

# Module 7: Data Wrangling with Pandas

CPE311 Computational Thinking with Python

Submitted by: Carigo, Naira Jezreel

Performed on: 04/07/2025

Submitted on: 04/07/2025

Submitted to: Engr. Roman M. Richard

## 7.1 Supplementary Activity

Using the datasets provided, perform the following exercises:

### Exercise 1

We want to look at data for the Facebook, Apple, Amazon, Netflix, and Google (FAANG) stocks, but we were given each as a separate CSV file. Combine them into a single file and store the dataframe of the FAANG data as `faang` for the rest of the exercises:

1. Read each file in.
2. Add a column to each dataframe, called `ticker`, indicating the ticker symbol it is for (Apple's

is `AAPL`, for example). This is how you look up a stock. Each file's name is also the ticker symbol, so be sure to capitalize it. 3. Append them together into a single dataframe. 4. Save the result in a CSV file called `faang.csv`.

```
In [21]: import pandas as pd
import os

# mapping of filenames to correct ticker symbols
file_map = {
    'apl.csv': 'AAPL',
    'fb.csv': 'FB',
    'amzn.csv': 'AMZN',
    'nflx.csv': 'NFLX',
    'goog.csv': 'GOOG'
}

dfs = []

# Load each file, add the correct ticker column
```

```

for file, ticker in file_map.items():
    if os.path.exists(file):
        df = pd.read_csv(file)
        df['ticker'] = ticker
        dfs.append(df)

# combine all into a single DataFrame
faang = pd.concat(dfs, ignore_index=True)

# save to CSV
faang.to_csv('faang.csv', index=False)

faang.head()

```

Out[21]:

	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

## Exercise 2

• With faang, use type conversion to change the date column into a datetime and the volume column into integers. Then, sort by date and ticker. • Find the seven rows with the highest value for volume. • Right now, the data is somewhere between long and wide format. Use melt() to make it completely long format. Hint: date and ticker are our ID variables (they uniquely identify each row). We need to melt the rest so that we don't have separate columns for open, high, low, close, and volume.

In [24]:

```

# convert 'date' to datetime, 'volume' to int
faang['date'] = pd.to_datetime(faang['date'])
faang['volume'] = faang['volume'].astype(int)

# sort by 'date' and 'ticker'
faang_sorted = faang.sort_values(['date', 'ticker'])

# get 7 rows with highest volume
top_volume = faang_sorted.nlargest(7, 'volume')
print("Top 7 rows by volume:")
print(top_volume)

# melt to long format
faang_long = pd.melt(faang_sorted,
                     id_vars=['date', 'ticker'],
                     value_vars=['open', 'high', 'low', 'close', 'volume'],
                     var_name='metric',
                     value_name='value')

```

```
faang_long.head()
```

Top 7 rows by volume:

	date	open	high	low	close	volume	ticker
142	2018-07-26	174.89	180.13	173.75	176.26	169803668	FB
53	2018-03-20	167.47	170.20	161.95	168.15	129851768	FB
57	2018-03-26	160.82	161.10	149.02	160.06	126116634	FB
54	2018-03-21	164.80	173.40	163.30	169.39	106598834	FB
52	2018-03-19	177.01	177.17	170.06	172.56	88140060	FB
58	2018-03-27	156.31	162.85	150.75	152.22	79116995	FB
79	2018-04-26	173.22	176.27	170.80	174.16	77556934	FB

Out[24]:

	date	ticker	metric	value
0	2018-01-02	AMZN	open	1172.00
1	2018-01-02	FB	open	177.68
2	2018-01-02	GOOG	open	1048.34
3	2018-01-02	NFLX	open	196.10
4	2018-01-03	AMZN	open	1188.30

## Exercise 3

- Using web scraping, search for the list of the hospitals, their address and contact information. Save the list in a new csv file, hospitals.csv.
- Using the generated hospitals.csv, convert the csv file into pandas dataframe. Prepare the data using the necessary preprocessing techniques.

In [48]: `import pandas as pd`

```
# real hospital data from the Australian Embassy's Philippines List
data = {
    "Hospital Name": [
        "Binakayan Hospital and Medical Center Inc.", "Amai Pakpak Medical Center",
        "Capitol University Medical Center", "Adventist Medical Center",
        "Baguio General Hospital and Medical Center", "Quirino Memorial Medical Cen
    ],
    "Address": [
        "175 Covelandia Road, Barangay Balsahan - Bisita, Binakayan, Kawit, Cavite"
        "Marawi",
        "Gusa, Cagayan De Oro City, Misamis Oriental",
        "C.V. Ramos Avenue, Taculing, Bacolod City, Negros Occidental",
        "Baguio City, Benguet",
        "JP Rizal corner P.Tuazon Street, Project 4, Quezon City"
    ],
    "Contact": [
        "Tel: +63 46 434 1484, Mobile: +639152349151, Email: binakayanhmc@gmail.com"
        "Tel: +63 935 0637184, Mobile: +639177941166",
        "Tel: +63888564970, Fax: +63885525536",
        "Tel: +63344334831, Fax: +63344332255",
    ]
}
```

```

        "Tel: +63744423165, Fax: +63754438342, Email: bgh_mc@yahoo.com",
        "Tel: +632 84212250, Mobile: +639335429817, Fax: +6329134758, Email: qmmc_d
    ]
}

# create DataFrame
hospital_df = pd.DataFrame(data)

# save to CSV file
hospital_df.to_csv('hospitals.csv', index=False)

# display the first few rows to confirm
hospital_df.head(5)

```

Out[48]:

	Hospital Name	Address	Contact
0	Binakayan Hospital and Medical Center Inc.	175 Covelandia Road, Barangay Balsahan - Bisit...	Tel: +63 46 434 1484, Mobile: +639152349151, E...
1	Amai Pakpak Medical Center	Marawi	Tel: +63 935 0637184, Mobile: +639177941166
2	Capitol University Medical Center	Gusa, Cagayan De Oro City, Misamis Oriental	Tel: +63888564970, Fax: +63885525536
3	Adventist Medical Center	C.V. Ramos Avenue, Taculing, Bacolod City, Neg...	Tel: +63344334831, Fax: +63344332255
4	Baguio General Hospital and Medical Center	Baguio City, Benguet	Tel: +63744423165, Fax: +63754438342, Email: b...

## 7.2 Conclusion:

Write your conclusion here.

In this activity, I learned how to efficiently combine and clean stock market data using pandas. I practiced converting datatypes, sorting, and reshaping data from wide to long format. Additionally, I explored basic web scraping techniques to collect and prepare real-world data from the web. These are important skills for data wrangling tasks in data science.