**HOUSING MARKET AND MEDIAN INCOME:**

**STATE-LEVEL AFFORDABILITY PANEL (2008–2023)**

**Michael L Monroe**

**Western Governors University**

**Table of Contents**

[**A. Project Overview 2**](#_dizjw960zj5b)

[**A1. Research Question or Organizational Need 2**](#_mv9sbqhsz3et)

[A2. Context and Background 2](#_436rln1fbtfl)

[A3. Summary of Published Works 3](#_fvkuacps21a1)

[A4. Summary of Data Analytics Solution 5](#_ligm769w1o5l)

[A5. Benefit to Organization and Decision-Making Process 6](#_9hispl2or6kf)

[**B. Data Analytics Plan 7**](#_5ughqiblfik)

[B1. Goals, Objectives, and Deliverables 7](#_4hkijvlqxxec)

[B2. Scope of Project 8](#_xxt8pbhqt4mp)

[B3. Standard Methodology 9](#_jf7qjaeak5pa)

[B4. Timeline and Milestones 11](#_8vl5ap33km1x)

[B5. Resources and Costs 12](#_br0wsejp3cj1)

[B6. Criteria for Success 12](#_wetxe6oa4s7)

[**C. Design of Data Analytics Solution 16**](#_mpkl7xvpi9k1)

[C1. Hypothesis 16](#_3dfd4dp058pi)

[C2. Analytical Method 16](#_lhb2kh3l0fly)

[C2a. Justification of Analytical Method 18](#_ghy7vquqrl18)

[C3. Tools and Environments of Solution 19](#_f5xf5g7mdlj4)

[C4. Methods and Metrics to Evaluate Statistical Significance 20](#_x9g3gdvfh8z6)

[C4a. Justification of Methods and Metrics 20](#_3087k4inp891)

[C5. Practical Significance 20](#_mih9zyplo7de)

[C6. Visual Communication 22](#_qpdawjlkvd8g)

[**D. Description of Datasets 22**](#_wc4yrxjssnw3)

[D1. Source of Data 22](#_ocfrxheini0h)

[D2. Appropriateness of Dataset 23](#_o1yoohs2dq8y)

[D3. Data Collection Methods 23](#_7magr2uw7nvj)

[D4. Data Quality 23](#_u9h9om390rc3)

[D5. Data Governance, Privacy and Security, Ethical, Legal, Regulatory Compliance 24](#_jjpkb417099r)

[D5a. Precautions 24](#_iwn3fsq58gsx)

[**E. Sources 25**](#_9u60wfd9csxa)

## 

## A. Project Overview

## A1. Research Question or Organizational Need

How has owner-occupied housing affordability changed in each U.S. state from 2008 to 2023 when combining three core drivers (median sale price, 30-year fixed mortgage rate, and median household income) into a single monthly principal and interest (P&I) payment burden percentage? The project provides a clear, reproducible state-level benchmark that replaces fragmented price-only or income-only commentary with an integrated affordability metric that stakeholders can compare across time and geography.

### A2. Context and Background

Public housing discussions often cite single metrics (price spikes, stagnant wages, higher rates) in isolation. Buyers actually experience affordability through the monthly mortgage payment, which depends on three things together: home price (loan size), interest rate (financing cost), and household income (capacity). The sharp rate rise after historically low pandemic levels pushed payments up even where price growth slowed. Stress has also spread beyond traditional coastal high‑cost states into interior regions. A clean, reproducible state panel from 2008 to 2023 (covering the post‑recession recovery, low‑rate era, and the recent rate shock) lets stakeholders see how payment burden changed, which states deteriorated most, and how many crossed standard burden thresholds.

### A3. Summary of Published Works

Harvard Joint Center for Housing Studies (2023) details how the rapid rise in mortgage rates after a prolonged low-rate period collided with already elevated post‑pandemic prices, producing a sharp jump in required monthly payments. The report also notes that worsening affordability is no longer confined to traditional high-cost coastal states; interior markets in the Mountain West and parts of the Sun Belt experienced some of the fastest recent stress increases. The report directly supports the nationwide scope of the project (all states instead of a coastal sample), justifies choosing an interval that includes the 2021–2023 rate shock, and reinforces focusing on a payment-based affordability construct (monthly principal and interest relative to income) rather than tracking sale prices alone. By highlighting geographic diffusion, the report underpins the decision to rank states by cumulative burden change and to count how many moved above chosen burden thresholds (Harvard Joint Center for Housing Studies, 2023).

The U.S. Department of the Treasury (2023) frames affordability as a multidimensional outcome shaped jointly by housing prices, financing costs, incomes, demographic trends, and supply constraints. This framing validates integrating price, mortgage rate, and income into a single composite burden metric rather than presenting separate unlinked charts. It also motivates transparent, reproducible transformation steps (aggregation of monthly prices to annual medians, annual median mortgage rate, and direct formula-based payment derivation) so that each component’s contribution can be inspected or updated. Treasury’s emphasis on structural drivers supports articulating limitations (for example, exclusion of taxes and insurance) and focusing inferential testing narrowly on whether average payment burden increased and whether the count of high burden states expanded, instead of attempting unsupported causal claims (U.S. Department of the Treasury, 2023).

Pew Research Center (2022) demonstrates the communication effectiveness of the percent of income framing for housing cost pressure and documents substantial interstate variation. This evidence supports expressing results as a share of median household income and selecting commonly recognizable burden bands (20, 25, 30 percent) for annual threshold counts. Pew’s focus on accessibility informs the inclusion of clear footnotes defining burden (principal and interest only) and the choice to present both continuous (point change) and categorical (threshold crossing) views, making the findings interpretable for non-technical stakeholders while preserving analytic rigor (Pew Research Center, 2022).

Collectively, these works justify the core design choices: nationwide panel coverage (Harvard diffusion finding), a composite monthly payment burden metric integrating prices, rates, and income (Treasury multidimensional framing), and communication through percent of income levels with threshold tracking (Pew clarity and interstate variation). They also support the selected timeframe (capturing the recent rate shock), the reliance on transparent arithmetic derivations instead of complex modeling, and the decision to evaluate both statistical (paired burden increase) and practical (threshold proliferation, ranking) significance aligned with the stated research question about how affordability has changed across states.

### A4. Summary of Data Analytics Solution

The project builds a 2008 to 2023 state year dataset combining annual median sale price (aggregated from monthly Redfin metro data), the yearly median national 30 year fixed mortgage rate (median of weekly FRED observations), state median household income (Census/FRED), and a calculated monthly principal and interest payment using the standard 30 year mortgage formula (360 months, full financing assumption). These components derive a burden percentage (monthly P&I divided by monthly median household income times 100). The analysis produces a time series of the burden percentage for every state, a first versus last year point change (2008 vs 2023, or first available vs last), year over year burden changes to show acceleration periods, annual counts of states at or above 20, 25, and 30 percent burden to track breadth of elevated pressure, a ranking of states by cumulative burden increase, and an exploratory comparison of cumulative price growth versus burden change to indicate where payment pressure outpaced price movement alone. All steps (ingestion, reshaping, aggregation, payment calculation, burden derivation, change and threshold computations, ranking, comparison view generation, and exports) run in a reproducible Jupyter notebook that outputs the final CSV, supporting tables, a data dictionary, and files prepared for Tableau dashboards.

### A5. Benefit to Organization and Decision-Making Process

Decision makers gain a transparent and auditable affordability benchmark that reduces reliance on selective indicators and allows faster price or rate shock claims validation. Realtors and housing advisors can frame individual client challenges against a consistent historical baseline and show how today's burden compares to low-rate and high-rate periods. Lenders can monitor aggregate borrower payment stress trajectories across jurisdictions, flag emerging concentration risk, and refine underwriting focus or proactive counseling. Policymakers and housing advocates can spot rapid deterioration clusters early and target supply incentives, zoning attention, down payment and interest buydown assistance rather than leaning on broad national averages. Prospective homeowners and relocating households receive a standardized comparative frame for evaluating cross-state affordability evolution and setting realistic expectations about payment-to-income trade-offs. Media and public communicators gain a straightforward method that reduces misinterpretation and separates structural trends (income lag, sustained price level) from episodic volatility (short-term rate swings). Organizationally, the reproducible design minimizes update costs because the annual refresh only requires downloading updated public series and re-running the notebook, supporting sustainable longitudinal monitoring and straightforward third-party replication. Academic and nonprofit analysts can also extend the dataset by layering taxes or insurance while preserving the core payment to income spine, reinforcing its role as a common baseline.

## B. Data Analytics Plan

### B1. Goals, Objectives, and Deliverables

This project aims to build a clear, reproducible state-level housing affordability dataset and a set of visuals that show how payment burden (principal and interest share of income) changed from 2008 to 2023.

The objectives for this goal are:

* Integrate core data (median sale price, 30-year fixed mortgage rate, median household income) and compute the monthly payment and burden percent.
  + The deliverable for this objective is a cleaned state-year dataset plus a data dictionary.
* Show each state’s annual burden trend.
  + The deliverable for this objective is a Tableau line dashboard (state trend view).
* Show how the burden changed between the first and last year and year over year.
  + The deliverable for this objective is a summary change table.
* Count how many states fall at or above the 30 percent burden yearly.
  + The deliverable for this objective is a threshold count table.
* Compare cumulative price growth to burden change to highlight rate and income effects.
  + The deliverable for this objective is an exploratory scatter plot (Tableau or notebook).
* Could you provide a fully reproducible workflow?
  + The deliverable for this objective is an annotated Jupyter notebook that produces all outputs.

### B2. Scope of Project

B2. Scope of Project

This project's scope will include building a reproducible state-level housing affordability dataset (2008 through 2023) using publicly available sources. The dataset will combine annual median home sale price (aggregated from monthly Redfin metro data to state year), the yearly median national 30-year fixed mortgage rate (median of weekly FRED observations), and state median household income (FRED/Census). It will calculate a standard 30-year fixed principal and interest monthly payment (full financing assumption) and the payment burden percentage (monthly principal and interest divided by monthly median household income). The outputs in scope are a cleaned state year CSV file, an annotated Jupyter notebook that performs all transformations, a data dictionary, and Tableau workbooks.

This project's scope will not include taxes, insurance, HOA fees, mortgage insurance, maintenance costs, rental market analysis, or alternative affordability indices. The scope will not involve sub-state geographies (county, metro drill downs) or demographic segment breakdowns. The scope will not include forecasting beyond the latest completed year, machine learning models, policy simulation, down payment scenario modeling, or multiple loan product types. The scope will not include statistical hypothesis testing, inflation adjustment of prices or incomes, cost of living adjustments across states, or construction of a web or mobile application beyond the Tableau workbook. The scope will not include monthly-level final reporting or detection of data for years after 2023.

### B3. Standard Methodology

Standard Methodology: Iterative CRISP-DM:

**Business Understanding**

Define the central question: how state-level mortgage principal and interest payment burden (share of household income) changes over the selected baseline period. Set success criteria: integrated dataset with documented fields, required analytical tables and visualizations, reproducible notebook, concise methodology description. Record assumptions (full financing, principal and interest only) and exclusions (taxes, insurance, fees, sub-state geographies, forecasting, alternative loan products, inflation adjustment). Confirm threshold bands after initial data profiling.

**Data Understanding**

Acquire raw source files (home price source, mortgage rate series, household income series). Inspect structure (column names, types, date formats, identifiers), temporal coverage, presence of all required years and states, and basic completeness. Log structural anomalies such as missing state year rows or malformed dates. Do not compute or publish summary statistics in this proposal text.

**Data Preparation**

Reshape and standardize inputs: convert wide monthly price data to long form, map metros to states, aggregate to annual state-level median (method documented), reduce weekly mortgage rate observations to a single yearly value, reshape income series to state-year. Join the state and year for the approved baseline. Create derived fields: monthly mortgage payment (standard 30-year amortization), monthly income (annual income divided by 12), payment burden percent (monthly payment divided by monthly income times 100), and placeholder fields for year-over-year change, first vs last change, and threshold flags. Preserve a precise sequence of steps in the notebook so each derived field is traceable.

**Modeling**

No predictive or machine learning models. Modeling here means applying deterministic formulas and structured calculations. Encapsulate the mortgage payment formula in a function. Only accepted calculations move into the primary workflow before definitions are frozen. Clarity and reproducibility take priority over performance tuning.

**Evaluation**

Apply a concise checklist before finalizing:

1. Completeness: all expected state year combinations present after exclusions.

2. Mandatory fields: no null values in core fields (state, year, price, rate, income, derived payment, burden).

3. Formula accuracy: spot check a small sample of monthly payment calculations manually or with an external calculator.

4. Logical behavior: burden responds directionally to combined changes in price and rate (qualitative inspection only).

5. Join integrity: no unintended duplication or loss of states after merges.

**Deployment**

Package fixed submission artifacts: processed dataset, annotated notebook, data dictionary, visualization assets or workbook file, and a brief methodology summary. Deployment means assembling these static deliverables for grading.

### B4. Timeline and Milestones

| Milestone | Start Date | End Date | Duration |
| --- | --- | --- | --- |
| Data acquisition and logging | 9/05/2025 | 9/05/2025 | 1 Day |
| Price and rate aggregation | 9/06/2025 | 9/06/2025 | 1 Day |
| Income integration and derivations | 9/06/2025 | 9/06/2025 | 1 Day |
| Validation | 9/07/2025 | 9/07/2025 | 1 Day |
| Change metrics and threshold counts | 9/07/2025 | 9/07/2025 | 1 Day |
| Visualization development | 9/08/2025 | 9/08/2025 | 1 Day |
| Report drafting and executive summary | 9/08/2025 | 9/08/2025 | 1 Day |
| Refinement, editorial, and visual polishing | 9/09/2025 | 9/09/2025 | 1 Day |
| Final packaging and submission | 9/10/2025 | 9/10/2025 | 1 Day |

### B5. Resources and Costs

Public data sources (Redfin, FRED, Census) have zero direct cost. The computational environment (Python, Jupyter, Tableau Public) is freely accessible. Hardware requirements are limited to a standard personal computer with sufficient memory to manipulate moderate-sized CSVs. Estimated labor is approximately 50 hours allocated across acquisition, transformation, validation, analysis, visualization, and documentation. No proprietary licensing, paid APIs, or hosting expenditures are anticipated.

### B6. Criteria for Success

**Data Source Confirmation**

All declared public sources (home price file, mortgage rate series, household income series) are cited and present in raw form exactly as downloaded (no overwriting, no manual edits). Each source has a brief description (file name, original URL, basic structure).

**Data Coverage Completeness**

The approved baseline period years appear for every included state. Total row count equals states multiplied by years after any documented exclusions.

**Structural Consistency**

The final dataset contains only the needed fields. Column names match the data dictionary. Data types are appropriate (numeric vs string).

**Transformation Traceability**

Every derived field has a plain formula or description in the data dictionary and an explicit code cell that creates it; no hidden transformations. Notebook reads top to bottom without reliance on prior hidden state.

**Mortgage Payment Calculation Integrity**

A single amortization function computes the monthly payment (principal = median sale price, rate = annual 30-year fixed, term = 360). Function parameters are clearly stated.

**Absence of Unresolved Nulls**

Mandatory fields have no null values in the final dataset. Any dropped rows due to missing upstream values are noted.

**Scope Adherence**

No out-of-scope items appear: taxes, insurance, HOA, PMI, forecasts, sub-state geographies, inflation adjustments, demographic splits, alternative loan products, down payment scenarios, or machine learning models. Exploratory branches not adopted are removed or labeled as excluded and do not feed outputs.

**Visualization Coverage**

Deliverables include:

* State trend of annual burden percent
* First vs last year change view or table
* Year-over-year change representation
* Threshold classification view (if thresholds retained)
* Ranking of cumulative burden change
* Comparative price change vs burden change (scatter or paired visual)

**Reproducible Notebook**

Running the notebook from scratch with raw files in place produces the final dataset and tables without manual intervention. Paths are relative and documented.

**Data Dictionary Completeness**

Each final dataset column has a name, plain definition, source or calculation, unit, and allowed value pattern if categorical; no sample numeric values are required.

**File Export and Naming**

The final dataset will be saved as final.csv in the designated processed (or project) directory. Notebook is clearly named as the primary workflow (e.g., main.ipynb). Data dictionary and methodology summary are present.

**Internal Consistency Checks**

All monthly payments are positive. All burden percentages are positive. The year-over-year change field (if present) equals the current year burden minus the prior year's burden; there are no circular or contradictory fields.

**Submission Package Integrity**

All required artifacts (raw data directory, final dataset, notebook, data dictionary, visualization workbook or images, methodology summary) are present, consistently named, and free of unrelated files.

**Ethical and Compliance Confirmation**

No human subjects data. No proprietary or restricted data. Will provide links to the data sources.

## C. Design of Data Analytics Solution

### C1. Hypothesis

The overall principal and interest burden (share of median household income) is higher in 2023 than in 2008. More states reached or exceeded the 30 percent burden level in 2023 than in 2008. States with larger median price growth from 2008 to 2023 show larger increases in burden.

### C2. Analytical Method

**Method**: descriptive time trend analysis of state‑level principal and interest (P&I) burden across the selected years.

**Core dataset construction**

* Start from raw public files: state annual median sale price (aggregated from monthly metro data), yearly mortgage rate (collapsed to one value per year), and state median household income (annual).
* Compute monthly payment using a single amortization function (30‑year, full financing, principal = median sale price, rate = annual mortgage rate).
* Derive monthly income (annual / 12) and burden percent (monthly payment / monthly income \* 100).
* Retain only required fields: state, year, price, rate, monthly payment, annual income, burden percent, and simple change fields.

**Deterministic transformations**

* Year-over-year burden change (current year burden minus prior year).
* Start vs end period burden change (first comparison year vs latest).
* Optional threshold flags (e.g., burden ≥ 20%, ≥ 25%, ≥ 30%).
* Ranking lists: (a) latest burden level, (b) total burden point increase.
* Cumulative price change vs burden point change pair for visual comparison (visual inspection only).

**Threshold and distribution views**

* Count states at or above the fixed burden thresholds for baseline and the latest year.
* Show change direction (increased, decreased, unchanged) in burden per state between anchor years.

**Visual outputs**

* Multi‑year burden trend lines per state (or small multiples / interactive filter).
* Start vs end burden change bar or table.
* Year-over-year change bar or slope view.
* Threshold distribution (counts by band).
* Scatter: cumulative price change vs burden point change.

**Quality and traceability**

* Every derived field is produced in a visible code cell in linear order.
* Simple spot checks: manually recompute at least three monthly payments and corresponding burden percentages and confirm exact equality with code output.
* No probabilistic tests, confidence intervals, p‑values, or tolerance bands.

**Output artifacts**

* Enriched state‑year CSV (final.csv).
* Supporting tables for rankings, threshold counts, and change metrics (separate exports or reproducible notebook cells).
* The visualization workbook / exported images are consistent with the fields above.
* Data dictionary documenting each final column and formula.

### C2a. Justification of Analytical Method

The goal is to show how principal and interest burdens have changed over time, how much, and how broadly across states—not to predict future values or make inferential claims. A descriptive time trend analysis with simple, repeatable formulas (monthly payment and burden percent) gives a direct, transparent answer: point levels by year, point changes between anchor years, and counts of states crossing fixed burden levels (e.g., 30%). These outputs map one‑to‑one to the research question's three needs: direction (up or down), magnitude (point difference), and geographic breadth (how many states and which ones).

Using only deterministic transformations avoids the ambiguity and extra assumptions of statistical testing or modeling. Point differences are significant enough to interpret plainly without p‑values, confidence intervals, or tolerance bands. Fixed threshold counts (≥20, ≥25, ≥30 percent if used) add a categorical lens consistent with standard public affordability framing, making the findings easier for non-technical reviewers to grasp. Listing state rankings and start‑to‑end changes highlights distribution spread without invoking distributional assumptions.

The side‑by‑side (scatter) view of cumulative price change versus cumulative burden change is qualitative; it helps visually assess whether price growth and burden shifts move together without asserting causality or estimating a coefficient. This restraint aligns the analysis with available data quality and the project's non-inferential success criteria.

This method minimizes complexity risk, is fully reproducible from raw public inputs, and maintains traceable logic that any reviewer can audit line by line. It maximizes interpretability and stakeholder trust while directly answering the research question.

### C3. Tools and Environments of Solution

Python (pandas) under Jupyter Notebook handles ETL, transformations, and the combination of datasets. Tableau Public provides interactive multi-view dashboards for stakeholder exploration.

### C4. Methods and Metrics to Evaluate Statistical Significance

No statistical significance tests will be used. The project relies on simple descriptive outputs from the final state year dataset: (1) point levels in anchor years, (2) point changes (later year minus earlier year), (3) counts of states at or above a fixed burden threshold (for example 30 percent), (4) lists of states increasing, decreasing, or unchanged, and (5) one scatter plot comparing cumulative price change to burden change for visual pattern checking. I will describe a result as supported when the deterministic metrics plainly show the observed directional and magnitude conditions.

### C4a. Justification of Methods and Metrics

We use simple descriptive comparisons because the dataset covers all states, so inferential statistics are unnecessary. Point levels, point changes, threshold counts, and a visual price vs burden check directly answer direction, size, and breadth of affordability shifts with transparent formulas anyone can recalculate.

### C5. Practical Significance

The findings are interpreted in plain point and dollar terms rather than in statistical significance. A burden point change (percentage of monthly income going to principal and interest) is directly meaningful because it reallocates household cash flow away from other essentials or savings.

Practical significance is established with four descriptive lenses:

* Magnitude of burden shifts: States showing increases of 5, 10, 15, or more points between the anchor years signal escalating affordability strain. Double-digit gains move households from manageable to constrained budgeting in standard housing guidelines.
* Threshold crossings: States moving into or further above a 30 percent payment-to-income level enter a commonly recognized cost burden zone used by housing policy stakeholders. Each additional state in that zone broadens the geographic reach of affordability stress.
* Dollar translation: Converting the burden percent back to an approximate monthly payment (payment formula already in the dataset) shows the extra outlay a typical state median household must allocate each month compared with the earlier anchor year. This frames results in amounts a non-technical audience can grasp.
* Breadth and clustering: Counting how many states increase and how many experience double-digit increases, and listing the top and bottom movers, indicates whether the issue is isolated or systemic. A widespread upward shift has greater practical relevance for national or regional policy attention than a small cluster of outliers.

Because the dataset covers all states, any sizable and widespread upward movement is decisive for practical interpretation. No additional inferential machinery is needed to argue real-world impact. Results will be summarized in the report as plain statements (for example, Number of states with a 10-plus point increase: X; States now at or above 30 percent: Y vs Z previously) so stakeholders can directly connect the metrics to budget pressure, potential increased housing instability risk, and the need for supply, financing, or income interventions.

### C6. Visual Communication

Visuals will be strictly descriptive, built from final.csv, and focused on three viewer needs: direction, magnitude, and breadth of affordability change. A ranked horizontal bar chart shows state point changes (latest minus earliest), and a small multiple (or interactive single) line view lets users see each state's burden path on a standard y scale for truthful comparison. A map colored by latest burden bands (<20, 20–24.9, 25–29.9, 30%+) and a companion column chart of the count of ≥30 percent states over time communicate geographic spread and threshold escalation. A before vs after distribution (paired histograms or box/strip) reveals whether the entire distribution shifted upward, not just extremes. A scatter of cumulative price percent change vs burden point change offers a qualitative structural check with median crosshairs and concise annotations for notable outliers, while a compact table lists top movers with start, end, point change, and approximate added monthly dollar burden (percent change times monthly income) to translate percentages into household impact.

## D. Description of Datasets

### D1. Source of Data

Redfin Data Center: Metro-level monthly median sale prices transformed into annual state medians. FRED: Weekly 30‑year fixed mortgage rate series (MORTGAGE30US) aggregated to annual medians. Census/FRED: State median household income time series (MEHOINUSxxA672N). Supplemental contextual (non-numeric) references: Harvard JCHS 2023, Treasury 2023, Pew 2022.

### D2. Appropriateness of the Dataset

Each data component is essential to the composite burden construct: sale prices determine principal magnitude, influencing payment size; mortgage rates set financing cost; median household income captures payment capacity at the state level. The triangulation avoids distortions inherent in single-factor analyses and improves inter-state comparability over time.

### D3. Data Collection Methods

Manual CSV retrieval from publicly accessible portals: Redfin monthly matrix, FRED weekly mortgage rates, and individual state income series. Income series were batch-downloaded, renamed by parsing state abbreviations from series codes, merged on observation dates, and converted to long format. A retrieval log stores source URLs, access timestamps, and SHA256 checksums to preserve integrity and lineage.

### D4. Data Quality

Coverage yields 799 integrated state-year observations (one missing price record for a single state-year). After pruning that row, derived fields (monthly payment, burden %) are complete. Numeric ranges are plausible: burdens remain <60% (consistent with P&I-only scope), rate series reflect widely reported trough and spike, and price dynamics align with known national patterns. Annual medians dampen intra-year noise while retaining macro shifts; national rate usage is documented as a limitation, but acceptable given the absence of uniform state-level public rate data.

### D5. Data Governance, Privacy and Security, Ethical, Legal, Regulatory Compliance

Governance: Structured directory segmentation (raw, processed, notebooks, docs, figures), retrieval log, and data dictionary ensure traceability and semantic clarity. Privacy: All inputs are aggregate public series with no personally identifiable information. Security: Local-controlled storage plus version control; integrity via checksums; no secret keys required. Ethical: Clear disclosure that the affordability metric excludes non-P&I ownership costs to prevent misinterpretation as total cost. Legal/Regulatory: Public licensing terms permit academic reuse with attribution; APA-formatted citations are supplied; the derived dataset is redistributed only with proper source acknowledgement.

### D5a. Precautions

Governance Precaution: Maintain an immutable retrieval log and tag repository at submission for reproducibility. Privacy Precaution: Explicitly affirm absence of PII in documentation. Security Precaution: Exclude unnecessarily large binary artifacts; use .gitignore for transient files. Ethical Precaution: Prominent footnotes on P&I scope in every visualization to deter overextension of conclusions. Legal Precaution: Retain the original field naming provenance references and cite all data sources. Quality Precaution: Automated schema, missingness, and range assertions execute before generating dashboards. Interpretive Precaution: Provide a straightforward narrative distinguishing the payment burden from the total ownership cost. Maintenance Precaution: Include annual refresh instructions (update source files, re-run notebook) to minimize divergence over time.

## E. Sources

Federal Reserve Bank of St. Louis. (2024). 30-year fixed rate mortgage average in the United States (MORTGAGE30US). <https://fred.stlouisfed.org/series/MORTGAGE30US>

Harvard Joint Center for Housing Studies. (2023, June 21). A new report shows that housing costs strain owners and renters alike, and millions are priced out. <https://www.jchs.harvard.edu/press-releases/new-report-shows-housing-costs-strain-owners-and-renters-alike-millions-priced-out>

Pew Research Center. (2022, March 23). Key facts about housing affordability in the U.S. <https://www.pewresearch.org/short-reads/2022/03/23/key-facts-about-housing-affordability-in-the-u-s/>

Redfin. (2024). Redfin Data Center: Metro\_median\_sale\_price\_now\_uc\_sfrcondo\_month.csv. <https://www.redfin.com/news/data-center/>

U.S. Census Bureau. (2024). State median household income series (accessed via FRED). <https://fred.stlouisfed.org/>

U.S. Department of the Treasury. (2023, June 27). Rent, house prices, and demographics. <https://home.treasury.gov/news/featured-stories/rent-house-prices-and-demographics>