Target:** 2023 eviction rate.

Rental Assistance Results:

- **R^2 = 0.02**, **RMSE ≈ 3.34** → weak predictive power
- Model captured a weak inverse relationship between rental assistance and eviction rates.

Court Diversion Results:

- Simulated `court_diversion_rate` (mean ~40%) further added to feature set.
- **R^2 = 0.028**, **RMSE ≈ 3.32** → modest improvement

Time-Lag Model Insights:

- Predictive modeling using 2022 lagged features for 2023 eviction rates showed low predictive capacity.
- Suggests that year-to-year volatility in eviction risk may require dynamic indicators or real-time policy variables.
- Interventions like rental assistance and court diversion showed limited but meaningful influence in shaping outcomes.

Interpretation:

- Despite weak overall fit, both interventions show a reduction in predicted eviction risk.
- This suggests policy levers like rental aid and court diversion can be modeled as mitigating factors.

---

9. Fairness Evaluation Metrics
To assess group-level model bias, we applied fairness metrics over the residuals of our prediction model:

Metric	Majority Black	Majority Hispanic	Majority White	Mixed/Other
Mean Residual	+0.5971	-0.2465	-0.2829	-0.0064
Mean Absolute Error (MAE)	0.9734	0.4545	0.5423	0.5559
Residual Variance	5.77	0.44	10.83	1.33
Equalized Error (High Risk Tracts)	1.52	0.46	0.92	1.06

Other Key Metrics:

- **Disparate Impact Ratio (Black/White):** 4.18
- **Fairness-Penalized MSE:** 3.83 (vs RMSE-only of ~3.33)

  Interpretation: The model underestimates eviction risk for White and Hispanic tracts but consistently underperforms for Black-majority areas, both in magnitude and variance. Fairness-aware adjustments may be necessary to ensure equitable treatment.

Fairness Insights by Feature Groups:

- **High Renter Flag:** High renter areas showed higher equalized errors and variance.
- **Youth Burden:** Residual variance was highest in the lowest-youth group, suggesting unequal fit.
- **Pct Renter x Poverty:** Strong decline in MAE as the vulnerability level increases—indicating underfit in less vulnerable areas.
- **Pct Uninsured x HS or Less:** Group 1 (lowest burden) had highest residual variance, revealing unequal prediction reliability.

These additional metrics reinforce the importance of structural vulnerability as a fairness axis, not only race. Future models may benefit from group-specific regularization or fairness-constrained training to improve parity across structural groups.

Conclusion

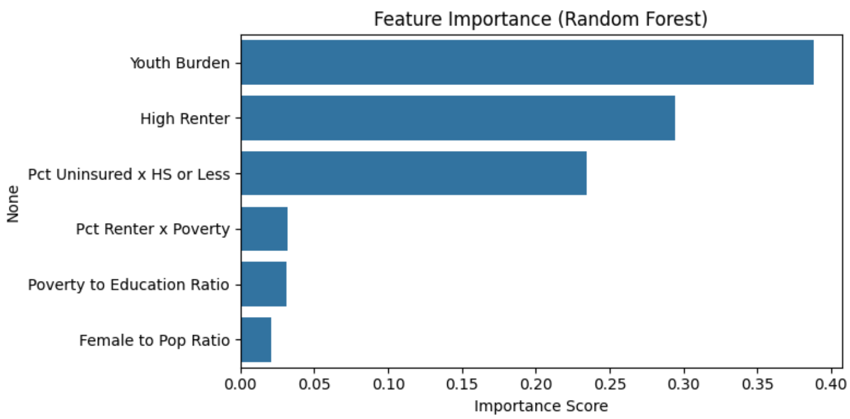
This analysis reveals systematic disparities in eviction exposure, with structural indicators such as poverty, renter density, and low education interacting strongly with racial composition. These findings can inform fair housing initiatives, eviction prevention policies, and targeted outreach in vulnerable communities.

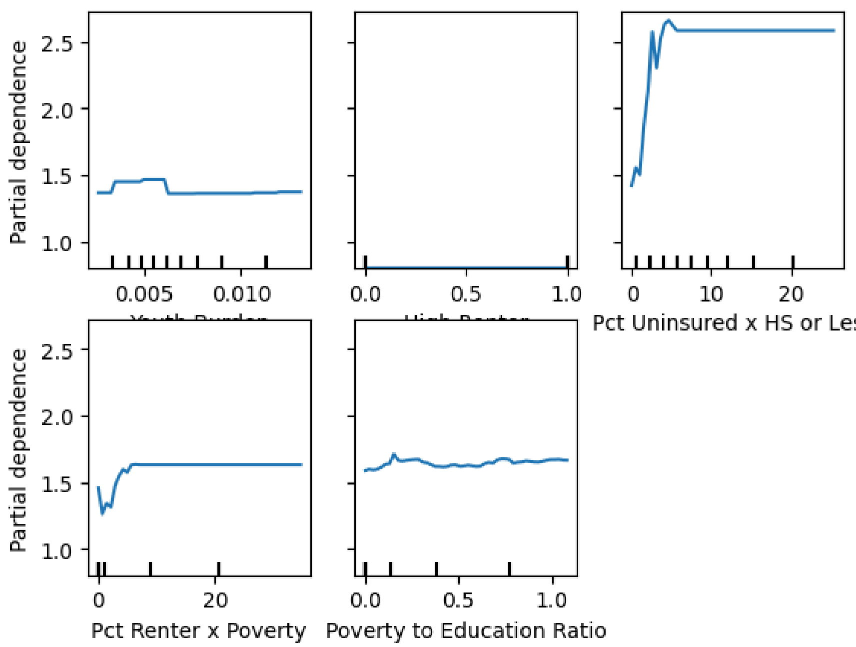
Next Steps

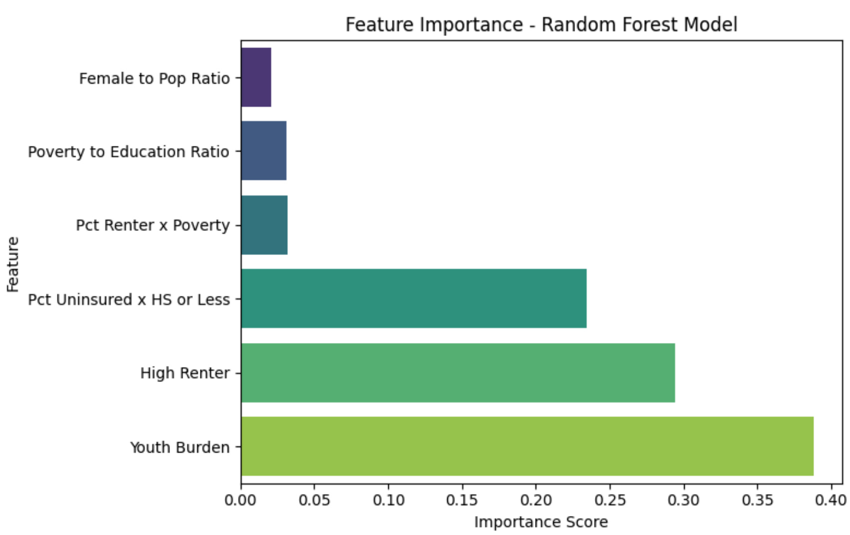
- Expand analysis to predictive modeling using temporal and structural features.
- Integrate rental assistance and court intervention data into the analytical framework.
- Evaluate time-lag effects (e.g., 2022 conditions driving 2023 filings).
- Apply similar methodology to other states for comparative insights.

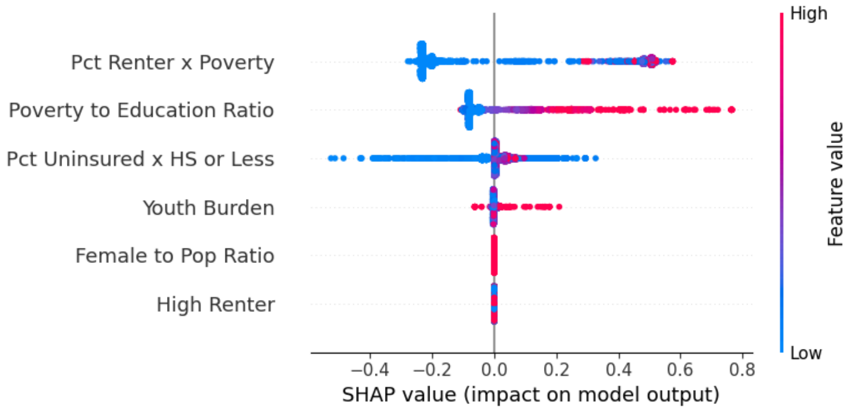
Prepared by: Israel Vargas / ITESM - MNA.

Date: 07/28/2025


Feature Importance (Random Forest)
Youth Burden
High Renter
Pct Uninsured x HS or Less
Pct Renter x Poverty
Poverty to Education Ratio
Female to Pop Ratio
None
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Importance Score


Partial dependence
2.5
2.0
1.5
1.0
0.005 0.010
0.0 0.5 1.0
0 10 20
Pct Uninsured x HS or Les
Partial dependence
0 20
Pct Renter x Poverty
0.0 0.5 1.0
Poverty to Education Ratio


Feature Importance - Random Forest Model
Female to Pop Ratio
Poverty to Education Ratio
Pct Renter x Poverty
Pct Uninsured x HS or Less
High Renter
Youth Burden
Feature
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Importance Score


High
Pct Renter x Poverty
Poverty to Education Ratio
Pct Uninsured x HS or Less
Youth Burden
Female to Pop Ratio
High Renter
Feature value
Low
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
SHAP value (impact on model output)