



# FORECASTING REAL ESTATE PRICES



# Overview

- Introduction **01**
- Business Understanding **02**
- Objectives **03**
- Exploratory Data Analysis **04**
- Modelling **05**
- Findings **06**
- Conclusion **07**



# Introduction

This project aims to accurately predict future prices which have a significant implications for buyers, sellers, and investors. The project uses time series analysis, for unravelling trends and patterns within sequential data.

We leverage historical real estate price data, and aim to develop a robust forecasting model capable of predicting future values with greater accuracy.

# Business Understanding and Problem Statement

The real estate market is complex and influenced by numerous factors, making the identification of prime investment locations a significant challenge. Traditional methods often rely on limited data and subjective expertise, leading to inconsistencies and reduced decision-making power.

This project aims to develop a robust and data-driven model to identify the top 5 best zip codes for real estate investment. The model will leverage advanced analytics and machine learning techniques to analyze historical data on property characteristics, market trends, and economic indicators, providing clear and actionable insights for strategic investment decisions.

# Objectives

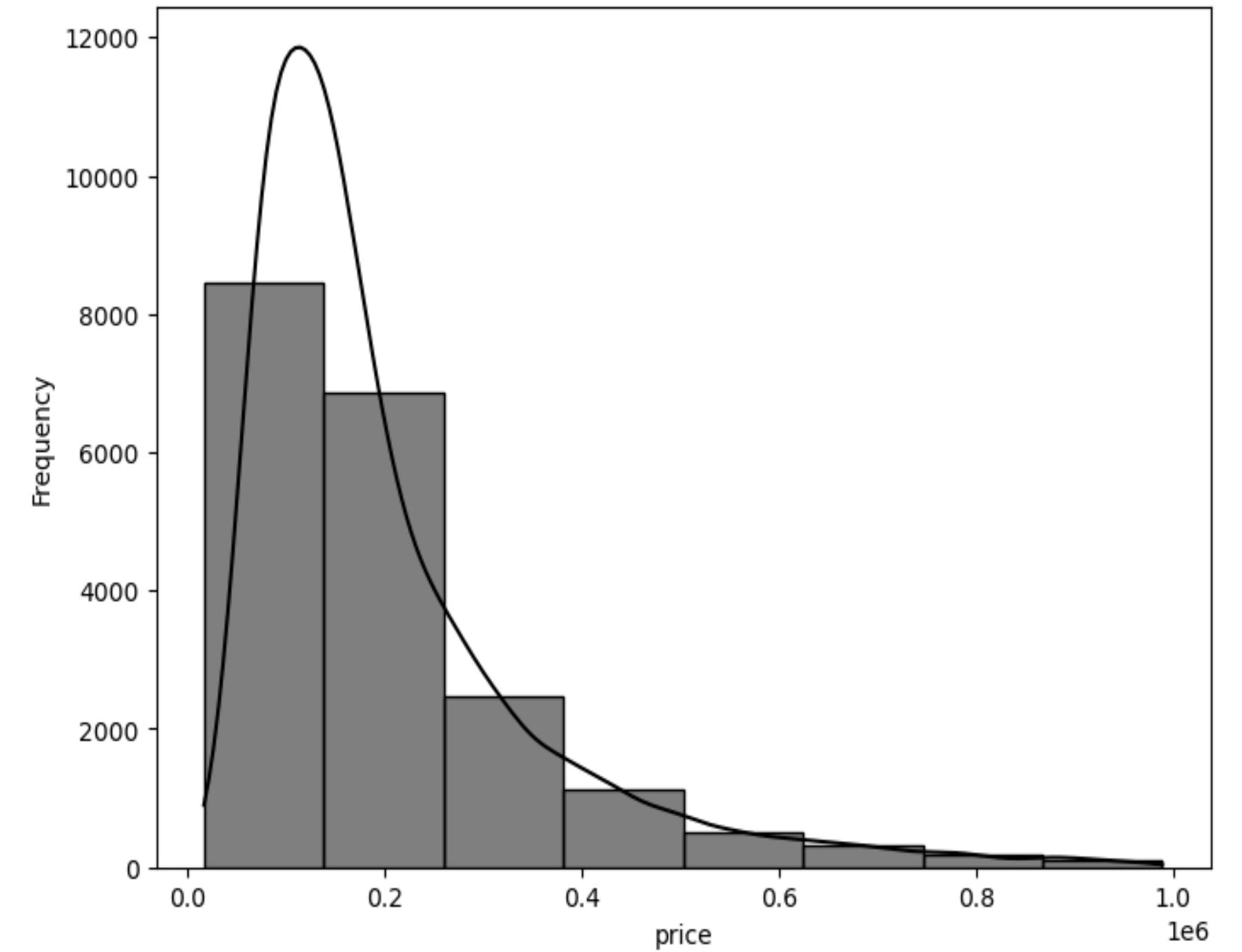
1. Provide a valuable and accurate predictive time series model with improved accuracy for stakeholders in the real estate industry.
2. Identify the most significant factors influencing property values in real Estate.
3. To achieve stationarity in the real estate price data which is crucial for many forecasting models.
4. To identify and account for potential trends in the real estate market:

# Exploratory Data Analysis

We have conducted the explanatory data analysis to check on the trends of the historical data set. From the graphs shown below, here are the observations made:

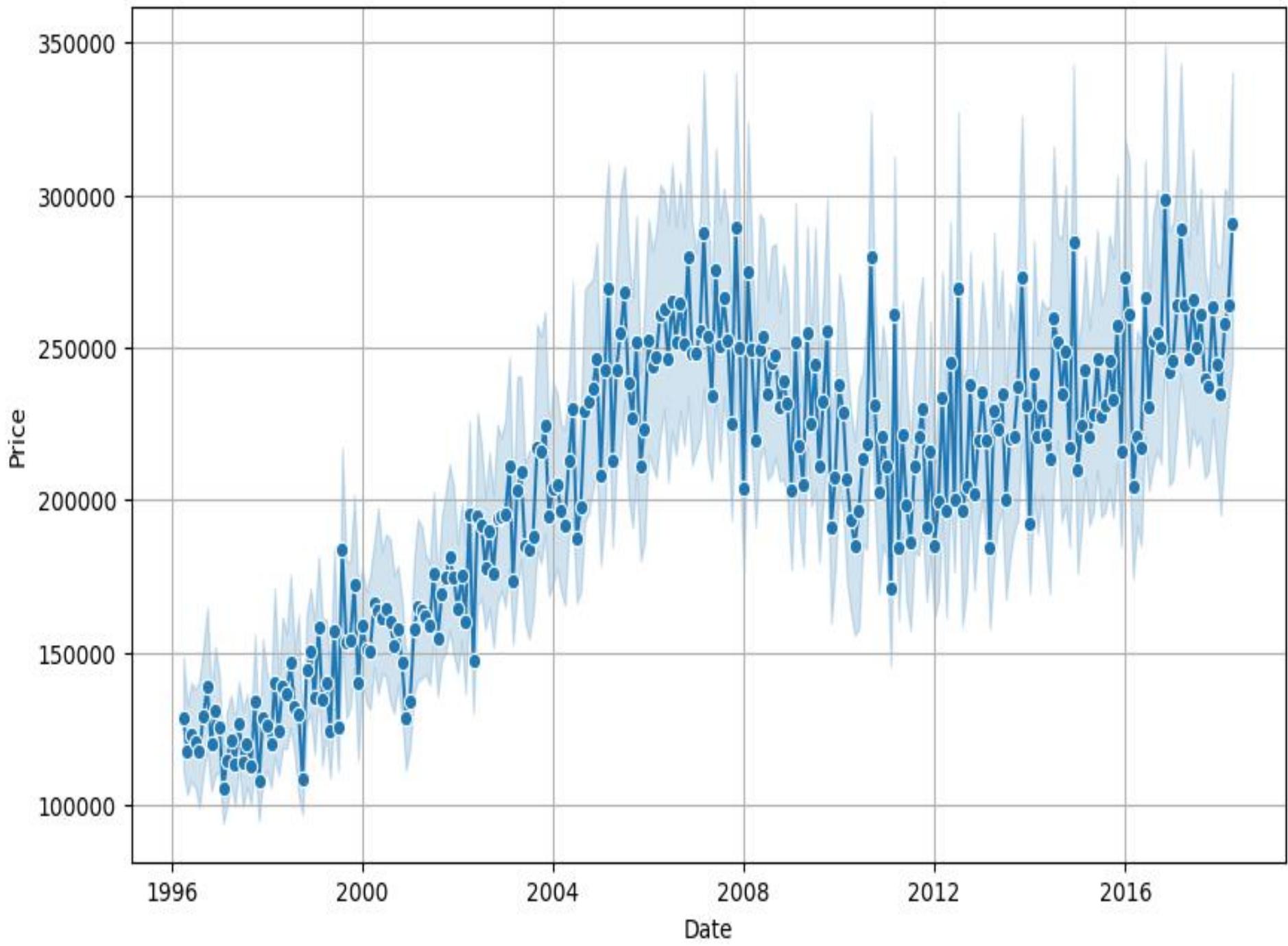
1. The sales prices have a right skew with majority of the houses selling at 2.5M to 3M (Graph I)
2. There has been an upward trend in the prices in the value of properties over time (Graph II)
3. The dataset is stationary which means that the statistical properties of the dataset do not change overtime (Graph III)

Histogram of House Prices

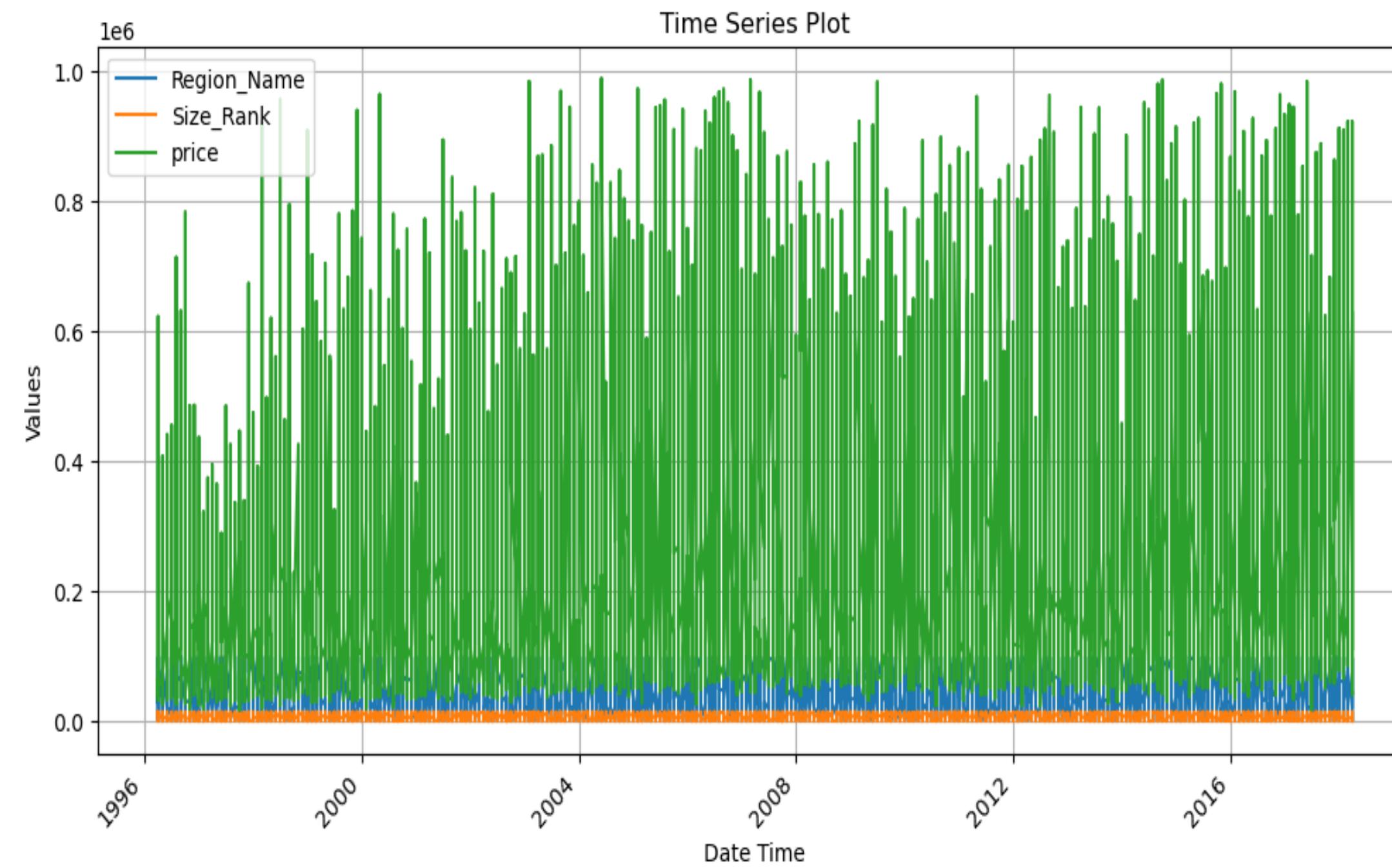


Graph I

House Prices Over Time



Graph II



Graph III

# Modelling

For the forecasting of the real estate prices,  
we shall use the Arima and Sarima models to  
predict the future prices of the properties.

# Arima Model

The model seems to capture the autoregressive nature of the differenced data (diff1) with statistically significant coefficients for past values (ar.L1 to ar.L4). However, the moving average terms (ma.L1 to ma.L3) don't appear to have a strong influence. The high variance of the error term (sigma2) suggests there might be room for improvement.

# SARIMAX Results

```
=====
Dep. Variable:      price  No. Observations:      20000
Model:              ARIMA(4, 1, 3)  Log Likelihood:   -266958.413
Date:          Sat, 29 Jun 2024  AIC:                 533932.826
Time:            21:27:46  BIC:                 533996.053
Sample:             0   HQIC:                533953.512
                   - 20000
=====
```

Covariance Type: opg

|        | coef      | std err  | z        | P> z  | [0.025   | 0.975]   |
|--------|-----------|----------|----------|-------|----------|----------|
| <hr/>  |           |          |          |       |          |          |
| ar.L1  | -1.6988   | 0.039    | -43.502  | 0.000 | -1.775   | -1.622   |
| ar.L2  | -0.9397   | 0.040    | -23.224  | 0.000 | -1.019   | -0.860   |
| ar.L3  | -0.0079   | 0.016    | -0.497   | 0.619 | -0.039   | 0.023    |
| ar.L4  | -0.0066   | 0.009    | -0.773   | 0.439 | -0.023   | 0.010    |
| ma.L1  | 0.7109    | 0.038    | 18.562   | 0.000 | 0.636    | 0.786    |
| ma.L2  | -0.7622   | 0.018    | -42.692  | 0.000 | -0.797   | -0.727   |
| ma.L3  | -0.9476   | 0.038    | -25.161  | 0.000 | -1.021   | -0.874   |
| sigma2 | 2.524e+10 | 1.24e-13 | 2.03e+23 | 0.000 | 2.52e+10 | 2.52e+10 |

Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 31603.95

Prob(Q): 0.99 Prob(JB): 0.00

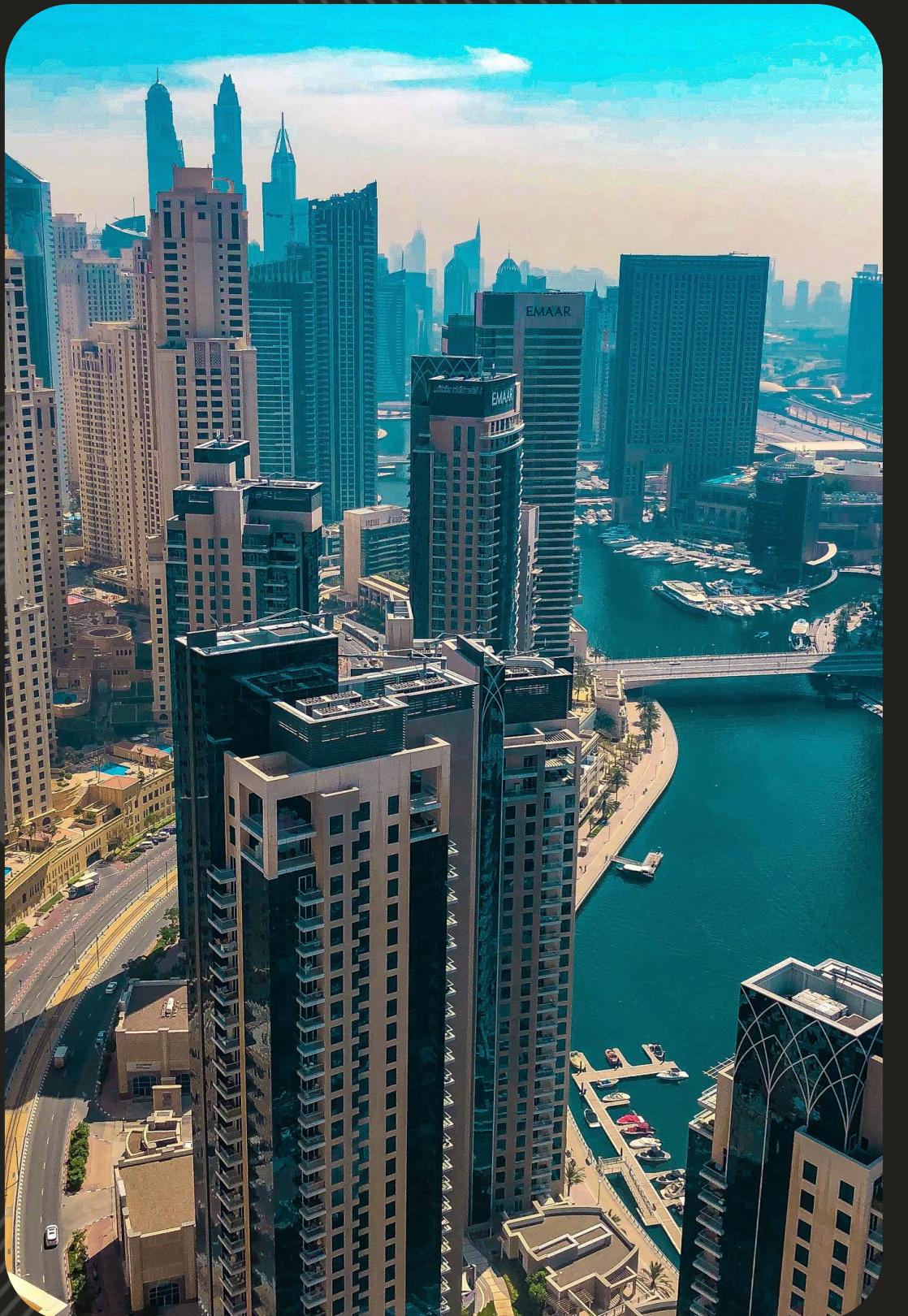
Heteroskedasticity (H): 1.02 Skew: 2.00

Prob(H) (two-sided): 0.49 Kurtosis: 7.69

Warnings:

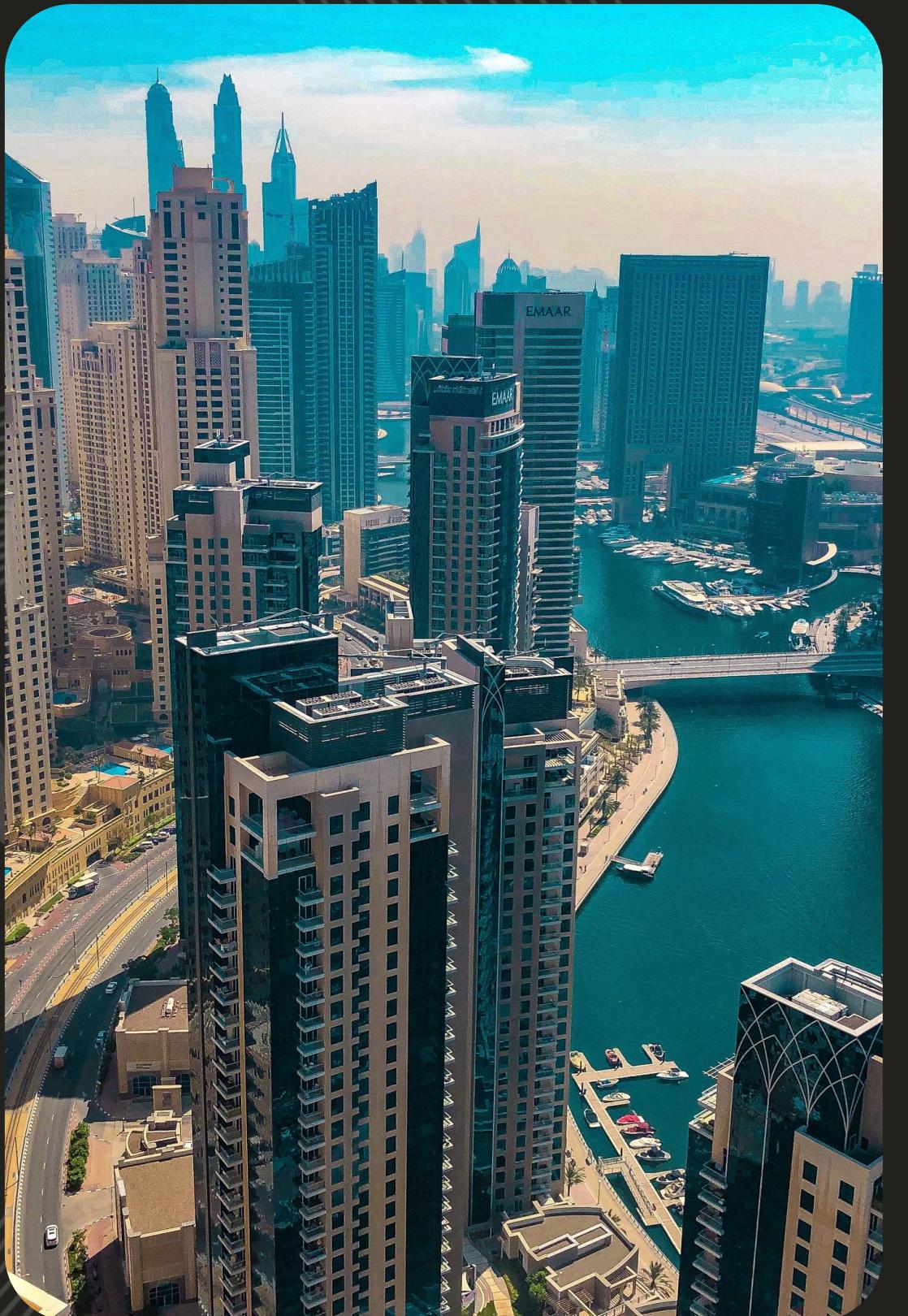
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

[2] Covariance matrix is singular or near-singular, with condition number 2.69e+39. Standard errors may be unstable.



# Findings

1. Price Trends
2. Market Shifts
3. Risk Assessment

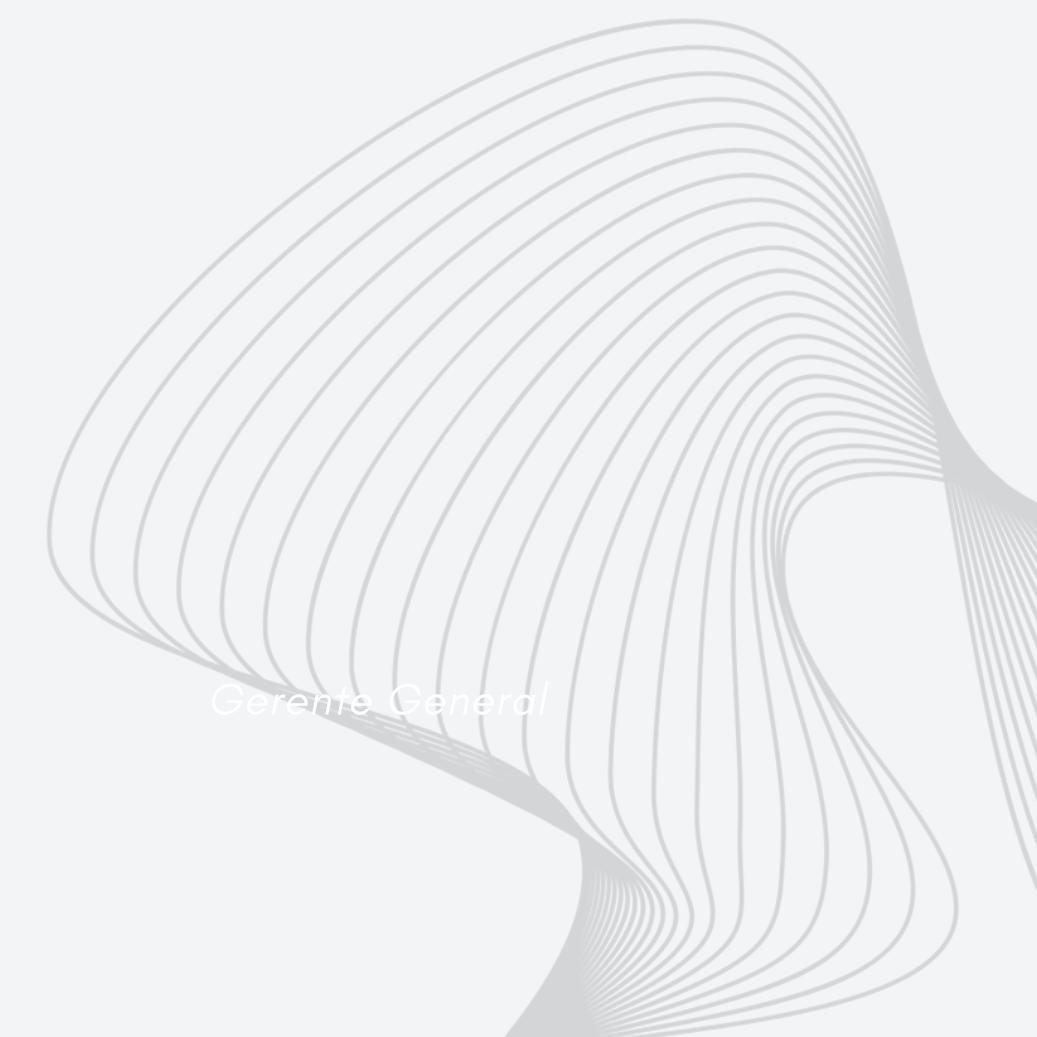


# Conclusion

In conclusion the project's value has reaffirmed how the time series analysis has provided valuable insights into market trends and price as the driver



Thank You



Gerente General

