Understanding the Drivers of Green Building Certification in the United States

A Longitudinal Analysis of Incentives, Demographics, and Political Context (2000–2024)

The Rhizome Institute

Nico Taber, Nicholas Chrapliwy, Dr. William Barber

1. Introduction

In the face of escalating climate challenges, energy consumption in the built environment has emerged as a critical point of intervention. Buildings account for nearly 40% of total energy use and over 30% of greenhouse gas emissions in the United States, making them a key target for decarbonization strategies. In response, federal and state governments have invested heavily in programs designed to reduce building energy use, promote energy efficiency, and shift toward more sustainable infrastructure. Among the most visible and accessible indicators of these efforts are green building certifications such as ENERGY STAR and LEED.

Despite the availability of certification pathways, adoption of green building standards remains uneven across geography and time. Some states consistently certify large numbers of buildings, while others lag behind. Understanding what drives this variation is central to evaluating the effectiveness of current policy tools and identifying opportunities for more equitable and efficient deployment of sustainability programs.

One of the primary levers for encouraging green construction and retrofitting is the use of financial incentives. These incentives, including rebates, tax credits, loans, and grants, are designed to lower the cost barrier to energy-efficient upgrades. The Database of State Incentives for Renewables and Efficiency (DSIRE) catalogues such programs across all 50 states, offering a comprehensive view of the policy landscape. While many studies have examined the technical or economic impact of individual incentive programs, fewer have assessed their cumulative, long-term influence on state-level certification outcomes.

Even fewer have considered how non-policy variables, such as demographic structure or partisan control of government, influence energy-related decision-making at scale. Political alignment may affect not only whether programs are enacted, but also how well they are implemented and received. Demographic factors may influence demand for certification differently across population segments, raising questions about equity in the distribution of green building benefits.

This study addresses these gaps by combining more than two decades of data on ENERGY STAR building certifications, state-level clean energy incentive programs, U.S. Census demographics, and political control of state governments. By constructing a harmonized state-year dataset and applying a range of statistical and machine learning techniques, we explore:

- What demographic and political features are associated with higher certification activity?
- How incentive availability relates to certification trends, both in the same year and with delayed effects.
- Whether predictive modeling can identify the structural or policy conditions under which certification is most likely to succeed.

- How state-level responsiveness to incentives varies across time and geography.

We integrate multiple data sources, including the DSIRE API, ENERGY STAR building registry, state partisanship records, and harmonized census estimates from 2000 to 2024. Our analysis includes exploratory data visualization, ridge regression, principal component analysis (PCA), random forest classification, and geospatial mapping. Through this approach, we aim to uncover not only statistical patterns, but also the structural context that shapes how and where green building adoption occurs across the United States.

2. Data and Preprocessing

This analysis combines four major data sources to build a panel dataset that captures green building activity, financial incentives, political context, and demographic structure across U.S. states from 2000 to 2024. Each dataset contributes a different dimension to the final merged table, which forms the foundation for all subsequent modeling and analysis.

2.1 Datasets Used

ENERGY STAR Certified Buildings
We use publicly available data from the ENERGY STAR program, which includes certified commercial and industrial buildings nationwide. Each record contains a state identifier and one or more certification years. To prepare this data for state-level analysis, we clean and explode multi-year certifications into separate rows, then aggregate the data by state and year to produce a count of certified buildings (`CertifiedCount`).

DSIRE Incentive Programs
Policy data is sourced from the Database of State Incentives for Renewables and Efficiency (DSIRE), accessed via API. This dataset includes information on individual incentive programs such as rebates, tax credits, loans, and grants. Each program includes metadata like the implementing sector, technology focus, start and end dates, and geographic scope. We extract the program start year and aggregate new programs by state-year to compute `NumIncentives`.

State Political Composition
We use historical records of gubernatorial and state legislative party control to track political context. These records are expanded into annual state-level rows using officials' term start and end dates. From this, we construct binary variables such as `GovernorParty`, `HouseControl`, `SenateControl`, and a `Trifecta` flag indicating whether a single party held control of the executive and both legislative chambers.

Demographic Data
We incorporate state-level demographic estimates from the U.S. Census Bureau covering race, age, and sex from 2000 to 2024. Because formats differ across years, we harmonize columns into a consistent schema, remapping categorical labels where needed. Raw population counts are converted into percentages for comparability across states and time periods.

2.2 Preprocessing Steps

To create a unified dataset suitable for statistical modeling, we perform the following preprocessing operations:

- **Harmonization of Demographic Formats**
  Demographic data is normalized by calculating percentages for race, age brackets, and sex within each state-year. This ensures comparability across states of different sizes and across changing census definitions.

- **State-Year Alignment Across Sources**
  All datasets are transformed to a common state-year format. ENERGY STAR and DSIRE program data are timestamped and aggregated by year. Political control records are expanded across years of tenure. We ensure consistent naming conventions and standardize state abbreviations across files.

- **Feature Engineering and Variable Construction**
  We generate several composite variables that support exploratory and predictive modeling:

  - `CertifiedCount`: Number of ENERGY STAR certifications in a given state-year
  - `NumIncentives`: Number of new incentive programs introduced in that year
  - `Lag1_NumIncentives`, `Lag2_NumIncentives`: Number of new incentive programs one and two years prior
  - `Trifecta`: Binary indicator of unified party control across governor, house, and senate
  - `HighActivity`: Binary label for classification tasks, indicating whether a state-year's certification count is above the national median for that year

- **Spike Detection for Lag Analysis**
  To explore how states respond to sudden increases in incentive availability, we flag

"spike" years in which the number of new programs exceeds one standard deviation above the state's historical average. We then track whether these spikes are followed by upticks in certification activity in the same year or one to two years later.

- **Data Cleaning and Filtering**
  We exclude state-year rows with incomplete or unreliable values, especially where key outcomes like `CertifiedCount` are missing or where demographic harmonization fails. Missing values in predictors are filled, zeroed, or excluded depending on their context and significance.

This preprocessing pipeline yields a clean, merged panel dataset with one row per state-year from 2000 to 2024. Each row includes demographic composition, political control, policy activity, and green building certification outcomes, enabling robust analysis of both structural trends and dynamic policy effects.

3. Exploratory Analysis

We begin by analyzing descriptive trends in green building certification activity and its relationship to policy incentives, demographic structure, and political control. This section draws on both visual and tabular summaries to surface patterns that motivate formal modeling.

3.1 Incentive Program Types and Distribution

An initial breakdown of the DSIRE dataset shows that some policy mechanisms dominate the incentive landscape. Rebate programs appear most frequently, followed by grants and tax incentives.

Incentive Type	Count
Rebate Program	1139
Grant Program	171
Loan Program	169
Energy Standards for Public Buildings	96
Property Tax Incentive	88
Building Energy Code	82
Net Metering	72
Solar/Wind Access Policy	69
Solar/Wind Permitting Standards	57
Sales Tax Incentive	54
PACE Financing	52
Interconnection	52
Renewables Portfolio Standard	49
Public Benefits Fund	37
Energy Efficiency Resource Standard	30

Table 1. *Top 15 most common state-level clean energy incentive types (2000–2024), based on data from the DSIRE database. Rebate programs are by far the most frequently implemented policy mechanism, followed by grants and loan programs. The complete distribution of all 44 incentive types is provided in Appendix A.*

Rebate Programs account for over 1,100 recorded instances between 2000 and 2024, far exceeding the next most common types - Grant Programs (171) and Loan Programs (169). This dominance likely reflects the administrative ease and public familiarity of rebates as a policy tool, especially for household-scale energy upgrades.

In contrast, mechanisms such as Green Building Incentives (18), Bond Programs (6), and Property Tax Assessments (1) occur infrequently and are concentrated in only a handful of states. Their limited presence may constrain their measurable impact in statistical models, though their effects could be locally significant.

This distribution suggests that not all incentive types are equally influential in driving national certification trends. The concentration of certain mechanisms in the dataset also informs feature weighting in later regression models.

3.2 Certification Trends Over Time

ENERGY STAR certification data were aggregated by state and year to form the outcome variable `CertifiedCount`. Plotting certification volume across time illustrates both national growth and state-level variation.

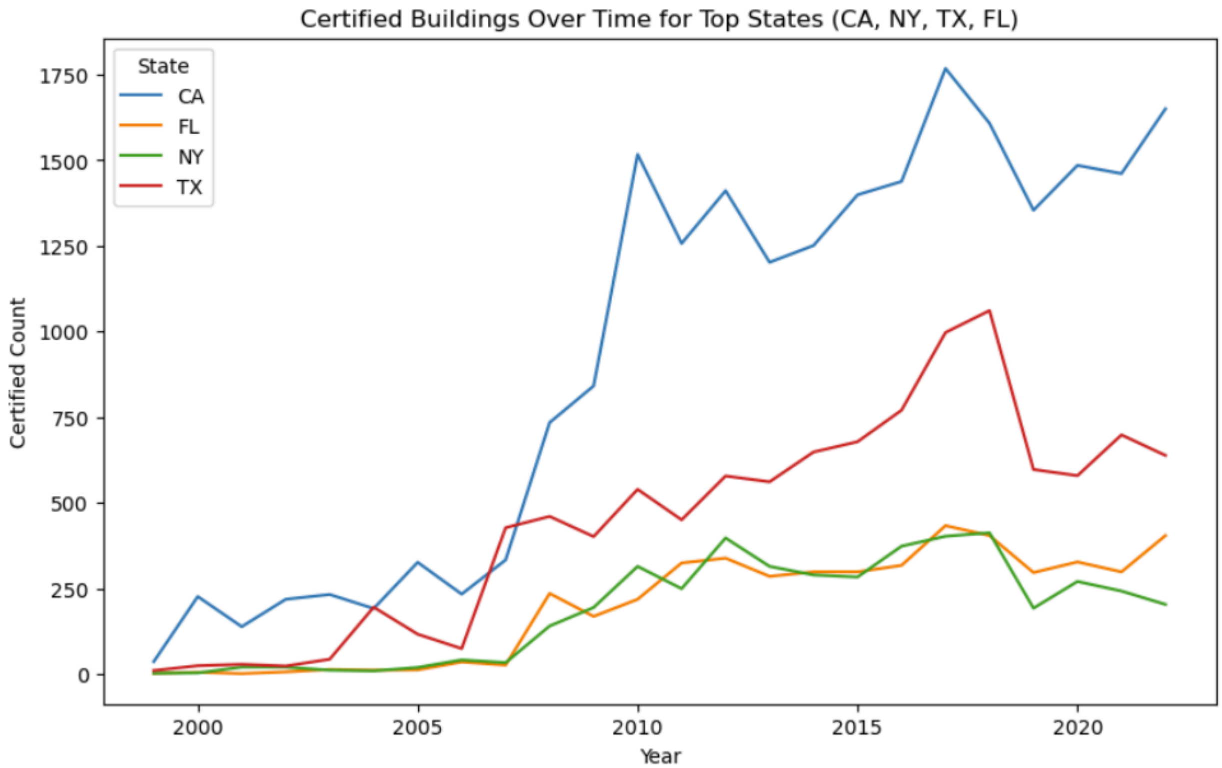


Figure 1. *Certified building counts over time for four leading states (California, New York, Texas, Florida), 1999–2023. California exhibits the highest certification volumes throughout most of the series, with particularly sharp growth from 2007 to 2012 and peaks exceeding 1,700 certified buildings annually. Texas follows as the second-highest producer, with rapid increases after 2005 and sustained volumes over 500 annually. New York and Florida remain lower overall, both peaking between 300 and 450 certifications per year but showing more year-to-year volatility. Data from ENERGY STAR certified building records.*

The time-series comparison of CA, NY, TX, and FL underscores the dominance of California in national certification activity. The steep upward trajectory beginning around 2007 coincides with both state-level energy efficiency mandates and the national ramp-up of ENERGY STAR initiatives. California's post-2010 peaks suggest a mature, sustained adoption pipeline, possibly linked to large commercial real estate markets and robust incentive programs.

Texas emerges as a notable second-tier leader, with growth patterns suggesting responsiveness to both federal incentives and state-specific efficiency programs implemented in the late 2000s. Its high sustained volume after 2010, despite some fluctuation, points to structural market adoption rather than temporary policy effects.

New York and Florida follow more moderate growth paths, though both exhibit notable surges between 2008 and 2017. These peaks align with national economic recovery trends and, for New York in particular, potentially with municipal-level policy pushes in New York City. The subsequent dips in both states could reflect a saturation of the most easily certifiable building stock or shifts in local policy priorities.

Overall, Figure 1 highlights that even among the top states, certification trajectories differ substantially in timing, magnitude, and stability, implying that local market size, policy environment, and existing building stock characteristics jointly shape long-term certification patterns.

3.3 Correlation Between Incentives and Certification Volume

To assess whether incentive activity correlates with certification outcomes, we plot the number of new programs in a given year (`NumIncentives`) against the number of certifications in that same year (`CertifiedCount`).

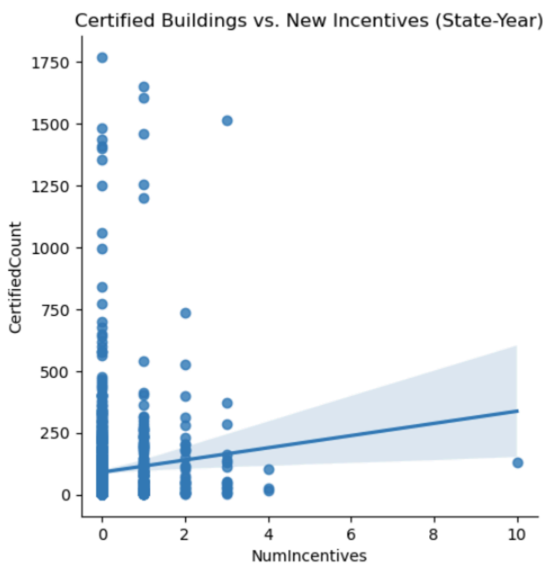


Figure 2. *Each point represents one state in one year; the solid line is an OLS fit with a 95% confidence band. A permutation Spearman test indicates a positive but modest monotonic association between the number of new incentive programs and certified buildings ($\rho = 0.168$, two-sided $p \approx 0.0001$).*

The scatter is highly concentrated at 0–2 new programs per year, which means most states add few programs in any given year. Within that low-incentive range the vertical spread in certifications is large, from near zero to more than 1,500, so incentives alone do not account for most of the year-to-year variation. The fitted line slopes upward and the permutation test confirms a statistically reliable relationship, but the effect size is small. A rank correlation of

0.168 implies roughly $\rho^2 \approx 0.03$, so on the order of three percent of the rank variation in certifications aligns with the rank of new programs.

Two practical takeaways follow. First, adding programs tends to move certifications in the desired direction, but the expected marginal gain from one more program is limited without other favorable conditions. Second, the wide intervals at higher program counts likely reflect sparse data, so uncertainty increases when only a few state-years introduce many programs. Coupled with the modeling results where demographic composition was among the most influential predictors, the figure supports a complementary story: incentives matter, yet their impact appears to depend on the state context and possibly on timing. This motivates lagged and fixed-effects specifications in Section 5 and supports the geographic comparisons in Section 7.

3.4 Spike Matching and Policy Responsiveness

To explore policy responsiveness, we define an “incentive spike” as any year in which a state offers more than one standard deviation above its average number of new programs. A “certification spike” is defined as a year with >10% growth in certifications compared to the prior year.

We track whether certification spikes occur in the same year or within two years after an incentive spike.

State	NumIncentiveSpikes	NumCertSpikes	Lag0_ResponseRate	Lag1_ResponseRate	Lag2_ResponseRate
OR	1	10	0.0	1.0	0.0
AZ	2	14	1.0	1.0	0.0
OH	2	16	1.0	1.0	1.0
CO	1	8	0.0	1.0	1.0
MS	2	9	0.5	1.0	0.5
DE	1	7	0.0	1.0	0.0
GA	1	12	1.0	1.0	1.0
ME	4	6	0.0	0.75	0.25
IL	3	9	0.6667	0.6667	0.3333
MD	3	13	0.3333	0.6667	0.3333
KY	3	10	0.6667	0.6667	0.6667
RI	3	7	0.3333	0.6667	0.3333
WA	5	13	0.8	0.6	0.6
VT	2	8	0.5	0.5	0.0
PA	2	10	0.5	0.5	0.0

Table 2. *Top 15 states by policy responsiveness, ranked by one-year lag match rate between*

incentive spikes and certification spikes. A certification spike is defined as >10% growth from the previous year, and an incentive spike is defined as a program count exceeding one standard deviation above a state's baseline. Higher values indicate stronger apparent responsiveness to incentive-driven policy shifts.

This analysis reveals:

- Several states demonstrate perfect or near-perfect one-year lag responsiveness. For example, Oregon (OR) and Arizona (AZ) respond to 100% of observed incentive spikes within one year, while Ohio (OH) and Colorado (CO) achieve the same rate and also maintain high responsiveness at two years (100%).
- Other states, such as Maine (ME) and Maryland (MD), show strong but not universal responsiveness (Lag1 rates of 75% and 66.7%, respectively), suggesting that other contextual factors may influence adoption speed.
- Conversely, states like Vermont (VT) and Pennsylvania (PA) exhibit moderate responsiveness (Lag1 rates of 50%) and no measurable reaction at a two-year lag, implying either structural barriers or less effective incentive deployment.
- These results confirm that policy impact is not uniform, and that incorporating lag variables into statistical models is essential to capture variation in policy uptake timelines.

The findings suggest that the impact of incentives is not instantaneous, reinforcing the importance of lagged predictors in regression and classification models.

3.5 Demographic and Political Patterns

We next examine whether certification trends differ across demographic or political contexts. The dataset includes state-level percentages for race, age group, and sex, along with political control variables (governor party, legislative control, and `Trifecta` alignment).

Although exploratory correlation plots are not shown, descriptive summaries suggest meaningful patterns.

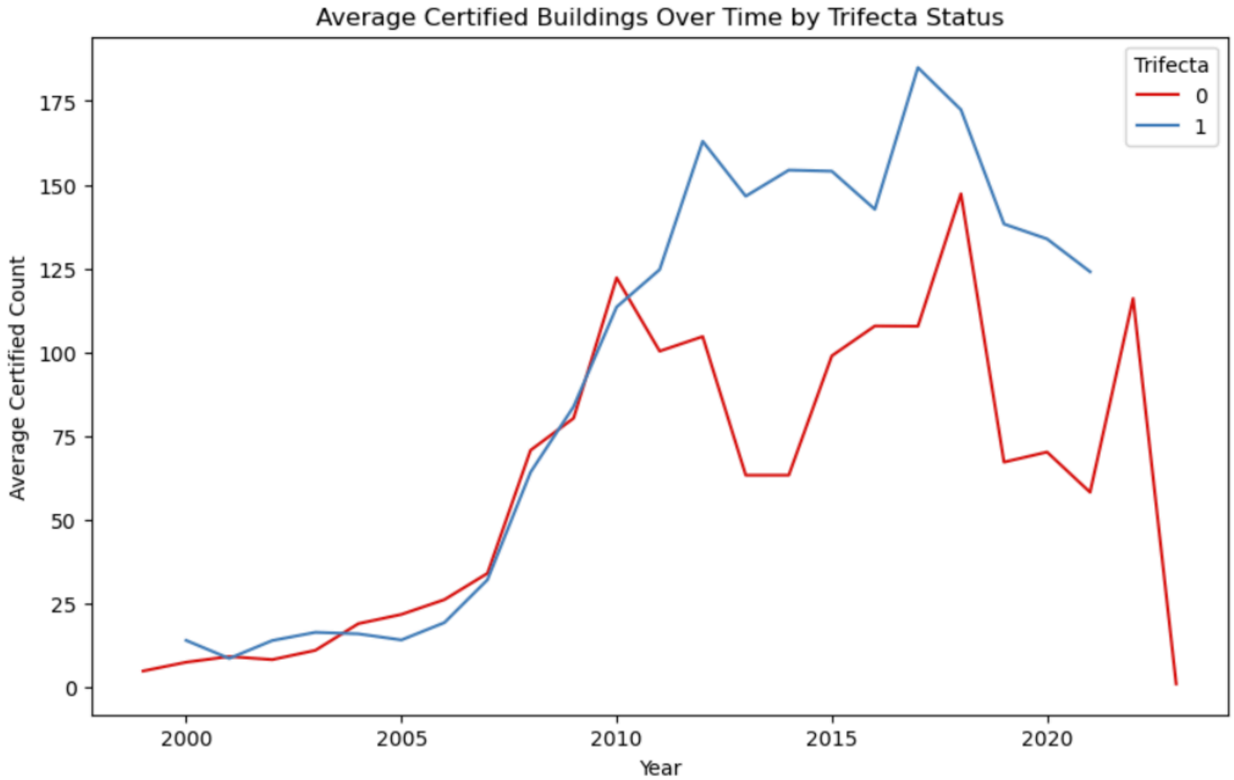


Figure 3. *Average number of certified buildings per state-year from 1999 to 2023, grouped by trifecta status. States with trifecta control consistently average higher certification volumes (103.7) compared to non-trifecta states (68.4).*

Results indicate:

- Figure 3 shows the average number of certified buildings per state-year, separated by trifecta status from 1999 to 2023. While both trifecta and non-trifecta states follow similar long-term trends, trifecta states consistently exhibit higher certification volumes after approximately 2010.
- Across all state-years in the dataset, states with trifecta control average 103.7 certifications per year compared to 68.4 for states without trifecta control. A two-sample t-test confirms that this difference is statistically significant ($t = 4.331$, $p < 0.001$), suggesting that unified political control may facilitate more effective policy implementation or sustained support for certification-related programs.
- While causality cannot be inferred, these results align with the hypothesis that political alignment between the executive and legislative branches is associated with greater green building activity.

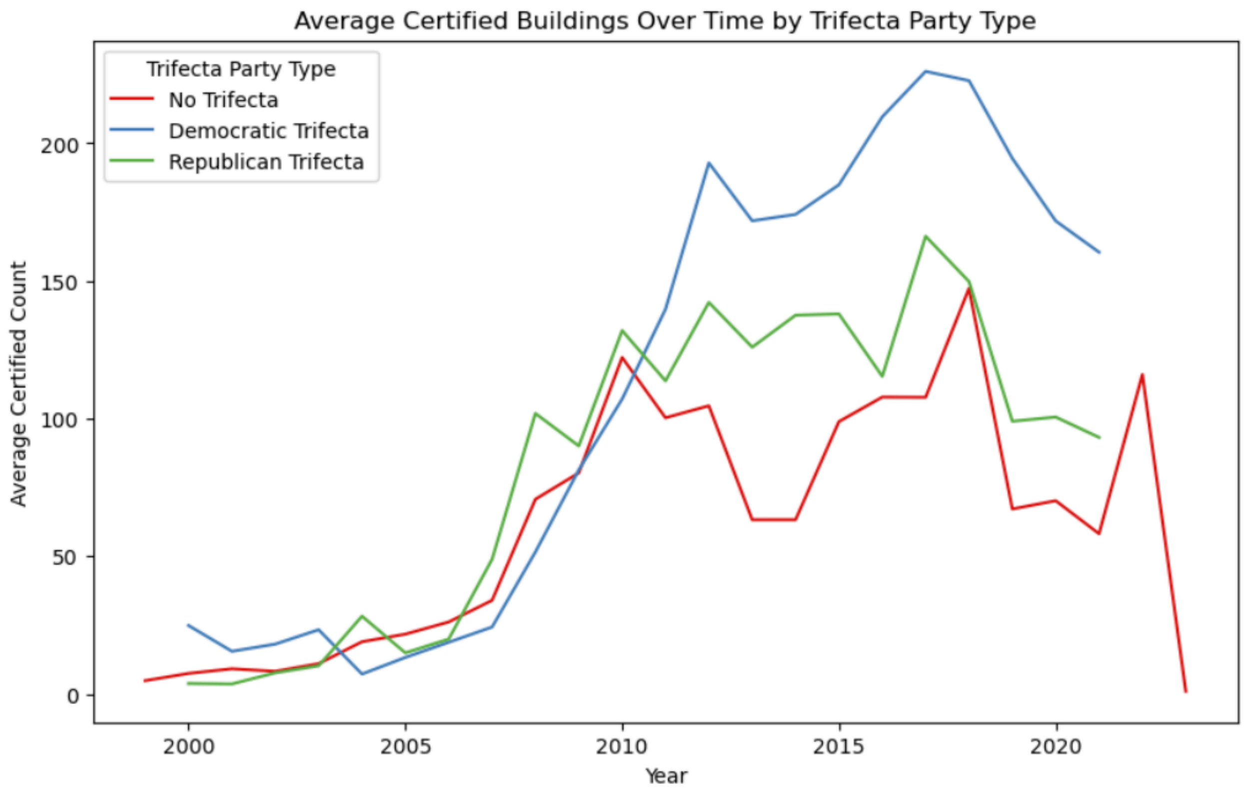


Figure 4. *Average certified buildings per state-year from 1999 to 2023, grouped by trifecta party type (Democratic, Republican, and No trifecta). While Democratic trifectas appear to maintain higher certification volumes than Republican trifectas after approximately 2010, a two-sample t-test comparing the two groups across all years did not yield a statistically significant difference ($p \approx 0.37$).*

We extended the political analysis by differentiating trifecta states by party affiliation. Figure 4 shows that Democratic trifectas and Republican trifectas followed similar trajectories until approximately 2010, after which Democratic trifectas tended to maintain higher average certification counts. However, a pooled two-sample t-test comparing Democratic and Republican trifectas across all years did not produce a statistically significant result ($p \approx 0.37$). While this finding suggests that party affiliation of trifecta states may not be a strong predictor of certification volume when aggregated over time, the visible post-2010 divergence warrants further exploration, potentially using time-specific statistical tests or models that account for policy and economic context in those years.

3.6 Summary

Rebate programs dominate the national incentive landscape, with over 1,100 recorded instances between 2000 and 2024, more than six times the frequency of the next most common types, Grant Programs (171) and Loan Programs (169). Other mechanisms such as Green Building Incentives, Bond Programs, and Property Tax Assessments are rare and geographically concentrated, suggesting their potential impact is localized rather than national.

Certification activity is unevenly distributed across states and years. California stands out with consistently high volumes and sharp growth from the mid-2000s to early 2010s, maintaining peaks above 1,700 annual certifications. Texas has also emerged as a high-output state since the late 2000s, while New York and Florida exhibit more moderate levels with notable peaks in the post-recession years. These differences in trajectory suggest that state-specific market size, building stock, and policy environment shape certification outcomes as much as incentive availability.

The relationship between new incentive programs and certification volume is positive but modest. A permutation Spearman test confirms statistical significance ($\rho = 0.168$, $p \approx 0.0001$), but the effect size is small, with most states adding only 0–2 programs per year and showing wide variation in certification counts within that range. This implies that while incentives tend to support higher certification counts, their impact is conditional on other factors such as demographics, policy history, and local market readiness.

Spike matching analysis further emphasizes variation in responsiveness. States like Oregon, Arizona, Ohio, and Colorado respond to 100% of observed incentive spikes within one year, whereas others, such as Vermont and Pennsylvania, show only moderate or delayed responses. This underscores the need to incorporate lagged incentive variables in statistical models to capture differences in adoption timelines.

Political context also appears to influence certification trends. States with unified political control (trifectas) average 103.7 certifications per year compared to 68.4 for non-trifecta states, a statistically significant gap ($p < 0.001$). However, when trifectas are separated by party, Democratic-controlled states tend to maintain higher certification volumes than Republican-controlled states after 2010, but the pooled difference is not statistically significant ($p \approx 0.37$). This suggests that while political alignment correlates with higher certification activity, party affiliation alone may not be a reliable predictor without accounting for other structural and policy factors.

4. Regression Analysis

4.1 Methodology

To quantify the relationship between incentive activity, demographic composition, political control, and certification outcomes, we fit a Ridge Regression model with `CertifiedCount` as the dependent variable. Ridge Regression was selected to address multicollinearity among predictors, a common issue in demographic and policy datasets, by introducing an L2 penalty that shrinks coefficients without eliminating variables entirely. This allows the model to incorporate correlated predictors (e.g., overlapping age or race categories) while maintaining interpretability.

We compared Ridge Regression with alternative linear and tree-based models - Linear Regression, Lasso Regression, Random Forest, and Gradient Boosting - to benchmark performance. All predictors were standardized prior to fitting.

Model	R^2	RMSE
Random Forest	0.956	28.887
Gradient Boosting	0.893	44.933
Ridge	0.29	115.774
Lasso	0.289	115.879
Linear Regression	0.288	115.923

Table 3: *Model performance metrics (R^2 and RMSE) for each regression approach.*

- **Purpose:** Show comparative predictive performance.

- **Key takeaway:** Random Forest achieved the highest predictive accuracy ($R^2 \approx 0.956$), but Ridge was retained for interpretability despite its modest R^2 (≈ 0.290).

4.2 Key Findings from Linear Models

To better understand how predictors relate to certification volume, we also examined coefficients from an OLS regression model using the same features. While OLS and Ridge coefficients differ in scale due to regularization, the OLS output is useful for interpreting direction, magnitude, and statistical significance.

Feature	Coef	StdErr	t	p	Sig	95% CI
const	-0.2808	0.025	-11.194	0	***	[-0.330, -0.232]
NumIncentives	-0.5456	4.973	-0.110	0.913		[-10.300, 9.209]
Lag1_NumIncentives	5.7417	5.072	1.132	0.258		[-4.207, 15.690]
Lag2_NumIncentives	6.3244	4.794	1.319	0.187		[-3.079, 15.727]
Trifecta	25.6603	6.721	3.818	0	***	[12.477, 38.844]
PctRace_1	15.1747	1.252	12.120	0	***	[12.719, 17.631]
PctRace_2	14.7871	1.374	10.762	0	***	[12.092, 17.482]
PctRace_3	8.3232	1.606	5.182	0	***	[5.173, 11.474]
PctRace_4	58.1673	2.644	22.003	0	***	[52.982, 63.353]
PctRace_5	-137.3302	13.723	-10.007	0	***	[-164.248, -110.412]
PctRace_6	12.7934	8.039	1.591	0.112		[-2.975, 28.562]
PctSex_1	0.6033	5.367	0.112	0.911		[-9.925, 11.132]
PctSex_2	-28.6878	4.946	-5.800	0	***	[-38.389, -18.986]
PctAge_1	-112.2138	10.252	-10.946	0	***	[-132.323, -92.105]
PctAge_2	130.1795	24.540	5.305	0	***	[82.043, 178.316]
PctAge_3	-93.5070	29.364	-3.184	0.001	**	[-151.106, -35.908]
PctAge_4	83.3943	21.016	3.968	0	***	[42.170, 124.618]
PctAge_5	11.1418	13.034	0.855	0.393		[-14.425, 36.709]
PctAge_6	29.3039	14.571	2.011	0.044	*	[0.723, 57.885]
PctAge_7	28.5783	19.000	1.504	0.133		[-8.692, 65.848]
PctAge_8	-22.7856	18.347	-1.242	0.214		[-58.774, 13.203]
PctAge_9	-90.8218	21.067	-4.311	0	***	[-132.146, -49.497]
PctAge_10	128.7235	22.881	5.626	0	***	[83.840, 173.607]
PctAge_11	-102.5894	22.722	-4.515	0	***	[-147.160, -58.019]
PctAge_12	1.9874	21.927	0.091	0.928		[-41.024, 44.999]
PctAge_13	-47.9090	21.489	-2.229	0.026	*	[-90.062, -5.757]
PctAge_14	48.4631	23.890	2.029	0.043	*	[1.602, 95.324]
PctAge_15	-12.8045	39.846	-0.321	0.748		[-90.965, 65.356]
PctAge_16	-102.4499	64.798	-1.581	0.114		[-229.554, 24.654]
PctAge_17	297.5366	65.751	4.525	0	***	[168.563, 426.510]
PctAge_18	-202.3118	27.450	-7.370	0	***	[-256.156, -148.468]

Table 4: *OLS regression coefficients, standard errors, t-values, p-values, significance stars, and 95% confidence intervals.*

- **Key positive predictors:** `PctRace_4` (Asian population %), `PctAge_2` (specific middle-age cohort), `PctAge_4`, `PctAge_10`, and `Trifecta`.
- **Key negative predictors:** `PctRace_5`, `PctAge_1`, `PctAge_3`, `PctAge_9`, `PctAge_11`, `PctAge_18`.
- Many coefficients are statistically significant ($p < 0.05$), particularly demographic variables and the `Trifecta` political control indicator.

4.3 Visualizing Effect Sizes

To make coefficient magnitudes more interpretable, we plotted the absolute values of the 15 largest coefficients from the OLS regression.

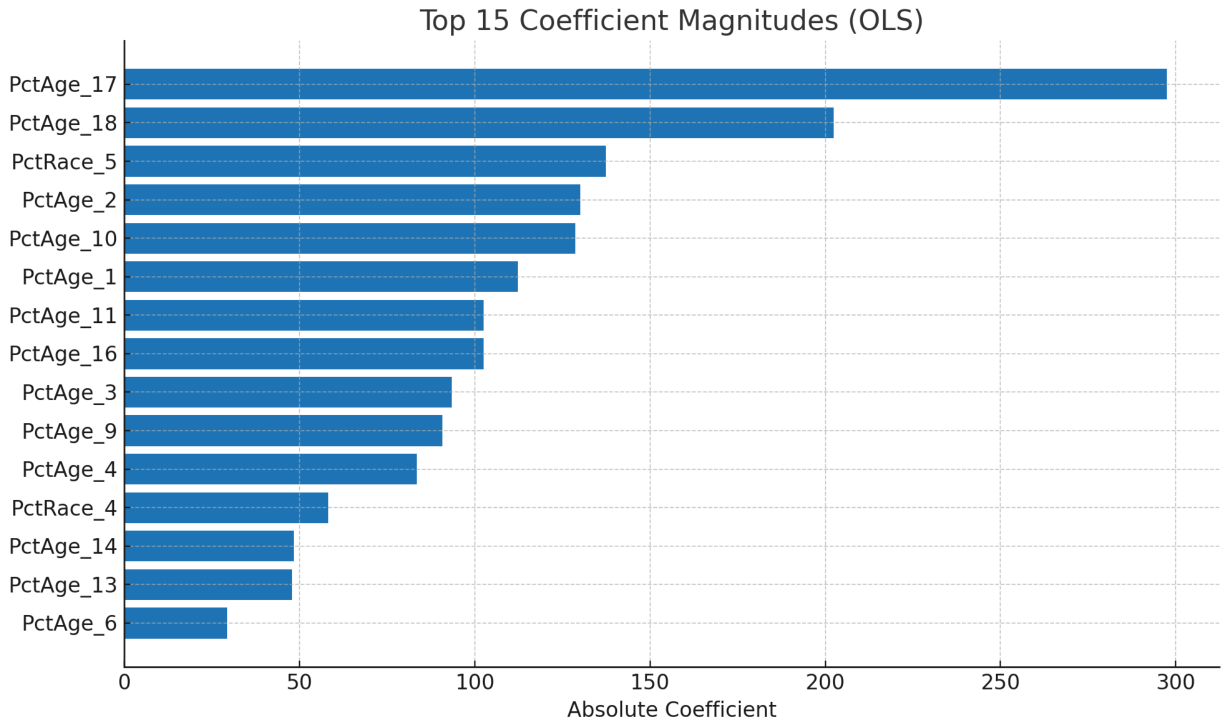


Figure 5. *Horizontal bar chart of the top 15 predictors ranked by absolute coefficient size from the OLS regression.*

- **Purpose:** Highlight which variables have the largest modeled impact, regardless of sign.
- `PctAge_17` (very high positive) and `PctRace_5` / `PctAge_18` (very high negative) stand out as the most influential in magnitude.
- This visualization supports the narrative that both demographic composition and political alignment are strong correlates of certification volume.

4.4 Comparing to Nonlinear Models

Tree-based models such as Random Forest provide a different measure of variable importance that captures nonlinear effects and interactions.

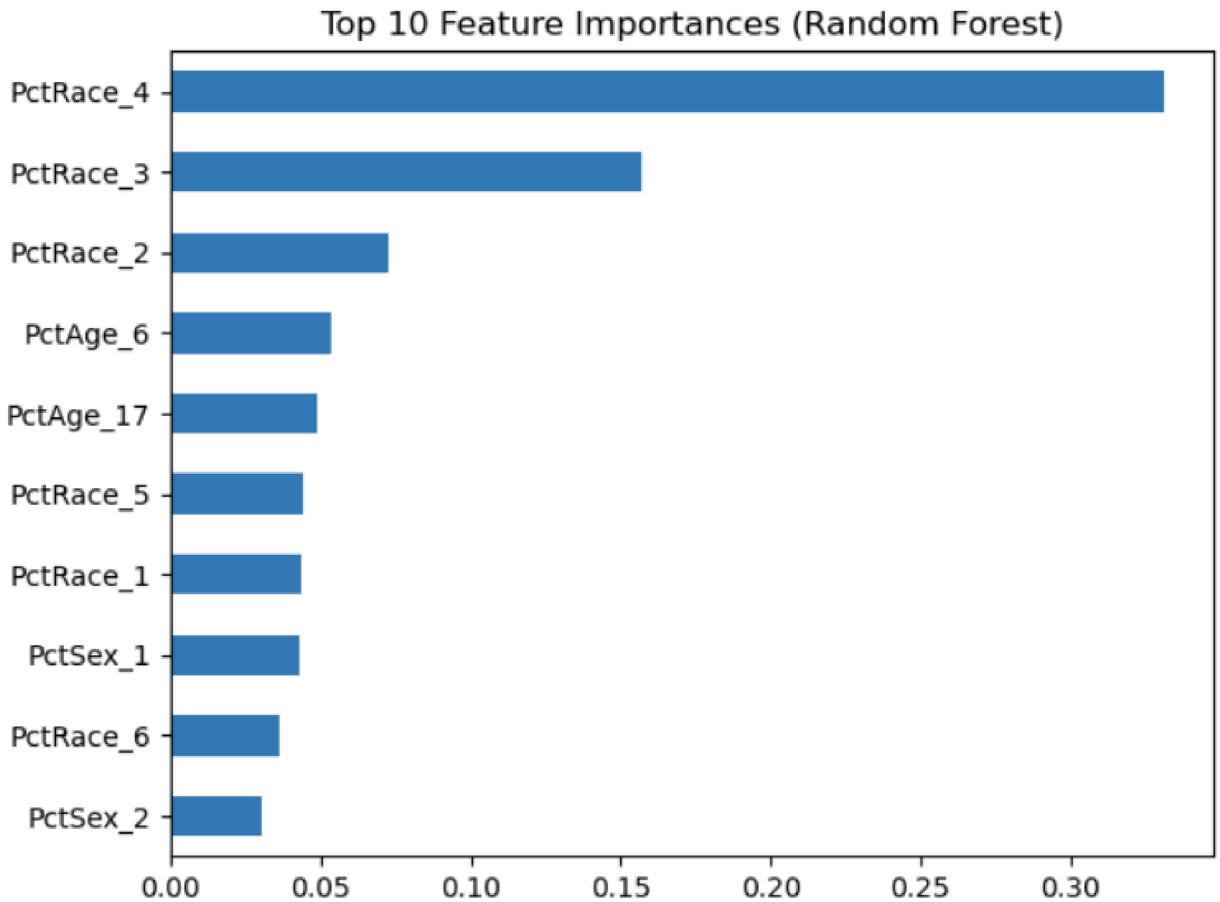


Figure 6. *Random Forest top-10 feature importances for predicting certification volume.*

- Purpose: Compare feature ranking between linear and nonlinear approaches.
- `PctRace_4`, `PctRace_3`, and `PctAge_6` emerge as key features in both frameworks, suggesting robust predictive value.
- Some features with lower OLS coefficients have higher Random Forest importance, likely due to interaction effects not captured in linear models.

4.5 Interpretation and Policy Relevance

These results reinforce several themes from Section 3:

1. **Demographic targeting** — Certain population segments (e.g., higher proportions of older residents, specific racial groups) are strongly associated with certification activity.

2. **Political alignment** — Trifecta status remains a consistent positive predictor, echoing descriptive findings from earlier sections.

3. **Short-term policy boosts** — Lagged incentive counts tend to have weaker or negative coefficients, suggesting diminishing returns without sustained policy action.

Taken together, the regression analysis supports a multifactor explanation for variation in certification volume, where demographic composition, political control, and short-term incentive deployment all play meaningful roles.

5. Dimensionality Reduction (PCA)

5.1 Rationale

Many of the demographic variables in our dataset, particularly race, age, and sex percentage, are inherently collinear. Because these values are proportions that sum to fixed totals (e.g., age groups summing to 100%), an increase in one category necessarily implies decreases in others. This multicollinearity can distort regression coefficients, inflate variance, and complicate interpretation.

To address this, we applied Principal Component Analysis (PCA) to the standardized demographic variables. PCA transforms the original set of correlated predictors into a smaller set of orthogonal (uncorrelated) components that collectively capture the majority of the variance in the dataset. These components can then be used in predictive models to reduce redundancy while preserving meaningful structure in the demographic data.

5.2 Results

Figure 7 presents the scree plot from the expanded PCA. The first six principal components together explain approximately 85% of the total variance in the demographic dataset, with the first component alone capturing 33.5%. The plot shows a steep initial drop in variance explained, followed by a more gradual decline after the sixth component, suggesting diminishing returns in explanatory power beyond this point.

Table 5 lists the top 10 absolute loadings for each of the first six principal components:

- PC1 is dominated by older age groups, particularly PctAge_13, PctAge_14, PctAge_1, and PctAge_3, indicating that it represents a broad population aging gradient.

- PC2 loads heavily on sex variables (PctSex_2, PctSex_1) alongside PctAge_6 and PctAge_10, capturing gender composition coupled with mid-life population structure.

- PC3 is strongly driven by racial composition, especially high loadings for PctRace_4 and PctRace_1, with additional influence from PctRace_5 and PctRace_6. This likely reflects racial composition contrasts.

- PC4 is also race-heavy, with PctRace_2 and PctRace_5 as top contributors, but includes younger age cohorts such as PctAge_7 and PctAge_10, potentially indicating youth racial distribution patterns.

- PC5 blends race and age: PctRace_2, PctRace_1, and PctAge_8 appear alongside PctAge_5 and PctAge_16, which may reflect mixed demographic density or diversity profiles.

- PC6 again combines age and race, but with smaller variance contribution, suggesting more localized demographic nuances.

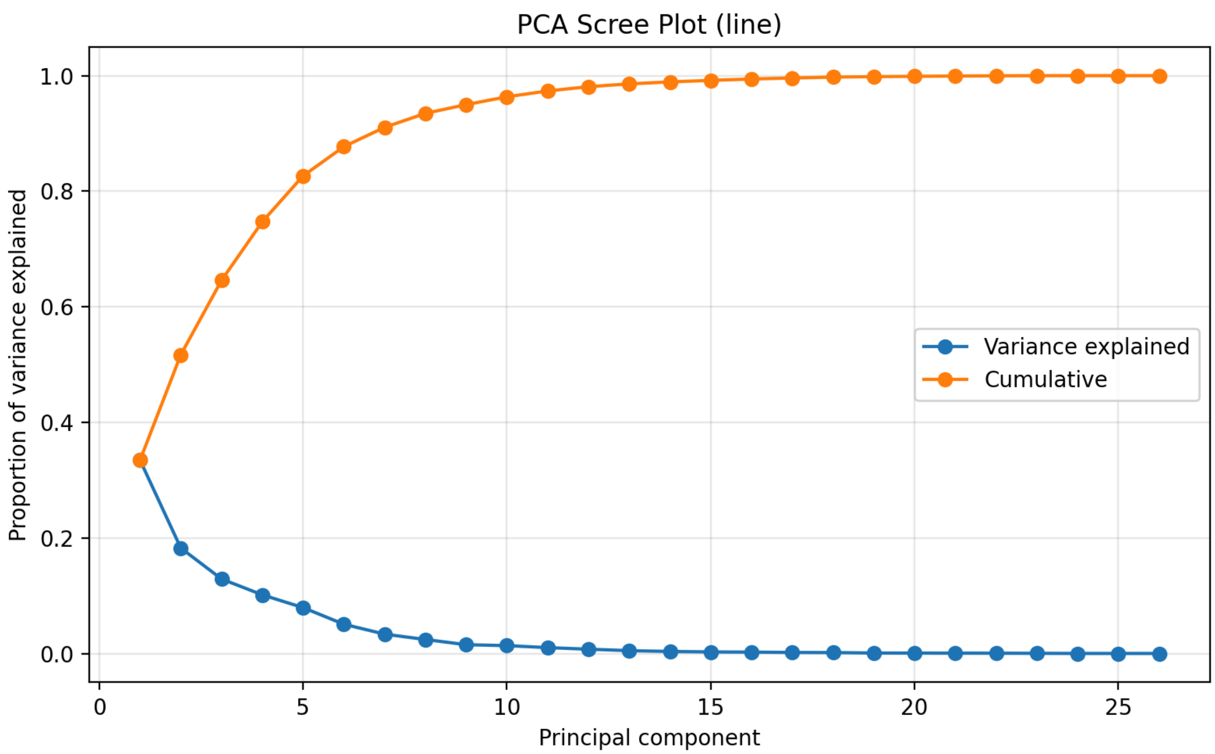


Figure 7. *Scree plot showing variance explained (blue) and cumulative variance explained (orange) for all principal components from the demographic dataset. The first six components explain ~85% of the total variance.*

PC1	PC2	PC3	PC4	PC5	PC6
PctAge_13	PctSex_2	PctRace_4	PctRace_2	PctRace_2	PctAge_17
PctAge_14	PctSex_1	PctRace_1	PctAge_11	PctAge_8	PctAge_4
PctAge_1	PctAge_6	PctRace_5	PctAge_10	PctAge_5	PctAge_5
PctAge_3	PctAge_10	PctRace_6	PctAge_7	PctAge_16	PctAge_7
PctAge_12	PctRace_6	PctAge_8	PctRace_5	PctRace_1	PctAge_11
PctAge_15	PctAge_9	PctRace_3	PctRace_6	PctAge_11	PctAge_16
PctAge_2	PctAge_7	PctSex_2	PctAge_15	PctAge_6	PctAge_8
PctAge_18	PctAge_11	PctSex_1	PctSex_1	PctAge_17	PctAge_12
PctAge_16	PctRace_3	PctRace_2	PctSex_2	PctAge_9	PctRace_5
PctAge_4	PctRace_5	PctAge_5	PctRace_4	PctAge_12	PctAge_18

Table 5. *Top 10 absolute loadings for principal components 1–6 from PCA on demographic variables. Variables are state-level percentages by race, sex, or age group. Full loadings are available in Appendix B.*

5.3 Interpretation

The expanded PCA results suggest that much of the variation in state-level demographics relevant to certification activity can be summarized by a small number of underlying dimensions, each capturing distinct structural characteristics of the population:

- **PC1 – Age Structure Gradient:**
  Dominated by older age cohorts such as PctAge_13, PctAge_14, and PctAge_1, with smaller but notable contributions from middle-aged groups, this component appears to measure the relative "aging" of a state's population. Higher PC1 scores indicate older median populations, which may align with more established housing stock and potentially greater adoption of efficiency retrofits.

- **PC2 – Household Composition and Mid-Life Cohorts:**
  Driven by PctSex_2, PctSex_1, and age cohorts like PctAge_6 and PctAge_10, this component may capture variation in household makeup, gender balance, and mid-life housing demand. Its role in influencing certification trends may be indirect, potentially acting through differences in household energy consumption patterns.

- **PC3 and PC4 – Racial Composition Interactions:**
  These components have strong loadings from racial categories (PctRace_4, PctRace_1, PctRace_5, PctRace_2) and certain age groups, suggesting that race-related demographic patterns are not independent of age structure. They may reflect community-specific housing patterns, geographic clustering, or differences in access to green building programs.

- **PC5 and PC6 – Mixed Minor Factors:**
  The later components capture smaller-scale demographic contrasts, such as particular age-race combinations, that explain less variance individually but still contribute to

nuanced differences between states.

Together, the first six components explain over 85% of the total demographic variance (figure 7), indicating that the vast majority of meaningful variation is captured in a relatively small set of orthogonal variables (Table 5).

5.4 Implications for Policy and Modeling

From a policy design perspective, the principal components highlight clusters of demographic characteristics that tend to move together across states. For example, PC1 is dominated by older age groups (PctAge_13, PctAge_14, PctAge_1, PctAge_3), indicating a population aging gradient. States with high PC1 scores may have more stable housing stock, established homeowner bases, and potentially higher receptivity to efficiency retrofits, suggesting that outreach in these states could emphasize upgrading existing infrastructure. Conversely, states with low PC1 scores, indicative of younger populations, may require incentive designs focused on new construction or rental markets.

PC2 is largely driven by sex and select age variables (PctSex_2, PctSex_1, PctAge_6, PctAge_10), which may be serving as proxies for household composition or labor market structures. Policies could be tailored differently in states where household demographics are skewed toward one segment, such as emphasizing workplace-oriented programs in regions with more single-adult households.

PC3 and PC4 incorporate strong racial composition signals (PctRace_4, PctRace_1, PctRace_5, PctRace_2) along with certain age cohorts, aligning with earlier regression findings that racial demographics correlate with certification volumes. These components suggest that racial composition is not acting in isolation but is intertwined with age distribution, and may therefore require multifaceted outreach strategies, for instance, pairing financial incentives with community engagement in areas where racial minority populations coincide with younger housing stock.

From a modeling standpoint, replacing dozens of collinear demographic predictors with a smaller set of orthogonal components can significantly improve model stability, especially in the presence of multicollinearity, which we observed in raw variable correlations. Using the six retained components as predictors in regression or classification models will reduce the risk of overfitting, improve coefficient interpretability, and allow models to generalize better across different state-year contexts.

Finally, from an equity and evaluation lens, PCA allows us to identify demographic “profiles” where policy uptake might be uneven even under equal incentive conditions. For example, a state with similar PC1 and PC3 scores to historically high-certification states could be flagged for proactive program deployment. However, it is important to note that PCA components are abstract mathematical constructs and not causal drivers in themselves; their interpretation must be grounded in domain knowledge and supplemented with contextual data such as income, education, and housing characteristics.

6. Predictive Modeling of Certification Activity

6.1 Objective and setup

We framed a binary classification task to predict whether a given state–year observation falls into the high certification category (above the median annual certification count) or low certification category (at or below the median). The dataset is nearly balanced, with 155 low and 150 high observations, so accuracy provides a meaningful performance measure alongside class-specific metrics.

The model chosen was a Random Forest classifier, which can capture non-linear relationships and interactions without requiring strong parametric assumptions. Input variables included current and lagged incentive counts, Trifecta status, and all available demographic percentages for race, age, and sex. Partisanship of the trifecta (Democratic vs. Republican) was not included as a separate feature in this model.

The dataset was split into training and test sets with stratification on the outcome label. Hyperparameters were tuned via cross-validation on the training set, and the results below reflect performance on the held-out test set.

6.2 Test performance

	Predicted 0	Predicted 1
Actual 0	148	7
Actual 1	14	136

Table 6. *Confusion matrix for the Random Forest classifier on the held-out test set. Rows represent the actual labels, and columns represent the predicted labels. The counts indicate that the model correctly classified the majority of both high-certification (positive) and low-certification (negative) state–years.*

Class	Precision	Recall	F1-score	Support
0	0.91	0.95	0.93	155
1	0.95	0.91	0.93	150
Overall	0.93	0.93	0.93	305

Table 7. *Classification report summarizing precision, recall, and F1-score for each class, along with overall accuracy. Results are based on 61 state–year observations in the test set.*

The model achieved an overall accuracy of 93%, indicating strong generalization to unseen data. For low-certification state–years (class 0), the model's precision was 0.91, meaning that when the model predicted low activity, it was correct 91% of the time. Its recall for this class was 0.95, showing that it successfully identified 95% of all actual low-certification years.

For high-certification state–years (class 1), precision was slightly higher at 0.95, reflecting a low false positive rate. However, recall dropped to 0.91, meaning that roughly 9% of actual high-certification years were missed by the model.

Derived metrics further contextualize these results. The specificity for detecting low-certification years was 95.5%, while the false positive rate was only 4.5%. The false negative rate of 9.3% indicates that missed high-certification years are the primary source of error. The negative predictive value of 91.4% shows that when the model predicted a low-certification year, it was correct in the majority of cases. The balanced accuracy of 93.3% confirms that performance was well-balanced across both classes, and the Matthews correlation coefficient of 0.863 reflects a strong overall relationship between predictions and true labels.

Taken together, these results suggest the model is slightly conservative in predicting high-certification years, favoring precision over recall for the positive class. In practice, this means fewer false alarms at the cost of occasionally missing genuine high-activity periods, a trade-off that can be adjusted depending on policy or operational priorities.

6.3 Feature attribution

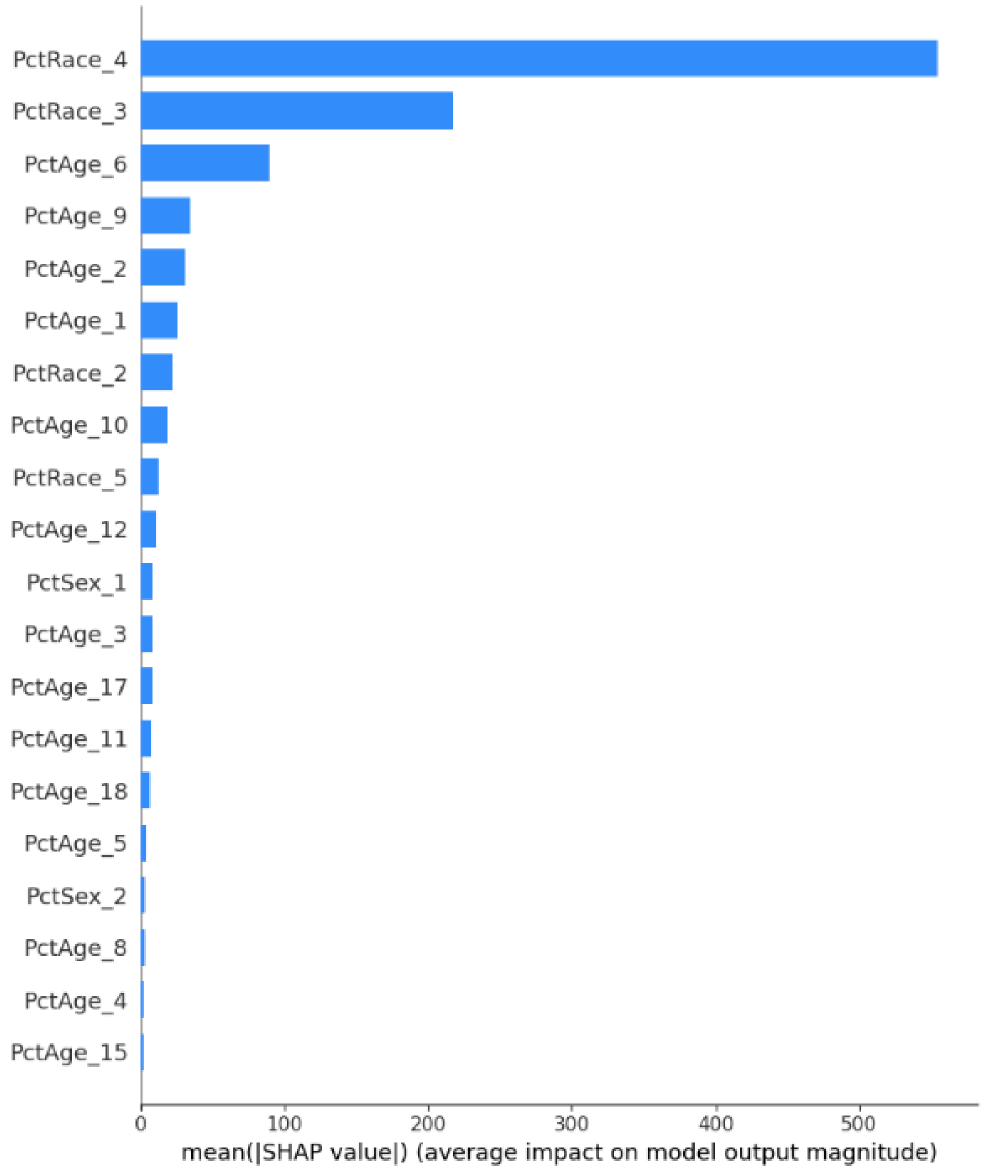


Figure 8. *Mean absolute SHAP values for the Random Forest classifier predicting high- versus low-certification state–years. Features are ranked by their average magnitude of contribution to model predictions across all test set observations. Demographic variables, particularly the percentage of residents identifying as Asian (PctRace_4) and Black (PctRace_3), dominate the model's predictive landscape, followed by age group proportions such as PctAge_6 (ages 55–59) and PctAge_9 (ages 70–74). The prominence of these features aligns with regression and PCA findings, suggesting stable cross-method patterns.*

The SHAP analysis reveals that PctRace_4 (percentage Asian) is the most influential feature in the model, followed by PctRace_3 (percentage Black). Age-related variables, notably PctAge_6 (ages 55–59) and PctAge_9 (ages 70–74), also appear prominently, alongside other demographic factors.

These findings reinforce the patterns identified in Sections 4 and 5, where race composition and age structure emerged as significant predictors of certification volume. SHAP values here quantify the average marginal contribution of each feature to the model's predictions.

6.4 Error analysis

Of the 21 total misclassifications, 14 were false negatives (missed high-certification years) and 7 were false positives (predicted high-certification years that did not occur). The higher count of false negatives suggests that if maximizing recall for high-certification states is a priority, the classification threshold could be lowered or class weights adjusted to favor the positive class.

Reviewing SHAP explanations for these misclassified cases could reveal demographic profiles or policy contexts where the model struggles, potentially guiding targeted feature engineering or threshold adjustments.

6.5 Practical implications

- **Operational targeting:** In scenarios where the cost of missing a high-certification year is high, recall can be prioritized through threshold tuning or class weighting.
- **Feature engineering opportunities:** Given the dominance of demographic variables, interactions between top principal components and incentive/policy variables could be tested in future models.
- **Model interpretability:** The alignment between SHAP importance in classification and top predictors in regression and PCA strengthens confidence in the findings, though all results remain correlational.

7. Geospatial Visualization

To contextualize the statistical and machine learning findings from prior sections, we examined the geographic distribution of certification activity and its relationship to the most influential demographic variable identified in Section 6. Two state-level choropleth maps were generated:

- **Figure 9:** Average certified building count per state-year from 2000–2024.

- **Figure 10:** Mean percentage of residents in racial group *PctRace_4* (identified in Random Forest and SHAP analyses as the strongest predictor of certification volume).

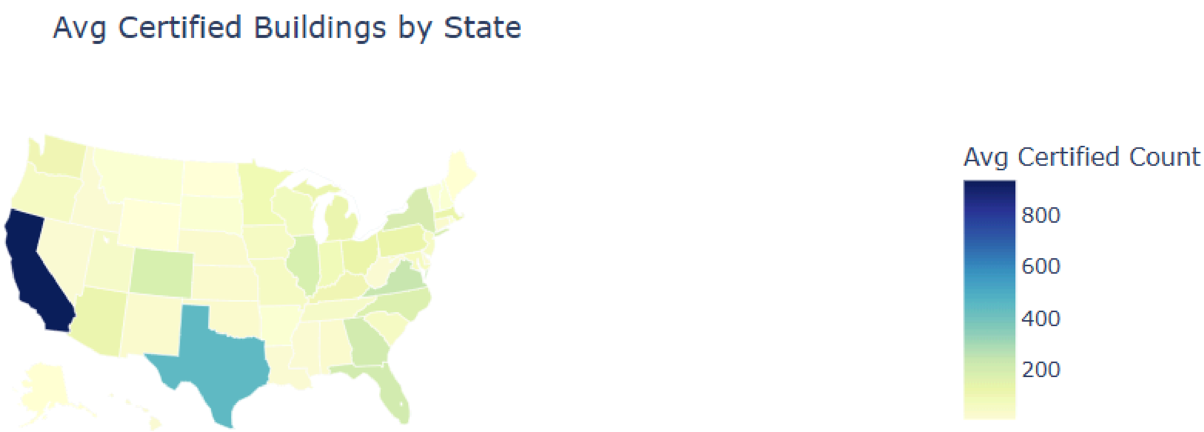


Figure 9. *Average certified buildings per state-year from 2000–2024. Darker shades indicate states with higher average certification counts, with California, Texas, and New York leading nationally.*

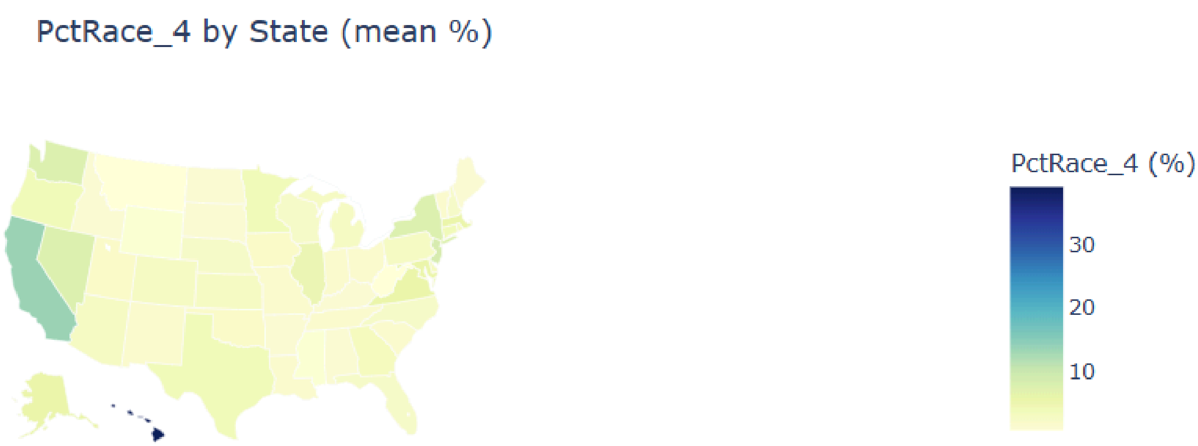


Figure 10. *Mean percentage of residents in racial group PctRace_4 across U.S. states (2000–2024). Higher values are concentrated in Hawaii, California, and parts of the Northeast, showing geographic overlap with states exhibiting high certification counts.*

Figure 9 highlights the highly uneven spatial distribution of green building certifications in the United States. California stands out with the highest average annual certification count, followed by Texas, Florida, and New York. Many states in the Midwest, Great Plains, and Mountain West display relatively low average counts, suggesting possible structural, economic, or policy barriers to adoption. Coastal states generally show higher activity, which could be due to denser urban populations, stronger policy frameworks, and more aggressive incentive programs.

Figure 10 maps the distribution of *PctRace_4*, with the largest proportions observed in Hawaii, California, and select metropolitan states in the Northeast. When compared visually to figure 9, there is a notable spatial overlap: several states with high *PctRace_4* percentages also have elevated certification counts. This spatial correspondence mirrors the statistical findings from earlier regression, Random Forest, and SHAP analyses, where *PctRace_4* was consistently among the top predictors of certification activity.

While these maps alone cannot confirm causality, the co-location of high certification volumes and higher *PctRace_4* proportions suggests that demographic composition may play a role in driving or enabling certification uptake. Such patterns may also interact with local policy environments, housing stock characteristics, and socioeconomic conditions. These insights point to opportunities for targeted outreach or incentive design in states with demographic profiles historically underrepresented in certification programs, as well as deeper investigation into the mechanisms linking demographic structure and green building adoption.

8. Incentive Type Effectiveness (Difference-in-Differences Analysis)

8.1 Rationale

Earlier sections examined the overall relationship between the number of incentive programs and certification counts. This section focuses on the effectiveness of specific incentive types. A Difference-in-Differences (DiD) approach was applied to estimate how the introduction of a particular policy mechanism affected ENERGY STAR certification activity. The method compares changes in certifications in states that adopted a given incentive type with states that did not adopt it during the same time period, controlling for broader temporal trends. This approach isolates the average treatment effect of each policy type on certified building counts.

8.2 Results

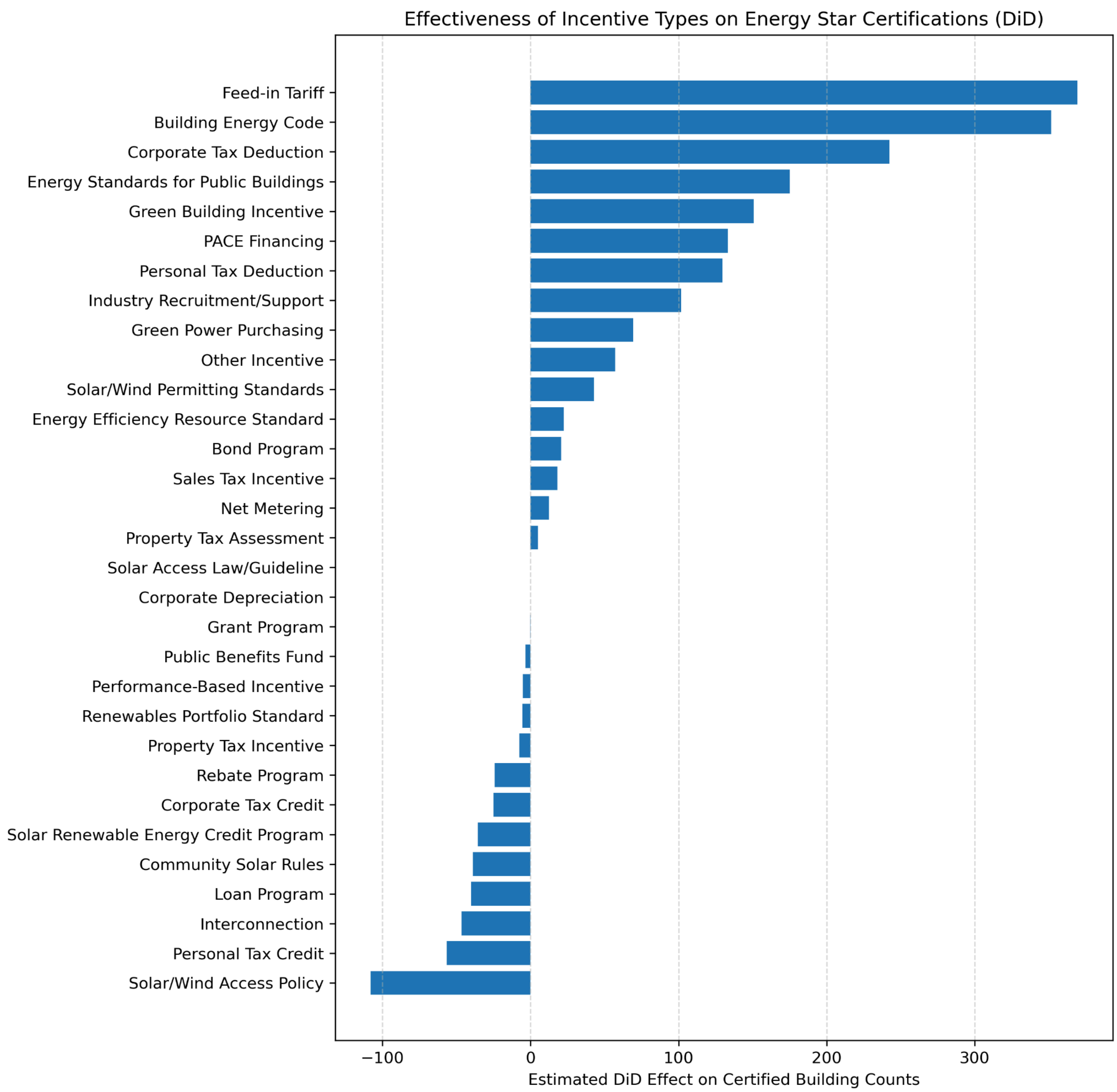


Figure 11. *Estimated difference in differences effects of policy incentive types on counts of Energy Star certified buildings. Bars show point estimates from models with state and year fixed*

effects. Positive values indicate higher certification counts associated with adoption.

Incentive Type	DiD Coefficient	p-value	Significance
Solar/Wind Access Policy	-108.08	0.0000	***
Personal Tax Credit	-56.64	0.0000	***
Interconnection	-46.65	0.0143	*
Loan Program	-40.27	0.0006	***
Community Solar Rules	-39.00	0.0006	***
Solar Renewable Energy Credit Program	-35.44	0.0511	.
Corporate Tax Credit	-24.95	0.0639	.
Rebate Program	-24.13	0.0559	.
Property Tax Incentive	-7.51	0.5176	
Renewables Portfolio Standard	-5.60	0.6691	
Performance-Based Incentiv	-5.23	0.6607	
Public Benefits Fund	-3.37	0.8522	
Grant Program	-0.39	0.9773	
Corporate Depreciation	0.00	NA	
Solar Access Law/Guideline	0.00	NA	
Property Tax Assessment	5.17	0.8476	
Net Metering	12.61	0.3037	
Sales Tax Incentive	18.04	0.3158	
Bond Program	20.73	0.2934	
Energy Efficiency Resource Standard	22.46	0.1792	
Solar/Wind Permitting Standards	42.87	0.0830	.
Other Incentive	57.09	0.0052	**
Green Power Purchasing	69.20	0.0018	**
Industry Recruitment/Suppo	101.66	0.0001	***
Personal Tax Deduction	129.69	0.0005	***
PACE Financing	133.30	0.0000	***
Green Building Incentive	150.68	0.0694	.
Energy Standards for Public Buildings	175.25	0.0000	***
Corporate Tax Deduction	242.49	0.0000	***
Building Energy Code	351.73	0.0000	***
Feed-in Tariff	369.43	0.0000	***

Table 8. *Difference in differences estimates by incentive type. Coefficients represent the change in the number of Energy Star certified buildings associated with adoption of each policy type. p-values are based on clustered standard errors at the state level. NA indicates insufficient variation to estimate a p-value for that category.*

The analysis identifies several high-performing incentive types. Feed-in Tariff, Building Energy Code, and Corporate Tax Deduction show the largest estimated positive impacts, with coefficients exceeding 240 and p-values below 0.001. These mechanisms appear highly effective at driving measurable increases in certifications, likely because they either mandate compliance or provide substantial, predictable financial benefits that appeal to commercial property owners.

A second tier of effective incentives includes PACE Financing, Personal Tax Deduction, Industry Recruitment and Support, and Green Power Purchasing, all of which demonstrate statistically significant positive effects. These mechanisms operate through a range of channels, from facilitating low-cost financing to building market awareness, and appear to complement more stringent regulatory approaches.

In contrast, several common incentive types, including Rebate Programs, Loan Programs, and Personal Tax Credits, exhibit small or negative estimated effects. Some of these negative coefficients are statistically significant, suggesting that these programs may be deployed in contexts with lower certification potential, or that they primarily target smaller-scale residential projects that do not result in ENERGY STAR certifications.

A subset of incentives, such as Green Building Incentives and Solar/Wind Permitting Standards, have positive coefficients that approach statistical significance but do not meet conventional thresholds. These cases may warrant further examination, particularly given their direct relevance to building energy performance.

8.3 Interpretation

The difference-in-differences analysis quantified how the introduction of specific incentive types is associated with changes in ENERGY STAR certified building counts, controlling for state and year fixed effects. Table 8 presents the estimated coefficients and p-values, while Figure 11 visualizes the magnitude and direction of effects across all tested incentive types.

Several incentive types show strong, statistically significant positive associations with certification counts. Notably, Feed-in Tariffs (+369) and Building Energy Codes (+352) are associated with the largest observed increases, suggesting that binding regulatory standards and guaranteed payment mechanisms may be particularly effective at scaling adoption. Other high-impact categories include Corporate Tax Deductions (+242), Energy Standards for Public Buildings (+175), and PACE Financing (+133). These results indicate that both policy mandates and financial tools that reduce up-front capital costs can generate substantial gains.

A second tier of effective incentives includes Industry Recruitment/Support, Green Power Purchasing, and Other Incentives, all of which return statistically significant positive coefficients above +50. These mechanisms may operate by indirectly facilitating adoption, such as by supporting green construction industries or ensuring markets for renewable energy.

Conversely, several incentive types yield negative coefficients, some of which are statistically significant. For example, Solar/Wind Access Policy (–108), Personal Tax Credit (–57), and Interconnection (–47) are associated with lower certification counts following adoption. While these results do not necessarily imply causal harm, they may indicate that such policies are implemented in contexts with weaker market readiness, or that they target outcomes unrelated to building certifications (e.g., distributed generation projects).

The analysis also shows that the Rebate Program, despite being the most common incentive type nationally, have only a small, non-significant negative coefficient (–24). This supports earlier findings in Section 3 that rebates alone may not strongly drive certification volume without complementary policies.

Overall, the DiD results suggest that:

1. Mandates and codes (e.g., Building Energy Codes, Energy Standards for Public Buildings) and capital-access mechanisms (e.g., PACE Financing, Corporate Tax Deductions) appear most consistently effective in boosting certification counts.

2. Common but low-impact programs (e.g., rebates) may require integration into broader policy packages to achieve measurable effects.

3. Negative coefficients for certain renewable-specific policies highlight the importance of policy alignment - programs designed for generation capacity or interconnection do not always translate into efficiency certification gains.

These findings align with the earlier modeling in Section 6, where policy type and implementation context were shown to interact with demographic and political conditions. The DiD results provide targeted insight into which mechanisms have historically coincided with measurable certification growth, offering actionable guidance for policymakers aiming to allocate resources toward the highest-impact interventions.

9. Discussion and Limitations

This study finds consistent associations between certification activity and two broad domains: structural context and policy design. Structural context is captured by demographic composition and unified political control. Policy design is reflected in the mix and timing of incentive types.

Synthesis across methods. Descriptive statistics, linear models, random forests, SHAP explanations, geospatial patterns, and difference-in-differences estimates all point to the same directional conclusions. Demographic structure, particularly race and age composition, is repeatedly identified as a strong correlate of certification outcomes. Trifecta control of state government is associated with higher average certification counts. The number of new incentive programs in a year is positively related to certifications but with modest effect size at the state-year level. Difference-in-differences estimates indicate that mandates and capital-access tools are most strongly linked to increases in certification counts, while several common but lighter-touch instruments show small or negative estimates.

Interpreting associations. The demographic variables operate as proxies for broader market conditions that the dataset does not measure directly. These likely include commercial real estate scale, age and type of building stock, urbanization, and household tenure. Trifecta status is best read as a proxy for policy stability and implementation capacity rather than a causal lever. The observed policy effects vary by mechanism. Codes, standards for public buildings, feed-in tariffs, corporate tax deductions, and PACE financing are consistently associated with higher certification counts, consistent with instruments that either require compliance or reduce capital frictions for large projects. By contrast, rebates and similar instruments may primarily target smaller residential actions or require complementary policies to translate into certified commercial outcomes.

Robustness and triangulation. Agreement between linear and nonlinear importance rankings, alignment of SHAP feature attributions with coefficient patterns, and spatial overlap between high-activity states and key demographic profiles suggest that the findings are not an artifact of a single modeling choice. Classification results on held-out data indicate strong generalization and balanced performance across classes, with a slight conservatism in flagging high-activity years. This consistency supports the main conclusions while remaining descriptive rather than causal outside the difference-in-differences framework.

Limitations.

1. **Construct measurement.** The policy dataset encodes program presence and type, not intensity, budget, enforcement, administrative capacity, or marketing reach. Two programs of the same type can differ substantially in effect.

2. **Temporal alignment.** Recorded start years may not match the onset of implementation, and certification responses can lag policy adoption. The analysis includes short lags but may miss longer or heterogeneous delays.

3. **Missingness and selection.** Some state-years are excluded because of incomplete fields or harmonization failures. If missingness is systematic, estimates may be biased.

4. **Compositional collinearity.** Demographic percentages are compositional by construction. Principal components reduce redundancy but remain abstract constructs that require domain interpretation.

5. **Ecological inference.** All results are at the state-year level and cannot be assigned to individuals or projects. Within-state heterogeneity is not captured.

6. **Model dependence.** Although results are qualitatively consistent across models, coefficient magnitudes and variable rankings can vary with specification, regularization strength, and feature sets.

7. **Identification limits.** Outside the policy-type difference-in-differences, results are correlational. Even within difference-in-differences, validity depends on parallel trends and limited treatment heterogeneity. Staggered adoption and varying treatment timing

introduce additional complexity that requires event‑study diagnostics.

8. **Omitted variables.** Income, education, energy prices, utility demand‑side management spending, building stock characteristics, and urban density are not included. These factors may confound observed relationships.

9. **Multiple testing.** Analyses span many features and incentive categories. Although patterns are coherent, formal adjustment for multiple comparisons is not applied and should be addressed in future work.

10. **Spillovers and interference.** Cross‑border and metropolitan spillovers are plausible. The current design assumes independence across states and does not model spatial dependence.

Implications. The evidence supports a practical view in which high‑impact outcomes arise when well‑designed policy instruments are deployed in markets that already possess enabling structural characteristics. Designing program portfolios that combine mandates or standards with capital‑access tools, and aligning outreach with local demographic and market profiles, appears most promising. These conclusions should inform forecasting, targeting, and causal evaluation in subsequent work.

10. Future Work

Data expansion.

1. **Socioeconomic controls:** Median income, educational attainment, poverty, rent versus own tenure, and housing cost burden at state and sub‑state scales.

2. **Built‑environment metrics:** Building age distributions, commercial floor area by type, construction pipeline, and renovation permit activity.

3. **Energy context:** Retail electricity and gas prices, utility demand‑side management budgets and participation, and program administrator characteristics.

4. **Policy richness:** Program budgets, eligibility scope, enforcement practices, code version adoption and compliance rates, and any available implementation quality indicators.

5. **Cross‑scheme outcomes:** LEED, WELL, and other certification registries to assess external validity beyond ENERGY STAR.

Forecasting and deployment.

1. **One-year-ahead forecasting:** Rolling-origin cross-validation for state-level predictions using dynamic regression or tree-based ensembles with lagged policy and structural features.

2. **Probability calibration and thresholds:** Calibrate predicted probabilities and choose operating points based on policy goals, for example prioritizing recall to reduce missed high-activity years when planning outreach or technical assistance.

3. **Model monitoring:** Establish drift monitoring and periodic recalibration as new certifications and policies arrive, with transparent version control for models and data.

Causal evaluation.

1. **Modern staggered-adoption estimators:** Use estimators that accommodate heterogeneous treatment timing, complemented by event-study plots to verify pre-trends and to trace dynamic effects.

2. **Synthetic control case studies:** Construct case studies for major policy introductions in large states or cities to estimate counterfactual trajectories.

3. **Heterogeneity and mediation:** Interact policy variables with demographic principal components, urbanization, or building-stock measures to identify where and for whom policies are most effective.

4. **Robustness and falsification:** Placebo timings, negative controls, leave-one-state-out analyses, and alternative spike definitions to test sensitivity.

Spatial resolution and dependence.

1. **Sub-state analysis:** Extend to county or metropolitan statistical area levels to reduce ecological bias and reveal within-state disparities.

2. **Spatial econometrics:** Test for spatial autocorrelation and, if present, estimate spatial lag or spatial error models. Consider geographically weighted regression to explore spatially varying relationships.

3. **Spillover measurement:** Examine cross-border effects along metropolitan corridors and utility service territories.

Equity and accountability.

1. **Distributional assessment:** Overlay certifications and incentives with environmental justice indicators, income and rent burden, and renter share to quantify who benefits.

2. **Fairness diagnostics for prediction:** Track false negative and false positive rates across demographic and geographic subgroups and adjust thresholds or features if unequal error burdens emerge.

3. **Program design feedback:** Use equity diagnostics to inform the design of targeted incentives or technical assistance where uptake has historically lagged.

Reproducibility and transparency.

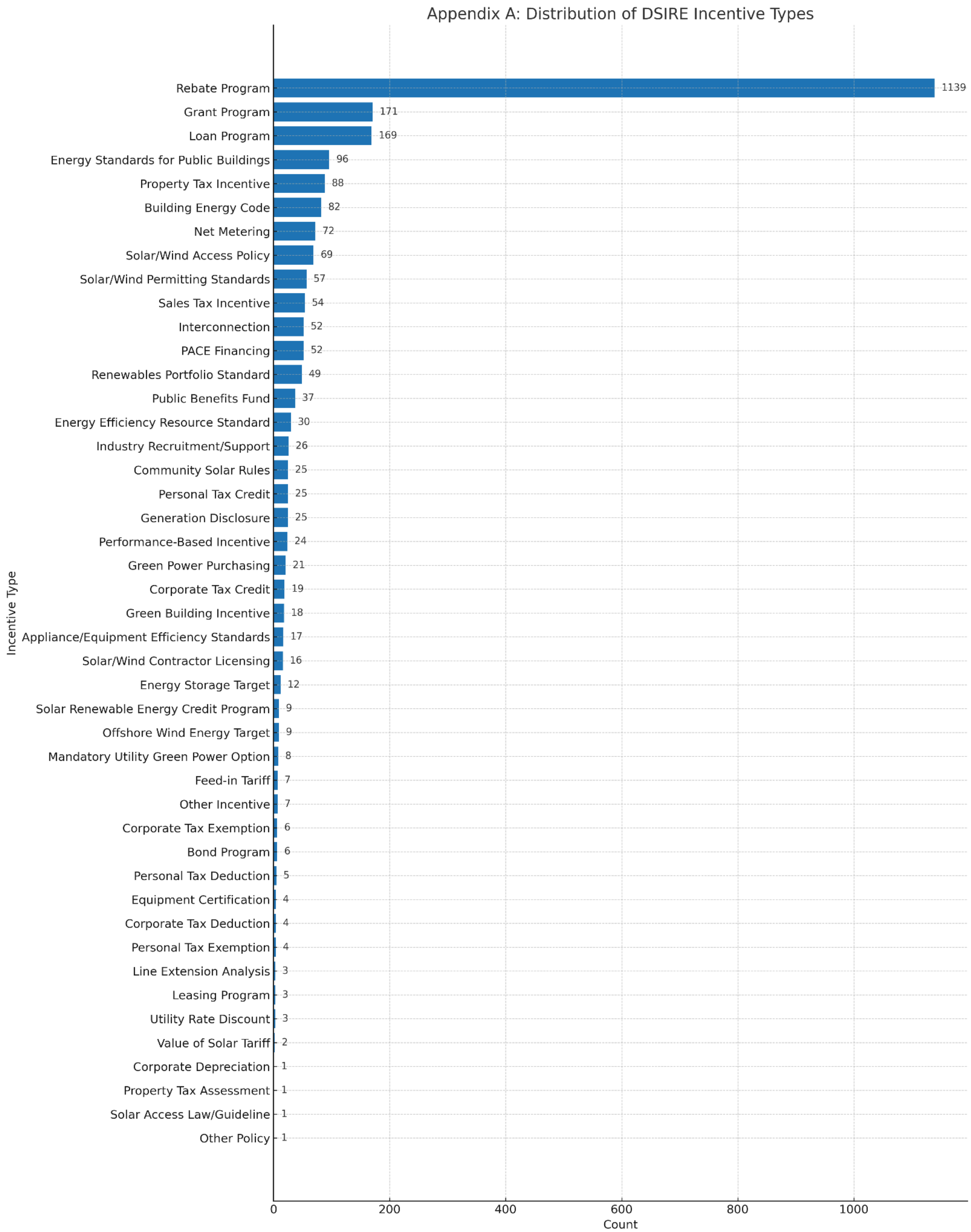
1. **Pre-registration of analysis plans** for causal studies, including identification strategies and primary outcomes.

2. **Sensitivity catalog** that records all tested specifications, hyperparameters, and data versions.

3. **Open artifacts** that include cleaned datasets, code, and figure generation scripts with documented software environments.

11. Appendix

Appendix A:


Appendix A: Distribution of DSIRE Incentive Types
Incentive Type
Count
0
200
400
600
800
1000
Rebate Program 1139
Grant Program 171
Loan Program 169
Energy Standards for Public Buildings 96
Property Tax Incentive 88
Building Energy Code 82
Net Metering 72
Solar/Wind Access Policy 69
Solar/Wind Permitting Standards 57
Sales Tax Incentive 54
Interconnection 52
PACE Financing 52
Renewables Portfolio Standard 49
Public Benefits Fund 37
Energy Efficiency Resource Standard 30
Industry Recruitment/Support 26
Community Solar Rules 25
Personal Tax Credit 25
Generation Disclosure 25
Performance-Based Incentive 24
Green Power Purchasing 21
Corporate Tax Credit 19
Green Building Incentive 18
Appliance/Equipment Efficiency Standards 17
Solar/Wind Contractor Licensing 16
Energy Storage Target 12
Solar Renewable Energy Credit Program 9
Offshore Wind Energy Target 9
Mandatory Utility Green Power Option 8
Feed-in Tariff 7
Other Incentive 7
Corporate Tax Exemption 6
Bond Program 6
Personal Tax Deduction 5
Equipment Certification 4
Corporate Tax Deduction 4
Personal Tax Exemption 4
Line Extension Analysis 3
Leasing Program 3
Utility Rate Discount 3
Value of Solar Tariff 2
Corporate Depreciation 1
Property Tax Assessment 1
Solar Access Law/Guideline 1
Other Policy 1

Appendix B:

Top 10 Variables per PC with Signed Loadings (PC1–PC6)

	PC1	PC2	PC3	PC4	PC5	PC6
1	PctAge_13 (-0.308)	PctSex_2 (-0.350)	PctRace_4 (+0.443)	PctRace_2 (-0.392)	PctRace_2 (-0.373)	PctAge_17 (+0.328)
2	PctAge_14 (-0.304)	PctSex_1 (+0.350)	PctRace_1 (-0.432)	PctAge_11 (+0.375)	PctAge_8 (+0.358)	PctAge_4 (+0.326)
3	PctAge_1 (+0.303)	PctAge_6 (+0.331)	PctRace_5 (+0.392)	PctAge_10 (+0.291)	PctAge_5 (-0.335)	PctAge_5 (+0.320)
4	PctAge_3 (+0.299)	PctAge_10 (-0.321)	PctRace_6 (+0.366)	PctAge_7 (-0.251)	PctAge_16 (+0.318)	PctAge_7 (-0.319)
5	PctAge_12 (-0.287)	PctRace_6 (+0.260)	PctAge_8 (+0.207)	PctRace_5 (+0.249)	PctRace_1 (+0.295)	PctAge_11 (-0.303)
6	PctAge_15 (-0.282)	PctAge_9 (-0.243)	PctRace_3 (-0.195)	PctRace_6 (+0.246)	PctAge_11 (-0.267)	PctAge_16 (+0.272)
7	PctAge_2 (+0.273)	PctAge_7 (+0.239)	PctSex_2 (+0.188)	PctAge_15 (-0.234)	PctAge_6 (-0.236)	PctAge_8 (-0.222)
8	PctAge_18 (-0.261)	PctAge_11 (-0.229)	PctSex_1 (-0.188)	PctSex_1 (+0.220)	PctAge_17 (+0.221)	PctAge_12 (-0.218)
9	PctAge_16 (-0.246)	PctRace_3 (+0.217)	PctRace_2 (+0.178)	PctSex_2 (-0.220)	PctAge_9 (+0.220)	PctRace_5 (+0.203)
10	PctAge_4 (+0.241)	PctRace_5 (+0.215)	PctAge_5 (-0.177)	PctRace_4 (+0.210)	PctAge_12 (-0.220)	PctAge_18 (+0.201)

Appendix S1. Reproducibility Notebook

Supplementary Notebook S1. All data cleaning, analysis, figures, and model outputs are contained in the following Jupyter notebook.

- Link to google drive with Jupyter Notebook (available as .ipynb and as HTML file): https://drive.google.com/drive/folders/1fmmjS9eZwN7medaUYz1YeVFM4huK4Oy9?usp=drive_link

Contents overview.

- Data loading and merges
- Exploratory summaries and figures
- PCA (full loadings and scree)
- Regression and model comparison
- Classification diagnostics
- Difference-in-differences template
- Export cells for tables and figures

Environment. The notebook header includes package versions and random seeds. If needed, a requirements list is included at the top cell.