DSC 680

Milestone 2

Armin Heldovac

9/25/2025

The main business problem that I am looking to address in this project is regarding the fact that housing affordability has been one of the biggest challenges in the United States. The prices of housing in America have gone up significantly over the past decade which is causing issues for families. Housing market trends are on the rise as well as rental affordability. The rise in home prices and rental prices makes it difficult for American families to purchase a new home. Many families are having trouble with managing their income due to the price of housing rising. Understanding housing trends is important for the public such as families in America, but also for real estate agents, mortgage institutions, and policy makers. Without an analysis of housing affordability in America, many policy makers may have issues with implementing new policies or assisting the public with housing. Also, construction companies and lending institutions may depend on the analysis of housing affordability to make better business decisions.

Historically, the housing market in the United States has been a rollercoaster. There have been many cycles of ups and downs throughout the years, such as the 2008 crisis where the prices of homes plummeted and caused chaos throughout the country. The main reason the housing crisis happened was due to the low interest rates and an excess of subprime mortgages available. Due to this incident, there was a period where the prices of homes rapidly increased, which led to many people selling their homes and eventually a collapse in housing prices. This eventually led to a global recession. In the past few years however, home prices have

dramatically increased, especially post COVID-19 pandemic. The interest rates were very low, there were supply shortages, and this increased the demand. Housing prices are now higher than they have been within the past 10 years. Since housing prices increased, rental prices followed the same trend. Rental prices have been rising in urban areas where the housing market didn't have an increase in population. Unfortunately, the wages have not been increasing at the same pace as the housing prices, which leads to housing affordability issues.

I have gathered two datasets that are from two main sources. One of them is a Zillow Reasearch dataset from Kaggle, which consists of housing prices and rental prices broken down according to city and state and number of bedrooms. My main focus will be on 2-bedroom houses. The other dataset is a dataset from Kaggle which includes the average income and rent in various states across the United States for the year 2022. To prepare my data for the project, I cleaned the inconsistent regional codes, I ensured all time periods are in the same format across the datasets, and I created a dictionary to include different variables. Since the Zillow Research dataset consists of many different types of data such as 1 bedroom, 2-bedroom, 3-bedroom, 4 bedrooms, etc., I will be using the 2-bedroom dataset since that is the most common to rent and own. Comparing the 2-bedroom dataset with the average income and rent across different states in the United States in 2022, will be the main focus of my analysis. Housing affordability will be determined by analyzing the median household income to the median housing prices.

Throughout my analysis, a few different methods will be used to analyze the datasets. This analysis will consist of descriptive methods and predictive methods. Descriptive methods such as analyzing the housing prices, analyzing the average rental price, and tracking the different amounts of income throughout the country. Time-series analysis will be used to analyze the different changes in housing prices and rental prices over time. Regression modeling will be used to analyze the relationship between housing costs and average income. Predictive modeling

will be used to analyze future housing prices and determine housing affordability based on the historical data within the datasets. The different types of models will help with analyzing housing affordability and allow for future planning on housing.

The analysis of the datasets showed the uneven distribution of housing costs across the United States. The bar chart of the least affordable states for housing shows that Hawaii is a big outlier. Almost 40% of the average household income is used on housing costs, specifically rental costs. California, Massachusetts, and New Jersey are also extremely high in non-affordable housing. This shows that houses on the coast are extremely expensive due to low supply and high demand. The most affordable states seem to be North Dakota, Illinois, and Virginia, where the income is only consumed by rent by about 14%. The histogram created shows that most states are within the 15%-25% range of rent consumption. The scatterplot shows the income vs rent which shows a positive relationship, higher income states have higher rent and housing prices.

The results and analysis of my project show that there are constantly growing challenges within housing affordability in the United States and it isn't getting any better anytime soon. Some of the regions of the United States might be somewhat affordable, the trend across the country shows that the price of housing is increasing faster than the average income of a household increase. Based off the predictive modeling used in the analysis, it is safe to assume that affordability will continue to decrease unless income rises or the housing market expands. In the major metropolitan areas of the country, housing affordability will continue to be an issue.

The analysis assumes that the datasets from Zillow and Kaggle are very accurate and show how Americans are struggling to afford housing due to income issues and housing supply.

It assumes that relationships between income and housing costs are going to be consistent in the future. Also, the analysis assumes that rent costs will continue to be a burden.

Limitations within the analysis include a few gaps within the datasets such as representation of smaller areas within the country. Also, housing costs might not be reported the same way in every region of the country. One thing that predictive models might lack is the limitation in accuracy, since housing markets are very complex.

A challenge that I came across during analysis was combining the datasets from multiple sources. Each dataset had different reporting structures within the data such as region names, income amounts, and housing costs, so it was challenging to clean and analyze the data. A challenge that I came across within the regression model was to make sure it doesn't overfit to historical data.

In the future, this project can be applied by various groups within the country. Local governments and policy makers can use this analysis to ensure housing programs are intact and are supporting the community. Real estate agents can use this analysis to identify which regions within their state require the highest demand for housing. Another way that this analysis could be used in the future is to address social issues and improve income for all Americans.

I recommend policymakers to prioritize the high-cost metropolitan areas of the country and expand the supply of affordable housing. I also recommend them to invest their money into rental assistance programs to support affordability issues. Real estate agents might also succeed with developing new affordable housing in the regions where it is too expensive.

This project will be implemented in different stages. It starts with finalizing the data preparation and analysis using regression and descriptive models. Following that, predictive models will be applied to analyze the future trends of affordability. All results will be presented

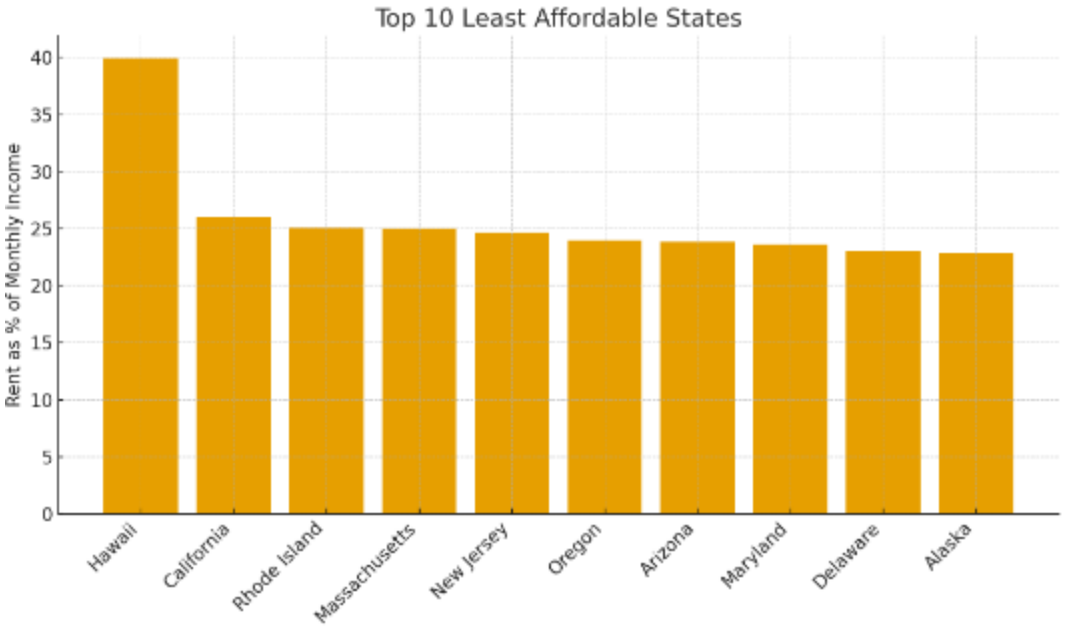
with visualizations such as graphs, scatterplots, and heat maps. The final stage will be the presentation along with a white paper.

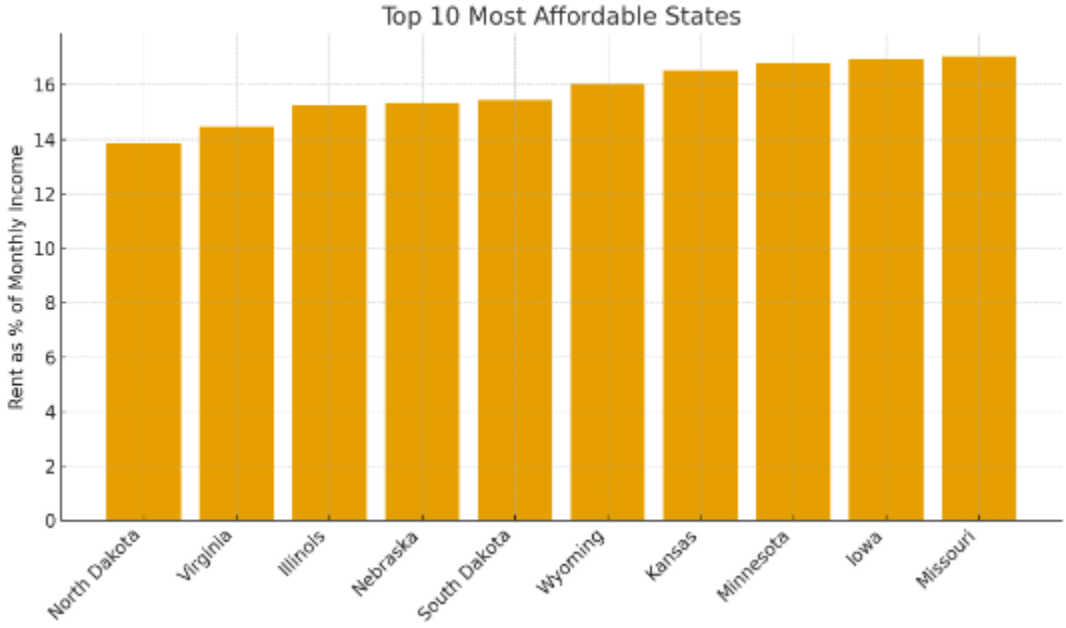
Since housing affordability is a sensitive topic within the social aspect, this project will be presented properly and responsibly. This project will avoid any misrepresentation of lower-income areas and will avoid blaming any housing affordability issues on personal struggles. Being transparent with the data being presented will be the focus. All findings within the project will be presented in a way that focuses on fairness and transparency.

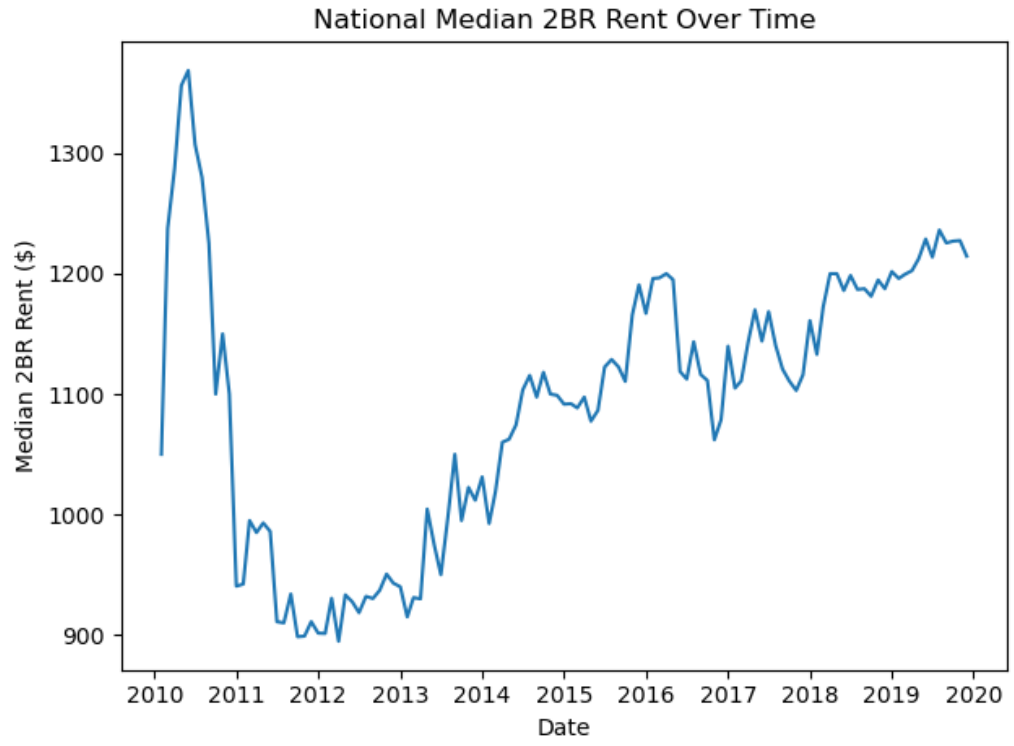
References -

- Mooney, P. (2020, December 8). *Zillow House Price Data*. Kaggle. https://www.kaggle.com/datasets/paultimothymooney/zillow-house-price-data
- Kabir, S. (2024, May 12). *Average income and rent in United States*. Kaggle. https://www.kaggle.com/datasets/shahriarkabir/average-income-and-rent-in-united-states

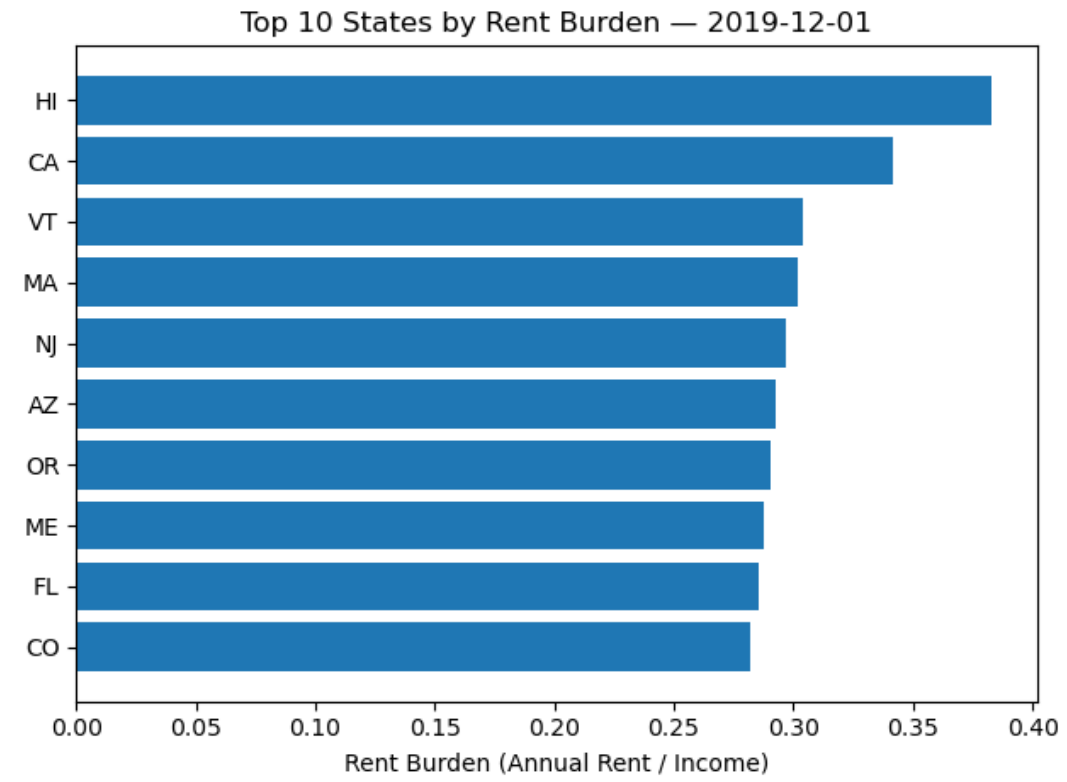
Visualizations -

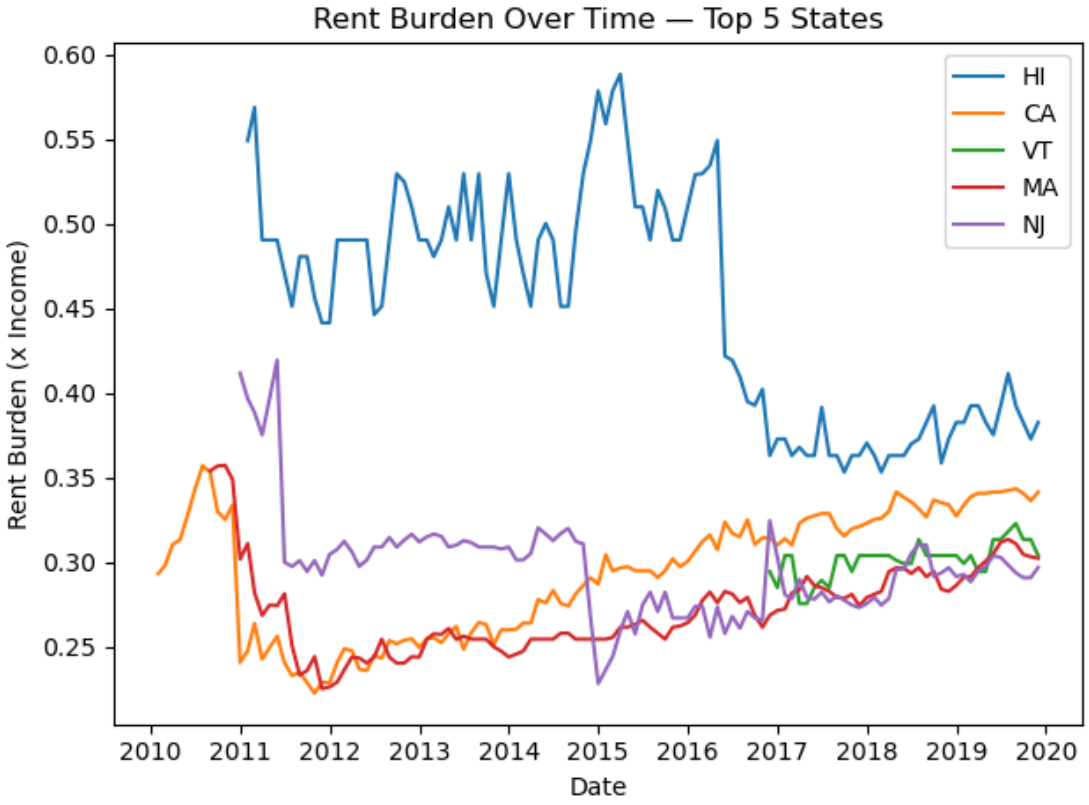

Top 10 Least Affordable States
Rent as % of Monthly Income
0
5
10
15
20
25
30
35
40
Hawaii
California
Rhode Island
Massachusetts
New Jersey
Oregon
Arizona
Maryland
Delaware
Alaska

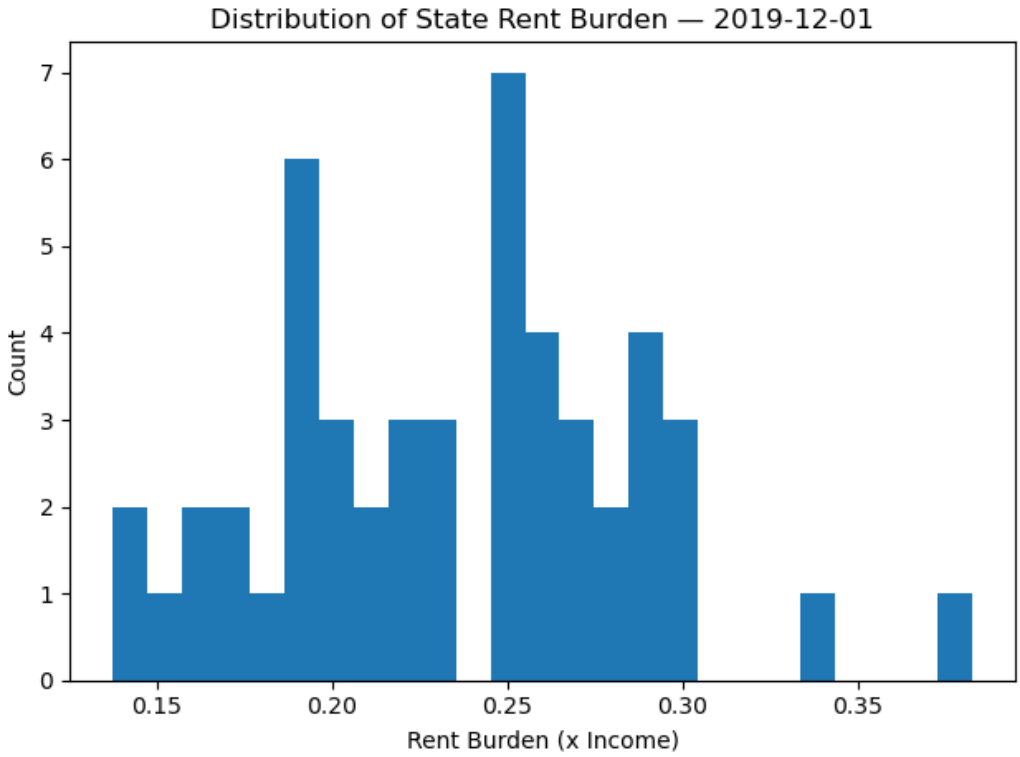

Top 10 Most Affordable States
Rent as % of Monthly Income
0
2
4
6
8
10
12
14
16
North Dakota
Virginia
Illinois
Nebraska
South Dakota
Wyoming
Kansas
Minnesota
Iowa
Missouri


National Median 2BR Rent Over Time
Median 2BR Rent ($)
1300
1200
1100
1000
900
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date


Income vs 2BR Rent — 2019-12-01 (State-level)
Median 2BR Rent ($)
2200
2000
1800
1600
1400
1200
1000
800
45000
50000
55000
60000
65000
70000
75000
80000
85000
Average Income ($)


Top 10 States by Rent Burden — 2019-12-01
HI
CA
VT
MA
NJ
AZ
OR
ME
FL
CO
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Rent Burden (Annual Rent / Income)


Rent Burden Over Time — Top 5 States
HI
CA
VT
MA
NJ
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
Rent Burden (x Income)
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date


Distribution of State Rent Burden — 2019-12-01
Count
Rent Burden (x Income)
0
1
2
3
4
5
6
7
0.15
0.20
0.25
0.30
0.35