Module 7: Data Wrangling with Pandas

CPE311 Computational Thinking with Python

Submitted by: Garcia, John Carlos

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 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.

```
1 import pandas as pd
2
3 tickers = ['aapl','amzn','fb','goog','nflx'] #This list is used to store the names of the csv files, listed in al
4 faang = pd.DataFrame() #We initialize an empty data frame for faang
5 for ticker in tickers: #We then loop as much as the number of strings there is in the tickers list
6    filepath = f"/content/data/{ticker}.csv" #We substitute each string in the tickers list to get the path of ea
7    df = pd.read_csv(filepath)
8    df['Ticker'] = ticker.upper() #We then add the corresponding ticker symbol for the current file, wherein the
9    faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang d
10
11 faang.to_csv('/content/data/faang.csv', index=False) #After the loop, we save the combined dataframe to a CSV fil
12
```

```
<ipython-input-32-cd6f04c1ac68>:9: FutureWarning: The frame.append method is deprecated and will be removed from
faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang da
<ipython-input-32-cd6f04c1ac68>:9: FutureWarning: The frame.append method is deprecated and will be removed from
faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang da
<ipython-input-32-cd6f04c1ac68>:9: FutureWarning: The frame.append method is deprecated and will be removed from
faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang da
<ipython-input-32-cd6f04c1ac68>:9: FutureWarning: The frame.append method is deprecated and will be removed from
```

faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang da <ipython-input-32-cd6f04c1ac68>:9: FutureWarning: The frame.append method is deprecated and will be removed from faang = faang.append(df, ignore_index=True) #We then add all the data of the current csv file to the faang da

1 faang = pd.read_csv('/content/data/faang.csv')

2 faang

	date	open	high	low	close	volume	Ticker	Ħ
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AAPL	ılı
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 rc	ws × 7 colum	ns						

Next steps:



View recommended plots

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.

Type Conversion

1 faang.dtypes

object
float64
float64
float64
float64
int64
object
object

- 1 faang['DateTime'] = pd.to_datetime(faang['date']) #We use .to_datetime() to convert date column into datetime
 2 faang['Volume'] = faang['volume'].astype(int) #We use .astype(int) to convert strings into integers
- 4 faang.drop(columns=['date', 'volume'], inplace=True) #We drop the previous date and volume
- 5 faang = faang.sort_values(by=['DateTime', 'Ticker']) #We then sort all the rows by the values in 'DateTime' and '
- 6 #This means that if there are rows with the same DateTime, it will then be sorted by their Ticker
- 7 faang

	open	high	low	close	Ticker	DateTime	Volume	
0	166.9271	169.0264	166.0442	168.9872	AAPL	2018-01-02	25555934	ılı
251	1172.0000	1190.0000	1170.5100	1189.0100	AMZN	2018-01-02	2694494	
502	177.6800	181.5800	177.5500	181.4200	FB	2018-01-02	18151903	
753	1048.3400	1066.9400	1045.2300	1065.0000	GOOG	2018-01-02	1237564	
1004	196.1000	201.6500	195.4200	201.0700	NFLX	2018-01-02	10966889	
•••								
250	157.8529	158.6794	155.8117	157.0663	AAPL	2018-12-31	35003466	
501	1510.8000	1520.7600	1487.0000	1501.9700	AMZN	2018-12-31	6954507	
752	134.4500	134.6400	129.9500	131.0900	FB	2018-12-31	24625308	
1003	1050.9600	1052.7000	1023.5900	1035.6100	GOOG	2018-12-31	1493722	
1254	260.1600	270.1001	260.0000	267.6600	NFLX	2018-12-31	13508920	
1255 ro	ws × 7 colum	nns						

1 faang.dtypes

open float64
high float64
low float64
close float64
Ticker object
DateTime datetime64[ns]
Volume int64
dtype: object

✓ Seven Highest Volume Values

1 faang.nlargest(7,'Volume')

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

✓ Melt()

```
1 id = ['DateTime', 'Ticker']
2 value = ['open', 'high', 'low', 'close', 'Volume']
3 long_faang = pd.melt(faang, id, value)
4 long_faang
```

	DateTime	Ticker	variable	value	\blacksquare				
0	2018-01-02	AAPL	open	1.669271e+02	ıl.				
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					
6274	2018-12-31	NFLX	Volume	1.350892e+07					
6275 rd	6275 rows × 4 columns								

Next steps: View recommended plots

✓ 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.

	CITY	CONTACT NUMBER	WEBSITE / EMAIL	FACEBOOK LINK	
NAME OF HOSPITAL					ıl.
NaN	LIST UPDATE	NaN	NaN	NaN	
NaN	15 SEPT 2021	NaN	NaN	NaN	
Caloocan City Medical Center	Caloocan	South 5310 7925, North 8282 3397, 0943 216 6963	NaN	https://www.facebook.com/Caloocan-City- Medical	
Dr. Jose N. Rodriguez Memorial Hospital and Sanitarium HESU / Tala Hospital	Caloocan	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.facebook.com/officialDJNRMHS	
MCU - FDT					

Next steps:

View recommended plots

1 df.dtypes

CITY object
CONTACT NUMBER object
WEBSITE / EMAIL object
FACEBOOK LINK object
dtype: object

1 df['ID'] = df.reset_index().index
2 df

MCII - EDT

	CITY	CONTACT NUMBER	WEBSITE / EMAIL	FACEBOOK LINK	ID
NAME OF HOSPITAL					
NaN	LIST UPDATE	NaN	NaN	NaN	0
NaN	15 SEPT 2021	NaN	NaN	NaN	1
Caloocan City Medical Center	Caloocan	South 5310 7925, North 8282 3397, 0943 216 6963	NaN	https://www.facebook.com/Caloocan-City- Medical	2
Dr. Jose N. Rodriguez Memorial Hospital and Sanitarium HESU / Tala Hospital	Caloocan	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.facebook.com/officialDJNRMHS	3

 \blacksquare

ıl.

1 df

	NAME OF HOSPITAL	CITY	CONTACT NUMBER	WEBSITE / EMAIL	FACEBOOK LINK	
0	NaN	LIST UPDATE	NaN	NaN	NaN	11.
1	NaN	15 SEPT 2021	NaN	NaN	NaN	
2	Caloocan City Medical Center	Caloocan	South 5310 7925, North 8282 3397, 0943 216 6963	NaN	https://www.facebook.com/Caloocan-City- Medical	
3	Dr. Jose N. Rodriguez Memorial Hospital and Sa	Caloocan	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.facebook.com/officialDJNRMHS	
4	MCU - FDT Medical Foundations	Caloocan	8367 2031	https://www.mcuhospital.org/	NaN	

1 df.dropna(axis=0, inplace=True)

2 df.dropna(axis=1, inplace=True)

3 df

	NAME OF HOSPITAL	CITY	CONTACT NUMBER	WEBSITE / EMAIL	
3	Dr. Jose N. Rodriguez Memorial Hospital and Sa	Caloocan	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.fac
5	Metro Balayan Medical Center	Caloocan	(043) 740 1350	http://www.metrobalayanmc.com.ph/	https://www.1
7	Las Pinas City Medical Center / City Med	Las Pinas	8806 2288, 8800 5695, 8800 5678	http://www.citymed.com.ph/	https://wv
8	Las Pinas Doctors Hospital	Las Pinas	0917 836 9466	lpdhinc.com	https://www.facebook.c
9	Las Pinas General Hospital	Las Pinas	8872 0509, 8873 0556, 8824 9435	https://lpghstc.doh.gov.ph/	https://www.faceboo
10	Perpetual Help Medical Center	Las Pinas	8874 8515	https://phmc.com.ph/	https://www.facebook.c
11	Makati Medical Center	Makati	8651 7800 loc 1149 or 1150	https://www.makatimed.net.ph/	https://www.f
13	St Clare's Medical Center	Makati	8831 6511 - 14, 8832 0057	https://www.stclares.ph/	https://www.facebook
14	VRP MC	Mandaluyong	84649999	http://www.vrp.com.ph/	https://www.face
16	Child Care Inc / Hospital of the Infant Jesus	Manila	87312771 – 76	http://hijmc.com/	https://ww
17	Chinese General Hospital and Medical Center	Manila	8711 4141	https://cghmc.com.ph/online/	https://www.t
18	Dr Jose Fabella Memorial Hospital	Manila	8734-5561	http://fabella.doh.gov.ph	https
19	Jose R. Reyes Memorial Medical Center	Manila	8711 9491	https://jrrmmc.gov.ph/	https://www.fa
20	Manila Doctors Hospital	Manila	85580888	http://www.maniladoctors.com.ph/	https://www.facebook
21	Mary Chiles General Hospital	Manila	87355341	http://marychiles.com/	https://www.facebool
22	Mary Johnston Hospital	Manila	5318 6600	http://maryjohnston.ph/	https://www.faceboo
23	Medical Center Manila / Manila Med	Manila	8523 8131	http://www.manilamed.com.ph/	https://ww
24	Metropolitan Medical Center	Manila	82541111	http://www.mmc.com.ph/	https://www
26	Our Lady of Lourdes Hospital	Manila	8716 3901 to 20	https://ollh-manila.com/	https://www.facebook.cc
	T 1 1 1 1				

Tondo Medical

32	Center	Manila	88659000	http://tmc.doh.gov.ph/	https://wv
34	Amang Rodriguez Memorial Medical Center	Marikina	8941 6289	http://www.armmc.doh.gov.ph/	https://www.f
35	Marikina Doctors Hospital	Marikina / Pasig	8535-3934	http://www.marikinadoctors.com/	https://www.fa
36	Marikina Valley Medical Center	Marikina	86822222	http://www.marikinavalleymedicalcenter.com/	https://www.facebook.
37	St. Anthony Medical Center Of Marikina, Inc.	Marikina	682 2000, 6823240, 6822000	http://www.saintanthony.com.ph/	https://www.facebook.c
38	Asian Hospital and Medical Center	Muntinlupa	8771 9000	https://www.asianhospital.com/	https://www.face
41	RITM	Muntinlupa	8807 2628, 8807 2632, 8072631/32/37	http://ritm.gov.ph/	https://v
42	Medical Center Paranaque	Paranaque	8825 6911, 8826 2096	http://www.medicalcenterparanaque.com/index.htm	https://www.facebook.c

¹ df_sorted = df.sort_values(by='CITY')

² df_sorted

	NAME OF HOSPITAL	CITY	CONTACT NUMBER	WEBSITE / EMAIL	
3	Dr. Jose N. Rodriguez Memorial Hospital and Sa	Caloocan	0966 549 2697, 8294 2571 to 73	http://djnrmh.doh.gov.ph/	https://www.fac
5	Metro Balayan Medical Center	Caloocan	(043) 740 1350	http://www.metrobalayanmc.com.ph/	https://www.1
7	Las Pinas City Medical Center / City Med	Las Pinas	8806 2288, 8800 5695, 8800 5678	http://www.citymed.com.ph/	https://ww
8	Las Pinas Doctors Hospital	Las Pinas	0917 836 9466	lpdhinc.com	https://www.facebook.c
9	Las Pinas General Hospital	Las Pinas	8872 0509, 8873 0556, 8824 9435	https://lpghstc.doh.gov.ph/	https://www.faceboo
10	Perpetual Help Medical Center	Las Pinas	8874 8515	https://phmc.com.ph/	https://www.facebook.c
11	Makati Medical Center	Makati	8651 7800 loc 1149 or 1150	https://www.makatimed.net.ph/	https://www.f
13	St Clare's Medical Center	Makati	8831 6511 - 14, 8832 0057	https://www.stclares.ph/	https://www.facebook
14	VRP MC	Mandaluyong	84649999	http://www.vrp.com.ph/	https://www.face
32	Tondo Medical Center	Manila	88659000	http://tmc.doh.gov.ph/	https://wv
26	Our Lady of Lourdes Hospital	Manila	8716 3901 to 20	https://ollh-manila.com/	https://www.facebook.cc
24	Metropolitan Medical Center	Manila	82541111	http://www.mmc.com.ph/	https://www
23	Medical Center Manila / Manila Med	Manila	8523 8131	http://www.manilamed.com.ph/	https://ww
22	Mary Johnston Hospital	Manila	5318 6600	http://maryjohnston.ph/	https://www.faceboo
20	Manila Doctors Hospital	Manila	85580888	http://www.maniladoctors.com.ph/	https://www.facebook
19	Jose R. Reyes Memorial Medical Center	Manila	8711 9491	https://jrrmmc.gov.ph/	https://www.fa
18	Dr Jose Fabella Memorial Hospital	Manila	8734-5561	http://fabella.doh.gov.ph	http:
17	Chinese General Hospital and Medical Center	Manila	8711 4141	https://cghmc.com.ph/online/	https://www.i
16	Child Care Inc / Hospital of the Infant Jesus	Manila	87312771 – 76	http://hijmc.com/	https://ww
21	Mary Chiles General	Manila	87355341	http://marvchiles.com/	https://www.facebool

	Hospital	mama	0,0000-1	map.//maryoniico.com/	парол/ питачесто
34	Amang Rodriguez Memorial Medical Center	Marikina	8941 6289	http://www.armmc.doh.gov.ph/	https://www.fa
36	Marikina Valley Medical Center	Marikina	86822222	http://www.marikinavalleymedicalcenter.com/	https://www.facebook.
37	St. Anthony Medical Center Of Marikina, Inc.	Marikina	682 2000, 6823240, 6822000	http://www.saintanthony.com.ph/	https://www.facebook.c
35	Marikina Doctors Hospital	Marikina / Pasig	8535-3934	http://www.marikinadoctors.com/	https://www.fa
41	RITM	Muntinlupa	8807 2628, 8807 2632, 8072631/32/37	http://ritm.gov.ph/	https://v
38	Asian Hospital and Medical Center	Muntinlupa	8771 9000	https://www.asianhospital.com/	https://www.face
46	UniHealth Paranaque	Paranaque	8832 0636, 8832 0641	http://unihealthparanaque.com/	https://www.facebc

7.2 Conclusion:

Write your conclusion here.

The only thing that I haven't done in this activity is to gather sensory, imagery, and voice data, yet it allowed me to understand a whole lot of things related to techniques related to preprocessing data—as evident as it is in the third exercise—as well as the complexity and challenge brought upon by the said processes. This activity helped me understand the importance of being versatile, quick, and smart in collecting and handling various types of data, to ensure a good and prepared data analysis. Lastly, this activity reminded me the importance of having a holistic approach in analyzing data to understand the data itself and derive meaningful insights from it—which is required if one wants to build a reliable data analysis.