Explore, Compare, and Predict Investment Opportunities through What-If Analysis: US Housing Market Investigation

Hongruyu Chen*
ETH Zürich
Zürich, Switzerland

Fernando Gonzalez*
ETH Zürich
Zürich, Switzerland

Oto Mraz*
ETH Zürich
Zürich, Switzerland

Sophia Kuhn*
ETH Zürich
Zürich, Switzerland

Cristina Guzman
ETH Zürich
Zürich, Switzerland

Mennatallah El-Assady
ETH Zürich
Zürich, Switzerland

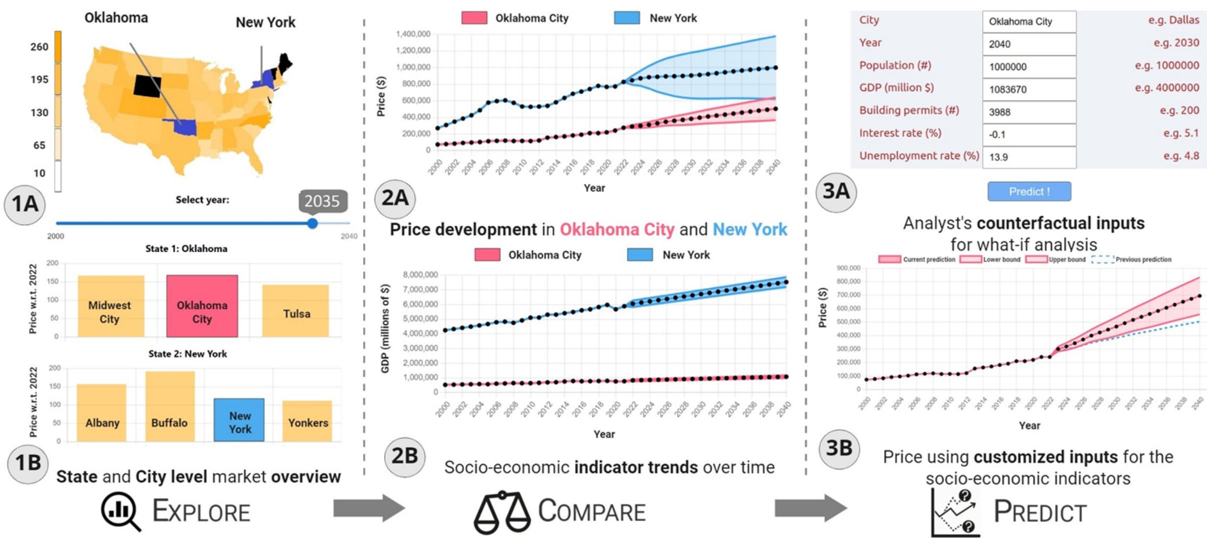

Figure 1: A human-computer interaction workflow for high-dimensional time series data analysis divided into three stages. (1) The exploration stage to get an overview of entities and features as in 1A and 1B. (2) The compare stage to compare the dynamic features over time as in 2A and 2B. (3) The prediction stage aimed for what-if analyses as in 3A and 3B.

ABSTRACT

A key challenge in data analysis tools for domain-specific applications with high-dimensional time series data is to provide an intuitive way for users to explore their datasets, analyze trends and understand the models developed for these applications through human-computer interaction. To address this challenge, we propose a three-stage workflow that allows domain experts to explore their data, compare the different entities' features, and predict the variable's long-term trend using what-if analyses. Based on this workflow, we created a data visualization workspace for real estate investment using data from the US housing market at state and city level. The underlying machine learning model ARIMAX uses house price data together with socio-economic data from 2000 to 2021 to learn the dependencies of the house prices on the socio-economic factors and make informative and robust predictions for future years.

*These authors contributed equally to this research.

VINCI 2023, September 22–24, 2023, Guangzhou, China

ACM ISBN 979-8-4007-0751-3/23/09…$15.00
https://doi.org/10.1145/3615522.3615542

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI); Interaction design**.

KEYWORDS

What-if analysis, Data visualization, Explainable artificial intelligence, Real estate investment, Time series data

ACM Reference Format:

Hongruyu Chen, Fernando Gonzalez, Oto Mraz, Sophia Kuhn, Cristina Guzman, and Mennatallah El-Assady. 2023. Explore, Compare, and Predict Investment Opportunities through What-If Analysis: US Housing Market Investigation. In *The 16th International Symposium on Visual Information Communication and Interaction (VINCI 2023), September 22–24, 2023, Guangzhou,* *China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3615522.3615542

1 INTRODUCTION

In recent years, Machine Learning (ML) applications continue to stretch into further decision-making fields including health care, education and business investment [4, 7, 8]. The growing abundance of data and models poses a challenge to keep such models interpretable and comprehensible to domain experts with little knowledge about ML theory. For the example of real estate (RE) investment, making wrong or suboptimal decisions based on poor understanding of ML model outputs could cause billion-dollar losses.

In particular, an RE investor will ask questions such as: *1) How can I easily identify the most profitable investment opportunities across the country? 2) How certain is the model about making a profit in city X? 3) From my knowledge, the population of city Y will decline twice as fast as it does now. How will this affect my revenue? 4) Which socio-economic driver affects the price development prediction the most?* This work proposes a workflow that answers these questions. More generally, it contributes to high-dimensional trend analysis following the logical thinking sequence of domain experts and guide them step-by-step through the ML decision-making process.

In this paper, we present, RealEstateGuru[1], a modular workspace that applies such a workflow to house market analysis and price forecasting in the United States (US). The tool is intended for an RE investor with sound expert knowledge in the field. RealEstateGuru aims to bridge the gap between visual analytics and interpretable ML methods, to assist the identification of the most profitable investment chances for RE investors. The tool uses house price and socio-economic indicator (SEI) data from 2000 to 2021 to predict the prices up to 2040. Furthermore, it allows users to use their domain knowledge to make customized predictions by feeding user-assessed values to the SEIs. As a result, users can empirically investigate the models' behavior in such counterfactual scenarios.

Our main contribution is two-fold: (1) we present a structured three-stage workflow to perform what-if analyses (see Figure 1) for high-dimensional trend analysis. (2) we introduce RealEstateGuru, a modular workspace that implements this workflow specifically for house price forecasting applications.

2 RELATED WORK

In the following paragraphs, we will introduce research from the Explainable AI (XAI), visualization, and RE fields tightly linked to our work. In numerous decision-making domains such as medicine or governmental policymaking, users strive for better control and understanding of the results produced by ML tools. Similarly, in the domain of RE, investors demand the ability to test alternate scenarios and a means to underpin the key model drivers.

What-If Analysis – What-if analysis has been widely used in causality, recommender systems, and XAI due to its efficiency in generating counterfactual hypotheses and causal inference. In the field of interactive ML, [30] developed the What-If Tool for testing model performance in hypothetical situations, and [10] designed Gaumut, a visual analytics system to support model interpretation.

[1] http://realestateguru.course-xai-iml23.isginf.ch/

Hypothesis Verification – When applying ML models to highly diverse fields, we must ensure the users have reasonable means to verify and scrutinize the outputs of the model. Naturally, RE investors will have limited knowledge about models such as ARIMAX. However, they can assess the reasonability of the predictions from domain expert knowledge. For example, [11] proposed a visualization-based method for hypothesis verification in the field of cosmology.

Structural Interpretability – Diving deeper, it is important to assess not only the model's decisions, but also why it makes them. ML experts have devised systematic theoretical techniques to this end, e.g., attribution methods [17, 24], activation and feature visualization [19]. However, for domain experts (in this case business investors), an empirical method is much more appropriate.

Real Estate Analysis – Numerous prior work has been dedicated to the efficient and accurate house price prediction. [12, 15] analyzed the influence of micro and macro factors on the RE prices. [16, 22, 28] explored conventional statistical and ML approaches such as SVMs, neural networks, or AdaBoost regression to develop housing price prediction models. In addition, visualization techniques have been applied to discover patterns in price data. [2, 20] presented a city-level map consisting of typical RE market territorial zones, [21] used a cartogram-based layout with PixelMaps to visualize geospatial, demographic, and price data, [13] deployed parallel coordinate systems to assess competitive pricing behavior.

However, limited attention has been given to Mixed-Initiative Systems which can tightly integrate visual analytics and model interpretability to support non-ML experts in leveraging their domain knowledge through human-AI interaction. The rising use of ML in novel domains necessitates additional studies on visual analytics in these interdisciplinary fields, which this work addresses. In particular, by closing the gap in a domain expert's understanding of an ML model, our work provides an effective means of communicating a model's sensitivity and confidence in a prediction.

3 REAL ESTATE DATASET

In this section, we introduce the data and how to clean and aggregate it for further analysis. RE market trends can be effectively predicted using SEIs, which influence the property depreciation and maintenance costs [9]. However, there is currently a lack of systematic documentation of RE market prices and SEIs. Therefore, we curated a dataset by aggregating multiple sources.

Price Data – We used a smoothed, seasonally adjusted measure of the typical home value (35th to 65th percentile), for our house price data. Such homes include single-family residences, condominiums, and co-operative homes. The data can be found on Zillow [31].

Socio-Economic Indicators – The selection of SEIs was made based on prior research in RE price prediction [3, 18]. We collected data for: a) Total city population (2000-21). b) Unemployment rate (2016-20). c) Annual Building Permits for US Core-Based Statistical Areas (2004-19) (a-c are available on the US Census website [27]). d) Annual GDP (2005-21) [5]. e) Lending Interest Rate (2000-21) [26].

Data Cleaning – To address missing values in the data, we used linear regression to approximate missing values for the years 2000-2021. For SEIs that were unavailable for certain time periods or cities, we estimated missing values using a statistical model to ensure consistent time intervals for all collected variables. If data for SEIs was not available for a specific city or state, we either used the (weighted) average of cities within the same state as a proxy or ignored the predictor entirely for price prediction in that city if no data was available for the entire state. We opted against using the US average to avoid introducing bias and misleading the model.

Data Aggregation – Given the disparate and heterogeneous nature of our data sources, we performed data aggregation to create a more comprehensive model input. Observing that GDP data was only available at the state level and interest rates are set centrally for the entire US, we replicated the values for all cities within the respective state to conform to the desired format.

4 REALESTATEGURU: A VISUAL ANALYTICS WORKSPACE

To align with investor investigations, our visual analytics workspace (RealEstateGuru) comprises three interlinked modules, namely Explore, showing a general overview of the US RE market, Compare enabling detailed comparisons of potentially profitable cities, and Predict where the user can perform customized dynamic predictions and what-if analyses. RealEstateGuru adopts a scrolling view arrangement that mirrors users' typical workflow. Furthermore, the three modules are interconnected via a clicking mechanism. The user's selections in the previous module populate the data in the next module. Consequently, the workspace enhances operational efficiency by means of direct transitions across modules.

4.1 Explore: High-Level Exploration

Explore provides a general overview of relevant entities and their features, facilitating high-level analyses. Its first submodule helps users understand the dynamic behavior of each data dimension and identify attractive entities for investigation. The second submodule enables high-level comparisons of individual entities within groups.

In the context of RE, Explore enhances the user's understanding of market development in different geographical regions of the US. It allows exploration of price and SEI evolution at the state and city level. The first submodule visualizes aggregated state-level statistics, considering variations in RE prices and investment revenues due to among others differing socioeconomic development, laws and property taxes among states [29]. The geospatial nature of our dataset advocates for a choropleth map in the first submodule, which depicts the distribution of the house prices, price change with respect to now, and the SEIs for each state in a chosen year. The color saturation represents the relative values compared to the other states and years for instant qualitative visual comparison. Hovering over the map displays the corresponding state values. Black states indicate missing data for the specified year.

Clicking on a state presents city-level data in bar plots, enabling high-level city comparisons (Figure 2). This granularity is crucial, as RE prices differ significantly across cities based on their attractiveness and human agglomeration. Besides a profound overview of the US RE market, Explore is especially useful for identifying the best investment opportunities by narrowing down regions with the highest expected investment return. As the profit of an RE investor comes from a rise in the market value, the price w.r.t to the purchase

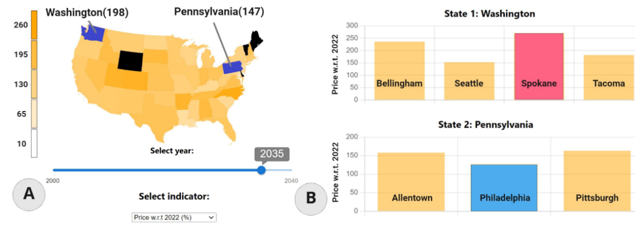

Figure 2: Interactive map of the US with bar plots displaying the percentage price change in 2035 with respect to 2022.

year is arguably the most informative value. It is determined by fixing the price today (2022) at 100% and then calculating the relative price in the remaining years. Consequently, the investor can select a year in which the property will be sold and directly acquires an overview of the percentage that the market values are expected to go up or down in the individual spatial regions, directly translating into investment yield. After the user identifies the potential cities to invest in, they can click on them and jump to the next module.

4.2 Compare: Detailed City Level Comparison

Compare facilitates a detailed comparison between entities of interest. It is highly application-dependent, and can be composed of several submodules fitting the nature of the data. Its primary function is to allow users to select relevant entities and compare their features over time. It also aids in formulating hypotheses to be investigated or verified in Predict.

With the identified investment opportunities and high-level market overview in mind, Compare facilitates RE investors in conducting detailed comparisons between two cities of interest. It assists users in performing an in-depth analysis to better understand how market prices change with the SEI development over time. Specifically, we utilized aligned line plots to depict the long-term price (Figure 1-2A) and selected SEI trends in the selected cities (Figure 1-2B). Users can select from around 130 US cities by clicking on the corresponding city bar plot in the previous module (Figure 2) or selecting from drop-down lists. Since we are predicting time series data, we deployed an ARIMAX model (described in §5).

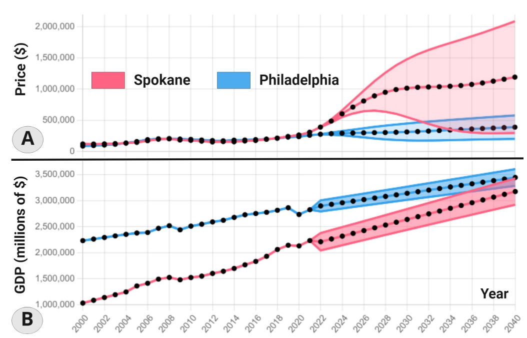

Figure 3: Trend comparison of the user-selected cities.

Users can scrutinize the predicted indicator development, especially when substantial data is missing, and validate it based on personal expertise and additional sources. This raises questions such as the potential impact of a doubled population growth rate in the next 10 years, which can be addressed using the following module. Furthermore, the absence of past and unknown future values underscores the inherent uncertainty in the model estimates. Hence, we constructed 95% confidence intervals (CI) to inform users about the certainty of the model predictions (colored region in Figure 3). Specifically, with 95% certainty, the true value falls within the predicted CI bounds, where the upper and lower bounds represent an optimistic and pessimistic scenario for the investor.

4.3 Predict: Customized Dynamic Prediction

In Predict, domain knowledge is utilized as input for the underlying ML model to generate personalized predictions based on customized estimates of the SEIs. The first submodule allows users to either use default values or specify alternative estimates for each feature. The second submodule displays the customized prediction and CIs. Predict serves the following purposes: conducting what-if analysis to validate counterfactual scenarios, enhancing the interpretability of the model, performing sensitivity analysis to assess model sensitivity to input changes, and providing recommendations for generating optimal outcomes as the next best action. This functionality is suitable for applications where the user has full control over the feature variables.

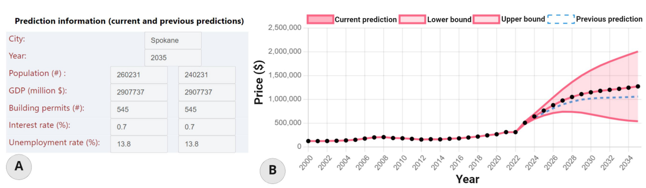

Figure 4: Personalized indicator table and dynamic house price predictions in Spokane. In this counterfactual example, population in the target year was increased by 10%.

The third module of RealEstateGuru facilitates forecasting via efficient human-AI interaction by integrating domain knowledge or non-public information into the model. In the first submodule (Figure 1-3A) each SEI can be assigned a user-defined value. The CIs from Compare serve as the guiding bounds for user-selected values; exceeding them may lead to highly unrealistic price predictions. The second submodule (Figure 1-3B) consists of a dynamic price line plot (including CIs) for the selected city and a table displaying the SEI values to keep track of the corresponding inputs that generate the prediction (Figure 4). Note that uncertainty analysis and risk assessment are crucial not only in RE, but also in other domains involving substantial financial or safety considerations [6]. Finally, to facilitate smooth interaction between the user and the model, a blue dashed line representing the previous prediction is included. Thanks to this design choice, the user can make an effortless comparison of two consecutive predictions.

5 ARIMAX PREDICTIVE MODEL

We employ the Autoregressive-Integrated Moving Average with Exogenous Variables (ARIMAX) model, extending ARIMA. A typical ARIMA model has three orders (as component parameters), p, d, and q, corresponding to the three components respectively: Autoregressive, Integrated, and Moving Average [14]. By aggregating the three components, ARIMA(p, d, q), leverages their strengths for analysis and forecasting [1]. The ARIMAX model extends ARIMA by adding exogenous variables to help measure the endogenous variable [25]. Exogenous factors are indirectly considered in the historical ARIMA model. However, including external data enables the model to respond more quickly to their effects compared to relying solely on lagging terms. The ARIMAX formula is given by:

$$\triangle^d Y_t = \beta + \sum_{s=1}^{r} \eta_s X_{s_t} + \sum_{k=1}^{p} \alpha_k \triangle^d Y_{t-k} + \sum_{k=1}^{q} \theta_k W_{t-k} + W_t, \quad (1)$$

where Y_t is the time series variable, t denotes the time step, $\triangle^d$ denotes the d-th order differencing, β is a constant, $\{\alpha_1, \alpha_2, \cdots, \alpha_p\}$ and $\{\theta_1, \theta_2, \cdots, \theta_q\}$ are the parameters, $\{W_{t-q}, W_{t-q+1}, \cdots, W_t\}$ are white noises, and $\{X_1, X_2, \cdots, X_r\}$ are exogenous variables (see Appendix for details). All exogenous series are assumed to be stationary. Otherwise, differencing all the non-stationary predictors is necessary during data preprocessing. We trained separate ARIMAX models for each city, where the endogenous variable is the price, and the exogenous variables include population, unemployment, GDP, building permits, and interest rate.

In Predict, the system incorporates user-specified SEI data as adjusted model inputs and generates personalized price forecasts. We linearly interpolate values between the last known year and the target prediction year instead of running linear regression, as linear regression would treat the new point as an outlier minimally affecting the predicted SEI values. Subsequently, the trained ARIMAX model is utilized to forecast prices in user-defined scenarios.

6 USE CASE

Next, we present a use case illustrating the integration of RE investors' domain knowledge into the model prediction through a human-AI interaction workflow using the example of Washington and Pennsylvania states. RealEstateGuru provides valuable insights by identifying less-known yet highly lucrative cities at first glance.

RE prices in Spokane are predicted to increase by 170% by 2035 in the first stage, vastly surpassing the expected 27% growth in well-known Philadelphia (Figure 2). In the second stage, investors can further investigate Spokane and Philadelphia. A line plot first displays the price development of the two cities to provide a global view (Figure 3), highlighting the previously identified higher profitability of Spokane. RealEstateGuru also prompts users to explore the influence of each SEI on the RE price, demonstrating its explainability power. E.g., the steeper increase in Spokane's GDP shown in Figure 3 is identified as a contributing factor to the local RE prices.

Human-AI interaction reaches its peak with what-if analysis in Predict. Investors can leverage specialized information on SEIs to examine market changes, such as Spokane's predicted population growth, which is caused by its growing, well-diversified economy and affordability [23]. RealEstateGuru allows investors to integrate this knowledge and conduct counterfactual what-if analyses (see Figure 4). Additionally, the investor can numerically test the model's sensitivity by modifying the input values. E.g., increasing the population by 10% in Spokane has a much greater effect on the price than doubling the number of building permits. This sensitivity analysis, together with CIs, provides insight into the significance of SEIs for future price predictions and enhances model explainability.

7 DISCUSSION

In this section, we discuss the important lessons learned, the identified limitations and promising future work opportunities.

Lessons Learned: Real Estate What-If Analysis – The development of the modular workspace has unleashed the underutilized potential of Mixed-Initiative Systems for ML applications in a broad range of domains. Many non-ML experts still treat ML models as a black-box. The modular workspace we propose assists domain experts in critically scrutinizing the model's behavior empirically via quick and simple experiments. While developed for the field of RE, the modular workspace has credit in substantially wider areas like urban transport design or Internet of Things ecosystem.

Limitation: Visualization Complexity – For each module we selected simple, intuitive visualizations to address the primary objective: workspace structure. While this choice limited our ability to simultaneously compare a larger volume of data by not harnessing multiple visualization channels, it also allows for flexibility. In specific use cases, individual visualizations can be substituted if they better serve the intended purpose.

Future Work: Expert evaluation – In order to deeper assess the benefits of our three-stage workflow for an RE investor, we plan to conduct comprehensive user studies with domain experts in the future. Through these studies, we aim to obtain a better grasp of the workspace's strengths and limitations.

8 CONCLUSION

In this paper, we propose a three-stage visual analytics workflow that helps users intuitively explore their data and perform what-if analyses. We implemented this workflow in a workspace named RealEstateGuru for an RE investment application. The workspace helps the user obtain a profound market overview, localize, understand, and evaluate predicted investment opportunities, and further make customized predictions. All in all, it enables the user to make robust decisions about when and where to invest in RE to maximize profit.

REFERENCES

[1] Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. 2014. Stock price prediction using the ARIMA model. In *UKSim-AMSS Int. Conf. on Comput. Modelling and Simulation*. IEEE, 106–112. https://doi.org/10.1109/UKSim.2014.67

[2] MA Bareicheva, EA Kubina, and NR Stepanova. 2021. Structural Analysis with Visualization Elements as a Factor in the Development of the Commercial Real Estate Market. In *IOP Conf. Series: Earth and Environ. Sci.*, Vol. 666. IOP Publishing. https://doi.org/10.1088/1755-1315/666/6/062130

[3] Alice Barreca, Rocco Curto, and Diana Rolando. 2018. Housing vulnerability and property prices: Spatial analyses in the Turin real estate market. *Sustainability* 10, 9 (2018), 3068. https://doi.org/10.3390/su10093068

[4] Rohan Bhardwaj, Ankita R Nambiar, and Debojyoti Dutta. 2017. A study of machine learning in healthcare. In *2017 IEEE 41st Annual Comput. Software and Applications Conf. (COMPSAC)*, Vol. 2. IEEE, 236–241. https://doi.org/10.1109/COMPSAC.2017.164

[5] Bureau of Economic Analysis. 2022. GDP and personal income. https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1&acrdn=1

[6] Sebastian Caldas, Jieshi Chen, and Artur Dubrawski. 2021. Using Machine Learning to Support Transfer of Best Practices in Healthcare. In *AMIA Annual Symp. Proc.*, Vol. 2021. American Medical Informatics Association, 265–274.

[7] Ana Isabel Canhoto and Fintan Clear. 2020. Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons* 63, 2 (2020), 183–193. https://doi.org/10.1016/j.bushor.2019.11.003

[8] Lijia Chen, Pingping Chen, and Zhijian Lin. 2020. Artificial Intelligence in Education: A Review. *IEEE Access* 8 (2020), 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

[9] Arthur Grimes and Andrew Aitken. 2007. House Prices and Rents: Socio-Economic Impacts and Prospects. *SSRN Electronic J.* (Mar. 2007). https://doi.org/10.2139/SSRN.975413

[10] Fred Hohman, Andrew Head, Rich Caruana, Rob DeLine, and Steven M. Drucker. 2019. Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. In *Proc. of the SIGCHI Conf. on Human Factors in Comput. Syst.* ACM. https://doi.org/10.1145/3290605.3300809

[11] Chung-Hsing Hsu, James P Ahrens, and Katrin Heitmann. 2010. Verification of the time evolution of cosmological simulations via hypothesis-driven comparative and quantitative visualization. In *IEEE Pacific Visualization Symp. (PacificVis)*. IEEE, 81–88. https://doi.org/10.1109/PACIFICVIS.2010.5429606

[12] Jun Kang, Hyun Jun Lee, Seung Hwan Jeong, Hee Soo Lee, and Kyong Joo Oh. 2020. Developing a forecasting model for real estate auction prices using artificial intelligence. *Sustainability* 12, 7 (2020), 2899. https://doi.org/10.3390/su12072899

[13] Bruce R. Klemz and Patrick M. Dunne. 2000. Exploratory Analysis using Parallel Coordinate Systems: Data Visualization in N-Dimensions. *Marketing Letters 2000 11:4* 11 (2000), 323–333. Issue 4. https://doi.org/10.1023/A:1008133112110

[14] Vijay Kotu and Bala Deshpande. 2019. Chapter 12 - Time Series Forecasting. In *Data Science* (2nd edition ed.). Morgan Kaufmann, 395–445. https://doi.org/10.1016/B978-0-12-814761-0.00012-5

[15] Sang-Hyang Lee, Jae-Hwan Kim, and Jun-Ho Huh. 2021. Land Price Forecasting Research by Macro and Micro Factors and Real Estate Market Utilization Plan Research by Landscape Factors: Big Data Analysis Approach. *Symmetry* 13, 4 (2021), 616. https://doi.org/10.3390/sym13040616

[16] CH. Raga Madhuri, G. Anuradha, and M. Vani Pujitha. 2019. House Price Prediction Using Regression Techniques: A Comparative Study. In *Int. Conf. on Smart Structures and Syst. (ICSSS)*. 1–5. https://doi.org/10.1109/ICSSS.2019.8882834

[17] Antonios Mamalakis, Elizabeth A Barnes, and Imme Ebert-Uphoff. 2022. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artif. Intell. for the Earth Syst.* 1, 4 (2022), e220012. https://doi.org/10.1175/AIES-D-22-0012.1

[18] Jen-Shi Ni, Shuen-Shi Huang, and Yu Wen. 2011. Interest rates, unemployment rate and house market in US. In *Int. Conf. on Social Sci. and Humanity. IACSIT Press, Singapore.*

[19] Chris Olah, Alexander Mordvintsev, and Ludwig Schubert. 2017. Feature Visualization. *Distill* (2017). https://doi.org/10.23915/distill.00007 https://distill.pub/2017/feature-visualization.

[20] Scott Orford. 2018. The Spatial Dynamics of an Urban Housing Market. *Valuing the Built Environ.* (2018), 103–155. https://doi.org/10.4324/9781315235134-5/SPATIAL-DYNAMICS-URBAN-HOUSING-MARKET-SCOTT-ORFORD

[21] Christian Panse, Mike Sips, Daniel Keim, and Stephen North. 2006. Visualization of Geo-spatial Point Sets via Global Shape Transformation and Local Pixel Placement. *IEEE Trans. on Visualization and Comput. Graphics* 12, 5 (2006), 749–756. https://doi.org/10.1109/TVCG.2006.198

[22] Byeonghwa Park and Jae Kwon Bae. 2015. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Syst. with Applications* 42, 6 (2015), 2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040

[23] Marco Santarelli. 2022. 21 best cities to invest in Real Estate 2022: Norada. https://www.noradarealestate.com/blog/best-places-to-invest-in-real-estate/.

[24] Sam Sattarzadeh, Mahesh Sudhakar, Konstantinos N Plataniotis, Jongseong Jang, Yeonjeong Jeong, and Hyunwoo Kim. 2021. Integrated grad-CAM: Sensitivity-aware visual explanation of deep convolutional networks via integrated gradient-based scoring. In *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*. IEEE, 1775–1779. https://doi.org/10.1109/ICASSP39728.2021.9415064

[25] Suhartono, Muhammad Hisyam Lee, and Dedy Dwi Prastyo. 2015. Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects. In *AIP Conf. Proc.*, Vol. 1691. AIP Publishing LLC, 050026. https://doi.org/10.1063/1.4937108

[26] The World Bank Group. 2022. Lending interest rate (%) - United States. https://data.worldbank.org/indicator/FR.INR.LEND?locations=US

[27] US Census Bureau. 2022. Census Data. https://data.census.gov/cedsci/

[28] Ayush Varma, Abhijit Sarma, Sagar Doshi, and Rohini Nair. 2018. House Price Prediction Using Machine Learning and Neural Networks. In *Int. Conf. on Inventive Commun. and Comput. Technol. (ICICCT)*. 1936–1939. https://doi.org/10.1109/ICICCT.2018.8473231

[29] John Joseph Wallis. 2001. A history of the property tax in America. *Property taxation and local government finance* (2001), 123–147.

[30] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. 2020. The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Trans. on Visualization and Comput. Graphics* 26, 1 (2020), 56–65. https://doi.org/10.1109/TVCG.2019.2934619

[31] Zillow research. 2022. Housing Data. https://www.zillow.com/research/data/