

WELCOME !

[Home](#)[About](#)[Service](#)[Listings](#)[Contact](#)

Zillow Real Estate Forecasting

Time Series Modelling Approach



Table Of Content



Search Here



Introduction



Problem Statement



Objectives



Data Understanding



Modelling



Evaluation



Recommendations



Conclusion



Introduction

- Zillow leverages time series models to analyze extensive historical real estate data, identifying trends and patterns for accurate market predictions.
- These analyses highlight prime investment opportunities and account for regional market variations, aiding buyers, sellers, and investors in making informed decisions

Business Understanding

Who cares?

- Real estate investment firm (wants to pick the best zip codes to invest in)
- Market analysts & advisors (help clients invest)
- Individual investors (make informed investment decisions)

Why is this important?

- Real estate can have a big impact on finances
- This project helps find the best areas to invest in using data
- It can:
 - Help make data-driven investment decisions
 - Find areas with the highest potential for growth (earn more money)
 - Understand market trends to manage risk and invest for the long term



"In real estate, you make money when you buy, not when you sell. You can't wait until the market is hot to start becoming a savvy investor." -Mark Cuban

MARK CUBAN - INVESTOR,
ENTREPRENEUR, OWNER

Objectives

[For Sale](#)[For Rent](#)[View All](#)

-  To develop time series models capable of accurately forecasting real estate prices for different zip codes.
-  To provide the investment firm with actionable recommendations regarding the top 5 zip codes for investment, supported by the rationale derived from the time series analysis.
-  To consider additional metrics beyond profit margins, including risk assessment, to ensure well-informed and balanced investment decisions.

Data Understanding

Zillow

Data Source

Contains 14,723 entries spanning over 20 years with monthly home values and location details making it ideal for time series analysis

22

Relevance

The extensive data (April 1996 - April 2018) empowers time series analysis to uncover trends and target promising investment areas.

Zip codes

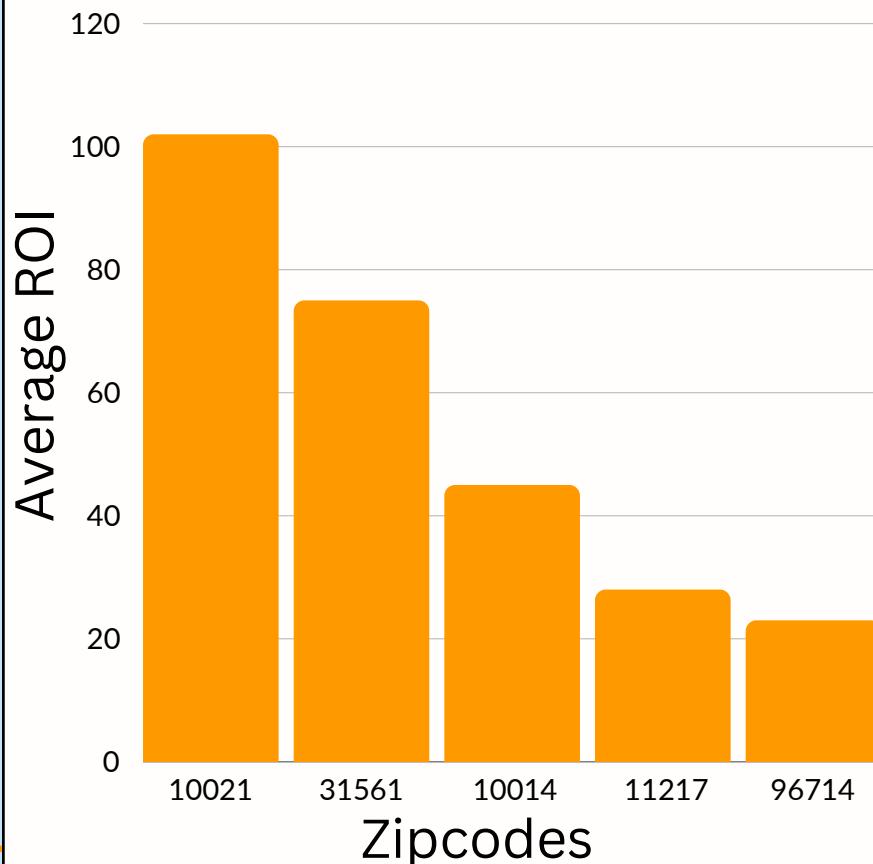
Features

It contains important data on real estate such as RegionID, RegionName (zip code), City, State, Metro, CountyName and monthly home values



Data Understanding

Zip codes with the highest ROI

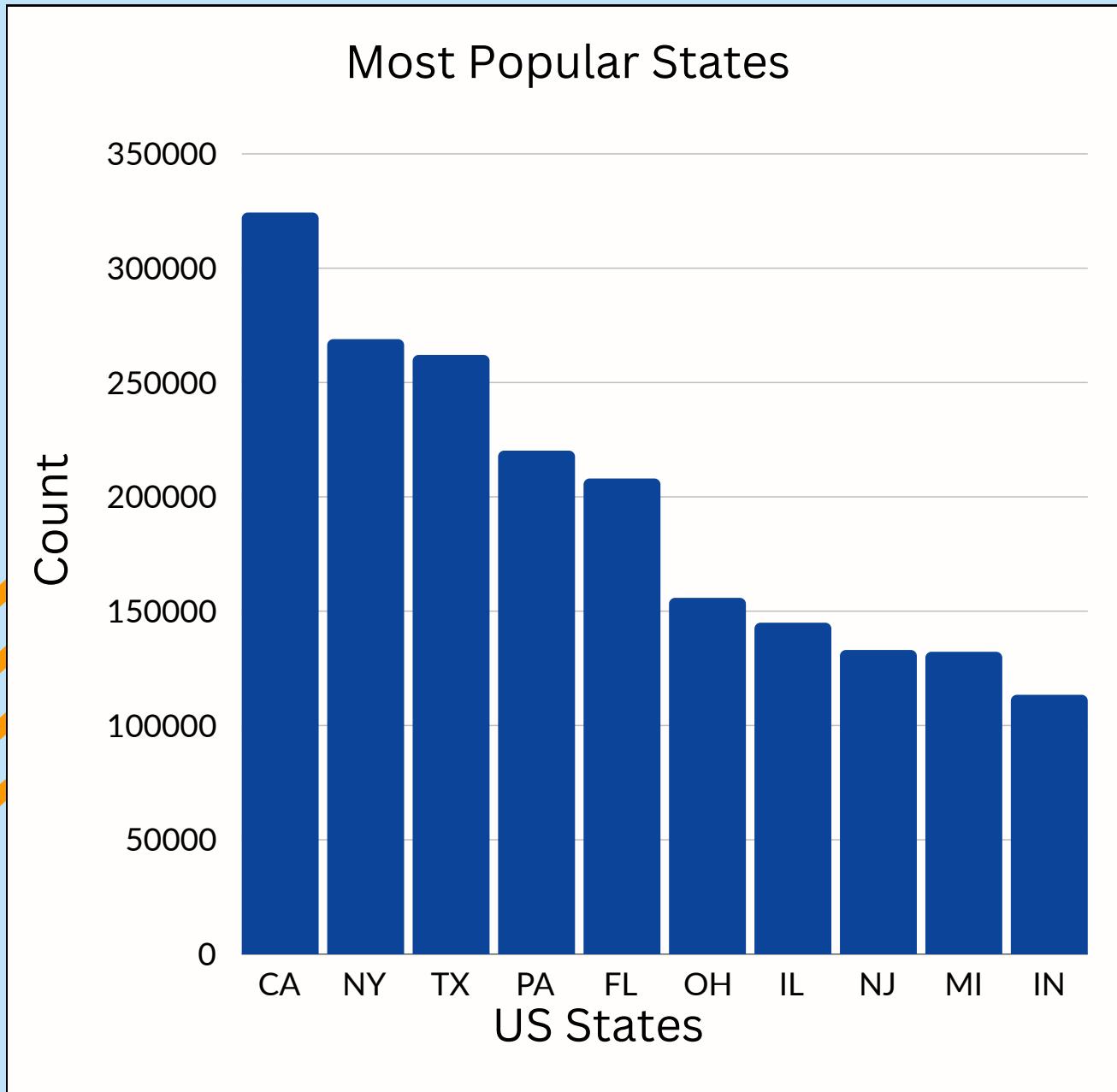


The zip code with the highest ROI:

- 10021

ROI stands for
Return on
Investment

Data Understanding

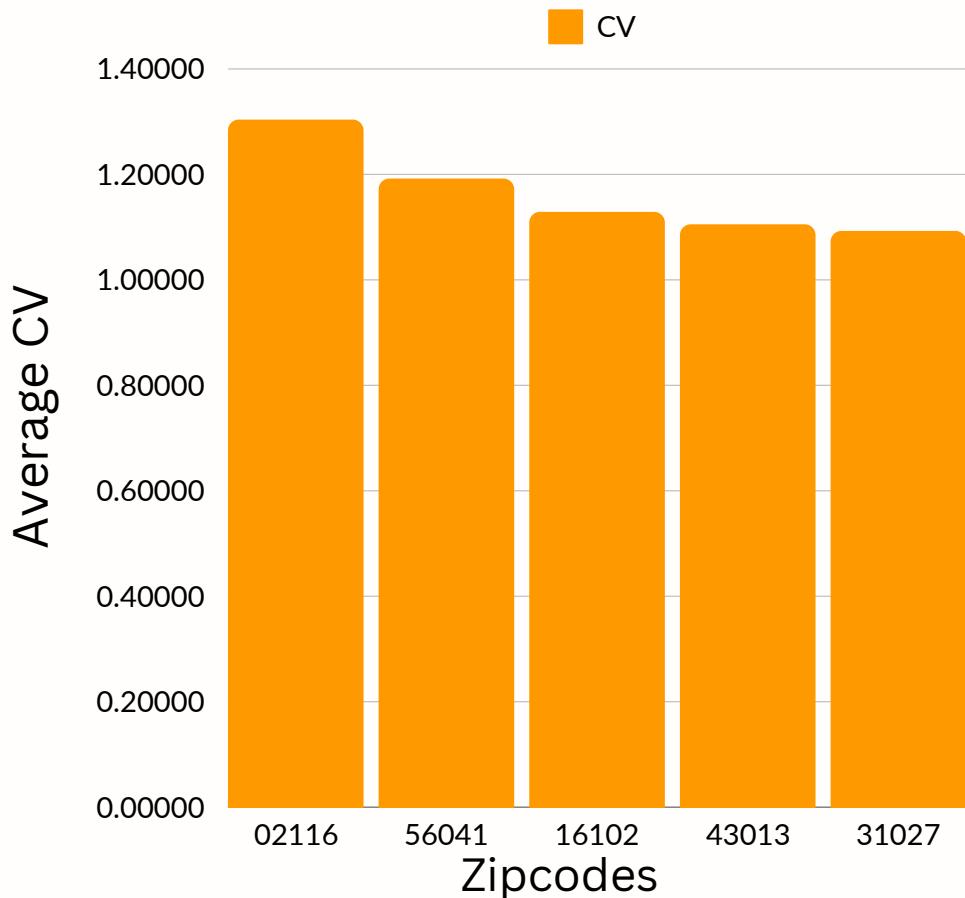


The most popular states are:

- California
- New York
- Texas
- Pennsylvania
- Florida

Data Understanding

Zip codes with the highest CV



The zip code with the highest CV:

- 02116

CV stands for
Coefficient Of
Variation



Modeling

1. Data Preprocessing

a. Seasonality

There was a straight line from 1996 to 2004 thus likely that data was missing and was filled through interpolation. The data was removed to avoid giving the wrong perception during modelling.

b. Stationarity

On checking for stationarity using dickey fuller test, the p_value was greater than 0.05, and the test_statistic was also greater than the critical values, therefore confirming that the time series is not stationary.

c. Detrending

Thus one lag differencing was done to remove the trend and the 12 lag for the seasonality in order to make the series stationary.



Modeling

2. Time Series Modeling

- SARIMA (Seasonal Autoregressive Integrated Moving Average) was used for modeling.
- SARIMA is suitable for time series data exhibiting seasonality and trends.
- It incorporates autoregressive terms (past values affecting current value), moving average terms (past errors affecting current value), and seasonal components to capture these patterns.

Evaluation

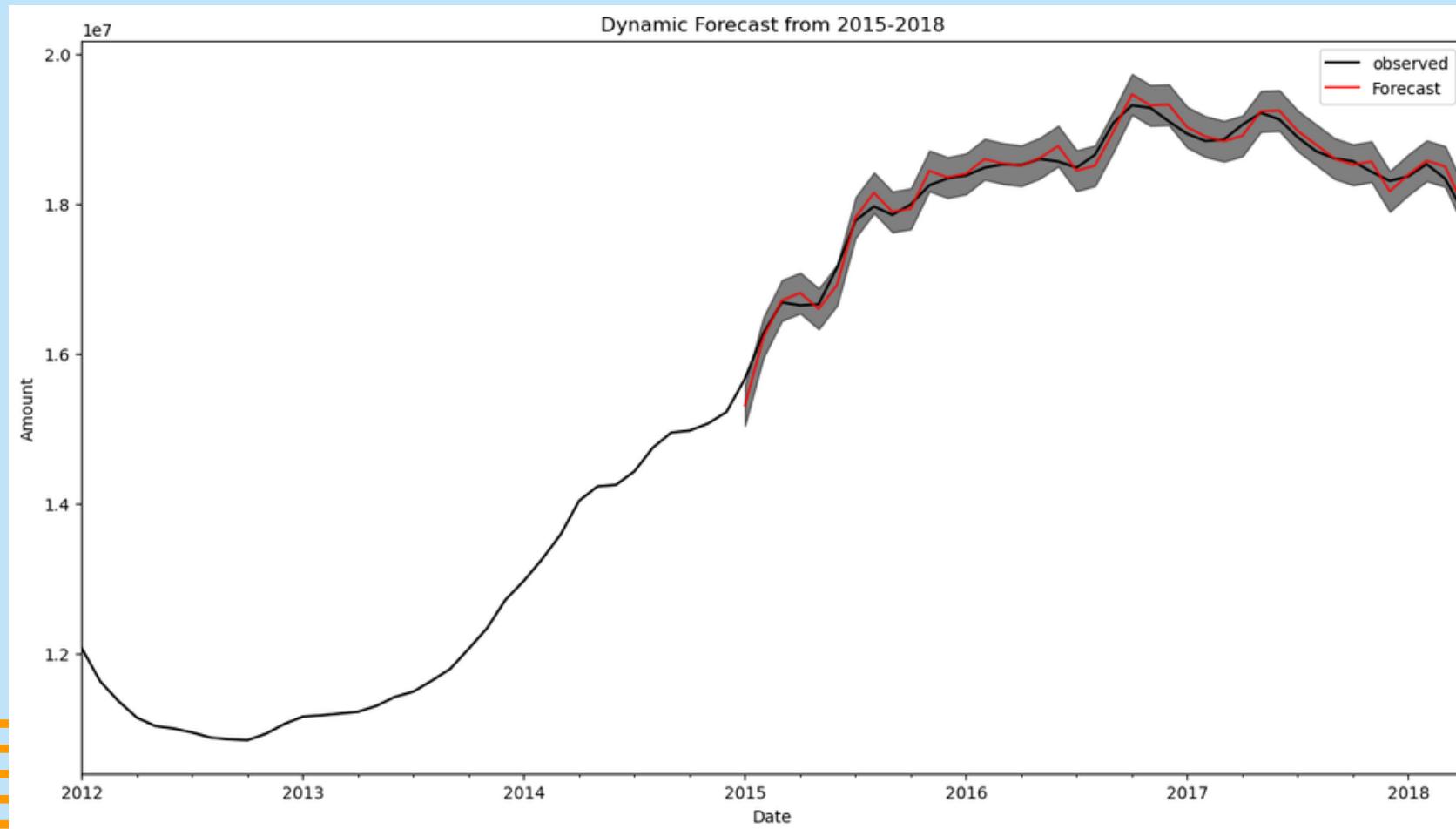
Predicted values were compared to real values of the time series this helps in understanding the accuracy of our forecasts from modeling



Evaluation

a) Dynamic Forecast

- Uses in-sample prediction. The model sequentially predicts one-step-ahead using the true value from previous time step instead of using predicted value.

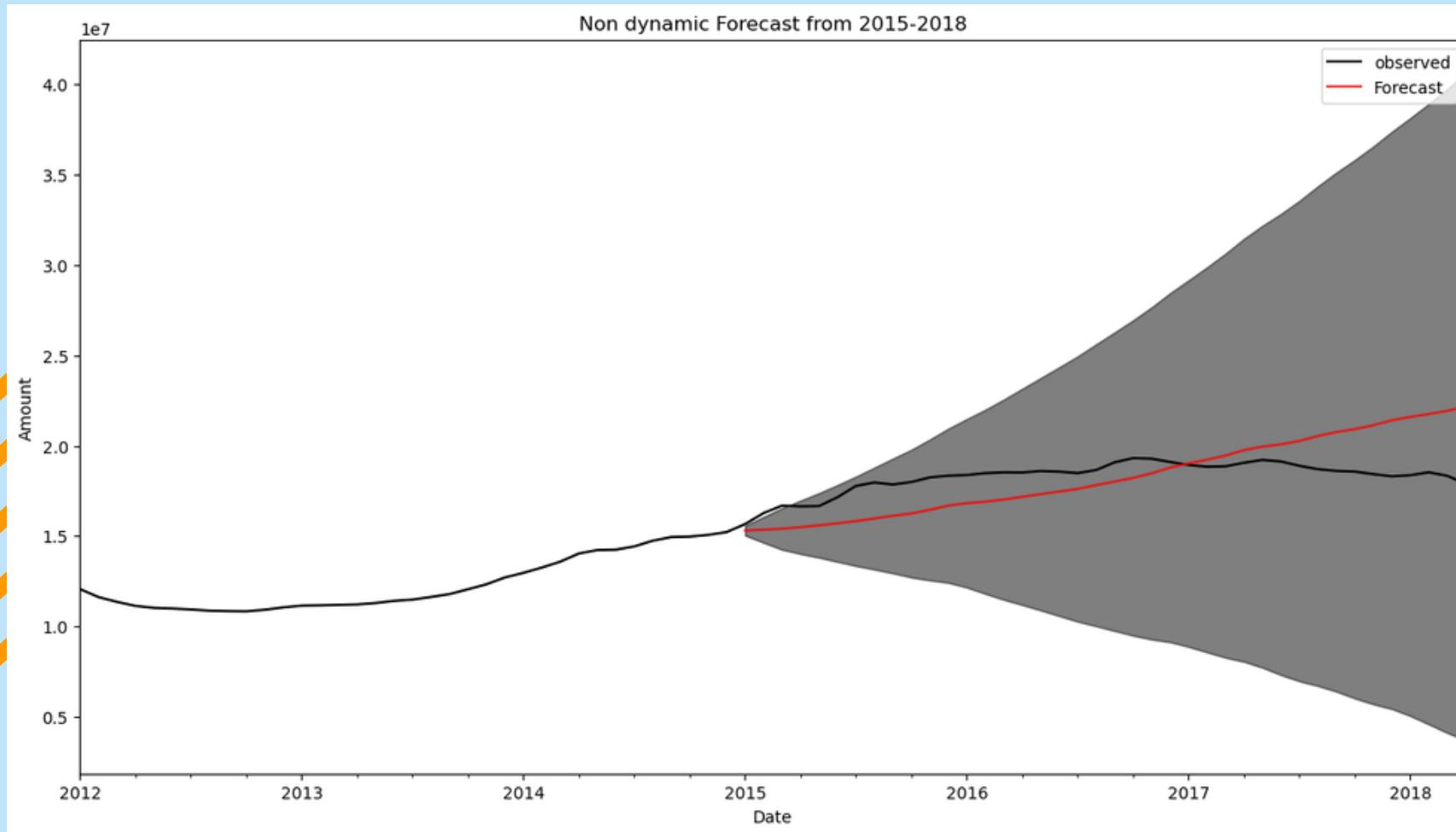


- The mean absolute percentage error (MAPE) was 0.005210916925129 336
- Black Line - Actual observed values
- Red Line(Forecast) - Predicted Values from trained model
- Grey Shaded Area - Confidence Interval of the Forecast

Evaluation

b) Non Dynamic Forecast

- Uses out-of-sample predictionThe model makes continuous one-step ahead predictions ($t+1$), then uses these predictions to expand the dataset for refitting and making two-step ahead forecasts ($t+2$).

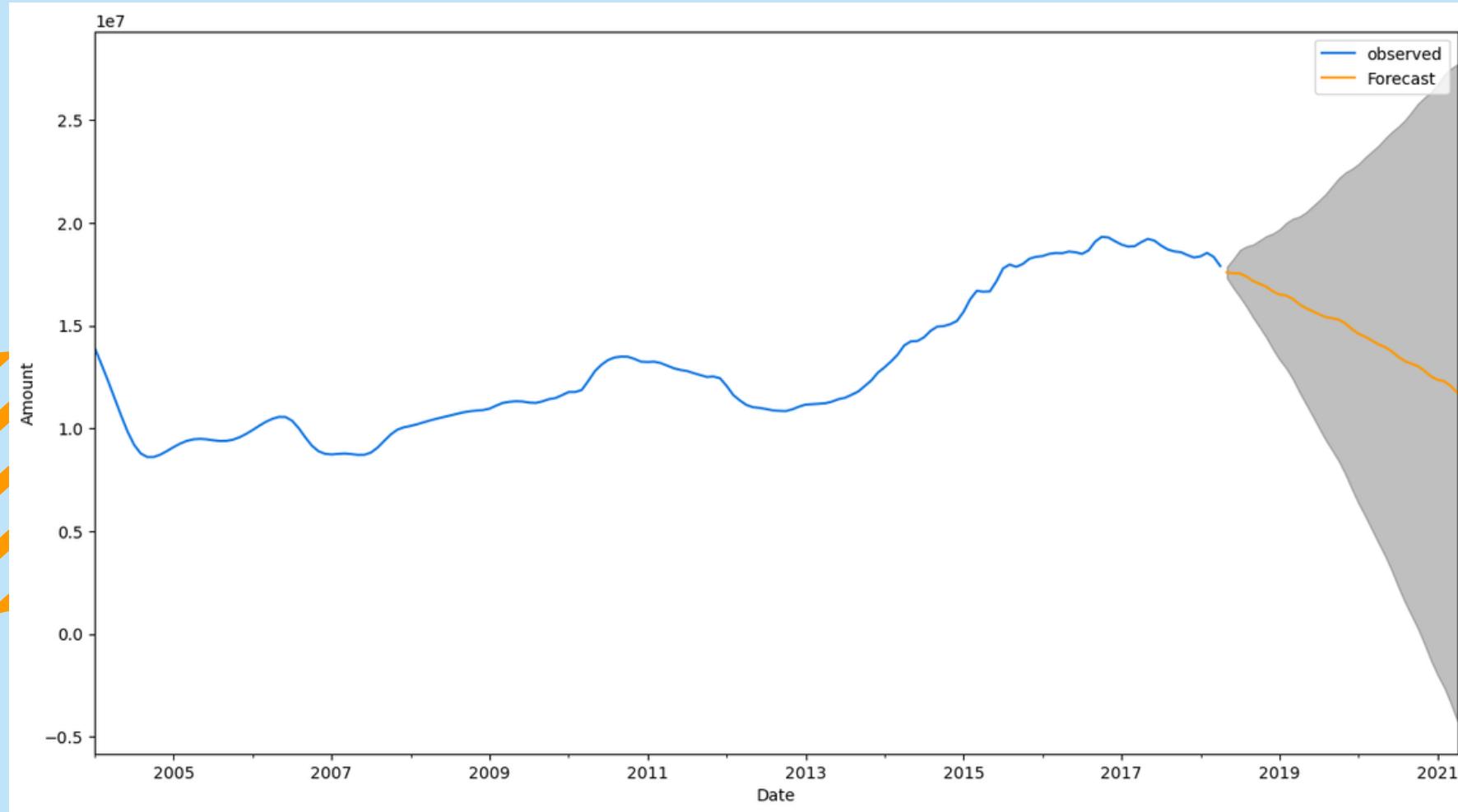


- The mean absolute percentage error (MAPE) was 0.084611622444 42985
- This large grey area indicates great uncertainty thus the model is less precise.

Evaluation

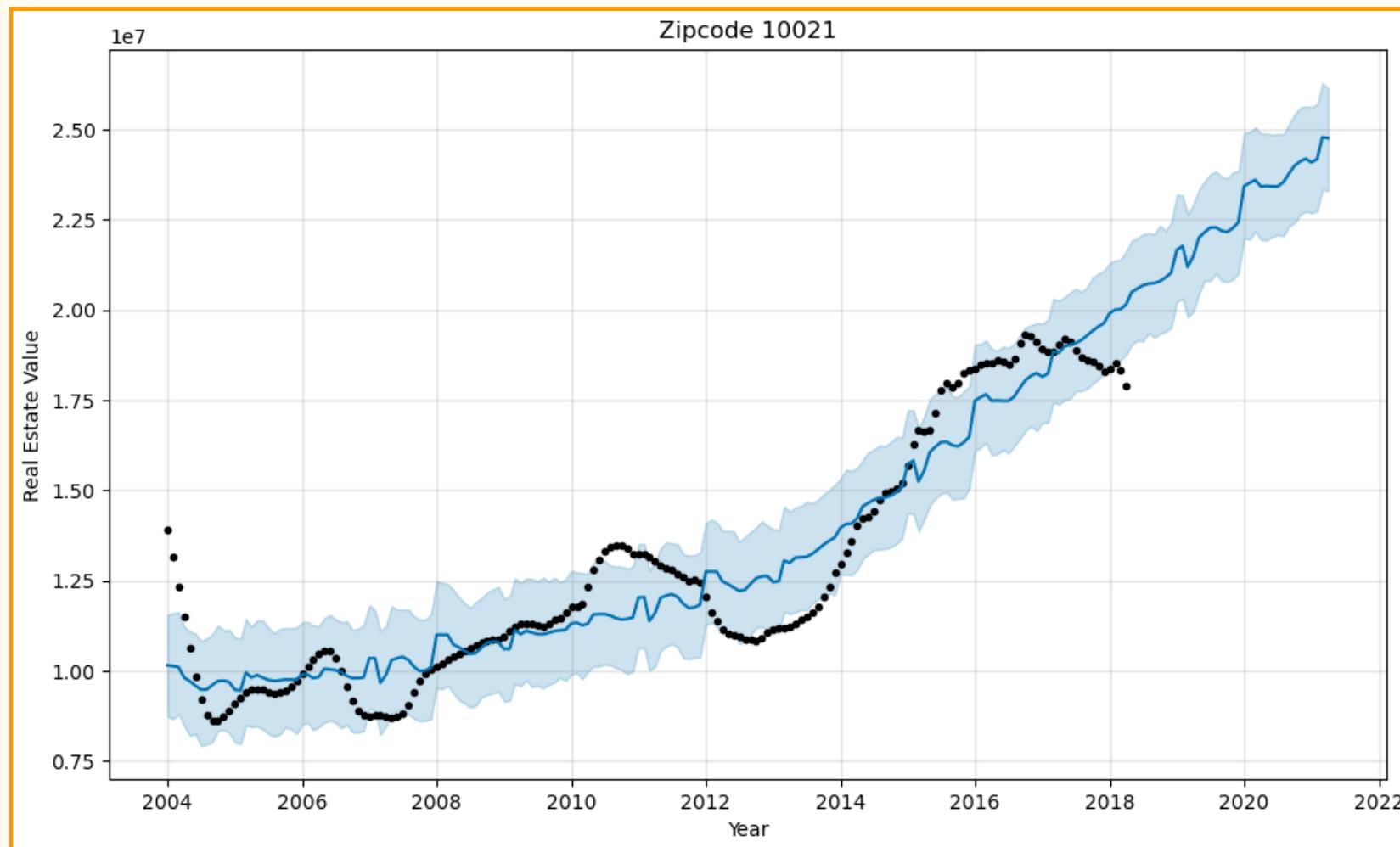
Future Prediction

- Our seasonal arima shows a drop in price in the coming years.



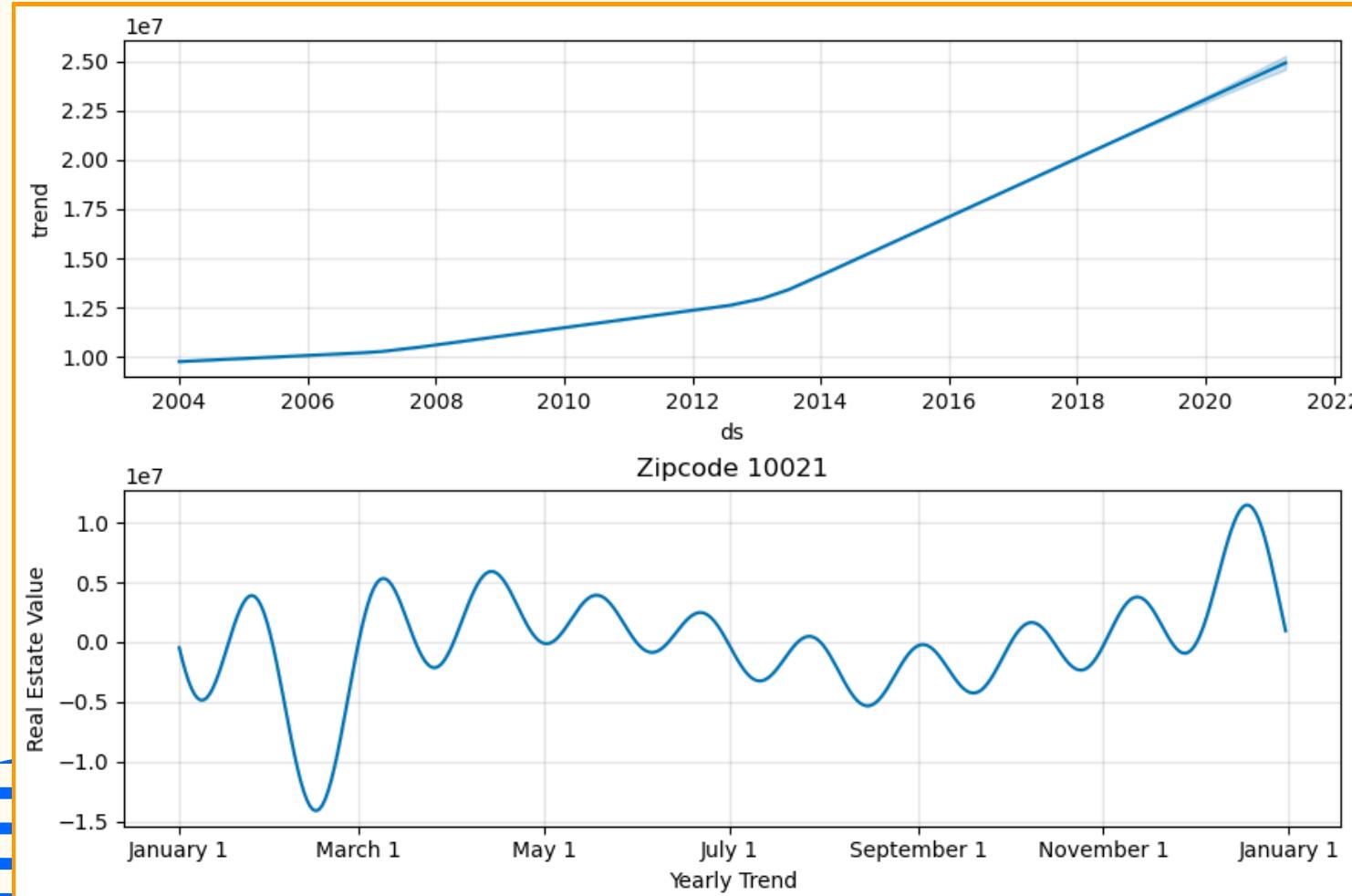
fbprophet Model

Series Graph



- **Black Dots** - Actual observed real estate values at different points in time.
- **Solid Blue Line** - Trend line that smooth out the fluctuations in the observed data to show the underlying trend in property values over time.
- **Blue Shaded Area** - Confidence interval, provides an estimate of the uncertainty around the trend line; a wider area suggests greater uncertainty.

fbprophet Model



- **Top Graph (Trend)** - shows a consistent increase over time in the data
- **Bottom Graph (Seasonal)** - shows the seasonal fluctuations within a single year.

Conclusions

Model comparison:

- The SARIMAX model outperforms Prophet model: Lower MAPE indicates better performance.

Future analysis:

- Analyze different zip codes: Current analysis discourages investment in the specific zipcode used.

Future analysis:

- Split data for training/testing: Improve analysis by splitting data into training and testing sets.

Limitations:

- Historical data only: Relies on assumptions about future market behavior.
- Lacks external factors: Doesn't include economic indicators that influence prices.
- Missing values: Might affect accuracy, requiring data preprocessing.



Recommendations

- Develop a framework for real-time monitoring of market indicators, enabling prompt adaptation to market changes and informed investment decisions.
 - Integrate external economic factors like interest rates and GDP growth into analysis for a comprehensive understanding of real estate market dynamics.
 - Perform thorough validation and sensitivity analyses to ensure reliability by testing against various scenarios and adjusting parameters.
- 
- 



THANK YOU!



Reimagine real estate to find
your way home

About

[About Us](#)
[Our Team](#)
[Features](#)
[Career](#)

Support

[Support Center](#)
[Feedback](#)
[FAQ](#)

Follow Us

[Instagram](#)
[Twitter](#)
[Facebook](#)
[YouTube](#)