Data collection:

Collecting the relevant data from the website : fred.stlouisfed.org/series/CSUSHPISA.  
1. Interest rate

2. Employment rate

3. GDP growth

4. House Units

Merging them using pandas.

Import pandas as pd

gdp\_data = pd.read\_csv('C:\\Users\\ELCOT\\Desktop\\HomeLLc\\gdp.csv')

employment\_data = pd.read\_csv('C:\\Users\\ELCOT\\Desktop\\HomeLLc\\employment.csv')

interest\_data = pd.read\_csv('C:\\Users\\ELCOT\\Desktop\\HomeLLc\\interest.csv')

house\_data=pd.read\_csv('C:\\Users\\ELCOT\\Desktop\\HomeLLc\\HOUST.csv')

merged\_data = pd.merge(gdp\_data, employment\_data, on='DATE', how='inner')

merged\_data = pd.merge(merged\_data, interest\_data, on='DATE', how='inner')

merged\_data = pd.merge(merged\_data, house\_data, on='DATE', how='inner')

merged\_data.to\_csv('C:\\Users\\ELCOT\\Desktop\\HomeLLc\\combined\_data.csv', index=False)

Findings:

1. Data Cleaning and Preprocessing:

No missing values were found in the dataset.  
  
print(data.isnull().sum())

data.fillna(data.mean(), inplace=True)

The date column needs to be formatted correctly for further analysis.

2. Exploratory Data Analysis (EDA):

GDP:

The distribution of GDP is right-skewed, indicating a positive economic trend.

There is a positive correlation between GDP and the number of house units.

Employment Rate:

The distribution of the employment rate shows a normal distribution.

A scatter plot suggests a positive correlation between the employment rate and the number of house units.

Interest Rate:

The distribution of interest rates appears to be relatively uniform.

The scatter plot indicates a weak correlation between interest rates and the number of house units.

Multivariate Analysis:

A pair plot was used to visualize relationships among GDP, employment rate, interest rate, and house units.  
  
# House units data

plt.figure(figsize=(12, 6))

plt.plot(data['DATE'], data['House Units'], label='House Units')

plt.xlabel('Date')

plt.ylabel('House Units')

plt.title('House Units Over Time')

plt.legend()

plt.show()

plt.scatter(data['GDP'], data['House Units'])

plt.xlabel('GDP')

plt.ylabel('House Units')

plt.title('Scatter Plot: GDP vs. House Units')

plt.show()

# Employment Rate uni - bi and multivariate

plt.figure(figsize=(10, 6))

sns.histplot(data['Emplyment rate'], kde=True, bins=30, color='blue')

plt.title('Distribution of Employment Rate')

plt.show()

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Emplyment rate', y='House Units', data=data, color='green')

plt.title('Employment Rate vs House Units')

plt.show()

sns.pairplot(data[['Emplyment rate', 'GDP', 'Interest rate', 'House Units']])

plt.show()

# Interest Rate uni- bi and multi variate analysis

plt.figure(figsize=(10, 6))

sns.histplot(data['Interest rate'], kde=True, bins=30, color='red')

plt.title('Distribution of Interest Rate')

plt.show()

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Interest rate', y='House Units', data=data, color='purple')

plt.title('Interest Rate vs House Units')

plt.show()

sns.pairplot(data[['Interest rate', 'GDP', 'Emplyment rate', 'House Units']])

plt.show()

3. Statistical Significance:

No specific statistical tests were mentioned in the provided information. Further tests may be required based on the context and goals of the analysis.

# lag creation

data['GDP\_Lag1'] = data['GDP'].shift(1)

data['EmploymentRate\_Lag1'] = data['Emplyment rate'].shift(1)

data['InterestRate\_Lag1'] = data['Interest rate'].shift(1)

data.dropna(inplace=True)

4. Model Building (if applicable):

No information was provided on model building in the given context. Depending on the goals, regression models could be explored to predict house prices based on economic indicators.  
#Model building

model = RandomForestRegressor(n\_estimators=100, random\_state=42)

model.fit(X\_train, y\_train)

5. Feature Importance:

If a predictive model was built, feature importance analysis would reveal which factors have the most significant impact on house prices.  
#Evaluation

y\_pred = model.predict(X\_test)

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print(f'Mean Squared Error: {mse}')

print(f'R-squared: {r2}'

# Final feature

feature\_importance = model.feature\_importances\_

feature\_names = X.columns

feature\_importance\_df = pd.DataFrame({'Feature': feature\_names, 'Importance': feature\_importance})

feature\_importance\_df.sort\_values(by='Importance', ascending=False, inplace=True)

print(feature\_importance\_df)

plt.barh(feature\_importance\_df['Feature'], feature\_importance\_df['Importance'])

plt.xlabel('Importance')

plt.title('Feature Importance')

plt.show()

6. Business Suggestions/Solutions:

Recommendations for policy changes or investment strategies based on the analysis and findings.

Conclusion:

Summarize the key insights obtained from the analysis.