



# PRIMENEST REAL ESTATE PROJECT

SOLPHINE JOSEPH

GRACE J. ROTICH

HILARY SIMIYU

MATHEW KIPROTICH

CLYDE OCHIENG

# BREAKDOWN

1. Business understanding
2. Data Understanding
3. Data Preparation
4. Explanatory Data Analysis
5. Modelling
6. Deployment
7. Conclusion
8. Next steps



# Business understanding

The project is dedicated to helping PrimeNest Investments select the best zip codes for real estate investment. Through a thorough analysis of historical data and the implementation of time series modeling, the objective is to provide robust insights to maximize returns on investment while effectively managing market risks and fluctuations, taking into account market risks and fluctuations.

# OBJECTIVES

- To provide a data-driven foundation for investment decisions with a model achieving 80% accuracy:
- To develop and deploy a FastAPI application to serve the real estate forecasting models, enabling seamless and real-time access to predictions for stakeholders. .
- Offer insights into market trends, enabling stakeholders to manage risks effectively and align investments with long-term financial goals.

# CHALLENGES

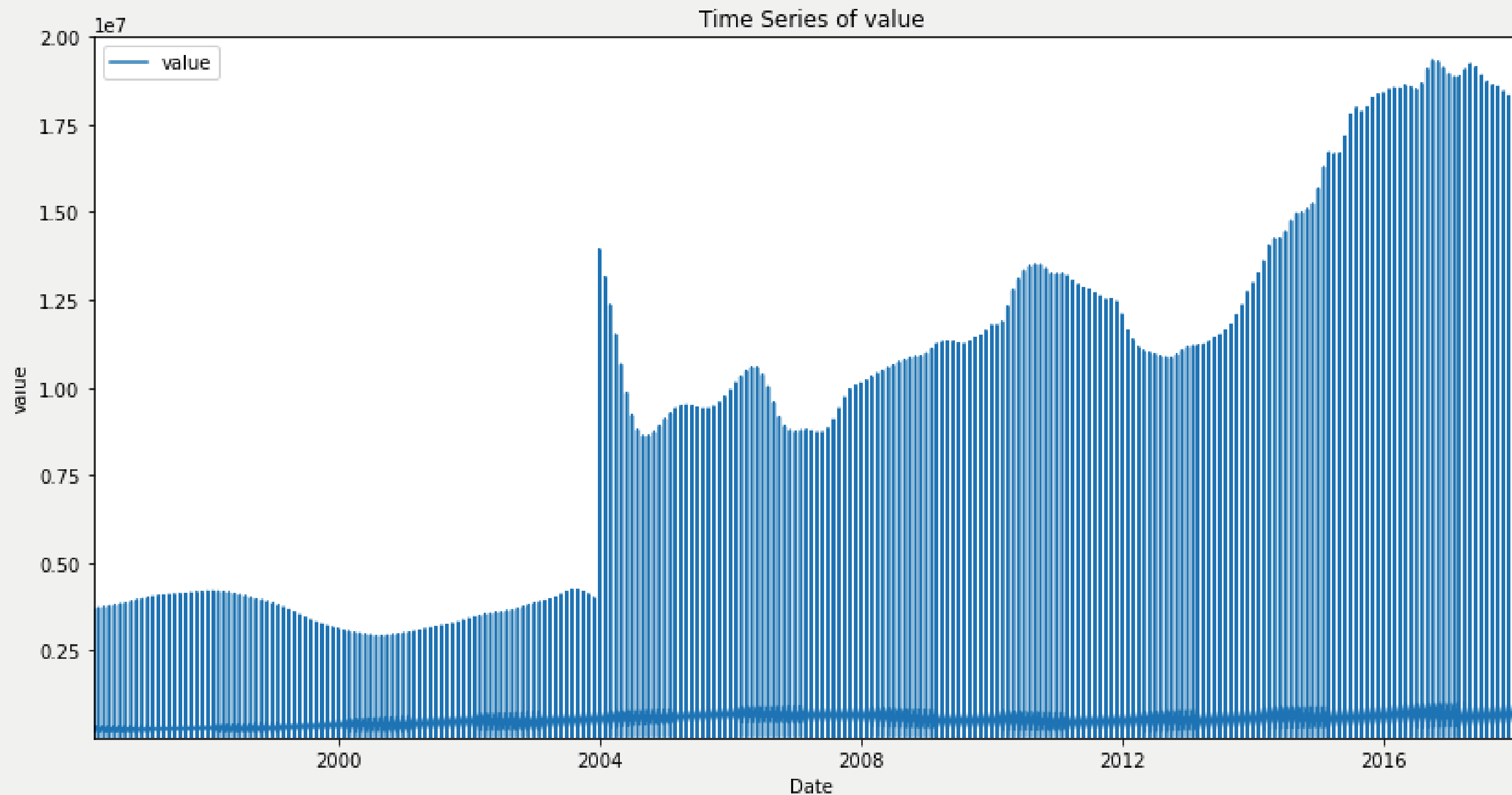
- Accurately forecasting real estate prices.
- Considering profit margins, risk, and investment horizons.
- Leveraging historical data for decision-making in the real estate market.

# PROPOSED SOLUTION

- Implement robust data cleaning methods to address missing values and unique values.
- Create time series plots to visualize trends and seasonal patterns.
- Employ SARIMA models to capture both seasonal patterns and trends in the time series data.
- Employ ARIMA models to capture trends
- Model evaluation using the Root Mean Squared Error (RMSE)

# Explanatory Data Analysis

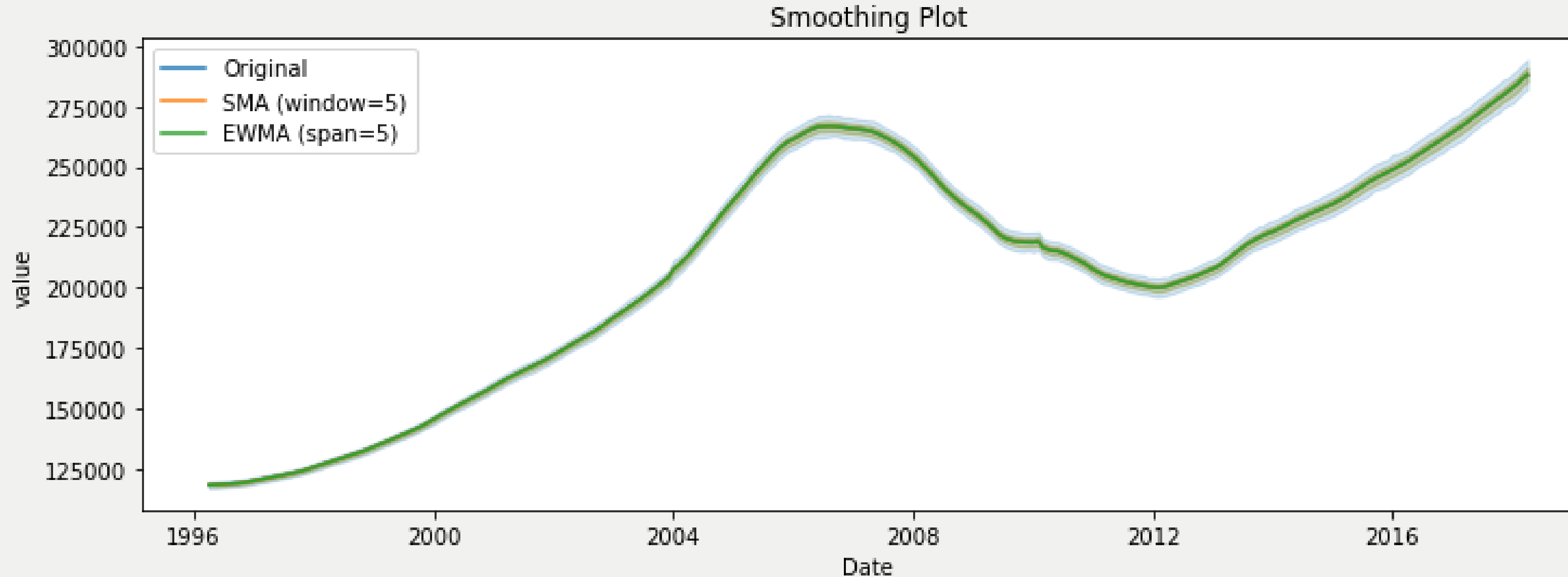
The trend of the entire dataset from 1996 to 2018.  
With the y-axis being the value of the houses(house prices) and the x-axis being the dates.



# Explanatory Data Analysis

## Time series smoothening:

Simple moving average (SMA) and Exponential weighted moving average (EWMA) are used below.

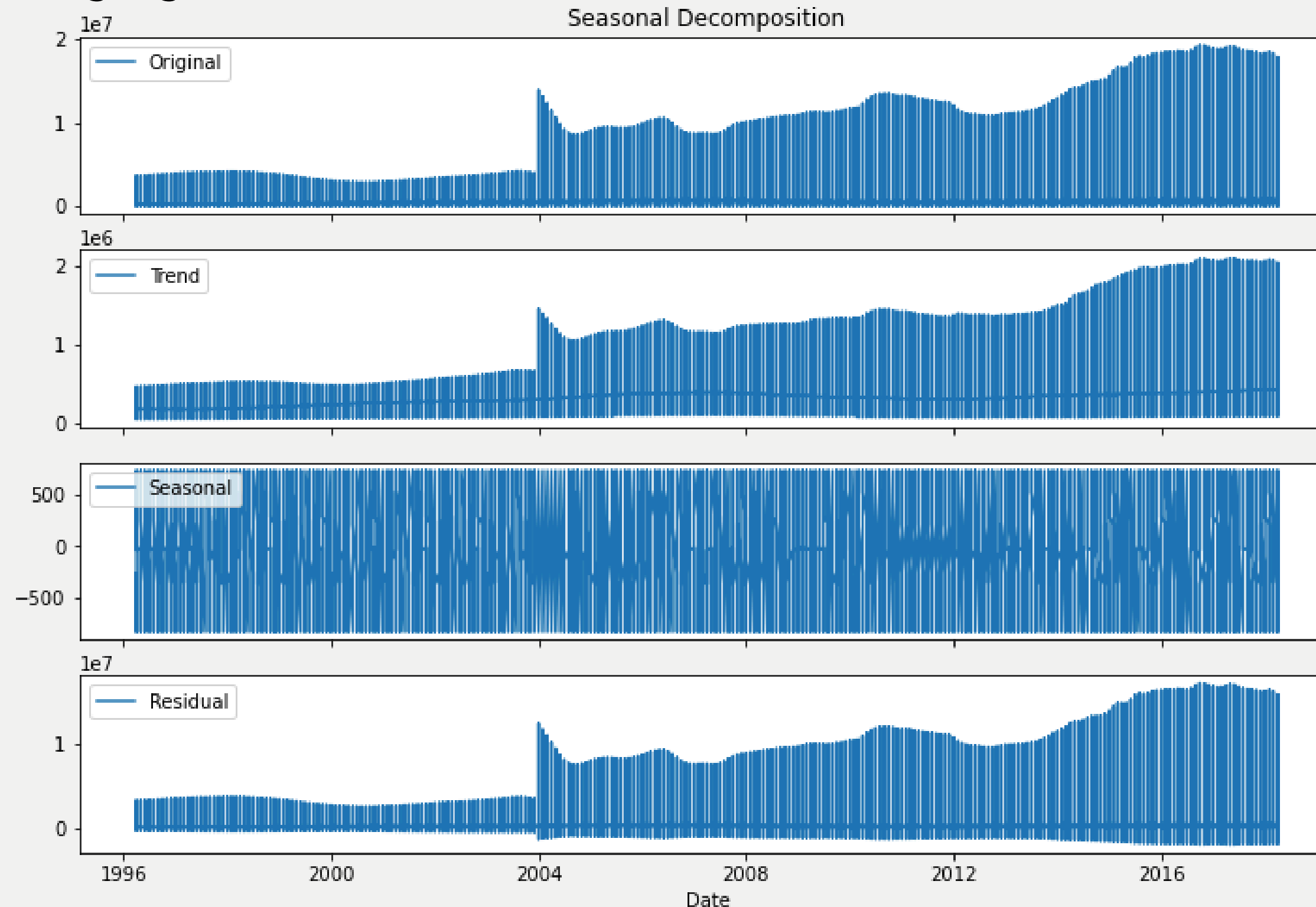




# Exploratory Data Analysis

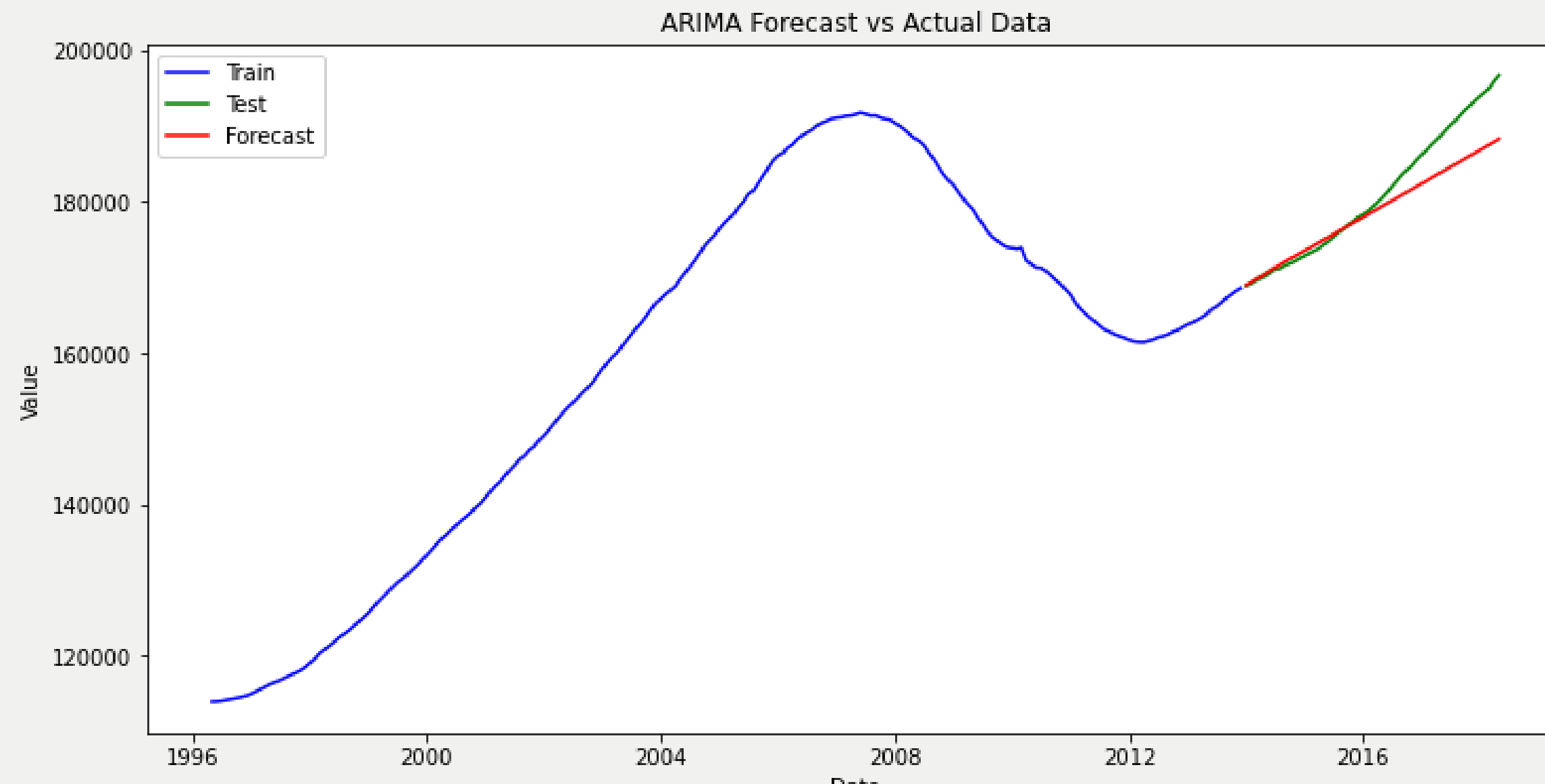
## Seasonal Decomposition:

Checking for seasonality in our dataset. This affects the type of machine learning algorithms to use.



# ARIMA MODEL

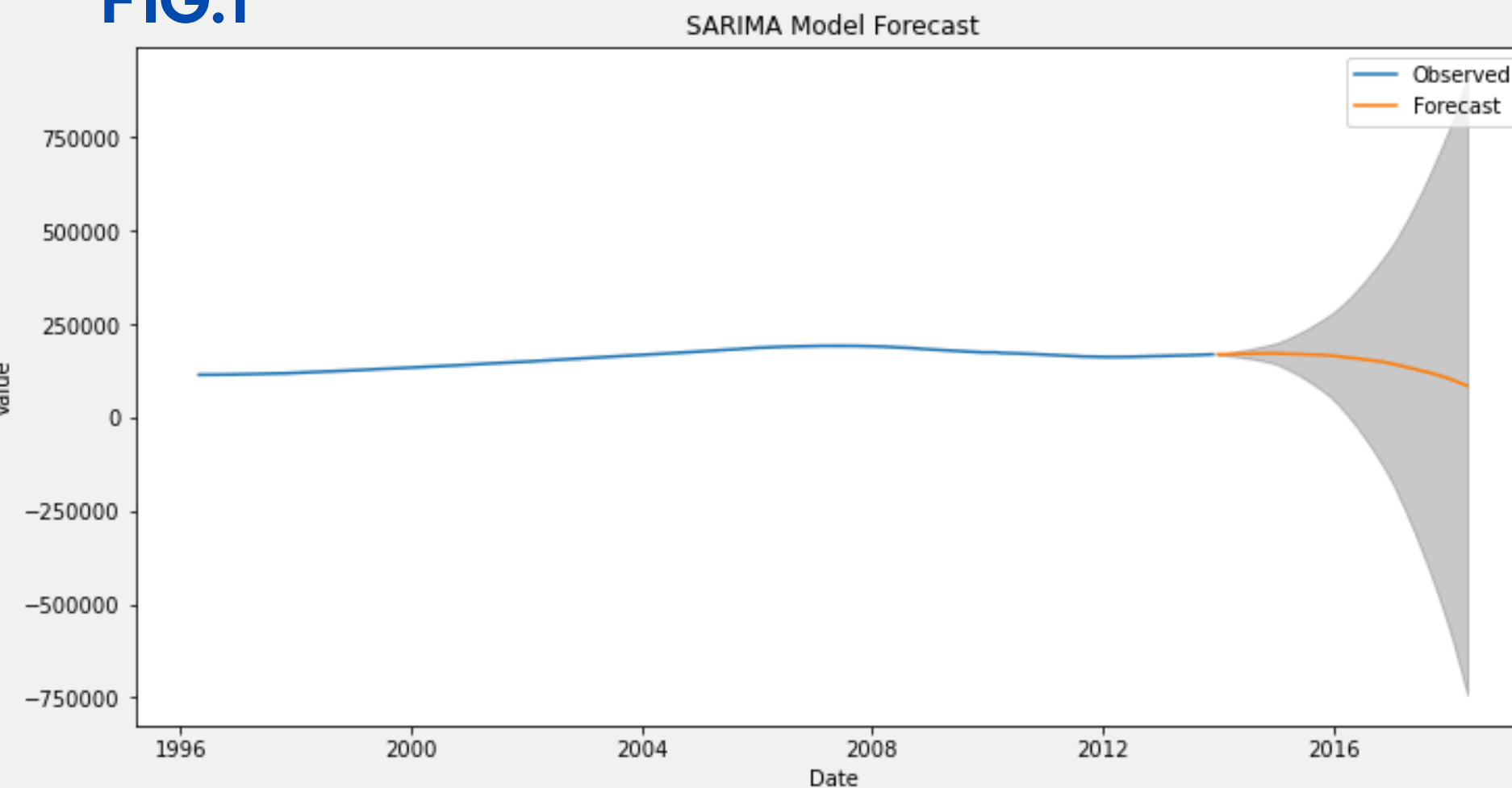
- The baseline model is an ARIMA model that captures the trend of the data.
- The model has an R2 Score of 81% and from the RMSE every forecast made it by 3657 dollars.
- However the model fails to capture the seasonality of the data ,hence need of Sarima model.



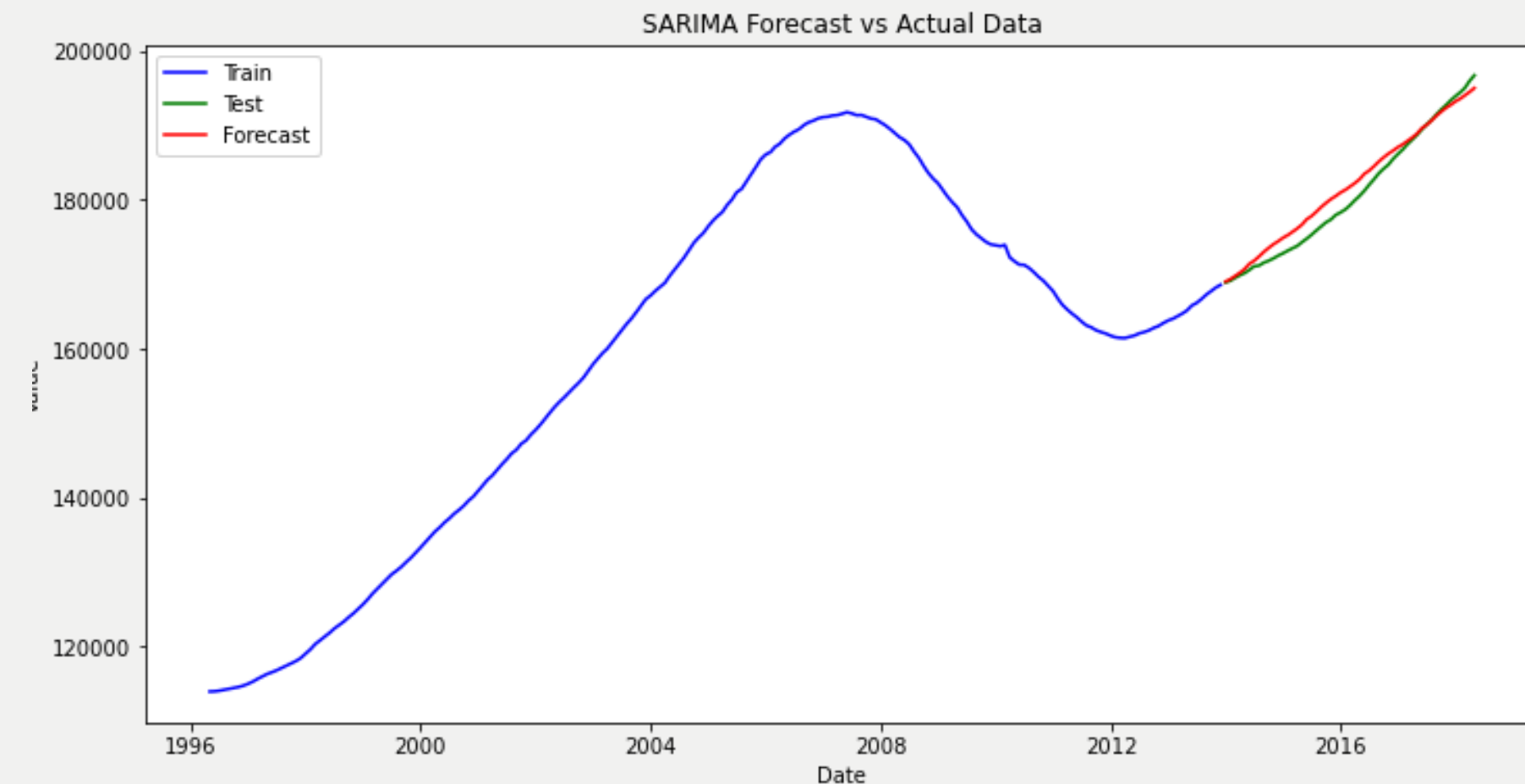
# SARIMA MODEL

- With an R2-squared of 96% the second iteration model works best compared to the ARIMA model with an r2-squared of 81% and the first SARIMA model. as shown below.
- An RMSE of 1663.02 indicates that each forecast is off by about 1670 dollars

**FIG.1**



**FIG.2**



# DEPLOYMENT

- The model was deployed for use to end users using python Framework Fast API
- A user inputs a date in form of a string
- THE model API has a predict route that returns the forecasted house values from 2013 November
- The input date should be greater than November 2013, otherwise an error message will be displayed
- Using the SARIMA Model

# CONCLUSION

- The SARIMA model forecasted house prices with a relatively low RMSE compared to the mean house price, indicating high accuracy.
- Time series analysis revealed consistent seasonal patterns in house prices.
- The SARIMA model effectively captured these seasonal patterns.

# NEXT STEPS

1. Continuously collect and preprocess new data.
2. Fine-tune SARIMA model parameters and explore hybrid models.
3. Present findings to stakeholders and refine strategy with feedback.
4. Implement and monitor the investment strategy, adjusting as needed.

THANK

**YOU**