

# Rent Inflation

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

**Name: Mohith Reddy Kandakatla**

**Regis University**



# Introduction and Background

## Investment Injection

- New businesses, with high injection of investment, cause the local economy to experience sudden change in the form of a surge in cashflows [1]. Such new businesses include:
  - Data centers
  - Home buying as investment
  - Short-stay rentals
- In order to exploit this surge in cashflows, owners of rental houses increase the rents, creating rent inflation [2].

## Rent Inflation

- Rent inflation is a term that refers to the rate at which rent prices for residential houses increase.
- Key drivers of rent inflation are [3]:
  - demand pull (that is high demand and low supply) and
  - cost push factors (that is higher property taxes, insurance and maintenance).

# Problems Arising

## Property Pricing

- Rent inflation causes rises in the rent prices at market level and is associated with increase in maintenance costs for property owners.
- Hence, property owners need to understand rent inflation as an anticipatory strategy for adjusting to both market prices and market costs.

## Property Development

- Rent inflation serves to modify the housing needs of tenants and they may shift location or move down the rental type and size.
- Property development for rental units is dependent on the rental needs in the market. Understanding the spread in rent inflation would allow for better judgement on when, where and what type of rental property to develop.

# Machine Learning Approach to Problems

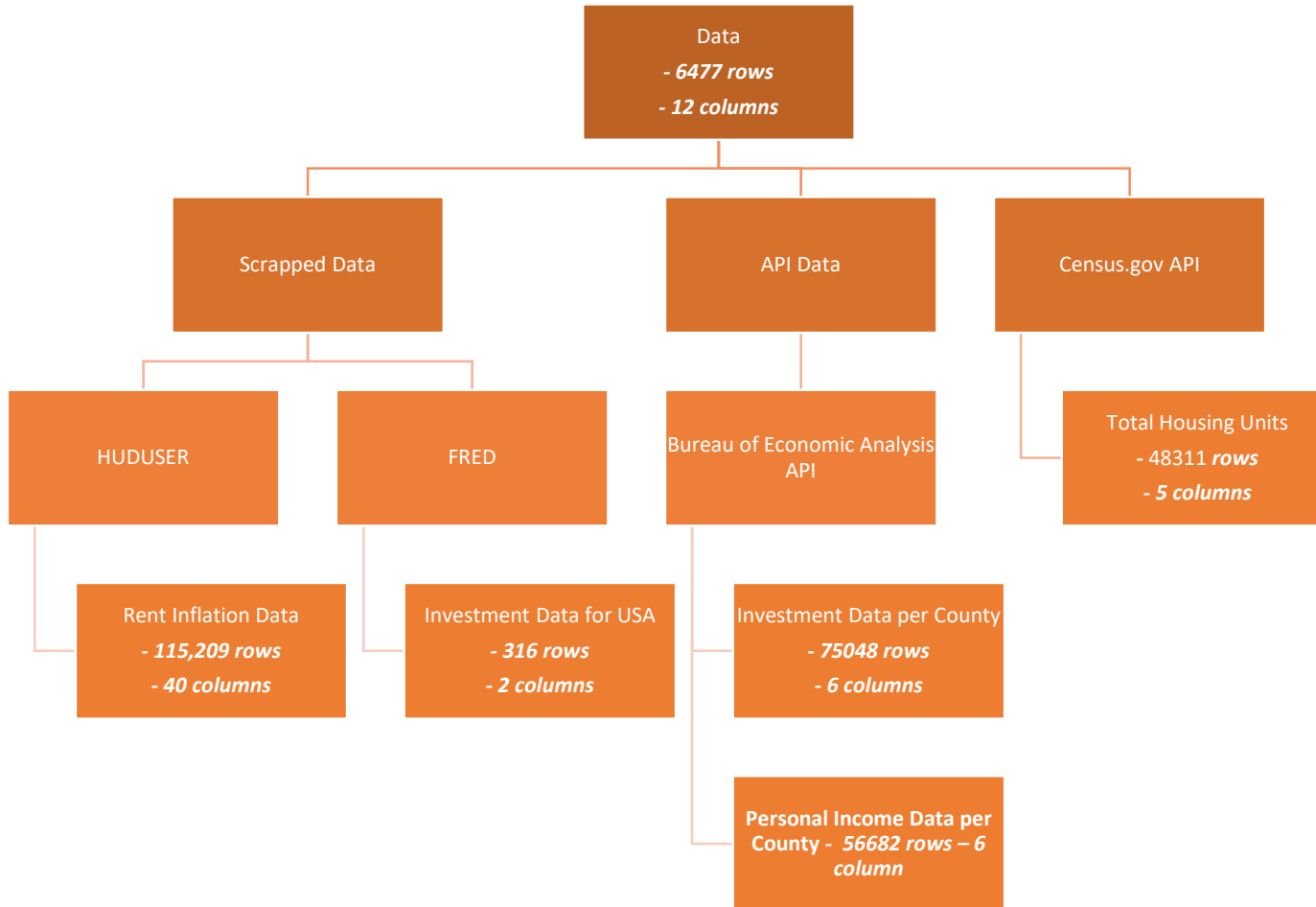
## Understanding Spread of Rent Inflation

- Diffusion modelling is an analytical approach for mapping behaviour spread through a population with the aim of understanding the nature (including the factors influencing the spread) and magnitude of the spread.
- In this context, increasing of rent by owners of rent units forms a behaviour with the spread in the tendency of increasing rent requiring investigation to understand the nature and magnitude of the spread.

## Forecasting Spread of Rent Inflation

- By factoring in the importance of proximity and time in the spread of rent increases, spatial analysis and diffusion modelling inform on how best property developers and managers adjust to the spread of rent inflation.
- Time series modelling can then borrow from the insights from both spatial analysis and diffusion modelling to forecast rent inflation and in turn improve decision making for property pricing and property development.

# Data: Sources, Features and Feature E

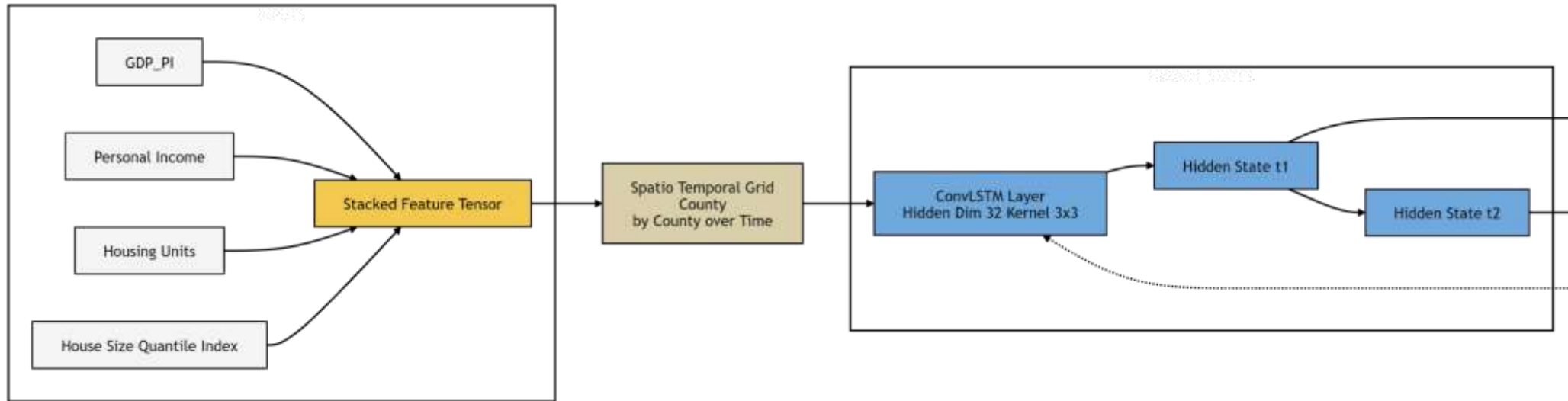


Variable	Description
State_Name	State name for only New York
county_name	County name for counties within New York State
Year	Year
Rent_50_0	Annual average median rent per county for living spaces not classified as 1-bedroom house and above.
Rent_50_1	Annual average median rent per county for living spaces 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.

# Data: Sources, Features and Feature Engineering

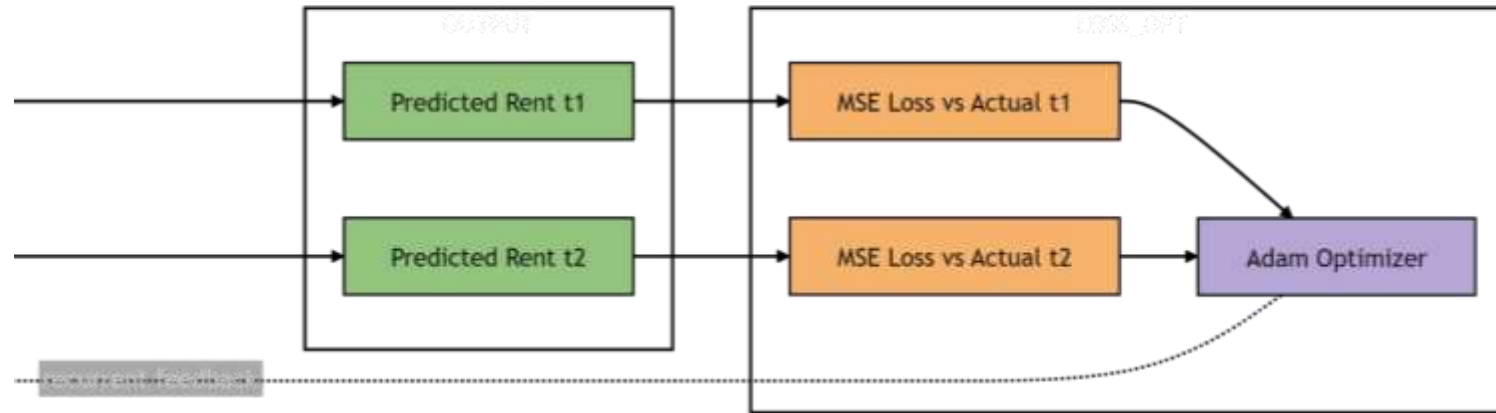
- *GDP\_PI*: Provides information on the injection of investment across the counties in the state; hence variation in this variable represented variation in the amount of investment injected into the local economies in the counties.
- *Housing Units*: Served as an indicator of the demand-pull driver of rent inflation.
- *Personal Income*: Served as an indicator 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.

# Methodology: Model Architecture: Diffusion Modelling - ConvLSTM(Inputs and Hidden States)



- The ConvLSTM model architecture consists of inputs, spatial-temporal grid, 2 hidden states, output and the loss optimization.
- The input layer shown above consists of the 4 input variables (GDP\_PI, Personal Income, Housing Units and House size) stacked 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.
- 2 hidden states exist in the hidden layer corresponding to 2 timestamps that correspond to 2 lags in the data.

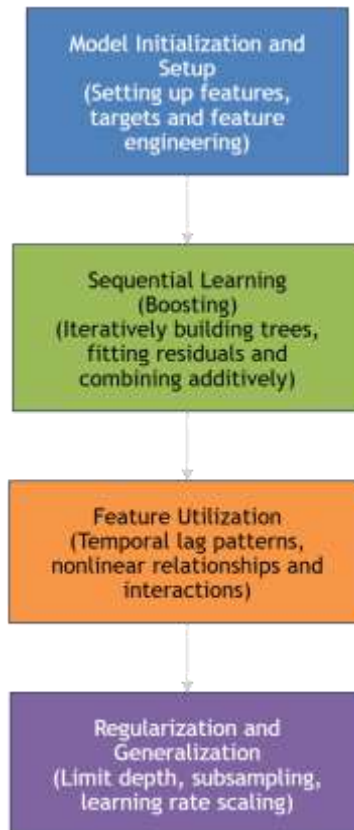
# Methodology Model Architecture: Diffusion Modeling - ConvLSTM (Outputs and Loss Optimization)



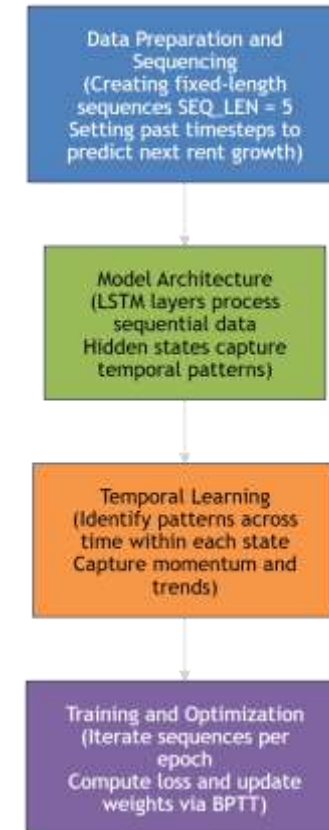
- The Output layer outputs the predicted percentage rent for each of the 2 timestamps from the hidden layer.
- Loss optimization is then applied in the loss optimization layer using the Adam Optimizer with the Huber(combined MAE and MSE) Loss computed for predicted vs actual across the predictions for 2 timestamps from the output layer.

# Methodology Model Architecture: Time Series Modeling - XGBoost and LSTM

## XGBoost Model



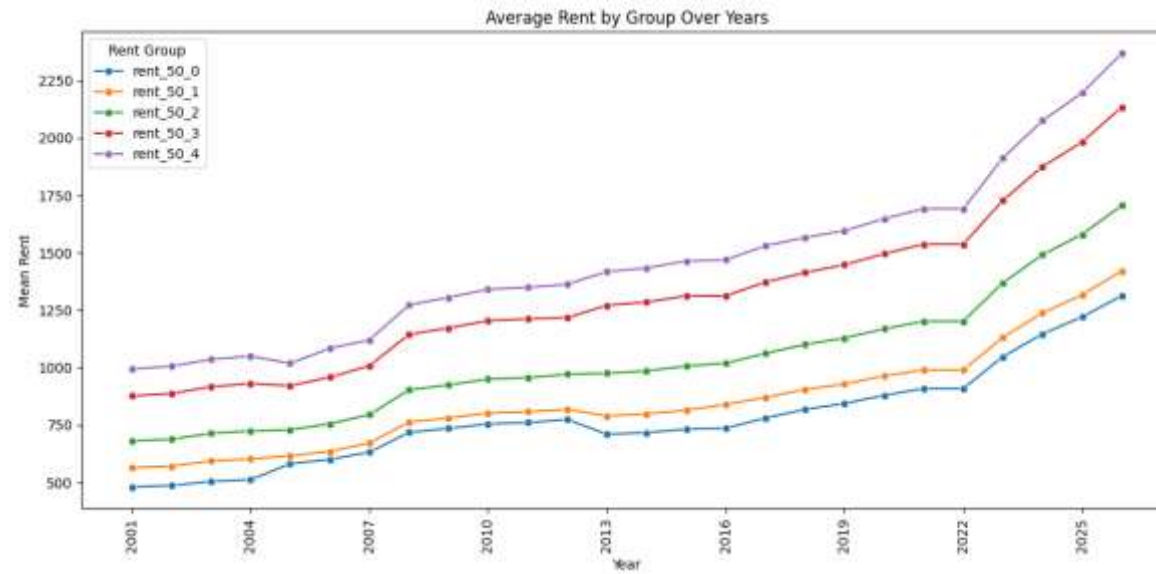
## Long-Short Term Model (LSTM)



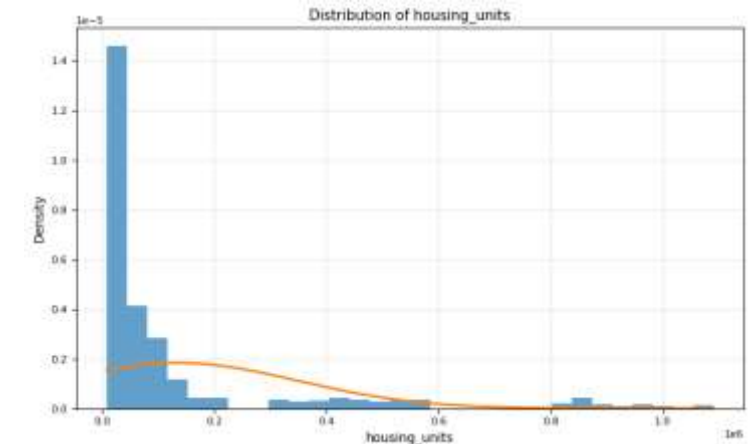
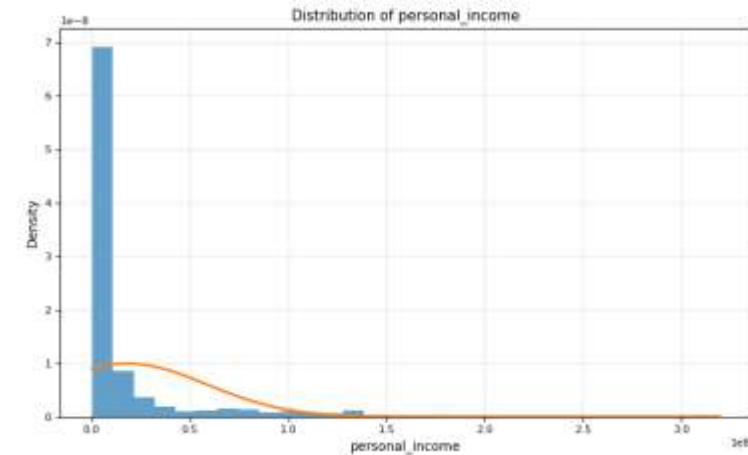
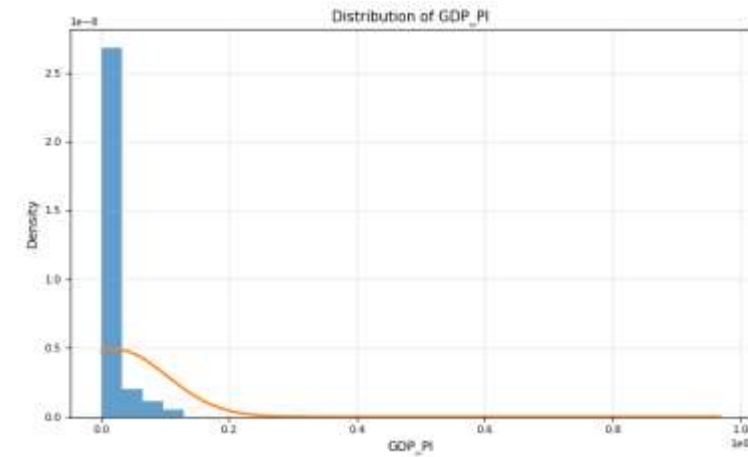
# Methodology: Features, Targets and Preprocessing

	ConvLSTM	XGBoost	LSTM
<b>Target</b>	Rent	Rent	Rent
<b>Features</b>	<ul style="list-style-type: none"> <li>• Spatial – Temporal Patterns (built into model)</li> <li>• GDP_PI</li> <li>• Personal Income</li> <li>• Housing Units</li> <li>• House Size</li> </ul>	<ul style="list-style-type: none"> <li>• GDP_PI</li> <li>• Personal Income</li> <li>• Housing Units</li> <li>• House Size</li> <li>• Rent Lag 1</li> <li>• Rent Lag 2</li> </ul>	<ul style="list-style-type: none"> <li>• GDP_PI</li> <li>• Personal Income</li> <li>• Housing Units</li> <li>• House Size</li> <li>• Rent Lag 1</li> <li>• Rent Lag 2</li> </ul>
<b>Data Preprocessing</b>	<ul style="list-style-type: none"> <li>• Building Spatial Weights (County Neighborhood Structure) using Queen’s Contiguity Weights.</li> <li>• Spatial Grid creation using PCA embeddings</li> <li>• Time-based Data Partitioning at year = 2015</li> <li>• Normalizing the data</li> <li>• Sequence Creation (length = 2)</li> </ul>	<ul style="list-style-type: none"> <li>• Shifting the rent variable by 1 year to create lag 1</li> <li>• Shifting the rent variable by 2 years to create lag 2</li> <li>• Time-based Data Partitioning at year = 2015</li> </ul>	<ul style="list-style-type: none"> <li>• Shifting the rent variable by 1 to create lag 1</li> <li>• Shifting the rent variable by 2 to create lag 2</li> <li>• Time-based Data Partitioning at year = 2015</li> <li>• Normalizing the data</li> <li>• Sequence Creation (length = 2)</li> </ul>

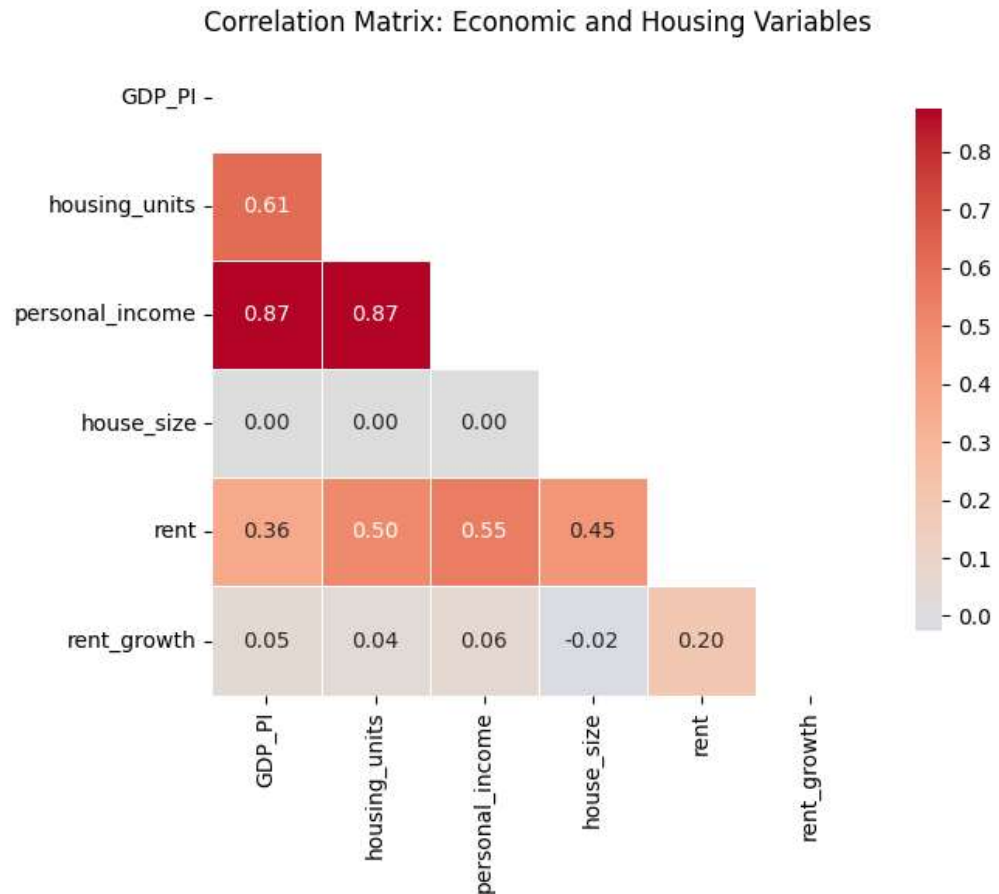
# Exploratory Data Analysis



- The time series chart on the left shows the trend in the growth in median (50<sup>th</sup> quantile) rent across the 5 house sizes. The chart shows that the rent growth across all house sizes has had an increasing trend from 2001 through to 2026. Also median rent for houses with 3 and 4 rooms have had a steeper increase as compared to smaller house sizes.
- The plots on the right show the distribution of the data on the GDP\_PI, Number of Housing Units and Personal Income. All the three distributions are skewed to the right which imply:
  - Majority of county level private investment, number of available housing units and average personal income are below the mean.
  - In very few counties and years there are high levels of private investment, number of available housing units and average personal income.

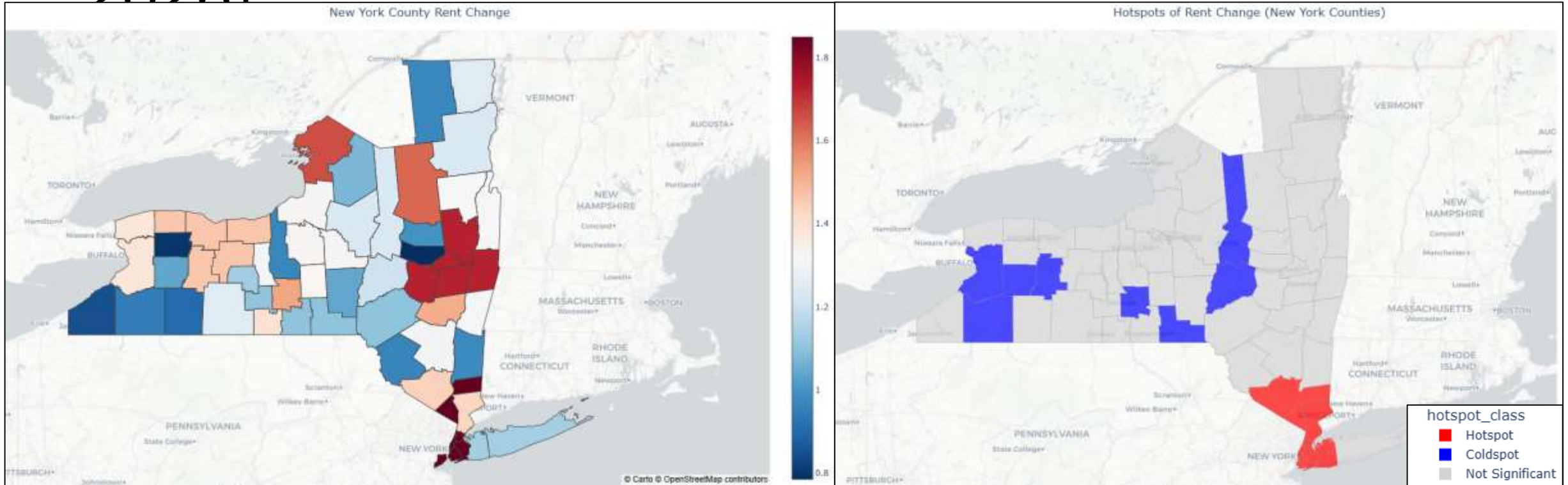


# Exploratory Data Analysis: Correlation



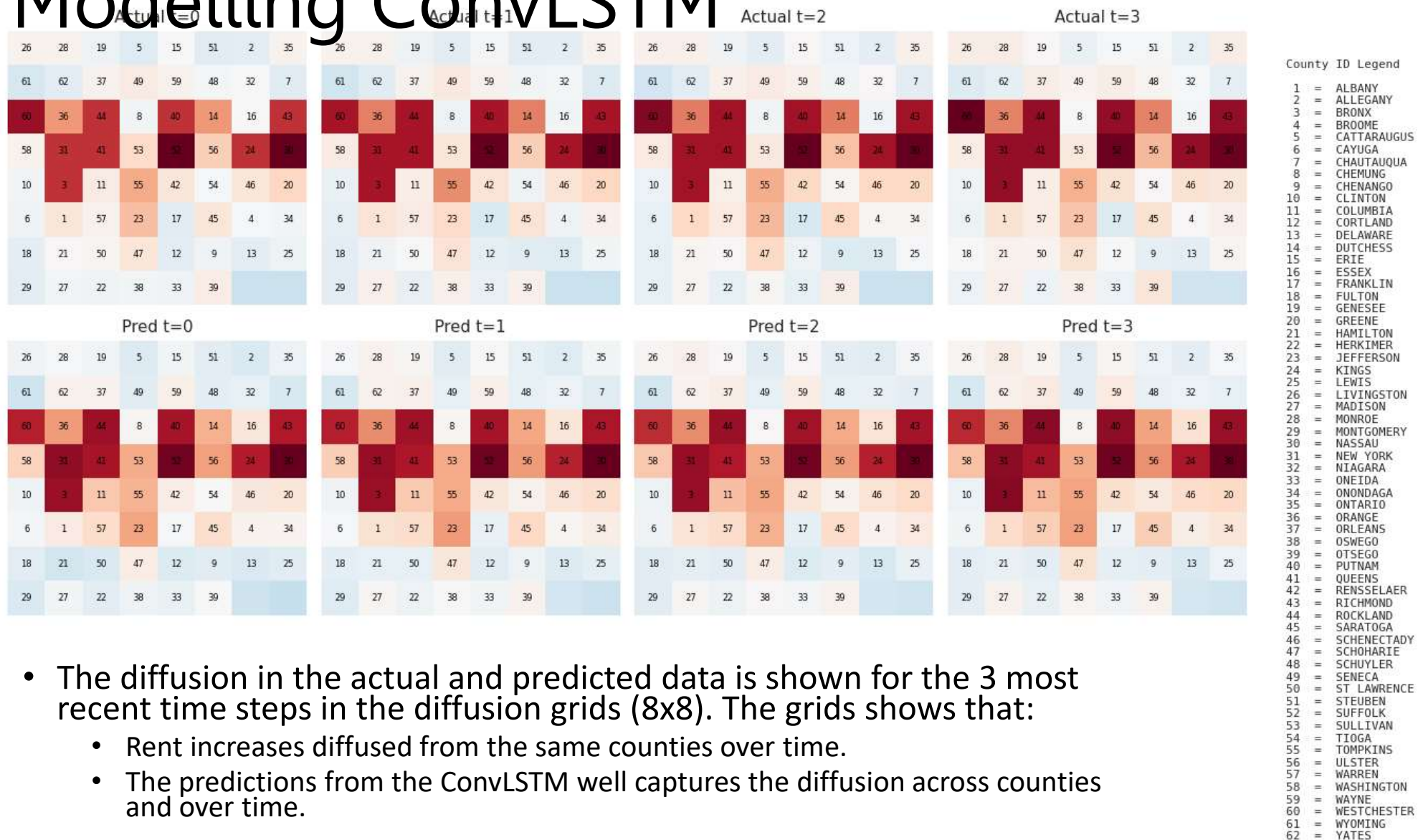
- The correlation analysis show two instances of high correlation:
  - Personal Income and GDP\_PI and
  - Personal Income and Housing Units.
- However, the ConvLSTM, XGBoost and LSTM are not sensitive to multicollinearity and hence would adequately manage multicollinearity.

# Spatial Analysis (2001 to 2026)



- Spatial analysis for the change in the median rent shows that higher changes are evident across the US; with existing clusters of high or low rent growths.
- The corresponding hotspot analysis on the right shows that 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.

# Results: Actual vs Predicted - Diffusion Modelling ConvLSTM



- The diffusion in the actual and predicted data is shown for the 3 most recent time steps in the diffusion grids (8x8). The grids shows that:
  - Rent increases diffused from the same counties over time.
  - The predictions from the ConvLSTM well captures the diffusion across counties and over time.

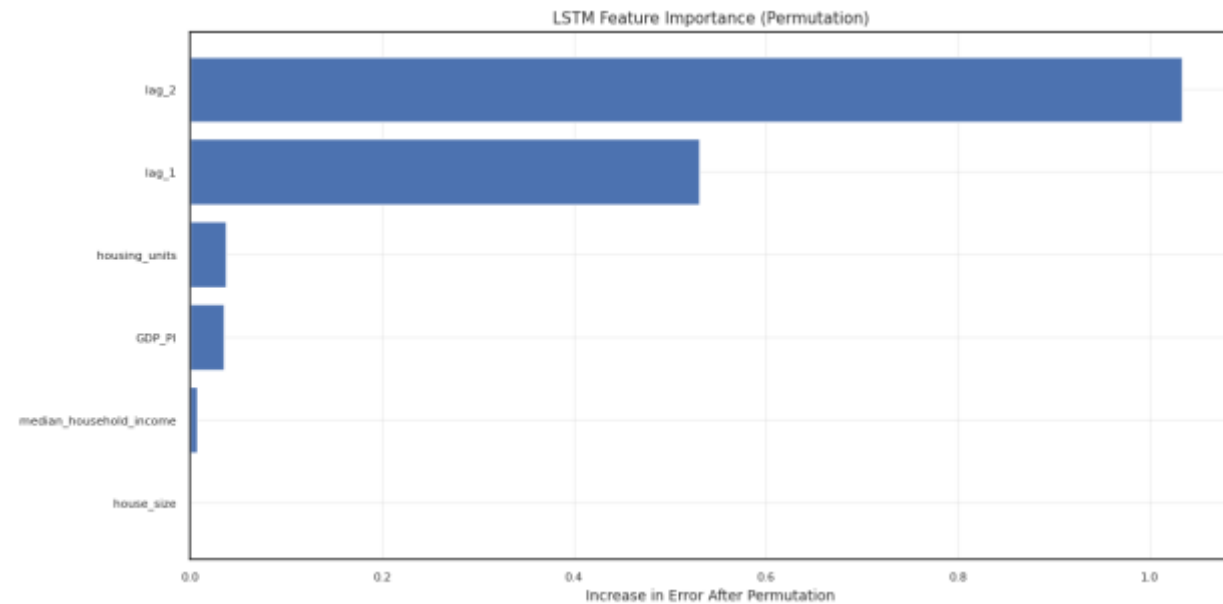
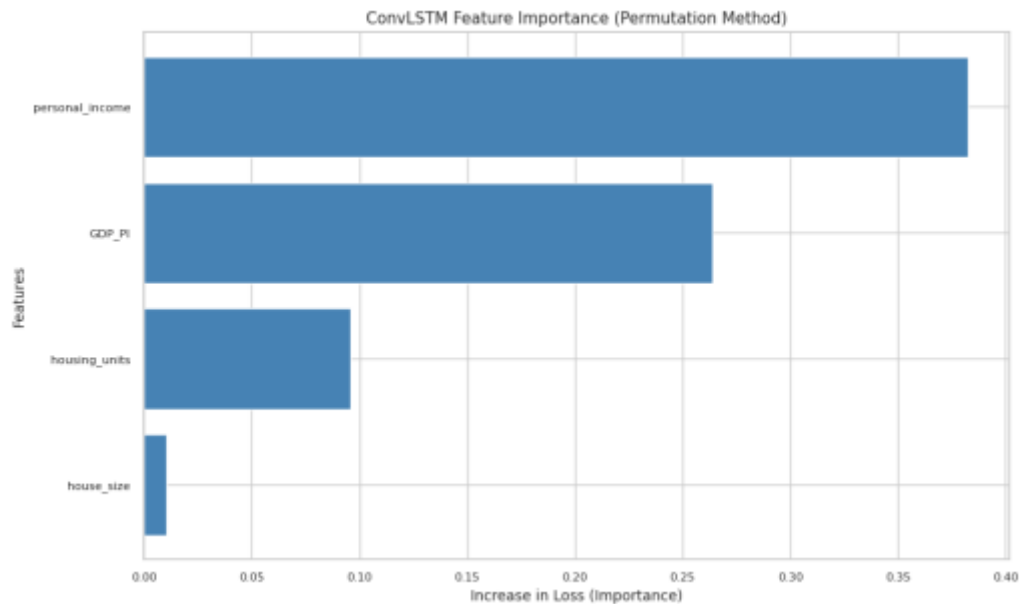
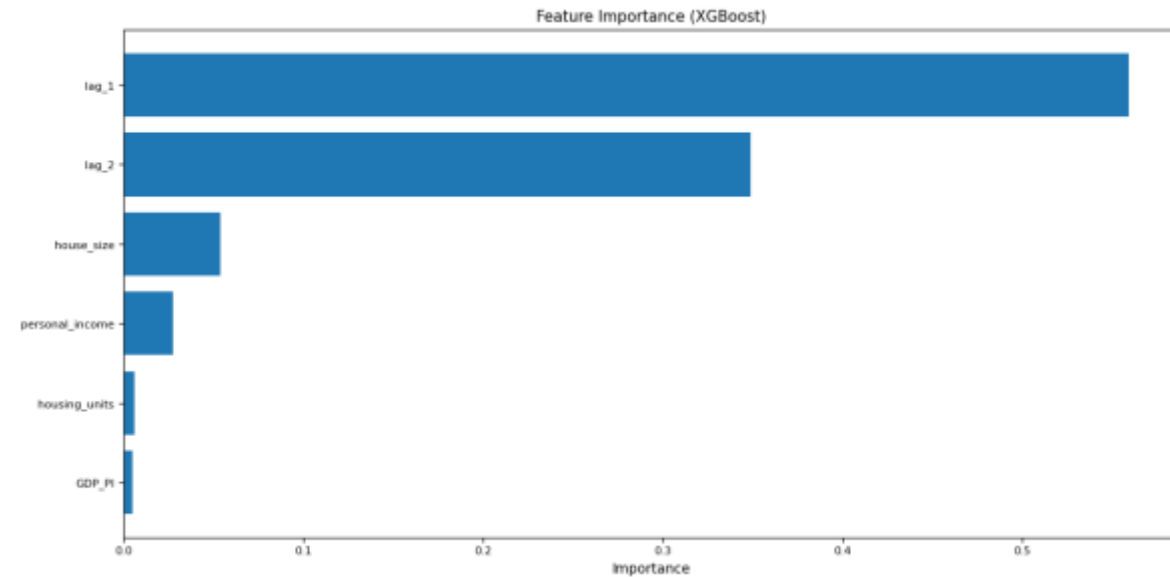
# Results: Model Forecasting Performance

- Spatial-Temporal patterns and GDP PI, Personal Income, Housing Units and House size account for 82.84% ( $R^2 = 0.828419$ ) of the variation (and hence diffusion) of rent.
- First and second rent lags, GDP PI, Personal Income, Housing Units and House size are shown to account for 88.58% ( $R^2 = 0.8858$ ) and 49.91% ( $R^2 = 0.4991$ ) of variation in rent by 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

# Results: Feature Importance

- The ConvLSTM shows that Personal Income followed by the GDP PI had the most influence on Rent.
- Lag 2 and Lag 1 lead as the most influential factors for Rent in both the XGBoost and LSTM models.



# Discussion

- The ConvLSTM showed that 82.84% of the variation in the rent (and hence diffusion) can be attributed to the features of the model; with Personal Income having the highest influence for rent diffusion across counties and overtime.
- The XGBoost outperforms the LSTM with a 88.58% forecasting accuracy as compared to 49.91% forecasting accuracy for the latter; with both models showing Lag 2 and Lag 1 as the top 2 most influential factors for rent diffusion over time.

# Conclusion and Recommendations: Property Pricing and Property Development

- Rent is shown to diffuse, both across counties and overtime where the county, GDP PI, Personal Income, Housing Units, House Size and Rent in previous years explain 82.84% of the rent.
  - Diffusion clusters imply the need for property owners and developers to understand the rent for housing in associated counties from the diffusion grid and the accompanying costs; and utilize them in the pricing strategy.
- Diffusion of rent overtime is shown to be mainly over 2-year periods (Lag 2) and to some extent over 1-year periods (Lag 1).
  - Observing rent over the past 2 years provides a good indication of the direction of present rent through which property owners and developers can anticipate the rent grows.

# References

- [1] R. Derven, “Data center real estate: Challenges and opportunities in the digital age,” *NAIOP Development Magazine*, Winter 2023–2024. 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. 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. <https://doi.org/10.1016/j.jue.2021.103356>
- [4] X. Zhang and Y. Yu, “Migration, transportation infrastructure, and the spatial transmission of COVID-19 in China,” *Journal of Urban Economics*, vol. 127, 2022, Art. no. 103351. doi: 10.1016/j.jue.2021.103351
- [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*, vol. 130, 2021, Art. no. 103326. doi: 10.1016/j.trc.2021.103326
- [6] S. Galiani, P. J. Gertler, and R. Undurraga, “Aspiration adaptation in resource-constrained environments,” *Journal of Urban Economics*, vol. 123, 2021, Art. no. 103326. doi: 10.1016/j.jue.2021.103326
- [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. doi: 10.1016/j.ins.2026.123404
- [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. doi: 10.1007/s13753-025-00685-8
- [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, Art. no. 105971, 2024. doi: 10.1016/j.envsoft.2024.105971.
- [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, Art. no. 1788, 2023. doi: 10.1186/s12889-023-16720-6.