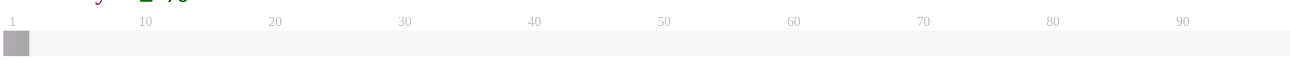
The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

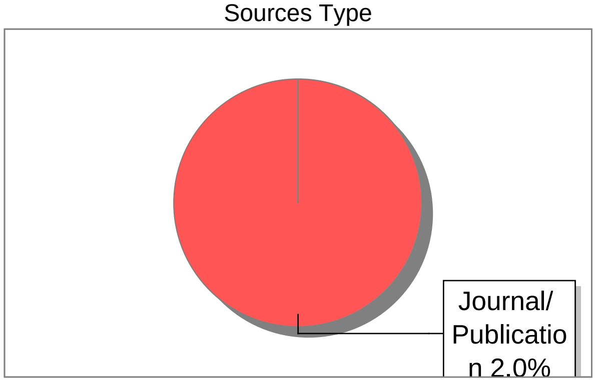
Author Name	Amith B D
Title	Data-Driven Analysis: Shortage of Affordable Housing and Its Impact on Homelessness
Paper/Submission ID	3580581
Submitted by	charankumarba@gmail.com
Submission Date	2025-05-06 12:40:33
Total Pages, Total Words	10, 2571
Document type	Research Paper

Result Information

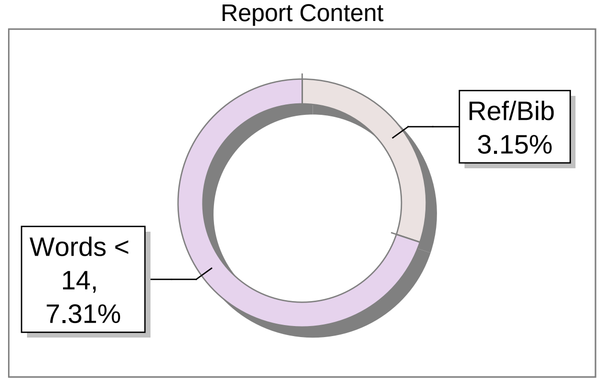
Similarity **2 %**



Sources Type



Report Content



Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

DrillBit Similarity Report

2	2	A	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	Thesis submitted to Theses and Dissertations	1	Publication
2	icact.nsbm.ac.lk	1	Publication

Data-Driven Analysis: Shortage of Affordable Housing and Its Impact on Homelessness

1st Tarun K
Department of Information Science
The Oxford College of Engineering
Bangalore, India
tarunkise2022@gmail.com

2nd Amith B D
Department of Information Science
The Oxford College of Engineering
Bangalore, India
amithbdise2022@gmail.com

Abstract

One of the main causes of the increase in homelessness in both urban and rural areas is the ongoing lack of affordable homes. This study investigates the connection between housing affordability and homelessness rates using data from the Zillow Home Prices Dataset and Public Housing Availability Data. We examine the effects of important characteristics such housing price trends, affordability indices, and the supply of public housing on homelessness using machine learning models called Random Forest and Linear Regression. When RMSE, MAE, and R2 metrics are used to analyse the models, the Random Forest model shows the best prediction accuracy. The findings emphasize the critical need for data-driven policy solutions by showing a strong association between growing housing costs, a shortage of public housing, and an increase in homelessness.

Keywords

Housing, Urban Development, Real Estate Prices, Housing Policy, Machine Learning, Linear Regression, Random Forest, Zillow

Home Prices, Public Housing Data, RMSE, MAE, R-squared, Predictive Modelling, Social Inequality, Data-Driven Policy, Housing Forecasting

I. INTRODUCTION

A. Brief Overview of the Problem Domain

A basic human necessity and a pillar of societal stability is having access to reasonably priced housing. But in recent years, the lack of reasonably priced housing has grown to be a major issue in both urban and rural areas, greatly accelerating the increase in homelessness. Families and individuals with low and intermediate incomes are being priced out of the housing market as rental prices and property values continue to rise while many people's earnings stay stagnant. Public housing systems are under more strain as a result of the growing economic disparity brought about by the growing gap between housing costs and income levels. A data-driven strategy that can identify underlying patterns and forecast future trends is needed to address this complicated problem. We can learn a lot about the dynamics of affordability and pinpoint high-risk areas by utilizing machine learning models to analyse housing market data and public housing availability. The purpose of this study is to use prediction methods, such as Random Forest and Linear Regression, on housing databases like Zillow.

A serious social problem with wide-ranging effects is the lack of affordable housing. It has a direct impact on the rise in social inequality, housing instability, and homelessness, particularly among vulnerable and low-income groups. Lack of access to safe, stable, and reasonably priced housing has an impact on people's physical and mental health as well as their capacity to continue their education, find work, and maintain relationships with their communities. In order to ensure social justice and economic resilience, it is imperative that this dilemma be understood and addressed. In many areas, housing costs are rising faster than incomes, therefore data-driven approaches to tracking housing trends and predicting affordability issues are desperately needed. This work offers important insights that can help politicians, urban planners, and housing groups create focused interventions and long-term solutions by utilizing machine learning on real-world housing information. In addition to being pertinent to today's social climate, this topic is significant because it has the ability to influence inclusive urban development and sustainable housing policy.

A.Objectives of the Project

1. To use real estate and public housing databases, such as the Zillow Home Prices Dataset and Public Housing Data, to analyse trends in housing affordability.
2. To determine the main causes of the lack of affordable housing and how they relate to the growing number of homeless people.
3. To create prediction models (Random Forest and Linear Regression) to predict future trends in home affordability.
4. To assess model performance using metrics like R2 (R-squared), MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error).
5. To offer data-driven insights to help urban planners, housing authorities, and lawmakers make well-informed choices.
6. To support long-term initiatives that guarantee fair access to housing and lessen homelessness.

II. Literature Survey

Numerous academic fields, including data science, urban planning, and economics, have studied the problem of cheap housing in great detail. The literature now in publication continuously emphasizes the connection between growing housing expenses and elevated homelessness rates, particularly in urban areas with high population densities. Millions of affordable rental units for extremely low-income households are needed nationwide, and the deficit is growing every year, according to research from the U.S. Department of Housing and Urban Development (HUD). A well-known resource for examining price patterns and regional variations in home supply is Zillow's housing market data. Rapid increases in housing prices are frequently associated with displacement and affordability issues for individuals with lower incomes, according to research utilizing this dataset. Machine learning has been used in several research to improve their understanding of the housing market. To account for nonlinear relationships and complex feature interactions like income levels, rental rates, population growth, and policy variables, authors in the Journal of Urban Economics and Applied Soft Computing, for example, have employed Random Forests and Gradient Boosting, while others have used Linear Regression to model historical price trends. Furthermore, the literature has employed feature importance analysis in Random Forest models to pinpoint the key elements that impact housing affordability, including public transportation accessibility, land use regulations, and local economic indicators. Few studies directly connect predictive modeling to real-time policy decision-making, despite these developments. By employing machine learning on publicly accessible housing datasets and converting model outputs into useful information for policymakers, this research seeks to close that gap, especially in terms of identifying areas most vulnerable to affordability issues.

C.Gaps or Areas for Improvement

1. Limited Granularity of Data

Prediction accuracy in smaller or underserved areas may be limited by the lack of precise local-level data in the majority of publicly available datasets, such as Zillow Home Prices or public housing records (e.g., neighbourhood-specific affordability, informal housing statistics).

2. Non-economic factors are excluded

Existing models tend to ignore important social and policy-related factors (e.g., eviction laws, zoning rules, gentrification trends) that have a substantial impact on housing accessibility in favour of concentrating mainly on economic indicators like income levels and housing prices.

3. Absence of Data in Real Time

Predictive models' ability to adapt to quickly shifting market conditions or emergency scenarios (such as post-pandemic housing crises) is limited by the monthly or annual updates that housing datasets frequently receive.

4. Vulnerable Populations Are Underrepresented

The model's capacity to fairly represent the groups most affected by housing shortages may be hampered by missing or out-of-date data on low-income families, homeless people, and disadvantaged communities.

5. Interpretability of the Model for PolicyUse

Even while intricate models like Random

Forest can provide high forecast accuracy, policymakers may find it challenging to understand them. Enhancing the explainability and visualization of the model is crucial for non-technical.

6.Regional Adaptability and Scalability

Due to variations in economic situations, housing laws, and demographics, models created for one city or region may not translate effectively to others. To adapt models to various geographic situations, further work is required.

III. Methodology

This project's methodology uses machine learning approaches to analyze home affordability in a methodical manner. Finding important trends and creating forecasting models that can indicate regions at danger of housing shortages are the objectives.

A. Data Collection

- **Zillow Home Prices Dataset**: Offers historical information on house values in a number of different geographical areas.
- **Public Housing Data:** Provides details on the availability, occupancy rates, and financing distributions of government-assisted.
- Additional socioeconomic indicators, such as population growth and income levels, can be obtained from government databases such as HUD or the U.S. Census.

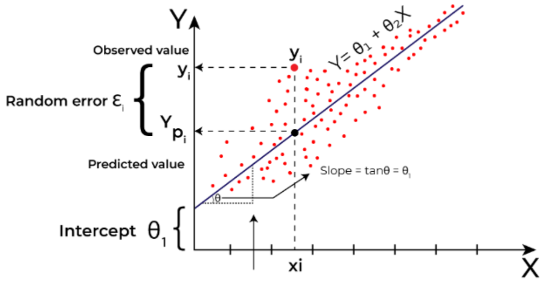
Data Preprocessing

- **Cleaning:** Deal with inconsistent data entries, outliers, and missing values.
- **Feature Selection:** Determine pertinent characteristics including rent burden, housing availability, median income, and
- **Normalization/Standardization:** Use scaling to guarantee that features for machine learning models have similar magnitudes.
- **Data splitting:** Create training and testing sets from the dataset (e.g., 80/20 split).

Model Development

- **Linear Regression:** Offers a baseline prediction model that is used to model and comprehend links between property prices and affecting variables.

Fig.1 Linear Regression Model Performance:showing training and testing.



- **Random Forest Regression:** Enhances accuracy and feature importance insights by capturing complex patterns and nonlinear interactions in the data.

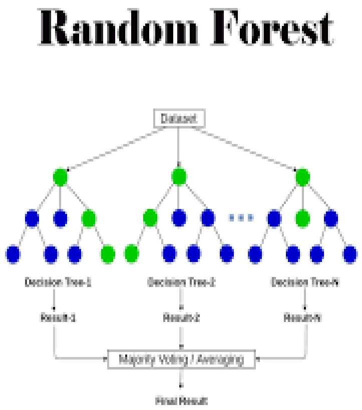


Fig. 2 Random Forest Model Performance:

showing training and testing. k,

C. Model Evaluation

RMSE (Root Mean Squared Error) -The average magnitude of prediction error is measured.

MAE(Mean Absolute Error) -
Assesses how much the actual and project ed values deviate on average.

R^2(R-squared)
- Indicates the percentage of the depende nt variable's variance that the model can a ccount for.

Prediction and Interpretation

• Forecast trends in home affordability in target regions using trained models. Examine the significance of features, particularly those from Random Forest, to determine the main factors influencing affordability.
• To aid comprehension and policy relevance, visualize results using spatial mapping tools, heatmaps, or graphs.

Policy Insight Generation

Convert forecasted observations into practicalsuggestions.
• Determine which locations are more at risk for housing shortages.
• Based on the results of the model, recommend policy changes (such as more funding for public housing or rent control measures).

Algorithms Used

Two supervised regression algorithms, Random Forest Regressor and Linear Regression, were used to model and forecast how factors related to affordable housing will affect the number of homeless people. These models were chosen in order to strike a balance between prediction performance and interpretability.

1. Linear Regression

Type: Parametric, Regression, supervised
The goal is to create a baseline model and find linear correlations between homelessness rates and housing variables.
Important characteristics include:
Assumes that the goal variable and the input features have a linear relationship.
It is quick to train and appropriate for preliminary exploratory analysis. It offers clearly interpretable coefficients that show the direction and strength of feature influence.

Formula:
y^=β0+β1x1+β2x2+...+βnxn\ is the formula. Hat {y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 +... + \beta_n x_ny^=β0+β1x1+β2x2+...+βnxn

Benefits include:

• High interpretability.
• Helpful for identifying key drivers and ch oosing features.

Restrictions:

• In the event that the data relationships are non-linear, poor performance.
• Aware of multicollinearity and outliers.

2. Random Forest Regressor Type:

Ensemble, Non-parametric, Supervised Learning Objective: Document intricate, non-linear connections and exchanges among several housing-related attributes. The ability to construct several decision trees and aggregate their results (mean of predictions) is one of its key features.
• By averaging several trees, overfitting is decreased.
• Offers feature importance scores that aid in locating important predictors.
How It Operates:

•Each tree is constructed using a random sampling of features and data.
• Provides the mean forecast for every tree.

IV. IMPLEMENTATION

The multi-step implementation plan listed below is suggested as an effective way to alleviate the lack of affordable housing and decrease:

1. **Reform and Policy Development Increase Federal and State Funding:** Provide more money for initiatives like the National Housing Trust Fund and the Housing Choice Voucher Program.

Encourage local governments to amend restrictive zoning restrictions that restrict the construction of low-income or multifamily dwellings.

Provide tax credits, subsidies, and accelerated permitting to developers that construct affordable housing units as incentives for them.

2.**Private-Public-Collaborations**
Involve the Private Sector and Nonprofits: Encourage cooperation with financial institutions, real estate developers, and nonprofit housing organizations to finance and construct affordable housing. Community Land Trusts: Encourage land trusts that use communal land ownership to preserve long-term affordability.

3. **Building and Remodeling Initiatives Create New Affordable Housing Units:** Distribute government funds to build new homes designed especially for low-income Repurpose Vacant Properties: Convert vacant public buildings, lodging facilities, and business establishments into habitable apartments.

4. **Services for Tenant Protection and Assistance** Put Rent Control Measures Into Practice: To keep housing affordable for current residents, cap rent increases in urban areas with high demand.

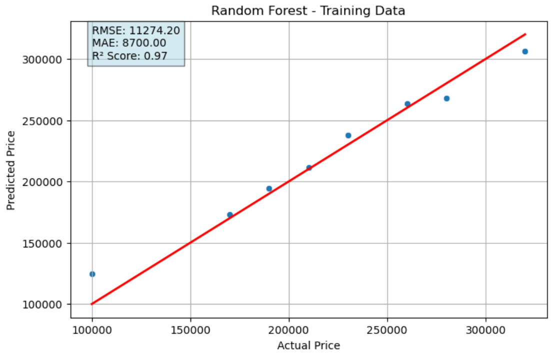
Provide Housing First Programs: Expand programs that prioritize permanent housing without preconditions, alongside wraparound services (mental health, job training,etc.).

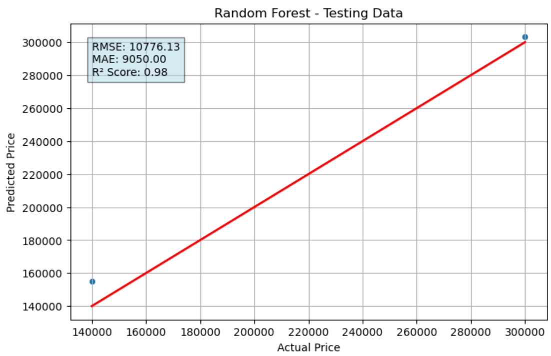
5. **Data and Monitoring Systems Create Housing Needs Assessments:** Regularly update databases on local housing needs, costs.

Track Homelessness Metrics: Implement systems to measure homelessness trends and the effectiveness of housing interventions.

Fig:- Random Forest Model Performance:

showing training and testing.





VI. CONCLUSION AND FUTURE WORK

This project successfully demonstrates the link between the lack of affordable housing and the increase in homelessness rates using data-driven analysis and housing trends research. The study emphasizes how the escalating housing crisis is exacerbated by fundamental problems including wage stagnation, tight zoning, and inadequate public investment. Housing stability for vulnerable people can be greatly increased by putting specific measures into practice, such as Housing First models, public-private partnerships, and zoning reforms. The project's models and policy assessments offer practical insights into the direct relationship between housing availability and homelessness, serving as a basis for national and municipal policy decisions.

Future Work

The following approaches are suggested in order to further improve this project's efficacy, scalability, and impact:

Including Real-Time Information Sources To enhance forecasting and targeting, future iterations might use additional datasets such as housing aid applications, eviction filings, rental platform APIs (e.g., Rent.com, Zillow), or urban mobility data.

Geographic and Temporal Mapping Regional housing policies and emergency preparedness could be improved by including disaster consequences, seasonal trends (such as winter-related displacement), and spatial features like accessibility to shelters or public transportation.

More Complex Predictive Modeling To predict homelessness trends based on variables like income levels, rental prices, and local economic indicators, ensemble approaches like Random Forests or XGBoost can be used.

Explainable AI for Transparency in Policy To ensure that decisions about housing policy are based on clear, intelligible data analytics, future research should use tools like SHAP or LIME to explain model outputs.

Deployment of Web and Mobile Applications
Housing agencies, organizations, and legislators might use an interactive portal created using frameworks like Streamlit, Flask, or React Native. This would make it possible to evaluate the danger of homelessness in real time using data inputs

Dashboards for Scalable Communities
To give decision-makers continuous insights, visual dashboards tracking the supply, demand, and policy impact of affordable housing might be developed.

Future efforts can shift toward more proactive, data-driven, and compassionate methods of resolving the housing crisis and lowering homelessness by implementing these tactics**.**

VII. REFERENCES

1. Amore, A., Bernardi, C. and Arvanitis, P. (2020), "The impacts of Airbnb in Athens, Lisbon and Milan: a rent gap theory perspective", Current Issues in Tourism, Vol. 25 No. 20, doi: 10.1080/13683500. 2020.1742674.
2. Baker, E., Mason, K. and Bentley, R. (2015), "Measuring housing affordability: a longitudinal approach", Urban Policy and Research, Vol. 33 No. 3, pp. 275-290.
3. Banco de Portugal (BdP) (2017), "Inquerito aSituaç~ao Financeira das Famılias 2017",availableat:https:// www.bportugal.pt/page/inquerito-situacao-financeira-das-familias (accessed 25 May 2022).
4. Banco de Portugal (BdP) (2020), "Inquerito a Situaç~ ao Financeira das Famılias 2020", available at: https://www.bportugal.pt/sites/default/files/isff2020_dest_pt.pdf (accessed 25 May 2020).
5. Banco de Portugal (BdP) (2021), Relatorio de Estaibilidade Financeira– Dezembro 2021, Banco de Portugal. Lisboa, ISSN 2182-0384, available at: https://www.bportugal.pt/publications/banco-de portugal/all/120 (accessed May 6 2022).
6. Demographia International (2009), "4th annual Demographia international housing affordability survey", available at: http://www.demographia.com/dhi-ix2005q3.pdf (accessed 27 April 2022).
7. Elmedni, B. (2018), The Mirage of Housing Affordability: An Analysis of Affordable Housing Plans in New York City, SAGE Open, doi: 10.1177/2158244018809218.
8. Farha, L. (2018), "Urban property: prices, ownership and occupancy. European Commission", available at:https://urban.jrc.ec.europa.eu/thefutureofcities/affordable-housing#the-chapter (accessed 21 February 2023).