Projecting HPI and Homelessness in the United States

Katon Minhas

I. Abstract

This paper examines the relationship between housing demand, cost, and homelessness in the United States. The study tests the hypothesis that more housing units per resident lead to lower prices and reduced homelessness. Two neural network models project future housing costs and homelessness based on different construction scenarios. The findings confirm the influence of housing availability on prices and homelessness, but exceptions exist. The models provide insights into the relative impact of housing and other factors on prices and homelessness, despite limitations in predicting future outcomes. The study underscores the need for a comprehensive approach to address these issues.

II. Introduction

In recent years, many states in the U.S. have been grappling with the issue of skyrocketing housing and rent costs. The substantial increase in cost of living has far-reaching consequences, pushing more and more people into homelessness every year. This effect is particularly seen in high-demand areas, where high-income jobs have fueled a population migration and tipped the supply-demand balance of the housing market unfavorably. New York City, San Francisco, Seattle, and Austin have become poster-cities for rental rates that would've been unfathomable years ago, and the corresponding homelessness crisis in these areas continues to make headlines.

While there are many factors that go into housing affordability, one basic underlying principle is presumed true: a higher number of housing units available per person puts less strain on the market and leads to generally lower prices. The goal of this project is to test this principle and examine the relationship between housing demand, housing cost, and homelessness. It is hypothesized that markets where there are more units available per resident also see lower home prices and lower rates of homelessness. Furthermore, two deep neural network models were developed to project the future of housing costs and homelessness based on potential new housing construction scenarios.

III. Background

Determining the cause of high housing prices in a region is not a one-size-fits-all problem. Supply and demand is considered the primary driver of housing prices, but interest rates, employment rates, land costs, government building regulation, and the prevalence of investment properties all play a significant role as well (Kahn 2008).

Causes of homelessness are also varied, but more straightforward than media may portray. While mental health, drug addiction, economic factors, and discrimination play a factor, studies have consistently found that the biggest correlating factor with homelessness is housing affordability. Major urban markets in California, New York, Hawaii, and Washington D.C. have extremely high housing costs and homeless rates, despite not having significantly higher rates of drug addiction or mental illness from other markets (Ortiz-Ospina and Roser, 2017). By reframing the discussion of homelessness as primarily housing-caused, Housing First models have seen higher success rates than traditional shelters and treatment facility models in reducing homelessness (Kertesz and Johnson, 2017).

IV. Approach

This project examines the relationship between four primary factors: population, housing inventory, housing price index, and homelessness. The factors are considered on a state level.

Data

Many topical datasets are publicly available through government agencies and other organizations.

a. Population

   Population figures were attained from the US Census Bureau (US Census Bureau, 2019-2022). State populations are officially measured every 10 years, with projections released yearly in between. For this analysis, population figures from 2010-2022 were considered.

b. Housing Inventory

   Housing inventory was also attained from Census data (US Census Bureau, 2019-2022). Data was available from 2010-2021. Note that figures refer to total number of housing units in each state, not necessarily new housing construction. Because of this, a few states have less housing year-over-year due to demolition and repurposing.

c. Housing Price Index (HPI)

   FHFA House Price Index is a standard measure of housing prices based on data from all states that has been used since 1975 to evaluate housing price fluctuations. A HPI score of 100 indicates the median US house price in 1975. HPI offers a standard comparison metric across time and region. Data was sourced from the Federal Housing Finance Agency and contains records from 1975 to 2022 (Federal Housing Finance Agency, 2022).

d. Homelessness

   Homelessness data was collected using Point-in-Time estimates from each state from 2007-2022 (HUD Exchange, 2023). Counts for 2021 are considerably lower due to many homeless people being housed during Covid-19, so accurate values for this year were interpolated as the mean of 2020 and 2022 records.

e. Population Projections

Projections were sourced from the Cooper Center Demographics Research Group and contain estimates of the population for each state on 5-year intervals between 2020 and 2040 (University of Virginia, 2018). The projections were released in 2018, so adjustments were made to factor in data points up to 2022. Intermediate year projections were interpolated.

Preprocessing

Initial data cleaning consisted of minor format manipulations and combinations of yearly data frames into a single output for model training and prediction. Data was also transformed into long format for visualization and ultimately condensed into two data sources: 'all_features.xlsx' and 'all_projections.xlsx'. Two Jupyter Notebooks were used for preprocessing: 'housing_data_cleaning.ipynb' and 'final_data_prep.ipynb'.

Modeling

The predictive modeling component of the project consisted of two separate neural network models, which can be found in 'housing_data_model_final.ipynb'. The purpose of the models is to predict HPI and homelessness, respectively, to 2040 based on different housing inventory scenarios. The scenarios are defined by set levels of housing construction between 25% and 300% of current construction pace each year, at 25% increments.

The HPI model was first trained on past data with the following variables: available housing inventory, population, year, and state (one-hot-encoded). HPI is the target variable. Many models were attempted, including random forest regression, k-nearest neighbors regression, and XGBoost regression. The model ultimately included as a Keras Sequential neural network with 3 hidden layers of 300, 200, and 100 neurons respectively. The network was trained over 100 epochs with a batch size of 32, using mean squared error as the loss function for evaluation. The validation root mean squared error (RMSE) during training was 32.2. The trained HPI network was used to generate predictions for each state, year, and housing construction level, with population projection data used to extend predictions to 2040. HPI predictions were then post-processed to align with 2022 measurements.

The neural network model to generate predictions of homelessness was nearly identical, but had the additional feature of HPI added to the training set. To generate predictions of homelessness until 2040, the outputs from the HPI network were used as features. The homelessness model achieved a validation root mean squared error of 5313, indicating lesser performance than the HPI model. Homelessness predictions were also post-processed.

Dashboards

To illustrate the findings, two Tableau dashboards were created. The first presents an overview of homelessness, housing, HPI, and population by state to illustrate the understood relationships between the four factors. This dashboard also allows for filtering on any of its visualizations.

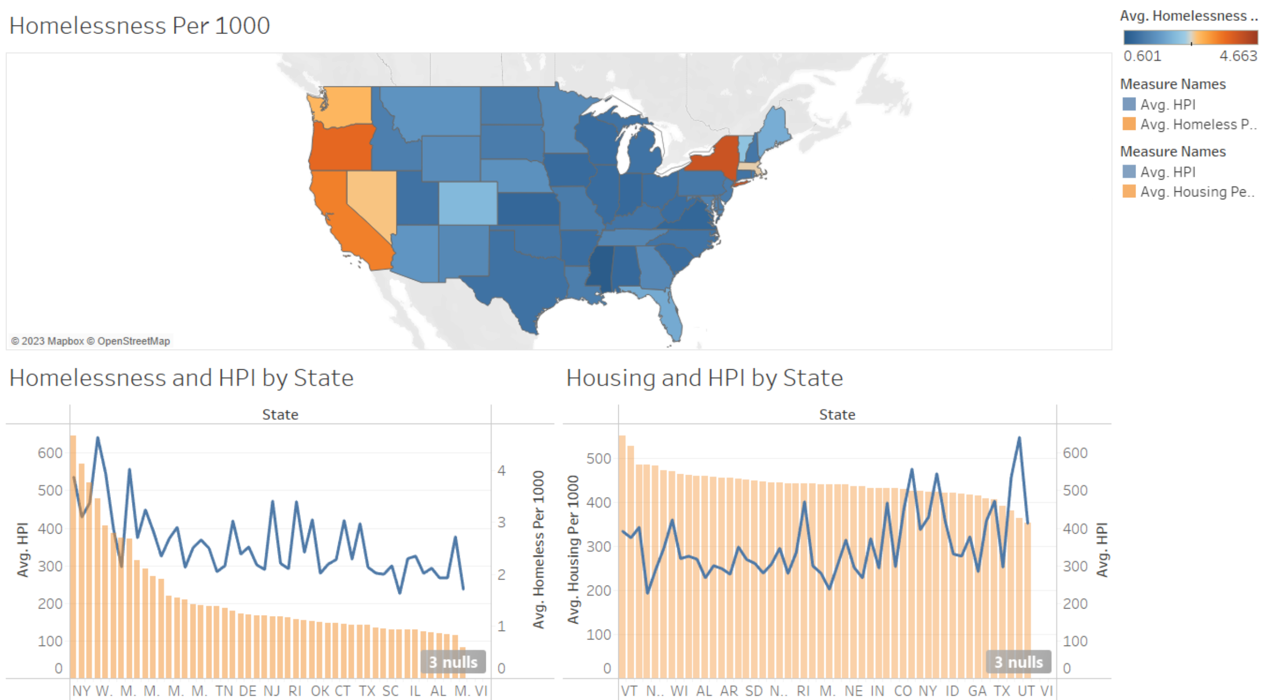

Figure 1. Overview Dashboard with no filters applied

The second dashboard contains yearly trends, as well as the results of the HPI and homelessness projection models. This dashboard is intended to examine one state at a time and should be viewed after applying a state filter on the overview dashboard. The dashboard shows the trend of homelessness and HPI for a given state over a year range, and shows future projections of homelessness and HPI based on hypothetical new housing construction rates.

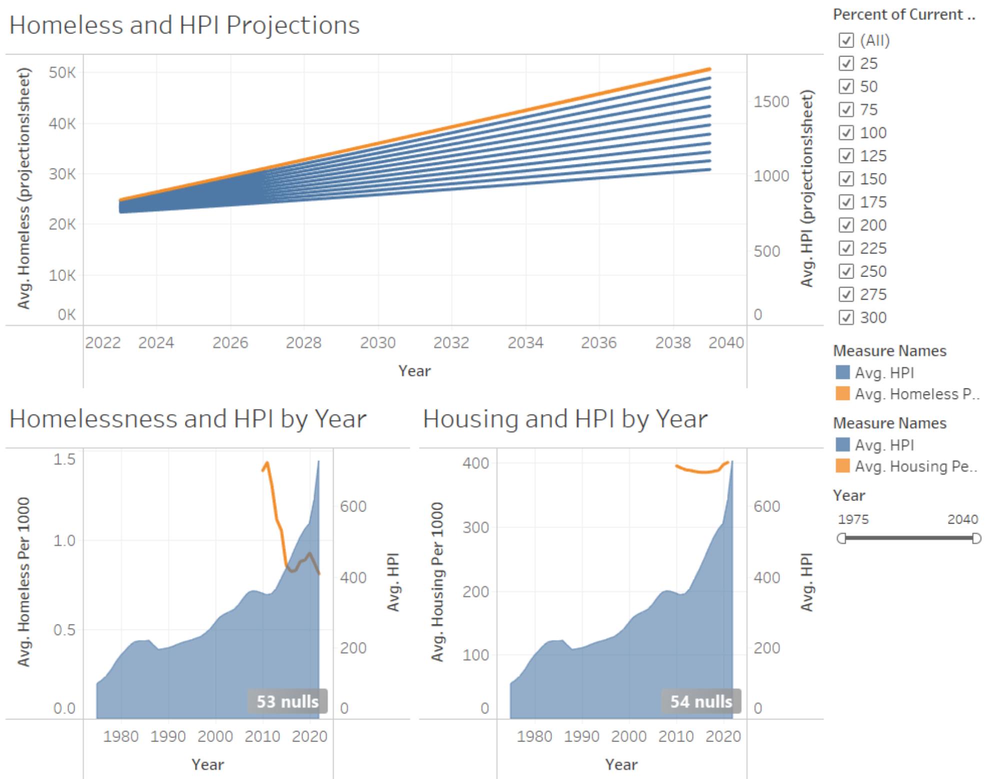

Figure 2. Yearly trend and projection dashboard for the state of Texas

V. Results and Discussion

The visualizations in the overview dashboard confirmed the expectation that regions with a lower house-to-resident ratio had a higher HPI, and regions with higher HPIs also saw higher rates of homelessness. This indirectly confirms the hypothesis that housing availability is a major factor in high homelessness rates. As figure 1 shows, this relationship is not directly linear and there exist exceptions to the trend. The degree to which these exceptions outlie can then be thought of as a measure of the impact of factors not related to housing and population.

The trends seen in the yearly trend and projection dashboard were more varying, especially the model projections. In general, the above-established relationship between linking housing to HPI and HPI to homelessness can be seen in most states. States like Texas, shown in figure 2, represent outliers in that steeply rising HPI values in recent years also corresponded to less homelessness. It is possible that in circumstances like this, the HPI rise was so due to external factors like increased jobs and economic activity in the state. This is confirmed when observing the trend seen in the relation of housing and HPI by year in Texas. Texas saw little movement in the past decade in terms of improving its low house-to-resident ratio, yet HPI increased dramatically over the same time period. In cases like these, external factors far outweigh the influence of housing- and population-related metrics.

Despite high validation performance, neither of the model predictions were in line with the hypothesis that increased housing production will lead to a decrease in HPI or homelessness. The HPI projections saw near-0 differences between each of the housing construction scenarios, while the homelessness projections saw the scenario with the highest amount of new housing leading to higher home values and more homelessness. This phenomenon was seen across nearly every state. There are a few likely explanations for this. The first is that the model was simply overfit, using too few data points and too few features to attempt to project something as complicated as homelessness or housing prices. This is exacerbated by the nature of the two values. Housing prices are based on broad market trends and data, leaving them less prone to huge surges and dips (except in the rare case of a market collapse). Change in housing prices due to construction is likely to be very slow as new units must first be built, then reacted to by the market. On the contrary, homelessness numbers based on point-in-time counts are prone to large fluctuations based on a variety of highly localized factors. Both of these metrics are based on much more than just housing. Despite housing still being the largest single factor in determining HPI and homelessness, it is far from having the 50% or greater explained variance that would lead to accurate future projections.

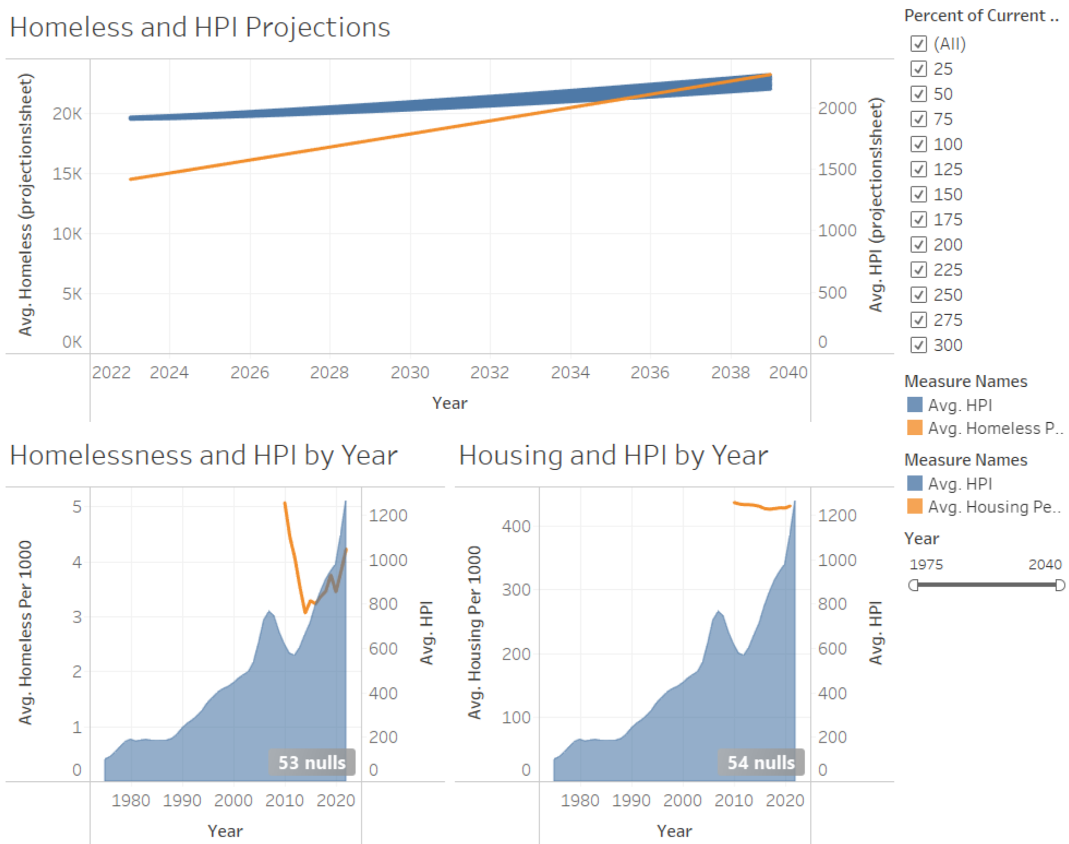

Figure 3. Yearly trends and projection dashboard for the stae of Oregon

Despite model performance not being explanatory enough to be used in real-world scenarios, an observation of the spread in different states can still lead to useful insights. Comparing the dashboard in figure 3 for Oregon with the one in figure 2 for Texas, it is clear that the homelessness and HPI scenarios in each state are very different and should be handled as such. In Texas, the homelessness model had a very large spread based on housing construction, indicating that housing is likely less of a contributor to homelessness in the state compared to other factors. This is confirmed by the fact that homelessness and HPI are completely unrelated or even conversely related. By contrast, the spread of homelessness predictions was very narrow, and the ‘Homelessness and HPI by Year’ chart shows a very close linear relationship between the two. This may indicate that the homelessness crisis in Oregon is much more directly related to housing, with other factors playing less of a role. If this model were to be used to make policy decisions (which, based on performance, is by no means a recommendation), it would suggest that Oregon's homelessness problem may be better fixed with a housing-first approach, while Texas could benefit from focusing efforts on other factors.

VI. Conclusion

To conclude, a detailed analysis of the relationship between housing, population, housing prices, and homelessness confirmed many widely held beliefs on the subject. Chief among these is the confirmation that the general principle of supply and demand holds true for the U.S. housing market, with more availability of housing leading to lower prices in most states. It was also confirmed that higher HPI was associated with higher homelessness, although this trend also saw notable outliers.

The experimental neural network models to predict HPI and homelessness were ultimately not able to accurately capture the full range of factors that go into either of those measures. However, the size of the gap in explainability could still be useful in determining how much of a state's HPI or homelessness rates can be explained by the housing market alone as opposed to other factors.

VII. References

Ananya Roy, Gary Blasi, Jonny Coleman, & Elana Eden. (2020). Hotel California: Housing the Crisis. *UCLA Luskin Institute on Inequality and Democracy*. https://escholarship.org/uc/item/0k8932p6#main

Edsall, R. M., Harrower, M., & Mennis, J. L. (2000). Tools for visualizing properties of spatial and temporal periodicity in geographic data. *Computers & Geosciences*, *26*(1), 109–118. https://doi.org/10.1016/S0098-3004(99)00037-0

Esteban Ortiz-Ospina & Max Roser. (2017). Homelessness. *Our World in Data*. https://ourworldindata.org/homelessness?source=post_page#

Federal Housing Finance Agency. (2022). House Price Index. https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx

HUD Exchange. (2023). PIT and HIC Data Since 2007. https://www.hudexchange.info/resource/3031/pit-and-hic-data-since-2007/

Jayant Madhavan, Shreeram Balakrishnan, Kathryn Brisbin, Hector Gonzalez, Nitin Gupta, Alon Halevy, Karen Jacqmin-Adams, Heidi Lam, Anno Langen, Hongrae Lee, Rod McChesney, Rebecca Shapley, & Warren Shen. (2012). Big Data Storytelling through Interactive Maps. *Google Inc.* https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/39959.pdf

John D. Landis. (2000). Raising the Roof: California Housing Development Projections and Constraints, 1997-2020. *Department of City & Regional Planning, UC Berkeley*. https://escholarship.org/uc/item/1391n947#main

James A. Kahn. (2008). What Drives Housing Prices?, *FRB of New York Staff Report No. 345.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1264048

Kertesz, S. G., & Johnson, G. (2017). Housing First: Lessons from the United States and Challenges for Australia: Housing First. *Australian Economic Review*, *50*(2), 220–228. https://doi.org/10.1111/1467-8462.12217

University of Virginia Weldon Cooper Center, Demographics Research Group. (2018). National Population Projections. https://demographics.coopercenter.org/national-population-projections

United States Census Bureau. (2019). State Population Totals and Components of Change: 2010-2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html

United States Census Bureau. (2022). State Population Totals and Components of Change: 2020-2022. https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html

United States Census Bureau. (2019). National, State, and County Housing Unit Totals: 2010-2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-housing-units.html

United States Census Bureau. (2021). National, State, and County Housing Unit Totals: 2020-2021. https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-housing-units.html