Evaluating the Impact of Corporate Investment on Columbus, Ohio Real Estate: A Machine Learning Perspective

By Siddharth Rajagopal

This study investigates the impact of Intel's $20 billion corporate development project on real estate prices and rent indices in Columbus, Ohio. The analysis employs advanced machine learning models, including Long Short-Term Memory (LSTM) networks and the Prophet forecasting model. The study compares pre- and post-announcement periods by modeling the pre-announcement data to forecast post-announcement trends, providing a benchmark for comparison. A one-tailed paired sample t-test was conducted to assess the statistical significance of the residuals. The results reveal a significant increase in residuals in Columbus post-announcement when compared to the broader US as comparison, suggesting accelerated market growth driven due to increased demand and positive sentiment. While isolating Intel's direct impact remains challenging, the findings provide a quantitative basis for understanding the broader effects of large-scale corporate investments on regional real estate markets. This research offers valuable insights for policymakers, investors, and stakeholders, and underscores the need for future studies with more granular data and diverse modeling techniques to fully capture long-term impacts.

Introduction

In recent years, Columbus, Ohio, has undergone a significant economic transformation, characterized by substantial investments from prominent technology companies such as Intel and Google. Intel's $20 billion development project, announced in January 2022, alongside Google's expanding data center operations, has positioned Columbus as an emerging hub in the Midwest. This influx of investment has profoundly impacted the local economic landscape, fostering growth in various ancillary industries. Most notably, this transformation has garnered national attention, enhancing Columbus's reputation as a favorable destination for corporate investment and expansion.

In addition to the technology sector, Columbus has experienced notable growth in its educational and health services sectors. The city's educational institutions, particularly The Ohio State University and the Wexner Medical Center, provide a robust talent pipeline that supports the region's expanding industries. This diversified economic base, coupled with infrastructure improvements and a conducive business environment, has made Columbus increasingly attractive to both investors and new residents. Anecdotal evidence from real estate professionals in Columbus suggests that the resulting population growth and economic activity have stimulated the local real estate market, with rising property values and rental rates becoming a prominent trend.

Columbus has recently been nationally recognized as one of the top emerging housing markets in the United States. Its affordability, coupled with a growing range of amenities and strong employment opportunities, has made it an appealing location for homebuyers and investors alike. This recognition has driven demand for residential and commercial properties, further invigorating the local real estate sector.

The overall economic environment in Columbus, shaped by strategic regional policies and favorable macroeconomic conditions, has been instrumental in sustaining this growth. The city's focus on creating a supportive ecosystem with abundant opportunities for the youth has attracted significant investment, supported by its advantageous location and infrastructure. As Columbus continues to establish itself as a key player in the national landscape, understanding the impacts of these developments on the local real estate market becomes increasingly important.

This investigation aims to quantify these impacts using machine learning models, focusing on changes in local real estate price indices following Intel's major investment announcement. By examining these dynamics, the research seeks to provide insights into the broader implications of corporate investments on regional real estate markets, offering valuable perspectives for policymakers, investors, and other stakeholders.

Review of Literature

The literature on the impact of corporate investments on real estate markets provides a robust theoretical foundation for understanding price dynamics and market behavior. Buyer and seller motivations significantly affect transaction prices in commercial real estate markets. Petrova highlights that conditions of sale, such as tax-deferred exchanges or portfolio sales, can lead to increased demand and higher selling prices due to the inelastic supply of available properties in commercial real estate markets. Specifically, Petrova notes, "Motivated buyers can create a temporary increase in demand in property markets. In perfect markets, a temporary increase in demand by motivated investors has no effect on market prices because supply can instantaneously respond. However, in commercial real estate markets the supply of available properties is less elastic to shocks in demand" (2006). Loewenstein and Willen (2022) also discuss how changes in beliefs about price growth, known as expectation shocks, can drive up the price-rent ratio, often leading to self-fulfilling price increases. They argue that "positive expectation shocks drive up the price-rent ratio," which can be a plausible explanation for housing market booms.

Chaney, Sraer, and Thesmar emphasize the collateral channel, demonstrating that real estate appreciation directly influences corporate investment decisions. Their study found that for every \$1 increase in a firm's real estate value, investment increased by approximately \$0.06, financed through additional debt issuance. This effect is particularly pronounced in firms likely to be credit constrained, underscoring the significant role real estate plays in broader economic activities. They state, “The impact of real estate shocks on investment is stronger when estimated on a group of firms which are more likely to be credit constrained” (2012).

McAllister and Nanda explore the impact of foreign investment on U.S. office real estate prices, finding that increased foreign investment significantly affects market cap rates. Their analysis across 38 U.S. metro areas revealed that a 100 basis point increase in the foreign share of total investment leads to an approximate 8 basis point decrease in market cap rates. They conclude that, “Foreign investment significantly impacts U.S. office real estate prices, demonstrating the substantial influence of global capital on local real estate markets” (2018). Gorback (2022) further explores the sensitivity of the U.S. housing market to global capital, finding that a 1% increase in instrumented foreign capital raises house prices at the zip code level by 0.27%. This highlights the interconnectedness of local real estate markets with global financial flows. “U.S. housing markets seem relatively inelastic in the short run,” Gorback notes, indicating the profound impact of foreign investment on local markets.

Hedonic Price Theory sets the foundation for real estate valuations, where property values are attributed from a decomposition from its individual characteristics. It explains how, location, macro-dynamics, and the asset itself, all are factors in the overall valuation. However, it is how these multiple independent variables are correlated with each other (multi-linearity) that make it difficult to predict the final value. Chou, Fleshman, and Truong discuss the prevalence of hedonic price theory in real estate as well as a basis for how machine learning can aid with forecasting and valuation. They highlight that while traditional hedonic models face issues like multi-linearity , machine learning techniques offer solutions by efficiently processing complex and non-linear data to predict house prices accurately. From the text, “many researchers have introduced techniques, such as those based on artificial intelligence, to solve these problems…machine learning is implemented to construct models that mimic human reasoning; consequently, it can deduce new facts from historical data and respond adaptively to changes of previously obtained information” (2022).

Geltner (2022) discusses the complexities of constructing real estate price indices, emphasizing the challenges posed by the heterogeneous nature of real property assets and the sparse transaction data available. He notes that these indices require sophisticated methodologies to account for volatility, cyclicality, and location-specific factors that influence price dynamics. “Indices cannot be computed directly from property market transactions data with simple matched samples or with stock market price indexing procedures,” Geltner points out.

Further basis of machine learning in forecasting real estate values has been well-documented in the literature. Parmezan et al. (2019) note how time series prediction methods rely on historical data patterns to forecast future values. These models, including ARIMA, LSTM, and Prophet, capture the dynamic relationships within the data, providing robust predictions for future trends. According to Parmezan, “The methods for time series prediction rely on the idea that historical data include intrinsic patterns which convey useful information for the future description of the phenomenon investigated” . The Prophet model, heavily used in this investigation, in particular, has been praised for its ability to handle time series data with strong seasonal effects and missing data, making it a versatile tool in real estate market analysis. Kramar and Alchakov highlight that “Prophet is a procedure for forecasting time series data based on an additive regression model, where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects” (2023).

Du et al. advocate for the use of combination-based ensemble models in time series forecasting, arguing that these models outperform single models by leveraging the strengths of multiple forecasting techniques. This approach mitigates the risk of model collapse and enhances predictive accuracy, making it particularly suitable for complex real estate markets. They explain, “The superiority of the forecast combination over single ones is evidenced… model selection and weighting can mitigate extreme situations like model collapse caused by structural misspecification” (2022). The integration of machine learning in real estate research is further supported by Choy and Ho (2023), who highlight the effectiveness of these methods in analyzing and forecasting property values. They note that machine learning algorithms, such as Extra Trees, k-Nearest Neighbors, and Random Forest, have been found to outperform traditional hedonic price models. “Accurate price signals in the property market predicted by machine learning algorithms play an important role in promoting sustainable production and consumption patterns,” they assert.

As shown, the literature provides a clear basis in terms of the significant impact of corporate investments and global capital on real estate markets. The application of advanced analytical methods, particularly machine learning, offers robust tools for forecasting and understanding market dynamics. These insights provide a valuable framework for continuing this investigation into the effects of Intel’s announcement on the local real estate market.

Methodology

The analytical framework for this study hinges on a comparative analysis of real estate index data from periods before and after the announcement of Intel's project. The core objective is to identify any significant deviations in real estate prices and rent indices that can be attributed to the announcement. To provide a basis, a primary analysis was used to identify if there is an increase in rate of change and absolute change post announcement. This primary analysis of simply taking means, of each set, showed that post-announcement data had higher means then pre-announcement. As COVID provided volatility in the market, machine learning forecast models were employed as a method to standardize this period of volatility for the Columbus region and Broader-US control region.

Three primary index data sources were used. The FHFA House Price Index (HPI) tracks single-family home value changes across the U.S. and Columbus, using data from over 400 cities since the mid-1970s. It employs a weighted, repeat-sales method for detailed national, regional, and local insights, retrieved from FRED, Federal Reserve Bank of St. Louis. The Redfin Home Price Index (RHPI) monitors the sale prices of the same homes over time, offering seasonally adjusted insights that control for home size and quality variables. The Market Asking Rent Index tracks asking rents for industrial, retail, and multifamily properties in Columbus and Broader US, providing a comprehensive view of rental market trends.

To accurately differentiate between these periods and understand whether observed changes were part of long-term macroeconomic trends (real estate boom post-COVID) or specific to the announcement, machine learning models were employed. These models were trained on pre-announcement data to learn the underlying patterns and subsequently forecast what the market would have looked like without the announcement, providing a benchmark for comparison.

The study initially segments the data into two distinct periods: the pre-announcement and post-announcement phases. The pre-announcement phase serves as the baseline, capturing the historical trends and behaviors in the real estate market prior to the announcement. This phase is critical for establishing a reference point against which subsequent changes can be measured. The post-announcement phase was then analyzed to identify any deviations from the baseline trends, which could potentially signal the influence of sentiment from the announcement on the market.

A central component in this is the residual analysis. Residuals, in this context, refer to the differences between observed market data and the predicted values from the models. These residuals are essential for understanding the extent to which the actual market outcomes deviate from the expected outcomes based on historical patterns. To assess the statistical significance of these deviations, a one-tailed paired sample t-test is employed. This statistical test compares the residuals from the Columbus market with those from a broader U.S. market control group. The underlying hypothesis posits that the residuals in Columbus, post-announcement, are significantly larger than those in the broader market, suggesting a localized impact due to the development project. The use of a one-tailed test is predicated on the expectation that the announcement would likely lead to an increase in real estate indices, driven by heightened demand and speculative activities.

In addition to statistical testing, the methodology includes processes for normalizing and standardizing the data. All absolute residuals are normalized to reflect percentage changes rather than absolute values. This normalization process is vital for ensuring that comparisons across different indices and regions are meaningful and accurate. It controls for variations in the scale and units of the indices, thereby eliminating potential distortions that could arise from comparing data on different scales.

To enhance the predictive accuracy and robustness of the analysis, the study employs advanced machine learning models. Initially, Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN) were utilized to model the pre-announcement data and forecast trends for the post-announcement period. LSTM RNNs are particularly well-suited for time-series forecasting due to their ability to capture long-term dependencies and sequential data patterns. These models can effectively handle the complexities and nuances of real estate market data, which often exhibit non-linear and volatile behaviors. To optimize the performance of the LSTM models, Bayesian optimization techniques were applied. This process involves systematically adjusting the hyperparameters of the models, such as the number of layers, neurons, and learning rates, to achieve the best possible fit to the data.

Despite the advanced capabilities of LSTM models, certain limitations were encountered. These included the potential for overfitting, where the model becomes too tailored to the training data and loses generalizability to new data. Additionally, the computational demands of LSTM models, in terms of processing power and time, presented challenges. This contributed to the shift to the Prophet machine learning model as the primary in this investigation.

This investigation majorly incorporated the Prophet machine learning model developed by Facebook, as an alternative forecasting tool. The Prophet model is specifically designed for time-series data with strong seasonal components and is robust to missing data and outliers. It decomposes time-series data into trend, seasonal, and holiday components, providing a clear and interpretable representation of the underlying patterns.

It’s form is shown in this equation:

$$y(t) = g(t) + s(t) + h(t) + t$$

> “Here g(t) is the trend function which models non-periodic changes in the value of the time series, s(t) represents periodic changes (e.g., weekly and yearly seasonality), and h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term t represents any idiosyncratic changes which are not accommodated by the model; later we will make the parametric assumption that t is normally distributed” (J. Taylor & Letham, 2017).

The Prophet model’s strengths lie in its simplicity and ease of use, making it accessible for quick iteration and exploration of different forecasting scenarios. It allows for the inclusion of seasonal effects at multiple granularities, such as daily, weekly, and yearly, which is particularly beneficial in real estate markets where seasonality can have a pronounced impact. Similar to the approach originally taken with LSTM, the model takes advantage of Bayesian optimization in which it automatically optimizes model characteristics. Furthermore, the Prophet model’s robustness to outliers ensures that sudden market anomalies, such as those caused by external economic shocks, do not unduly influence the results.

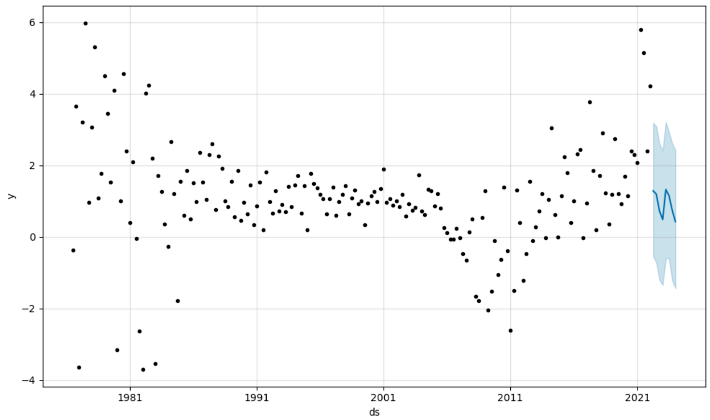
The primary hypothesis of this study is that Intel's announcement of its $20 billion corporate development project significantly impacted the Columbus real estate market. Specifically, it hypothesizes that the residuals (differences between observed and predicted values) for Columbus are significantly larger than those for the broader U.S. market following the announcement. This hypothesis stems from the expectation that such a substantial investment would drive up local real estate prices and rents due to increased demand and positive market sentiment.

To test this hypothesis, a one-tailed paired sample t-test was conducted. This statistical test compares the means of two related groups to determine whether there is a statistically significant difference between them. In this context, the t-test was used to compare the residuals of the Columbus real estate market with those of the broader U.S. market. The following steps outline the statistical analysis process:

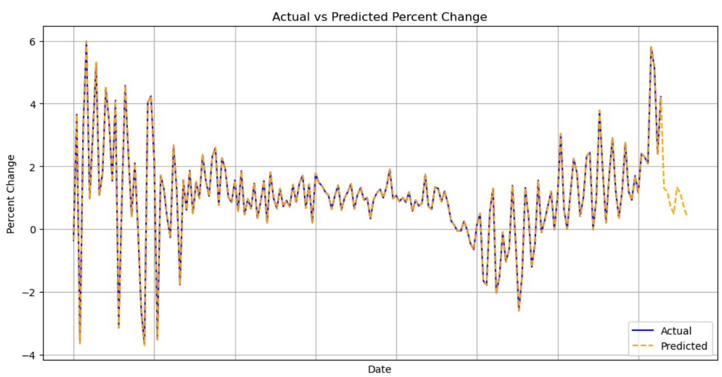
1. **Calculate Residuals**: Residuals were calculated for each dataset (FHFA HPI, Redfin HPI, and Market Asking Rent Index) by subtracting the predicted values from the actual observed values.
2. **Normalize Residuals**: The residuals were normalized to ensure consistency and comparability across different indices and datasets.
3. **One-Tailed Paired Sample T-Test**: The t-test was applied to the normalized residuals to determine if the residuals for Columbus were significantly larger than those for the U.S. market. The null hypothesis (H0) posited that there is no significant difference between the residuals of the two markets, while the alternative hypothesis (H1) posited that the residuals for Columbus are significantly larger.
4. **Significance Level**: The test was conducted at a 0.05 significance level. A p-value less than 0.05 would lead to the rejection of the null hypothesis, supporting the claim that Intel's announcement had a significant impact on the Columbus real estate market.
5. **Results Interpretation**: The results were interpreted to determine whether the observed differences in residuals were statistically significant, providing insights into the impact of Intel's development project on local real estate prices and rents.

To visualize the data, descriptions of each significant chart are presented below.

Prophet Visualization Charts

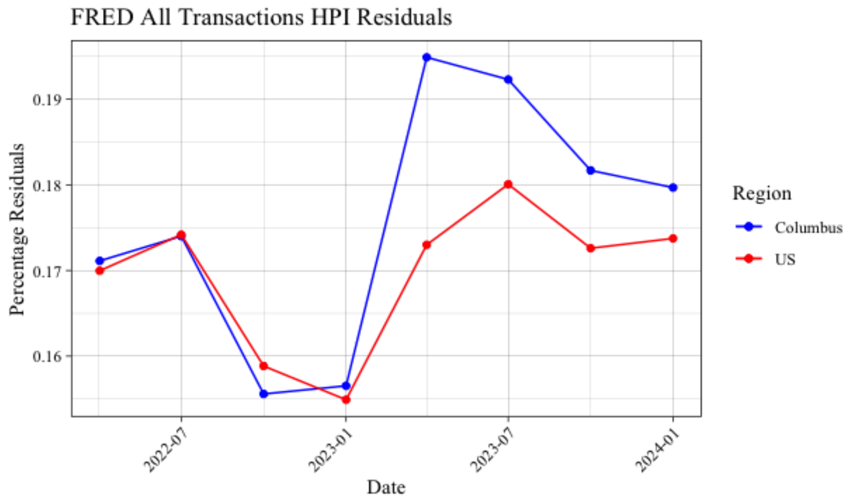


This time series plot visualizes the index over time. The chart depicted includes the variable y on the y -axis and the date ds on the x -axis. The black dots represent the observed change points, while the blue line, accompanied by a shaded region, indicates the forecasted values along with their uncertainty intervals. The inclusion of uncertainty intervals provides insight into the confidence levels of these forecasts.

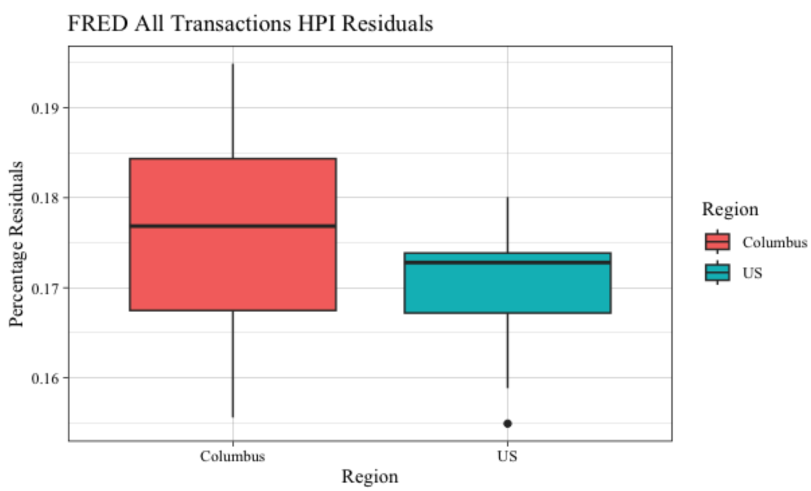


The chart presented displays the change on the y -axis and the date on the x -axis. The red/blue line represents the actual observed percent changes, while the dashed orange line indicates the predicted percent changes produced by our model. It aids to the first chart above with a clearer point of view of the seasonality and characteristics of the model.

Residuals Analysis Charts

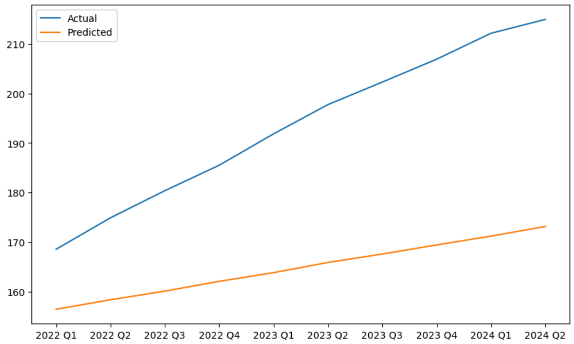


The chart shown illustrates the residuals on the y -axis and the date on the x -axis. The blue line represents the residuals for Columbus, while the red line depicts the residuals for the entire United States. The between Columbus and the national data helps to contextualize the local real estate market's performance over time. This visualization aids in identifying any systematic biases or patterns in the residuals.

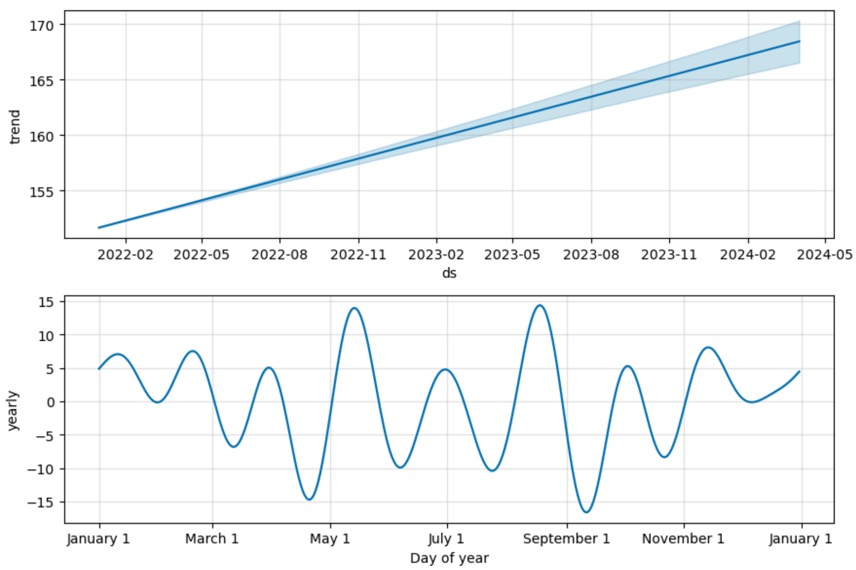


The chart presented shows the percentage residuals on the y -axis, with separate box plots for Columbus and the US on the x -axis. The red box plot represents the residuals for Columbus, while the teal box plot depicts the residuals for the entire United States. These box plots allow for understanding the variability and central tendency of the residuals from our predictive models.

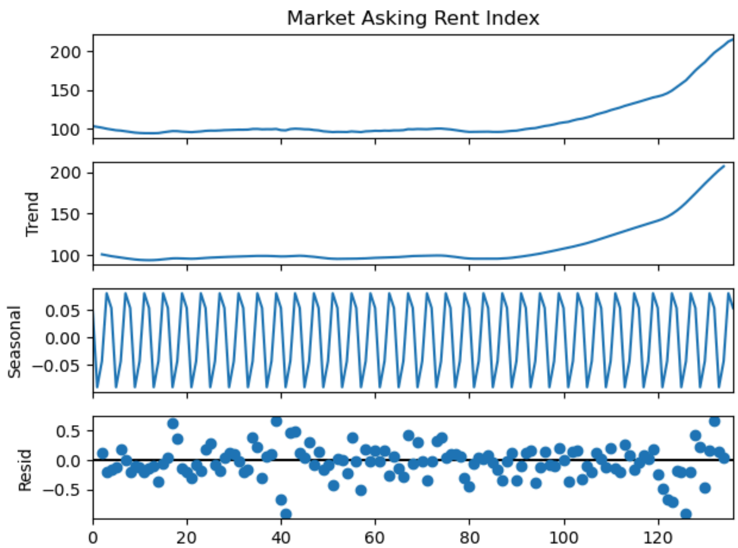
Other Charts Used



This chart presented shows the index values on the y -axis and the quarters on the x -axis, spanning from Q1 2022 to Q2 2024. The blue line represents the actual observed values of the index, while the orange line indicates the values predicted by our model. By directly comparing the actual data with the predicted values, we can assess how well our model captures the trends and variations over time. The clear visual representation of actual versus predicted values provides a framework for the further analysis.



This trend and seasonal decomposition plot was used for analyzing each dataset. The chart presented includes two subplots: the trend component and the yearly seasonal component. The top panel displays the trend component over time, indicating the long-term progression of the index from early 2022 to mid-2024. The blue line represents the trend, and the shaded region illustrates the confidence interval around the trend prediction. The bottom panel shows the yearly seasonal component, capturing the recurring seasonal patterns within the data on a yearly basis.

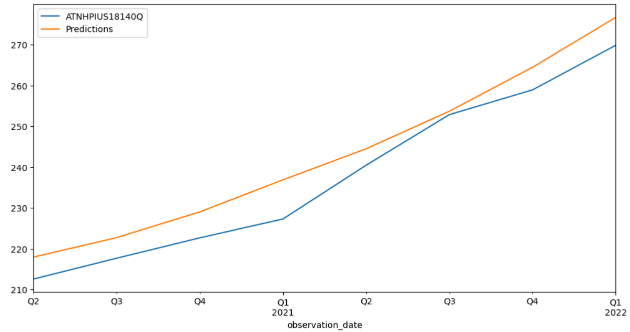


This decomposition plot was utilized to analyze the components of each index. The chart presented is divided into four subplots: the original data, trend, seasonal, and residual components. The top panel shows the index over time. The second panel depicts the trend component, illustrating the underlying long-term movement in the data. The third panel displays the seasonal component, capturing the repeating patterns or cycles within the data. The bottom panel shows the residual component, which represents the remaining variability after removing the trend and seasonal effects.

Full data for each model is included in the attached appendices.

Results

LSTM Results Summary



Throughout the extensive analysis phase, a Long Short-Term Memory (LSTM) model was initially employed to predict real estate price indices. The process began with a thorough exploration of LSTM architecture. Different models were built using several layers, including LSTM layers with varying nodes (units) and activation functions, and were optimized using the Adam optimizer. The model underwent multiple trials and errors, focusing on hyperparameter optimization to improve its predictive accuracy.

The LSTM model was trained over 50 epochs, with an initial configuration of nodes and learning rates. The accompanying chart (Figure 1) illustrates the training and validation loss across epochs. Despite multiple attempts to fine-tune the hyperparameters, including adjustments to the number of nodes, learning rate, and batch size, the model failed to achieve satisfactory convergence. The training loss remained near zero, while the validation loss plateaued at a significantly higher value, indicating potential overfitting. During the training phase, a series of hyperparameter optimization techniques were employed, including grid search and Bayesian optimization, to identify the optimal configuration. These methods involved systematically varying the number of epochs, batch size, and the number of LSTM units. Despite these rigorous optimization efforts, the model's performance on the test set remained suboptimal.

The subsequent chart (Figure 2) compares the actual and predicted values for the test set. While it showed strength in the testing set, this would not remain evident when applied to the full index datasets. Several other challenges were encountered throughout the LSTM modeling phase. These included extensive trials with different hyperparameters, such as the number of LSTM units, learning rate, and dropout rates, which yielded minimal improvement in model performance. The model exhibited signs of overfitting, where it performed well on the training data but poorly on the validation and test data. Issues related to data scaling and sequence generation were meticulously addressed, yet the model's predictive power remained insufficient. Despite multiple consultations with experts and iterative debugging sessions, the complexity of LSTM tuning proved to be beyond the project's expertise level.

Given the expertise limitations observed with the LSTM model, a strategic decision was made to transition to the Prophet forecasting model. Prophet, developed by Facebook, is known for its simplicity and effectiveness in capturing seasonal trends and handling missing data points, making it a more suitable choice for the given dataset and project scope. This transition was necessitated by the realization that the intricate tuning and debugging of LSTM models required a higher level of expertise than was available within the project's constraints.

Forecasting Results

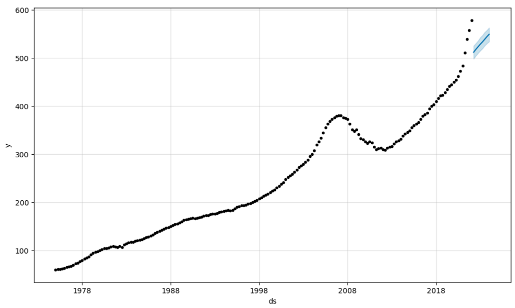
Dataset	P-Value
FRED Full Index	0.0370560728
FRED Percent Change	0.4110931809
REDFIN HPI MoM	0.0000001500
MARI Industrial	0.0000024464
MARI Industrial ROC	0.0000001573
MARI Multifamily	0.0107367470
MARI Multifamily ROC	0.0003583070
MARI Retail	0.0033230000
MARI Retail ROC	0.0001540780

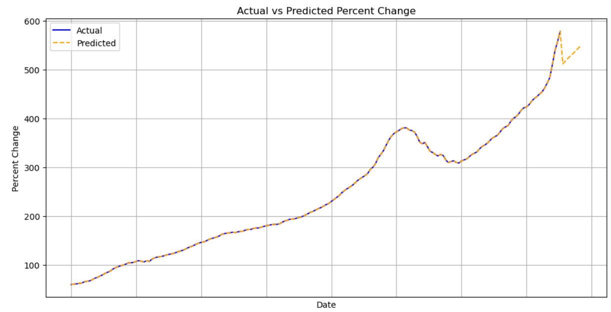
Dataset	RMSE
FRED Full Index	
- US	108.8103326986
- Columbus	53.6412022617
FRED Percent Change	
- US	2.2581253838
- Columbus	2.7572270301
REDFIN HPI MoM	
- US	0.0138617964
- Columbus	0.0085850374
MARI Industrial	
- US	13.1843284443
- Columbus	30.3719089777
MARI Industrial ROC	
- US	0.0083418806
- Columbus	1.5441291318
MARI Multifamily	
- US	7.2163510394
- Columbus	9.3473712490
MARI Multifamily ROC	
- US	0.0093971147
- Columbus	0.0106927934
MARI Retail	
- US	3.0635842072
- Columbus	4.4421086393
MARI Retail ROC	
- US	0.005061441
- Columbus	0.006793042

Prophet Visualizations

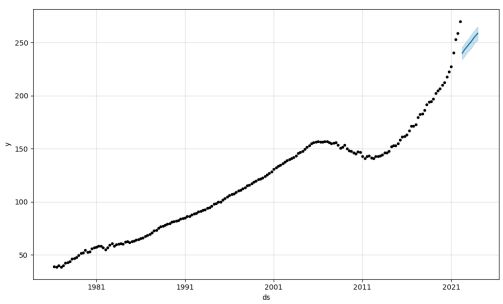
FRED All Transactions Housing Price Index

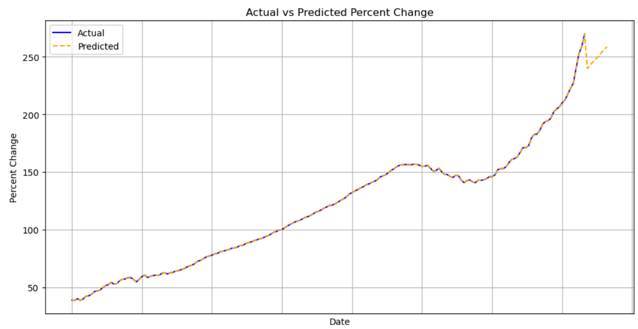
US





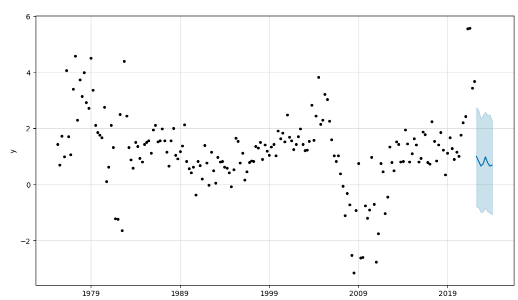
Columbus

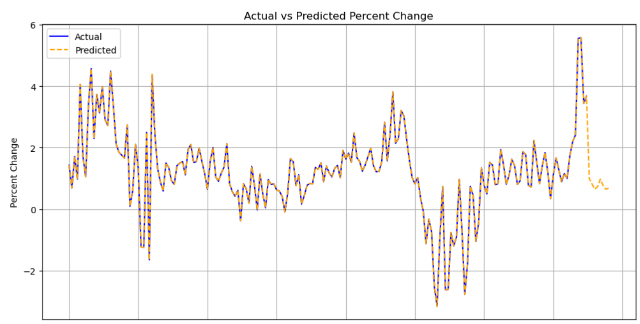




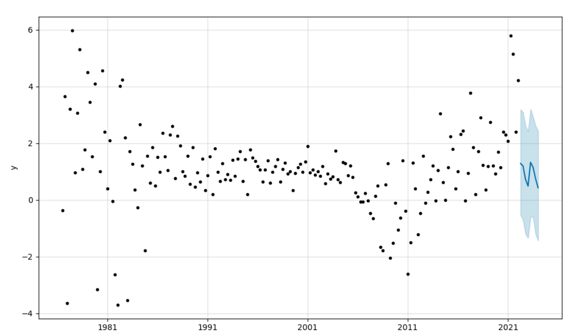
FRED All Transactions Housing Price Index - Percent Change

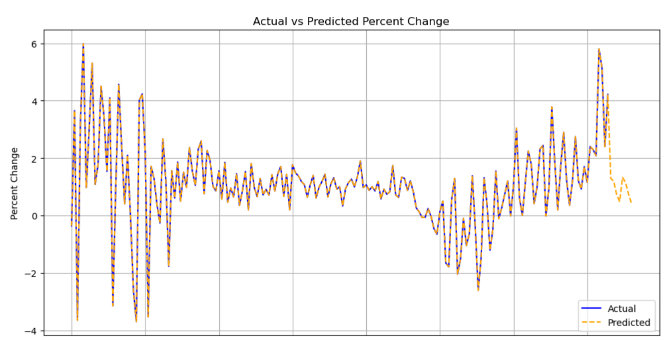
US





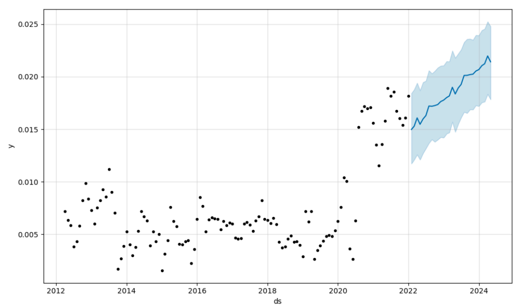
Columbus

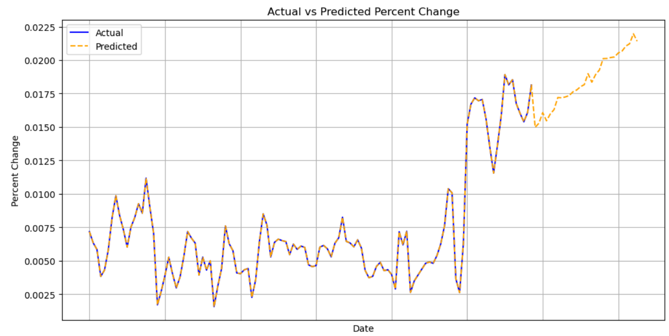




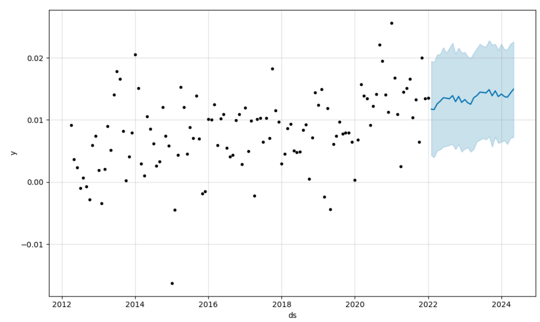
Redfin HPI MoM

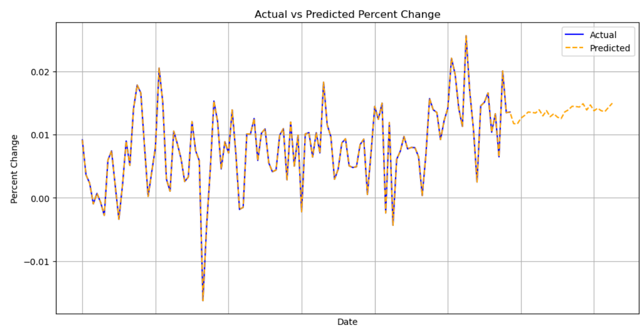
US





Columbus

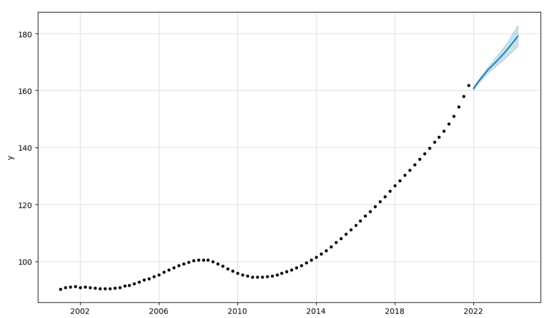


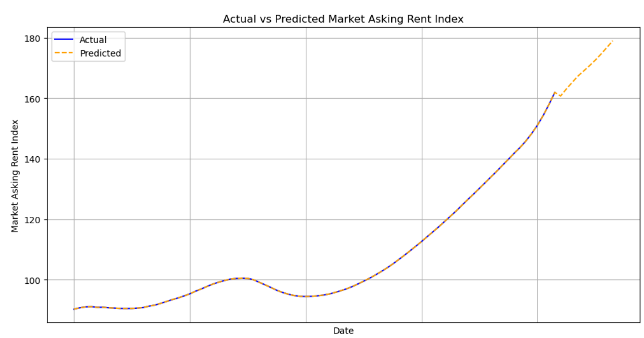


u

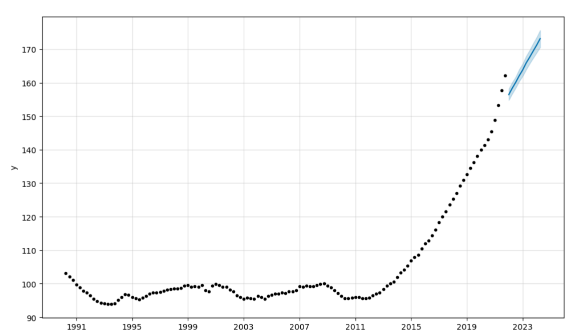
Market Asking Price Index (MARI) — Industrials

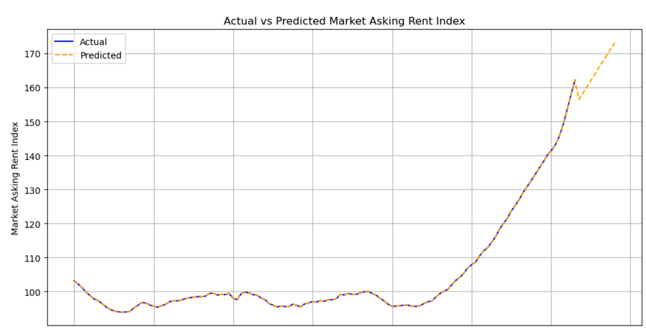
US





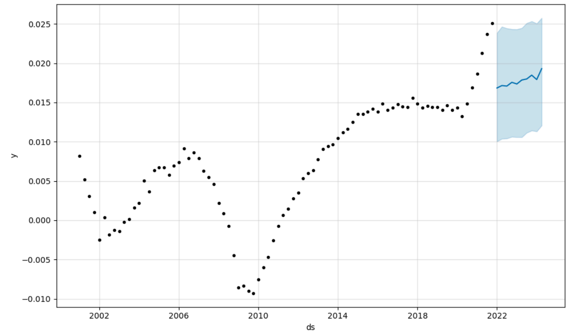
Columbus

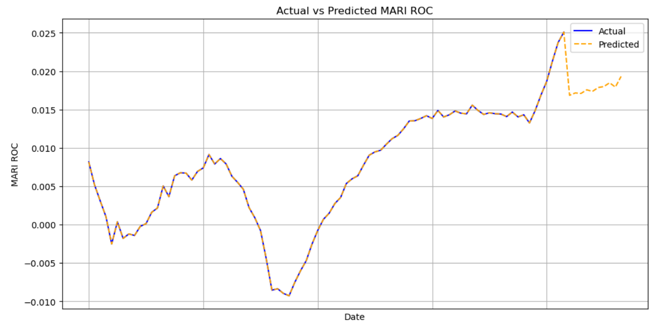




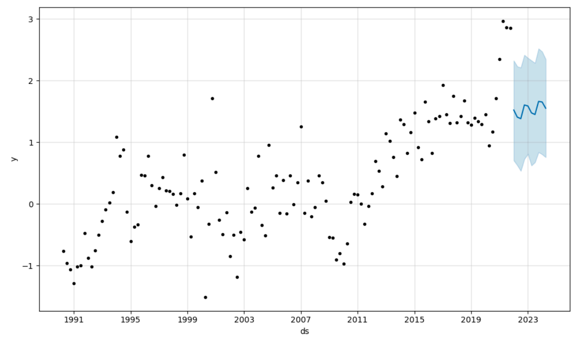
Market Asking Price Index (MARI) — Industrials — Rate of Change (ROC)

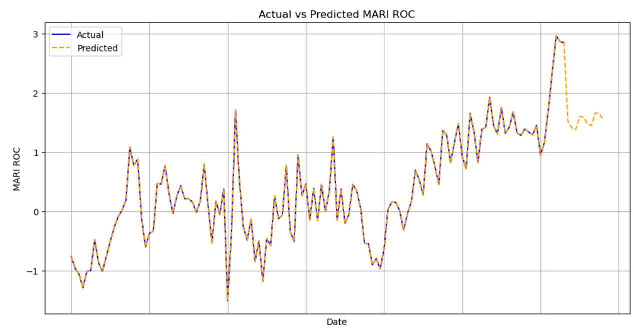
US





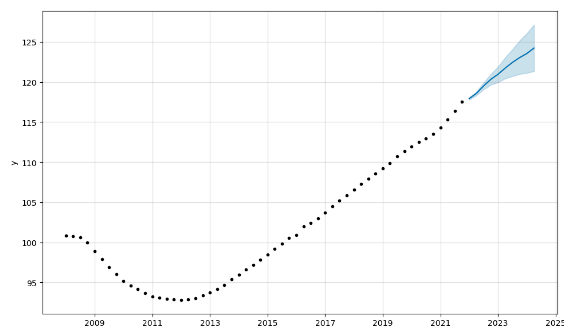
Columbus

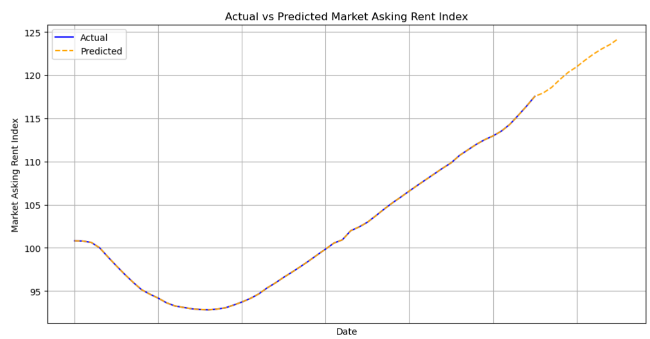




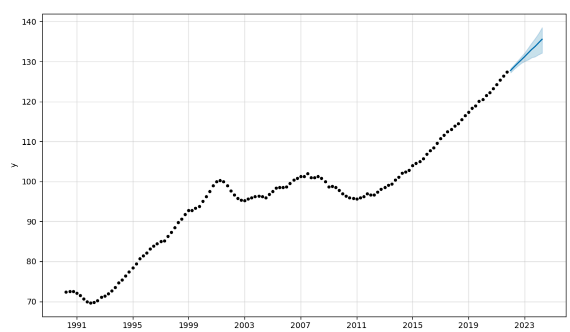
MARI — Retail

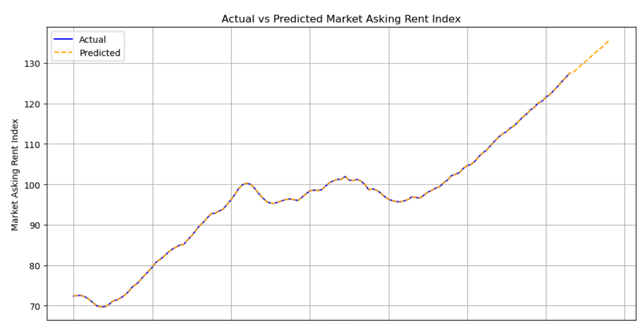
US





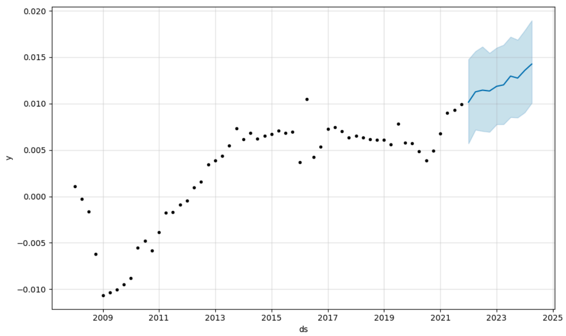
Columbus

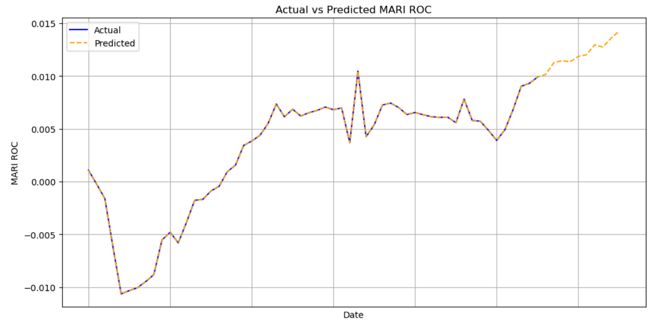




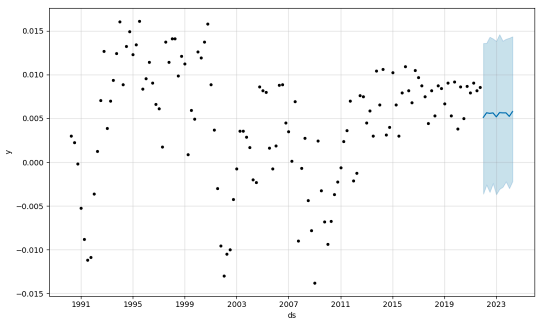
MARI — Retail — ROC

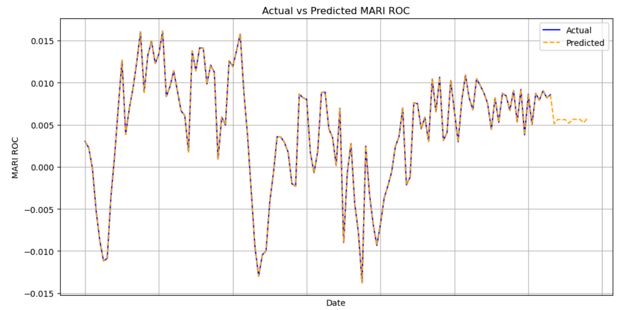
US





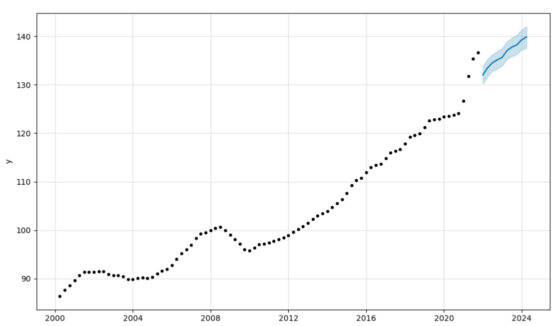
Columbus

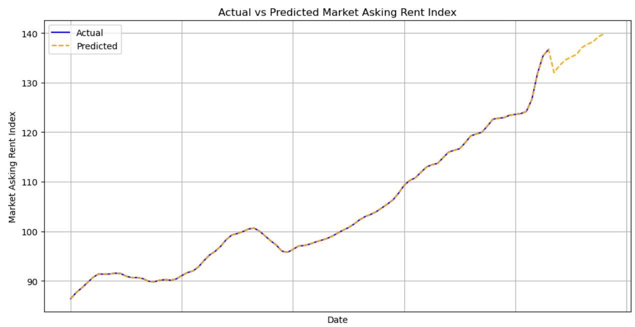




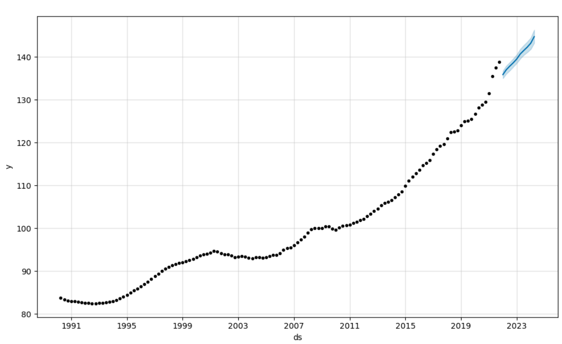
MARI — Multi-Family

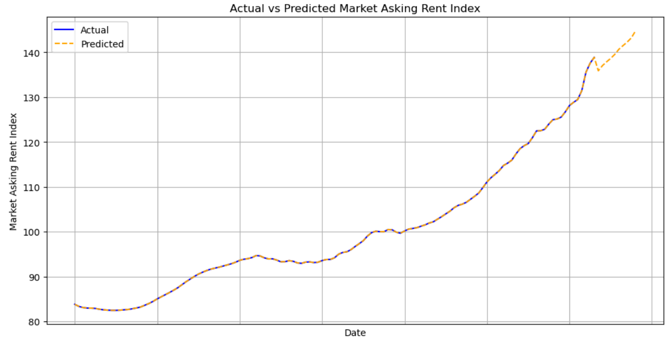
US





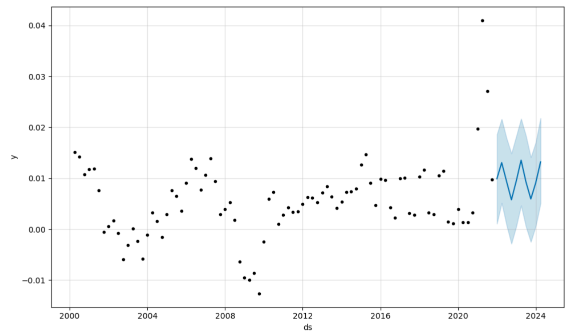
Columbus

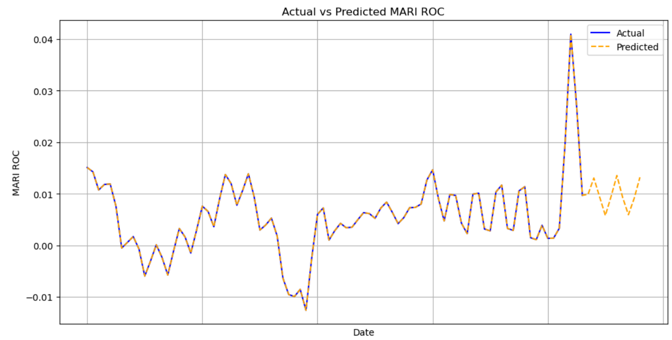




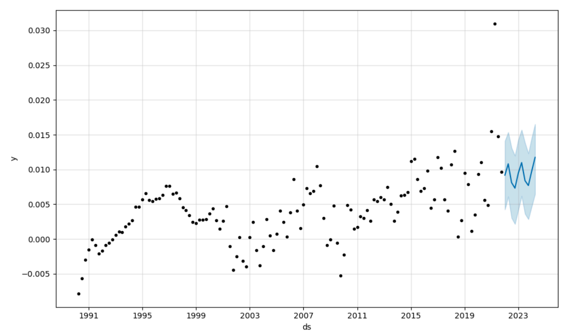
MARI — Retail — ROC

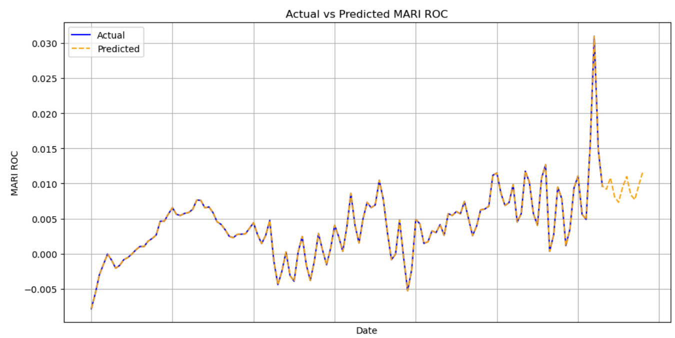
US





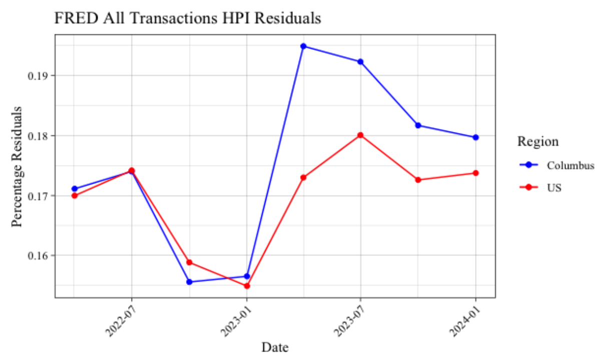
Columbus

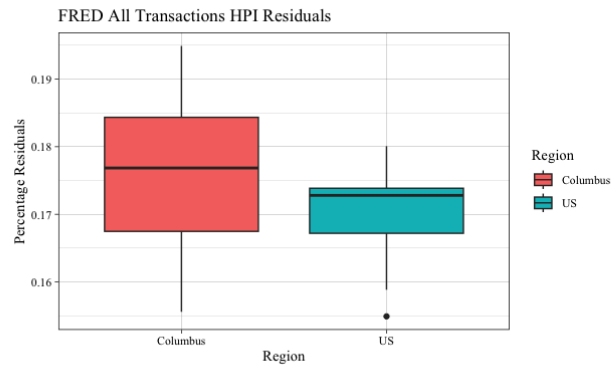




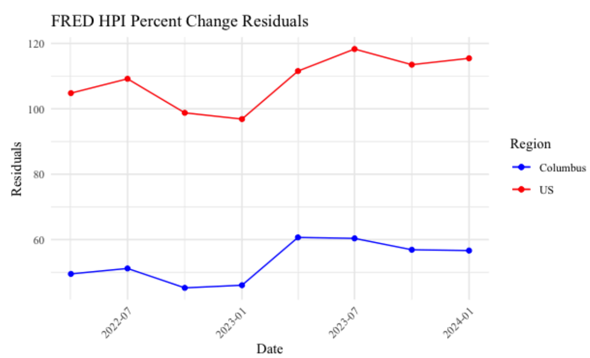
Residuals Visualizations

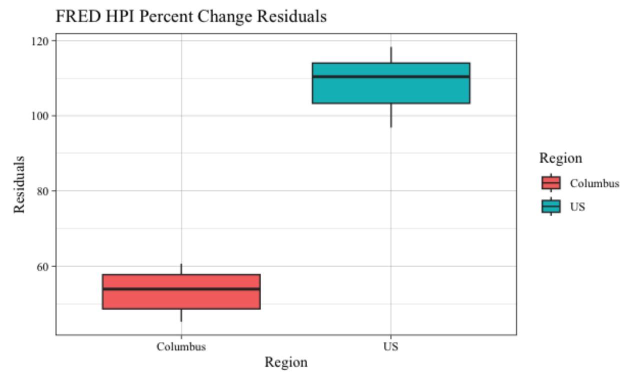
FRED All Transactions Housing Price Index



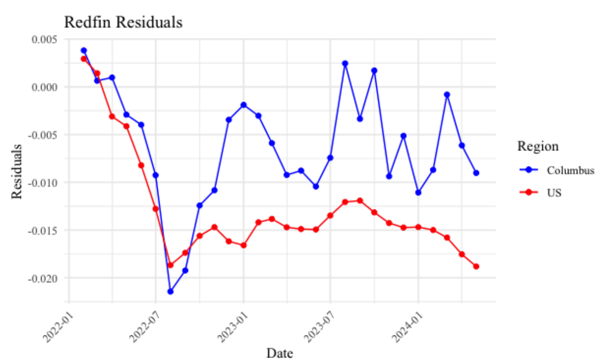


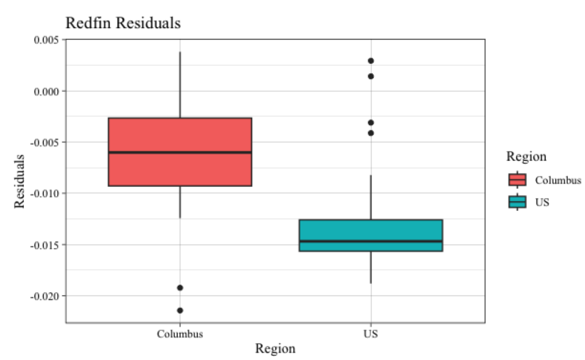
FRED All Transactions Housing Price Index - Percent Change



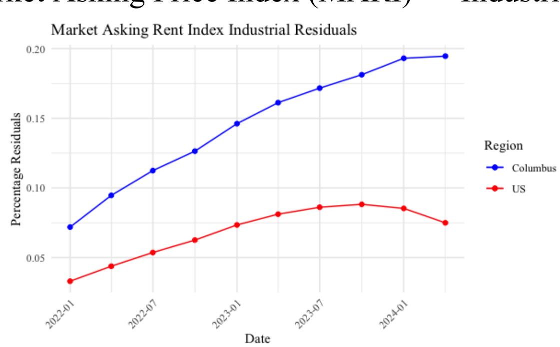


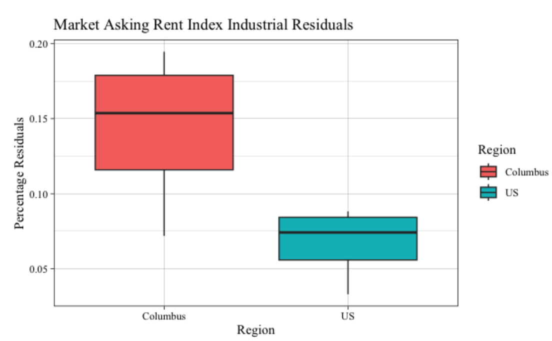
Redfin HPI MoM



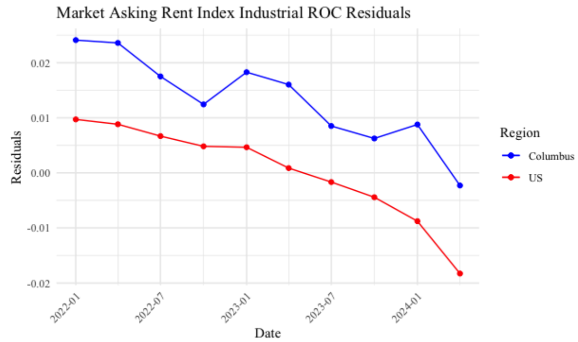


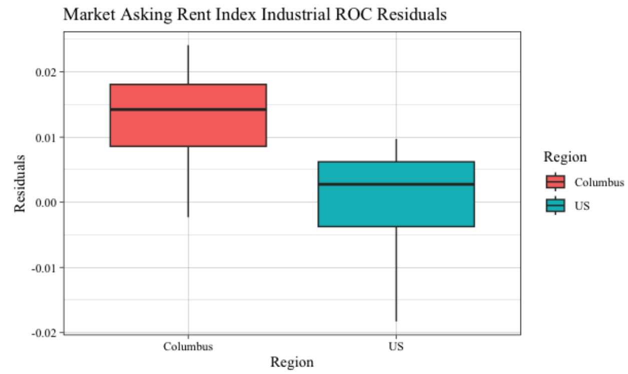
Market Asking Price Index (MARI) — Industrials



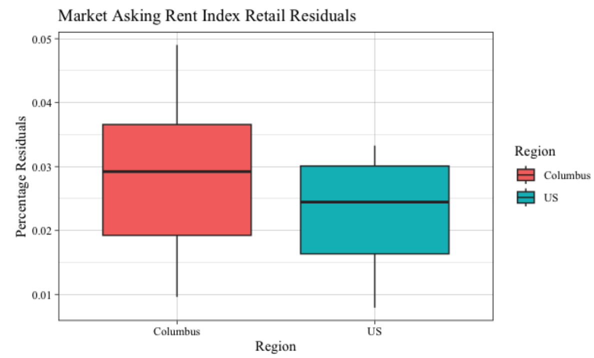


Market Asking Price Index (MARI) — Industrials — Rate of Change (ROC)

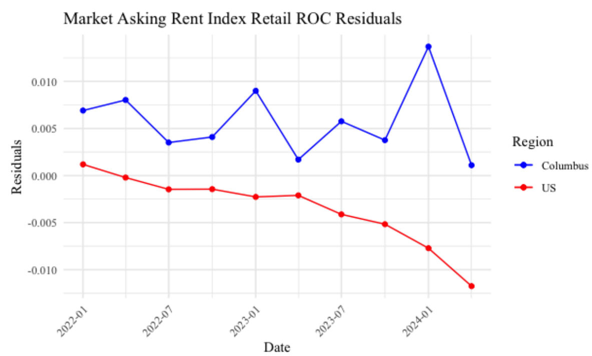


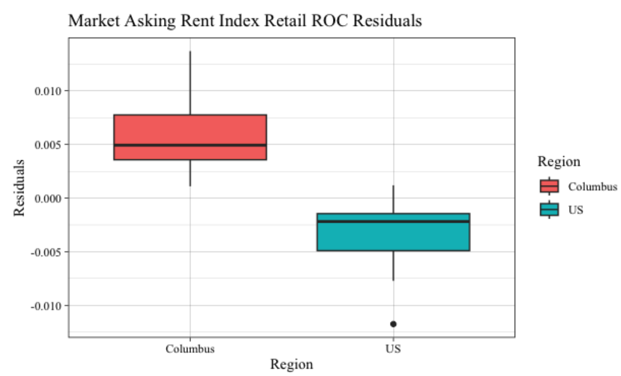


MARI — Retail

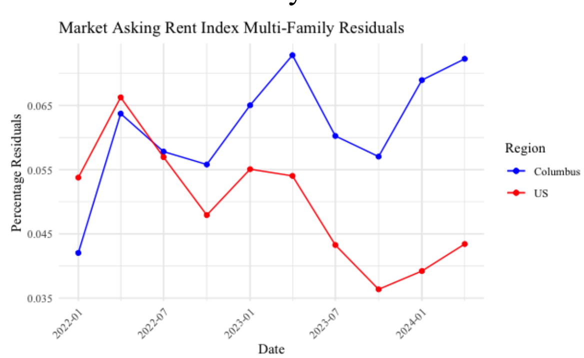


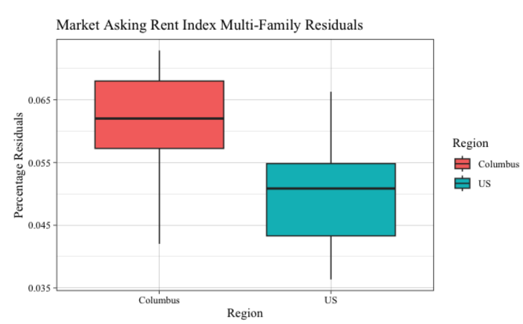
MARI — Retail — ROC



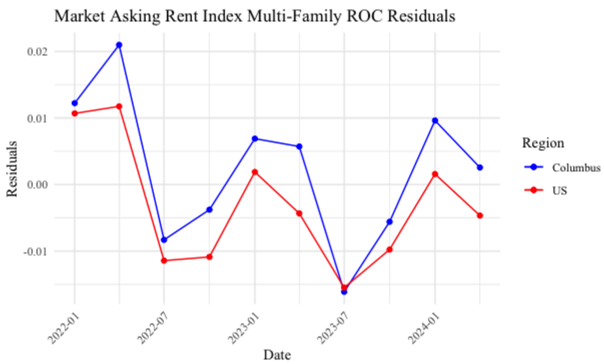


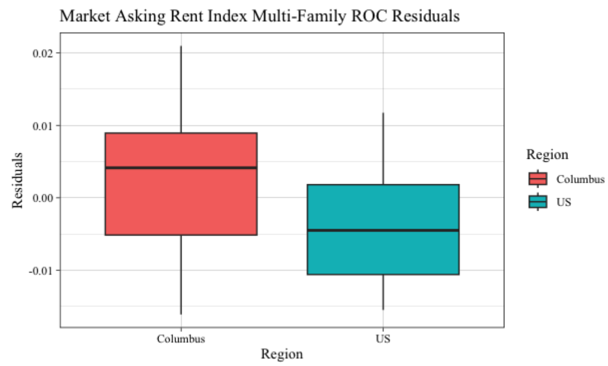
MARI — Multi-Family





MARI — Retail — ROC





Discussion

Summary of Key Findings & Interpretation of Results

The analysis revealed that, for 8 out of 9 datasets, the null hypothesis can be rejected, indicating that the residuals for Columbus are statistically significantly larger than those for the entire U.S. As shown in Figure 1, this statistical significance is evident across multiple indices. Figure 2 presents the Root Mean Squared Error (RMSE) values, demonstrating consistency between the models applied to both the U.S. and Columbus datasets. Root Mean Squared Error (RMSE) is a standard metric used to measure the accuracy of a predictive model. It calculates the square root of the average squared differences between the predicted values and the actual observed values, providing a measure of how well the model's predictions match the actual data.

In the "Prophet Visualizations" section, raw predictions made by the model are displayed, providing a clear forecast of trends. The "Residuals Visualizations" section shows the percent residuals for absolute index predictions, plotted over time in line graphs and collectively in a box plot for deeper insights. These visualizations clearly indicate that the residuals for Columbus are substantially larger than those for the broader U.S.

In these visualizations, it is clear that residuals for Columbus are much larger than residuals for the broader US. While it is complex to directly attribute these changes to Intel's impact, the investigation suggests a significant acceleration and growth in the Columbus real estate market. Recent years have seen Columbus increasingly recognized as a burgeoning tech hub. This study identifies specific data points that coincide with the dates of the announcement, marking the start of heightened national attention to the Columbus market. These findings provide a viable quantitative basis for understanding this growth and offer a foundation for future research to build upon and explore further

Theoretical Frameworks Review

The findings in this investigation align with several key theoretical frameworks discussed in the literature, providing insights into the operation of how large-scale corporate investments impact local real estate markets.

As noted by Petrova (2006), price models typically assume homogeneous motivations among buyers and sellers, yet distinct motivations, such as tax-deferred exchanges or out-of-state buyers, often affect commercial real estate transactions. Motivated buyers can create a temporary increase in demand, and in markets with inelastic supply, such as commercial real estate, this can lead to significant price changes. The basis for which is distinct from perfect markets where perfectly elastic supply leads to no price change. This study observed similar phenomena in Columbus, Ohio, where the announcement of Intel's $20 billion investment generated significant market activity. The subsequent demand increase would not be met by the existing supply, leading to notable price changes. This inelasticity in supply, particularly severe in retail and office properties, forced equilibrium prices to adjust, reflecting the heightened demand. The case of Columbus exemplifies how specific sale conditions, such as a major announcement, can catalyze price increases due to the inelastic nature of the real estate market, supporting the theoretical framework that distinct buyer motivations and market conditions significantly influence transaction prices.

The basis for this investigation, identifying how sentiment around this large investment can meaningfully influence metrics like the price index, is referenced in the literature as well. The concept of expectation shocks, discussed by Loewenstein and Willen (2022), is also relevant. Expectation shocks refer to changes in beliefs about future price growth that can lead to self-fulfilling price increases. Intel's announcement likely created positive expectation shocks in the Columbus real estate market. Investors and potential homeowners, anticipating future economic growth and increased property values, may have driven up prices even before the actual impact of the project materialized.

Limitations

This investigation has clear limitations, including the difficulty of isolating Intel's announcement's impact from other macroeconomic factors. While robust, the models used may have limitations in predicting future trends with absolute accuracy. Expertise limitations with machine learning models also hindered the accuracy, as shown by the inability to complete the LSTM Bayesian-optimized predictive model. Additionally, using univariate time series data restricts the models to trends and features present in the training data, limiting the ability to extrapolate new data. This restricts the application of models such as exponential smoothing, gradient boosting, or regression trees (e.g., random forest, XGBoost) that may not predict beyond trained values.

Another limitation involves the data sources, which were publicly available indices with limited features and measurement details. While applicable for this investigation, the raw univariate data may need to be revised for more detailed future research.

Future Research

Future research could explore similar economic developments in other regions for comparative analysis. Employing more granular data and different modeling techniques could enhance predictive accuracy and address current study limitations. Longitudinal studies would be beneficial in assessing the long-term effects of corporate investments on real estate markets.

Further research with LSTM models offers another opportunity for deeper insights, given their strength in handling sequential time series data and making supported extrapolations. While expertise limitations prevented the completion of this approach in this investigation, future research could overcome these challenges.

Additionally, applying machine learning techniques beyond regression analysis, such as classification, could provide new perspectives. This would require other forms of data, potentially necessitating funding or access to private institution measurements. Comparing these results with those of this investigation could provide policymakers, investors, and other stakeholders with insightful information for future economically-driven opportunities.

Bibliography

Burroughs, A. (2024). Increasing investment in central ohio. *Smart Business Columbus, 30*(3), 10-11. Retrieved from http://libproxy.usc.edu/login?url=https://www.proquest.com/trade-journals/increasing-investment-central-ohio/docview/2909512656/se-2

Chaney, T., Sraer, D., & Thesmar, D. (2012). The collateral channel: How real estate shocks affect corporate investment. *The American Economic Review, 102*(6), 2381-2409. doi:https://doi.org/10.1257/aer.102.6.2381

Chou, J., Fleshman, D., & Truong, D. (2022). Comparison of machine learning models to provide preliminary forecasts of real estate prices. *Journal of Housing and the Built Environment, 37*(4), 2079-2114. https://doi.org/10.1007/s10901-022-09937-1

Choy, L. H. T., & Ho, W. K. O. (2023). The use of machine learning in real estate research. *Land, 12*(4), 740. https://doi.org/10.3390/land12040740

Danielle Hale, H. J. (2024, April 25). *The spring 2024 Wall Street Journal/Realtor.com Housing Market Ranking*. Realtor.com Economic Research. https://www.realtor.com/research/april-2024-wsj-rdc-housing-market-ranking/

Du, L., Gao, R., Suganthan, P. N., & Wang, D. Z. W. (2022). Bayesian optimization based dynamic ensemble for time series forecasting. *Information Sciences, 591*, 155-175. https://doi.org/10.1016/j.ins.2022.01.010

Geltner, D. (2014). Real estate price indices and price dynamics: An overview from an investments perspective. SSRN. Available at https://ssrn.com/abstract=2540742 or http://dx.doi.org/10.2139/ssrn.2540742

Gorback, C. S., & Keys, B. J. (2020). Global capital and local assets: House prices, quantities, and elasticities (Working Paper No. 27370). National Bureau of Economic Research. https://doi.org/10.3386/w27370

Kramar, V., & Alchakov, V. (2023). Time-series forecasting of seasonal data using machine learning methods. *Algorithms, 16*(5), 248. https://doi.org/10.3390/a16050248

Loewenstein, L., & Willen, P. S. (2023). *House prices and rents in the 21st century* Retrieved from http://libproxy.usc.edu/login?url=https://www.proquest.com/working-papers/house-prices-rents-21st-century/docview/2835254049/se-2

McAllister, P., & Nanda, A. (2015). Does foreign investment affect U.S. office real estate prices? *The Journal of Portfolio Management, 41*(6), 38-47. https://doi.org/10.3905/jpm.2015.41.6.038

Nguyen, T. T., Nguyen, H. G., Lee, J. Y., Wang, Y. L., & Tsai, C. S. (2023). The consumer price index prediction using machine learning approaches: Evidence from the United States. *Heliyon, 9*(10), e20730. https://doi.org/10.1016/j.heliyon.2023.e20730

Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences, 164*, 105126. https://doi.org/10.1016/j.cageo.2022.105126

Oo, Z. Z., & Phyu, S. (2020). Time Series Prediction Based on Facebook Prophet: A Case Study, Temperature Forecasting in Myintkyina. International Journal of Applied Mathematics Electronics and Computers, 8(4), 263-267. https://doi.org/10.18100/ijamec.816894

Parmezan, A. R. S., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences, 484*, 302-337. https://doi.org/10.1016/j.ins.2019.01.076

Petrova, M. T. (2006). *Do buyer and seller motivations affect transaction prices in commercial real estate markets? evidence from tax deferred exchanges and other conditions of sale* (Order No. 3228805). Available from ProQuest Central; ProQuest Dissertations & Theses Global; ProQuest One Academic; ProQuest One Business. (305330170). Retrieved from http://libproxy.usc.edu/login?url=https://www.proquest.com/dissertations-theses/do-buyer-seller-motivations-affect-transaction/docview/305330170/se-2

Taylor, S. J., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints, 5*, e3190v2. https://doi.org/10.7287/peerj.preprints.3190v2

Williams, M. (2023, May 05). Google to build 2 more data centers in ohio: Projects in columbus, lancaster push tech giant's investment in state past $2 billion. *Lancaster Eagle - Gazette* Retrieved from http://libproxy.usc.edu/login?url=https://www.proquest.com/newspapers/google-build-2-more-data-centers-ohio/docview/2809361108/se-2

FHFA house price index. FEDERAL HOUSING FINANCE AGENCY. (n.d.). https://www.fhfa.gov/data/hpi

All-transactions house price index for the United States. FRED. (2024, May 28). https://fred.stlouisfed.org/series/USSTHPI

All-transactions house price index for Columbus, OH (MSA). FRED. (2024a, May 28). https://fred.stlouisfed.org/series/ATNHPIUS18140Q