Module 7: Data Wrangling with Pandas

CPE311 Computational Thinking With Python

Name: Borja, Angelo Louis C.

Section: CPE22S3

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.

```
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

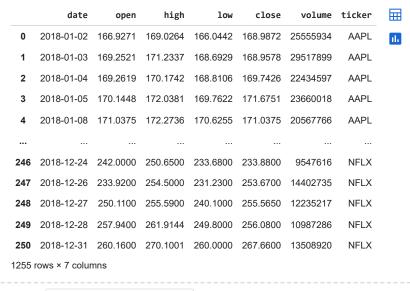
apple_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/aapl.csv')
amzn_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/amzn.csv')
fb_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/fb.csv')
google_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/goog.csv')
netflix_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/nflx.csv')
```

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.

```
apple_df ['ticker'] = ('AAPL')
amzn_df ['ticker'] = ('AMZN')
fb_df ['ticker'] = ('FB')
google_df ['ticker'] = ('GOOG')
netflix_df ['ticker'] = ('NFLX')
```

3.Append them together into a single dataframe.

```
merged_df = pd.concat([apple_df, amzn_df, fb_df, google_df, netflix_df])
merged_df
```



Next steps:

View recommended plots

4. Save the result in a CSV file called faang.csv

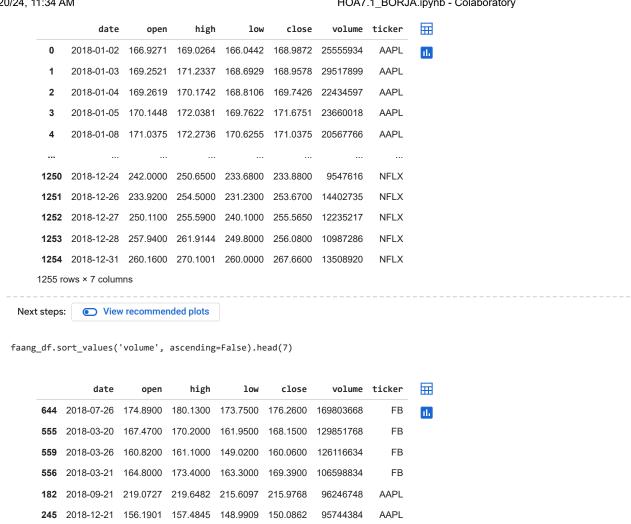
```
merged_df.to_csv('/content/drive/MyDrive/HOA7 files/faang.csv', index = False)
```

Exercise 2

faang_df

- · With faang, use type conversion to change the date column into a datetime and the volume column into integers. Then, sort by date and
- · Find the seven rows with the highest value for volume.
- · Right now, the data is somewehere between long and wide format. Use melt to make it completely long format. Hint: date and ticker are our ID Variables (the 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_df = pd.read_csv('/content/drive/MyDrive/HOA7 files/faang.csv')
faang_df['date'] = pd.to_datetime(faang_df["date"])
faang_df.sort_values(['date', 'ticker'], ascending=(True, True))
faang_df.dtypes
     date
               datetime64[ns]
                      float64
     open
                      float64
     high
     low
                      float64
     close
                      float64
                        int64
     volume
     ticker
                       object
     dtype: object
```



91328654

AAPL

melted_faang = pd.melt(faang_df, id_vars=['date','ticker'], var_name='Type', value_name='value') melted faang



212 2018-11-02 207.9295 211.9978 203.8414 205.8755

Exercise 3

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

```
from bs4 import BeautifulSoup
import requests as rq
req = rq.get("https://sulit.ph/list-of-hospitals-in-metro-manila-with-contact-details-website-and-social-media-accounts/")
bSoup = BeautifulSoup(req.content, 'html.parser')
rowheader = []
for x in bSoup.find_all('tr'):
  for y in x.find_all('th'):
    rowheader.append(y.text)
    hospitalTable = []
for x in bSoup.find_all('tr')[3:]:
  tdfind = x.find_all('td')
  tdval = [y.text for y in tdfind]
  hospitalTable.append(tdval)
hospitals = pd.DataFrame(hospitalTable, columns= rowheader).drop(["WEBSITE / EMAIL","FACEBOOK LINK"], axis = 1)
hospitals
                CITY
                                                NAME OF HOSPITAL
                                                                                                  CONTACT NUMBER
                                                                                                                    丽
           Caloocan
                                        Caloocan City Medical Center South 5310 7925, North 8282 3397, 0943 216 6963
      0
            Caloocan Dr. Jose N. Rodriguez Memorial Hospital and Sa...
                                                                                    0966 549 2697, 8294 2571 to 73
       1
                             MCU - FDT Medical Foundations Hospital
       2
                                                                                                        8367 2031
            Caloocan
       3
            Caloocan
                                       Metro Balayan Medical Center
                                                                                                    (043) 740 1350
       4
           Las Pinas
                                             Alabang Medical Center
                                                                                             8807 8189, 8850 8719
      91
                                            Medical Center of Taguig
                                                                                                        8888 6284
              Taguig
                             Allied Care Experts (ACE) Medical Center
                                                                      direct line to Admission 0917 844 3654, 8332 0...
      92 Valenzuela
      93 Valenzuela
                                     Fatima University Medical Center
                                                                                                        8291 6538
      94 Valenzuela
                                   Valenzuela Citicare Medical Center
                                                                                             8860 9300, 8860 9300
      95 Valenzuela
                                           Valenzuela Medical Center
                                                                                                        8294 6711
     96 rows × 3 columns
```

7.2 Conclusion

Next steps:

View recommended plots

In this hands-on activity, I learned how to apply data-wrangling techniques using the Pandas library. In more detail, these are the things I performed on the given four csv tables: I've added an additional column named 'ticker' to all four of the dataframes while also adding a value for that column to all of the entries, merged all of them into one dataframe, changed a column's data type, performed sorting, and reshaped a dataframe by using the melt() method. The most important thing I have learned is how to perform web scraping using the BeautifulSoup library.