



# Golden State Housing Insights

Welcome to our predictive analysis of California's housing market. This presentation provides valuable insights for real estate stakeholders, using data-driven methods to guide investment decisions in California Metro Areas.



# Executive Summary

## 1 Housing Prices and ROI Analysis

Examining the current trends and potential returns on investment in the California housing market.

## 2 Housing Feature Analysis

Identifying key features that influence housing prices and their impact on market value.

## 3 Housing Interest Rate Analysis

Analyzing the effects of interest rates on housing prices and market dynamics.

## 4 Team Goal

What are the key interrelationships between housing price as dictated by home features and interest rates as well as housing price for California metro area returns over time?

# Team Approach



## Strategic Task Allocation

Assigned tasks based on individual strengths to optimize team contributions.



## Organized Workflow

Utilized separate notebooks for efficient organization and collaboration.



## Detailed Analysis

Examined data-analysis techniques and modeling thoroughly, fostering diverse perspectives.

# Housing Price Prediction with Zillow Data Analysis

## Data Pre-Processing

- Utilized NumPy and Pandas for data manipulation
- Used Matplotlib for visualization
- Applied data melting for reshaping datasets
- Excluded metro areas with high p-values
- Incorporated time series analysis for temporal trends

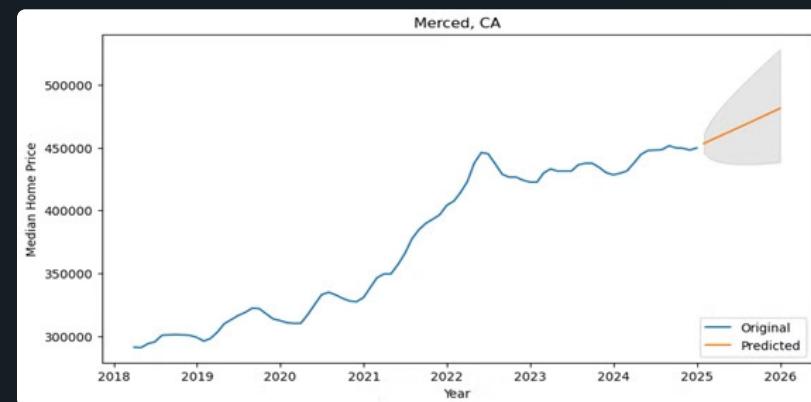
```
ARIMA(0, 0, 0) RMSE=77082.425
ARIMA(0, 0, 1) RMSE=39108.126
ARIMA(0, 0, 2) RMSE=23543.552
ARIMA(0, 1, 0) RMSE=3002.889
ARIMA(0, 1, 1) RMSE=3795.657
ARIMA(0, 1, 2) RMSE=2972.305
ARIMA(0, 2, 0) RMSE=3254.428
ARIMA(0, 2, 1) RMSE=3252.892
ARIMA(0, 2, 2) RMSE=3261.778
ARIMA(1, 0, 0) RMSE=3052.797
ARIMA(1, 0, 1) RMSE=2767.805
ARIMA(1, 0, 2) RMSE=2432.907
ARIMA(1, 1, 0) RMSE=2979.649
ARIMA(1, 1, 1) RMSE=3183.208
ARIMA(1, 1, 2) RMSE=3251.832
ARIMA(1, 2, 0) RMSE=3253.021
```

## Model Training

- We used ARIMA and SARIMA models for housing price prediction. Comparing both models improved prediction accuracy and robustness.

## Results

- ARIMA model: 90% ROI accuracy; 10% off observed values
- SARIMA model: 85% ROI accuracy; 15% off observed values



# Housing Feature Analysis

1

## Data Pre-Processing

Removed outliers from key features, used heatmaps to identify impactful variables, and split data: 80% training, 20% testing. Standardized values with StandardScaler.

2

## Model Training

Trained a Linear Regression model and made predictions, showing general alignment with actual prices but indicating areas for improvement.

3

## Feature Importance

Used Random Forest and recursive feature elimination, evaluating performance with RMSE, R<sup>2</sup>, and MAE metrics to understand home price factors.

4

## Results

The top five factors influencing house prices are Median Income, House Age, Average Number of Rooms, exterior quality, and Garage size.



	year_month	Rate
0	2017-01	0.649677
1	2017-02	0.656786
2	2017-03	0.786129
3	2017-04	0.896000
4	2017-05	0.907419
5	2017-06	1.040000
6	2017-07	1.150645
7	2017-08	1.157097
8	2017-09	1.153333
9	2017-10	1.153871
10	2017-11	1.156897

# Interest Rate Prediction Analysis

## Data Pre-Processing

Utilized Pandas for data manipulation, removed nulls, and grouped data. Used Matplotlib for visualization.

## Model Training

Utilized Linear Regression and KNN models for feature analysis. The model was trained using the `model.fit(X_train, y_train)` method, allowing it to learn from the training data.

## Results

The model predicts that interest rates will rise through February 2025, providing valuable insights for investment decisions.

# Future Development



By leveraging a larger dataset, we aim to enhance the reliability and precision of our investment recommendations. This will allow us to provide more accurate and valuable insights for real estate stakeholders in California's dynamic housing market.

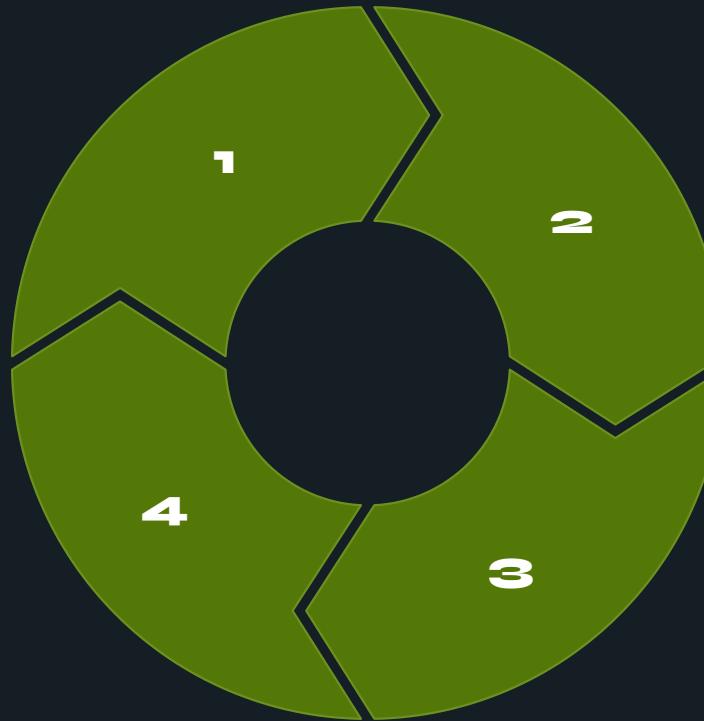
# Conclusion

## Valuable Insights

Uncovered investment opportunities

## Informed Decisions

Guide economic policies



## Interconnected Factors

ROI, features, and interest rates

## Continuous Optimization

Need for model refinement

Our housing market analysis highlights the interconnected effects of ROI, housing features, and interest rates on California's real estate prices, emphasizing the need for continuous model optimization to guide policies and investments.

# Project Contributors



**Chris Gilbert**



**Will Atwater**



**Dexter Johnson**



**Joel Freeman**



*Paper Not Included*

**Rod Burroughs**



**Jacinto  
Cepeda Quiroz**

# Demonstrations

1

## StreamLit Demonstration

Interactive showcase of predictive models

2

## Visualizations

Dynamic display of project findings

3

## Results Presentation

Comprehensive overview of project outcomes