Management (Montevideo). 2025; 3:115
doi: 10.62486/agma2025115
ISSN: 3046-4048

ORIGINAL BRIEF

Predictive Analytics for Housing Market Trends and Valuation

Análisis predictivo de tendencias y valoración del mercado inmobiliario

Awais Azam[1], Sakshi Rai[1], Shams Raza[2]

[1]Department of computer science LNCT University, Bhopal, India.
[2]Academic Counselor, IGNOU International Division, India.

Cite as: Awais Azam M, Rai S, Shams Raza M. Predictive Analytics for Housing Market Trends and Valuation. Management (Montevideo). 2025; 3:115. https://doi.org/10.62486/agma2025115

Submitted: 16-02-2024 **Revised:** 09-06-2024 **Accepted:** 15-10-2024 **Published:** 01-01-2025

Editor: Misael Ron

Corresponding author: Awais Azam

ABSTRACT

Introduction: the demand for housing in major cities is exceptionally high due to the concentration of offices and economic hubs in these areas. The combination of limited available land and increased demand drives house prices upward.
Objective: to accommodate this, developers are increasingly constructing residential areas on the outskirts of cities, offering easier access to transportation such as trains and highways. These developers compete by offering competitive pricing, diverse housing options, simplified mortgage processes, and attractive promotions like zero down payments. Buying a house is a significant long-term investment, as property values typically appreciate over time. Therefore, a thorough analysis is crucial when purchasing a home. Several key factors, such as location, land size, building area, and property type, play a role in determining house prices.
Method: this study adopts a quantitative approach, which involves systematically investigating phenomena by collecting measurable data and analyzing it through statistical, mathematical, or computational methods.
Result: this paper discusses the most effective techniques for data collection, pre-processing, feature extraction, model training, and evaluation. The purpose of this research method is to develop theoretical frameworks related to real-world phenomena.
Conclusions: measurement plays a pivotal role in this quantitative study, as it is central to understanding the data and drawing meaningful conclusions. Finally, we evaluate the current state of research, identifying trends and gaps in the field.

Keywords: Data; Computational Methods; House Prediction.

RESUMEN

Introducción: la demanda de vivienda en las grandes ciudades es excepcionalmente alta debido a la concentración de oficinas y centros económicos en estas zonas. La combinación de la escasez de suelo disponible y el aumento de la demanda impulsa al alza los precios de la vivienda.
Objetivo: para hacer frente a esta situación, los promotores inmobiliarios construyen cada vez más zonas residenciales en las afueras de las ciudades, que ofrecen un acceso más fácil a medios de transporte como trenes y autopistas. Estos promotores compiten ofreciendo precios competitivos, diversas opciones de vivienda, procesos hipotecarios simplificados y promociones atractivas como el pago inicial cero. Comprar una casa es una inversión importante a largo plazo, ya que el valor de los inmuebles suele revalorizarse con el tiempo. Por lo tanto, un análisis exhaustivo es crucial a la hora de adquirir una vivienda. Varios factores clave, como la ubicación, el tamaño del terreno, la superficie edificable y el tipo de propiedad, intervienen en la determinación del precio de la vivienda.

Método: este estudio adopta un enfoque cuantitativo, que implica la investigación sistemática de fenómenos mediante la recopilación de datos mensurables y su análisis a través de métodos estadísticos, matemáticos o computacionales.
Resultados: en este trabajo se analizan las técnicas más eficaces para la recopilación de datos, el preprocesamiento, la extracción de características, el entrenamiento de modelos y la evaluación. El objetivo de este método de investigación es desarrollar marcos teóricos relacionados con fenómenos del mundo real.
Conclusiones: la medición desempeña un papel fundamental en este estudio cuantitativo, ya que es esencial para comprender los datos y extraer conclusiones significativas. Por último, evaluamos el estado actual de la investigación, identificando tendencias y lagunas en este campo.

Palabras clave: Datos; Métodos Computacionales; Predicción Doméstica.

INTRODUCTION

The real estate market is one of the most critical sectors globally, where house prices play a pivotal role in determining economic conditions. The importance of predicting house prices accurately has grown with the increased availability of large datasets encompassing various attributes such as location, size, and market trends.[1] This has led to the emergence of data analytics as a tool to process these vast datasets and develop predictive models that can forecast house prices with increasing accuracy. In recent years, advancements in machine learning and artificial intelligence have revolutionized data-driven decision-making in real estate. Predicting house prices through data analytics involves gathering relevant datasets, pre-processing the data, and employing machine learning algorithms to derive meaningful predictions. The predictive models assist stakeholders in making informed decisions, mitigating risks, and maximizing returns.[2]

This paper seeks to review the methods and technologies employed in house price prediction using data analytics, shedding light on the challenges and opportunities presented by current methodologies.

METHOD

Data Sources

The dataset in question is a comprehensive collection of housing data, primarily used for predicting house prices. It contains various features related to the properties, such as physical characteristics, location, and sale conditions. The primary objective of analyzing this dataset is to understand the factors influencing real estate values and to develop predictive models for house prices. House price data is often obtained from multiple sources, including government databases, real estate websites, and proprietary datasets maintained by real estate companies.[3] Common attributes collected include the square footage of the house, the number of bedrooms, the location, proximity to schools and amenities, market trends, and interest rates. Publicly available datasets like the one provided by Kaggle (e.g., "House Prices - Advanced Regression Techniques") are often used for research and experimentation:[4] https://www.kaggle.com/datasets/srikanthladda/house-price-prediction

Data Analysis

The analysis involves several steps:

1. Data Exploration: Initial exploration of the dataset to understand its structure and contents.
2. Visualization: Creating visualizations to identify patterns and relationships between different features and the target variable, SalePrice.
3. Statistical Analysis: Applying statistical methods to quantify relationships and assess the significance of different features.
4. Modeling: Developing predictive models to estimate house prices based on the available features.

RESULT AND DISCUSSION

- **Distribution of Sale Prices**

The histogram shows the frequency of different sale prices, indicating a right-skewed distribution with most properties priced below a certain threshold (figure 1).

- **Living Area vs Sale Price**

The scatter plot illustrates a positive correlation between the ground living area and sale price, suggesting that larger living areas tend to have higher prices (figure 2).

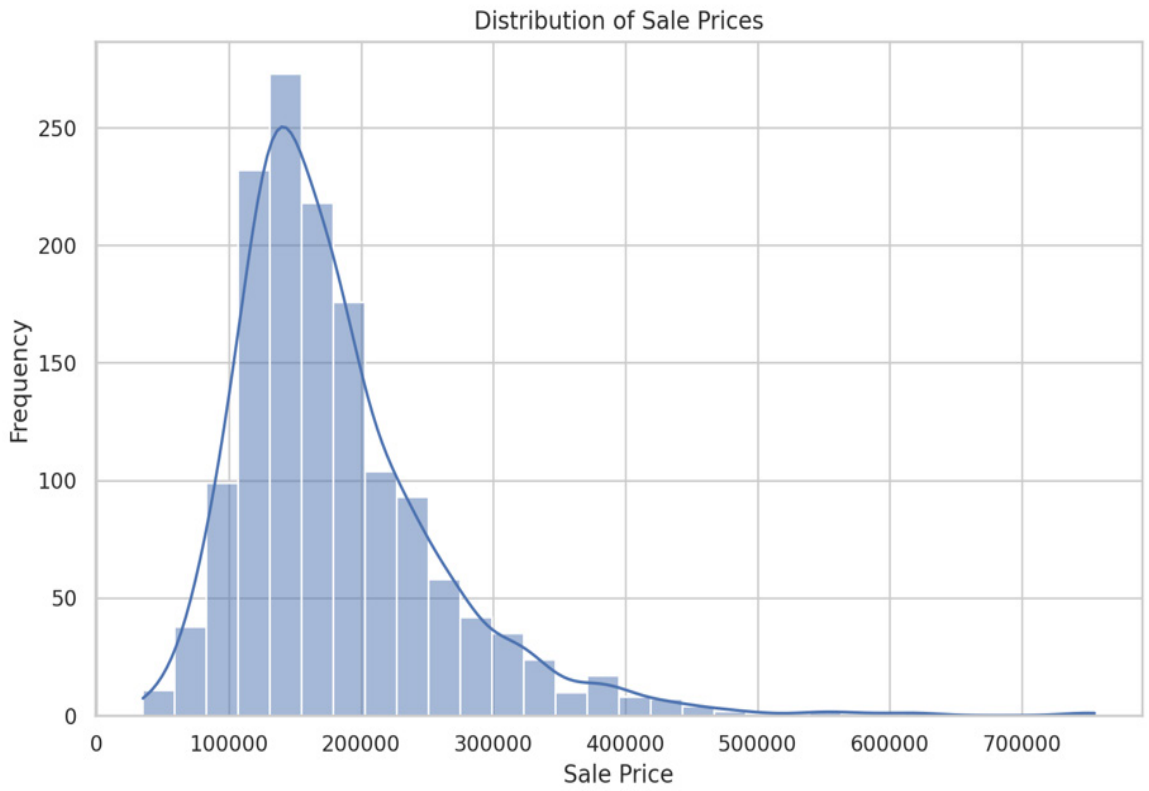

Figure 1. Distribution of Sale Prices[4]

Figure 2. Living Area vs Sale Price[4]

- **Sale Price by Neighborhood**

The boxplot reveals significant variation in sale prices across different neighborhoods, highlighting the impact of location on property values (figure 3).

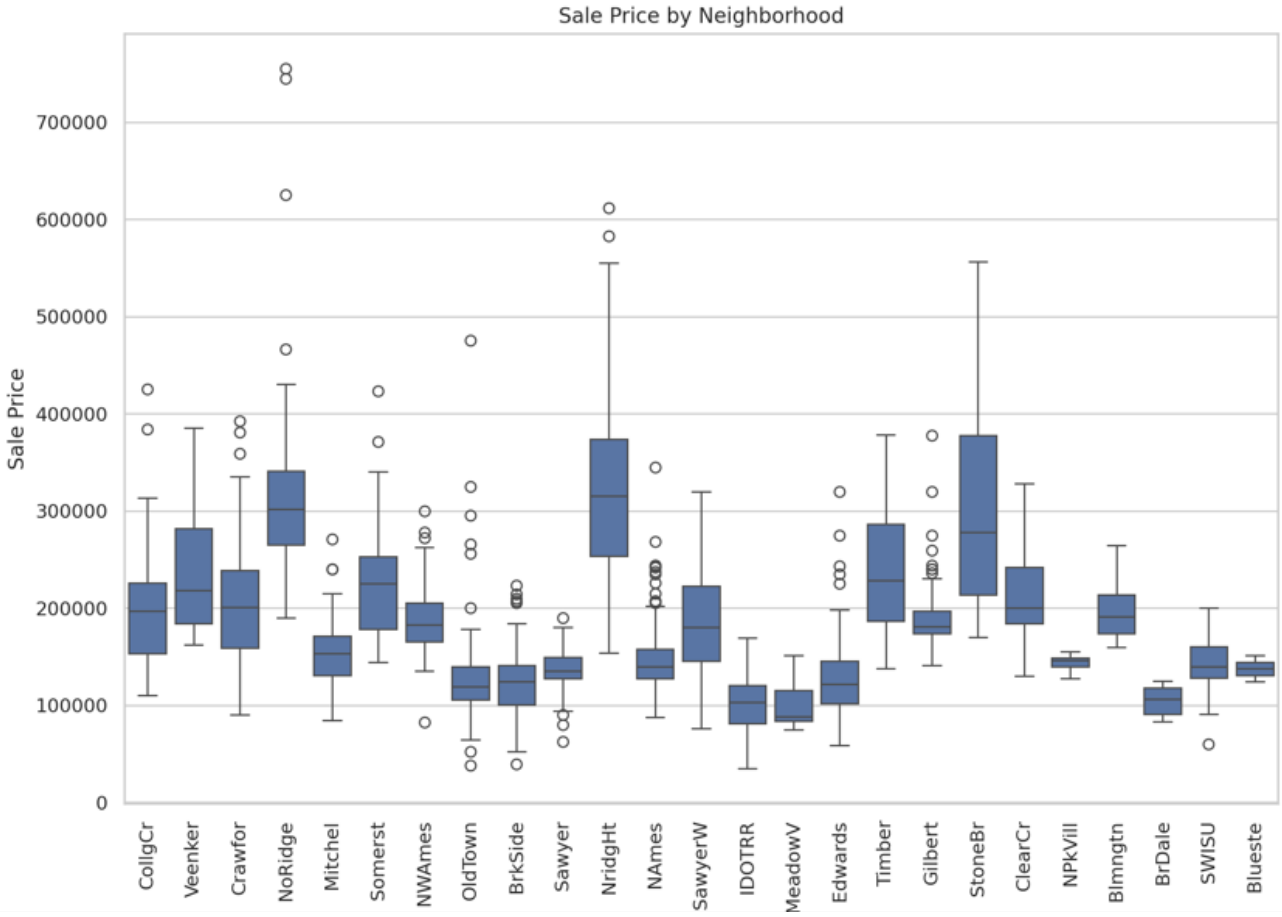

Figure 3. Sale Price by Neighborhood[4]

DISCUSSION

The results of the data analysis indicate that both the size of the living area and the neighborhood significantly influence housing prices. The histogram of sale prices shows a right-skewed distribution, suggesting that while most properties fall within a moderate price range, a small number of high-value properties considerably drive up the average price. This reflects the heterogeneity of the housing market, where luxury properties or those in premium locations can distort the overall market perception.

The positive correlation between the living area and sale price aligns with the expectation that buyers value space. This relationship suggests that size remains a crucial factor in purchase decisions, which has implications for developers and sellers as they consider the design and features of new properties. Additionally, the significant variation in prices across neighborhoods, as depicted in the boxplot, highlights the influence of location on real estate valuation. Properties in certain neighborhoods command higher prices due to factors such as proximity to amenities, schools, transportation, and the overall reputation of the area.

These findings emphasize the importance of considering multiple variables in real estate pricing. For stakeholders, such as real estate agents, buyers, and policymakers, understanding the impact of these factors is key to making informed decisions. For example, developers could use this insight to target investment in areas where larger living spaces are in demand, while buyers can focus on neighborhoods that match their investment goals. Policymakers, on the other hand, could leverage this data to inform urban development strategies, ensuring that new residential areas are equipped with the necessary amenities to attract potential buyers.

However, the study also faces certain limitations. The dataset, although comprehensive, does not account for all possible factors affecting house prices. Aspects like property age, recent renovations, zoning regulations, and proximity to upcoming infrastructure projects were not included in this analysis. Moreover, external economic factors, such as interest rates and market cycles, were also omitted. This exclusion can limit the accuracy of the predictive models, as real estate prices are often sensitive to broader economic conditions.

Additionally, the predictive models used are constrained by the quality and scope of the input data. While the study successfully demonstrates the utility of data analytics in identifying key determinants of house prices, improving model accuracy requires addressing challenges like data completeness, the inclusion of real-time market changes, and adopting more sophisticated algorithms. Incorporating additional data sources, such

as economic indicators, real-time sales data, and social factors, could enhance the model's robustness.

This study provides valuable insights into the key factors influencing house prices, reinforcing the importance of living area size and neighborhood characteristics in real estate valuation. While the use of data analytics in housing market prediction offers significant benefits for various stakeholders, its accuracy is contingent upon the inclusion of diverse and comprehensive datasets. Future research should aim to integrate additional variables and explore advanced modeling techniques to further refine predictions. Addressing these challenges will be crucial in developing more reliable tools for buyers, sellers, developers, and policymakers, ultimately contributing to a more informed and efficient housing market.

REFERENCES

1. Cox N. DISTPLOT: Stata module to generate distribution function plot. 2017.

2. Liu J. Dataset Analysis and House Price Prediction. Highlights Sci Eng Technol. 2024;81:363–7.

3. Lahmiri S, Bekiros S, Avdoulas C. A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. Decision Analytics Journal 2023;6:100166. https://doi.org/10.1016/j.dajour.2023.100166.

4. Lorenz F, Willwersch J, Cajias M, Fuerst F. Interpretable machine learning for real estate market analysis. Real Estate Economics 2023;51:1178–208. https://doi.org/10.1111/1540-6229.12397.

5. Kuchler T, Piazzesi M, Stroebel J. Housing market expectations□. In: Bachmann R, Topa G, van der Klaauw W, editors. Handbook of Economic Expectations, Academic Press; 2023, p. 163–91. https://doi.org/10.1016/B978-0-12-822927-9.00013-6.

6. Lee J, Lee KO. Online listing data and their interaction with market dynamics: evidence from Singapore during COVID-19. J Big Data 2023;10:99. https://doi.org/10.1186/s40537-023-00786-5.

7. Gourevitch JD, Kousky C, Liao Y (Penny), Nolte C, Pollack AB, Porter JR, et al. Unpriced climate risk and the potential consequences of overvaluation in US housing markets. Nat Clim Chang 2023;13:250–7. https://doi.org/10.1038/s41558-023-01594-8.

8. Kuchler T, Piazzesi M, Stroebel J. Housing market expectations□. In: Bachmann R, Topa G, van der Klaauw W, editors. Handbook of Economic Expectations, Academic Press; 2023, p. 163–91. https://doi.org/10.1016/B978-0-12-822927-9.00013-6.

9. Dou M, Gu Y, Fan H. Incorporating neighborhoods with explainable artificial intelligence for modeling fine-scale housing prices. Applied Geography 2023;158:103032. https://doi.org/10.1016/j.apgeog.2023.103032.

10. Shi S, Phillips PCB. Diagnosing housing fever with an econometric thermometer. Journal of Economic Surveys 2023;37:159–86. https://doi.org/10.1111/joes.12430.

11. Graham J, Makridis CA. House Prices and Consumption: A New Instrumental Variables Approach. American Economic Journal: Macroeconomics 2023;15:411–43. https://doi.org/10.1257/mac.20200246.

12. Javaid HA. Improving Fraud Detection and Risk Assessment in Financial Service using Predictive Analytics and Data Mining. Integrated Journal of Science and Technology 2024;1.

13. Tannady H, Pahlawi N, Hernawan MA, Arta DNC, Yusuf SD. Role of Stock Performance as an Intervening Variable in a Relationship Between Profitability, Leverage, Growth and Company Value. JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi) 2023;9:220–5. https://doi.org/10.35870/jemsi.v9i2.1016.

14. Yang L, Liang Y, He B, Yang H, Lin D. COVID-19 moderates the association between to-metro and by-metro accessibility and house prices. Transportation Research Part D: Transport and Environment 2023;114:103571. https://doi.org/10.1016/j.trd.2022.103571.

15. Devine A, Yönder E. Impact of Environmental Investments on Corporate Financial Performance: Decomposing Valuation and Cash Flow Effects. J Real Estate Finan Econ 2023;66:778–805. https://doi.org/10.1007/s11146-021-09872-y.

16. Klimberg R. Fundamentals of Predictive Analytics with JMP, Third Edition. SAS Institute; 2023.

17. Alam I, Ali Y. Studying the effects of Türkiye earthquake disaster and its impact on real estate industry: A risk analysis based on input-output & non-linear optimization models. International Journal of Disaster Risk Reduction 2023;96:103920. https://doi.org/10.1016/j.ijdrr.2023.103920.

18. Al-Okaily A, Teoh AP, Al-Okaily M. Evaluation of data analytics-oriented business intelligence technology effectiveness: an enterprise-level analysis. Business Process Management Journal 2023;29:777–800. https://doi.org/10.1108/BPMJ-10-2022-0546.

19. Xu X, Zhang Y. Contemporaneous causality among residential housing prices of ten major Chinese cities. International Journal of Housing Markets and Analysis 2022;16:792–811. https://doi.org/10.1108/IJHMA-03-2022-0039.

20. Xu X, Zhang Y. Network analysis of comovements among newly-built residential house price indices of seventy Chinese cities. International Journal of Housing Markets and Analysis 2024;17:726–49. https://doi.org/10.1108/IJHMA-09-2022-0134.

21. Deng QS, Alvarado R, Cheng FN, Cuesta L, Wang CB, Pinzón S. Long-run mechanism for house price regulation in China: Real estate tax, monetary policy or macro-prudential policy? Economic Analysis and Policy 2023;77:174–86. https://doi.org/10.1016/j.eap.2022.11.009.

22. Cox N. STRIPPLOT: stata module for strip plots (one-way dot plots). 2023;

23. House-price-prediction [Internet]. Available from: https://www.kaggle.com/datasets/srikanthladda/house-price-prediction

AVAILABILITY OF DATA AND MATERIALS

The datasets used in this research are publicly available(Kaggel) and properly cited in our dataset section for transparency and ease of replication: https://www.kaggle.com/datasets/srikanthladda/house-price-prediction

FUNDINGS

None.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Conceptualization: Awais Azam, Sakshi Rai and Shams Raza.
Investigation: Awais Azam, Sakshi Rai and Shams Raza.
Methodology: Awais Azam, Sakshi Rai and Shams Raza.
Writing - original draft: Awais Azam, Sakshi Rai and Shams Raza.
Writing - review and editing: Awais Azam, Sakshi Rai and Shams Raza.