**Final Project Report**

**Project Title: PROPERTY APPRECIATION ESTIMATION AND RECOMMENDATION FOR STRATEGIC REAL ESTATE INVESTMENTS**

**1. Introduction**

Problem Statement and Motivation

The U.S. housing market is influenced by a multitude of factors—macroeconomic indicators (e.g., mortgage rates, GDP), local and seasonal variations, and even political cycles. For real estate investors, identifying locations with the highest potential returns and understanding the trajectory of property values is critical. This project aims to forecast property value trends at the zip code level and determine the top areas to invest in over a one-year horizon. By leveraging historical price data, mortgage rates, and other indicators, we seek to empower data-driven investment decisions.

**Objectives**

1. Forecast Housing Price Changes:

Use time series analysis and regression models to predict short-term (1 month), medium-term (1 quarter), and longer-term (1 year) changes in property values.

2. Examine Influential Factors:

Investigate how mortgage rates, seasonal patterns, and political/election cycles correlate with property value trends.

3. Identify Optimal Investment Regions:

Pinpoint top U.S. zip codes that are likely to yield the highest return on investment over a one-year period.

**Scope**

This analysis focuses on U.S. housing prices at the zip code level from January 2000 through September 2024, supplemented by national mortgage rate data and political timeline markers. The results inform both individual investors and corporate stakeholders about market dynamics and potential investment hotspots.

**2. Data**

**Primary Data Sources:**

1. Zillow Housing Data (Zillow Home Value Index - ZHVI):

• Description: Monthly median home values for thousands of U.S. zip codes.

• Size: Approximately 26,338 rows × 305 columns.

• Coverage: County, city, state, and zip code-level granularity spanning over two decades.

• Link/Reference: Zillow’s official website (dataset originally obtained via downloadable CSV).

2. Mortgage30US Dataset (FRED - Federal Reserve Economic Data):

• Description: Monthly 30y fixed mortgage rates in the US from 2000 to 2024.

• Size: Approximately 1,297 rows × 2 columns.

• Link/Reference: https://fred.stlouisfed.org/series/MORTGAGE30US

**Data Cleaning & Preprocessing:**

• Zillow Data:

• Removed redundant metadata columns.

• Handled missing values (NaN) using forward-fill techniques.

• Converted date columns into numeric or DateTime indices for time series analysis.

• Computed monthly percentage changes and yearly average percentage changes to normalize comparisons across regions.

• Mortgage Data:

• Filtered mortgage rate data to the 2000–2024 period for consistency.

• Reversed the mortgage rate time series to compare it inversely against property values, reflecting the known inverse relationship (as mortgage rates decrease, property appreciation often increases).

Data Submission:

• Due to size constraints, cleaned data files or data cleaning scripts are provided separately.

• Scripts used for data cleaning and transformations are included in the submitted code files.

3. Exploratory Data Analysis (EDA)

**Key Questions Explored:**

1. Influence of Macroeconomic Indicators:

We examined mortgage rates, GDP trends, and Federal interest rates to identify correlations with property value changes. Visualizations showed an inverse relationship between mortgage rates and property appreciation: as mortgage rates hit record lows (e.g., 2020–2021), property values surged.

2. Seasonal Trends in Property Prices (2000–2025):

By categorizing each month into a season (Fall, Winter, Spring, Summer), we analyzed seasonal price patterns. Surprisingly, while slight variations exist, all seasons displayed similar overall upward trajectories. After about 2015, seasonal effects diminished, indicating that broader economic factors overshadow minor seasonal patterns.

3. Election Cycles and Housing Prices:

Over multiple election years, we observed that certain administrations and their policies influenced housing markets. For example, policies that favored homeownership and tax breaks correlated with periods of price growth. The impact of the 2008 financial crisis and subsequent government interventions were also evident in the data’s historical patterns.

Notable Visualizations:

Need to add

**4. Modeling and Statistical Analysis**

**Techniques Used:**

1. ARIMA (AutoRegressive Integrated Moving Average) Model:

• Goal: Forecast property price changes for the next month, quarter, and year at the zip code level.

• Data Window: Predictions were tested using different historical training spans (e.g., last year, last 5 years, last 10 years, full dataset from 2000 onward).

• Findings:

• Directional accuracy (predicting if prices go up or down) hovered around 35%–50% depending on horizon and training history.

• Short-term forecasts (1 month ahead) were more stable, while 1-year-ahead forecasts had larger errors.

• Quarterly predictions performed worst in terms of accuracy, suggesting ARIMA struggled with that particular horizon.

2. Linear Regression with Lag Features:

• Goal: Explore a simpler model’s ability to predict price movements using lagged property values and potentially macroeconomic indicators.

• Findings:

• Achieved about 70% directional accuracy over the full dataset (2000–2024).

• Although less sophisticated than ARIMA, the regression model benefited from feature engineering (e.g., lags) and simpler interpretations.

**Additional Analyses:**

Tested the inclusion of mortgage rates as a feature in the regression model. While it slightly improved accuracy, the improvement was not statistically significant.

**Summary of Modeling Results:**

• ARIMA: Suitable for time-series forecasting but yielded moderate performance.

• Linear Regression (with lags): Outperformed ARIMA in terms of directional accuracy, indicating simpler models with well-chosen features can sometimes be more effective.

With these two distinct modeling techniques plus the incorporation of macroeconomic indicators, we fulfill the ML/Stats requirements and produce inferences valuable for understanding market dynamics.

**5. Results and Insights**

**Correlations and Inferences:**

• Mortgage Rates vs. Property Values: Inverse relationship confirmed. When mortgage rates dropped to historical lows (2020–2021), property values appreciated strongly.

• Seasonality: Minimal seasonal effect on price trends, especially after 2015. Market forces and economic conditions overshadowed seasonal patterns.

• Political and Economic Context:

• Certain administrations’ policies (e.g., tax incentives, stimulus acts) coincided with periods of growth or stability in housing markets.

• Post-2008 financial crisis interventions influenced the recovery trajectory.

• By 2024, with projected interest rate adjustments and a new political regime, lower mortgage rates are expected to sustain or increase housing demand.

**Top 5 Zip Codes for Investment:**

Based on the forecasts and analysis, the top 5 recommended zip codes to invest in (for a 1-year horizon) included:

• 48505 (Flint, MI) - Genesee County

• 61605 (Peoria, IL) - Peoria County

• 36610 (Prichard, AL) - Mobile County

• 71103 (Shreveport, LA) - Caddo Parish

• 62914 (Cairo, IL) - Alexander County

These selections are derived from analyzing forecasted price appreciation and market stability metrics, offering potential higher ROI opportunities.

**6. Visualizations Provided**

Need to add

**7. Discussion and Next Steps**

**Model Limitations & Future Enhancements:**

• Model Complexity: ARIMA alone, trained solely on price history, provided limited predictive power. Incorporating mortgage rates and other macroeconomic indicators improved performance slightly, but more sophisticated models (e.g., Gradient Boosted Trees, LSTM networks) might yield better results.

• Feature Engineering: Adding new predictive features (regional economic health indicators, unemployment rates, inventory levels, etc.) may improve forecasts.

• Hyperparameter Tuning & Model Selection: Further refinement of ARIMA parameters or using advanced ML methods could narrow the accuracy gap between our predictions and Zillow’s estimators.

**Evaluation Criteria:**

Future iterations aim for:

• Improved predictive accuracy (beyond current 50–70% directional accuracy).

• Better capture of price trends through new features and advanced modeling.

• Robust generalization across different real estate market conditions.

**8. Reflection**

**Challenges:**

• Data Complexity: Handling and transforming large, multi-year datasets was time-consuming. EDA was critical to discovering underlying relationships.

• Interpreting Results: Decoupling the effects of mortgage rates, policies, and seasonal factors was non-trivial. Each factor’s influence varied over time.

**Key Insights:**

• A clear inverse relationship between mortgage rates and property appreciation.

• Minimal seasonal effects in recent years.

• Political and economic policies can indirectly influence housing markets, though these effects are interwoven with broader macroeconomic trends.

**Progress and Continuing Work:**

• While initial modeling approaches provided a foundation, there is room to improve accuracy and interpretability.

• Future efforts will involve testing new models, refining features, and iterating on evaluation metrics to better meet investor needs.

**Conclusion**

This project leveraged a complex dataset of U.S. housing prices, supplemented by mortgage rates and contextualized by economic and political timelines, to forecast property value changes and identify investment opportunities. While ARIMA models captured certain patterns, linear regression models with lagged features performed better, highlighting the importance of careful feature engineering and model selection. The results confirmed the inverse relationship between mortgage rates and property appreciation and found that seasonal factors play a limited role in long-term trends.

With these insights, we identified several promising zip codes for future investments. Going forward, integrating more robust data points, experimenting with advanced machine learning architectures, and refining our approach to model evaluation will further enhance our predictive capabilities and provide deeper insights for real estate investors.

**Peer Evaluation**