> # Module 7: Data Wrangling with pandas

CPE311 Computational Thinking With Python

Submitted by: Vista, Jens Liam P.

Performed on: 03/20/2024

Submitted on: 03/20/2024

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 fil 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 tickerm indicating the ticker sumbol it is for (Apple's AAPL, for example). This is how you look up a stock. Each file's name is also the tickwer sumbol, 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
```

```
# Read each file in
Facebook = pd.read_csv('/content/fb.csv')
Apple = pd.read_csv('/content/aapl.csv')
Amazon = pd.read_csv('/content/amzn.csv')
Netflix = pd.read_csv('/content/nflx.csv')
Google = pd.read_csv('/content/goog.csv')
```

```
Facebook
```

	date	open	high	low	close	volume
0	2018-01-02	177.68	181.58	177.5500	181.42	18151903
1	2018-01-03	181.88	184.78	181.3300	184.67	16886563
2	2018-01-04	184.90	186.21	184.0996	184.33	13880896
3	2018-01-05	185.59	186.90	184.9300	186.85	13574535
4	2018-01-08	187.20	188.90	186.3300	188.28	17994726
...	...	...	...	...	...	...
246	2018-12-24	123.10	129.74	123.0200	124.06	22066002
247	2018-12-26	126.00	134.24	125.8900	134.18	39723370
248	2018-12-27	132.44	134.99	129.6700	134.52	31202509
249	2018-12-28	135.34	135.92	132.2000	133.20	22627569
250	2018-12-31	134.45	134.64	129.9500	131.09	24625308

251 rows × 6 columns

```
# Add a column to each dataframe, called tickerm indicating the ticker sumbol it is for (Apple's AAPL, for example).
# This is how you look up a stock. Each file's name is also the tickwer sumbol, so be sure to capitalize it.
Facebook.insert(loc = 0, column = 'ticker', value = 'FB')
Apple.insert(loc = 0, column = 'ticker', value = 'AAPL')
Amazon.insert(loc = 0, column = 'ticker', value = 'AMZN')
Netflix.insert(loc = 0, column = 'ticker', value = 'NFLX')
Google.insert(loc = 0, column = 'ticker', value = 'GOOG')
```

```
Facebook
```

	ticker	date	open	high	low	close	volume
0	FB	2018-01-02	177.68	181.58	177.5500	181.42	18151903
1	FB	2018-01-03	181.88	184.78	181.3300	184.67	16886563
2	FB	2018-01-04	184.90	186.21	184.0996	184.33	13880896
3	FB	2018-01-05	185.59	186.90	184.9300	186.85	13574535
4	FB	2018-01-08	187.20	188.90	186.3300	188.28	17994726
...	...	...	...	...	...	...	...
246	FB	2018-12-24	123.10	129.74	123.0200	124.06	22066002
247	FB	2018-12-26	126.00	134.24	125.8900	134.18	39723370
248	FB	2018-12-27	132.44	134.99	129.6700	134.52	31202509
249	FB	2018-12-28	135.34	135.92	132.2000	133.20	22627569
250	FB	2018-12-31	134.45	134.64	129.9500	131.09	24625308

251 rows × 7 columns

```
# Append them together into a single dataframe
df = pd.concat([Facebook, Apple, Amazon, Netflix, Google])
df
```

	ticker	date	open	high	low	close	volume
0	FB	2018-01-02	177.68	181.58	177.5500	181.42	18151903
1	FB	2018-01-03	181.88	184.78	181.3300	184.67	16886563
2	FB	2018-01-04	184.90	186.21	184.0996	184.33	13880896
3	FB	2018-01-05	185.59	186.90	184.9300	186.85	13574535
4	FB	2018-01-08	187.20	188.90	186.3300	188.28	17994726
...	...	...	...	...	...	...	...
246	GOOG	2018-12-24	973.90	1003.54	970.1100	976.22	1590328
247	GOOG	2018-12-26	989.01	1040.00	983.0000	1039.46	2373270
248	GOOG	2018-12-27	1017.15	1043.89	997.0000	1043.88	2109777
249	GOOG	2018-12-28	1049.62	1055.56	1033.1000	1037.08	1413772
250	GOOG	2018-12-31	1050.96	1052.70	1023.5900	1035.61	1493722

1255 rows × 7 columns

```
# Save the result in a CSV file called faang.csv.
df.to_csv('faang.csv')
```

Exercise 2

- With faang, use type conversion to change the date column into 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 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.

```
# With faang, use type conversion to change the date column into datetime and the volume column into integers. Then, sort
df.dtypes
```

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

```
df.loc[:,'date'] = pd.to_datetime(df.date)
df.dtypes
```

```
<ipython-input-109-80606e5f8dec>:1: DeprecationWarning: In a future version, `df.iloc[:, i] = newvals` will attempt to
  df.loc[:,'date'] = pd.to_datetime(df.date)
ticker            object
date      datetime64[ns]
open             float64
high             float64
low              float64
close            float64
volume             int64
dtype: object
```

```
# Find the seven rows with the highest value for volume.
df.nlargest(7, 'volume')
```

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

```
df = df.melt(id_vars = ['ticker','date'])
df.head()
```

	ticker	date	variable	value
0	FB	2018-01-02	open	177.68
1	FB	2018-01-03	open	181.88
2	FB	2018-01-04	open	184.90
3	FB	2018-01-05	open	185.59
4	FB	2018-01-08	open	187.20

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 requests
from bs4 import BeautifulSoup
url = 'https://en.wikipedia.org/wiki/List_of_hospitals_in_the_Philippines'
soup = BeautifulSoup(requests.get(url).text, 'html')
```

```
table = soup.find('table', class_ = 'wikitable')
```

```
hospitals = table.find_all('th')
hospitals
```

```
[<th>Name of Hospital
 </th>,
 <th>Location
 </th>,
 <th>Class
 </th>]
```

```
df = pd.DataFrame(columns = [title.text.strip() for title in hospitals])
df
```

```
   Name of Hospital  Location  Class
```

```
for row in table.find_all('tr')[1:]:
  data = row.find_all('td')
  results = [datas.text.strip() for datas in data]
  df.loc[len(df)] = results
```

```
df
```

	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
5	Justice Jose Abad Santos General Hospital	Numancia St., Binondo, Manila	LGU
6	Ospital ng Sampaloc	677 Geronimo St., cor. Carola St., Sampaloc, M...	LGU
7	Navotas City Hospital	M. Naval St., Brgy. San Jose, Navotas City	LGU
8	Ospital ng Parañaque	0440 Quirino Ave., La Huerta, Parañaque City	LGU
9	Ospital ng Parañaque District II	187 Taiwan Extension Corner Doña Soledad Avenu...	LGU
10	Novaliches District Hospital	Quirino Highway, San Bartolome, Novaliches, Qu...	LGU
11	San Juan Medical Center	N. Domingo St., San Juan City	LGU
12	Army General Hospital	Fort Andres Bonifacio, Taguig City	AFP
13	Manila Naval Hospital	Naval Station, Jose Francisco, Fort Bonifacion...	AFP
14	Taguig-Pateros District Hospital	East Service Road, Western Bicutan, Taguig	LGU
15	Santa Ana Hospital	New Panaderos St., Sta. Ana, Manila	LGU
16	Mandaluyong City Medical Center	605 Boni Avenue, Mandaluyong City	LGU
17	Air Force General Hospital	Gozar St., Colonel Jesus Villamor Air Base, Pa...	PAF
18	Pasig City Children's Hospital – Child's Hope	Industria St. cor. Alcalde Jose St., Kapasigan...	LGU
19	PNP General Hospital	Camp Crame, Quezon City	PNP
20	Rosario Maclang Bautista General Hospital	IBP Road, Batasan Hills, District 2, Quezon City	LGU
21	Dr. Jose N. Rodriguez Memorial Hospital and Sa...	St. Joseph Avenue (Dr. Uyguanco Street), Tala,...	DOH Retained
22	Las Piñas General Hospital and Satellite T...	Bernabe Compound, Pulanglupa, Las Piñas City	DOH Retained

22	Las Pinas General Hospital and Satellite Traum...	Bernabe Compound, Pulanglupa, Las Pinas City	DOH Retained
23	Dr. Jose Fabella Memorial Hospital	Lope de Vega St., Sta. Cruz, Manila	DOH Retained
24	Jose R. Reyes Memorial Medical Center	San Lazaro Compound, Rizal Avenue, Sta. Cruz, ...	DOH Retained
25	San Lazaro Hospital	Quiricada St., Sta. Cruz, Manila	DOH Retained
26	Tondo Medical Center	Honorio Lopez Boulevard., Balut, Tondo, Manila	DOH Retained
27	Philippine General Hospital	Taft Avenue, Ermita, Manila	University
28	Ospital ng Maynila Medical Center	Pres. Quirino Avenue, cor. Roxas Blvd., Malate...	LGU
29	National Center for Mental Health	#9 De Febrero St., Mandaluyong City	DOH Retained
30	Ospital ng Makati	Sampaguita St. cor. Gumamela St., Brgy. Pembo....	LGU

```
df.to_csv('hospitals.csv', index=False)
```

```
df.dtypes
```

```
     Name of Hospital    object
     Location            object
     Class               object
     dtype: object
```

```
# Using the generated hospitals.csv, convert the csv file into pandas dataframe. Prepare the data using the necessary prepr
df["Class"].value_counts()
```

```
     LGU              22
     DOH Retained     16
     GOCC              4
     AFP               3
     PAF               1
     PNP               1
     University        1
     DND               1
     Name: Class, dtype: int64
```

ANOTHER TRY

```
url = 'https://sulit.ph/list-of-hospitals-in-metro-manila-with-contact-details-website-and-social-media-accounts/?fbclid=Iw
response = requests.get(url)
response
sopas = BeautifulSoup(response.content, 'html.parser')
```

```
table1 = sopas.find('table', class_ = 'has-fixed-layout')
tables= table1.find_all('th')
```

```
tables
```

```
[<th>CITY</th>,
 <th>NAME OF HOSPITAL</th>,
 <th>CONTACT NUMBER</th>,
 <th>WEBSITE / EMAIL</th>,
 <th>FACEBOOK LINK</th>]
```

```
dfss = pd.DataFrame(columns = [title.text.strip() for title in tables])
dfss
```

```
CITY  NAME OF HOSPITAL  CONTACT NUMBER  WEBSITE / EMAIL  FACEBOOK LINK
```

```
for i in table1.find_all('tr')[1:]:
  data = i.find_all('td')
  resultss = [datas.text.strip() for datas in data]
  dfss.loc[len(dfss)] = resultss
```

```
dfss
```

	CITY	NAME OF HOSPITAL	CONTACT NUMBER	WEBSITE / EMAIL	FACEBOOK LINK
0	LIST UPDATE				
1	15 SEPT 2021				
2	Caloocan	Caloocan City Medical Center	South 5310 7925, North 8282 3397, 0943 216 6963		https://www.facebook.com/Caloocan-City-Medical...
3	Caloocan	Dr. Jose N. Rodriguez Memorial Hospital and Sa...	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.facebook.com/officialDJNRMHS
4	Caloocan	MCU – FDT Medical Foundations Hospital	8367 2031	https://www.mcuhospital.org/	
...	...	...	...	...	...
93	Taguig	Medical Center of Taguig	8888 6284		https://www.facebook.com/mctadminofficial/
		Allied Care Experts	direct line to		https://www.facebook.com/ACEMC-

A clean one where I disregard email and fb link

```
table1 = sopas.find('table', class_ = 'has-fixed-layout')
tables= table1.find_all('th')[:3]

tables
```

```
[<th>CITY</th>, <th>NAME OF HOSPITAL</th>, <th>CONTACT NUMBER</th>]
```

```
dfs = pd.DataFrame(columns = [title.text.strip() for title in tables])
dfs
```

```
        CITY  NAME OF HOSPITAL  CONTACT NUMBER
```

```
for i in table1.find_all('tr')[3:]:
  data = i.find_all('td')
  resultss = [datas.text.strip() for datas in data[:3]]
  dfs.loc[len(dfs)] = resultss
```

```
dfs
```

	CITY	NAME OF HOSPITAL	CONTACT NUMBER
0	Caloocan	Caloocan City Medical Center	South 5310 7925, North 8282 3397, 0943 216 6963
1	Caloocan	Dr. Jose N. Rodriguez Memorial Hospital and Sa...	0966 549 2697, 8294 2571 to 73
2	Caloocan	MCU – FDT Medical Foundations Hospital	8367 2031
3	Caloocan	Metro Balayan Medical Center	(043) 740 1350
4	Las Pinas	Alabang Medical Center	8807 8189, 8850 8719
...	...	...	...
[illegible]	Taguig	Medical Center of Taguig	[illegible]