

- **OBJECTIVE:**

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.

- **AIMS:**

The project aims to build a data science model to predict U.S. home prices based on these factors over the last 20 years.

- **STEPS:**

**1) Data Collection:** Utilized the S&P Case-Schiller Home Price Index as a proxy for home prices, sourced from the Federal Reserve Economic Data (FRED) website.

**2) Key Factors:** Gathered publicly available data on factors influencing home prices nationally, including Per Capita GDP, Consumer Price Index (CPI), Construction Material Costs, Median Income, Subsidies, and Population Demographics, etc.

**3) Data Cleaning and Processing:** Cleaned and processed the data, addressing missing values, converting date formats, and handling outliers.

**4) Exploratory Data Analysis (EDA):** Conducted EDA to understand the distribution of variables, identify correlations, and visualize trends over time.

**5) Model Selection:** Explored various regression models, including Linear Regression, Elastic Net, Random Forest, Gradient Boosting, Support Vector Regression (SVR), and XG Boost.

**6) Model Training and Evaluation:** Trained each model using a subset of the data, evaluated performance using metrics such as Mean Squared Error (MSE) and R-squared.

**7) Feature Importance:** Analysed feature importance for models like Random Forest, XG Boost, and Gradient Boosting to understand the factors influencing home prices.

**8) Model Comparison:** Compare the performance of different models based on metrics such as Mean Squared Error (MSE) and R-squared.

Select the best-performing model that provides accurate predictions and insights into the factors influencing home prices over the last 20 years.

**9) Visualization:** Create visualizations to illustrate the relationships between actual and predicted home prices for each model.

Visualize the importance of different features or coefficients in influencing home prices.

**10) Conclusion:** Identified strong contender for the best model, considering their low MSE and high R-squared values.

Draw conclusions about the key factors that have historically influenced US home prices.

**11) Overall Implication:** The project contributes to understanding the key factors influencing U.S. home prices over the last 20 years and provides a foundation for building robust predictive models in the real estate domain.