Forecasting Optimal Real Estate Investments Using Time Series Modeling

By Sam Odongo

Business Problem

Investing in real estate in the US is a complex journey.14723 unique Zip Codes available

The client/investor is looking for the BEST 5 Zip Codes to invest in.

The aim of this project is to identify the top 5 zip codes for investment by combining data-driven modeling, risk assessment, and a transparent decision-making process.

The goal: empower informed, confident real estate investment decisions in a dynamic market.

Main Objective

Recommend the top 5 best zip codes for investment within the real estate market

Analyze Real Estate Market Data, Implement Time Series Models, and Provide Invaluable Insights for Real Estate Investors, Sellers, and Industry Experts

Specific Objectives

Implement Time Series Models for Real Estate Market Analysis.

Identify Crucial Factors Affecting Property Prices in the Chosen Market.

What's "best"? Refine data and build a robust real estate investment model using evaluation criteria, including historical price appreciation, SizeRank, Coefficient of Variation (CV), and Return on Investment (ROI), to assess long-term trends, market significance, volatility, and profitability of ZIP codes for optimal investment selection..

Data

The dataset, sourced from [Housing Data - Zillow Research](#) , consists of 14723 rows and 272 columns, providing comprehensive information on house sales and property values from April 1996 to April 2018.

ZIP Code selection will be conducted to identify relevant variables among the 14723 rows, focusing on key factors like house prices, location (region, city, state), and property size rank, aligning with project objectives.

Data limitation;Absence of external factors (e.g., economic indicators, demographic trends) impacting real estate markets. Additionally, the dataset's size and complexity may present challenges in data preprocessing and modeling, which need to be addressed for project success.

Data Preparation

Proactively addressed data preprocessing to enhance time series modeling.

Utilized robust statistical methods to detect and manage outliers effectively.

Employed tailored strategies to handle missing data.

Conducted Zip Code relevance evaluation and removed non-relevant Zip Codes

Ensured dataset quality to improve modeling outcomes.

10 Zip Codes selected

Definition of "BEST"

Three key metrics were chosen to define "best":

Historical Price Appreciation: Provides insights into ZIP code performance over time, aiding in long-term trend analysis, risk assessment, and profit potential.

SizeRank: Measures ZIP code popularity and size relative to others, aiding in assessing market significance, diversification opportunities, and growth potential.

Coefficient of Variation (CV): Assesses data volatility to aid in risk assessment and investment strategy, ensuring portfolio diversity.

Return on Investment (ROI): Quantifies investment profitability by evaluating median housing price changes from April 1996 to April 2018, enabling us to rank and select ZIP codes for optimal returns in our real estate investment analysis.

Why Time Series Model?

Why Time Series Model?

Temporal Insights: Captures time-dependent factors shaping real estate markets.

Seasonality & Trends: Identifies and accounts for recurring patterns and long-term trends.

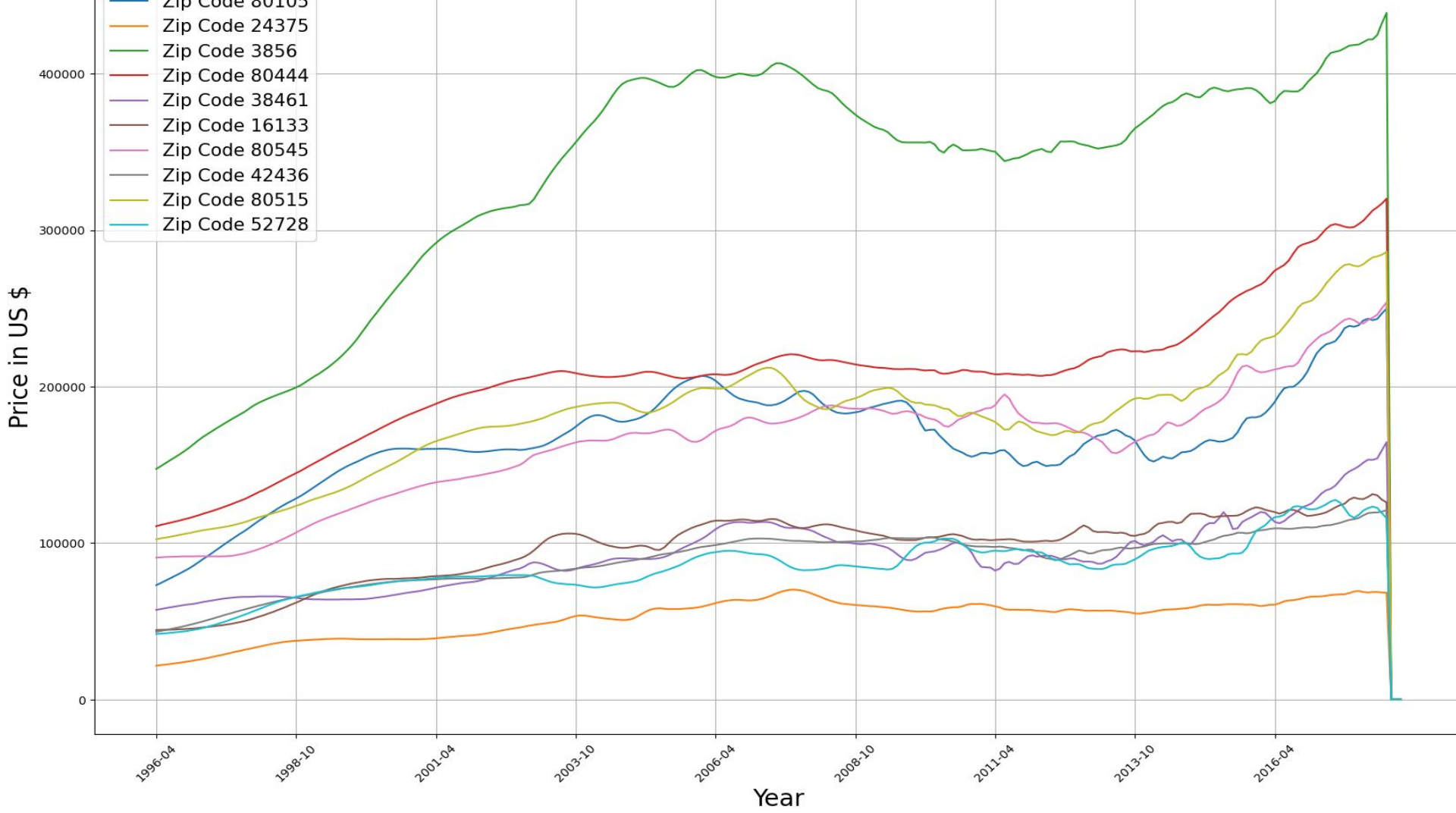**Historical Data**: Effectively utilizes historical price data for informed decisions.

Future Forecasting: Predicts future values crucial for spotting investment opportunities.

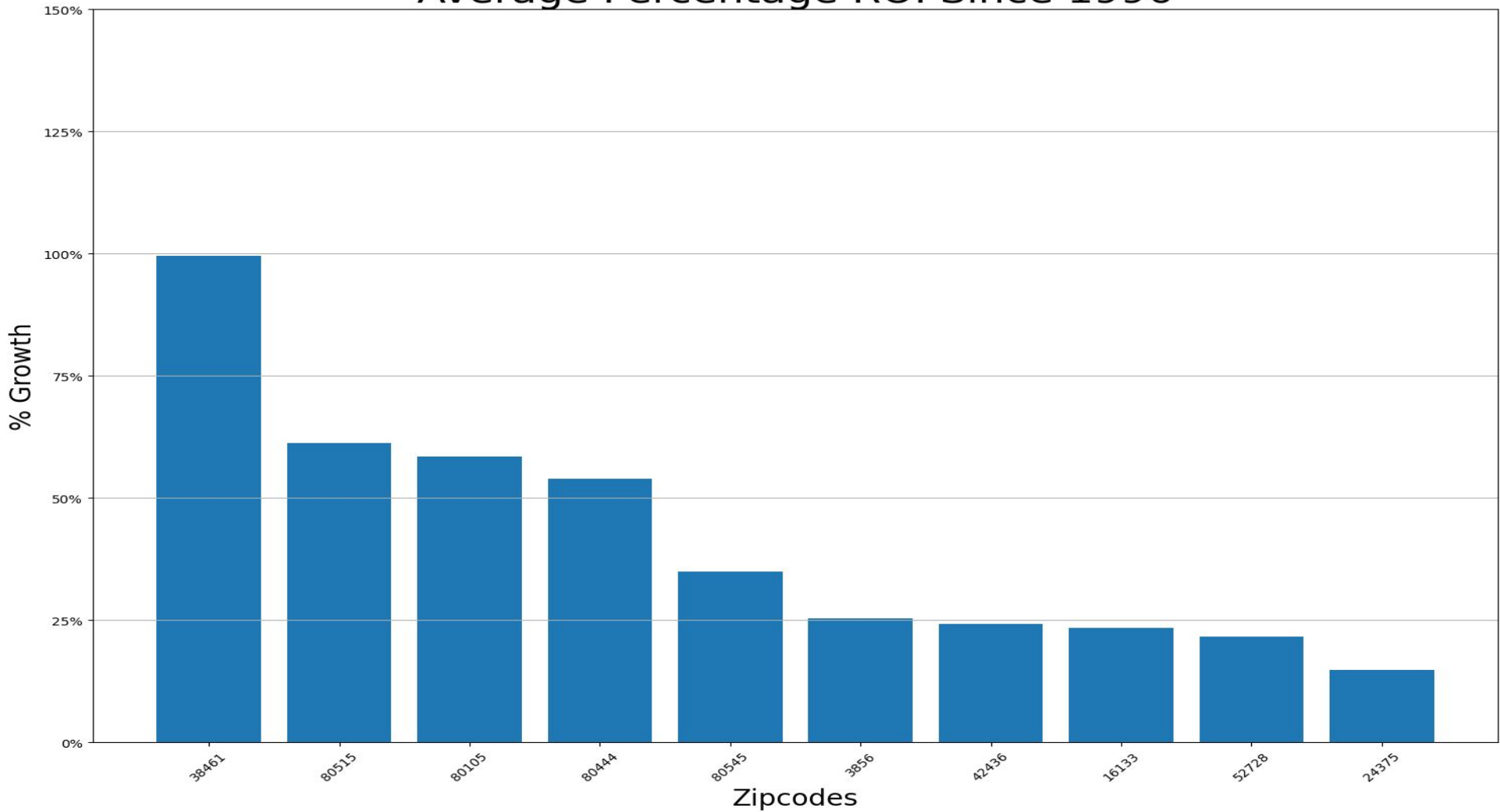**Multiple Variables**: Integrates diverse factors, improving prediction accuracy.

Continuous Monitoring: Offers ongoing market forecasting for adaptive strategies.

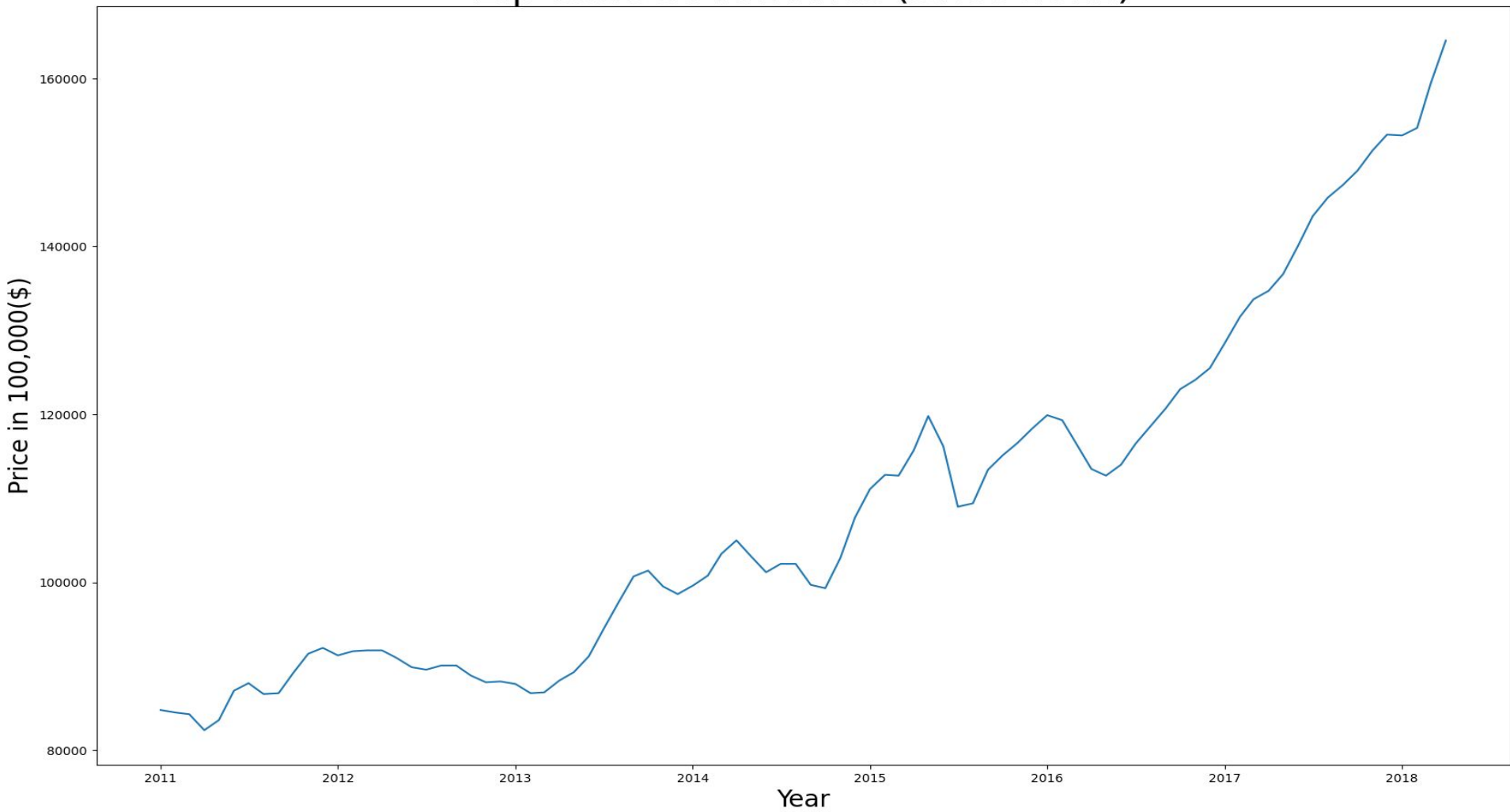# Analysis

Average Percentage ROI Since 1996

Modelling

Evaluation

Its observed that the market conditions before the 2010s do not necessarily mirror the current market situation. Hence, I have decided to focus the Time Series modeling solely on data from 2011 onwards.

ZIP Code 38461 selected for modelling;Has the highest ROI

Zip Code 38461 Price (since 2011)

Model Results

```
                         SARIMAX Results
==========================================================================
Dep. Variable:                       y   No. Observations:              64
Model:                SARIMAX(3, 0, 2)   Log Likelihood            -551.253
Date:                Sat, 16 Sep 2023   AIC                       1116.506
Time:                         03:48:58   BIC                       1131.618
Sample:                      01-01-2011   HQIC                      1122.460
                           - 04-01-2016
Covariance Type:                   opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
intercept   1269.1426   4623.002      0.275      0.784   -7791.774    1.03e+04
ar.L1          0.7509      0.187      4.014      0.000       0.384       1.117
ar.L2         -0.1198      0.206     -0.582      0.561      -0.523       0.284
ar.L3          0.3559      0.119      2.993      0.003       0.123       0.589
ma.L1          1.5123      0.135     11.223      0.000       1.248       1.776
ma.L2          0.8402      0.127      6.629      0.000       0.592       1.089
sigma2      1.535e+06     14.045   1.09e+05      0.000    1.53e+06    1.54e+06
==========================================================================
Ljung-Box (L1) (Q):                0.03   Jarque-Bera (JB):            5.01
Prob(Q):                           0.87   Prob(JB):                    0.08
Heteroskedasticity (H):            2.75   Skew:                       -0.20
Prob(H) (two-sided):               0.02   Kurtosis:                    4.31
==========================================================================
```
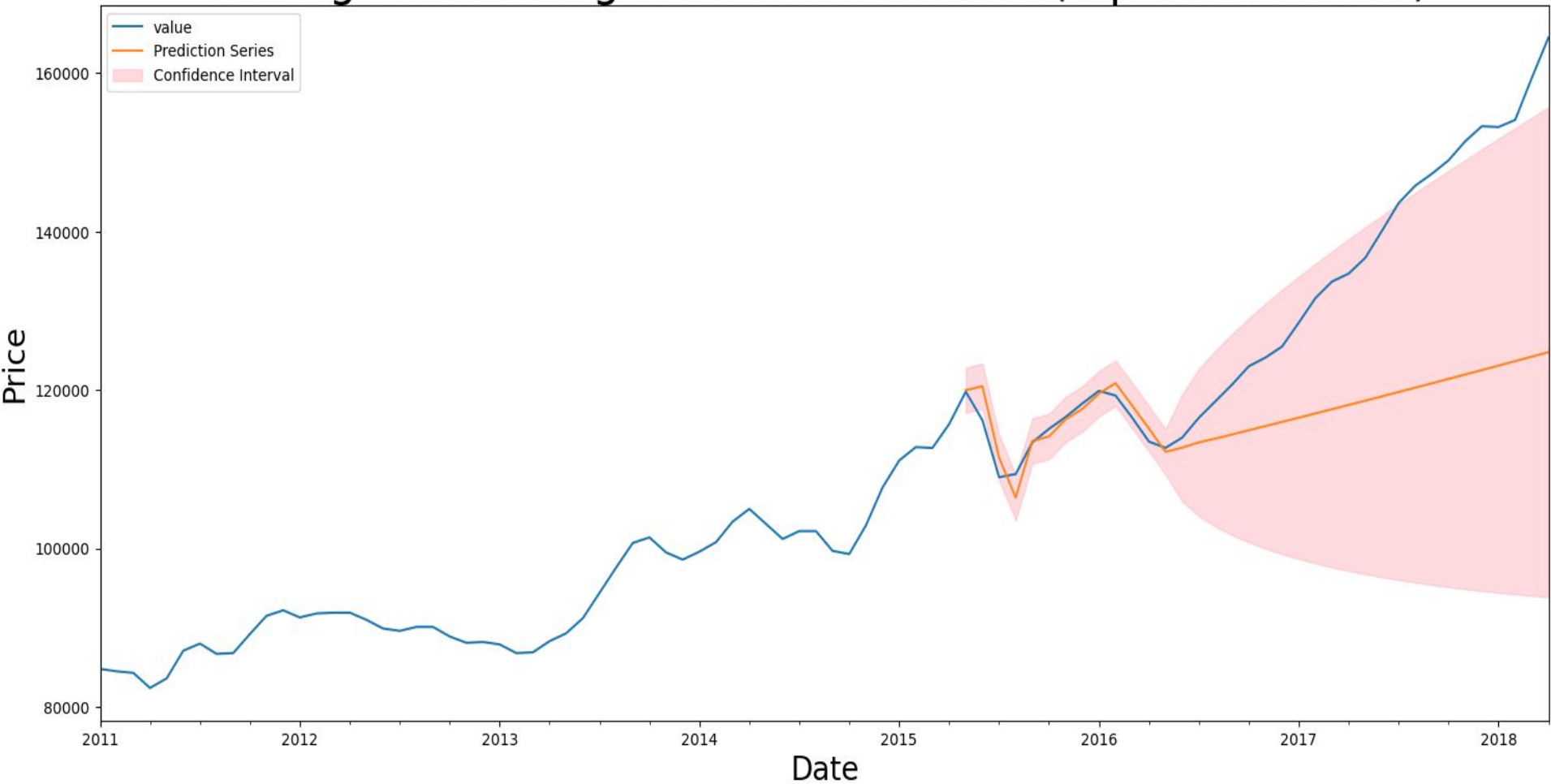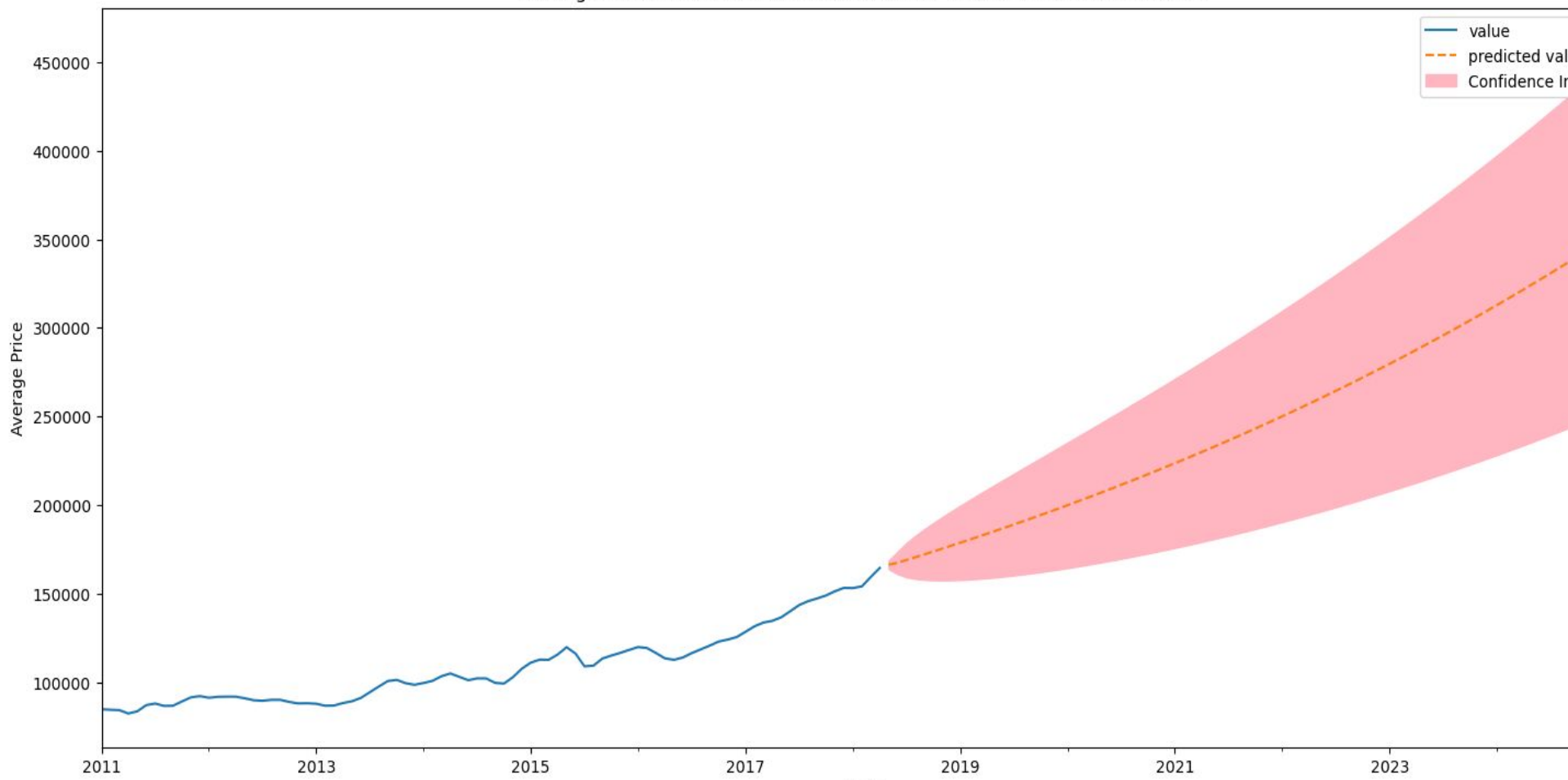
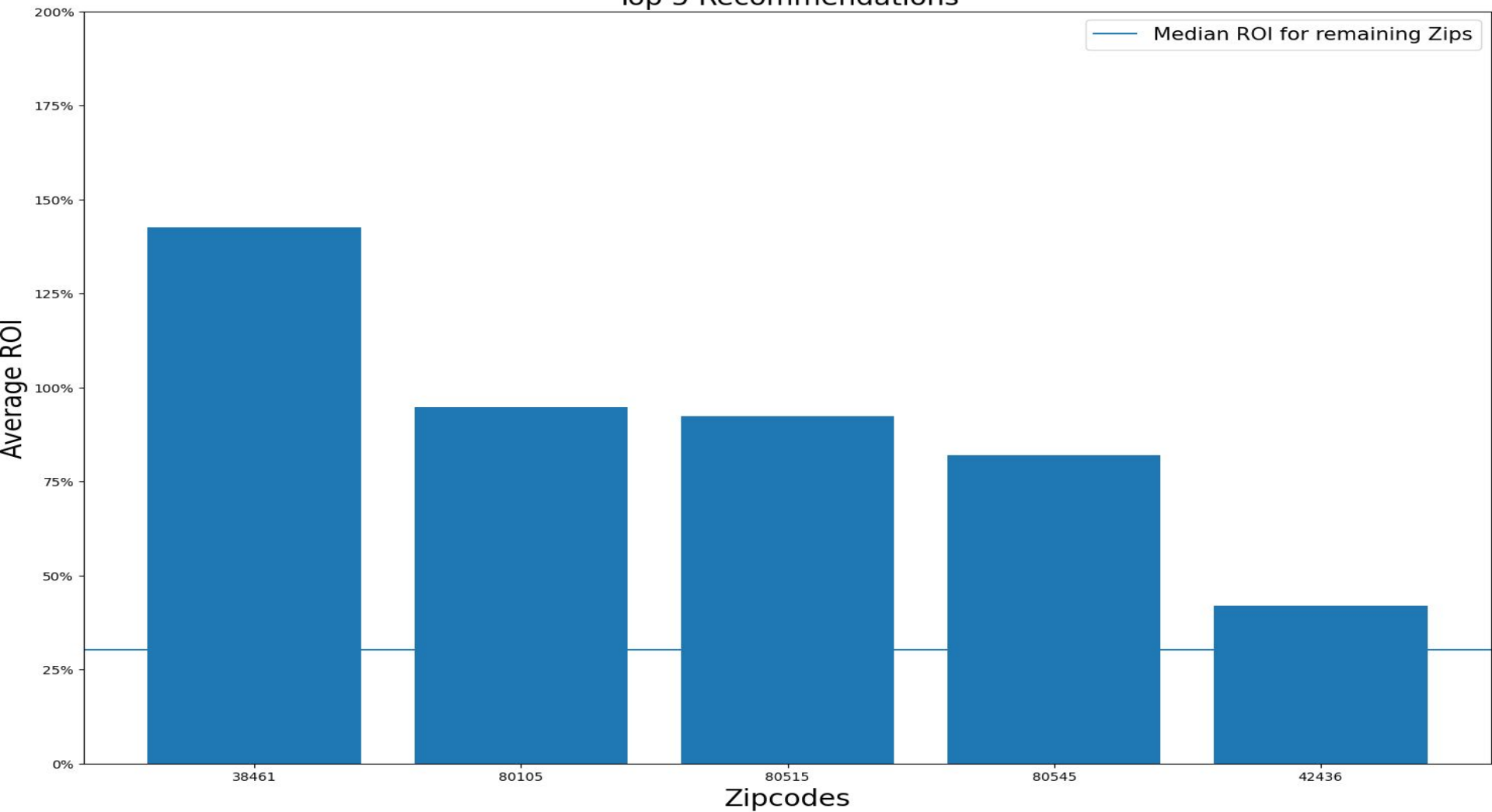Testing Forecasting Model Performance(Zip Code 38461)

Forecasting

Average Home Price - 38461 - With Forcasted Value & Confidence Intervals
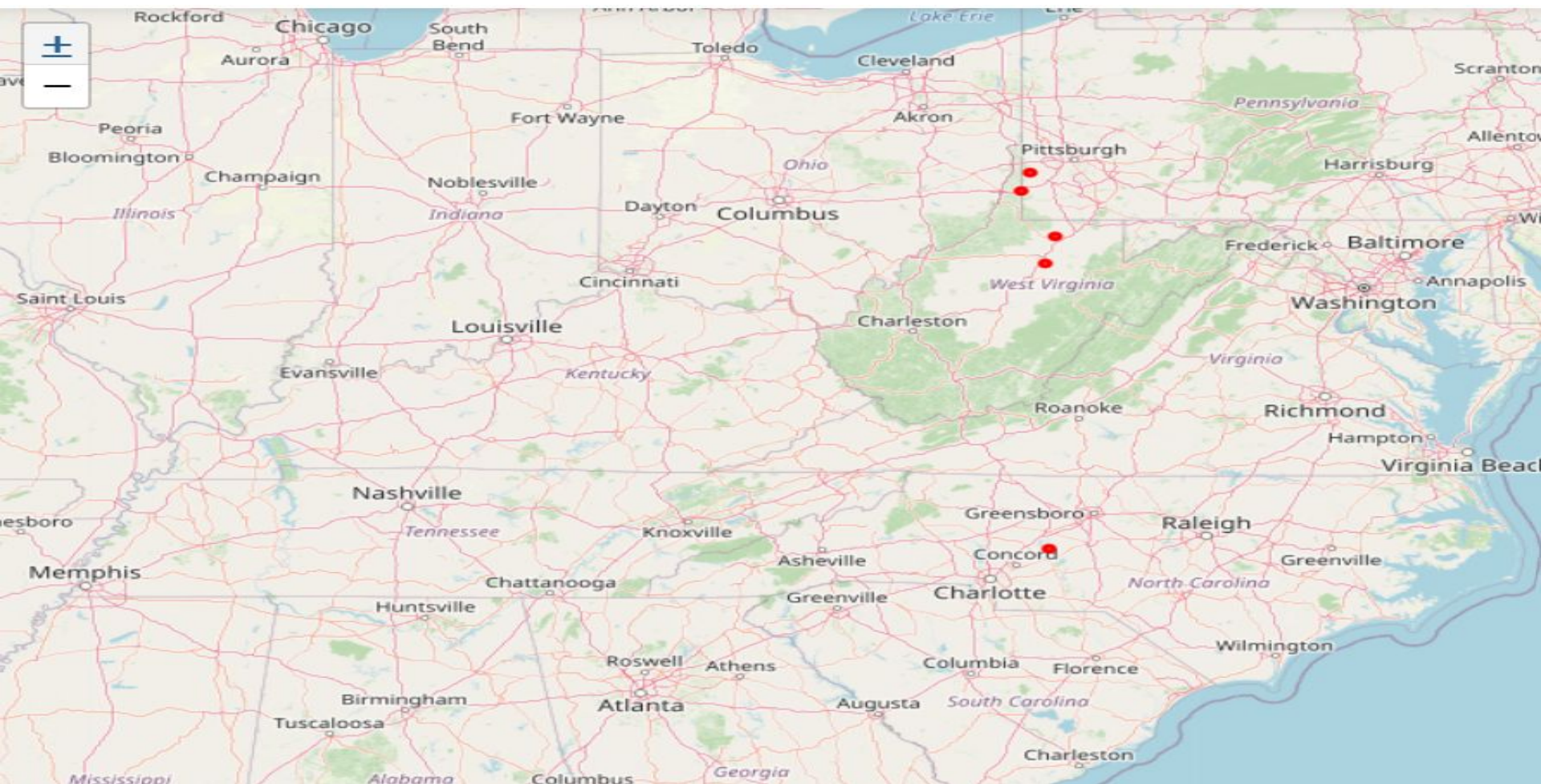
Recommendation

Top 5 Recommendations

hat are the top 5 best zip codes for us to invest in?

Zipcode	Expected Growth Rate in 2025
38461	142.62%
80105	94.74%
80515	92.4%
80545	82.08%
42436	41.96%

Location

By 2025

The Highest Growing Zipcode: 38461 is expected to grow by 142.62%

The Second Highest Growing Zipcode: 80105 is expected to grow by 94.74%

The Third Highest Growing Zipcode: 80515 is expected to grow by 92.4%

The Fourth Highest Growing Zipcode: 80545 is expected to grow by 82.08%

The Fifth Highest Growing Zipcode: 42436 is expected to grow by 41.96%

Other Factors

Financial forecasting relies on historical monthly returns but may not consistently predict future outcomes due to unforeseen events and changing conditions.

Investments and economic conditions are subject to uncertainties and risks beyond historical data, including policy changes, technological advancements, and global events.

To enhance accuracy, forecasting models should incorporate external factors such as economic indicators, demographic trends, and market sentiment.

Combining historical data with external factors improves forecast quality and equips decision-makers to navigate uncertainties in predicting future outcomes.