

## HOME.LLC SCREENING TASK

AYUSHMANT BHARADWAJ

9958493981

ayushmant.bharadwaj@gmail.com

I have broken down the task into 4 different parts:

### 1) Download datasets showing the value of factors across the last 20 years.

I have used datasets from <https://fred.stlouisfed.org/> for 4 different factors: Federal Funds Effective Rate, Unemployment Rates, CPI and Average Cost of Housing Inventory. They have been uploaded in the datasets folder for reference.

### 2) Clean and process the data: Make sure all factors align in dates, use mean to fill initial empty values and then use linear interpolation to fill the remaining empty ones.

```
9
10 # Align all datasets to the date range of home price index
11 start_date = '2004-01-01'
12 end_date = '2024-01-01'
13
14 cpi_aligned = cpi[start_date:end_date]
15 federal_funds_rate_aligned = federal_funds_rate[start_date:end_date]
16 housing_inventory_aligned = housing_inventory[start_date:end_date]
17 unemployment_rate_aligned = unemployment_rate[start_date:end_date]
18
19 # Resample CPI to monthly data (forward filling the values)
20 cpi_monthly = cpi_aligned.resample('M').ffill()
21
22 # Resample Federal Funds Effective Rate to monthly data (taking the average for each month)
23 federal_funds_rate_monthly = federal_funds_rate_aligned.resample('M').mean()
24
25 # Resample Housing Inventory Estimate to monthly data (forward filling the values)
26 housing_inventory_monthly = housing_inventory_aligned.resample('M').ffill()
27
28 # Combine all datasets into a single DataFrame
29 combined_data = pd.concat([
30     home_price_index,
31     cpi_monthly,
32     federal_funds_rate_monthly,
33     housing_inventory_monthly,
34     unemployment_rate_aligned
35 ], axis=1)
```

Here is a snippet from the code file **data\_processing.py**, which shows how I've aligned all dates to the Home Price Index dataset. I have resampled CPI and Federal Funds Effective rate to monthly using the mean of each month's values, so that it fits the dates. Then, I have combined all datasets to a single dataset.

```

46 # Fill initial NaN values by taking the mean for the month across all years
47 combined_data = combined_data.apply(lambda x: x.fillna(x.groupby(x.index.month).transform('mean')))
48
49 # Interpolate missing values using linear interpolation after the first non-NaN value in each column
50 combined_data = combined_data.apply(lambda x: x.interpolate(method='linear', limit_area='inside'))
51
52 # Save the combined data to a CSV file
53 combined_data.to_csv('combined_data_last_20_years.csv')

```

Here is another snippet from the same file. I have applied mean to fill some initial null values, and then used linear interpolation to fill the other values. Then, I have combined all values into a single data file.

```

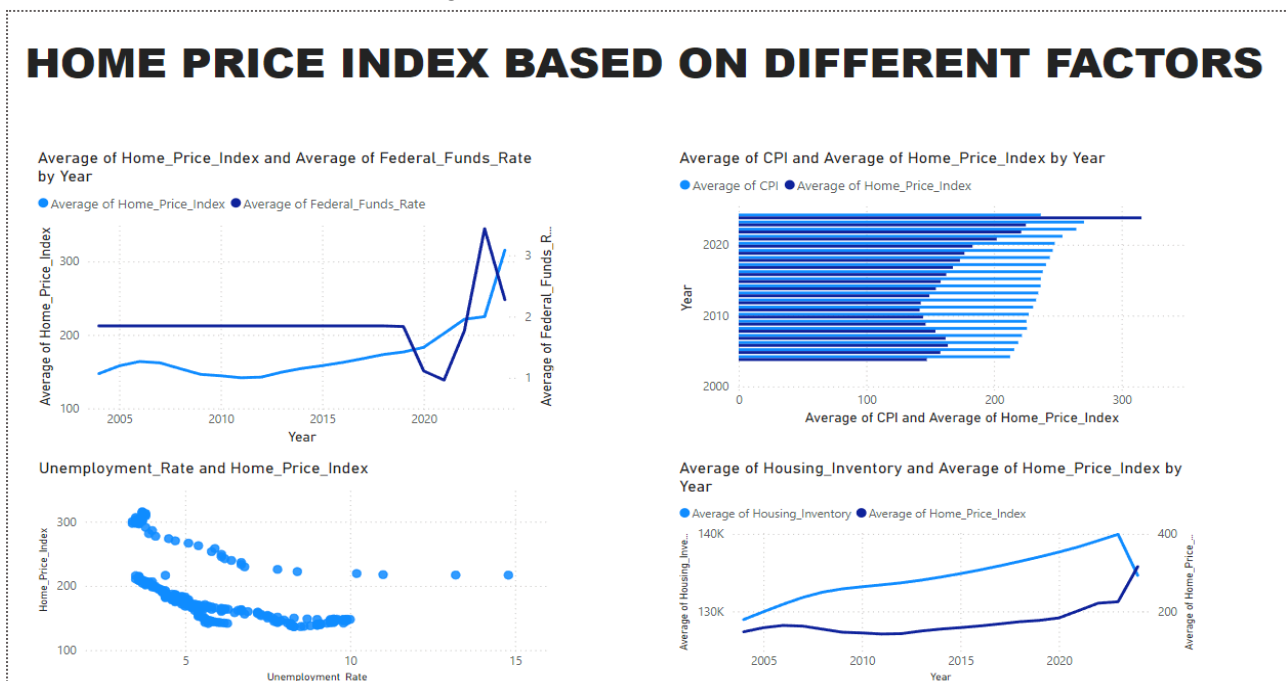
1  import pandas as pd
2
3  # Load the combined data
4  combined_data = pd.read_csv('combined_data_last_20_years.csv', parse_dates=['DATE'], index_col='DATE')
5
6  # Filter the DataFrame to keep only data from the last 20 years
7  start_date = '2004-01-01'
8  end_date = '2024-01-01'
9  filtered_data = combined_data.loc[start_date:end_date]
10 |
11 # Save the filtered data to a new CSV file
12 filtered_data.to_csv('combined_data_last_20_years_filtered.csv')

```

Here is a snippet from the file **extra\_dates.py**. Here, I have trimmed the dataset down so that it gives me values for only the last 20 years.

### 3) Use Power BI to graph the factors and show the correlation between the CASE-SHILLER Index and the other variables.

Here is a Power BI dashboard showing the correlation of the Home Price Index with the factors:



This dashboard has been uploaded in the GitHub repository, so that it can be used interactively for different years.

#### 4) Use linear regression to show the coefficient by which the factors affect the CASE-SHILLER Index.

From the dashboard in the previous step, we can conclude that the CASE-SHILLER index is:

1. Directly proportional to CPI.
2. Directly proportion to Federal Funding Effective Rates.
3. Directly proportional to Average cost of Housing Inventory.
4. Inversely proportional to Unemployment Rates.

We can confirm these relationships by performing linear regression, taking the CASE-SHILLER index as the dependent variable(y), and the other factors as independent variables(x).

Here's the output of the code file **coefficient\_finder.py**, which is a Python file performing linear regression using the scikit-learn library.

```
● PS C:\Users\ayush\Pictures\Home LLC> & 'c:\Python312\python.exe' 'c:\Users\ayush\.vscode\extensions\ms-python
.debugpy-2024.6.0-win32-x64\bundled\libs\debugpy\adapter\..\..\debugpy\launcher' '60504' '--' 'c:\Users\ayush\
Pictures\Home LLC\coefficient_finder.py'
Coefficient for CPI: 0.012556262218143375
Coefficient for Federal_Funds_Rate: 1.116624047469781
Coefficient for Housing_Inventory: 5.483642054968281e-05
Coefficient for Unemployment_Rate: -12.952362543363542
○ PS C:\Users\ayush\Pictures\Home LLC>
```

The coefficients of CPI, Federal\_Funds\_Rate and Housing\_Inventory are **positive**, and the coefficient of Unemployment\_Rate is **negative**, hence confirming our findings.