

From Permits to Units*

Factors Associated with Housing Approvals in California Cities

Abhishek Rasikbhai Shekhada

December 9, 2025

This paper explores why some California cities approve more housing units each year than others, using application level data from the state's Housing Element Annual Progress Report (APR). After cleaning and aggregating the data into a city-year format, I examine how past approval activity, the share of affordable units and the use of density bonuses relate to annual housing approvals. I fit a pooled OLS model as well as a city fixed effects model with clustered standard errors to account for differences across cities and repeated observations over time. The results show that cities with higher affordable housing shares tend to approve fewer total units, and prior year approvals show a small mean-reversion pattern once fixed effects are included. These findings highlight how both long-term city characteristics and year-to-year shifts in development activity play a role in housing production across California.

Introduction

California's ongoing housing shortage continues to shape the lives of people across the state. Even though the issue is often discussed at a statewide level, the actual number of new homes built each year depends heavily on decisions made by individual cities. Some cities consistently approve large amounts of new housing, while others approve very little, even when facing similar economic or policy conditions. This uneven pattern raises an important question: *what factors are associated with how many housing units a city approves in a given year?*

Understanding these differences matters because local approval decisions play a major role in determining where housing is built, how quickly communities grow and whether cities are meeting their housing needs. In recent years, California has introduced several policies such as density bonuses and streamlined approval pathways designed to encourage more housing. At the same time, cities vary in how many affordable units they approve and they also show very

*Project repository available at: <https://github.com/arshekhada/Math-261A-Project-1>.

different levels of development activity from year to year. These differences suggest that city level approval behavior is shaped by more than just state policy and that examining variation across cities may help us better understand housing production patterns.

While a lot of research focuses on statewide housing totals, broader market conditions or the effects of specific policies, there is less work that looks closely at how approval activity changes *within the same city over time*. Studying this city-by-city variation can highlight the local factors that influence development patterns and reveal whether past approval behavior or affordable housing commitments are related to annual production levels.

In this paper, I use application-level permit data from the California Housing Element Annual Progress Report (APR) (California Department of Housing and Community Development 2025) and convert it into a city–year dataset. I then analyze how past approvals, affordable housing shares and density bonus usage relate to the total number of units a city approves in a given year. I estimate both a pooled OLS model and a city fixed-effects model with clustered standard errors to account for long-term differences between cities and repeated observations over time. The main findings of the analysis show that cities approving a higher share of affordable units tend to approve fewer total units overall and that prior-year approvals show a mild “mean-reversion” pattern once fixed effects are included.

The rest of the paper is organized as follows: the **Data** section describes the dataset and explains how each variable was constructed; the **Methods** section outlines the regression models and key assumptions; the **Results** section presents the main findings using tables and figures; and the **Discussion** section summarizes the implications of the study, acknowledges limitations and suggests possible areas for future research.

Data

The dataset for this project comes from the *Housing Element Annual Progress Report (APR) – Table A* (California Department of Housing and Community Development 2025), which is published each year by the California Department of Housing and Community Development (California Department of Housing and Community Development 2025). The APR tracks every housing development application submitted in a given year, so each row in the raw dataset represents a single project application made by a city or county.

Observational Units

Because my research question focuses specifically on cities, I restrict the dataset to **cities only** and remove all county-level records. I then aggregate the application-level entries into a **city–year** format, meaning that each row in my final dataset represents one city during one specific year. This structure makes it possible to study how approval activity changes over

time within the same city, which is important for the fixed-effects analysis used later in the paper.

Variables Used

I construct several variables from the APR that are directly relevant to the regression models:

- **TotalApprovedUnits:** The total number of housing units approved in a city in a given year. I compute this by summing the “approved units” column across all applications in that city-year.
- **PriorApprovedUnits:** The total units approved in the previous year for the same city. This is created by sorting the dataset by city and year and then taking the lag of TotalApprovedUnits. This variable helps capture whether approval activity carries momentum from one year to the next.
- **VeryLowUnits, LowUnits, ModUnits, AboveModUnits:** APR reports approved units by income category. I sum these categories to create totals by income group.
- **LowIncomeShare:** The share of approved units that are affordable to very-low, low or moderate income households.

It is calculated as:

$$\text{LowIncomeShare}_{j,t} = \frac{\text{VeryLowUnits}_{j,t} + \text{LowUnits}_{j,t} + \text{ModUnits}_{j,t}}{\text{VeryLowUnits}_{j,t} + \text{LowUnits}_{j,t} + \text{ModUnits}_{j,t} + \text{AboveModUnits}_{j,t}}.$$

This variable gives a sense of how much of a city’s production in a given year is oriented toward affordable housing.

- **DensityBonusCount:** The number of applications in that city-year that used a state density bonus. This is computed by counting how many application records have the density bonus flag marked “Yes.”
- **YearFactor:** A categorical version of the year, used to control for statewide changes such as economic shifts or policy changes that affect all cities in a given year.
- **City:** The city identifier, later used for fixed effects.

Data Cleaning and Transformations

A few important steps were needed to prepare the dataset for analysis.

First, many unit-count variables were stored as character strings in the original APR files, so I converted them to numeric and replaced blank entries with zeros when appropriate. Because cities appear in multiple years, some city-years naturally have missing values for PriorApprovedUnits (for example, the first year a city shows up in the dataset). These entries were dropped when necessary since they do not have a lagged value.

Another issue is that APR reporting is not always perfectly consistent across jurisdictions, which means some income category fields occasionally contain missing or incomplete values. Wherever possible, I converted missing numeric entries to zero so the totals would still reflect the overall approved units submitted through that year's applications.

After filtering to cities, aggregating the applications, and constructing the derived variables the final cleaned dataset includes **2,367 city–year observations**.

Alternative or Additional Data Sources

Although the APR dataset is one of the most detailed sources available for housing approvals in California, it does not include everything that might matter for explaining why some cities approve more units than others. For example, zoning information, political data (like city council composition), or regional housing need allocations (RHNA) could all provide additional context. Future work could potentially merge APR data with Census, ACS or local planning datasets to provide a more complete picture.

Key Data Patterns

To understand the basic structure of the dataset before running any models. I created three exploratory visualizations. The first shows the distribution of total approved units across all city-years, which is highly right skewed meaning most cities approve only a small number of units, while a few approve very large amounts. The second graph compares prior-year approvals to current-year approvals, showing a clear positive pattern. The third graph plots the share of affordable units against total approvals, revealing that city-years with higher affordable shares often approve fewer units overall. These patterns helped inform the model choices and give context to the results presented later in the paper.

Methods

To study which factors are associated with the number of housing units a city approves in a given year, I used a regression approach designed for panel data. Since my dataset follows

the same cities over multiple years, I wanted a method that could account for the fact that cities tend to behave in consistent ways over time. For this reason, I estimated two models: a pooled ordinary least squares (OLS) model and a city fixed-effects model with clustered standard errors. The pooled model gives a broad, overall view of the relationships across cities, while the fixed-effects model lets me compare each city to itself over time by holding constant all of the city's unchanging characteristics.

Model 1: Pooled OLS

The pooled OLS model treats each city-year as an independent observation and does not attempt to control for long-term differences across cities. The model is written as:

$$\begin{aligned} \text{TotalApprovedUnits}_{j,t} = & \beta_0 + \beta_1 \text{PriorApprovedUnits}_{j,t-1} \\ & + \beta_2 \text{LowIncomeShare}_{j,t} + \beta_3 \text{DensityBonusCount}_{j,t} \\ & + \gamma_{\text{Year}(t)} + \varepsilon_{j,t}. \end{aligned}$$

Where:

- $\text{TotalApprovedUnits}_{j,t}$ is the total number of housing units approved in city j during year t .
- $\text{PriorApprovedUnits}_{j,t-1}$ is the number of units approved in the previous year for that same city.
- $\text{LowIncomeShare}_{j,t}$ is the fraction of approved units that were affordable.
- $\text{DensityBonusCount}_{j,t}$ is the number of projects using a density bonus.
- $\gamma_{\text{Year}(t)}$ captures statewide time effects using year indicators.
- $\varepsilon_{j,t}$ is the error term.

This model is simple and easy to interpret but it does not address the fact that cities differ in their long-term development patterns.

Model 2: City Fixed Effects with Clustered Standard Errors

Since cities have characteristics that do not change much over time for example, geography, zoning history or local development culture—I also estimated a city fixed-effects model. This model controls for all time invariant differences across cities by giving each city its own intercept:

$$\begin{aligned}\text{TotalApprovedUnits}_{j,t} = & \beta_1 \text{PriorApprovedUnits}_{j,t-1} \\ & + \beta_2 \text{LowIncomeShare}_{j,t} + \beta_3 \text{DensityBonusCount}_{j,t} \\ & + \gamma_{\text{Year}(t)} + \alpha_j + \varepsilon_{j,t},\end{aligned}$$

Where:

- α_j is the fixed effect for city j , capturing everything about that city that stays the same over time.
- Standard errors are *clustered by city* to account for correlation among observations within the same city over multiple years.

This model lets me focus on how approval activity changes within a city from one year to the next, rather than comparing high-building cities to low-building cities.

Assumptions

Both models rely on common regression assumptions, including:

1. **Linearity:** Predictors influence the expected approval counts in linear ways.
2. **No perfect multicollinearity:** Predictors should not be exact combinations of one another.
3. **Mean-zero errors:** The error term should not systematically favor higher or lower approval counts.
4. **Independence across cities:** One city's approval activity does not directly affect another city's in the model.
5. **Cluster-robust errors (Model 2):** Observations within the same city may be correlated and clustering compensates for this.

These assumptions seem reasonable for this dataset, especially after reviewing the exploratory visuals and checking changes across models.

Model Selection and Validation

I estimated both the pooled OLS and fixed-effects models to see how the results changed when controlling for unobserved city characteristics. Comparing the two models helped me understand whether the relationships were driven by differences across cities or by changes

within cities. The fixed effects model had a much higher overall R^2 and produced more consistent interpretations after controlling for city specific patterns, so I chose it as the primary model for discussing results.

For validation, I checked whether the estimated coefficients matched the patterns seen in the Data section, whether the signs made sense in context and whether standard errors behaved as expected once clustering was added.

Software

All analysis was done in **R** (R Core Team 2025) using the following packages:

- **fixest** for fitting fixed-effects models and computing clustered standard errors (Bergé (2023)),
- **dplyr** for data cleaning and transformation (Wickham et al. (2023)),
- **ggplot2** for creating all visualizations (Wickham (2016)),
- **readr** for loading the APR dataset (Wickham, Hester, and Bryan (2024)).

Limitations

As with any observational analysis, the results represent associations rather than causal effects. Important city-level factors—such as zoning rules, RHNA obligations or political sentiment toward development are not included in the dataset. The APR data may also contain reporting inconsistencies across cities or years. Despite these limitations, the models still highlight meaningful patterns in how affordable housing share and prior approval activity are associated with annual housing approvals.

Results

The goal of this analysis is to understand which factors are related to how many housing units a California city approves in a given year. To study this question, I fitted two regression models: a pooled OLS model and a city fixed effects model with clustered standard errors. The pooled model gives a broad view across cities, while the fixed effects model allows me to compare each city to itself over time by holding constant all long-standing city characteristics. Because the fixed effects model most accurately reflects the structure of the data. I focus on that model when interpreting the results.

Below is a summary table showing the key coefficients from the fixed-effects model.

Regression Table (Fixed-Effects Model)

Parameter	Estimate	Std. Error	t-statistic	p-value
PriorApprovedUnits	-0.107	0.054	-1.99	0.047
LowIncomeShare	-91.303	28.581	-3.19	0.0015
DensityBonusCount	3.028	17.799	0.17	0.865

These results show how each variable is associated with the number of units approved in a city-year after controlling for year effects and city specific characteristics.

Interpretation of Model Estimates

PriorApprovedUnits has a small negative and statistically significant coefficient. This means that when a city has an unusually high approval year, the following year tends to dip slightly below average. Instead of showing strong year-to-year momentum, approval activity seems to “settle back” toward a typical level for each city.

LowIncomeShare is strongly negative and highly significant. City-years with a higher share of affordable units tend to approve fewer total units. One possible explanation is that affordable housing projects require more coordination, funding and time, which may reduce the total number of units a city approves in that year. This relationship also appeared clearly in the exploratory scatterplot from the Data section.

DensityBonusCount is not statistically significant in the fixed effects model. Although the pooled model suggested a positive relationship, this effect disappears once city fixed effects are added. This implies that density bonuses are used more often in cities that already build more housing, rather than density bonuses themselves increasing year-to-year approvals.

Visual Summary of Patterns in the Data

A four panel visualization helps illustrate these relationships. The distribution of total approvals is heavily right-skewed, with a few cities approving very large numbers of units. The scatterplot of prior-year versus current year approvals shows a positive pattern overall, but the fixed-effects model reveals that this relationship does not hold once comparing cities to themselves. The plot of affordable unit share versus total approvals reinforces the negative relationship seen in the regression. Finally, the time trend shows a clear increase in approvals in 2022 followed by a drop in 2023, which matches the estimated year effects.

Putting all these pieces together, the results give a pretty good idea of what drives yearly differences in housing approvals.

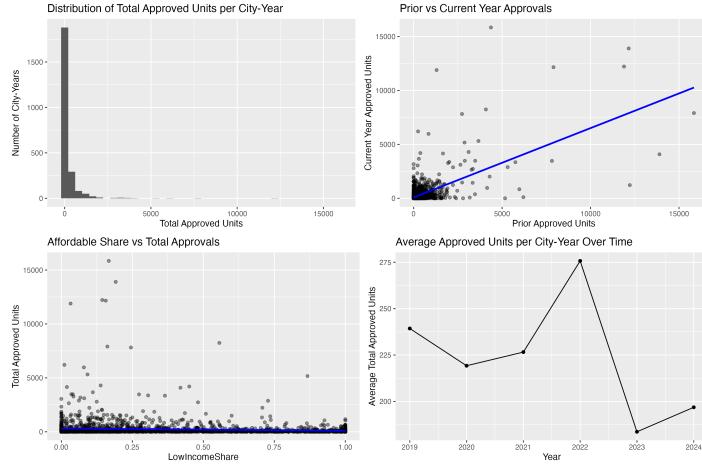


Figure 1: Four-panel visualization of housing approval patterns across California cities.

Connecting Back to the Research Question

This study asked: *Which factors are associated with how many housing units a city approves each year?*

Based on the fixed-effects model, two main conclusions emerge:

1. **Cities that approve a higher share of affordable housing tend to approve fewer total units.**

This is consistent across all models and visualizations and may reflect the added complexity and resources required for affordable housing production.

2. **Approval activity does not consistently carry momentum from year to year.**

Cities tend to return to their usual approval levels after an unusually high year, showing a mild mean reversion pattern rather than a strong upward trend.

Overall, these results suggest that city-specific characteristics and local policy decisions both play important roles in shaping housing production. They also highlight the importance of examining how approval behavior changes *within* cities over time rather than only comparing cities to one another.

Discussion

This project set out to understand which factors are associated with how many housing units a California city approves in a given year. Using application level data from the Housing Element Annual Progress Report (APR), I built a city-year dataset and examined how prior approval activity, the share of affordable units, and density bonus usage relate to annual approval totals.

By estimating both a pooled OLS model and a city fixed effects model. I was able to explore not only broad differences across cities but also how approval activity changes within each city over time.

The key findings suggest that affordable housing share and past approval levels play meaningful roles in explaining yearly approval totals. Cities that approve a higher share of affordable units tend to approve fewer total units, which may reflect the added time, funding and coordination required for affordable housing projects. In addition, approval activity shows a mild “mean-reversion” pattern rather than strong momentum. When a city has an unusually high approval year, it tends not to repeat that level the following year. While density bonus usage initially appeared to be associated with higher approval totals, this relationship disappeared after adding city fixed effects, suggesting that cities using density bonuses are often the same cities that already build more housing, rather than density bonuses themselves driving year-to-year increases.

These findings have several implications for the research question. They show that variation in city-level housing approvals cannot be explained by short-term policy tools alone. Instead, approval patterns seem to be shaped by city specific traits and the type of housing being approved in a given year. The results also highlight the importance of examining within city changes rather than only comparing cities to each other. Looking at cities over time provides a more nuanced understanding of approval behavior and helps explain why some statewide policies may not have uniform effects across jurisdictions.

At the same time, this study has a few limitations. First, the APR dataset is observational, so the associations found here should not be interpreted as causal. Second, several important factors—such as zoning regulations, local political attitudes, developer capacity, and RHNA targets—were not available in the dataset but likely influence approval decisions. Third, APR reporting may vary slightly across cities and years, which can introduce measurement inconsistencies. Finally, the fixed-effects model captures time-invariant city characteristics but does not identify what those characteristics actually are.

Future research could extend this work in a few different ways. One direction would be to incorporate zoning or planning data to better understand the structural constraints that affect approval patterns. Another extension could examine the role of local politics or administrative capacity in shaping housing approvals. Merging APR data with economic conditions, demographic changes or RHNA progress might also provide a more complete picture of why approval levels fluctuate from year to year. Exploring these additions would help clarify the underlying mechanisms driving city level housing production and offer stronger guidance for policymakers working to address California’s housing shortage.

References

- Bergé, Laurent. 2023. *Fixest: Fast Fixed-Effects Estimation Tools*. <https://cran.r-project.org/package=fixest>.
- California Department of Housing and Community Development. 2025. “Housing Element Annual Progress Report (APR) Data.” <https://data.ca.gov/dataset/housing-element-annual-progress-report-apr-data-by-jurisdiction-and-year>.
- R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.