Modeling the Spread of Rent Inflation across Counties in the state of New York through a Combined Spatial Analysis and Diffusion Modeling Approach

Mohith Reddy Kandakatla
Data Science
Regis University
Regis University , Denver, CO, USA
mkandakatla@regis.edu

Abstract

Spread of rent inflation across the US has reduced the affordability of housing, especially for young people in the country. Despite injection of investment being a factor for rent inflation, it is important for economic growth. Federal and state governments hence face a dilemma of continued economic growth at the expense of house affordability. Hence, there is a need to observe rent inflation from the perspective of how it diffuses from target areas of investment injection, as a management approach rather than discouraging injection of investment all together. This project employs spatial analysis, diffusion modeling and time series modeling in order to provide an understanding of how rent inflation spreads overtime and across counties in the state of New York. The ConvLSTM finds that spatial-Temporal patterns, GDP PI, Personal Income, Housing Units and House size account for 82.84% ($R2 = 0.828419$) of the variation of rent. On the other hand, first and second rent lags, GDP PI, Personal Income and Housing Units are shown to account for 88.58% ($R2 = 0.8858$) and 49.91% ($R2 = 0.4991$) of variation in rent for XGBoost and LSTM respectively.

I. Introduction/Background

Injection of both foreign and non-local investment, in forms such as tourism, real estate and data center investments, have changed the rental market in cities across the world. A mismatch between the local economy and the level of investment injection, for investments such as setting up of data centers, contributes to a jump in the rent prices following investment [1]. This high level of investment is accompanied by the employment of professionals, by the new businesses, who are paid higher wages than the averages in the local economy. The ripple effect is that local rent prices are increased to exploit the high wages paid to professionals working for the new businesses.

The case of real estate investment shows that the US economy may have not learnt lessons from the real estate bubble in the 2008 Great Recession. Individuals, often non-locals purchase homes as investments and at prices well beyond market prices [2]. This consequently push the housing prices up for both purchasing and renting as more people are priced out of home ownership. Additionally, tourism has increasingly moved from hotels to rental apartments since the setting up of short-stay rental services such as Airbnb [3]. The rental services operate within the same market as the conventional house renting for local residents. With short-stay rental services being more lucrative, apartment owners tend to rise the rent prices for conventional tenants as a trade-off for not converting the apartments for short-stay rental services.

Whereas big cities were initially key targets for injection of foreign and non-local investment, the investment landscape has currently changed. Smaller cities, as well as rural towns, are presently targeted for investment. Affordability of housing has, as a result, become a persistent policy issue for governments both at national and local levels.

II. Problem Statement

Rent prices are often consistent within neighborhoods, making proximity a factor in how house rent pricing is determined. Hence, injection of investment in an area not only leads to rent inflation in the given area, instead, by proximity, surrounding areas may as well experience rent inflation. This implies a diffusion of the rent inflation to areas surrounding the target area of investment. Managing the rent inflation hence requires understanding of how rent inflation diffuses and what is the nature of the diffusion in terms of magnitude.

Spatial analysis would provide context for the rent inflation across counties in the state of New York while diffusion modeling would build on the spatial analysis to model the spread of rent inflation across the state, both overtime and across counties. Data for the project was collected from:

- Census.gov for housing units.
- Bureau of Economic Analysis for historical investment data and personal income data per county.
- FRED for historical investment data for the US.
- HUDUSER for the historical rent inflation data per county.

Spatial analysis and Diffusion modeling require extensive feature engineering of the initial data in order to enable effective utilization. Hence, the data from the three sources underwent data cleaning, data transformation and data aggregation through merging across time and counties. The merged data then underwent feature engineering as data preparation for spatial analysis and diffusion modeling.

III. Related Work

Rent inflation represents a shock in the local economy, since it deviates from the expected trajectory of rent prices in the local market. Migration and infrastructure networks are important aspects of consideration in terms of how diffusion of shocks occurs across markets [4]. Given that markets are tied to geographical locations, spatial modeling represents an important concept in diffusion analysis. The spatial general equilibrium model appropriately analyzes shock diffusion across geographical regions through migration and infrastructure networks, with inter-regional connectivity significantly amplifying spatial transmission effects [4]. Although spatial-temporal effects model diffusion, uncertainty plays an important role in diffusion analysis. Bayesian deep learning frameworks can model trajectory diffusion using probabilistic methods with incorporation of spatial-temporal structure and uncertainty propagation [5]. Diffusion modeling considers how behavior spills over between neighbors. Localized spillover effects are an important aspect on how housing interventions affects neighboring household's behavior [6]. Proximity drives short-term increases in housing aspirations, but these effects decay over time, indicating temporary diffusion of behavioral responses [6]. Spatial and temporal analysis, and factors that capture uncertainty should hence be considered along with diffusion modeling in investigating the diffusion of behavior, as in this case rent prices increases that lead to rent inflation.

IV. Methodology

Three models were employed to investigate the spread of rent inflation overtime and across the counties in the state of New York; these models are one diffusion model and two temporal models. The Convolutional Long Short-Term Memory (ConvLSTM) was used for the diffusion modeling. Inputs, spatial-temporal grid, hidden states, output and the loss optimization form the model architecture [7]- [8]. Figure 1 below gives the summary for the ConvLSTM.

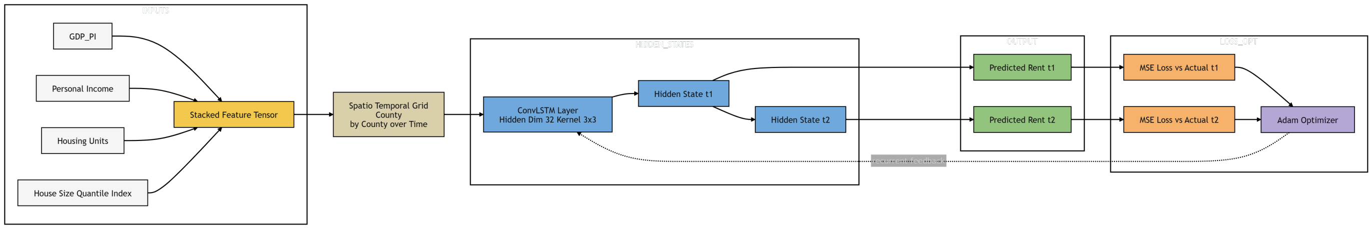

Fig. 1: Model Architecture: ConvLSTM

The input layer consists of the 4 input variables (GDP PI, Personal Income, Housing Units and House size) staked together into a feature tensor for the model. The spatial-temporal grid (8x8) connects the inputs layer with the hidden layers and it arranges the data by county over time. Two hidden states exist in the hidden layer, corresponding to 2 timestamps that represent 2 lags in the data. The Output layer outputs the predicted percentage rent for each of the 2 timestamps from the hidden layer. Loss optimization is then applied in the loss optimization layer using the Adam Optimizer with the Huber (combined MAE and MSE) Loss computed for predicted vs actual across the predictions for 2 timestamps from the output layer.

The XGBoost model and the Long Short-Term Memory (LSTM) models were employed for the temporal diffusion analysis. Four stages involved in effective utilization of XGBoost modeling are model initialization and setup, sequential learning (the boosting stage), feature utilization and, regularization and generalization [9]. The chart on the left in Figure 2 below shows the summary of the stages of effective utilization as employed in XGBoost. For the LSTM, the stages involved in the effective modeling are data preparation and sequencing, model architecture, temporal learning and, training and optimization [10]. The chart on the right in Figure 2 below shows the summary of the stages of effective utilization as employed in LSTM.

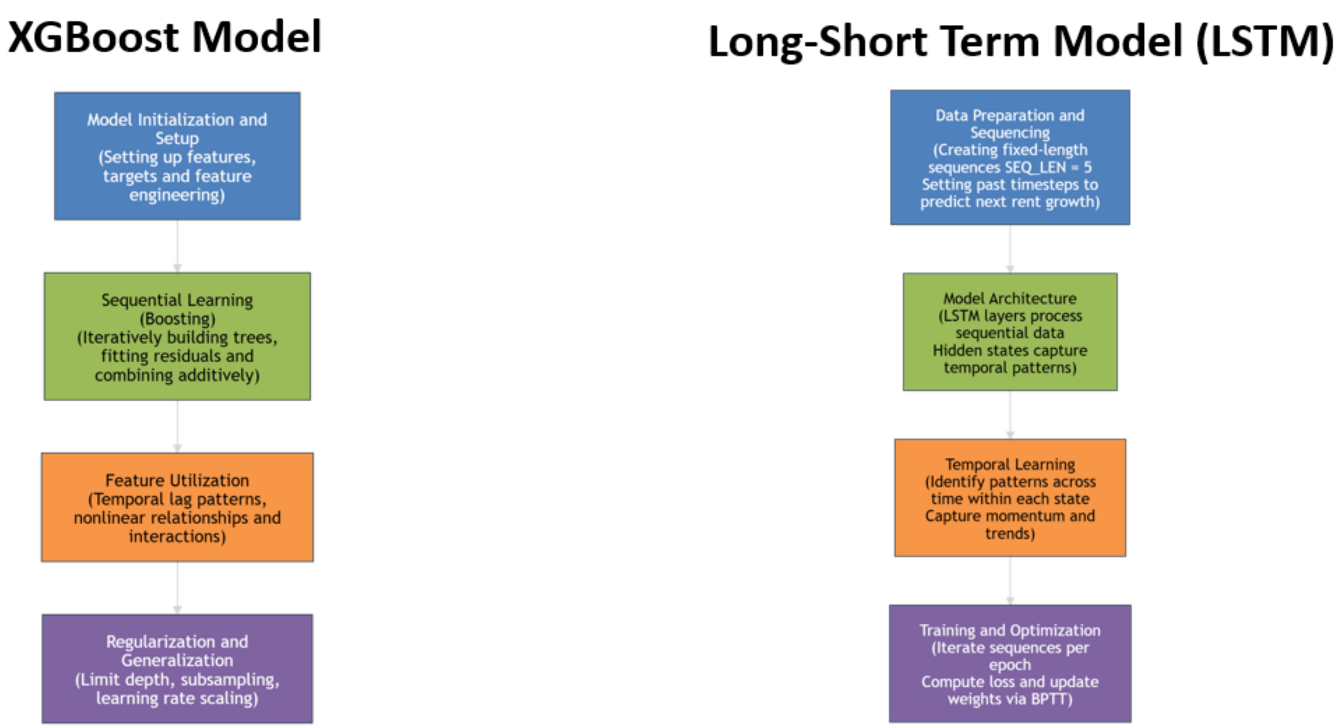

Fig. 2: Model Architecture: XGBoost and LSTM

Table 1 below gives the features, targets and data pre-processing for the three models above. GDP PI provides information on the injection of investment across the counties in the state; hence variation in this variable represented variation in the amount of investment injected into the local economies in the counties. Housing Units served as an indicator of the demand-pull driver of rent inflation. Personal Income served as an indictor of demand-pull in terms of the ability of the customers in the housing market to afford rent in a given county. House Size served as an indicator of the cost push factors in terms of the associated costs for different house sizes. Rent gives variations in the house rents providing an indication of rent inflation and hence formed the target variable.

	ConvLSTM	XGBoost	LSTM
Target	Rent	Rent	Rent
Features	• Spatial – Temporal Patterns (built into model) • GDP_PI • Personal Income • Housing Units • House Size	• GDP_PI • Personal Income • Housing Units • House Size • Rent Lag 1 • Rent Lag 2	• GDP_PI • Personal Income • Housing Units • House Size • Rent Lag 1 • Rent Lag 2
Data Preprocessing	• Building Spatial Weights (County Neighborhood Structure) using Queen's Contiguity Weights. • Spatial Grid creation using PCA embeddings • Time-based Data Partitioning at year = 2015 • Normalizing the data • Sequence Creation (length = 2)	• Shifting the rent variable by 1 to create lag 1 • Shifting the rent variable by 2 to create lag 2 • Time-based Data Partitioning at year = 2015	• Shifting the rent variable by 1 to create lag 1 • Shifting the rent variable by 2 to create lag 2 • Time-based Data Partitioning at year = 2015 • Normalizing the data • Sequence Creation (length = 2)

TABLE I: Targets, Features and Data Pre-processing

V. Data Analysis

A. Data Collection and Preparation

Data for the project was collected from:

- Census.gov for housing units
- Bureau of Economic Analysis for historical investment and personal income per county.
- FRED for historical investment data for the US.
- HUDUSER for the historical rent inflation data per county.

The chart below in Figure 3 gives the summary of the data preparation process.

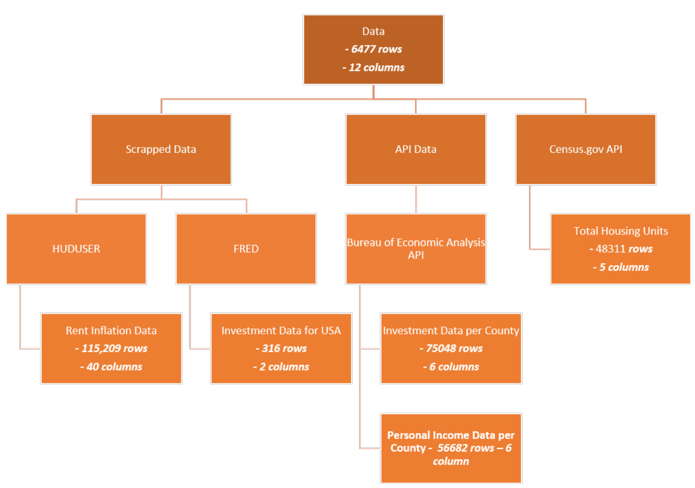

Fig. 3: Data Pipeline

Merging of the four datasets was completed by county name and year which were available across all four variables. In the feature engineering, the cleaned data was converted from wide format to long format; where the median rent for different house sizes were placed in the rent column with the house size variable created. From the rent column, the rent growth feature was then created to for annual growth by county and house size. The data resulting from the data preparation process, including the feature engineering, is shown in the summary of the description of variables in Table 2 below. The variables in bold represent the variables in the final dataset. Lagging formed a notable data transformation where data on rent was shifted forward by one year to produce lag 1 and shifted forward by 2 years to produce lag 2.

Variable	Description
State_Name	**State name for only New York**
county_name	**County name for counties within New York State**
Year	**Year**
Rent_50_0	Annual average median rent per county for living spaces not classified as 1-bedroom house and above.
Rent_50_1	Annual average median rent per county for living spaces classified as 1-bedroom house.
Rent_50_2	Annual average median rent per county for living spaces classified as 2-bedroom house.
Rent_50_3	Annual average median rent per county for living spaces classified as 3-bedroom house.
Rent_50_4	Annual average median rent per county for living spaces classified as 4-bedroom house.
GDP_PI	**Annual GDP Private Investment per county**
Housing_Units	**Total number of available housing units per county for the given year.**
Personal_Income	**Average personal income per county for the given year.**
House Size	**Feature engineered through the conversion of the data from wide format (with rent for different house types as separate column) to long format with a column for house size and a new column for rent.**
Rent	**Feature engineered through the conversion of the data from wide format (with rent for different house types as separate column) to long format with a column for house size and a new column for rent.**
Rent Growth	**Feature engineered as the annual percentage change in the rent price computed by year, county and house size used in spatial analysis.**

TABLE II: Variables' Description

B. Exploratory Data Analysis

The analysis of the trend in the median rent growth over the years across the five different house sizes is given below in Figure 4. Rent growth across all house sizes has had an increasing trend from 2001 through to 2026. Also, the median rent for houses with 3 and 4 rooms have had a steeper increase as compared to smaller house sizes.

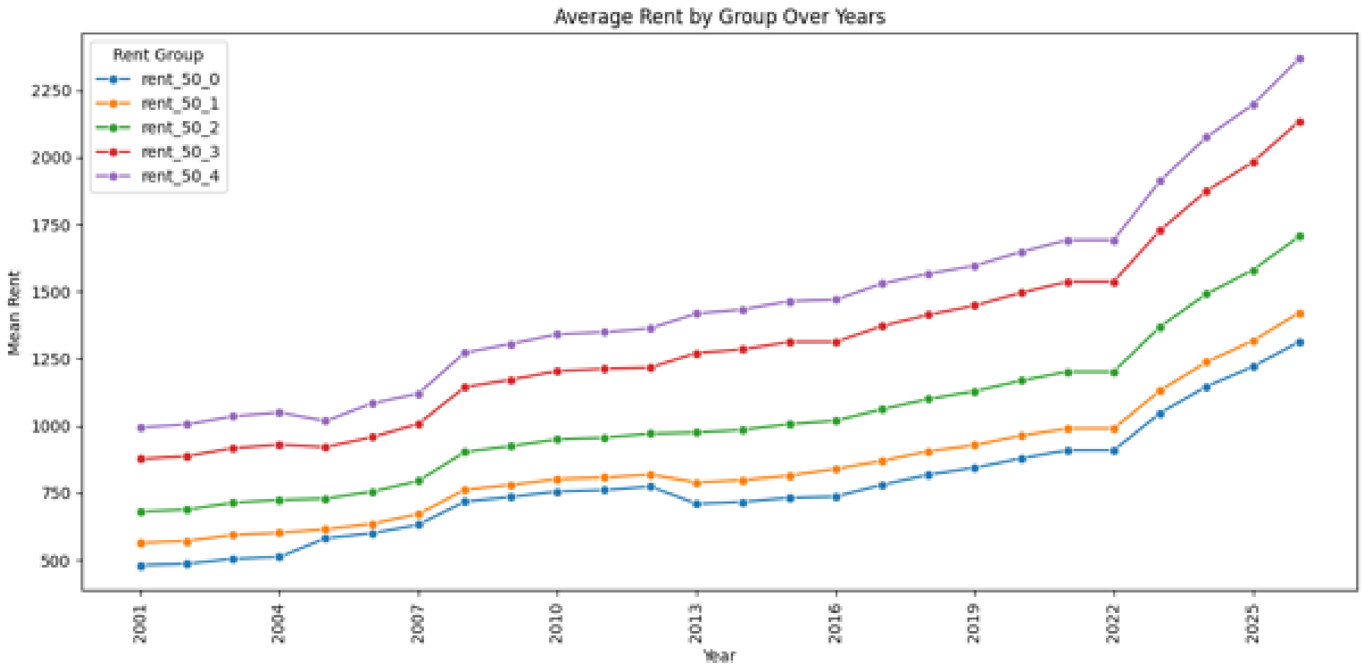

Fig. 4: Trends in Rent Growth

The investigation of the distribution of data for the features, GDP PI, Housing Units and Median Household Income, is presented below through the three histogram plots in Figures 5, 6 and 7 respectively. From the histograms, we note a positive skewness in the distribution of data across all three features. Hence, majority of the data on GDP PI, Housing Units and Median Household Income fell below the mean.

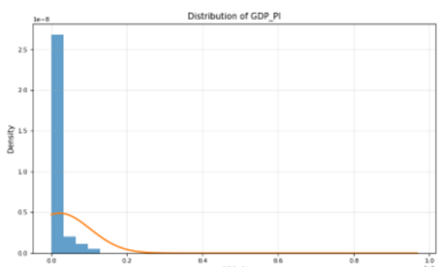

Fig. 5: GDP PI Distribution

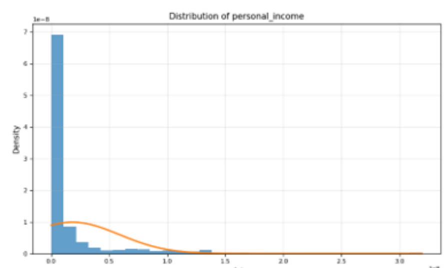

Fig. 6: Housing Units Distribution

Fig. 7: Personal Income Distribution

The results of the spatial analysis for spatial diffusion of rent over study period, that is proportion of rent change between 2026 and 2001, is presented below in Figure 8; High proportions of rent change are observed across different counties. Clustering of rent change proportions is visible with group of counties falling within the same range of rent change; this is indicative of existence of diffusion in rent change. In Figure 9, the hotspot analysis corresponding to the spatial distribution is presented. The hotspot analysis indicates that hotspot for rent change exists in the the counties to the south within and around New York City form the cluster for hotspot while the counties in the middle and towards the west of the state has clusters of cold spots.

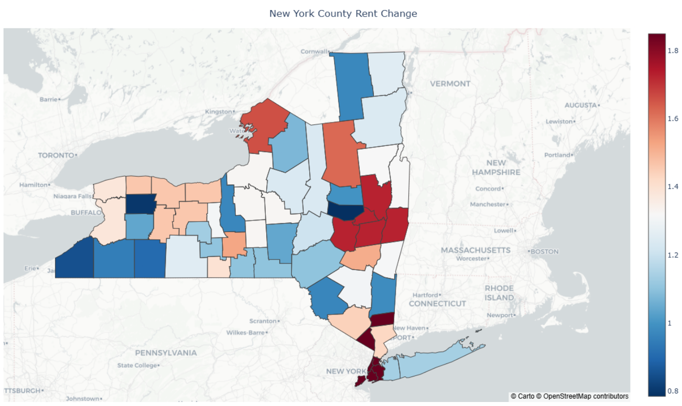

Fig. 8: Rent Inflation Spatial Analysis

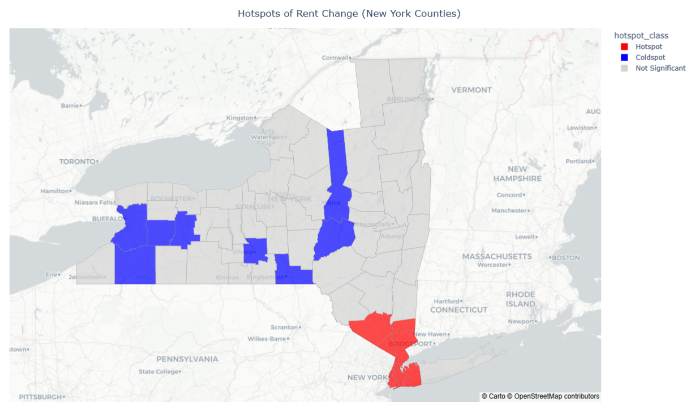

Fig. 9: Rent Inflation Hot Spot Analysis

C. Diffusion Modeling Results: ConvLSTM

Figure 10 below presents the rent growth diffusion visualization across three recent timestamps in the top row from the actual data and the predictions made by the model for the corresponding timestamps in the bottom row. The diffusion grids places counties adjacent to each other based on co-increases in rent and hence have association and co-influence in rent price growths. The grids at the top show the existence of diffusion across counties within each timestamp and how this diffusion evolves across timestamps; rent increases diffused from hotspots consistently in recent time. Comparing actual and predicted diffusions per timestamp, the model is able to correctly captures the hot spots and cold spots for rent inflation.

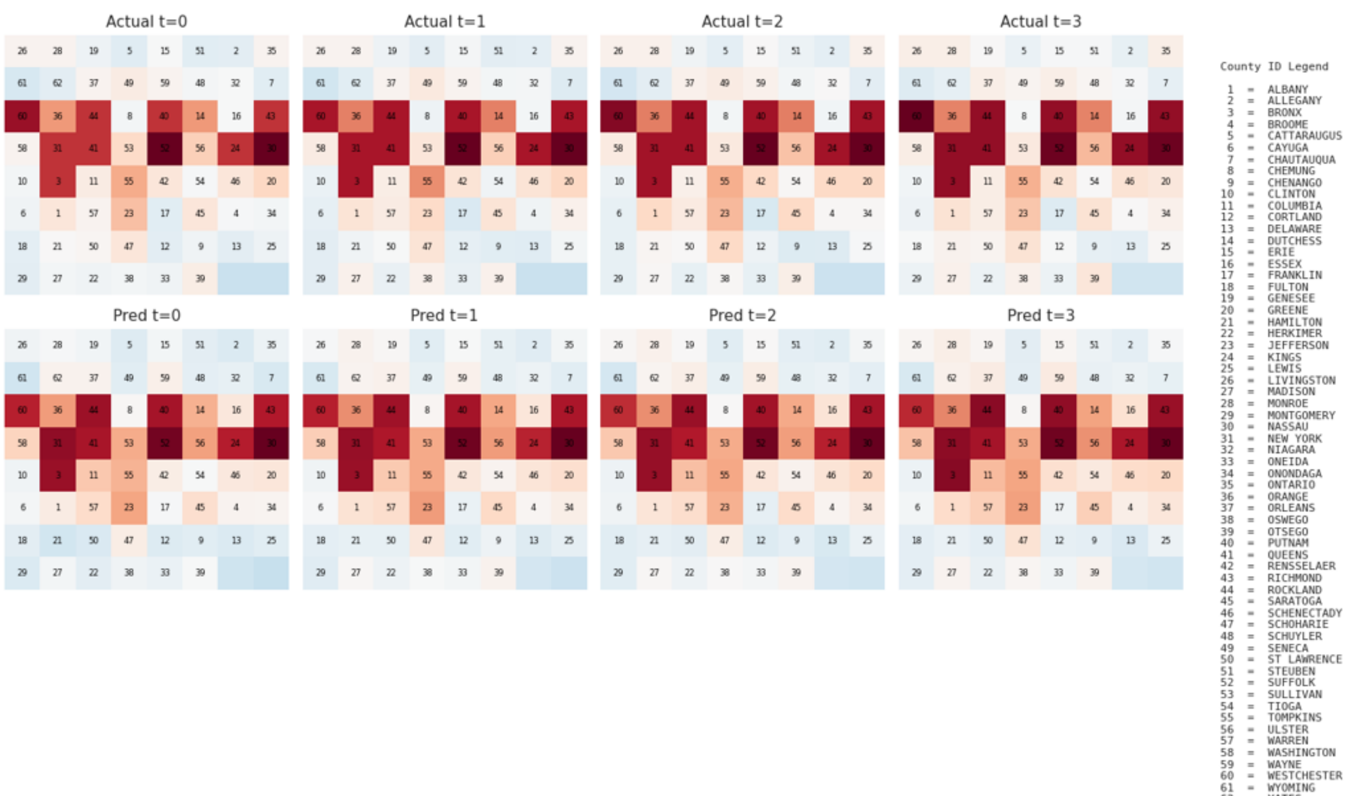

Fig. 10: Diffusion Grids: Actual vs Predictions

D. Diffusion Modeling Results: Forecasting Diffusion Over Time

The comparison between the forecasting performance of the diffusion and time series models is shown below in Table 3. The results show that Spatial-Temporal patterns, GDP PI, Personal Income, Housing

Units and House size account for 82.84% (R2 = 0.828419) of the variation of rent growth into the future. Also, from the results, first and second rent growth lags, GDP PI, Personal Income and Housing Units are shown to account for 88.58% (R2 = 0.885800) and 49.91% (R2 = 0.4991) of variation in rent for the XGBoost and LSTM respectively.

	Diffusion Model	Time Series Models	
	ConvLSTM	XGBoost	LSTM
RMSE	314.672238	231.240965	484.892804
MAE	211.155426	117.932873	278.995544
R2	0.828419	0.885800	0.4991

TABLE III: Model Performance Comparison

Feature importance analysis is presented in the bar charts in Figures 11, 12 and 13 below for the ConvLSTM, XGBoost and LSTM respectively. For the ConvLSTM, the plot shows that the Personal Income followed by the GDP PI had the most influence on Rent Growth. Lag 2 and Lag 1 lead as the most influential factors for Rent Growth in both the XGBoost and LSTM models. House size is shown to have no influence across time series modeling.

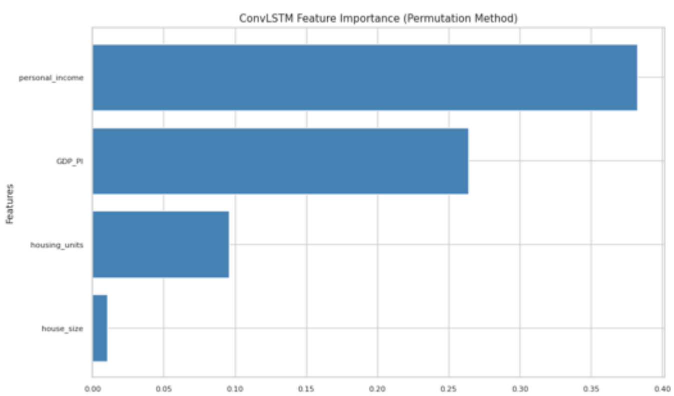

Fig. 11: Feature Importance (ConvLSTM)

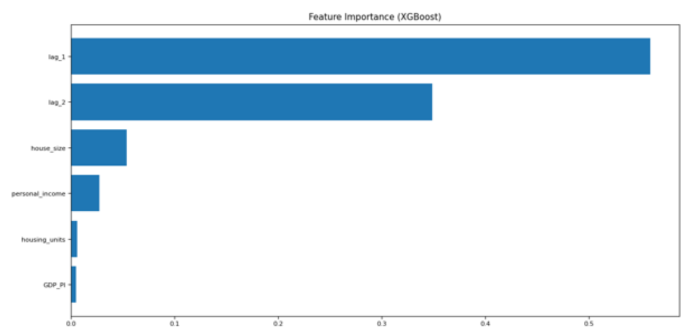

Fig. 12: Feature Importance (XGBoost)

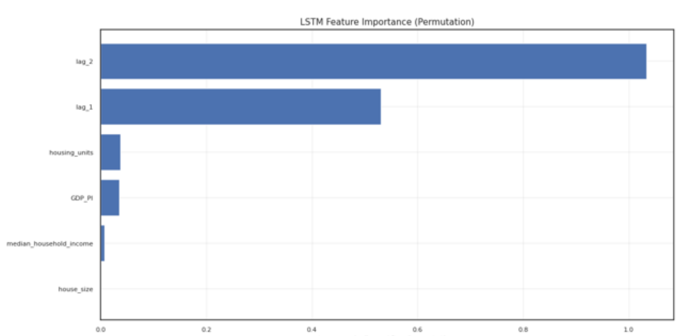

Fig. 13: Feature Importance (LSTM)

VI. Timeline

Week 1 - Project Proposal

Week 2 - Data Collection, Data Cleaning, Data Transformation, Data Aggregation and Feature Engineering

Week 3 – Spatial Analysis of Rent Inflation

Week 4 – Diffusion Modeling for Spread of Rent Inflation

Week 5 – Time Series Modeling of Rent Inflation

Week 6 – Combined Diffusion and Time Series Modeling

Week 7 – Report and Slide Presentation

Week 8 - Project Review

VII. Conclusion

The diffusion grid from the ConvLSTM provides the mapping of the diffusion of rent inflation across the counties in the state of New York and can be used for monitoring how rent inflation spread. The grid also shows that rent growth diffuses both across counties and overtime where the county, GDP for Private Investment, Personal Income, Number of Housing Units, House size and rent in previous years explain the rent growth. In temporal terms, diffusion of rent overtime occurs mainly over 2-year periods (Lag 2) and to some extent 1-year periods (Lag 1); this is since Lag 2 and Lag 1 of rent growth are shown to be most influential for rent growth in the LSTM and XGBoost models. Therefore, observing rent growth over the past 2 years provides a good indication of the direction of present rent growth; through which property owners and developers can anticipate the rent growths. The study also identified diffusion clusters among counties, which implies the need for property owners and developers to understand the rent growth for housing in associated counties from the diffusion grid and the accompanying costs; and utilize this information in the pricing strategy development.

References

[1] R. Derven, "Data center real estate: Challenges and opportunities in the digital age," NAIOP Development Magazine, Winter 2023–2024, 2024, [Online; accessed 2026]. [Online]. Available: https://www.naiop.org/research-and-publications/magazine/2023/winter-2023-2024/development-ownership/data-center-real-estate-challenges-and-opportunities-in-the-digital-age/

[2] F. Sá, "The effect of foreign investors on local housing markets: Evidence from the uk," *Journal of Economic Geography*, vol. 25, no. 3, pp. 329–349, May 2025. [Online]. Available: https://academic.oup.com/joeg/article/25/3/329/7901283

[3] H. R. A. Koster, J. van Ommeren, and N. Volkhausen, "Short-term rentals and the housing market: Quasi-experimental evidence from airbnb in los angeles," *Journal of Urban Economics*, vol. 124, p. 103356, Jul. 2021.

[4] X. Zhang and Y. Yu, "Migration, transportation infrastructure, and the spatial transmission of covid-19 in china," *Journal of Urban Economics*, vol. 127, p. 103351, 2022.

[5] Y. Pang, X. Zhao, H. Yan, and Y. Liu, "Data-driven trajectory prediction with weather uncertainties: A bayesian deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 130, p. 103326, 2021.

[6] S. Galiani, P. J. Gertler, and R. Undurraga, "Aspiration adaptation in resource-constrained environments," *Journal of Urban Economics*, vol. 123, p. 103326, 2021.

[7] F. Yuan, C. Leng, H. Yan, Z. Pei, and J. Peng, "Mema-convlstm: Spatiotemporal prediction via multi-scale autocorrelation memory and hierarchical fusion," *Information Sciences*, 2026.

[8] J. Xiao, Z. Wang, Y. Liao *et al.*, "A convlstm-based model for urban flood prediction under dynamic rainfall patterns and exploration on its extrapolation capability," *International Journal of Disaster Risk Science*, vol. 16, pp. 1057–1073, 2025.

[9] M. A. Mosavi, S. Ardabili, and A. R. Várkonyi-Kóczy, "Applications of xgboost in water resources engineering: A systematic literature review (dec 2018–may 2023)," *Environmental Modelling & Software*, vol. 174, p. 105971, 2024.

[10] A. B. Amendolara, D. Sant, H. G. Rotstein, and E. Fortune, "Lstm-based recurrent neural network provides effective short-term flu forecasting," *BMC Public Health*, vol. 23, p. 1788, 2023.