7-1-data-collection-and-wrangling

March 20, 2024

1 Module 7: Data Wrangling with Pandas

1.1 CPE311 Computational Thinking with Python

Submitted by: Detchosa, Ralph Christian D.

Performed on: 03/20/2024

Submitted on: 03/20/2024

Submitted to: Engr. Roman M. Richard

2 7.1 Supplementary Activity

Using the datasets provided, perform the following exercises:

2.1 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.

[3]:
```
# 1.
import pandas as pd

fb_file = '/content/fb.csv'
aapl_file = '/content/aapl.csv'
amzn_file = '/content/amzn.csv'
nflx_file = '/content/nflx.csv'
goog_file = '/content/goog.csv'

fb_df = pd.read_csv(fb_file)
aapl_df = pd.read_csv(aapl_file)
```

```
amzn_df = pd.read_csv(amzn_file)
nflx_df = pd.read_csv(nflx_file)
goog_df = pd.read_csv(goog_file)
```

[4]:
```
# 2.
fb_df = fb_df.assign(ticker = 'FB')
aapl_df = aapl_df.assign(ticker = 'AAPL')
amzn_df = amzn_df.assign(ticker = 'AMZN')
nflx_df = nflx_df.assign(ticker = 'NFLX')
goog_df = goog_df.assign(ticker = 'GOOG')

fb_df.head()
```

[4]:

	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

[5]:
```
# 3.
faang = pd.concat([fb_df, aapl_df, amzn_df, nflx_df, goog_df], sort = False)
faang
```

[5]:

	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
..	…	…	…	…	…	…	…
246	2018-12-24	973.90	1003.54	970.1100	976.22	1590328	GOOG
247	2018-12-26	989.01	1040.00	983.0000	1039.46	2373270	GOOG
248	2018-12-27	1017.15	1043.89	997.0000	1043.88	2109777	GOOG
249	2018-12-28	1049.62	1055.56	1033.1000	1037.08	1413772	GOOG
250	2018-12-31	1050.96	1052.70	1023.5900	1035.61	1493722	GOOG

[1255 rows x 7 columns]

[6]:
```
# 4.
faang.to_csv('faang.csv')
```

3 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.

[7]:
```
faang = faang.assign(
    date = pd.to_datetime(faang.date),
    volume = lambda x: x.volume.astype('int')
).sort_values(by=['date', 'ticker'], ascending = False)

faang
```

[7]:

	date	open	high	low	close	volume	ticker
250	2018-12-31	260.1600	270.1001	260.0000	267.6600	13508920	NFLX
250	2018-12-31	1050.9600	1052.7000	1023.5900	1035.6100	1493722	GOOG
250	2018-12-31	134.4500	134.6400	129.9500	131.0900	24625308	FB
250	2018-12-31	1510.8000	1520.7600	1487.0000	1501.9700	6954507	AMZN
250	2018-12-31	157.8529	158.6794	155.8117	157.0663	35003466	AAPL
..	…	…	…	…	…	…	…
0	2018-01-02	196.1000	201.6500	195.4200	201.0700	10966889	NFLX
0	2018-01-02	1048.3400	1066.9400	1045.2300	1065.0000	1237564	GOOG
0	2018-01-02	177.6800	181.5800	177.5500	181.4200	18151903	FB
0	2018-01-02	1172.0000	1190.0000	1170.5100	1189.0100	2694494	AMZN
0	2018-01-02	166.9271	169.0264	166.0442	168.9872	25555934	AAPL

[1255 rows x 7 columns]

[8]:
```
faang.sort_values(by = 'volume', axis = 0, ascending=False).head(7)
```

[8]:

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

[9]:
```
faang = pd.melt(
    faang,
    id_vars = ['date', 'ticker'],
    value_vars = ['open', 'high', 'low', 'close', 'volume']
)

faang
```

[9]:
```
            date ticker variable         value
0     2018-12-31   NFLX     open  2.601600e+02
1     2018-12-31   GOOG     open  1.050960e+03
2     2018-12-31     FB     open  1.344500e+02
3     2018-12-31   AMZN     open  1.510800e+03
4     2018-12-31   AAPL     open  1.578529e+02
…            …    …       …            …
6270  2018-01-02   NFLX   volume  1.096689e+07
6271  2018-01-02   GOOG   volume  1.237564e+06
6272  2018-01-02     FB   volume  1.815190e+07
6273  2018-01-02   AMZN   volume  2.694494e+06
6274  2018-01-02   AAPL   volume  2.555593e+07

[6275 rows x 4 columns]
```

3.1 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.

[69]:
```
!pip install -q tabula-py
```

[93]:
```
import tabula

hsptl_pdf = 'https://new-axa-prod.s3.amazonaws.com/axa-com-ph/
  ↪81ee51cb-9236-476c-80b5-d56814b1d9e3_GHA-Hospitals.pdf'

table = tabula.read_pdf(hsptl_pdf ,pages = 'all')
df = pd.DataFrame(table[0])
df

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

[94]:
```
df.rename(
    columns={
        'BERMUDEZ POLYMEDIC HOSPITAL, INC.' : 'Hospital',
        '(02) 8961-3229' : 'Contact Number',
        '#391 MALARIA ROAD CAMIA' : 'Address',
        'CALOOCAN CITY' : 'City'
    }, inplace = True
)
df.loc[len(df.index)] = ['BERMUDEZ POLYMEDIC HOSPITAL, INC.', '(02) 8961-3229',␣
  ↪'#391 MALARIA ROAD CAMIA', 'CALOOCAN CITY']

df
```

[94]:

	Hospital	Contact Number	Address	City
0	ACEBEDO GENERAL HOSPITAL	(02) 8806-4298, (02) 8935-9139, (02) 8983-5363	849 GEN LUIS ST.	CALOOCAN CITY
1	MARTINEZ MEMORIAL HOSPITAL	(02) 8288-4574, (02) 8288-8861, (02) 8288-8863	198 A. MABINI ST.,	CALOOCAN CITY
2	MCU - FDT MEDICAL FOUNDATION INC.	(02) 8367-2031 [1104] ADMITTING, (02) 8367-\r2…	SAMSON ROAD	CALOOCAN CITY
3	NODADO GENERAL HOSPITAL - CALOOCAN	(02) 8962-8021 [124] ADMITTING, (02) 8962-8021…	AREA A	CALOOCAN CITY
4	NORTH CALOOCAN DOCTOR'S HOSPITAL	(02) 7501-9924 HMO, (02) 8961-5213 ADMITTING	L31, B10 QUIRINO HIGHWAY	CALOOCAN CITY
5	OUR LADY OF GRACE HOSPITAL, INC.	(02) 8361-1124, (02) 8361-1138	8TH AVENUE, COR.	CALOOCAN CITY
6	SAN LORENZO HOSPITAL HEALTH MANAGE-\rMENT CO.,…	(02) 8939-6042		
7	POPE JOHN PAUL II HOSPITAL AND MEDICAL\rCENTER…	(02) 8826-0285		
8	CLINIC SYSTEMS INC. (AMC) - TALON BRANCH	(02) 8873-6464, (02) 8874-0164, (02) 8874-2506		
9	LAS PIÑAS CITY MEDICAL CENTER	(02) 8800-5524, (02) 8800-5613, (02) 8800-5678…		
10	LAS PIÑAS DOCTOR'S HOSPITAL	(02) 8825-5236 [1066, 165,166], (0917) 851-443…		
11	UNIVERSITY OF PERPETUAL HELP DALTA MEDI-\rCAL …	(02) 8874-2582, (02) 8874-8515		
12	MAKATI MEDICAL CENTER	(02) 8815-9911, (02) 8888-8999, (02) 8892-5544		
13	ST. CLARE'S MEDICAL CENTER INC	(02) 8831-6511, (02) 8831-6512, (02) 8831-6513…		
14	DR. VICTOR R. POTENCIANO MEDICAL CENTER	(02) 8533-4188 ADMITTING, (02) 7621-4665 HMO,\…		
15	CHINESE GENERAL HOSPITAL & MEDICAL\rCENTER	(02) 8711-4141 [234] ADMITTING, (02) 8711-4141…		
16	OUR LADY OF LOURDES HOSPITAL	(02) 8716-8002, (02) 8716-8003, (02) 8716-8004…		
17	BERMUDEZ POLYMEDIC HOSPITAL, INC.	(02) 8961-3229		

```
6                          SUSANO ROAD     CALOOCAN CITY
7             545 ALABANG ZAPOTE ROAD,    LAS PIÑAS CITY
8                  ALABANG-ZAPOTE ROAD    LAS PIÑAS CITY
9           1314 MARCOS ALVAREZ AVENUE,    LAS PIÑAS CITY
10            8009 J. I. AGUILAR AVENUE    LAS PIÑAS CITY
11         REAL ST, ALABANG - ZAPOTE RD    LAS PIÑAS CITY
12  2 AMORSOLO ST. COR. DELA ROSA\rST.       MAKATI CITY
13                     1838 DIAN STREET       MAKATI CITY
14                              163 EDSA  MANDALUYONG CITY
15              286 BLUMENTRITT STREET       MANILA CITY
16                46 P. SANCHEZ STREET               NaN
17             #391 MALARIA ROAD CAMIA     CALOOCAN CITY
```

[95]:
```
df.dtypes
df
```

[95]:
```
                                          Hospital  \
0                           ACEBEDO GENERAL HOSPITAL
1                         MARTINEZ MEMORIAL HOSPITAL
2                   MCU - FDT MEDICAL FOUNDATION INC.
3                  NODADO GENERAL HOSPITAL - CALOOCAN
4                    NORTH CALOOCAN DOCTOR'S HOSPITAL
5                    OUR LADY OF GRACE HOSPITAL, INC.
6   SAN LORENZO HOSPITAL HEALTH MANAGE-\rMENT CO.,…
7   POPE JOHN PAUL II HOSPITAL AND MEDICAL\rCENTER…
8            CLINIC SYSTEMS INC. (AMC) - TALON BRANCH
9                       LAS PIÑAS CITY MEDICAL CENTER
10                        LAS PIÑAS DOCTOR'S HOSPITAL
11  UNIVERSITY OF PERPETUAL HELP DALTA MEDI-\rCAL …
12                              MAKATI MEDICAL CENTER
13                    ST. CLARE'S MEDICAL CENTER INC
14             DR. VICTOR R. POTENCIANO MEDICAL CENTER
15         CHINESE GENERAL HOSPITAL & MEDICAL\rCENTER
16                        OUR LADY OF LOURDES HOSPITAL
17                  BERMUDEZ POLYMEDIC HOSPITAL, INC.

                                       Contact Number  \
0      (02) 8806-4298, (02) 8935-9139, (02) 8983-5363
1      (02) 8288-4574, (02) 8288-8861, (02) 8288-8863
2   (02) 8367-2031 [1104] ADMITTING, (02) 8367-\r2…
3   (02) 8962-8021 [124] ADMITTING, (02) 8962-8021…
4        (02) 7501-9924 HMO, (02) 8961-5213 ADMITTING
5                      (02) 8361-1124, (02) 8361-1138
6                                       (02) 8939-6042
7                                       (02) 8826-0285
8      (02) 8873-6464, (02) 8874-0164, (02) 8874-2506
9   (02) 8800-5524, (02) 8800-5613, (02) 8800-5678…
```

```
10  (02) 8825-5236 [1066, 165,166], (0917) 851-443…
11                     (02) 8874-2582, (02) 8874-8515
12     (02) 8815-9911, (02) 8888-8999, (02) 8892-5544
13  (02) 8831-6511, (02) 8831-6512, (02) 8831-6513…
14  (02) 8533-4188 ADMITTING, (02) 7621-4665 HMO,\…
15  (02) 8711-4141 [234] ADMITTING, (02) 8711-4141…
16  (02) 8716-8002, (02) 8716-8003, (02) 8716-8004…
17                                    (02) 8961-3229
```

	Address	City
0	849 GEN LUIS ST.	CALOOCAN CITY
1	198 A. MABINI ST.,	CALOOCAN CITY
2	SAMSON ROAD	CALOOCAN CITY
3	AREA A	CALOOCAN CITY
4	L31, B10 QUIRINO HIGHWAY	CALOOCAN CITY
5	8TH AVENUE, COR.	CALOOCAN CITY
6	SUSANO ROAD	CALOOCAN CITY
7	545 ALABANG ZAPOTE ROAD,	LAS PIÑAS CITY
8	ALABANG-ZAPOTE ROAD	LAS PIÑAS CITY
9	1314 MARCOS ALVAREZ AVENUE,	LAS PIÑAS CITY
10	8009 J. I. AGUILAR AVENUE	LAS PIÑAS CITY
11	REAL ST, ALABANG - ZAPOTE RD	LAS PIÑAS CITY
12	2 AMORSOLO ST. COR. DELA ROSA\rST.	MAKATI CITY
13	1838 DIAN STREET	MAKATI CITY
14	163 EDSA	MANDALUYONG CITY
15	286 BLUMENTRITT STREET	MANILA CITY
16	46 P. SANCHEZ STREET	NaN
17	#391 MALARIA ROAD CAMIA	CALOOCAN CITY

[96]:
```
df['Hospital'] = df['Hospital'].astype('string')
df.dtypes
```

[96]:
```
Hospital          string
Contact Number    object
Address           object
City              object
dtype: object
```

[97]:
```
df['Hospital']= df['Hospital'].apply(str.lower).str.capitalize()
df['Address']= df['Address'].apply(str.lower).str.capitalize()
df['City']= df['City'].str.lower().str.capitalize()
df['Contact Number']= df['Contact Number'].str.lower().str.capitalize()
df
```

[97]:

	Hospital	\
0	Acebedo general hospital	
1	Martinez memorial hospital	

```
2                  Mcu - fdt medical foundation inc.
3                 Nodado general hospital - caloocan
4                   North caloocan doctor's hospital
5                   Our lady of grace hospital, inc.
6   San lorenzo hospital health manage-\rment co.,…
7   Pope john paul ii hospital and medical\rcenter…
8           Clinic systems inc. (amc) - talon branch
9                     Las piñas city medical center
10                      Las piñas doctor's hospital
11  University of perpetual help dalta medi-\rcal …
12                             Makati medical center
13                    St. clare's medical center inc
14           Dr. victor r. potenciano medical center
15         Chinese general hospital & medical\rcenter
16                      Our lady of lourdes hospital
17                  Bermudez polymedic hospital, inc.

                                        Contact Number  \
0      (02) 8806-4298, (02) 8935-9139, (02) 8983-5363
1      (02) 8288-4574, (02) 8288-8861, (02) 8288-8863
2   (02) 8367-2031 [1104] admitting, (02) 8367-\r2…
3   (02) 8962-8021 [124] admitting, (02) 8962-8021…
4        (02) 7501-9924 hmo, (02) 8961-5213 admitting
5                      (02) 8361-1124, (02) 8361-1138
6                                       (02) 8939-6042
7                                       (02) 8826-0285
8      (02) 8873-6464, (02) 8874-0164, (02) 8874-2506
9   (02) 8800-5524, (02) 8800-5613, (02) 8800-5678…
10  (02) 8825-5236 [1066, 165,166], (0917) 851-443…
11                     (02) 8874-2582, (02) 8874-8515
12     (02) 8815-9911, (02) 8888-8999, (02) 8892-5544
13  (02) 8831-6511, (02) 8831-6512, (02) 8831-6513…
14  (02) 8533-4188 admitting, (02) 7621-4665 hmo,\…
15  (02) 8711-4141 [234] admitting, (02) 8711-4141…
16  (02) 8716-8002, (02) 8716-8003, (02) 8716-8004…
17                                      (02) 8961-3229

                              Address             City
0                     849 gen luis st.     Caloocan city
1                   198 a. mabini st.,     Caloocan city
2                          Samson road     Caloocan city
3                               Area a     Caloocan city
4              L31, b10 quirino highway     Caloocan city
5                      8th avenue, cor.     Caloocan city
6                           Susano road     Caloocan city
7              545 alabang zapote road,    Las piñas city
8                   Alabang-zapote road    Las piñas city
```

```
9          1314 marcos alvarez avenue,    Las piñas city
10           8009 j. i. aguilar avenue    Las piñas city
11        Real st, alabang - zapote rd    Las piñas city
12  2 amorsolo st. cor. dela rosa\rst.       Makati city
13                    1838 dian street       Makati city
14                             163 edsa  Mandaluyong city
15              286 blumentritt street       Manila city
16                46 p. sanchez street              NaN
17             #391 malaria road camia     Caloocan city
```

[98]:
```
df = df.ffill()
df
```

[98]:
```
                                          Hospital  \
0                           Acebedo general hospital
1                          Martinez memorial hospital
2                   Mcu - fdt medical foundation inc.
3                  Nodado general hospital - caloocan
4                    North caloocan doctor's hospital
5                    Our lady of grace hospital, inc.
6   San lorenzo hospital health manage-\rment co.,…
7   Pope john paul ii hospital and medical\rcenter…
8            Clinic systems inc. (amc) - talon branch
9                       Las piñas city medical center
10                        Las piñas doctor's hospital
11  University of perpetual help dalta medi-\rcal …
12                              Makati medical center
13                     St. clare's medical center inc
14            Dr. victor r. potenciano medical center
15         Chinese general hospital & medical\rcenter
16                        Our lady of lourdes hospital
17                  Bermudez polymedic hospital, inc.

                                       Contact Number  \
0      (02) 8806-4298, (02) 8935-9139, (02) 8983-5363
1      (02) 8288-4574, (02) 8288-8861, (02) 8288-8863
2   (02) 8367-2031 [1104] admitting, (02) 8367-\r2…
3   (02) 8962-8021 [124] admitting, (02) 8962-8021…
4        (02) 7501-9924 hmo, (02) 8961-5213 admitting
5                      (02) 8361-1124, (02) 8361-1138
6                                      (02) 8939-6042
7                                      (02) 8826-0285
8      (02) 8873-6464, (02) 8874-0164, (02) 8874-2506
9   (02) 8800-5524, (02) 8800-5613, (02) 8800-5678…
10  (02) 8825-5236 [1066, 165,166], (0917) 851-443…
11                     (02) 8874-2582, (02) 8874-8515
12     (02) 8815-9911, (02) 8888-8999, (02) 8892-5544
```

```
13  (02) 8831-6511, (02) 8831-6512, (02) 8831-6513…
14  (02) 8533-4188 admitting, (02) 7621-4665 hmo,\…
15  (02) 8711-4141 [234] admitting, (02) 8711-4141…
16  (02) 8716-8002, (02) 8716-8003, (02) 8716-8004…
17                                    (02) 8961-3229
```

	Address	City
0	849 gen luis st.	Caloocan city
1	198 a. mabini st.,	Caloocan city
2	Samson road	Caloocan city
3	Area a	Caloocan city
4	L31, b10 quirino highway	Caloocan city
5	8th avenue, cor.	Caloocan city
6	Susano road	Caloocan city
7	545 alabang zapote road,	Las piñas city
8	Alabang-zapote road	Las piñas city
9	1314 marcos alvarez avenue,	Las piñas city
10	8009 j. i. aguilar avenue	Las piñas city
11	Real st, alabang - zapote rd	Las piñas city
12	2 amorsolo st. cor. dela rosa\rst.	Makati city
13	1838 dian street	Makati city
14	163 edsa	Mandaluyong city
15	286 blumentritt street	Manila city
16	46 p. sanchez street	Manila city
17	#391 malaria road camia	Caloocan city

4 7.2 Conclusion:

In conclusion to this Hands-on Activity 7.1, it introduces the basic manipulation of datasets using pandas and performing certain operation and exercises. Learning the basic manipulation of datas using pandas is important for it serves as a backbone of data analysis. Pandas offers a lot of option to the programmer to do certain tasks and solutions to a problem, therefore, pandas plays a vital role in data analysis.