U.S. and the living-cost expenses. Assessing the ability of households to afford house prices.

1. Introduction

1.1. A look into the U.S. states, the households' expenses and their housing affordability.

Since the last few years, the U.S states are suffering the "cost of living crisis": their daily basic expenses are rising faster than their income levels. The U.S. households are facing difficulties in maintaining a good quality of life, particularly when it comes to their ability to afford buying/renting and maintaining a house.[1] This project addresses this exact *sensitive* topic and analyzes the effect of the latest living costs of U.S. residents in this regard. It suggests that there is a strong connection between the essential living expenses of the U.S. households and their ability to afford a house. One main question will be considered by analyzing the underlying patterns of data: How do essential living expenses impact housing affordability for households across the U.S.? Two main data sources were utilized by the ETL Pipeline to uncover data trends.

2. Data

2.1 Data License

Both datasets are sourced from Kaggle and are under the same standard open-data license. Creative Commons Zero v1.0 Universal (CC0 1.0). *Respecting the license, I have uploaded, used and transformed the datasets in a responsible manner, without taking for granted the openness and accessibility it offers.*

2.2 Data Used: Pipeline Output

The cleaned and well-structured data given as output from the ETL Pipeline is stored in an SQLite database named train_data.sqlite. This database consists of two tables, cost_of_living and house_listings, each of which contains relevant columns necessary for the analysis.

2.2.1 Structure of the data: Cost_of_living Table

Cost of Living in the U.S

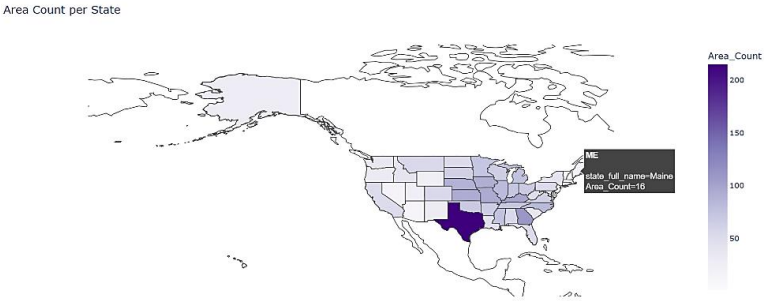
household_id	state	areaname	housing_expenses	food_expenses	transport_expenses	healthcare_expenses	other_necessities_expenses	childcare_expenses	household_taxes	median_family_income
1	AL	Montgomery, AL MSA	8505.72876	3454.91712	10829.16876	5737.47984	4333.81344	0.0	6392.94504	73010.4140625
1	AL	Montgomery, AL MSA	12067.5024	5091.70788	11588.19288	8659.5564	6217.45896	6147.8298	7422.07836	73010.4140625
1	AL	Montgomery, AL MSA	12067.5024	7460.20308	12361.7772	11581.6326	7075.65816	15824.694	9769.56228	73010.4140625
1	AL	Montgomery, AL MSA	15257.1504	9952.23924	13452.186	14503.7076	9134.3562	18802.1892	13101.7032	73010.4140625
1	AL	Montgomery, AL MSA	15257.1504	12182.214	13744.5984	17425.7856	9942.36396	18802.1892	13469.2188	73010.4140625
1	AL	Montgomery, AL MSA	10180.2942	6334.01436	12861.8868	11474.95968	5983.78524	0.0	8236.73076	73010.4140625
1	AL	Montgomery, AL MSA	12067.5024	7883.31888	13589.112	14397.0372	7228.96944	6147.8298	9459.9024	73010.4140625
1	AL	Montgomery, AL MSA	12067.5024	9984.05268	14723.6076	17319.1128	7990.1484	15824.694	11168.75028	73010.4140625
1	AL	Montgomery, AL MSA	15257.1504	12189.7704	14994.6	20241.1872	9945.10176	18802.1892	13210.1484	73010.4140625
1	AL	Montgomery, AL MSA	15257.1504	14917.3584	15064.2636	23163.2652	10933.41504	18802.1892	13417.2192	73010.4140625
2	AL	Daphne-Fairhope-Foley, AL MSA	8616.0	3714.29484	10731.65256	5593.47984	4467.7518	0.0	6455.71512	77884.7578125

This table has the necessary information about households living in different areas of the U.S. states. There's a total of 31429 rows of data. There are 2 categorial columns representing: states and areas. *This table includes a total of 50 U.S. states, along with District of Colombia (DC) federal district.* The other 7 columns represent the different expense categories in numerical floating-point format ($). The last column's data gives the households' incomes ($). The household income is given by the *raw* dataset as median income, which means it represents a *typical* family's outcome, area-based. The data is in a cleaned format, transformations are applied to remove redundant columns, and to fill table with data when needed. The data is annual and represents 2023 expenses.

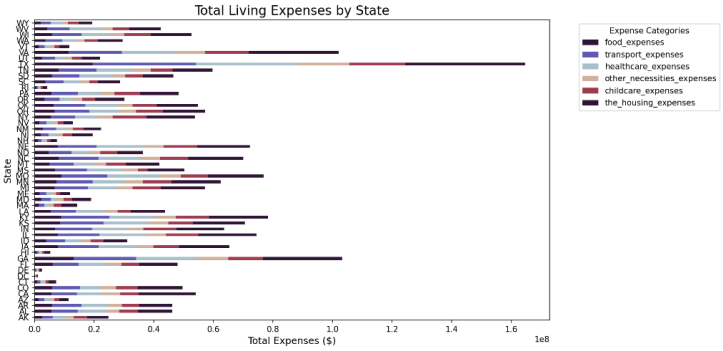
To give a more comprehensive level of insights on the data, various graph representations as below are created to visualize the data.

[1] The Heritage Foundation. Facing a Cost-of-Living Crisis? You're Not Alone. https://www.heritage.org/budget-and-spending/commentary/facing-cost-living-crisis-youre-not-alone ;

2.2.2 Meaning of the data: Cost_of_living Table



As a general overview, states have different number of areas that they cover. For instance, the light purple color shows a state with very few areas, e.g., Maine: 16, and the strongest purple color shows the state with the highest count of areas; e.g., Texas: 215.



The legend shows the expense categories. Represented by the stack bar; in most of the U.S. states the housing expenses section takes up a large portion of total expenses, meaning that a considerable part of the budget goes towards that direction. This observation will be explored in the later analysis.

2.2.3 Structure of the data: House_listings Table

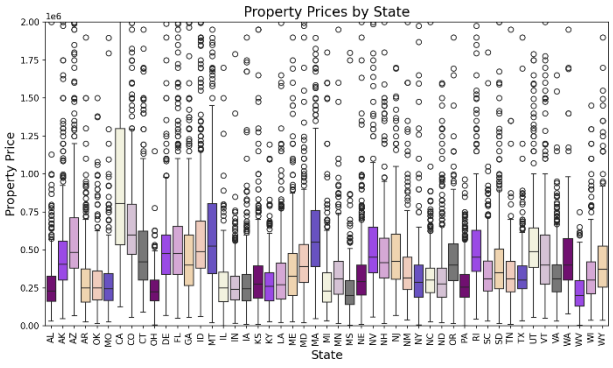
This table has information about properties in the U.S. states. There are 22732 rows of data. There is 1 categorial columns representing states. *There is a total of 49 states, excluding Hawaii and District of Colombia.* The other 3 columns represent the properties' attributes: area (m), price/m^2 ($), and the property price ($) in numerical floating-point format. The last column's data is important for later analysis on housing affordability according to market evaluations. This data is in a cleaned format, and transformations are applied to fill with necessary data. The data is annual and represents 2023 prices.

House Prices in the U.S

state	property_area_meters	price_per_sq_meter	property_price
AL	1614.0	148.636926889715	239900.0
AL	1800.0	144.38888888888889	259900.0
AL	1250.0	274.0	342500.0
AL	2224.0	150.6294964028777	335000.0
AL	1225.0	204.08163265306123	250000.0
AL	1564.0	96.54731457800511	151000.0
AL	1717.0	139.1962725684333	239000.0
AL	1674.0	149.2831541218638	249900.0
AL	2190.0	134.70319634703196	295000.0
AL	3030.0	173.23432343234325	524900.0
AL	2099.0	94.80705097665556	199000.0
AL	1194.0	125.6281407035176	150000.0

2.2.4 Meaning of the data: House_listings Table

The boxplot gives information about the price ranges in different states according to housing market. There's a diversity in the property prices, where some states e.g., California (CA) are skewed towards higher prices, and some other states e.g., Mississipi (MS) have a more uniform housing market. Data appears to have a huge number of outliers, for that reason a median measure of central tendency is considered for the later evaluations.



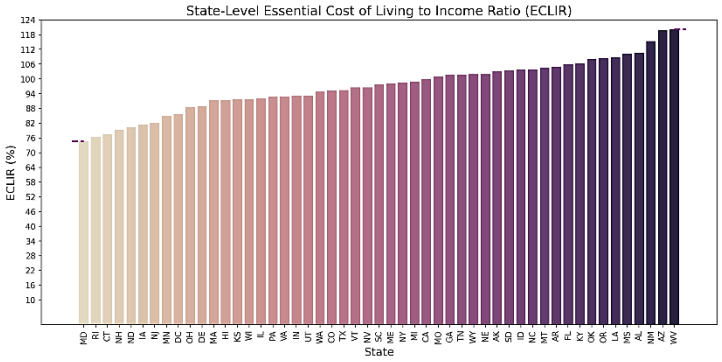
3. Analysis

For answering the main question, this project conducts a literature-based analysis for addressing the housing affordability issue and the variability between states. It focuses on 3 main variables to give a comprehensive assessment of housing affordability: ***essential expenses*** *(ECLIR),* ***housing expenses (as one of the essential expenses)*** *(HCB) and* ***housing costs (as a potential housing expense)*** *(PIR).*

3.1 Essential Cost of Living to Income Ratio (ECLIR)

$$ECLIR(area) = \frac{Median\ Essential\ Expenses}{Median\ Household\ income} \times 100$$

$$ECLIR(state) = \frac{\sum(ECLIR(area)x\ Total\ Area\ Expenses)}{\sum(Total\ Area\ Expenses)}$$

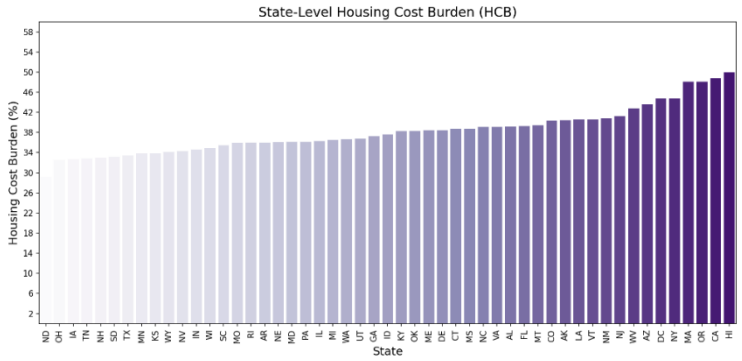


The essential cost of living to income metric represents the portion of a typical household's income that is required to cover essential expenses **(housing, food, transport, healthcare, childcare, other**) which helps assessing the financial strain on households. A high percentage shows that households in that state spend a large amount of their income for their essential needs. Many studies[2] support the argument that less than 50% of the income should go towards living expenses for financial wellbeing. Anything above the margin suggests potential financial stress, and if the ratio exceeds 80%, it may indicate serious to severe financial stress. This graph shows result in incremental order. It starts it representation with states of MD, RI, CT in the range of 75-80%, indicating **moderate financial stress**. Next, it represents 33 states in the range of 80%- 99%, which implies a **serious financial stress**. It then continues with the last 15 states in an **severe financial strain** situation. Overall, the U.S. appears to be in **a non-healthy financial situation**, where households can't comfortably cover essential expenses, especially in the states of NM, AZ, WV where the ratio exceeds even 100%, which potentially suggests that households require assistance, or need debts to cover their expenses.

3.2 Housing Cost Burden (HCB)

$$HCB(area) = \frac{Median\ Housing\ Expenses}{Median\ Household\ Income} \times 100$$

$$State - Level\ HCB = \frac{\sum(HCB(area)x\ Total\ Area\ Expenses)}{\sum(Total\ Area\ Expenses)}$$



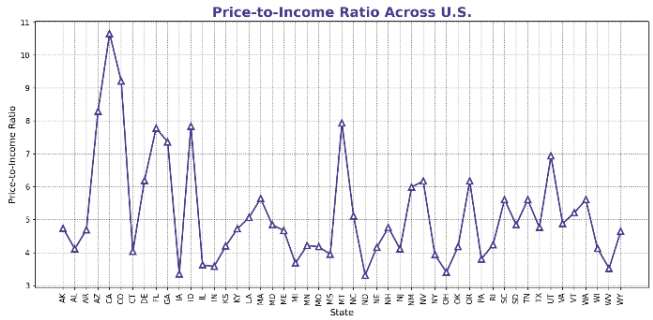
Housing is one of the core essential expenses, along with food, healthcare, transportation, and childcare. Thus, a household's housing expenses directly impacts housing affordability. The Housing Cost Burden specifically measures the exact portion of income that is being spent on housing expenses. By the conventional measure in the U.S.[3], affordable housing is typically defined as housing that costs 30% or less of a household's income. This bar plot chart visualizes the results in incremental order. The method used starts its representation with results of North Dakota (ND) as **non**-house-burdened with HCB around 28%, and **all other** U.S. states as house burdened. 48 states are included in the batch of **moderate housing burden**, including New York (NY), Florida (FL), and Louisiana (LA), which have housing cost burdens in the range of 40%-44%. Lastly, nearly 50% of the households' expenses in California (CA) goes toward housing, which indicates **serious housing cost** burden. 50 out of 51 states require households to allocate a large part of their income to cover housing expenses, conventionally **pointing** to **low housing affordability** across U.S. states.

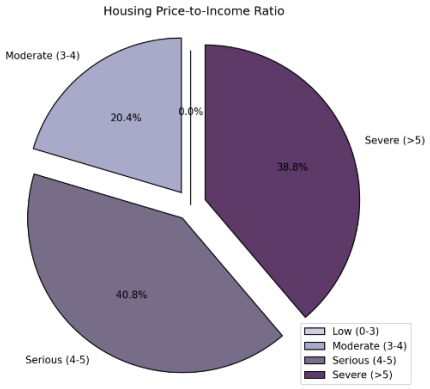
3.3 House Price – Income ratio (PIR)

$$PIR(state) = \frac{Median\ House\ Price}{Median\ Household\ Income}$$

[2] Calculating Your Expenses to Income Ratio: What You Need to Know ; The_50_30_20_Budget_Investopedia.pdf

[3] Housing Cost Burden, Material Hardship, and WellBeing https://nlihc.org/sites/default/files/Housing-Cost-Burden-Material-Hardship-and-Well-Being.pdf





After the assessment of essential living expenses and the housing expenses, this study combines both datasets and measures the **exact impact essential expenses have on housing affordability**. As shown by ECLIR, essential living expenses consume a large portion of income. At the same time HCB highlighted that housing expenses were quite high overall; straining household budgets. To finalize this report, the Price to Income Ratio **directly** measures the housing affordability by **comparing** the house market prices with household incomes. In the housing market[4], housing is generally considered affordable if the cost of it does not exceed three times the median annual income of a household. In this case, there are 0 states that have a PIR 3 or less, indicating that the entire U.S. faces housing affordability as a **challenge**. A PIR value between 3 and 4 is considered as moderate affordability. **20.4%** of U.S. states have moderate housing affordability, emphasizing they are not at immediate risk. The largest proportion of states is at range 4-5 with a value of **40.8%** indicating a serious housing affordability concern for households in these regions. The very last subset of high PIR values indicates a poor affordability for PIR values above 5, and affects **38.8%** of U.S. states; CA (California) is one of the outlier states with extremely high PIR value above 10.

4. Conclusions

Essential living expenses are reducing the financial flexibility.

- The Essential Cost of Living to Income Ratio *metric* potentially links to housing affordability challenges. It showed that all U.S. states can't comfortably cover essential expenses. This indicates that households are unable to save for housing expenses: down payments, cover rising rents, invest on buying houses, or maintain housing quality; worsening core aspects of housing affordability.

Housing expenses exceed the affordable thresholds (30%) in most states.

- Housing Cost Burden *metric* reveals that 50 out of 51 states are house burdened. This implies low housing quality, safety, or location. It also implies a risk of long-term housing affordability. In particular, the West region emphasizes the need for affordable housing solutions.

House prices are inaccessible relative to incomes.

- PIR reveals no state achieves the affordability standard of PIR<=3. This confirms that housing affordability is a systematic issue in the market.

This study highlights major challenges in housing affordability according to the represented analysis. It supports the main question with coherent results from metrics. High essential expenses, housing costs, and house prices together confirm housing affordability challenges in the U.S. However, it does not 100% answer the research question due to limitations listed below.

5. Limitations and Uncertainties

- The entire study is descriptive and doesn't include predictive modeling for housing future challenges.
- The analysis doesn't consider dynamic factors e.g., inflation, wage growth, evolving housing market prices, because the data used is static from year 2023.
- The house_listings table lacks data for Hawaii and District of Columbia, resulting in an incomplete view of house affordability across the *entire* United States.
- Data is aggregated at the state level and uses median values of expenses which can lead to a potential miss of variations related to urban versus rural areas.
- The calculations are based on median values of income which can give a less precise result; especially for states with high gaps in income across different areas.
- The analysis does not take into consideration the household size and its possible effect to the analysis.
- The analysis does not consider external factors like government policies or federal programs that help householders in regard to housing affordability.

[4]Housing Affordability: A Conceptual Overview for House Price Index
https://www.researchgate.net/publication/271881506_Housing_Affordability_A_Conceptual_Overview_for_House_Price_Index

6. Appendix 1: U.S Abbreviation and Names

State Abbreviation	State Name
AL	Alabama
AK	Alaska
AZ	Arizona
AR	Arkansas
CA	California
CO	Colorado
CT	Connecticut
DE	Delaware
DC	District of Columbia
FL	Florida
GA	Georgia
HI	Hawaii
ID	Idaho
IL	Illinois
IN	Indiana
IA	Iowa
KS	Kansas
KY	Kentucky
LA	Louisiana
ME	Maine
MD	Maryland
MA	Massachusetts
MI	Michigan
MN	Minnesota
MS	Mississippi
MO	Missouri
MT	Montana
NE	Nebraska
NV	Nevada
NH	New Hampshire
NJ	New Jersey
NM	New Mexico
NY	New York
NC	North Carolina
ND	North Dakota
OH	Ohio
OK	Oklahoma
OR	Oregon
PA	Pennsylvania
RI	Rhode Island
SC	South Carolina
SD	South Dakota
TN	Tennessee
TX	Texas
UT	Utah
VT	Vermont
VA	Virginia
WA	Washington
WV	West Virginia
WI	Wisconsin
WY	Wyoming