MODULE 7: DATA WRANGLING WITH PANDAS

CPE311 Computational Thinking With Python

submitted by: Dolores, Marc Joseph S.

Performed on: 03/20/24

Submitted on: 03/20/24

Submitted to: Engr. Roman M. Richard

Exercise 1

```
import pandas as pd
filepath1 = '/content/aapl.csv'
filepath2 = '/content/amzn.csv'
filepath3 = '/content/fb.csv'
filepath4 = '/content/goog.csv'
filepath5 = '/content/nflx.csv'

appl = pd.read_csv(filepath1)
amzn = pd.read_csv(filepath2)
fb = pd.read_csv(filepath3)
goog = pd.read_csv(filepath4)
nflx = pd.read_csv(filepath5)

dfs = [appl, amzn, fb, goog, nflx]
```

```
appl['ticker'] = 'AAPL'
appl
```

	date	open	high	low	close	volume	ticke
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AAP
1	2018-01-03	169.2521	171.2337	168.6929	168.9578	29517899	AAP
2	2018-01-04	169.2619	170.1742	168.8106	169.7426	22434597	AAP
3	2018-01-05	170.1448	172.0381	169.7622	171.6751	23660018	AAP
4	2018-01-08	171.0375	172.2736	170.6255	171.0375	20567766	AAP
...	...	...	...	...	...	...	.

Next steps: View recommended plots

```
amzn['ticker'] = 'AMZN'
amzn
```

	date	open	high	low	close	volume	ticker
0	2018-01-02	1172.00	1190.00	1170.51	1189.01	2694494	AMZN
1	2018-01-03	1188.30	1205.49	1188.30	1204.20	3108793	AMZN
2	2018-01-04	1205.00	1215.87	1204.66	1209.59	3022089	AMZN

3	2018-01-05	1217.51	1229.14	1210.00	1229.14	3544743	AMZN
4	2018-01-08	1236.00	1253.08	1232.03	1246.87	4279475	AMZN
...	...	...	...	...	...	...	...

Next steps: View recommended plots

```
fb['ticker'] = 'FB'
fb
```

	date	open	high	low	close	volume	ticker
0	2018-01-02	177.68	181.58	177.5500	181.42	18151903	FB
1	2018-01-03	181.88	184.78	181.3300	184.67	16886563	FB
2	2018-01-04	184.90	186.21	184.0996	184.33	13880896	FB
3	2018-01-05	185.59	186.90	184.9300	186.85	13574535	FB
4	2018-01-08	187.20	188.90	186.3300	188.28	17994726	FB
...	...	...	...	...	...	...	...

Next steps: View recommended plots

```
goog['ticker'] = 'GOOG'
goog
```

	date	open	high	low	close	volume	ticker
0	2018-01-02	1048.34	1066.94	1045.23	1065.00	1237564	GOOG
1	2018-01-03	1064.31	1086.29	1063.21	1082.48	1430170	GOOG
2	2018-01-04	1088.00	1093.57	1084.00	1086.40	1004605	GOOG
3	2018-01-05	1094.00	1104.25	1092.00	1102.23	1279123	GOOG
4	2018-01-08	1102.23	1111.27	1101.62	1106.94	1047603	GOOG
...	...	...	...	...	...	...	...

Next steps: View recommended plots

```
nflx['ticker'] = 'NFLX'
nflx
```

	date	open	high	low	close	volume	ticker
0	2018-01-02	196.10	201.6500	195.4200	201.070	10966889	NFLX
1	2018-01-03	202.05	206.2100	201.5000	205.050	8591369	NFLX
2	2018-01-04	206.20	207.0500	204.0006	205.630	6029616	NFLX
3	2018-01-05	207.25	210.0200	205.5900	209.990	7033240	NFLX
4	2018-01-08	210.02	212.5000	208.4400	212.050	5580178	NFLX
...	...	...	...	...	...	...	...

Next steps: View recommended plots

```
faang = pd.DataFrame()

for i in dfs:
  faang = faang.append(i, ignore_index=True)
```

```
<ipython-input-34-046a64803c77>:4: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a
  faang = faang.append(i, ignore_index=True)
```

```
faang
```

	date	open	high	low	close	volume	tick
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AA
1	2018-01-03	169.2521	171.2337	168.6929	168.9578	29517899	AA
2	2018-01-04	169.2619	170.1742	168.8106	169.7426	22434597	AA
3	2018-01-05	170.1448	172.0381	169.7622	171.6751	23660018	AA
4	2018-01-08	171.0375	172.2736	170.6255	171.0375	20567766	AA
...	...	...	...	...	...	...	

Next steps: View recommended plots

```
faang.to_csv('/content/faang.csv', index=False)
```

Exercise 2

```
faang
```

	date	open	high	low	close	volume	ticker
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AAPL
1	2018-01-03	169.2521	171.2337	168.6929	168.9578	29517899	AAPL
2	2018-01-04	169.2619	170.1742	168.8106	169.7426	22434597	AAPL
3	2018-01-05	170.1448	172.0381	169.7622	171.6751	23660018	AAPL
4	2018-01-08	171.0375	172.2736	170.6255	171.0375	20567766	AAPL
...	...	...	...	...	...	...	...
1250	2018-12-24	242.0000	250.6500	233.6800	233.8800	9547616	NFLX
1251	2018-12-26	233.9200	254.5000	231.2300	253.6700	14402735	NFLX
1252	2018-12-27	250.1100	255.5900	240.1000	255.5650	12235217	NFLX
1253	2018-12-28	257.9400	261.9144	249.8000	256.0800	10987286	NFLX
1254	2018-12-31	260.1600	270.1001	260.0000	267.6600	13508920	NFLX

1255 rows × 7 columns

Next steps: View recommended plots

```
faang.dtypes
```

```
date        object
open       float64
high       float64
low        float64
close      float64
volume       int64
ticker      object
dtype: object
```

```
faang['date'] = pd.to_datetime(faang['date'])
faang
```

	date	open	high	low	close	volume	tick
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AA
1	2018-01-03	169.2521	171.2337	168.6929	168.9578	29517899	AA
2	2018-01-04	169.2619	170.1742	168.8106	169.7426	22434597	AA
3	2018-01-05	170.1448	172.0381	169.7622	171.6751	23660018	AA
4	2018-01-08	171.0375	172.2736	170.6255	171.0375	20567766	AA
...	...	...	...	...	...	...	

Next steps: View recommended plots

```
faang['volume'] = faang['volume'].astype(int)
```

```
faang.dtypes
```

```
date      datetime64[ns]
open             float64
high             float64
low              float64
close            float64
volume             int64
ticker            object
dtype: object
```

```
faang = faang.sort_values(by=['date', 'ticker'])
faang
```

	date	open	high	low	close	volume
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934
251	2018-01-02	1172.0000	1190.0000	1170.5100	1189.0100	2694494
502	2018-01-02	177.6800	181.5800	177.5500	181.4200	18151903
753	2018-01-02	1048.3400	1066.9400	1045.2300	1065.0000	1237564
1004	2018-01-02	196.1000	201.6500	195.4200	201.0700	10966889
...	...	...	...	...	...	...

Next steps: View recommended plots

```
dfhighvlm = faang.nlargest(n=7, columns='volume')
dfhighvlm
```

	date	open	high	low	close	volume	tick
644	2018-07-26	174.8900	180.1300	173.7500	176.2600	169803668	
555	2018-03-20	167.4700	170.2000	161.9500	168.1500	129851768	
559	2018-03-26	160.8200	161.1000	149.0200	160.0600	126116634	

Next steps: View recommended plots

```
faangmelt = pd.melt(faang, id_vars=['date','ticker'])
faangmelt
```

	date	ticker	variable	value
0	2018-01-02	AAPL	open	1.669271e+02
1	2018-01-02	AMZN	open	1.172000e+03
2	2018-01-02	FB	open	1.776800e+02
3	2018-01-02	GOOG	open	1.048340e+03
4	2018-01-02	NFLX	open	1.961000e+02
...	...	...	...	...
6270	2018-12-31	AAPL	volume	3.500347e+07
6271	2018-12-31	AMZN	volume	6.954507e+06
6272	2018-12-31	FB	volume	2.462531e+07
6273	2018-12-31	GOOG	volume	1.493722e+06

Next steps: View recommended plots

Exercise 3

sorry po sir di ko po gets pano gagawin kahit aralin ko po kaya di ko po magawa yung exercise, di na po ako naglagay kasi kung maglalagay po ako di rin po manggagaling sakin.

Conclusion:

i conclude that there are many ways to clean data and add index to the completed data. in this exercise we learned to put 5 csv file into one dataframe and convert it into csv and also sort in order to make it neat and use melt to simplify the data