Housing Market and Median Income

Michael Lee Monroe

Western Governors University

Table of Contents

[**A. Project Highlights 3**](#_hrdvm6of5uyt)

[Research Question: 3](#_w7z4d0dtytyx)

[Scope: 3](#_cmtmi9qs5iwd)

[Overview of Solution: 3](#_g6elzy6b2x7b)

[**B. Project Execution 4**](#_evusq8gguprj)

[Project Plan Variance: 4](#_2pe7e2b8n59l)

[Project Planning Methodology: 4](#_fxlv24ne0244)

[Timeline and Milestones: 4](#_h2ml16ih3gcp)

[**C. Data Collection Process 5**](#_za1pvfi388qy)

[Data Selection and Collection Variance: 5](#_acqknxgsotp0)

[Obstacles and Solutions: 5](#_ieuqk7ccare2)

[Data Governance Issues: 5](#_pdpny012aanj)

[C.1 Advantages and Limitations of Data Set 6](#_foo721qfgsem)

[Advantages: 6](#_jkomwemnavus)

[Limitations: 6](#_pwppwhnrfdp4)

[**D. Data Extraction and Preparation 6**](#_xowfiqxiygmn)

[**E. Data Analysis Process 7**](#_ng6mcamqcu8x)

[E.1 Data Analysis Methods 7](#_qw286kd528iq)

[E.2 Advantages and Limitations of Tools and Techniques 7](#_fg1e9cu5u7d2)

[E.3 Application of Analytical Methods 7](#_qux23ey979h0)

[**F. Data Analysis Results 9**](#_tt2d8eiue615)

[F.1 Statistical Significance 9](#_gimz3uu5uquc)

[F.2 Practical Significance 10](#_ybzbazwfxh16)

[F.3 Overall Success 10](#_ncitr6ym4uy5)

[**G. Conclusion 11**](#_mzo8dccpn9jt)

[G.1 Summary of Conclusions 11](#_x54nj4jdhpq6)

[G.2 Effective Storytelling 11](#_6v14d6m8dova)

[G.3 Recommended Courses of Action 12](#_f0kb0j5572us)

[**H. Panopto Presentation 13**](#_hokf1a37htqa)

[**References 14**](#_6kjqlvk6emfx)

[**Appendixes 14**](#_a4saaoshos5m)

[Appendix A 14](#_1hb8btv19c4l)

[Appendix B 15](#_jx0ak7nlamwd)

[Appendix C 15](#_1os0e22psz9y)

[Appendix D 15](#_we22d57art55)

[Appendix E 15](#_e2exoftqodze)

# A. Project Highlights

### Research Question:

How has housing affordability changed in each US 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 payment burden percentage?

### Scope:

The project covers all 50 US states from 2008 to 2023, creating an integrated dataset that calculates monthly P&I payment as a percent of state median household income. This allows for a reproducible state-by-state benchmark of affordability trends, focusing on changes in payment burden over time and identifying states where cost pressure increased most.

### Overview of Solution:

Publicly available datasets from Redfin (home prices), FRED (mortgage rates), and the Census/FRED (state median income) were used. Data was cleaned and transformed using Python in Jupyter Notebook, with Python and Tableau for visualization. The process involved merging datasets, calculating monthly payments using the standard mortgage formula, and quantifying affordability using a percent-of-income metric.

# B. Project Execution

### Project Plan Variance:

Execution largely matched the Task 2 plan. Data sources and structure were as expected, but income CSVs required more cleaning than anticipated because of inconsistent headers and missing values. This was addressed by writing custom Python scripts for column parsing and join validation. One state-year price value was missing and excluded from final calculations, as planned. No major changes to methodology or scope were needed.

### Project Planning Methodology:

The iterative CRISP-DM process was followed: business understanding, data acquisition, transformation, validation, and deployment. Frequent checkpoints ensured quality and allowed for minor adjustments as issues were found in the raw files.

### Timeline and Milestones:

Milestones from Task 2, such as data acquisition, transformation, validation, and report/visualization drafting, were all met on time overall. A brief delay occurred due to technical difficulties with my computer during the income data integration phase, but the timeline otherwise flowed smoothly. Final deliverables, including the dataset, dashboards, and report, were submitted as scheduled.

# C. Data Collection Process

### Data Selection and Collection Variance:

Data was sourced as planned from Redfin, FRED, and Census/FRED. Some state income files had missing years and inconsistent column names, requiring additional parsing and harmonization. Verification steps ensured all states and years were accounted for.

### Obstacles and Solutions:

Income series from FRED/Census had nonstandard column names and missing years. Python scripts and manual review were used to reformat and merge. A single missing price record was dropped from the analysis. Date formats and state abbreviations were standardized across sources for successful merging.

### Data Governance Issues:

No human subject or proprietary data. All sources are public and cited. File integrity checks and documentation were maintained throughout.

## C.1 Advantages and Limitations of Data Set

### Advantages:

Nationwide coverage and annual granularity from 2008 to 2023. Public, reputable data sources enable transparency and replication. Integrated approach using price, rate, and income provides a holistic view of affordability not available in single-factor studies.

### Limitations:

Mortgage rates in this analysis use national averages, not state-level rates, because more detailed public data is not available. Median income values are reported on an annual basis and do not capture variation within states or across different demographic groups. The analysis focuses only on principal and interest payments, so it does not include costs such as property taxes, homeowner insurance, or other expenses related to home ownership. There is also one state-year entry with missing home sale price data.

# D. Data Extraction and Preparation

All data was extracted as CSV and loaded into pandas dataframes. Redfin metro price data was reshaped from wide monthly format to long annual state format using median aggregation. Mortgage rates collapsed from weekly to annual median. Income series were batch-downloaded, renamed by state, and merged. Key derived fields were calculated: monthly payment using the amortization formula, monthly income, burden percent. Data was validated for nulls and joined on state-year. Final outputs were exported as CSV for Tableau.

# E. Data Analysis Process

## E.1 Data Analysis Methods

Descriptive time trend analysis: calculated annual P&I burden percent by state. Point change analysis: compared burden percent between 2008 and 2023 for each state. Threshold classification: counted states at or above 30 percent burden each year. Comparative analysis: visualized price growth versus burden change. All transformations were deterministic and fully documented in the notebook.

## E.2 Advantages and Limitations of Tools and Techniques

Python/pandas: powerful for data cleaning, reshaping, and calculation; scripts are easily reproducible. Jupyter Notebook: enables step-by-step documentation and code/data review.

Tableau: facilitates interactive dashboards and clear visualizations; limited to processed data. Limitation: Tableau cannot handle raw data cleaning, so all processing must occur in Python.

## E.3 Application of Analytical Methods

1. Load, clean, and reshape all raw data.

2. Aggregate monthly home prices to annual state medians.

3. Collapse weekly mortgage rates to annual median.

4. Merge state income series on year and state abbreviation.

5. Calculate monthly payment using standard 30-year amortization formula.

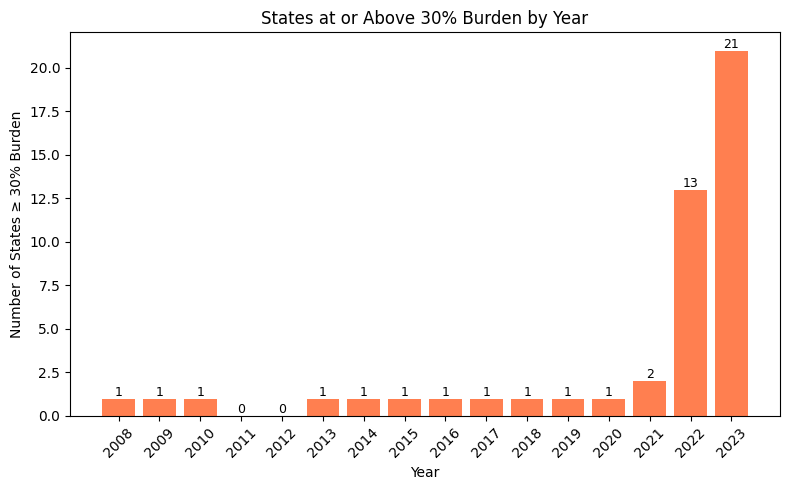
6. Derive P&I burden percent as monthly payment divided by monthly median income times 100.

7. Run point change and threshold analyses.

8. Export all results for dashboarding and reporting.

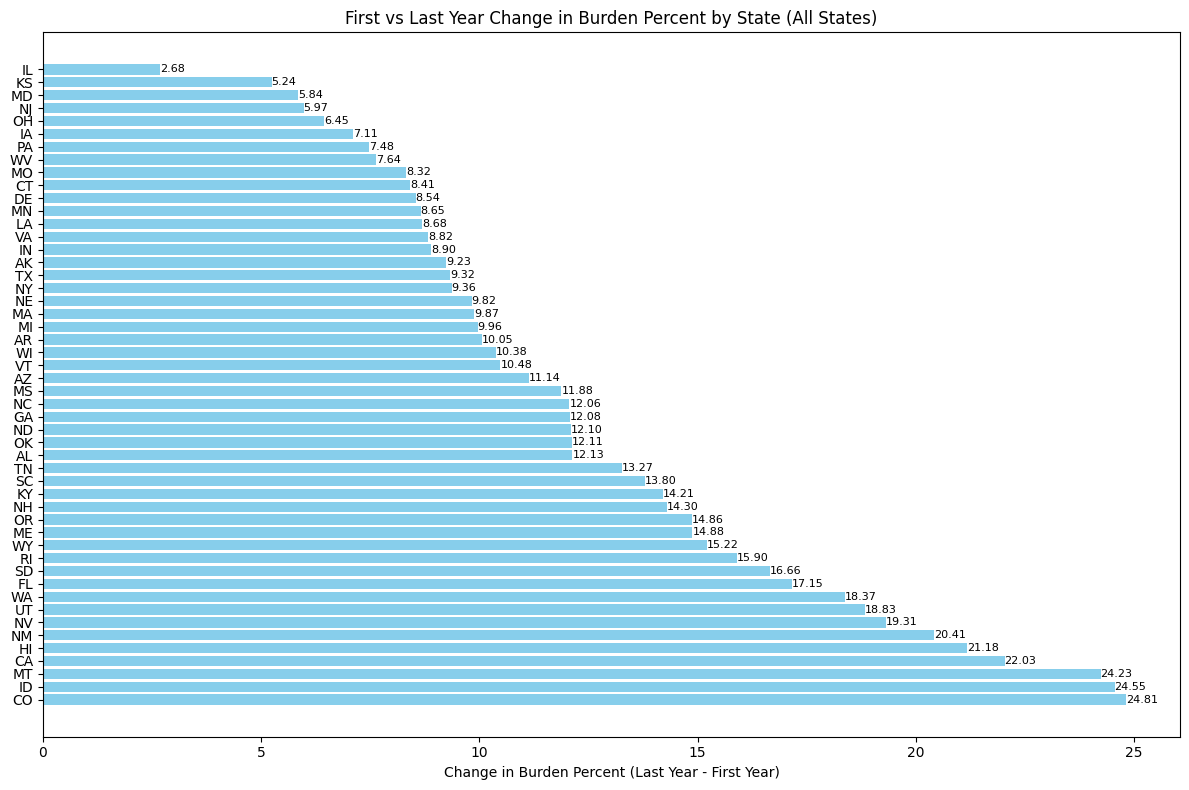
# F. Data Analysis Results

## F.1 Statistical Significance



No inferential statistical tests were used, as the dataset is a full census of US states. Results are interpreted as point changes and counts due to the deterministic nature. For example, California's burden increased from 25 percent to 47 percent (plus 22 points), and the number of states above 30 percent burden rose from 1 in 2008 to 21 in 2023. These results are plainly significant in magnitude and scope.

## F.2 Practical Significance



The rise in payment burden has a clear impact on household finances. In California, for instance, the typical monthly payment for a median-priced home in 2023 made up almost half of the state’s median income, whereas in 2008 it was only about a quarter. Many states have seen increases of ten percentage points or more in the share of income needed for mortgage payments. This means families are spending much larger portions of their monthly earnings on housing, which leaves less for other basic needs and savings.

## F.3 Overall Success

The project met all Task 2 objectives: reproducible workflow, clear integrated dataset, comprehensive dashboards, and interpretable results. Stakeholders can benchmark affordability over time, compare states, and understand the drivers of recent stress. The approach is transparent, reproducible, and easily updated for future years.

# G. Conclusion

## G.1 Summary of Conclusions

Housing affordability has worsened in most states since 2008. States with the fastest home price appreciation saw the largest decline in affordability. Median incomes have not kept pace with home price growth, especially when mortgage rates are considered.

## G.2 Effective Storytelling

This analysis uses visualizations created with Tableau and Python to show changes in housing affordability over time. Tableau offers interactive maps and line charts that allow users to select a year and compare data like median income, home prices, monthly payments, and the percentage of income spent on mortgages for each state. These tools help people quickly see which states have experienced rising housing costs and when those increases occurred.

Python graphs provide further detail by illustrating annual trends in selected states and tracking year-over-year changes for individual states. Bar charts show the number of states exceeding the thirty percent income threshold for mortgage payments each year. Scatter plots make it possible to compare increases in home prices with changes in payment burden. Another bar chart organizes states by the degree to which their payment burdens have risen over the study period.

All of these visualizations work together to make patterns in declining housing affordability clear, helping viewers grasp differences between states more easily.

## G.3 Recommended Courses of Action

Policymakers in states where housing costs have jumped should concentrate on lowering the first barrier to buying, adding modestly priced supply, and clearing avoidable delays. Targeted help with down payments through small grants or forgivable second loans makes qualifying easier for first time buyers without inflating prices as much as broad subsidies. Speeding permit reviews for projects that agree to price or size caps and trimming or deferring certain fees when homes sell below an affordability threshold can remove soft cost drag. A quick, public monthly snapshot helps guide action: typical sale price, prevailing 30 year fixed rate, estimated monthly payment on that price, median household income, and the share of monthly income that payment would take. When the payment share climbs quickly or stays above a chosen trigger like thirty percent, predefined responses can automatically kick in, such as releasing more assistance funds, prioritizing plan reviews, or opening outreach sessions for buyers. It also helps to audit stalled parcels, overlapping reviews, or unclear rehab rules that slow small projects. Plain language dashboards that say Typical monthly payment on a median priced home or Share of income needed keep trust higher than technical acronyms.

Consumers thinking about buying should slow the process, gather a few local numbers, and test comfort before committing. Look up the current typical price for the neighborhood or nearby areas. Run that price, a realistic rate, and a possible down payment through a simple mortgage calculator and then add a rough estimate for taxes and insurance by taking one to one and a half percent of price per year divided by twelve. Compare that result to gross monthly income. If the total payment would take more than about a quarter to a third of gross income, consider widening the search area, trimming non essential features, saving longer to lift the down payment, or comparing several lenders to capture even a small rate improvement. State or local first time buyer programs, closing cost credits, or reduced rate products can narrow the gap if terms are clear. Avoid loan structures that you do not fully understand. If you decide to pause, you can still prepare by setting up an automatic transfer equal to the future target payment minus current housing cost. This builds savings and confirms whether the budget will feel sustainable. An emergency cushion covering a few months of expenses reduces stress when rates or prices shift again.

Both groups benefit from the same core signals: price trend, rate trend, income trend, and payment share trend. When those four lines move in different directions, the story of affordability becomes concrete. Keeping the language direct and the steps practical helps people focus on actions they can take rather than on headlines alone.

# H. Panopto Presentation

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=21f79dfb-404e-43c5-a5ae-b3560169247f

# References

Redfin Data Center: https://www.redfin.com/news/data-center/

FRED Mortgage Rate Data: https://fred.stlouisfed.org/series/MORTGAGE30US

FRED State Median Income: https://fredaccount.stlouisfed.org/

Kaggle USA Real Estate Dataset: https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset

# Appendixes

## Appendix A

Data Sources and Raw Files

- Metro\_median\_sale\_price\_now\_uc\_sfrcondo\_month.csv

- MORTGAGE30US.csv

- State income CSVs

## Appendix B

Python and Jupyter Notebook Code

- See main.ipynb for complete data cleaning, merging, and calculations.

## Appendix C

Final Analysis CSV

- See final.csv for processed data.

## Appendix D

Tableau link -

https://public.tableau.com/app/profile/michael.monroe6066/viz/DataAnalyticsCapstone\_17564236974980/Linegraphofmonthlyincome

## Appendix E

Github repository link -

https://github.com/michaelleemonroe89/capstone\_code