Documentation

Task: Find publicly available data for key factors that influence US home prices nationally. Then, build a data science model that explains how these factors impacted home prices over the last 20 years.

1. *Project Overview:*

- This project aims to analyze historical data and build regression models to predict the Housing Price Index (HPI) in the United States.
- By leveraging key economic, demographic and housing market factors, I gained insights into HPI trends.
- The analysis includes data collection, preprocessing, feature engineering, and modeling, with a focus on accurate HPI predictions and meaningful findings.

2. *Data Collection:*

- Data Collected from - https://fred.stlouisfed.org/categories
- The data has 14 key features which includes –
  - **Economic Indicators**
    - *Date* (2002-2022 | 20 years)
    - *GDP (Gross Domestic Product)*
      - GDP represents the total monetary value of all goods and services produced within a country's borders. It serves as a measure of the nation's economic health and growth.
    - *CPI (Consumer Price Index)*
      - CPI is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. It reflects changes in the cost of living and inflation.

- *HPI (Housing Price Index)*
  - HPI tracks changes in the prices of residential properties over time. It is a key indicator of the housing market's performance and trends.
- *ECI (Employment Cost Index)*
  - ECI measures changes in the labor costs for employers and provides insights into wage and benefit trends, which can impact the overall economy.
- *UNRATE (Unemployment Rate)*
  - The unemployment rate indicates the percentage of the labor force that is currently without a job. It is a crucial indicator of the job market's health.
- **Demographic Indicators**
  - *Working Age Distribution*
    - Working age distribution reflects the proportion of the population within the age group typically associated with being in the workforce. It impacts labor force dynamics and productivity.
  - *Income Distribution*
    - Income distribution measures how income is spread across a population. It sheds light on wealth disparities and socioeconomic trends.
  - *Population Level*
    - Population level quantifies the total number of residents in a region. It plays a central role in understanding demographic trends and their impact on housing demand.
  - *FED (Federal Fund Rate)*
    - The Federal Fund Rate is the interest rate at which banks lend to one another in the United States. It is set by the Federal Reserve and influences overall interest rates.

- **Mortgage**
  - Mortgage rates represent the interest rates applied to loans for purchasing homes. They significantly influence housing affordability and demand.
  - **Housing Market Indicators**
    - **Homeownership Rate**
      - Homeownership rate denotes the percentage of households that own their homes. It reflects the prevalence of homeownership in a given region.
    - **Rental Vacancy Rate**
      - Rental vacancy rate measures the proportion of rental properties that are unoccupied. It offers insights into the demand for rental housing.
    - **Housing Inventory Estimate**
      - Housing inventory estimate provides an estimate of the number of available homes for sale in a specific market. It affects housing supply and pricing.

3. *Exploratory Data Analysis (EDA):*

Feature Engineering:

- After performing correlation analysis for feature selection, the following output was generated –

```
GDP                          0.873220
FED                          0.869696
Income Distribution          0.836804
ECI                          0.825638
Housing Inventory Estimate   0.778069
Rental Vacancy Rate          0.752992
Population Level             0.745547
CPI                          0.624177
```

```
Working Age Distribution       0.622487
UNRATE                         0.512042
Mortgage                       0.315767
Homeownership Rate             0.301625
```

4. Data Modeling:

- Choice of Regression model –
    - o **Linear Regression**
        - ▪ Linear Regression is selected as one of the regression models for its simplicity and interpretability. It assumes a linear relationship between the target variable (HPI) and the features.
        - ▪ Linear Regression serves as the baseline model, providing straightforward insights into the relationship between key economic and demographic factors and HPI.

- Cross-Validation & Evaluation –
    - o **K-Fold Cross-Validation**
        - ▪ The dataset is divided into multiple subsets or "folds," typically with a common value of k (5)
        - ▪ The model is trained on k-1 folds and evaluated on the remaining fold, repeated k times.
        - ▪ Cross-validation helps estimate how well the model generalizes to unseen data, provides insights into its predictive accuracy, and helps identify any issues of overfitting or underfitting.
        - ▪ Evaluation metrics used in cross-validation include R-squared, Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to assess the goodness of fit and prediction accuracy.

```
        R2 Score:  0.8892863562863811
   Mean Squared Error:  47.64049121790354
        Accuracy:  88.9286356286381 %

 Linear Regression Mean R-squared:  0.639074876684388
 Linear Regression Std R-squared:  0.40552036530842744
```
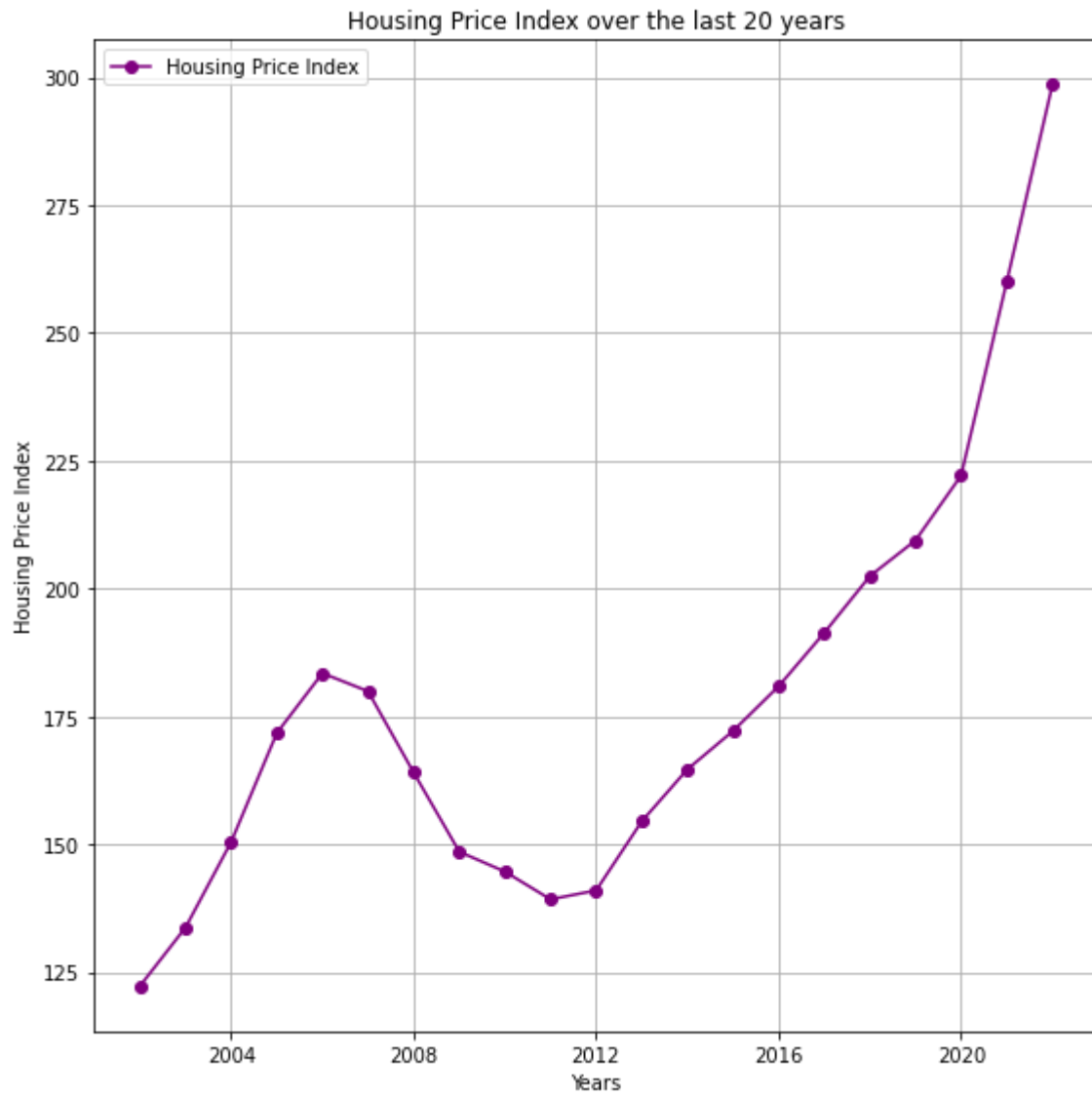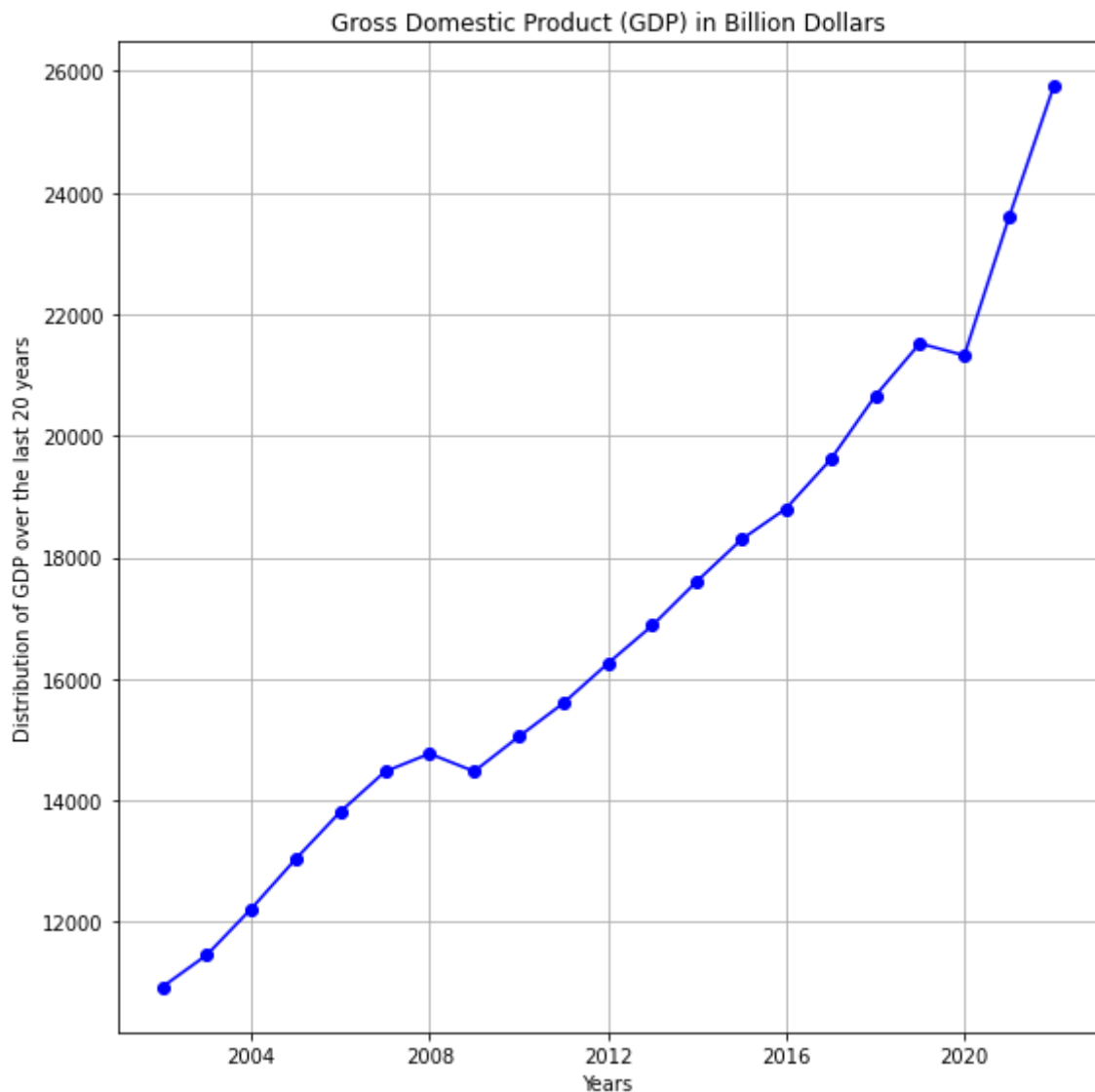
5. *Visualization & Findings:*

Housing Price Index

1. **Economic Indicators:**
   a. **Gross Domestic Product (GDP)**

Gross Domestic Product (GDP) in Billion Dollars

Effects of GDP on Housing Price

A rising GDP is likely to positively affect housing prices. Economic growth and increased prosperity often lead to higher demand for housing, which can drive up prices as people have more financial means to invest in real estate.

b. Consumer Price Index (CPI)

Consumer Price Index (CPI) in Billion Dollars
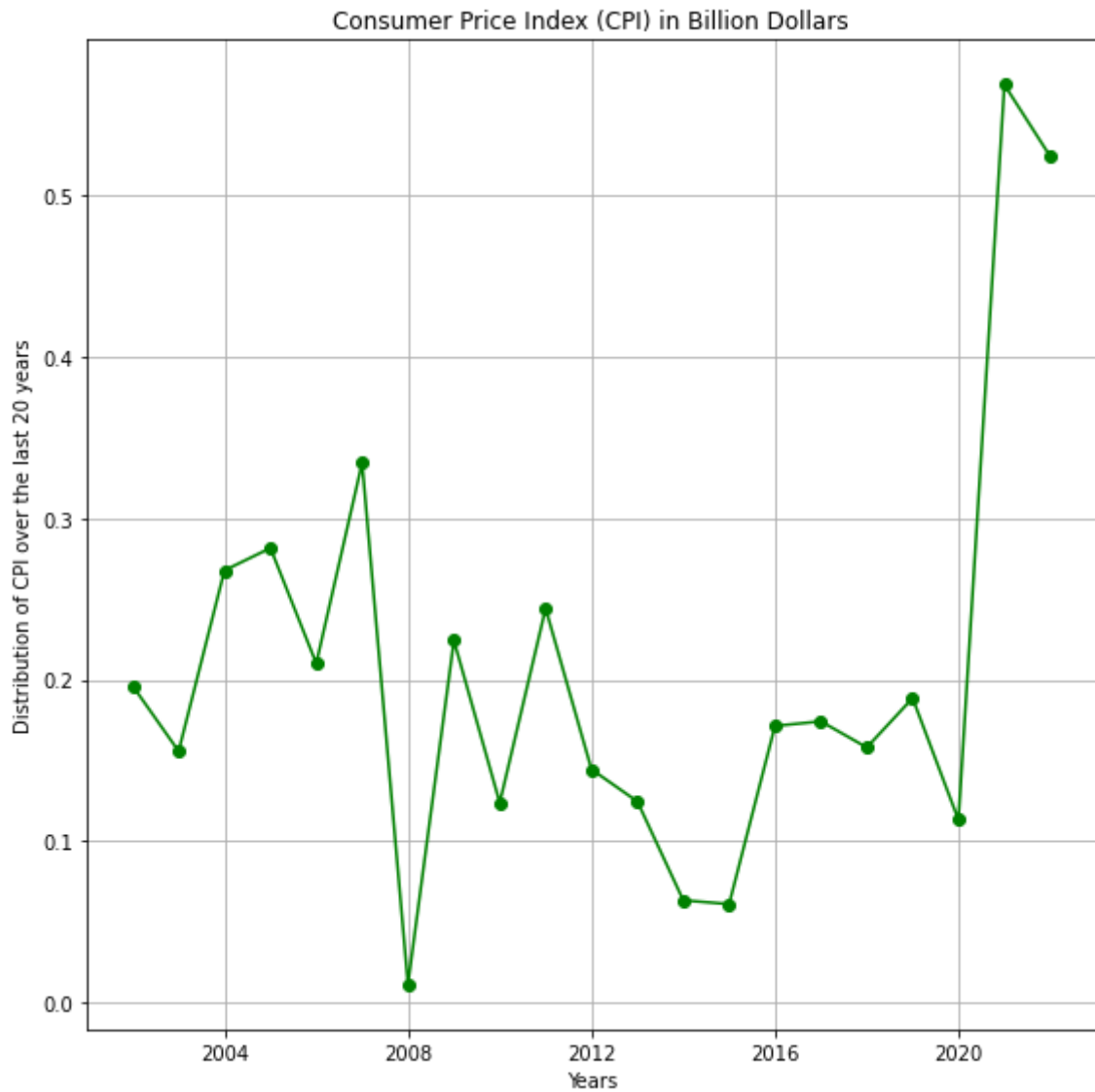
Effects of CPI on Housing Price

The Consumer Price Index (CPI) can influence housing prices by reflecting changes in the cost of living. An increase in CPI may lead to higher housing costs, impacting affordability and potentially contributing to rising housing prices.

c. Unemployment Rate

Unemployment Rate over the last 20 years
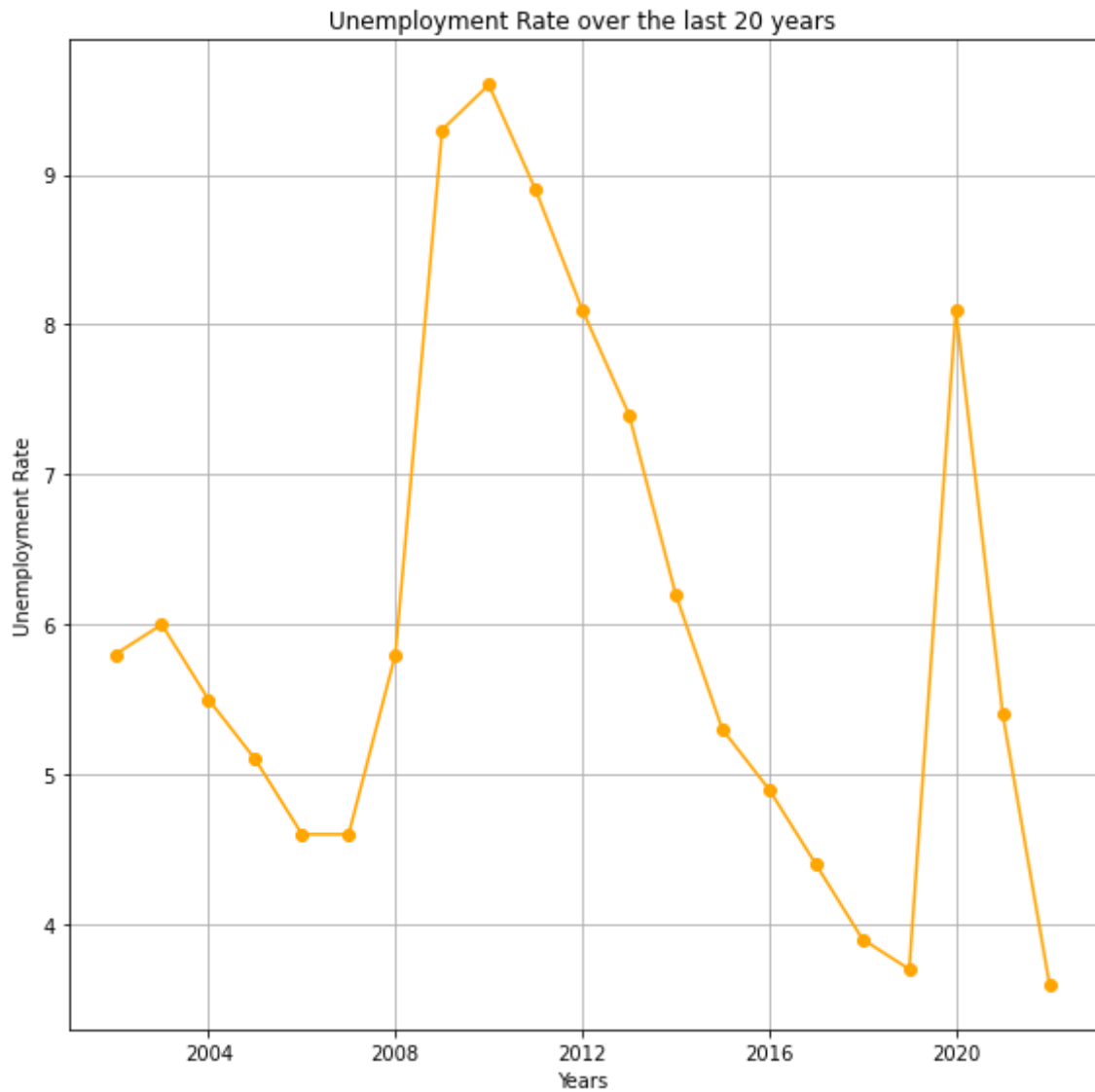
Effects of Unemployment on Housing Price

A higher unemployment rate may negatively impact housing prices by reducing people's ability to afford homes, leading to decreased demand and potential downward pressure on prices.
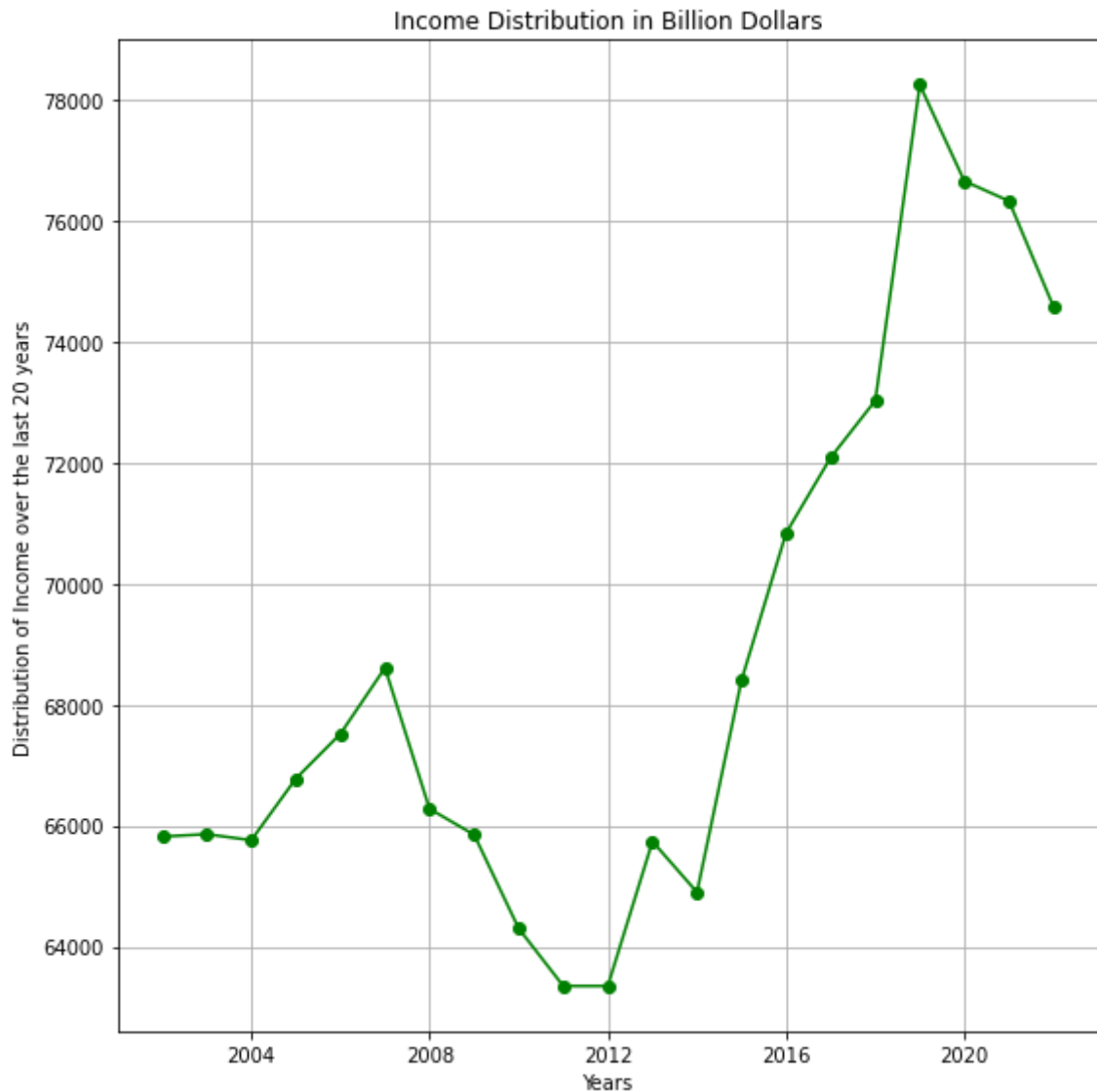
2. **Demographic Indicators:**

    a. **Population Growth**

Population Growth over the last 20 years
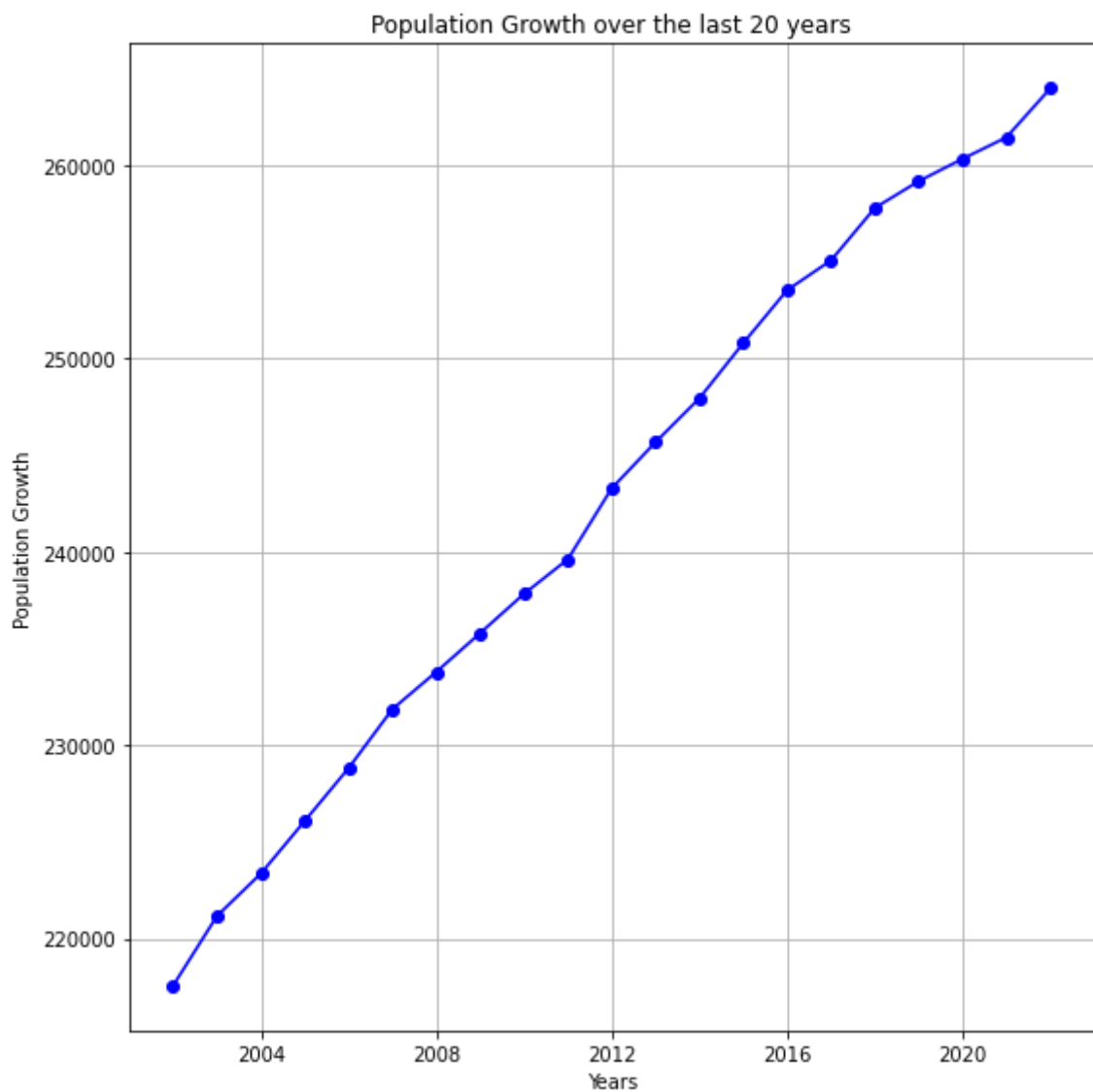
Effects of population growth on housing price

Population growth often leads to increased housing demand, which can drive up prices as supply struggles to keep pace with a growing population.

b. Income Distribution

Income Distribution in Billion Dollars

Effects of Income on housing price

Higher income levels tend to positively influence housing prices, as people with greater incomes can afford more expensive homes and are willing to pay higher prices for desirable properties. But there has been a drop in income near 2018.

c. Mortgage Rate

Average Mortgage Rate over the last 20 years
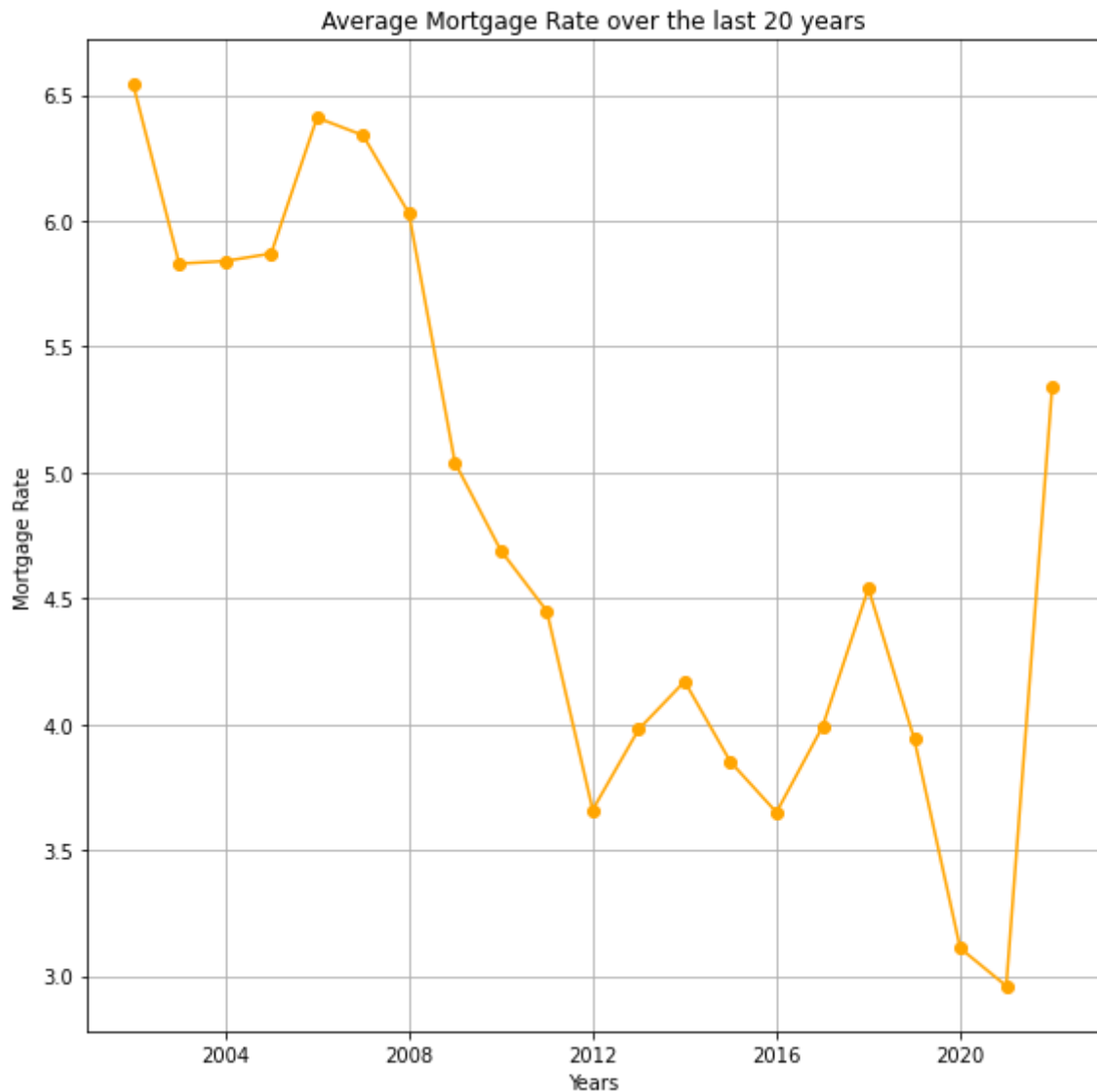
Effects of mortgage rate on housing price

Mortgage rates can impact housing prices. Lower mortgage rates can stimulate housing demand, potentially driving up prices, while higher rates may reduce affordability and dampen demand, affecting prices accordingly.

3. **Housing Market Indicators:**

   a. **Homeownership Rate**

Homeownership Rate over the last 20 years
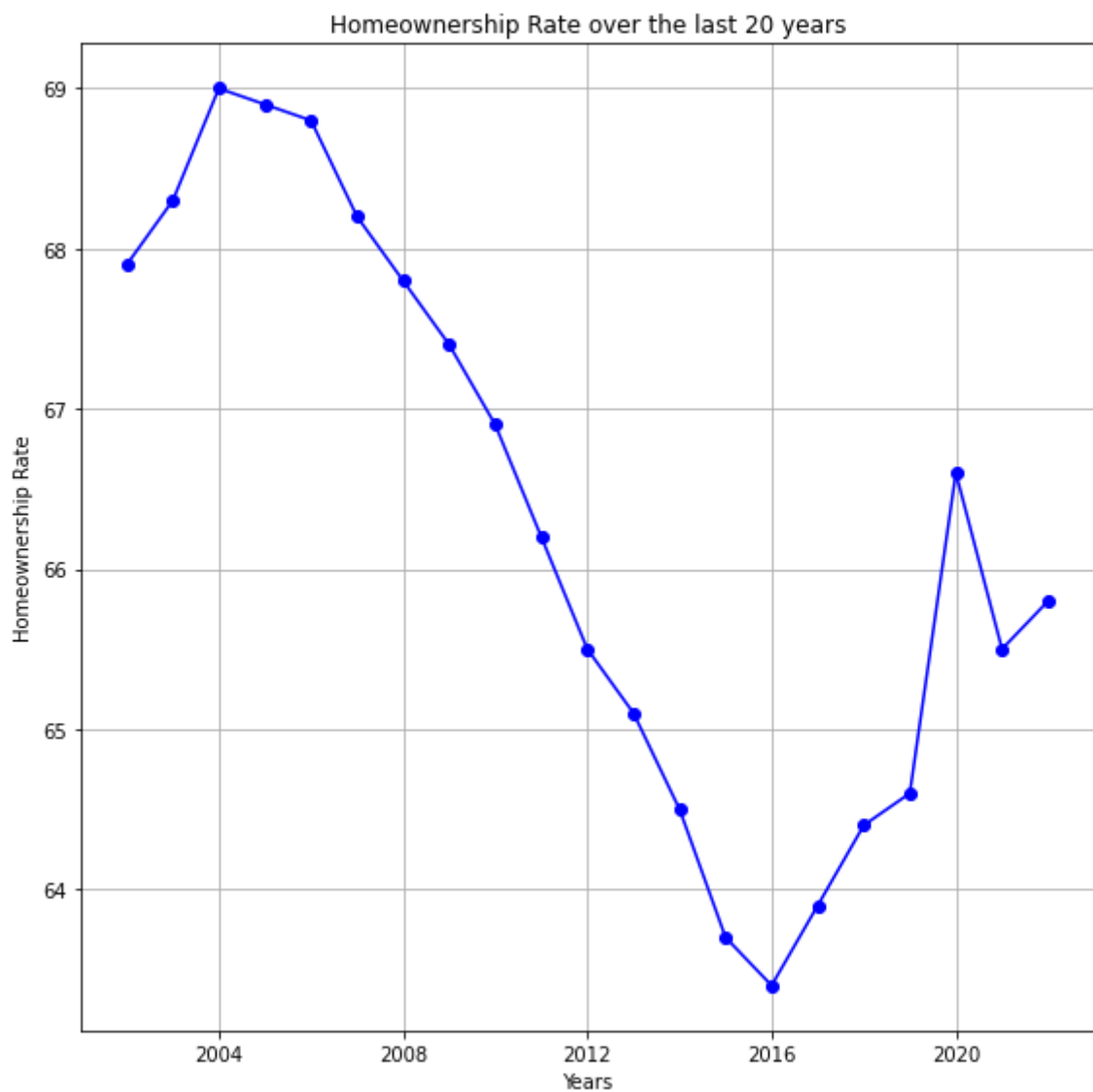
Effects of homeownership rate on housing price

A higher homeownership rate may lead to more stable and potentially higher housing prices, as it reflects a higher demand for homes and a sense of property value maintenance in the community.

b. Rental Vacancy Rate

Rental Vacancy Rate over the last 20 years
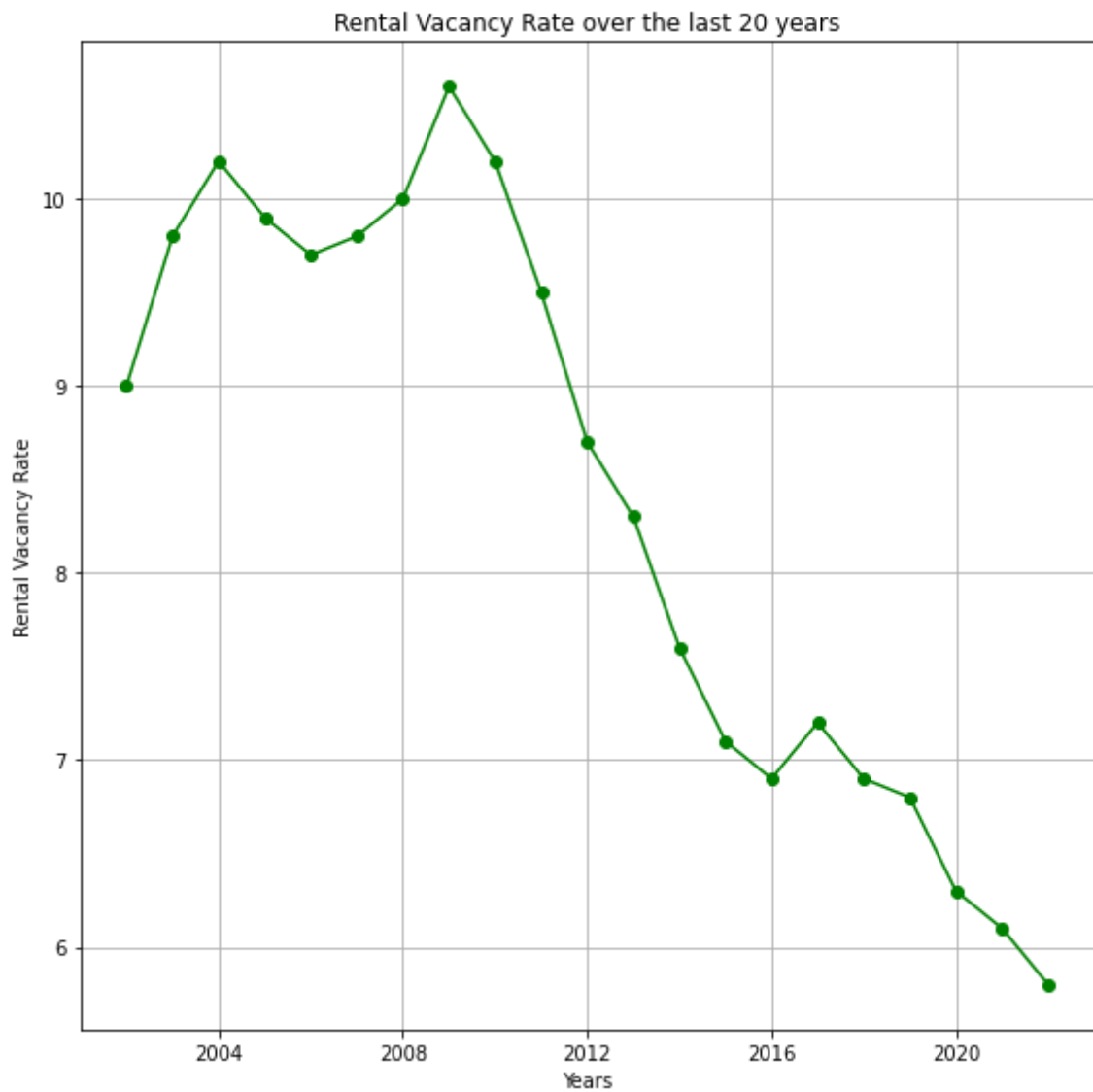
Effects of rental vacancy rate on housing price

A lower rental vacancy rate can indicate a stronger demand for housing, potentially driving up prices as individuals and families seek to buy homes rather than rent due to limited rental options.

c. Housing Inventory Estimate

Housing Inventory Estimate over the last 20 years
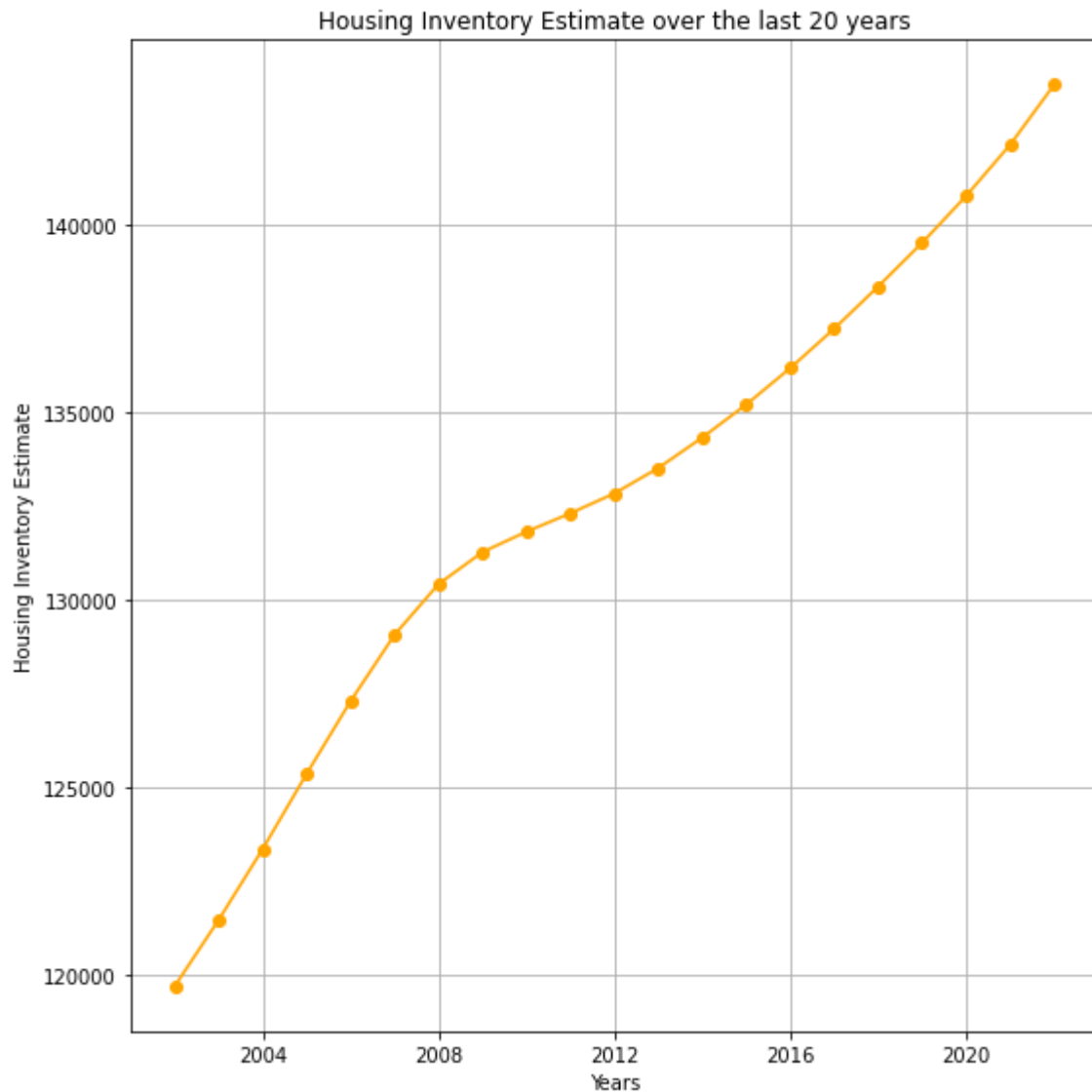
Effectors of housing inventory estimate

An increasing housing inventory estimate typically suggests a growing supply of homes, which can put downward pressure on housing prices as supply outpaces demand.

6. Conclusion:

In conclusion, the analysis of various factors influencing housing prices in the United States over the last 20 years reveals several key insights. Economic indicators, such as GDP, CPI, and unemployment rates, have played a substantial role in shaping housing market dynamics. A growing GDP and lower unemployment rates tend to boost housing demand and drive

prices upward, while rising CPI can strain affordability and impact prices. Moreover, population growth has consistently driven demand, leading to higher prices, as communities struggle to meet the housing needs of their expanding populations. Income levels have also proven to be a significant driver of housing prices, as higher incomes enable individuals to invest in more expensive homes, thus increasing the value of properties. Additionally, mortgage rates and homeownership rates show sensitivity to housing prices, with lower rates and increasing homeownership contributing positively to price trends. Finally, the rental vacancy rate and housing inventory estimate have demonstrated that limited supply can propel prices upward, highlighting the complex interplay of supply and demand dynamics in the housing market. Understanding these relationships is crucial for homeowners, real estate professionals, and policymakers as they navigate the ever-evolving landscape of housing prices in the United States.