Module 7: Data Wrangling with Pandas

CPE311 Computational Thinking with Python

Submitted by: Sanchez, Justin Bjorn

Performed on: 04/07/2025

Submitted on: 04/07/2025

Submitted to: Engr. Roman M. Richard

+ Code + Text

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 separate CSV file. Combine them into a single file and store the dataframe of the FAANG as faang for the rest of the exercises:

1. Reach 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

```
import pandas as pd   #Import all csv files.

apple = pd.read_csv('/content/aapl.csv')
amazon = pd.read_csv('/content/amzn.csv')
FB = pd.read_csv('/content/fb.csv')
Google = pd.read_csv('/content/goog.csv')
Netflix = pd.read_csv('/content/nflx.csv')
```

```
apple['Ticker'] = 'AAPL'     #Put a Ticker column with a symbol to indicate what stock
amazon['Ticker'] = 'AMZN'
FB['Ticker'] = 'FB'
Google['Ticker'] = 'GOOG'
Netflix['Ticker'] = 'NFLX'
```

```
faang = pd.concat([apple,amazon,FB,Google,Netflix])   #Append together into a single data frame with the Ticker column added
faang.head()
```

	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

Next steps: View recommended plots New interactive sheet

```
faang = faang.to_csv("faang.csv")   #Save new faang data frame to a csv filefa
```

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 tickers 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.

```
faang = pd.read_csv('/content/faang.csv') #import faang.csv


faang['date'] = pd.to_datetime(faang['date'])  #change dtypes
faang['volume'] = faang['volume'].astype(int)

faang = faang.sort_values(by=['volume'], ascending = False)

faang.dtypes
```

	0
Unnamed: 0	int64
date	datetime64[ns]
open	float64
high	float64
low	float64
close	float64
volume	int64
Ticker	object

dtype: object

```
HighVol = faang.nlargest(7, 'volume') #find top 7 highest volume
HighVol
```

	Unnamed: 0	date	open	high	low	close	volume	Ticker
644	142	2018-07-26	174.8900	180.1300	173.7500	176.2600	169803668	FB
555	53	2018-03-20	167.4700	170.2000	161.9500	168.1500	129851768	FB
559	57	2018-03-26	160.8200	161.1000	149.0200	160.0600	126116634	FB
556	54	2018-03-21	164.8000	173.4000	163.3000	169.3900	106598834	FB
182	182	2018-09-21	219.0727	219.6482	215.6097	215.9768	96246748	AAPL
245	245	2018-12-21	156.1901	157.4845	148.9909	150.0862	95744384	AAPL
212	212	2018-11-02	207.9295	211.9978	203.8414	205.8755	91328654	AAPL

Next steps: View recommended plots | New interactive sheet

```
FaangMelt = pd.melt(faang, id_vars=['date','Ticker'],value_vars=['open','high','low','close','volume']) #melting to make long format
FaangMelt
```

	date	Ticker	variable	value
0	2018-07-26	FB	open	174.8900
1	2018-03-20	FB	open	167.4700
2	2018-03-26	FB	open	160.8200
3	2018-03-21	FB	open	164.8000
4	2018-09-21	AAPL	open	219.0727
...	...	...	...	...
6270	2018-08-09	GOOG	volume	848601.0000
6271	2018-07-10	GOOG	volume	798412.0000
6272	2018-05-24	GOOG	volume	766773.0000
6273	2018-11-23	GOOG	volume	691462.0000
6274	2018-07-03	GOOG	volume	679034.0000

6275 rows × 4 columns

Next steps: View recommended plots | New interactive sheet

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

```
import pandas as pd
import requests                    #Get list of hosp using web url
from bs4 import BeautifulSoup

url = 'https://en.wikipedia.org/wiki/List_of_hospitals_in_the_Philippines'
request = requests.get(url)

request

soup = BeautifulSoup(request.content, 'html.parser')

table = soup.find('table', {'id': 'example'})


table = soup.find_all('table', {'class': 'wikitable'})[0]      #Extract data from wiki using wikitable as class using beautifulsoup
headers = [th.text.strip() for th in table.find_all('th')]
headers

rows = []

for row in table.find_all('tr')[1:]:
  cells = row.find_all('td')
  row_data = [cell.text.strip() for cell in cells]
  rows.append(row_data)


df = pd.DataFrame(rows,columns=headers) #convert to csv
df.to_csv('hospitals.csv', index=False)


hospitals = pd.read_csv('/content/hospitals.csv') #import to display new hospitals.csv dataframe

hospitals.head()
```

	Name of Hospital	Location	Class
0	Caloocan City Medical Center	450 A. Mabini St., Caloocan City	LGU
1	Ospital ng Malabon	F. Sevilla Boulevard, Tañong, Malabon City	LGU
2	San Lorenzo Ruiz General Hospital	O. Reyes St., Rosita Subdivision, Santulan, Ma...	DOH Retained
3	Gat Andres Bonifacio Memorial Medical Center	8001 Delpan St., Tondo, Manila	LGU
4	Ospital ng Tondo	Jose Abad Santos Avenue, Tondo, Manila	LGU

Next steps: View recommended plots | New interactive sheet

```
hospitals['Class'].unique() #find error values
```

```
array(['LGU', 'DOH Retained', 'AFP', 'PAF', 'PNP', 'University', 'DND',
       'GOCC'], dtype=object)
```

```
hospitals['Location'].unique() #find error values
```

```
array(['450 A. Mabini St., Caloocan City',
       'F. Sevilla Boulevard, Tañong, Malabon City',
       'O. Reyes St., Rosita Subdivision, Santulan, Malabon City',
       '8001 Delpan St., Tondo, Manila',
       'Jose Abad Santos Avenue, Tondo, Manila',
       'Numancia St., Binondo, Manila',
       '677 Geronimo St., cor. Carola St., Sampaloc, Manila',
       'M. Naval St., Brgy. San Jose, Navotas City',
       '0440 Quirino Ave., La Huerta, Parañaque City',
       '187 Taiwan Extension Corner Doña Soledad Avenue, Don Bosco, Parañaque City',
       'Quirino Highway, San Bartolome, Novaliches, Quezon City',
       'N. Domingo St., San Juan City',
       'Fort Andres Bonifacio, Taguig City',
       'Naval Station, Jose Francisco, Fort Bonifacion, Taguig City',
       'C-6 Road, Hagonoy, Taguig',
```

```
       'East Service Road, Western Bicutan, Taguig',
       'New Panaderos St., Sta. Ana, Manila',
       '605 Boni Avenue, Mandaluyong City',
       'Gozar St., Colonel Jesus Villamor Air Base, Pasay City',
       'Industria St. cor. Alcalde Jose St., Kapasigan, Pasig City',
       'Camp Crame, Quezon City',
       'IBP Road, Batasan Hills, District 2, Quezon City',
       'St. Joseph Avenue (Dr. Uyguanco Street), Tala, Caloocan City',
       'Bernabe Compound, Pulanglupa, Las Piñas City',
       'Lope de Vega St., Sta. Cruz, Manila',
       'San Lazaro Compound, Rizal Avenue, Sta. Cruz, Manila',
       'Quiricada St., Sta. Cruz, Manila',
       'Honorio Lopez Boulevard., Balut, Tondo, Manila',
       'Taft Avenue, Ermita, Manila',
       'Pres. Quirino Avenue, cor. Roxas Blvd., Malate, Manila',
       '#9 De Febrero St., Mandaluyong City',
       'Sampaguita St. cor. Gumamela St., Brgy. Pembo, Makati City',
       'Sumulong Highway, Brgy. Sto. Niño, Marikina City',
       'Civic Drive, Filinvest Corporate City, Alabang, Muntinlupa City',
       'Filinvest Corporate City, Alabang, Muntinlupa City',
       'P. Burgos St., Pasay City',
       'Pasig Boulevard, Bagong Ilog, Pasig City',
       'M. Eusebio Avenue, Maybunga, Pasig City',
       'North Avenue, Diliman, Quezon City',
       'East Avenue, Diliman, Quezon City',
       'Quezon Avenue corner Senator Miriam P. Defensor-Santiago Avenue, Diliman, Quezon City',
       'V. Luna Road, Quezon City', 'East Avenue, Quezon City',
       'Ma. Clara St. cor. Banawe St., Quezon City',
       'Seminary Road, EDSA, Quezon City', 'Quezon Avenue, Quezon City',
       '#226 E. Rodriguez Sr. Boulevard, Quezon City',
       'Katipunan Road, Project 4, Quezon City',
       'Padrigal St., Karuhatan, Valenzuela City'], dtype=object)
```

```
hospitals['Name of Hospital'].unique() #find error values
```

```
array(['Caloocan City Medical Center', 'Ospital ng Malabon',
       'San Lorenzo Ruiz General Hospital',
       'Gat Andres Bonifacio Memorial Medical Center', 'Ospital ng Tondo',
       'Justice Jose Abad Santos General Hospital', 'Ospital ng Sampaloc',
       'Navotas City Hospital', 'Ospital ng Parañaque',
       'Ospital ng Parañaque District II', 'Novaliches District Hospital',
       'San Juan Medical Center', 'Army General Hospital',
       'Manila Naval Hospital', 'Taguig City General Hospital',
       'Taguig–Pateros District Hospital', 'Santa Ana Hospital',
       'Mandaluyong City Medical Center', 'Air Force General Hospital',
       "Pasig City Children's Hospital – Child's Hope",
       'PNP General Hospital',
       'Rosario Maclang Bautista General Hospital',
       'Dr. Jose N. Rodriguez Memorial Hospital and Sanitarium',
       'Las Piñas General Hospital and Satellite Trauma Center',
       'Dr. Jose Fabella Memorial Hospital',
       'Jose R. Reyes Memorial Medical Center', 'San Lazaro Hospital',
       'Tondo Medical Center', 'Philippine General Hospital',
       'Ospital ng Maynila Medical Center',
       'National Center for Mental Health', 'Ospital ng Makati',
       'Amang Rodriguez Memorial Medical Center', 'Ospital ng Muntinlupa',
       'Research Institute for Tropical Medicine',
       'Pasay City General Hospital', 'Rizal Medical Center',
       'Pasig City General Hospital', 'Veterans Memorial Medical Center',
       'Philippine Heart Center',
       'National Kidney and Transplant Institute',
       "Philippine Children's Medical Center",
       'Victoriano Luna Medical Center', 'East Avenue Medical Center',
       'Philippine Orthopedic Center', 'Quezon City General Hospital',
       'Lung Center of the Philippines', 'National Children’s Hospital',
       'Quirino Memorial Medical Center', 'Valenzuela Medical Center'],
      dtype=object)
```

7.2 Conclusion:

To conclude this activity, it has helped me learn how to combine csv files and use the .melt, also this activity helped me how to manipulate multiple data frames and combine it into one and append values and how to edit it. also this activity has enhanced my skills in coding with dataframes.