The Impact of Key Factors on US Home Prices Over the Last 20 Years

WORK FLOW

■ **01** **Collect data**

■ **02** **Prepare data**

■ **03** **Evaluate model**

■ **04** **Build model**

■ **05** **Interpret model**

key factors that influence US home prices nationally.

S&P Case-Schiller Home Price Index
Data Sourced: https://fred.stlouisfed.org/series/CSUSHPISA

Recession
Data Sourced: https://fred.stlouisfed.org/series/RECPROUSM156N

Inflation (Annual %)
Data Sourced: https://fred.stlouisfed.org/series/FPCPITOTLZGUSA

Interest Rates (Annual %)
Data Sourced: https://fred.stlouisfed.org/series/FEDFUNDS

GDP Growth (Annual %)
Data Sourced: https://www.macrotrends.net/countries/USA/united-states/gdp-growth-rate

Population Growth (Annual %)
Data Sourced: https://www.macrotrends.net/countries/USA/united-states/population-growth-rate

Real Median Household Income in the USA
Data Sourced: https://fred.stlouisfed.org/series/MEHOINUSA672N

Construction Spending (Million $)
Data Sourced: https://www.census.gov/construction/c30/data/index.html

Unemployment Rates (Annual %)
Data Sourced: https://fred.stlouisfed.org/series/UNRATE

Housing Starts(New Housing Project)
Data Sourced: https://fred.stlouisfed.org/series/HOUST

Mortgage Rates(%)
Data Sourced: https://fred.stlouisfed.org/graph/?g=zneW

Step 1: Data Collection and Preparation

In our first step, we gathered data from multiple sources, each representing a key factor affecting home prices. These include factors like household income, construction spending, GDP, and more.

Method: Data Collection and Preprocessing

We focused on a 20-year period from 2002 to 2022, ensuring the data was relevant.
We converted the 'DATE' column to a datetime format and sorted the data by date.

Step 2: Data Visualization

We believe that visualizing the data is crucial. It helps us understand how these key factors have evolved over time.

Methods: Time Series Visualization

- We utilized various types of plots:
  - Line plots to illustrate trends (e.g., GDP)
  - Bar plots for comparisons (e.g., new housing projects)
  - Area plots for cumulative data (e.g., mortgage rates)
  - Scatter plots for exploring relationships (e.g., income vs. home prices)

Step 3: Correlation Analysis

To identify the key factors that strongly correlate with home prices, we created a heatmap.
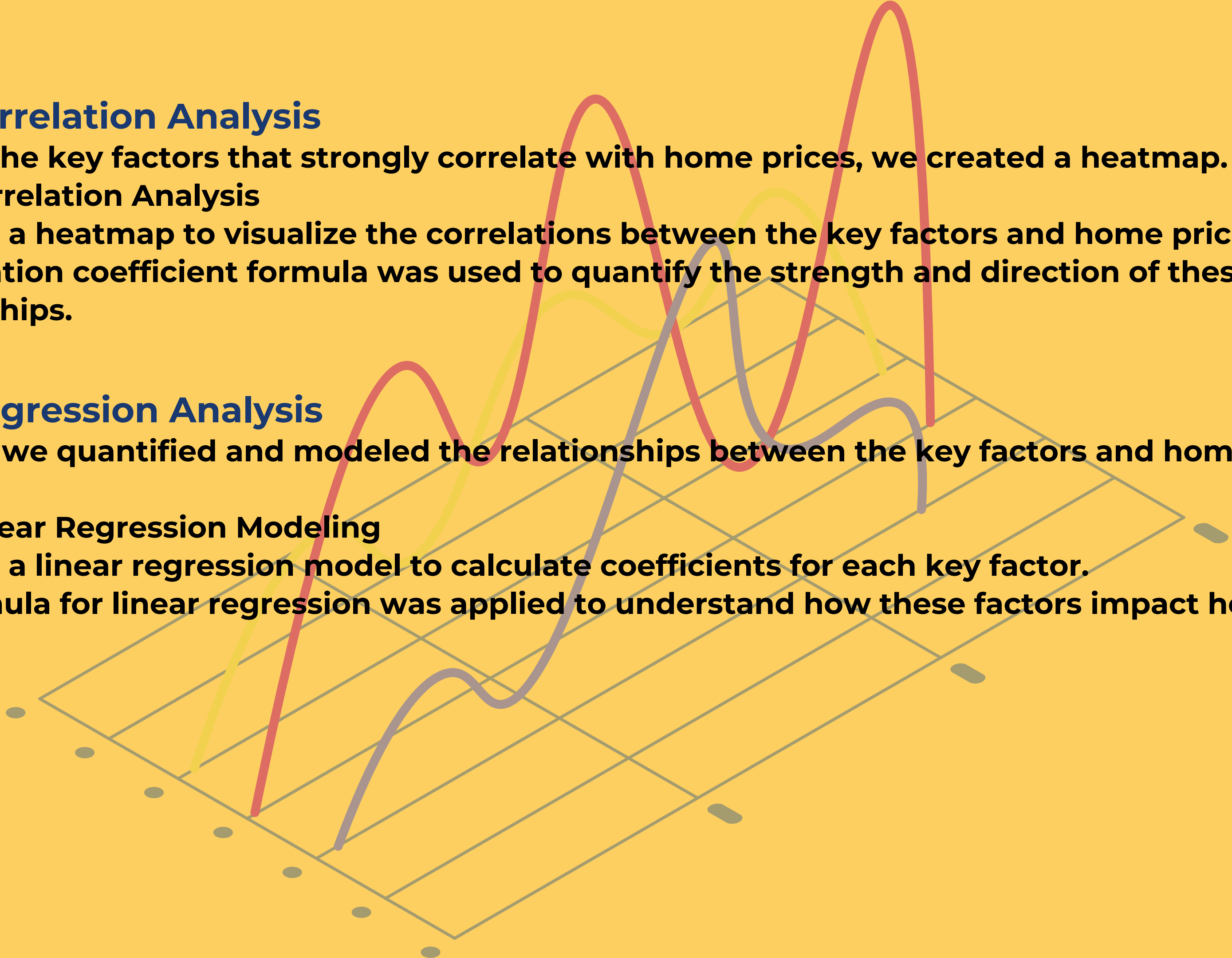
Method: Correlation Analysis

- **We used a heatmap to visualize the correlations between the key factors and home prices.**
- **A correlation coefficient formula was used to quantify the strength and direction of these relationships.**

Step 4: Regression Analysis

In this step, we quantified and modeled the relationships between the key factors and home prices.

Method: Linear Regression Modeling

- **We used a linear regression model to calculate coefficients for each key factor.**
- **The formula for linear regression was applied to understand how these factors impact home prices.**

Step 5: Data Exploration

To explore the relationships between each key factor and home prices, we turned to scatter plots.
Method: Scatter Plot Analysis
- **We generated scatter plots to visually assess how each key factor relates to home prices.**
- **These plots allowed us to evaluate the direction and strength of these relationships.**

Step 6: Visualizing Key Factors Over Time

We decided to visually present each key factor's changes over time. This step included various types of plots.
Method: Time Series Visualization
- **Line plots for time-based factors (e.g., GDP)**
- **Bar plots for comparisons (e.g., new housing projects)**
- **Box plots to explore data distribution (e.g., unemployment rates)**
- **Regression plots to model the relationship between each factor and home prices.**

SUMMERY

Regarding the accuracy of your model, the evaluation metrics provided in my code include the Mean Squared Error (MSE) and the R-squared (R2) value. These metrics quantify how well linear regression model fits the data:

The specific values for MSE and R2 were provided in your code output:

MEAN SQUARED ERROR: 224.52
R-SQUARED: 0.8077

An R2 value of 0.8077 is relatively high, indicating that themodel explains a significant portion of the variance in home prices. However, it's essential to interpret these results in the context of the specific data used.