

1. Introduction

The primary objective of this project is to analyse the impact of bank interest rates on the House Price Index (HPI) in the United States. This analysis uses historical data on interest rates and housing prices from the years 2022 to 2024. Given the importance of interest rates in the economy and their influence on the housing market, this project provides valuable insights for investors and policymakers.

The main research question is: How do changes in bank interest rates affect house prices, and what are the measurable impacts of these changes on housing market trends?

2. Data Sources

- House Price Index (HPI):

- Extracted from CSUSHPINSAprice0.csv. This file contains dates and housing price index values. The data was downloaded from the FRED (Federal Reserve Economic Data) website and covers the years 2022 to 2024.
- Structure: Two columns:
 - DATE: Date (text format).
 - CSUSHPINSA: Housing Price Index (float).
- Quality: The dataset is mostly complete, with minimal missing values or outliers.
- License: This dataset is sourced from a public economic database and is allowed for research and analysis purposes. For more details, visit [FRED Licensing Page](#).

- Prime Interest Rate:

- Extracted from PRIMEloan.csv. The data was downloaded from the FRED website and covers the years 2022 to 2024.
- Structure: Two columns:
 - DATE: Date (text format).
 - PRIME: Interest rate (float).
- Quality: The dataset has good coverage with consistent data entries.
- License: The dataset is sourced from FRED and can be used for research. For more details, visit [FRED Licensing Page](#).

3. Data Pipeline

The data pipeline was implemented using Python, utilizing libraries such as pandas and NumPy for efficient data processing.

Key steps include:

1. Data Import:

- Reading CSV files and importing them as Data Frames using pandas.

2. Cleaning and Preparation:

- Renaming columns for consistency.
- Converting date formats to datetime to facilitate merging and analysis.

3. Merging Data:

- Combining the datasets using the date column.
- Merge method: Outer join to include all dates and ensure no data is lost during merging.

4. Handling Missing Values:

- Removing records with missing housing price index values to maintain data integrity.

5. Managing Errors and Changing Data:

- Implemented error handling for missing or inconsistent data entries.
- Automated alerts for changes in data formats to ensure seamless updates.

6. Output Creation:

- The combined data was saved as a CSV file for further analysis.

The pipeline is robust to handle new incoming data and any format changes, ensuring continuous processing without manual intervention.

4. Analysis and Results

• House Price Index Trends:

- The HPI showed a steady increase over time, with notable fluctuations influenced by interest rate changes.
- A notable trend was the relationship between rising interest rates and a decrease in HPI growth rate.

- **Interest Rate Trends:**

- Interest rates varied across periods, directly affecting demand and housing prices. When interest rates were higher, housing price growth showed signs of slowing.

Below is a chart depicting the trends in the House Price Index and Interest Rate

5. Result and Limitations

The output of the data pipeline is a merged dataset containing both the HPI and interest rate data for the years 2022 to 2024.

- **Output Data Structure and Quality:**

- The output dataset contains three columns: DATE, CSUSHPINSA, and PRIME.
- The data format is CSV, chosen for its simplicity and compatibility with various data analysis tools.

- **Critical Reflection on Data:**

- The data is largely complete; however, some fluctuations may be attributed to external economic factors not covered in this dataset.
- Potential issues include the limited time frame (2022-2024), which may not capture long-term trends or rare economic events.

Adding relevant figures and tables that visually depict trends in interest rates and the HPI over the analysis period can greatly enhance reader understanding. Moreover, including a flowchart of the data pipeline structure will make the process more transparent and engaging.

Below is an example of a summary table showcasing key statistics:

Metric	Value
Average Interest Rate	3.5%
Average HPI	210.45
Max HPI	225.30
Min Interest Rate	2.0%

