Structural Exposures and Returns: Forecasting Total Shareholder Return of Public Real Estate Companies by Operating Region and Articulating Value

1) Problem statement

1.1) Problem definition & method (what we will build)

Goal. Predict *next-month* total shareholder return (TSR) for six leading, publicly traded real-estate firms, each linked to one metro: Boston, New York, Washington, Miami, Chicago, Los Angeles.
TSR. (price_{t+1} − price_t + dividends_{t+1}) ÷ price_t.

Unit of analysis. Company-month panel. Predictors are at month *t*; the target is TSR at *t+1*.
Mapping. First pass uses one company → one "Local" metro to keep assumptions simple.

Data on hand (monthly). Prices, dividends, city CPI, MSA unemployment, U.S. 10-year Treasury yield.

Features (time-ordered).

- Local CPI: YoY and MoM change
- Local unemployment: level and MoM change
- 10-year yield: level and MoM change
- Dividend yield (TTM ÷ price), momentum (lag-1 TSR)
- Month dummies if needed

Models.

- Logit, LDA, QDA, Naïve Bayes and KNN, Decision Tree.
- CART decision tree (main). Captures simple interactions (e.g., CPI up helps when unemployment falls; rate spikes hurt); **Bagging:** Random Forest Classifier; **Boosting:** Gradient-boosted trees (XGBoost/LightGBM); **BART** (classification); **SVM (SVC)** with RBF

Training & evaluation. Time-aware rolling split; choose shallow depth / minimum leaf by validation. Metrics: MAE, RMSE, and sign hit rate. Report feature importance and 2–3 short decision paths for interpretability.

Hygiene. Align to month-end; lag all predictors by one month; forward-fill at most one missing month; set a random seed.
Limits (acknowledged now). One-metro link, small sample, no earnings/news shocks.

Concise problem statement

- Method card: variables, split, models, metrics
- Hypothesis table (H1–H4):
    - H1 Labor: higher unemployment → lower next-month TSR
    - H2 Prices: faster CPI with stable/declining unemployment → higher TSR
    - H3 Rates: higher 10-year → lower TSR
    - H4 Anchors: higher dividend yield and positive momentum → slightly higher TSR

1.2) Articulation of value (why this is useful)

This project creates value by turning place-based information into a forecast that is simple to use and explain.

1. Earlier signals. Metro unemployment and CPI often move differently across cities. Our model reads those differences directly, so we can notice turning points for each company before they appear in earnings commentary or national data.
2. Clear explanations. A decision tree shows *why* a forecast changes (for example, "if LA unemployment rose and the 10-year jumped, expected TSR falls"). These rules translate neatly into short notes or slides and match the level of explanation expected in a graduate course.
3. Better risk sense. Because the model watches both local demand and rates, it can flag when conditions deteriorate in a company's key metro. Even if return gains are small, avoiding a few bad months is valuable.
4. Low complexity, easy updates. All inputs—TSR, metro CPI, metro unemployment, and the 10-year yield—are public and already in our files. The process can be run monthly with minimal effort, making the project realistic to maintain through the term.
5. Learning outcome. The project connects concepts we learned—decision trees and testable hypotheses—to a real finance question. We end with an interpretable model, a compact feature set, and a clear evaluation plan.

1.3) Dataset

We built a monthly panel for six leading public real-estate companies tied to six metros (Boston, New York, Washington DC, Miami, Chicago, Los Angeles). For each company–month, the file includes adjusted price, cash dividends, computed next-month TSR, local CPI (YoY and 1-month change), local unemployment (level and change), the U.S. 10-year Treasury yield (level and change), lagged TSR, and dividend yield.
Data come from public, reproducible sources: standard market data portals for prices/dividends, official

U.S. statistical releases for metro CPI and unemployment, and the Federal Reserve for the 10-year. I aggregated to month-end and linked each company to its primary metro (or averaged two when clearly split).

Due to the large size of the dataset, in order to avoid confusion, we will select some parts of it, as well as the data from the past 5 years of the U.S. 10-year Treasury yield, for processing to enhance the efficiency of the model.

2）Assumption and potential economic value

We are studying real estate companies in six different cities, whose primary business concentrated in specific cities or metropolitan areas. Therefore, the future Total Shareholder Return (TSR) of these companies is closely tied to the economic fundamentals of those regions. Based on this, we propose three core hypotheses:

2.1) **Unemployment Rate Hypothesis:**

When a city's unemployment rate rises, it signals a decline in local economic vitality, leading to reduced demand for rental housing and commercial real estate. This results in lower occupancy rates and falling rents. Consequently, the company's cash flow and profits from that city are negatively impacted, driving down its stock price and lowering the following month's TSR.

2.2) **Local Economic Vitality (GDP and Population) Hypothesis:**

If a city experiences continuous population inflows and strong GDP growth, it typically reflects robust demand for real estate. Housing prices and rents tend to rise, improving the company's performance → stock price appreciation → TSR increase.

Conversely, population outflows or declining GDP indicate weakening demand, putting pressure on company performance and leading to a decrease in TSR.

2.3) **Interest Rate (10-Year Treasury Yield) Hypothesis:**

Real estate companies are highly dependent on financing, and investors often compare their dividends with the risk-free rate.

When the 10-year Treasury yield rises: Financing costs increase, placing downward pressure on company profitability; Higher risk-free rates make bonds more attractive, prompting investors to shift away from real estate stocks. As a result → real estate company stock prices decline → TSR decreases.

3）Business Case related

1.Logic

Find out a real world example of a REIT who is using a model which takes region and customer into account to measure the company's total value or shareholder return. The result may be the company knows more about expanding business in a new area. Or the example can be a stakeholder is curious about his (or her) current investment in REITS filed, so he is using a model to see the return and analyze the firms' value.

2. Current content of business case

Simon Property Group (SPG), the world's largest retail REIT, has consistently pursued geographic diversification and tenant diversification to stabilize returns, mitigate regional market risks and enhance Total Shareholder Return.

Despite the disruptions in the retail landscape driven by e-commerce, SPG's premium assets and high-quality tenant mix drove resilient performance: its five-year TSR (2019–2023) reached 51%, outperforming the FTSE Nareit All Equity REITs Index at 35%.

This business case reveals that by modeling structural exposures in operating regions and tenant positioning does enhance TSR and reduce capital costs, which brings tangible economic

Academic research has clearly revealed the systemic impact of tenant credit and geographical location on real estate cap rate. In the case of Letdin et al. (2022), the study analyzed more than 8,200 single-tenant net lease (STNL) transactions between 2005 and 2019 and found that tenant credit ratings, parent company listing status, lease term, and tenant default risk were all highly correlated with cap rates. Specifically, a tenant's credit rating downgrade would result in an average cap rate increase of 40–60 basis points, reflecting a higher risk compensation for future cash flow stability. Similarly, if the tenant is unlisted or lacks a long-term credit history, the cap rate will tend to be 30–50 basis points higher than that of a listed blue-chip tenant.

In addition to tenant characteristics, location quality is also an important factor in determining cap rate. Research shows that properties located in population-growing, economically active first-tier areas such as Austin, Texas or Orlando, Florida tend to have cap rates 70–100 basis points lower than in the Midwest or depressed regions. This is because first-tier markets have higher rental growth expectations, lower vacancy rates, and investors are willing to accept lower returns in exchange for more stable cash flow. Conversely, in areas with weak economic foundations or population exodus, investors demand higher cap rates to offset potential risks.

2.1) Background and challenge of Simon Property Group

Simon Property Group is the world's largest retail real estate investment trust, managing over 180 shopping malls and outlet malls. As a leading retail REIT, it faces two major structural challenges: the long-term impact of e-commerce and tenant defaults and declining customer traffic caused by the pandemic. These factors have led investors to widely question the sustainability of its Total Shareholder Return. Regional and tenant exposure are key determinants of TSR: excessive concentration in low-growth markets or reliance on a single retailer can significantly impact returns.

2.2) Strategy

In its official financial report, SPG emphasized "differentiation in product types, geographies, and property formats." As of 2023, the company will own 195 retail properties in the United States (93 shopping centers, 69 premium outlets, etc.), 35 international outlet/designer outlets in 13 countries worldwide, and a 22.4% stake in France's Klépierre, indirectly covering 14 European countries. This global presence effectively hedges against the risks of economic fluctuations in a single region.

In terms of customer structure, SPG leverages partnerships with highly reputable global retailers (such as LVMH, Gucci, Apple, and Nike) and continuously introduces growing brands originally online (such as Warby Parker, Glossier, and Vuori) to optimize its tenant mix, ensure stable rental income, and improve sales per square meter.

2.3) Data

According to the annual report of SPG (2023), Consolidated revenue was $5.66 billion (up 7% year-over-year); FFO was $4.69 billion, a record high; NOI from U.S. properties increased 4.8% to $5.26 billion; NOI from international properties increased 4.9% (at constant currency); Retailer sales were $743 per square foot; Occupancy was 95.8% (up 90 basis points year-over-year); Cash dividends were $2.8

billion ($7.45 per share), up 8% year-over-year. In addition, the company has a total market capitalization of $90 billion, has distributed over $42 billion in dividends to shareholders over 30 years, and has a long-term TSR of 3,100%.

These figures clearly demonstrate the stable cash flow and capital returns generated by its diverse geographic and tenant base.

2.4) Strategic significance

Based on 2023 data, if NOI growth were to decline by 2% due to regional concentration risk, it would result in a decrease of approximately $120 million in NOI, further compressing FFO and shareholder dividends. SPG's multi-regional footprint effectively mitigates such a decline, effectively preserving hundreds of millions of dollars in shareholder value. Furthermore, an optimized tenant mix and an increased proportion of high-credit customers reduce the risk of rental defaults and lower capital market risk premiums.

Project Plan（GPT）

Week 1: Problem Definition and Environment Setup

We will clarify our research objectives: predicting the Total Shareholder Return (TSR) for six major listed real estate companies in the following month, and examining the relationship between macroeconomic variables (unemployment rate, CPI, interest rates) and corporate financial variables (dividend yield, momentum, etc.). Complete the business case and research hypotheses (H1–H4), and establish a GitHub repository to centrally store code, data, and documentation. Simultaneously, we will configure the Jupyter/Colab environment and ensure all required Python libraries (pandas, scikit-learn, xgboost, statsmodels, etc.) are fully installed.

Week 2: Data Acquisition and Preliminary Preparation

We will collect and organize data, including company stock prices, dividends, regional CPI, regional unemployment rates, and 10-year U.S. Treasury yields. Check for missing values and outliers, then perform preliminary alignment to ensure time indices are unified into a “company-month” panel format.

Week 3: Exploratory Data Analysis (EDA)

Plot each company's TSR trends to observe differences in macroeconomic variable patterns across regions. Calculate correlations between variables and conduct descriptive statistical analysis to preliminarily identify features potentially strongly correlated with TSR.

Week 4: Data Modeling Preparation

Shift all predictor variables one month backward to ensure correct causality order. Construct the panel dataset and split it into training and testing sets (using time-based segmentation to prevent information leakage). Standardize or normalize variables for subsequent modeling.

Week 5: Feature Engineering

Generate economic features (e.g., CPI year-over-year change, unemployment rate month-over-month change, interest rate spread) and financial features (dividend yield, lagged TSR, month dummy variables). Attempt to construct interaction terms (e.g., “CPI × unemployment rate interaction”) to capture combined effects of economic conditions.

Week 6: Baseline Model

Establish the simplest regression baseline model (OLS or ridge regression), evaluated using RMSE, MAE, and directional prediction accuracy (hit rate). This model serves as a benchmark to measure improvements over more complex models.

Week 7: Intermediate Model Development

Employ tree-based models (decision trees, random forests) to explore nonlinear relationships and variable interactions. Fine-tune parameters (e.g., tree depth, minimum leaf node size) through cross-validation.

Week 8: Advanced Model Development

Experiment with gradient boosting (XGBoost, LightGBM) and support vector machines (SVM). These models better capture complex relationships but require consideration of interpretability. Document performance differences across models at this stage.

Week 9: Model Selection and Comparison

Compare all models (OLS, Random Forest, Boosting, SVM), evaluating both predictive performance and interpretability. Select the top-performing model as the final choice. Write a progress report justifying the selection.

Week 10: Data-Centric AI Approach

Re-examine data quality and feature engineering. Consider data augmentation or generating new features (e.g., volatility-based metrics) and evaluate whether they can further enhance model performance.

Week 11: Risk and Ethics Analysis

Analyze potential risks and biases in the model, such as sample size imbalances across cities potentially leading to a bias toward companies in larger cities. Simultaneously discuss ethical issues related to using financial market data (e.g., potential market sensitivity). Write the model risk and ethical considerations section.

Week 12: Model Preservation and Deployment Preparation

Save the final model (in joblib/pickle format) and document input/output formats to ensure future accessibility. Design a model monitoring plan covering data drift, performance degradation, and potential bias detection.

Week 13: Integration and Submission

Consolidate all reports, code, and notebooks into a final PDF report (10 pages of main text + appendices). Ensure the GitHub repository fully contains all deliverables and conduct a comprehensive end-to-end test (verifying notebooks can reproduce results from start to finish).

- Identify the type of modeling that will solve the problem you identified: supervised or unsupervised. If supervised, is this a classification or regression problem? If it is classification, is it binary or multi class? If it is unsupervised what type of unsupervised learning is it?

In this project, we will use a supervised learning model, and it will be defined as a regression problem. This is because our prediction target is the total shareholder return (TSR) for the following month, and it is a continuous variable. It is jointly determined by stock price changes and dividend income. Supervised learning leverages historical data mappings to train models and capture the patterns by which macroeconomic indicators (CPI, unemployment rate, interest rates) and company characteristics influence TSR. Since TSR is expressed as a continuous percentage return, a regression model is more suitable for our task than a classification model. By solving these regression problems, we can obtain numerical forecasts for future returns and quantify the marginal effects of various economic and financial factors on TSR.

Github repo: https://github.com/zhaoyangLin1008/test