

# Final Report: Real Estate Market Analysis and Loan-to-Value Adjustment Framework

## 1. Overview

This project presents a comprehensive framework for analyzing real estate market dynamics at the ZIP-code level. The primary objectives include the calculation of **Market Scores** for assessing real estate competitiveness and the adjustment of **Loan-to-Value (LTV)** ratios to mitigate risks in high-demand areas. By combining statistical analysis, geospatial mapping, and a transparent rule-based scoring model, the framework delivers actionable insights for stakeholders in the real estate and lending industries.

## 2. Data Source

The data used in this project is sourced from Redfin's Housing Market Data, which provides up-to-date, reliable, and granular real estate market data. Redfin, as a real estate brokerage, has direct access to data from local multiple listing services (MLS) and insights from agents across the U.S., ensuring the data's accuracy and timeliness.

This rich dataset allowed the development of a highly localized analysis, focusing on California ZIP codes and the 2024 real estate market.

## 3. Objectives

1. **Evaluate Market Competitiveness:**
  - Develop a scoring mechanism to rank ZIP codes based on real estate competition levels.
2. **Dynamically Adjust LTV Ratios:**
  - Modify LTV ratios based on competitive conditions to balance risk and opportunity.
3. **Provide Actionable Insights:**
  - Deliver visual and statistical outputs to guide lenders, policymakers, and investors.

## 4. Understanding Competition and LTV

The relationship between a house's competition and its LTV ratio is dynamic, influenced by market conditions, borrower behavior, and lender policies:

### 4.1 Key Concepts

- **Loan-to-Value (LTV) Ratio:**
  - $LTV = (\text{Loan Amount} / \text{Property Value}) \times 100$ .
  - Higher LTV indicates a higher loan amount relative to property value, posing greater risk to lenders.
  - Lower LTV suggests a safer loan, with a smaller loan amount compared to property value.

### 4.2 Market Conditions and LTV

| Market Condition                          | Effect on Property Value   | Impact on LTV  |
|---|--|--|
| <b>High Competition</b> (Seller's Market) | Higher property values due to bidding wars.                      | Higher LTV to encourage borrowing due to stable demand and price growth.                     |
| <b>Low Competition</b> (Buyer's Market)   | Lower property values as sellers lower prices to attract buyers. | Lower LTV to mitigate risks associated with slower price growth or potential price declines. |

## 5. Model Overview

### 5.1 Why a Rule-Based Model?

The decision to employ a rule-based model was driven by the need for:

- **Interpretability:** Stakeholders require clear, transparent scoring mechanisms for decision-making.
- **Flexibility:** Scoring rules and weights can be easily adjusted to reflect changing market conditions or priorities.
- **Domain-Driven Approach:** The model leverages domain knowledge to define feature importance and scoring logic, ensuring relevance to real estate dynamics.

### 5.2 Research and Feature Selection

Through detailed research and statistical analysis, the following features were identified as the most impactful in assessing market competitiveness:

1. **SOLD\_ABOVE\_LIST**: Indicates how frequently homes sell above their list price, showcasing demand intensity.
2. **AVG\_SALE\_TO\_LIST**: Measures the average sale-to-list price ratio, reflecting price competitiveness.
3. **PENDING\_SALES\_YOY**: Captures year-over-year growth in pending sales, signaling market momentum.
4. **MEDIAN\_DOM**: Represents the median number of days on market; fewer days suggest stronger demand.

These features were selected based on their statistical significance, practical relevance, and interpretability, as highlighted in the correlation analysis and feature importance outputs.

### 5.3 Scoring Framework

The **Market Score** is calculated as a weighted combination of the normalized feature values:

- **Weights:**
  - **SOLD\_ABOVE\_LIST**: 30%
  - **AVG\_SALE\_TO\_LIST**: 30%
  - **PENDING\_SALES\_YOY**: 20%
  - **MEDIAN\_DOM**: 20% (inverted so lower values indicate higher competitiveness).
- **Normalization:**
  - All features are scaled to a 0–1 range using Min-Max normalization.
- **Score Scaling:**
  - The final score is scaled to a 0–100 range, with higher scores indicating more competitive markets.

### 5.4 Adjusted LTV Framework and Rationale

The **Adjusted LTV Framework** dynamically modifies LTV ratios based on the Market Score to manage risk and incentivize lending activity:

#### Rationale

1. **Higher Competition = Lower Risk:**
  - In highly competitive markets, strong demand and price appreciation reflect market stability, reducing the risk of defaults. Higher LTVs encourage lending activity in these areas.

## 2. Lower Competition = Higher Risk:

- In less competitive markets, lower LTVs are used to mitigate the risk of lower demand.

### Adjustment Logic

The following thresholds are used to adjust the LTV:

| Market Score Range | Adjustment to Base LTV (70%) | Final Adjusted LTV | Rationale   |
|--------------------|------------------------------|--------------------|---|
| 90 - 100           | Base + 10                    | 80%                | High competition reflects stable markets. Higher LTV encourages borrowing.                              |
| 80 – 89            | Base + 7                     | 77%                | Strong demand but slightly less competitive than top-tier markets.                                      |
| 60 – 79            | Base + 3                     | 73%                | Moderately competitive; slight increase in LTV.   |
| 40 – 59            | Base                         | 70%                | Neutral markets; no adjustment.   |
| 20 – 39            | Base - 10                    | 65%                | Low competition signals potential risks. Reduced LTVs protect lenders from price stagnation or decline. |
| 0 - 19             | Base - 10                    | 60%                | Very low competition; Significantly reduced LTVs mitigate risks in these volatile markets.              |

### Rationale

#### 1. High Competition:

- Competitive markets are seen as stable, with robust demand and price appreciation, warranting higher LTVs to encourage borrowing.

#### 2. Low Competition:

- Riskier markets require lower LTVs to mitigate risks of price stagnation or decline.

## 6. Insights and Outputs

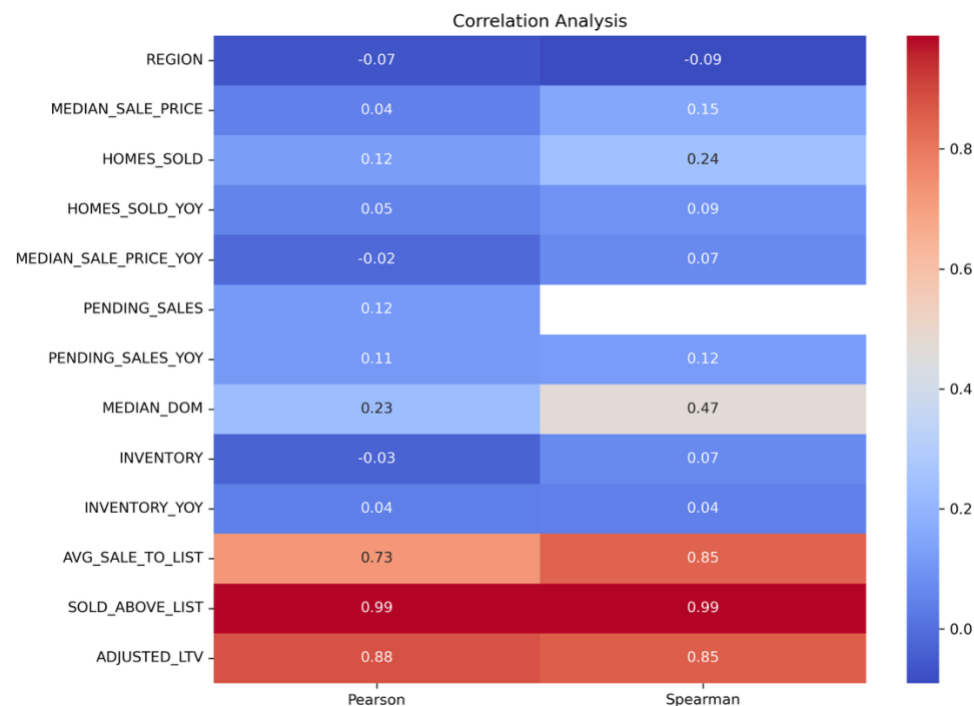
### 6.1 Statistical Insights

#### Correlation Analysis

- **Key Observations:**
  - SOLD\_ABOVE\_LIST and AVG\_SALE\_TO\_LIST exhibit the highest correlation with Market Scores, emphasizing their role in determining competitiveness.
  - Moderate correlations exist between PENDING\_SALES\_YOY and HOMES\_SOLD\_YOY and the Market Score, indicating their relevance in identifying demand shifts.

#### Feature Importance

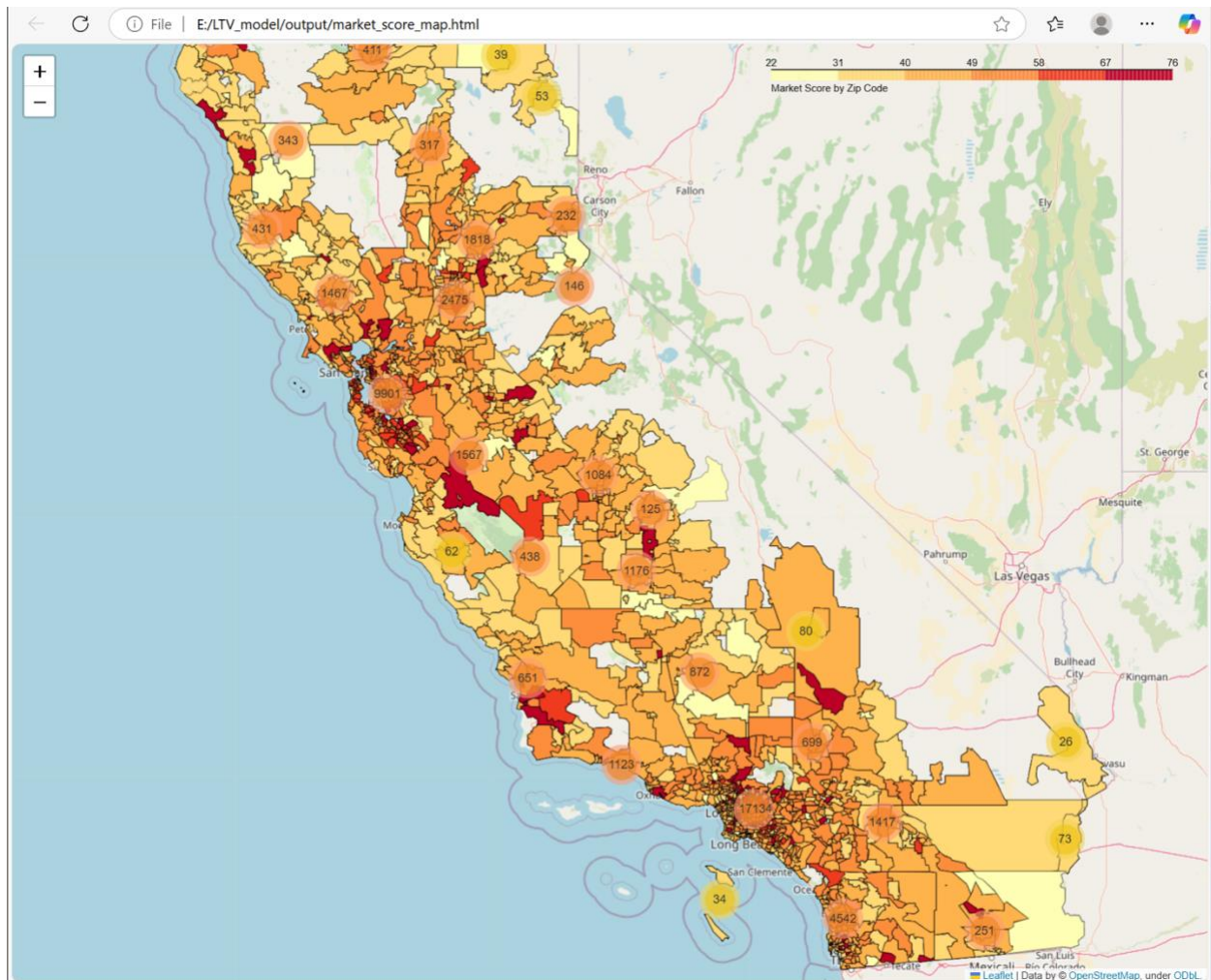
- **Top Contributors:**
  - SOLD\_ABOVE\_LIST was identified as the most significant feature, contributing over twice the importance of AVG\_SALE\_TO\_LIST.
  - This confirms the importance of metrics that reflect buyer behavior and competition in the market.



### 6.2 Geospatial Insights

## Market Score Map

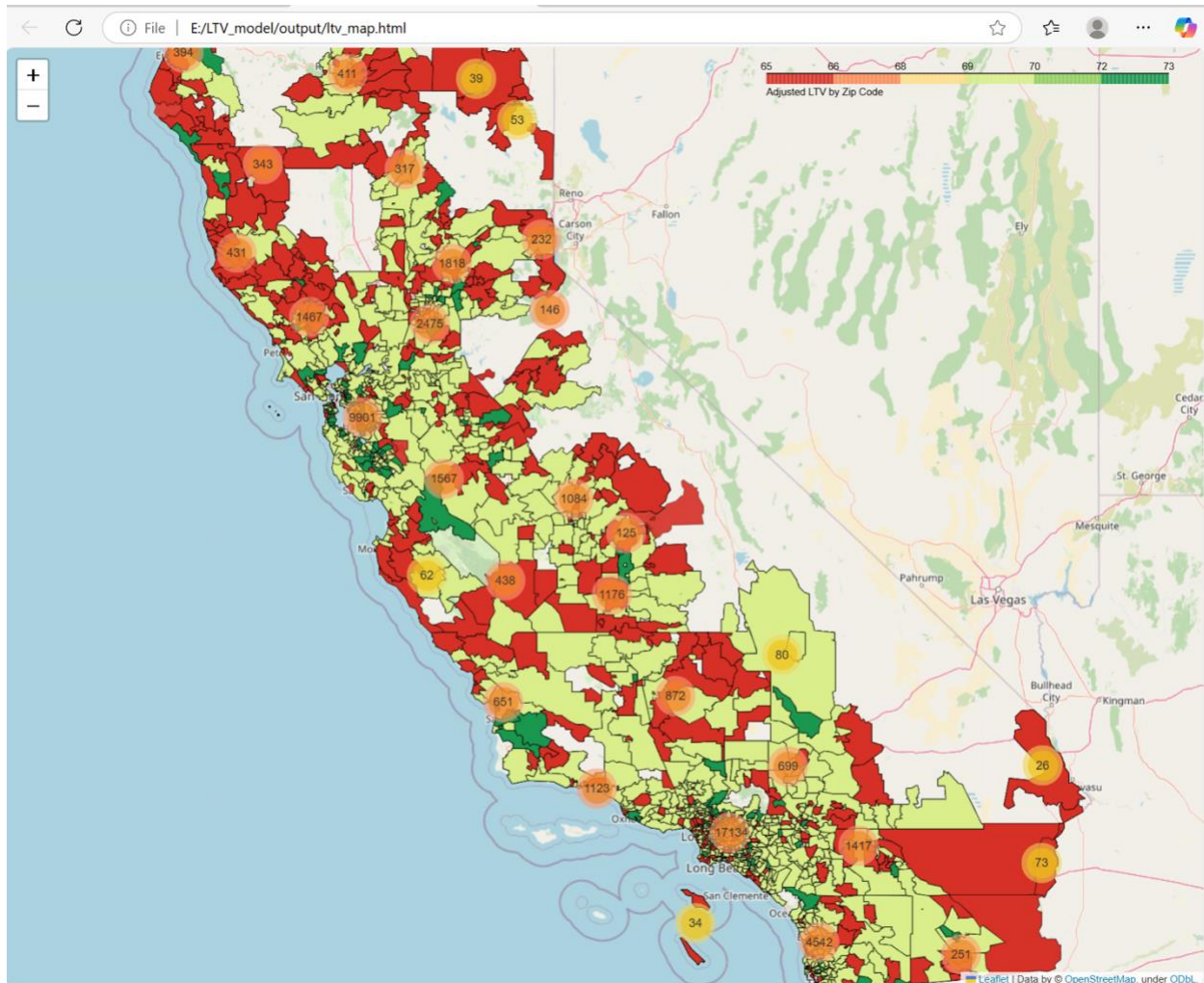
- **Observations:**
  - High Market Scores (red areas) are concentrated in urban and suburban regions, such as Los Angeles, San Francisco, and San Diego.
  - Lower scores (yellow areas) are prevalent in rural or less competitive markets.
- **Insights:**
  - Competitive ZIP codes align with economic hubs, where demand outpaces supply.
  - These areas are prime targets for lenders seeking opportunities with reduced risk.



## Adjusted LTV Map

- **Observations:**
  - LTV adjustments positively correlate with Market Scores, with competitive markets receiving higher LTVs (green zones), and less competitive markets receiving lower LTVs (red zones).

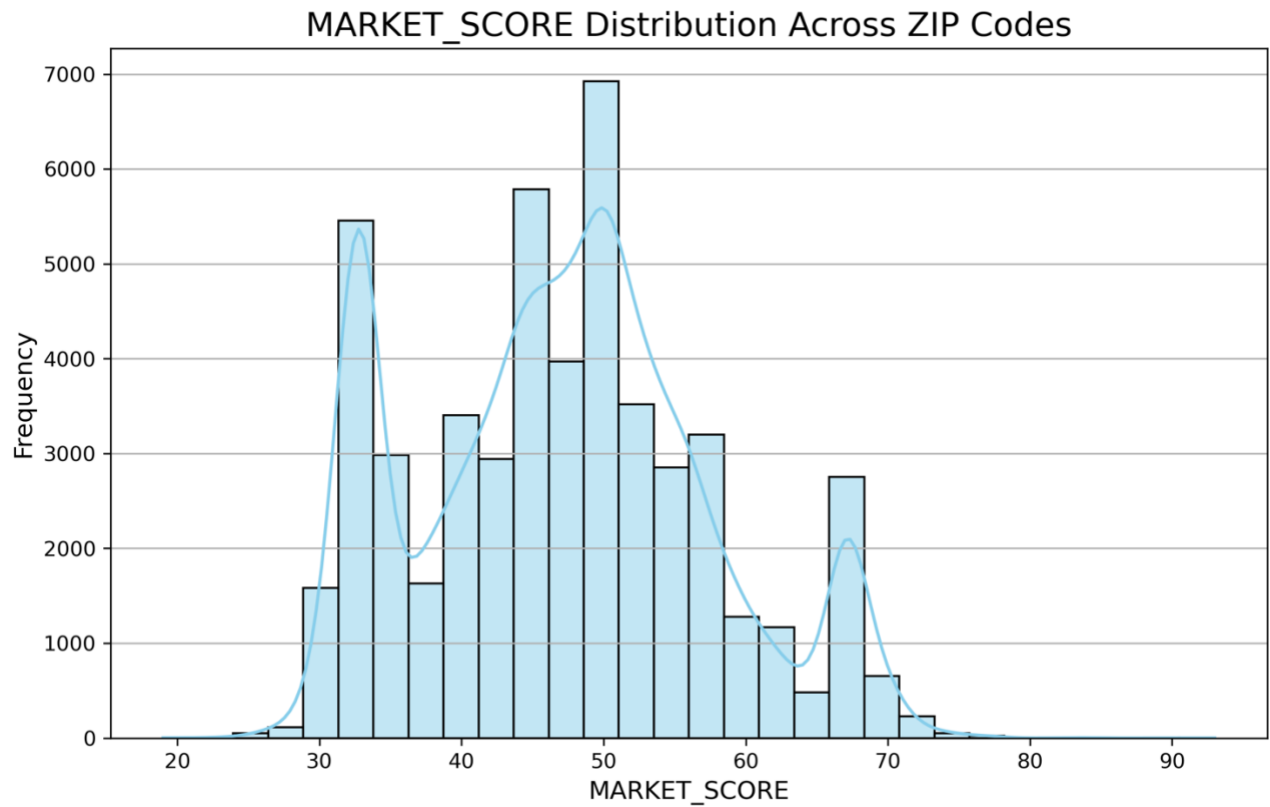
- This approach maximizes lending opportunities in areas with rapid price growth and heightened competition, leveraging market stability and demand.
- **Insights:**
  - The LTV adjustment framework protects lenders in less competitive markets by reducing risk exposure while maximizing loan volume in competitive markets through higher LTVs.



### 4.3 Visual Reports

- **Market Score Distribution:**

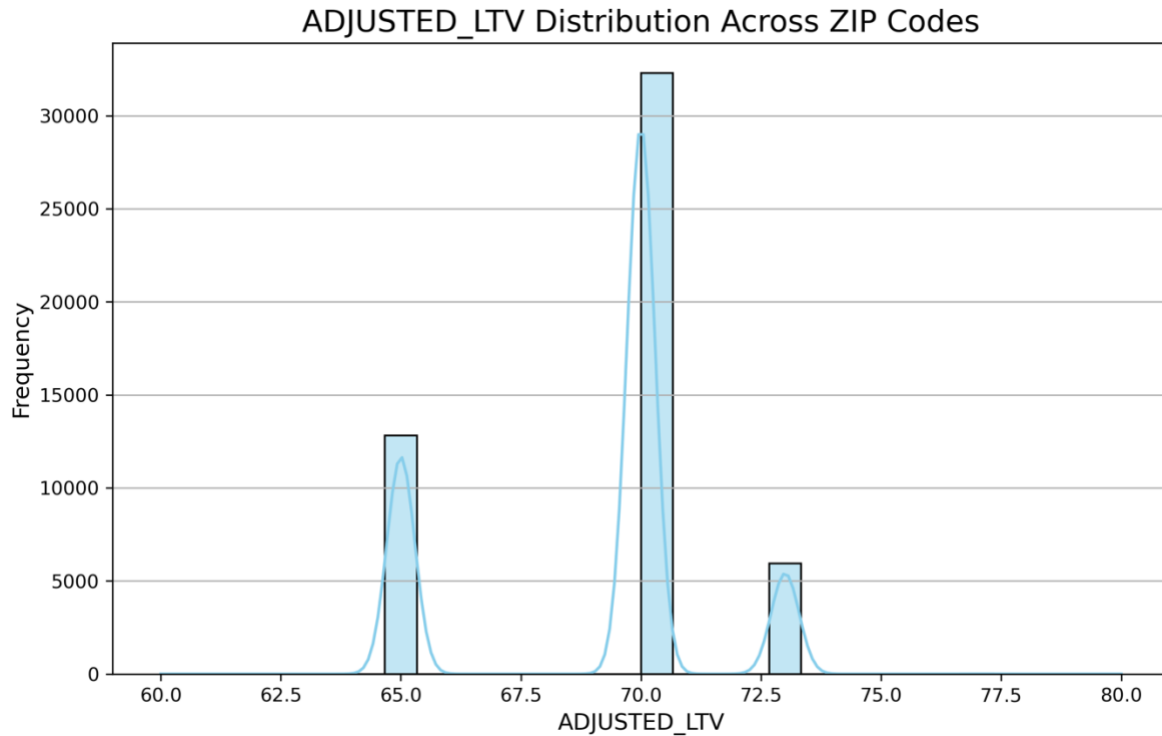
- Most ZIP codes have Market Scores between 40 and 60, with fewer outliers above 70.



- **Adjusted LTV Distribution:**

- LTV values cluster around 65–70%, with competitive markets experiencing values closer to 70%.





## 5. Implementation Details

### 5.1 Preprocessing

- The dataset was cleaned and filtered to focus on California real estate metrics for 2024.
- Missing values were addressed:
  - Continuous variables were filled with their median values.
  - Discrete variables were filled with zero to indicate no activity.

### 5.2 Rule-Based Scoring

- The scoring logic was implemented to calculate Market Scores based on the weighted feature contributions.
- Normalization ensured all features were scaled consistently, and `MEDIAN_DOM` was inverted for interpretability.

### 5.3 LTV Adjustment

- LTV values were adjusted dynamically based on Market Scores:
  - Competitive markets (high Market Scores) received higher LTVs to maximize loan volume while leveraging market stability.

- Less competitive markets received lower LTVs to mitigate risk.

## 5.4 Geospatial Mapping

- Geospatial maps were created to visualize Market Scores and LTV adjustments across ZIP codes in California.
- Simplified geometries improved map rendering performance, ensuring usability for stakeholders.

# 6. Next Steps and Recommendations

## 6.1 Enhancements

1. **Expand Feature Set:**
  - Include additional metrics such as rental yields, price appreciation rates, and affordability indices.
2. **Dynamic Weight Adjustment:**
  - Incorporate a machine learning component to dynamically adjust feature weights based on market trends.

## 6.2 Operational Deployment

1. **Integration with Loan Underwriting:**
  - Deploy the scoring and LTV adjustment framework into loan underwriting systems for automated decision-making.
2. **Real-Time Updates:**
  - Implement a pipeline for real-time data ingestion and score recalculation to reflect changing market conditions.

## 6.3 Future Research

1. **Predictive Modeling:**
  - Leverage predictive models to forecast Market Scores and LTV adjustments based on historical trends.
2. **Regional Analysis:**
  - Explore regional variations in Market Scores to identify emerging trends and opportunities.

## **7. Conclusion**

The rule-based scoring framework provides a robust, interpretable, and actionable approach to analyzing real estate markets at the ZIP-code level. By combining statistical rigor with domain knowledge, the framework delivers insights that are both practical and transparent. The interactive geospatial maps and visual reports enable stakeholders to make informed decisions, ensuring balanced risk and opportunity in lending practices.

This project lays the foundation for ongoing improvements and operational deployment, with the flexibility to adapt to evolving market conditions.

## References

Redfin, "Downloadable Housing Market Data," Redfin Data Center, accessed Jul 14 2025, <https://www.redfin.com/news/data-center/>.