Data Import

August 31, 2022

[11]:
```
#Data import

import pandas as pd
import numpy as np
import math
import os
import yfinance as yf

tickers_dict = {'ENERGY': ['COP', 'CVX', 'SHEL', 'XOM'], 'GOLD': ['CDE', 'HL',␣
↪'NEM', 'RGLD'],
          'HEALTHCARE': ['CVS', 'ELV', 'HCA', 'UNH'], 'TECH': ['AAPL', 'AMZN',␣
↪'GOOG', 'MSFT'],
         'UTILITY': ['D', 'DUK', 'NEE', 'SO']}

for sector in tickers_dict:
    for symbol in tickers_dict[sector]:

        ticker = yf.Ticker(symbol)

        data = ticker.history(start='2012-01-01', end='2022-01-01')

        path = 'Stock Data (2012-2022)/{}/{}_data.csv'.format(sector, ticker.
↪ticker)

        data.to_csv(path)
```

[16]:
```
#Create a single dataframe of just the date and close price info for all stocks␣
↪of each sector

COMPLETE_df = pd.DataFrame()

for filename in os.listdir('/home/gandhi96/ROP/Stock Data (2012-2022)'):
    path = '/home/gandhi96/ROP/Stock Data (2012-2022)/{}'.format(filename)
    if os.path.isdir(path):
        directory = filename
        for file in os.listdir(path):
```

```
            if file[-3:] == 'csv':
                file_path = '/home/gandhi96/ROP/Stock Data (2012-2022)/{}/{}'.
 ↪format(directory, file)
                curr_data = pd.read_csv(file_path, parse_dates = True).dropna()
                ticker = file.strip('_data.csv')
                COMPLETE_df['Date'] =  np.array(curr_data["Date"])
                COMPLETE_df[ticker] = np.array(curr_data["Close"])

COMPLETE_df = COMPLETE_df.set_index('Date')

COMPLETE_df
```

[16]:

Date	COP	CVX	SHEL	XOM	CDE	HL	\
2012-01-03	39.503304	71.152664	41.390953	55.358418	25.330000	5.446543	
2012-01-04	39.316895	71.030167	41.418819	55.371304	25.590000	5.474912	
2012-01-05	39.002655	70.333916	40.688267	55.203938	25.840000	5.437089	
2012-01-06	38.699070	69.824661	41.134403	54.791969	25.639999	5.342529	
2012-01-09	38.848198	70.585335	41.257107	55.036602	25.510000	5.361442	
…	…	…	…	…	…	…	
2021-12-27	71.226372	115.546692	43.536983	60.007717	5.100000	5.231403	
2021-12-28	71.138817	115.322968	43.232529	59.813797	5.080000	5.191546	
2021-12-29	70.944229	114.729622	42.898605	59.290226	4.870000	5.032112	
2021-12-30	70.321579	114.223824	42.623619	58.941174	5.000000	5.131758	
2021-12-31	70.419136	114.146004	42.623619	59.329006	5.040000	5.201510	

Date	NEM	RGLD	CVS	ELV	HCA	\
2012-01-03	50.008633	60.809380	32.779312	57.656933	16.601580	
2012-01-04	49.855579	60.003040	33.000370	57.972462	15.957155	
2012-01-05	50.024742	60.169575	32.960880	58.424438	16.333073	
2012-01-06	49.920025	60.975925	32.731953	60.317593	17.138603	
2012-01-09	49.525307	61.650799	32.992477	60.701366	16.609251	
…	…	…	…	…	…	
2021-12-27	59.269157	103.681671	100.744537	457.053223	252.447067	
2021-12-28	59.377491	103.493401	101.108551	461.718079	255.421494	
2021-12-29	59.702492	103.285309	102.023514	465.159485	257.142456	
2021-12-30	60.657814	103.780777	102.003838	464.642273	256.764435	
2021-12-31	61.081303	104.256416	101.492256	461.051666	255.580658	

Date	UNH	AAPL	AMZN	GOOG	MSFT	\
2012-01-03	43.781525	12.540046	8.951500	16.573130	21.527195	
2012-01-04	44.419247	12.607436	8.875500	16.644611	22.033812	
2012-01-05	44.716866	12.747403	8.880500	16.413727	22.258972	
2012-01-06	44.878410	12.880663	9.130500	16.189817	22.604755	
2012-01-09	44.827393	12.860233	8.928000	15.503389	22.307230	

```
…                  …          …          …          …          …
2021-12-27  496.256500  179.586868  169.669495  148.063995  340.227112
2021-12-28  499.674164  178.551147  170.660995  146.447998  339.034882
2021-12-29  502.296997  178.640778  169.201004  146.504501  339.730377
2021-12-30  501.154480  177.465622  168.644501  146.002502  337.117432
2021-12-31  498.879364  176.838242  166.716995  144.679504  334.136902

                   D         DUK         NEE         SO
Date
2012-01-03  34.734024   40.991581  10.976557  27.804970
2012-01-04  34.449535   40.801960  10.948627  27.786449
2012-01-05  34.376755   40.783012  11.088282  27.749407
2012-01-06  34.026112   40.783012  10.970974  27.465359
2012-01-09  33.979797   40.707176  10.984007  27.755581
…                 …          …          …          …
2021-12-27  76.172012  100.425606  89.866829  65.096268
2021-12-28  76.644218  101.670105  90.417915  65.844841
2021-12-29  77.145927  101.854843  90.526169  66.175385
2021-12-30  77.195122  101.884010  91.293755  66.457306
2021-12-31  77.283653  101.990952  91.874367  66.671188

[2517 rows x 20 columns]
```

[17]:
```
#Export the data to csv

COMPLETE_df.to_csv("Stock Data (2012-2022)/COMPLETE_data.csv")
```

[18]:
```
#import market data

ticker = yf.Ticker("SPY")

data = ticker.history(start='2012-01-01', end='2022-01-01')

data.to_csv("Stock Data (2012-2022)/{}_data.csv".format(ticker.ticker))
```