1. Data Collection Overview

We collected data for predicting US house prices by gathering 13 key economic and housing market indicators from a reliable source: the S&P CoreLogic Case-Shiller U.S. National Home Price Index. This platform provides accurate, historical time-series data crucial for building robust predictive models.

2. Data Source

- **S&P CoreLogic Case-Shiller U.S. National Home Price Index**: Provides data for national home price trends and related indicators.

3. Collection Process

- Identified 13 important features influencing house prices.

- Downloaded historical data in CSV format for a consistent time frame (e.g., 20 years).

- Cleaned the data by removing duplicates, filling missing values, and standardizing date formats.

- Merged and aligned data features into a single comprehensive dataset.

- Validated the data by cross-checking trends and ensuring integrity.

4. Features and Their Descriptions

Below are the features collected, their descriptions, and their importance:

1. **CSUSHPISA (S&P/Case-Shiller U.S. National Home Price Index)**

   o **Description**: Measures changes in residential real estate values.

   o **Importance**: Directly reflects housing price trends and is critical for price prediction.

2. **DAXRNSA (Private Residential Fixed Investment)**

   o **Description**: Tracks spending on residential construction.

   o **Importance**: Indicates investment trends and housing demand.

3. **GDP (Gross Domestic Product)**

   o **Description**: Represents the overall economic health of the country.

   o **Importance**: Strong GDP correlates with increased housing demand and higher prices.

4. **HOUST (Housing Starts)**

   o **Description**: The number of new residential construction projects.

   o **Importance**: Reflects supply-side activity in the housing market.

5. **LXXRNSA (Real Estate Loans Outstanding)**

   o **Description**: Volume of real estate loans.

   o **Importance**: Highlights credit availability for homebuyers and its effect on demand.

6. **MIXRNSA (Mortgage Delinquency Rate)**

- o **Description**: Percentage of delinquent mortgage payments.

- o **Importance**: Indicates financial stress and potential downward price pressure.

7. **MORTGAGE30US (30-Year Fixed Mortgage Rate)**

   - o **Description**: Average interest rate for 30-year fixed mortgages.

   - o **Importance**: Affects housing affordability and demand.

8. **NYXRSA (Existing Home Sales)**

   - o **Description**: Volume of previously owned homes sold.

   - o **Importance**: Reflects liquidity and demand in the housing market.

9. **PERMIT (Building Permits)**

   - o **Description**: Number of permits issued for new construction.

   - o **Importance**: A leading indicator of future housing supply.

10. **POPTHM (Population in Thousands)**

    - o **Description**: Tracks population growth trends.

    - o **Importance**: Higher population drives housing demand and prices.

11. **SFXRNSA (Single-Family Home Starts)**

    - o **Description**: Number of single-family homes under construction.

    - o **Importance**: Focuses on supply dynamics for single-family housing.

12. **UNRATE (Unemployment Rate)**

    - o **Description**: Percentage of unemployed individuals in the workforce.

    - o **Importance**: High unemployment reduces purchasing power and suppresses housing demand.

13. **CPI (Consumer Price Index)**

    - o **Description**: Tracks inflation trends.

    - o **Importance**: Reflects changes in housing costs and overall affordability.

5. Conclusion

By collecting these 13 features from the S&P CoreLogic Case-Shiller U.S. National Home Price Index, we ensured high-quality data for analyzing and predicting US house prices. Each feature captures critical economic or housing market dynamics, providing a comprehensive understanding of the factors driving price changes.