

# Time Series Analysis Zillow Data

By Group 5



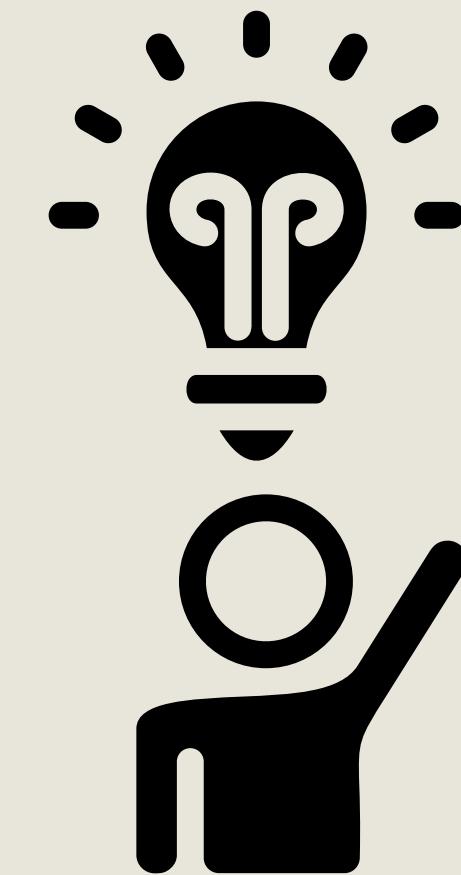
# Overview

This project aims to analyze historical housing data to identify trends and patterns that can inform investment decisions. By developing a predictive model, we aim to forecast future returns on investment (ROI) for different zip codes, providing valuable insights for prospective investors.



# Business Understanding

In the competitive world of real estate investment, making informed decisions is crucial for maximizing returns and minimizing risks. The real estate market is influenced by numerous factors, including economic conditions, demographic trends, and local market dynamics. The project aims to analyze historical housing data to identify trends and patterns that can inform investment decisions. By developing a predictive model, we aim to forecast future returns on investment (ROI) for different zip codes, providing valuable insights for prospective investors.



# Problem Statement

**The real estate market is experiencing unpredictable fluctuations in property prices. This uncertainty makes it difficult for prospective investors to make informed decisions about investing in property. The company needs to find an informed way to decide on which properties to invest in.**



# Objectives



01

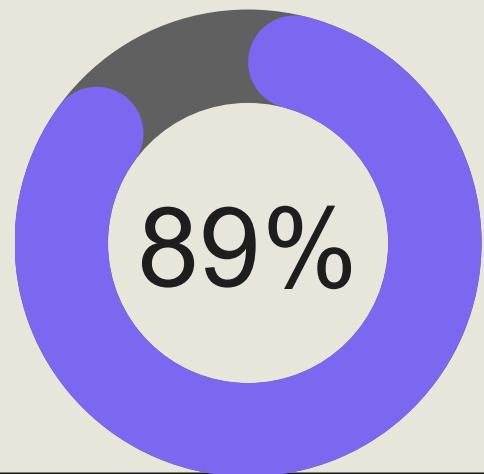
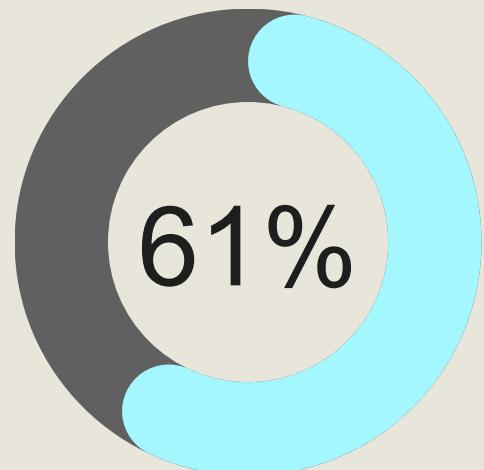
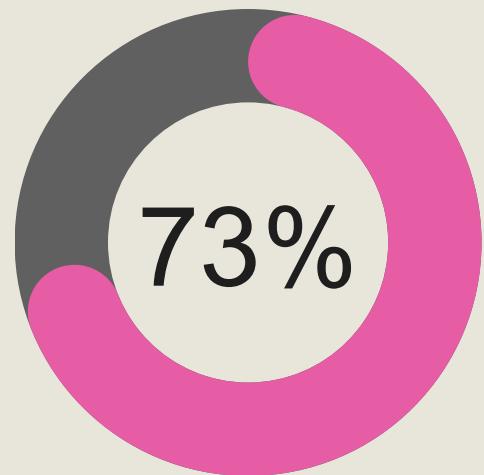
To analyze the zillow housing dataset, identify key patterns, trends and relationships in the data and yield visualizations that will aid in data-driven decisions.

02

To develop a predictive timeseries model that will forecast return on investment(ROI) of the various Zip codes.

# Data Understanding

- This study makes use of the Zillow House data csv.
- It contains 14723 entries and 272 features.
- It consists of various columns including region information and monthly real estate prices from April 1996 to April 2018.



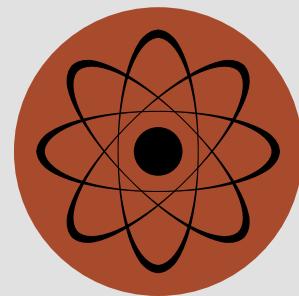
# Data Preparation

Steps include:

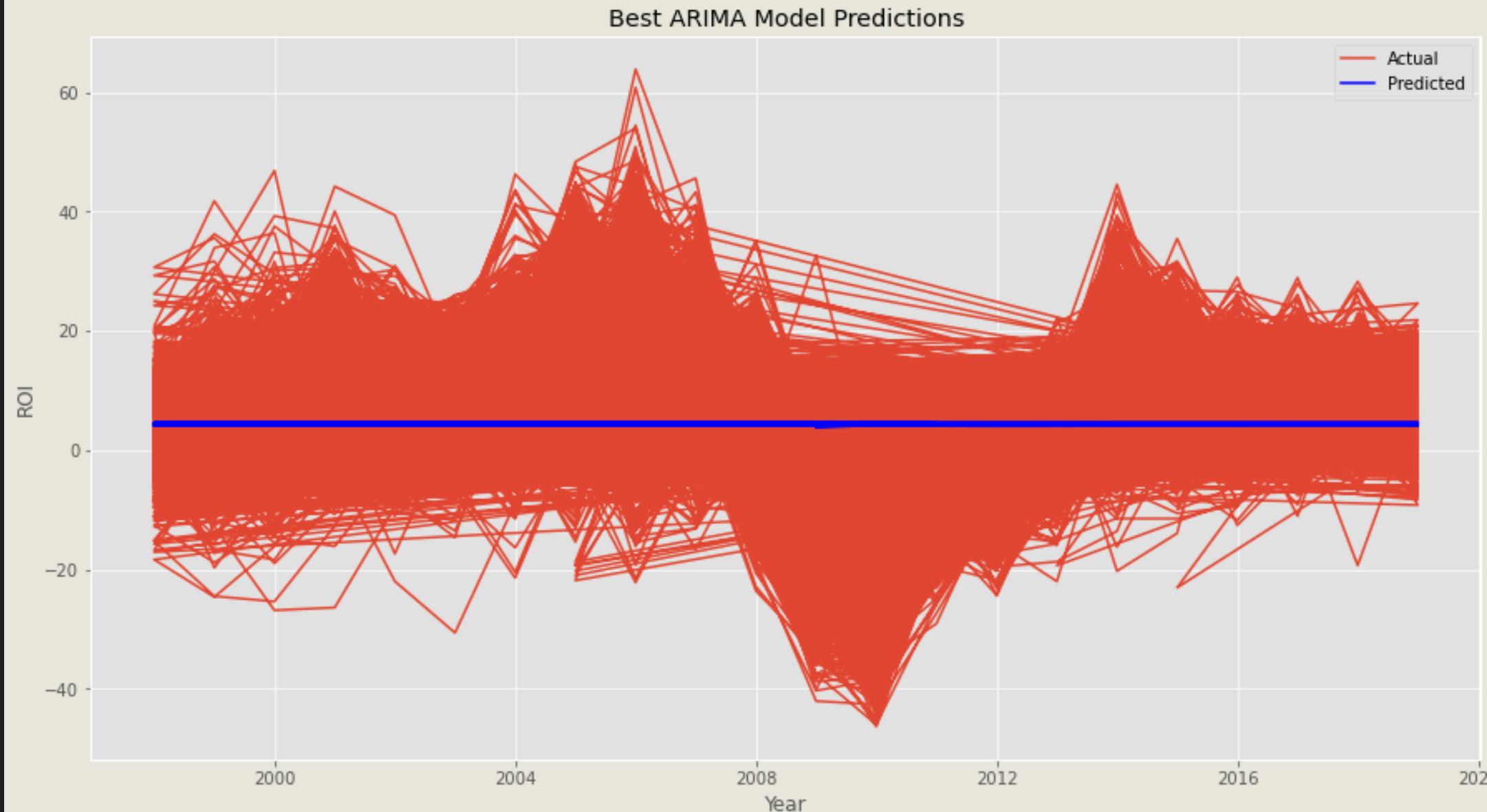
- Data Cleaning
- Exploratory Data Analysis



# Modeling



# Baseline Model



**Based on the results of the ARIMA(2,1,0) model, we can conclude that it is the best fit for the data. The lower AIC and BIC values, along with the statistically significant AR terms, indicate that the ARIMA(2,1,0) model captures the underlying patterns in the data more effectively and makes accurate forecasts.**

# Conclusion

- California dominates the market, followed by New York and Texas.
- Historical data reveals a general upward trend in real estate prices with a notable peak around 2006, followed by a decline and subsequent recovery.
- Return on Investment (ROI) varies significantly across cities and timeframes, emphasizing the complex nature of real estate investments. The ARIMA(2,1,0) model demonstrated superior performance compared to the baseline ARIMA(1,0,0) model, effectively capturing the underlying patterns in the data.



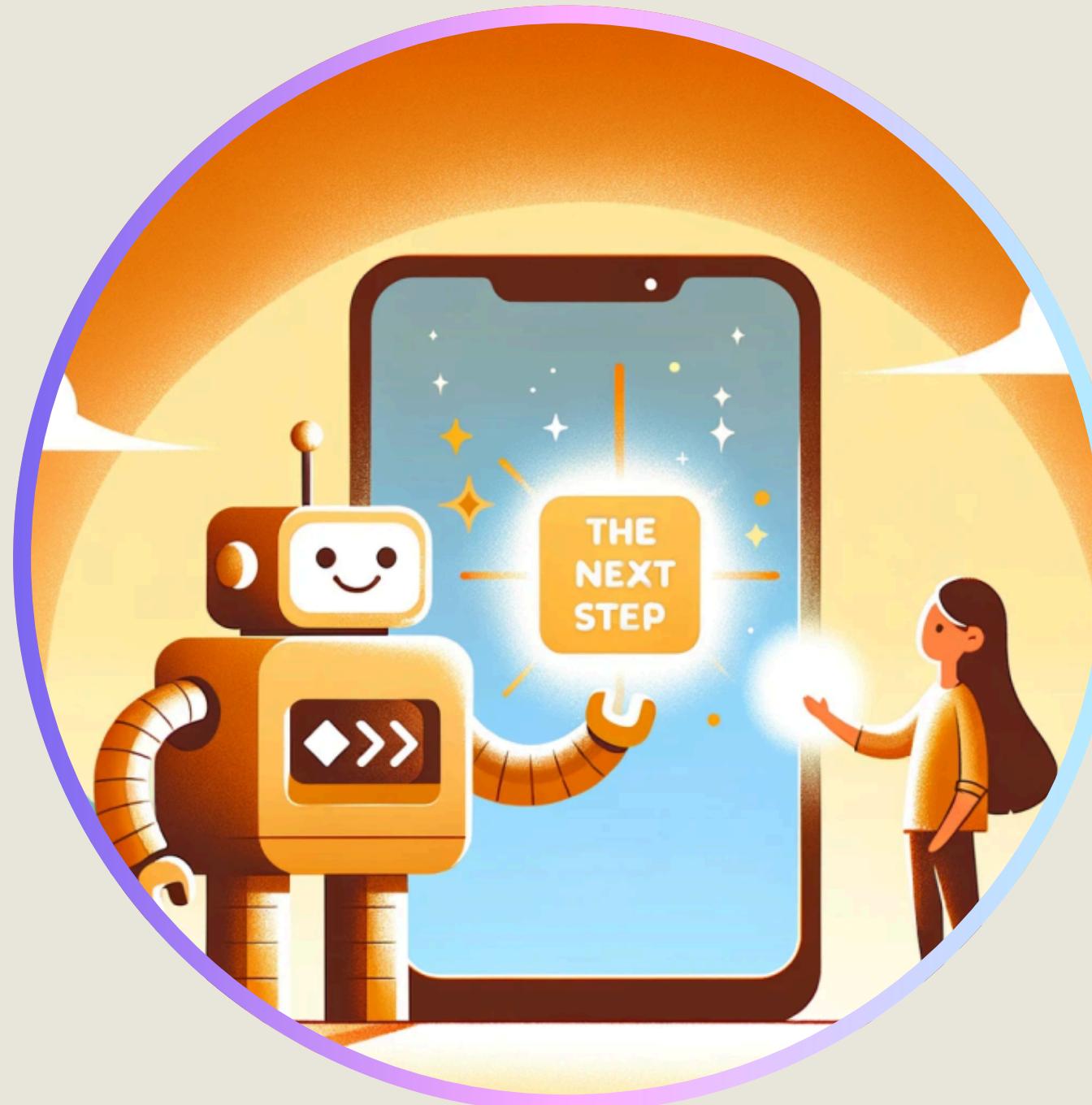
# Recommendations



- **Target High-Growth Regions.**
- **Diversify Investment Portfolio.**
- **Long-Term Investment Horizon.**
- **Continuous Monitoring.**
- **Risk Assessment.**

# Next Steps

- Incorporate External Factors
- Advanced Modeling Techniques:
- Risk Modeling:
- Portfolio Optimization:
- Real-Time Data Integration



Thank you



 **Zillow**<sup>®</sup>