

```
In [1]: ▶ import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [2]: ▶ file = pd.read_excel(r"E:\assignment\New folder (2)\half_ready.xlsx")
```

```
In [3]: ▶ file
```

Out[3]:

	observation_date	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_Pc
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921e+0
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433e+0
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055e+0
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807e+0
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272e+0
...
280	2023-05-01	302.566	5.06	1534	337.473	16818.5	3.7	2.086128e+0
281	2023-06-01	304.593	5.08	1492	337.336	16809.5	3.6	2.087069e+0
282	2023-07-01	306.767	5.12	1334	334.576	16796.9	3.5	2.087792e+0
283	2023-08-01	309.155	5.33	1370	333.980	16799.7	3.8	2.089066e+0
284	2023-09-01	311.175	5.33	1478	332.224	16804.8	3.8	2.091172e+0

285 rows × 9 columns



Data Collection

```
In [4]: file1 = pd.read_excel(r"E:\assignment\New folder (2)\population.xls")
```

```
In [5]: file['observation_date'] = pd.to_datetime(file['observation_date'])
file['year'] = file['observation_date'].dt.year
file['month'] = file['observation_date'].dt.month
file1['observation_date'] = pd.to_datetime(file1['observation_date'])
file1['year'] = file1['observation_date'].dt.year
merged_df = pd.merge(file, file1, on='year')
final = pd.DataFrame(merged_df.drop(['observation_date_y'], axis=1))
```

```
In [ ]:
```

```
In [ ]:
```

```
In [6]: file2 = pd.read_excel(r"E:\assignment\New folder (2)\SUBSIDIES.xls")
```

```
In [7]: file2['observation_date'] = pd.to_datetime(file2['observation_date'])
file2['year'] = file2['observation_date'].dt.year
final1 = pd.merge(final, file2, on='year')
```

In [8]: `final1`

Out[8]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921e
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433e
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055e
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807e
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272e
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707e
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536e
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312e
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219e
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245e

276 rows × 14 columns



In [9]: `file3 = pd.read_csv(r"E:\assignment\New folder (2)\united-states-inflation-rate-cpi.csv")`

In [10]: `file3['date'] = pd.to_datetime(file3['date'])`

In [11]: `file3['year'] = file3['date'].dt.year
final2 = pd.merge(final1, file3, on='year')`

In [12]: `file4 = pd.read_csv(r"E:\assignment\New folder (2)\real gdp per capita.csv")`

```
In [13]: file4['observation_date_x'] = pd.to_datetime(file4['observation_date_x'])
```

```
In [14]: file4
```

Out[14]:

	observation_date_x	GDP
0	2000-01-01	49335
1	2000-04-01	50109
2	2000-07-01	50024
3	2000-10-01	50190
4	2001-01-01	49911
...
90	2022-07-01	65462
91	2022-10-01	65783
92	2023-01-01	66078
93	2023-04-01	66341
94	2023-07-01	67083

95 rows × 2 columns

```
In [15]: final3 = final2.merge(file4, how='left', on='observation_date_x')
```

```
In [16]: final3
```

Out[16]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 18 columns



```
In [17]: final4 = final3.drop(['date', 'observation_date'], axis=1)
```

In [18]: `final4`

Out[18]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 16 columns



In []:

In [19]: `file5 = pd.read_excel(r"E:\assignment\New folder (2)\mort.xlsx")`

In [20]: `file5['Date'] = pd.to_datetime(file5['Date'])`
`file5['year'] = file5['Date'].dt.year`

```
In [22]: final5 = pd.merge(final4, file5, left_on='year', right_on='year')
```

```
In [24]: final6 = final5.drop(['Date'], axis=1)
```

```
In [25]: final6
```

Out[25]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 17 columns



```
In [ ]:
```

```
In [26]: file0 = pd.read_csv(r"E:\assignment\New folder\EMPLOYMENT RATE 15-64 ALL PERSON.csv")
```

```
In [27]: ▶ file0.head(3)
```

Out[27]:

	observation_date_x	Employment Rate
0	1/1/2000	74.264657
1	2/1/2000	74.283062
2	3/1/2000	74.242355

```
In [28]: ▶ file0['observation_date_x'] = pd.to_datetime(file0['observation_date_x'])
```

```
In [29]: ▶ final7 = final6.merge(file0, how='left', on='observation_date_x')
```


In [30]: `final7`

Out[30]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 18 columns



In []:

In [31]: `file6 = pd.read_csv(r"E:\assignment\New folder\Employment Rate Aged 25-54 Males.csv")`
`file7 = pd.read_csv(r"E:\assignment\New folder\Employment Rate Aged 25-54 Females.csv")`

```
In [32]: file6['observation_date_x'] = pd.to_datetime(file6['observation_date_x'])
file6['year'] = file6['observation_date_x'].dt.year

file7['observation_date_x'] = pd.to_datetime(file7['observation_date_x'])
file7['year'] = file7['observation_date_x'].dt.year
```

```
In [33]: final8 = final7.merge(file6, how='left', on='observation_date_x')

final9 = final8.merge(file7, how='left', on='observation_date_x')
```

```
In [34]: final10 = final9.drop(['year_y', 'year', 'year_x'], axis=1)
```

```
In [35]: final10
```

Out[35]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921e
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433e
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055e
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807e
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272e
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707e
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536e
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312e
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219e
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245e

276 rows × 9 columns



In []: ▶

```
In [36]: ▶ file10 = pd.read_csv(r"E:\assignment\New folder (4)\labour force participation rate.csv")
file11 = pd.read_csv(r"E:\assignment\New folder (4)\monthly supply of new houses.csv")
file12 = pd.read_csv(r"E:\assignment\New folder (4)\personal saving rate.csv")
file13 = pd.read_excel(r"E:\assignment\New folder (4)\property tax.xls")
```

```
In [37]: ▶ file10['observation_date_x'] = pd.to_datetime(file10['observation_date_x'])
file11['observation_date_x'] = pd.to_datetime(file11['observation_date_x'])
file12['observation_date_x'] = pd.to_datetime(file12['observation_date_x'])
file13['observation_date_x'] = pd.to_datetime(file13['observation_date_x'])
```

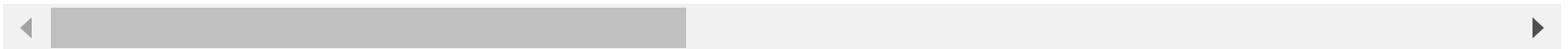
```
In [38]: ▶ final11 = final10.merge(file10, how='left', on='observation_date_x')
final12 = final11.merge(file11, how='left', on='observation_date_x')
final13 = final12.merge(file12, how='left', on='observation_date_x')
```

In [39]: `final13`

Out[39]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 22 columns



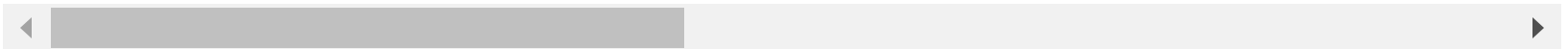
In [40]: `final14 = final13.merge(file13, how='left', on='observation_date_x')`

In [41]:  final14


Out[41]:


	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 23 columns



In []: 

In [42]:  `file14 = pd.read_csv(r"E:\assignment\New folder (4)\real_money_supply.csv")
file15 = pd.read_csv(r"E:\assignment\New folder (4)\velocity money m2.csv")`

In [43]:  `file14['observation_date_x'] = pd.to_datetime(file14['observation_date_x'])
file15['observation_date_x'] = pd.to_datetime(file15['observation_date_x'])`

```
In [44]: ► final15 = final14.merge(file14, how='left', on='observation_date_x')  
        ► final16 = final15.merge(file15, how='left', on='observation_date_x')
```

```
In [ ]: ►
```

```
In [45]: ► file16 = pd.read_csv(r"E:\assignment\New folder (5)\cpi ubran house.csv")  
        ► file17 = pd.read_csv(r"E:\assignment\New folder (5)\government expenditure.csv")
```

```
In [48]: ► file16['observation_date_x'] = pd.to_datetime(file16['observation_date_x'])  
        ► file17['observation_date_x'] = pd.to_datetime(file17['observation_date_x'])
```

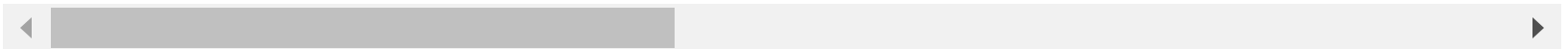
```
In [50]: ► final17 = final16.merge(file16, how='left', on='observation_date_x')  
        ► final18 = final17.merge(file17, how='left', on='observation_date_x')
```

In [51]: `final18`

Out[51]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 27 columns



In []:

Missing value

In [52]: `final19 = final18.fillna(method='ffill')`

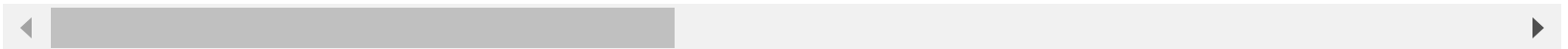
In [62]: `final19.rename(columns={' Inflation': 'Inflation'}, inplace=True)`

In [63]: `final19`

Out[63]:

	observation_date_x	House_Price_Index	Interest_Rate	No_of_property_Introduced	PPI	Income	Unemployment	Working_
0	2000-01-01	100.551	5.45	1574	144.100	9799.9	4.0	1.782921€
1	2000-02-01	101.339	5.73	1677	144.700	9837.9	4.1	1.783433€
2	2000-03-01	102.127	5.85	1704	145.400	9864.0	4.0	1.784055€
3	2000-04-01	102.922	6.02	1610	145.600	9913.7	3.8	1.785807€
4	2000-05-01	103.678	6.27	1682	144.900	9954.5	4.0	1.787272€
...
271	2022-08-01	301.473	2.33	1355	342.753	16161.4	3.7	2.073707€
272	2022-09-01	299.353	2.56	1438	336.464	16184.9	3.5	2.074536€
273	2022-10-01	298.873	3.08	1348	333.796	16223.5	3.7	2.074312€
274	2022-11-01	298.269	3.78	1543	330.369	16229.6	3.6	2.075219€
275	2022-12-01	297.413	4.10	1390	326.449	16265.1	3.5	2.075245€

276 rows × 27 columns



In [64]: `final19.to_csv('final19.csv', index=False)`

In []: