Dataset_Building

December 29, 2022

[1]:
```
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
import pandas as pd
import numpy as np
```

[3]:
```
monthly_df = pd.read_csv('Natural Gas Futures Historical Data Monthly.csv')
```

[4]:
```
monthly_df
```

[4]:

	Date	Price	Open	High	Low	Vol.	Change %
0	01/01/1995	1.354	1.760	1.795	1.250	332.65K	-21.51%
1	02/01/1995	1.483	1.345	1.555	1.341	270.04K	9.53%
2	03/01/1995	1.685	1.485	1.725	1.417	273.23K	13.62%
3	04/01/1995	1.662	1.700	1.745	1.585	224.41K	-1.36%
4	05/01/1995	1.718	1.670	1.824	1.635	256.06K	3.37%
..	…	…	…	…	…	…	
331	08/01/2022	9.127	7.925	10.005	7.559	86.29K	15.15%
332	09/01/2022	7.065	9.194	9.454	6.562	1.30M	-22.59%
333	10/01/2022	6.355	6.800	7.188	4.750	2.01M	-10.05%
334	11/01/2022	6.930	6.270	7.604	5.614	2.01M	9.05%
335	12/01/2022	5.079	6.978	7.210	4.779	87.38K	-26.71%

[336 rows x 7 columns]

[5]:
```
monthly_df['Date'] = pd.to_datetime(monthly_df['Date'])
```

[6]:
```
df = monthly_df[['Date','Price']]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 336 entries, 0 to 335
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    336 non-null    datetime64[ns]
 1   Price   336 non-null    float64
```

```
dtypes: datetime64[ns](1), float64(1)
memory usage: 5.4 KB
```

[7]:
```
df = df.rename(columns={'Price':'Natural Gas Price'})
df
```

[7]:

	Date	Natural Gas Price
0	1995-01-01	1.354
1	1995-02-01	1.483
2	1995-03-01	1.685
3	1995-04-01	1.662
4	1995-05-01	1.718
..	…	…
331	2022-08-01	9.127
332	2022-09-01	7.065
333	2022-10-01	6.355
334	2022-11-01	6.930
335	2022-12-01	5.079

[336 rows x 2 columns]

[8]:
```
crude_oil = pd.read_csv('Crude Oil WTI Futures Historical Data Monthly.csv')
crude_oil['Date'] = pd.to_datetime(crude_oil['Date'])
crude_oil
```

[8]:

	Date	Price	Open	High	Low	Vol.	Change %
0	1995-01-01	18.39	17.65	18.85	17.27	791.63K	3.55%
1	1995-02-01	18.49	18.38	18.98	18.13	653.68K	0.54%
2	1995-03-01	19.17	18.52	19.35	17.82	950.55K	3.68%
3	1995-04-01	20.38	19.12	20.82	18.90	760.45K	6.31%
4	1995-05-01	18.89	20.49	20.80	18.45	1.00M	-7.31%
..	…	…	…	…	…	…	…
331	2022-08-01	89.55	98.46	98.65	85.73	6.61M	-9.20%
332	2022-09-01	79.49	88.83	90.39	76.25	5.81M	-11.23%
333	2022-10-01	86.53	81.02	93.64	80.87	5.33M	8.86%
334	2022-11-01	80.56	86.17	93.74	73.61	5.68M	-6.90%
335	2022-12-01	79.81	80.56	83.33	70.11	NaN	-0.93%

[336 rows x 7 columns]

[9]:
```
crude_oil_df = crude_oil[['Date','Price']]
crude_oil_df = crude_oil_df.rename(columns={'Price':'Crude Oil Price'})
crude_oil_df
```

[9]:

	Date	Crude Oil Price
0	1995-01-01	18.39
1	1995-02-01	18.49

```
2   1995-03-01            19.17
3   1995-04-01            20.38
4   1995-05-01            18.89
..         …                …
331 2022-08-01            89.55
332 2022-09-01            79.49
333 2022-10-01            86.53
334 2022-11-01            80.56
335 2022-12-01            79.81

[336 rows x 2 columns]
```

[10]:
```
new_data = df.merge(crude_oil_df,on='Date')
new_data
```

[10]:
```
           Date  Natural Gas Price  Crude Oil Price
0   1995-01-01              1.354            18.39
1   1995-02-01              1.483            18.49
2   1995-03-01              1.685            19.17
3   1995-04-01              1.662            20.38
4   1995-05-01              1.718            18.89
..         …                …                …
331 2022-08-01              9.127            89.55
332 2022-09-01              7.065            79.49
333 2022-10-01              6.355            86.53
334 2022-11-01              6.930            80.56
335 2022-12-01              5.079            79.81

[336 rows x 3 columns]
```

[11]:
```
dollar_index_df = pd.read_csv('US Dollar Index Historical Data.csv')
dollar_index_df['Date'] = pd.to_datetime(dollar_index_df['Date'])
dollar_index_df = dollar_index_df[['Date','Price']]
dollar_index_df = dollar_index_df.rename(columns={'Price':'US Dollar Index'})
dollar_index_df
```

[11]:
```
           Date  US Dollar Index
0   1995-01-01            87.99
1   1995-02-01            85.93
2   1995-03-01            81.90
3   1995-04-01            81.80
4   1995-05-01            82.57
..         …                …
331 2022-08-01           108.85
332 2022-09-01           112.12
333 2022-10-01           111.54
334 2022-11-01           105.87
```

```
335 2022-12-01           104.06

[336 rows x 2 columns]
```

[12]:
```
new_data = new_data.merge(dollar_index_df,on='Date')
new_data
```

[12]:

	Date	Natural Gas Price	Crude Oil Price	US Dollar Index
0	1995-01-01	1.354	18.39	87.99
1	1995-02-01	1.483	18.49	85.93
2	1995-03-01	1.685	19.17	81.90
3	1995-04-01	1.662	20.38	81.80
4	1995-05-01	1.718	18.89	82.57
..	…	…	…	…
331	2022-08-01	9.127	89.55	108.85
332	2022-09-01	7.065	79.49	112.12
333	2022-10-01	6.355	86.53	111.54
334	2022-11-01	6.930	80.56	105.87
335	2022-12-01	5.079	79.81	104.06

[336 rows x 4 columns]

[13]:
```
GDP = pd.read_csv('GDP.csv')
GDP = GDP.tail(112)
GDP = GDP.reset_index()
GDP = GDP.drop(['index'],axis=1)
GDP
```

[13]:

	DATE	GDP
0	1994-10-01	7455.288
1	1995-01-01	7522.289
2	1995-04-01	7580.997
3	1995-07-01	7683.125
4	1995-10-01	7772.586
..	…	…
107	2021-07-01	23550.420
108	2021-10-01	24349.121
109	2022-01-01	24740.480
110	2022-04-01	25248.476
111	2022-07-01	25723.941

[112 rows x 2 columns]

[14]:
```
id = np.arange(0,112)
df = pd.DataFrame(id, columns = ['id'])
```

[15]:
```
GDP = pd.concat([df['id'],GDP['GDP'],], axis=1,ignore_index=False)
GDP_yearly = GDP.append([GDP]*2,ignore_index=True).sort_values(by='id')
GDP_yearly.reset_index(inplace= True)
GDP_yearly = GDP_yearly.drop(['index'],axis=1)
```

[16]:
```
GDP_yearly
```

[16]:
```
       id        GDP
0       0   7455.288
1       0   7455.288
2       0   7455.288
3       1   7522.289
4       1   7522.289
..    …         …
331   110   25248.476
332   110   25248.476
333   111   25723.941
334   111   25723.941
335   111   25723.941

[336 rows x 2 columns]
```

[17]:
```
final_data = pd.concat([new_data,GDP_yearly['GDP']], axis=1)
final_data
```

[17]:
```
           Date  Natural Gas Price  Crude Oil Price  US Dollar Index        GDP
0    1995-01-01              1.354            18.39            87.99   7455.288
1    1995-02-01              1.483            18.49            85.93   7455.288
2    1995-03-01              1.685            19.17            81.90   7455.288
3    1995-04-01              1.662            20.38            81.80   7522.289
4    1995-05-01              1.718            18.89            82.57   7522.289
..          …                  …                …                …         …
331 2022-08-01              9.127            89.55           108.85  25248.476
332 2022-09-01              7.065            79.49           112.12  25248.476
333 2022-10-01              6.355            86.53           111.54  25723.941
334 2022-11-01              6.930            80.56           105.87  25723.941
335 2022-12-01              5.079            79.81           104.06  25723.941

[336 rows x 5 columns]
```

[18]:
```
df = pd.read_excel("Natural_Gas_Overview.xlsx")
df = df.iloc[9:]
df
```

[18]:
```
    U.S. Energy Information Administration                    Unnamed: 1  \
9                                    Month  Natural Gas Gross Withdrawals
10                                     NaN           (Billion Cubic Feet)
```

```
11                    1973-01-01 00:00:00                 Not Available
12                    1973-02-01 00:00:00                 Not Available
13                    1973-03-01 00:00:00                 Not Available
..                                    …                             …
603                   2022-05-01 00:00:00                    3667.147
604                   2022-06-01 00:00:00                    3556.574
605                   2022-07-01 00:00:00                    3689.978
606                   2022-08-01 00:00:00                    3695.024
607                   2022-09-01 00:00:00                    3617.812

                                Unnamed: 2            Unnamed: 3  \
9    Natural Gas Marketed Production (Wet)       NGPL Production
10                    (Billion Cubic Feet)  (Billion Cubic Feet)
11                                    1948                    79
12                                    1962                    79
13                                    1907                    77
..                                     …                     …
603                             3295.871               265.773
604                             3214.637               259.484
605                             3330.463               275.788
606                             3345.322               269.945
607                             3262.022               264.538

                      Unnamed: 4                 Unnamed: 5  \
9    Natural Gas Production (Dry)  Supplemental Gaseous Fuels
10           (Billion Cubic Feet)        (Billion Cubic Feet)
11                           1869               Not Available
12                           1883               Not Available
13                           1830               Not Available
..                            …                           …
603                     3030.098                        5.61
604                     2955.154                       4.034
605                     3054.675                       5.887
606                     3075.377                       5.955
607                     2997.484                       4.423

              Unnamed: 6            Unnamed: 7               Unnamed: 8  \
9     Natural Gas Imports   Natural Gas Exports  Natural Gas Net Imports
10   (Billion Cubic Feet)  (Billion Cubic Feet)     (Billion Cubic Feet)
11                 92.694                 5.808                   86.887
12                  83.87                 6.079                   77.791
13                 91.581                 4.021                    87.56
..                    …                     …                        …
603               230.744               614.309                -383.564
604               228.726               550.542                -321.816
605               257.052               556.495                -299.443
606               235.709               554.813                -319.104
```

```
607               233.781               525.904                -292.123

                             Unnamed: 9                 Unnamed: 10  \
9    Natural Gas Net Storage Withdrawals  Natural Gas Balancing Item
10                   (Billion Cubic Feet)       (Billion Cubic Feet)
11                          Not Available              Not Available
12                          Not Available              Not Available
13                          Not Available              Not Available
..                                    …                          …
603                             -403.136                    -4.164
604                             -323.613                    11.171
605                             -180.038                    15.668
606                             -206.239                     4.941
607                             -436.689                    33.811

                Unnamed: 11
9    Natural Gas Consumption
10      (Billion Cubic Feet)
11                      2348
12                      2126
13                      2015
..                       …
603                2244.844
604                2324.929
605                 2596.75
606                2560.929
607                2306.907

[599 rows x 12 columns]
```

[19]:
```
# df = df.reset_index()
# df = df.drop(['index'],axis=1)
# df
```

[20]:
```
df = df.iloc[2:]
df = df.drop(['Unnamed: 1','Unnamed: 2','Unnamed: 3','Unnamed: 5','Unnamed:␣
↪6','Unnamed: 7',
              'Unnamed: 9','Unnamed: 10'],axis=1)
```

[21]:
```
df = df.rename(columns={'U.S. Energy Information Administration':
↪'Date','Unnamed: 4':'Natural Gas Production',
                        'Unnamed: 11':'Natural Gas Consumption','Unnamed: 8':
↪'Net Imports'})
df['Date'] = pd.to_datetime(df['Date'])
df = df.tail(336)
df = df.reset_index()
df = df.drop('index',axis=1)
```

```
df
```

[21]:

	Date	Natural Gas Production	Net Imports	Natural Gas Consumption
0	1994-10-01	1560.033	208.822	1514.35
1	1994-11-01	1603.862	206.435	1748.263
2	1994-12-01	1663.964	227.778	2119.285
3	1995-01-01	1598.615	239.7	2432.305
4	1995-02-01	1425.792	222.862	2230.289
..	…	…	…	…
331	2022-05-01	3030.098	-383.564	2244.844
332	2022-06-01	2955.154	-321.816	2324.929
333	2022-07-01	3054.675	-299.443	2596.75
334	2022-08-01	3075.377	-319.104	2560.929
335	2022-09-01	2997.484	-292.123	2306.907

[336 rows x 4 columns]

[22]:
```
final_df = pd.concat([final_data,df[['Natural Gas Production','Net␣
↪Imports','Natural Gas Consumption']]], axis=1)
final_df
```

[22]:

	Date	Natural Gas Price	Crude Oil Price	US Dollar Index	\
0	1995-01-01	1.354	18.39	87.99	
1	1995-02-01	1.483	18.49	85.93	
2	1995-03-01	1.685	19.17	81.90	
3	1995-04-01	1.662	20.38	81.80	
4	1995-05-01	1.718	18.89	82.57	
..	…	…	…	…	
331	2022-08-01	9.127	89.55	108.85	
332	2022-09-01	7.065	79.49	112.12	
333	2022-10-01	6.355	86.53	111.54	
334	2022-11-01	6.930	80.56	105.87	
335	2022-12-01	5.079	79.81	104.06	

	GDP	Natural Gas Production	Net Imports	Natural Gas Consumption
0	7455.288	1560.033	208.822	1514.35
1	7455.288	1603.862	206.435	1748.263
2	7455.288	1663.964	227.778	2119.285
3	7522.289	1598.615	239.7	2432.305
4	7522.289	1425.792	222.862	2230.289
..	…	…	…	…
331	25248.476	3030.098	-383.564	2244.844
332	25248.476	2955.154	-321.816	2324.929
333	25723.941	3054.675	-299.443	2596.75
334	25723.941	3075.377	-319.104	2560.929
335	25723.941	2997.484	-292.123	2306.907

```
[336 rows x 8 columns]
```

[23]:
```
final_df[['Natural Gas Production','Net Imports','Natural Gas Consumption']] =␣
↪final_df[
   ['Natural Gas Production','Net Imports','Natural Gas Consumption']].mul(0.
↪01)
```

[24]:
```
final_df
```

[24]:

	Date	Natural Gas Price	Crude Oil Price	US Dollar Index	GDP	Natural Gas Production	Net Imports	Natural Gas Consumption
0	1995-01-01	1.354	18.39	87.99	7455.288	15.60033	2.08822	15.1435
1	1995-02-01	1.483	18.49	85.93	7455.288	16.03862	2.06435	17.48263
2	1995-03-01	1.685	19.17	81.90	7455.288	16.63964	2.27778	21.19285
3	1995-04-01	1.662	20.38	81.80	7522.289	15.98615	2.397	24.32305
4	1995-05-01	1.718	18.89	82.57	7522.289	14.25792	2.22862	22.30289
..	…	…	…	…	…	…	…	…
331	2022-08-01	9.127	89.55	108.85	25248.476	30.30098	-3.83564	22.44844
332	2022-09-01	7.065	79.49	112.12	25248.476	29.55154	-3.21816	23.24929
333	2022-10-01	6.355	86.53	111.54	25723.941	30.54675	-2.99443	25.9675
334	2022-11-01	6.930	80.56	105.87	25723.941	30.75377	-3.19104	25.60929
335	2022-12-01	5.079	79.81	104.06	25723.941	29.97484	-2.92123	23.06907

```
[336 rows x 8 columns]
```

[25]:
```
ng_df = pd.read_excel('NG_ENR_SUM_DCU_NUS_A.xls')
ng_df = ng_df.iloc[2:,0:2]
ng_df = ng_df.dropna()
ng_df = ng_df.rename(columns={'Back to Contents':'Date',
                              'Data 1: U.S. Natural Gas Reserves Summary as of␣
↪Dec. 31':'US total natural gas proved reserves (trillion cubic feet)'})
ng_df['Date'] = pd.to_datetime(ng_df['Date'])
ng_df = ng_df.tail(26)
ng_df = ng_df.reset_index()
ng_df = ng_df.drop('index',axis=1)
```

```
ng_df['US total natural gas proved reserves (trillion cubic feet)'] = ng_df['US␣
 ↪total natural gas proved reserves (trillion cubic feet)'].mul(0.001)
ng_df
```

[25]:

	Date	US total natural gas proved reserves (trillion cubic feet)
0	1995-06-30	173.476
1	1996-06-30	175.147
2	1997-06-30	175.721
3	1998-06-30	172.443
4	1999-06-30	176.159
5	2000-06-30	186.51
6	2001-06-30	191.743
7	2002-06-30	195.561
8	2003-06-30	197.145
9	2004-06-30	201.2
10	2005-06-30	213.308
11	2006-06-30	220.416
12	2007-06-30	247.789
13	2008-06-30	255.035
14	2009-06-30	283.879
15	2010-06-30	317.647
16	2011-06-30	348.809
17	2012-06-30	322.67
18	2013-06-30	353.994
19	2014-06-30	388.841
20	2015-06-30	324.303
21	2016-06-30	341.133
22	2017-06-30	464.292
23	2018-06-30	504.501
24	2019-06-30	494.911
25	2020-06-30	473.285

[26]:
```
df = pd.DataFrame(id, columns = ['id'])
id = np.arange(0,26)
ng_concat = pd.concat([df['id'],ng_df['US total natural gas proved reserves␣
 ↪(trillion cubic feet)']], axis=1,ignore_index=False)

ng_concat = ng_concat.append([ng_concat]*12,ignore_index=True).
 ↪sort_values(by='id')
ng_concat.reset_index(inplace= True)
ng_concat = ng_concat.drop(['index'],axis=1)
ng_concat = ng_concat.dropna()
ng_concat
```

[26]:

	id	US total natural gas proved reserves (trillion cubic feet)
0	0	173.476
1	0	173.476

```
2     0                                                 173.476
3     0                                                 173.476
4     0                                                 173.476
..   ..                                                     …
333  25                                                 473.285
334  25                                                 473.285
335  25                                                 473.285
336  25                                                 473.285
337  25                                                 473.285

[338 rows x 2 columns]
```

[27]:
```
final_dataframe = pd.concat([final_df,ng_concat['US total natural gas proved␣
↪reserves (trillion cubic feet)']], axis=1)
final_dataframe = final_dataframe.dropna()
final_dataframe
```

[27]:
```
           Date  Natural Gas Price  Crude Oil Price  US Dollar Index  \
0    1995-01-01              1.354            18.39            87.99
1    1995-02-01              1.483            18.49            85.93
2    1995-03-01              1.685            19.17            81.90
3    1995-04-01              1.662            20.38            81.80
4    1995-05-01              1.718            18.89            82.57
..          …                  …                …                …
331  2022-08-01              9.127            89.55           108.85
332  2022-09-01              7.065            79.49           112.12
333  2022-10-01              6.355            86.53           111.54
334  2022-11-01              6.930            80.56           105.87
335  2022-12-01              5.079            79.81           104.06

            GDP  Natural Gas Production  Net Imports  Natural Gas Consumption  \
0      7455.288                15.60033      2.08822                   15.1435
1      7455.288                16.03862      2.06435                  17.48263
2      7455.288                16.63964      2.27778                  21.19285
3      7522.289                15.98615        2.397                  24.32305
4      7522.289                14.25792      2.22862                  22.30289
..          …                       …            …                         …
331   25248.476                30.30098     -3.83564                  22.44844
332   25248.476                29.55154     -3.21816                  23.24929
333   25723.941                30.54675     -2.99443                   25.9675
334   25723.941                30.75377     -3.19104                  25.60929
335   25723.941                29.97484     -2.92123                  23.06907

    US total natural gas proved reserves (trillion cubic feet)
0                                                 173.476
1                                                 173.476
2                                                 173.476
```

```
3                                                  173.476
4                                                  173.476
..                                                     …
331                                                473.285
332                                                473.285
333                                                473.285
334                                                473.285
335                                                473.285

[336 rows x 9 columns]
```

[28]:
```
final_dataframe['Natural Gas Production'] = pd.
 ↪to_numeric(final_dataframe['Natural Gas Production'])
final_dataframe['Net Imports'] = pd.to_numeric(final_dataframe['Net Imports'])
final_dataframe['US total natural gas proved reserves (trillion cubic feet)'] =␣
 ↪pd.to_numeric(final_dataframe['US total natural gas proved reserves␣
 ↪(trillion cubic feet)'])
```

[29]:
```
final_dataframe.to_csv('Final_Data.csv',index=False)
```

[30]:
```
data = pd.read_csv('Final_Data.csv')
data
```

[30]:
```
           Date  Natural Gas Price  Crude Oil Price  US Dollar Index  \
0    1995-01-01              1.354            18.39            87.99
1    1995-02-01              1.483            18.49            85.93
2    1995-03-01              1.685            19.17            81.90
3    1995-04-01              1.662            20.38            81.80
4    1995-05-01              1.718            18.89            82.57
..          …                  …                …                …
331  2022-08-01              9.127            89.55           108.85
332  2022-09-01              7.065            79.49           112.12
333  2022-10-01              6.355            86.53           111.54
334  2022-11-01              6.930            80.56           105.87
335  2022-12-01              5.079            79.81           104.06

           GDP  Natural Gas Production  Net Imports  Natural Gas Consumption  \
0     7455.288                15.60033      2.08822                 15.14350
1     7455.288                16.03862      2.06435                 17.48263
2     7455.288                16.63964      2.27778                 21.19285
3     7522.289                15.98615      2.39700                 24.32305
4     7522.289                14.25792      2.22862                 22.30289
..         …                       …            …                        …
331  25248.476                30.30098     -3.83564                 22.44844
332  25248.476                29.55154     -3.21816                 23.24929
333  25723.941                30.54675     -2.99443                 25.96750
334  25723.941                30.75377     -3.19104                 25.60929
```

```
335  25723.941                29.97484     -2.92123                 23.06907

     US total natural gas proved reserves (trillion cubic feet)
0                                                    173.476
1                                                    173.476
2                                                    173.476
3                                                    173.476
4                                                    173.476
..                                                       …
331                                                  473.285
332                                                  473.285
333                                                  473.285
334                                                  473.285
335                                                  473.285

[336 rows x 9 columns]
```

[ ]:

[ ]: