Khyatee Desai

Research Proposal: Projecting Gentrification in New York City Using Multivariate Time Series Forecasting

1. Introduction

Gentrification is a socioeconomic phenomenon characterized by rising property values, changing demographics, and shifts in cultural and economic landscapes. In New York City these changes have profound impacts on neighborhood composition, housing affordability, and local businesses. This research aims to develop a multivariate time series forecasting model to predict gentrification patterns in NYC by census tract, incorporating various socio-economic, demographic, and real estate factors.

2. Problem Statement

Current efforts to understand gentrification are often retrospective, making it difficult for policymakers to proactively address the negative impacts of this phenomena, such as displacement, housing affordability, and cultural disintegration. By forecasting gentrification trends, city planners can be better equipped to make informed decisions about housing policies, community initiatives, and infrastructure development.

3. Research Objectives

- identify the key indicators (ex. housing prices, income inequality, business closure) that drive gentrification
- develop a multivariate time series forecasting model that predicts future gentrification trends across all NYC census tracts
- evaluate the model's accuracy in forecasting gentrification, and assess its potential utility for urban planning and policy-making

4. Data

Data will be sourced from public datasets such as the NYC Open Data portal, the U.S. Census Bureau, and Zillow rental prices. Key variables include housing prices, household income, population demographics, crime rates, business closures, and education levels.

5. Modeling

The first modeling iteration for this project will employ Vector Autoregression (VAR,) which is a popular framework for multivariate time series forecasting. Pending the availability of enough data, a deep learning approach (Long Short-Term Memory RNN) will also be used. The models will be evaluated on historical data using error metrics such as Root Mean Squared Error (RMSE).

6. Timeline

- **September**: Literature review and data collection
- **October**: Data preprocessing and feature engineering
- **November**: Model development and training
- **December**: Performance evaluation, writing, final submission