#### **DSCI 560 Lab 3 Report**

#### **Group Code Submission Link:**

https://github.com/shubhamdarekar/DSCI560---Shubham/tree/main/Lab 3

#### Code README:

DSCI560---Shubham/Lab 3/README.md at main · shubhamdarekar/DSCI560---Shubham

### **Meeting Notes Link:**

https://docs.google.com/document/d/1Yx0zCHnpvug2czdTALt-7H9Z32BcRcTYsCPhUPD4qEw/edit?usp=sharing

# Team Details: Group 1

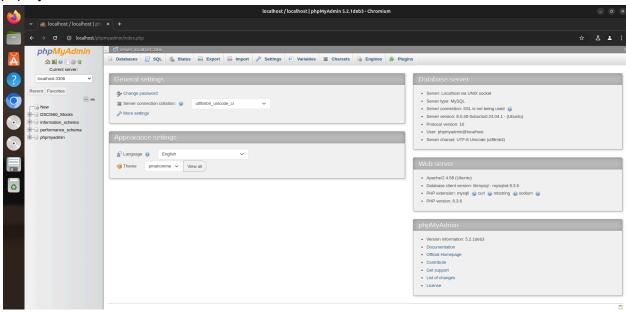
Saavani Vaidya 9385579920 Yuxuan Liu 4780355176 Shubham Darekar 1641138809

# Task 1: Database Installation and Configuration

#### Console:

```
shubhamdarekar@shubham-darekar-DSCI560:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or ackslash g .
Your MySQL connection id is 29
Server version: 8.0.40-0ubuntu0.24.04.1 (Ubuntu)
Copyright (c) 2000, 2024, Oracle and/or its affiliates.
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> show databases;
| Database
| DSCI560 Stocks
 information_schema
 mysql
 performance_schema
 phpmyadmin
6 rows in set (0.00 sec)
mysql>
```

### phpMyAdmin:



# Task 2: Technologies Used

Database Storage: MySQL

Database Interface: phpMyAdmin Scraping: Python and yFinancepytho

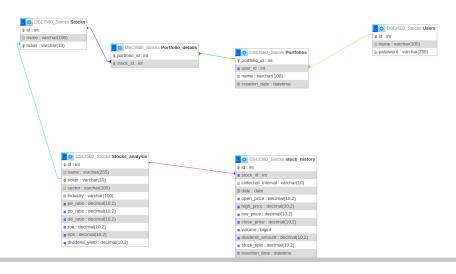
**Environment:** 

```
shubhamdarekar@shubham-darekar-DSCI560:~/DSCI560---Shubham/Lab_3$ source .venv/bin/activate
(.venv) shubhamdarekar@shubham-darekar-DSCI560:~/DSCI560---Shubham/Lab_3$ pip list
Package
                        Version
beautifulsoup4 4.12.3
certifi 2024.12.14
charset-normalizer 3.4.1
frozendict 2.4.6
html5lib 1.1
idna 3.10
lxml 5.3.0
multitasking 0.0.11
mysql-connector-python 9.2.0
numpy
pandas
                         2.2.3
                        3.17.8
peewee
                        24.0
4.3.6
pip
platformdirs
python-dateutil
                     2.9.0.post0
pytz
                        2024.2
requests
                          2.32.3
                         1.17.0
six
soupsieve
                         2.6
                          2025.1
tzdata
urllib3
                          2.3.0
webencodings
                          0.5.1
yfinance
                          0.2.52
```

# Task 3: Data Collection / Storage

#### **Database Schema:**

```
Users:
2
         Id
3
         Name
         Password
4
5
6
     Portfolios:
         portfolio id
8
         user_id
9
         name
         creation date
12
     Portfolio_details:
13
         portfolio_id
14
         stock id
    Stocks_analysis:
16
17
         id
18
         name
19
         ticker
20
         sector
21
         industry
22
         pe ratio
23
         pb ratio
         de_ratio
24
25
         roe
26
         eps
27
         dividend_yield
28
29
30
     stock history:
31
         stock id,
32
         date,
         open_price,
33
34
         high price,
35
         low price,
36
         close_price,
37
         volume,
38
         dividend amount,
         stock split
39
40
```



#### **Functions:**

1. Create a new user

```
(.venv) shubhamdarekar@shubham-darekar-DSCI560:~/DSCI560---Shubham$ python3 Lab_3/scripts/scrape_stocks.py
Database initialized successfully

1. Create User
2. Login
3. Exit
Enter your choice: 1
Enter your username: grace
Create new password: Grace@123
User created successfully
```

#### 2. Login

- 1. Create User
- 2. Login
- 3. Exit

Enter your choice: 2

Enter your username: grace

Enter your password: Grace@123

Logged in successfully

- 3. User password and name validation
  - 1. Create User
  - 2. Login
  - 3. Exit

Enter your choice: 2

Enter your username: graceL

User does not exist

- 1. Create User
- 2. Login
- 3. Exit

Enter your choice: 2

Enter your username: grace Enter your password: 1234

Invalid password

#### 4. Create a new portfolio

- 1. Create Portfolio
- 2. Select Existing Portfolio
- Logout

Enter your choice: 1

Enter portfolio name: Big Pharma Portfolio created successfully

### 5. Select and list stocks in portfolio

- 1. Create Portfolio
- 2. Select Existing Portfolio
- Logout

Enter your choice: 2

Select a portfolio

1. Big Pharma

Enter your choice: 1

No stocks in portfolio Big Pharma Stocks in Portfolio Big Pharma

# Operations

- 1. Add Stock
- 2. Delete Stocks
- 3. Go to previous menu

Enter your choice:

## 6. Add stock to portfolio

## Operations

- 1. Add Stock
- 2. Delete Stocks
- 3. Go to previous menu

Enter your choice: 1

Enter ticker name: PFE

Stock added successfully

Stocks in Portfolio Big Pharma 1. Pfizer, Inc.

# Operations

- 1. Add Stock
- 2. Delete Stocks
- 3. Go to previous menu

Enter your choice: 1

Enter ticker name: JNJ

Stock added successfully

Stocks in Portfolio Big Pharma

- Pfizer, Inc.
- Johnson & Johnson

# Operations

- 1. Add Stock
- 2. Delete Stocks
- 3. Go to previous menu

Enter your choice:

#### 7. Delete stock from portfolio

```
Operations
1. Add Stock
2. Delete Stocks
3. Go to previous menu
Enter your choice: 1
Enter ticker name: JNJ
Stock added successfully
Stocks in Portfolio Big Pharma
1. Pfizer, Inc.
2. Johnson & Johnson
 Operations
1. Add Stock
2. Delete Stocks
3. Go to previous menu
Enter your choice: 2
Enter stock number to delete: 1
Stock deleted from portfolio successfully
Stocks in Portfolio Big Pharma
1. Johnson & Johnson
 Operations
1. Add Stock
2. Delete Stocks
3. Go to previous menu
Enter your choice:
```

### 8. Stocks ticker validation

```
Operations

1. Add Stock

2. Delete Stocks

3. Go to previous menu
Enter your choice: 1
Enter ticker name: Abek

404 Client Error: Not Found for url: https://query2.finance.yahoo.com/v10/finance/quoteail&corsDomain=finance.yahoo.com&formatted=false&symbol=ABEK&crumb=dCeQikO2Al0
Ticker not recognized
```

9. Fetch the stock price data for an input date range and save it to database

```
Stocks in Portfolio Big Pharma

    Pfizer, Inc.

Johnson & Johnson
3. AbbVie Inc.
 Operations
1. Add Stock
Delete Stocks
3. Show/Save Stock History
4. Go to previous menu
Enter your choice: 3
Enter period (1d, 1mo, 1y, 5y): 3d
Enter interval (1d, 1wk, 1mo): 1d
Saved_history of Stocks: PFE Period: 3d Interval: 1d
Stock: PFE Period: 3d Interval: 1d History saved in db, showing first 5 rows
                                                            Volume Dividends Stock Splits
                              Open High
                                            Low
                                                    Close
Date
2025-01-29 00:00:00-05:00 26.799999 27.07 26.57 26.620001 34418200
                                                                          0.0
                                                                                       0.0
2025-01-30 00:00:00-05:00 26.660000 27.01 26.58 26.910000 32289600
                                                                          0.0
                                                                                       0.0
2025-01-31 00:00:00-05:00 26.959999 27.01 26.42 26.520000 37269448
                                                                          0.0
                                                                                       0.0
Saved_history of Stocks: JNJ Period: 3d Interval: 1d
Stock: JNJ Period: 3d Interval: 1d History saved in db, showing first 5 rows
                                                                       Volume Dividends Stock Splits
                               Open
                                          High
                                                      Low
                                                                Close
2025-01-29 00:00:00-05:00 150.500000 151.880005 150.000000 151.149994 7156400
                                                                                                  0.0
                                                                                    0.0
2025-01-30 00:00:00-05:00 151.910004 154.139999 151.360001 152.869995 8911000
                                                                                                 0.0
2025-01-31 00:00:00-05:00 152.619995 153.179993 152.054993 152.149994 5281291
                                                                                    0.0
                                                                                                 0.0
Saved_history of Stocks: ABBV Period: 3d Interval: 1d
Stock: ABBV Period: 3d Interval: 1d History saved in db, showing first 5 rows
                                                                Close Volume Dividends Stock Splits
                               Open
                                          High
                                                      Low
Date
2025-01-29 00:00:00-05:00 174.500000 176.919998 174.300003 175.259995 3748300
                                                                                    0.0
                                                                                                  0.0
2025-01-30 00:00:00-05:00 176.009995 176.119995 173.979996
                                                           175.649994
                                                                      5209200
                                                                                    0.0
                                                                                                  0.0
2025-01-31 00:00:00-05:00 187.300003 191.735001 183.669998 183.899994 9434227
                                                                                    0.0
                                                                                                  0.0
```

### Task 4: Data Preprocessing

- 1. Converting Date Format: Converted date column into a proper datetime format. Ensured chronological sorting by stock id and date.
- 2. Handling Missing Values: Used forward fill to fill missing values based on previous data. Used backward fill as a backup for remaining missing values. Dropped any rows where NaN values still remained.
- 3. Feature Engineering:

Daily Return (daily return): Measures percentage change in closing price.

Log Return (log\_return): More stable alternative to percentage change. Moving Averages (5-day & 20-day): Identifies short-term and long-term trends. Volatility (5-day standard deviation of daily returns): Measures risk. Cumulative Return (cumulative\_return): Helps analyze overall trends. Volume Change Percentage (volume\_change): Detects shifts in trading activity.

stock\_history

id	stock_id	collected_interval	date	open_price	high_price	low_price	close_price	volume	dividend_amount	stock_split	insertion_time
1	1	1d	2024-01-31	186.13	186.19	183.45	183.50	55467800	0.00	0.00	2025-01-31 09:40:31
2	1	1d	2024-02-01	183.09	186.04	182.92	185.95	64885400	0.00	0.00	2025-01-31 09:40:31
3	1	1d	2024-02-02	178.98	186.42	178.38	184.94	102518000	0.00	0.00	2025-01-31 09:40:31
4	1	1d	2024-02-05	187.23	188.33	184.93	186.77	69668800	0.00	0.00	2025-01-31 09:40:31
5	1	1d	2024-02-06	185.95	188.39	185.86	188.38	43490800	0.00	0.00	2025-01-31 09:40:31
6	1	1d	2024-02-07	189.71	190.12	187.69	188.49	53439000	0.00	0.00	2025-01-31 09:40:31
7	1	1d	2024-02-08	188.47	188.62	186.44	187.40	40962000	0.00	0.00	2025-01-31 09:40:31
8	1	1d	2024-02-09	187.97	189.31	187.32	188.17	45155200	0.24	0.00	2025-01-31 09:40:31
9	1	1d	2024-02-12	187.74	187.99	186.12	186.48	41781900	0.00	0.00	2025-01-31 09:40:31
10	1	1d	2024-02-13	185.10	185.54	182.85	184.37	56529500	0.00	0.00	2025-01-31 09:40:31
11	1	1d	2024-02-14	184.65	184.86	181.78	183.49	54630500	0.00	0.00	2025-01-31 09:40:31
12	1	1d	2024-02-15	182.89	183.82	180.70	183.20	65434500	0.00	0.00	2025-01-31 09:40:31
13	1	1d	2024-02-16	182.76	184.18	181.02	181.65	49701400	0.00	0.00	2025-01-31 09:40:31
14	1	1d	2024-02-20	181.13	181.77	179.35	180.91	53665600	0.00	0.00	2025-01-31 09:40:31
15	1	1d	2024-02-21	181.28	182.23	180.01	181.66	41529700	0.00	0.00	2025-01-31 09:40:31
16	1	1d	2024-02-22	182.82	184.29	181.80	183.71	52292200	0.00	0.00	2025-01-31 09:40:31
17	1	1d	2024-02-23	184.34	184.37	181.57	181.86	45119700	0.00	0.00	2025-01-31 09:40:31
18	1	1d	2024-02-26	181.58	182.10	180.00	180.51	40867400	0.00	0.00	2025-01-31 09:40:31
19	1	1d	2024-02-27	180.45	183.26	178.91	181.97	54318900	0.00	0.00	2025-01-31 09:40:31
20	1	1d	2024-02-28	181.85	182.46	179.48	180.77	48953900	0.00	0.00	2025-01-31 09:40:31
21	1	1d	2024-02-29	180.62	181.91	178.88	180.10	136682600	0.00	0.00	2025-01-31 09:40:31
22	1	1d	2024-03-01	178.90	179.88	176.74	179.01	73488000	0.00	0.00	2025-01-31 09:40:31
23	1	1d	2024-03-04	175.51	176.26	173.16	174.47	81510100	0.00	0.00	2025-01-31 09:40:31
24	1	1d	2024-03-05	170.14	171.42	169.01	169.51	95132400	0.00	0.00	2025-01-31 09:40:31

stock_id collected_interval	date	open_price	high_price	low_price	close_price	volume	dividend_amount stoc	ck_split	insertion_time	daily_return	log_return	moving_avg_5d	moving_avg_20d	volatility_5d	cumulative_return	volume_change
1 1d	2024-01-31	186.13	186.19	183.45	183.5	55467800	0.0	0.0	2025-01-31 09:40:31							
1 1d	2024-02-01	183.09	186.04	182.92	185.95	64885400	0.0	0.0	2025-01-31 09:40:31	0.01335149863760220	0.013263152876454100				0.01335149863760220	0.1697849923739540
1 1d	2024-02-02	178.98	186.42	178.38	184.94	102518000	0.0	0.0	2025-01-31 09:40:31	-0.0054315676257057400	-0.005446372221586390				0.007919931011896480	0.57998563621400
1 1d	2024-02-05	187.23	188.33	184.93	186.77	69668800	0.0	0.0	2025-01-31 09:40:31	0.009895101113874820	0.009846465176023120				0.01781503212577130	-0.32042373046684
1 1d	2024-02-06	185.95	188.39	185.86	188.38	43490800	0.0	0.0	2025-01-31 09:40:31	0.008620228088022540	0.008583286069155270	185.908			0.026435260213793800	-0.375749259352824
1 1d	2024-02-07	189.71	190.12	187.69	188.49	53439000	0.0	0.0	2025-01-31 09:40:31	0.0005839261068054660	0.0005837556882943460	186.906		0.0076537488346986600	0.027019186320599300	0.228742630625328
1 1d	2024-02-08	188.47	188.62	186.44	187.4	40962000	0.0	0.0	2025-01-31 09:40:31	-0.005782800148549040	-0.005799585278643570	187.196		0.0074677764794154600	0.02123638617205030	-0.233481165440970
1 1d	2024-02-09	187.97	189.31	187.32	188.17	45155200	0.24	0.0	2025-01-31 09:40:31	0.0041088580576307300	0.004100439752231320	187.842		0.006367054106684370	0.025345244229681000	0.10236804843513
1 1d	2024-02-12	187.74	187.99	186.12	186.48	41781900	0.0	0.0	2025-01-31 09:40:31	-0.008981240367752610	-0.009021814829051490	187.78400000000000		0.007162652721276040	0.016364003861928400	-0.074704574445468
1 1d	2024-02-13	185.1	185.54	182.85	184.37	56529500	0.0	0.0	2025-01-31 09:40:31	-0.011314886314886200	-0.011379386645202200	186.98200000000000		0.006478432932114870	0.005049117547042180	0.352966236576
1 1d	2024-02-14	184.65	184.86	181.78	183.49	54630500	0.0	0.0	2025-01-31 09:40:31	-0.0047730107935130000	-0.004784437985437140	185.98200000000000		0.005889736709879880	0.0002761067535291820	-0.03359307971943
1 1d	2024-02-15	182.89	183.82	180.7	183.2	65434500	0.0	0.0	2025-01-31 09:40:31	-0.0015804676004143000	-0.0015817178568324400	185.142		0.006108078667562470	-0.001304360846885120	0.1977649847612
1 1d	2024-02-16	182.76	184.18	181.02	181.65	49701400	0.0	0.0	2025-01-31 09:40:31	-0.008460698689956250	-0.008496693572818860	183.838		0.0038407543835759500	-0.00976505953684137	-0.24044044044044
1 1d	2024-02-20	181.13	181.77	179.35	180.91	53665600	0.0	0.0	2025-01-31 09:40:31	-0.004073768235618000	-0.004082088634041230	182.72400000000000		0.003841970991641950	-0.013838827772459400	0.07976032868289
1 1d	2024-02-21	181.28	182.23	180.01	181.66	41529700	0.0	0.0	2025-01-31 09:40:31	0.004145707810513600	0.004137138040897800	182.182		0.004668476724976250	-0.009693119961945770	-0.22613927730240
1 1d	2024-02-22	182.82	184.29	181.8	183.71	52292200	0.0	0.0	2025-01-31 09:40:31	0.011284817791478600	0.011221619246778300	182.226000000000000		0.007662648655277270	0.0015916978295328300	0.2591518840733
1 1d	2024-02-23	184.34	184.37	181.57	181.86	45119700	0.0	0.0	2025-01-31 09:40:31	-0.010070219367481300	-0.010121267023078900	181.958		0.008997619373780760	-0.00847852153794848	-0.137161947670
1 1d	2024-02-26	181.58	182.1	180.0	180.51	40867400	0.0	0.0	2025-01-31 09:40:31	-0.0074232926426922600	-0.00745098239739679	181.730000000000000		0.008805001898511470	-0.01590181418064070	-0.09424486421673
1 1d	2024-02-27	180.45	183.26	178.91	181.97	54318900	0.0	0.0	2025-01-31 09:40:31	0.00808819455985832	0.008055660424781390	181.942000000000000		0.009476376709866290	-0.00781361962078242	0.32914988474921
1 1d	2024-02-28	181.85	182.46	179.48	180.77	48953900	0.0	0.0	2025-01-31 09:40:31	-0.0065944935978457000	-0.006616333338388220	181.764000000000000	184.209	0.00985294003233495	-0.014408113218628100	-0.09876856858294
1 1d	2024-02-29	180.62	181.91	178.88	180.1	136682600	0.0	0.0	2025-01-31 09:40:31	-0.003706367206948170	-0.0037132528048527200	181.042000000000000	184.03900000000000	0.007097381460910570	-0.018114480425576300	1.7920676391462
1 1d	2024-03-01	178.9	179.88	176.74	179.01	73488000	0.0	0.0	2025-01-31 09:40:31	-0.006052193225985600	-0.00607058197981022	180.472000000000000	183.692	0.006425614524670600	-0.024166673651561900	-0.4623456094630
1 1d	2024-03-04	175.51	176.26	173.16	174.47	81510100	0.0	0.0	2025-01-31 09:40:31	-0.025361711636221400	-0.025688863110289700	179.264000000000000	183.1685	0.012003536957931000	-0.04952838528778330	0.1091620400609
1 1d	2024-03-05	170.14	171.42	169.01	169.51	95132400	0.0	0.0	2025-01-31 09:40:31	-0.02842895626755320	-0.028840884961415500	176.772	182.3055	0.011845327905384500	-0.0779573415553364	0.16712407419448
1 1d	2024-03-06	170.44	170.62	168.07	168.51	68587700	0.0	0.0	2025-01-31 09:40:31	-0.005899356970090250	-0.005916826917943880	174.32	181.31200000000000	0.011957940716696600	-0.08385669852542670	-0.2790290164024
1 1d	2024-03-07	168.54	170.11	167.88	168.39	71765100	0.0	0.0	2025-01-31 09:40:31	-0.0007121239095603250	-0.0007123775902334740	171.978	180.307	0.012650342220354500	-0.08456882243498700	0.04632609053809
1 1d	2024-03-08	168.39	173.07	168.33	170.11	76114600	0.0	0.0	2025-01-31 09:40:31	0.010214383276916800	0.010162568999274000	170.198	179.44250000000000	0.016487137176636500	-0.07435443915807020	0.06060745404103
1 1d	2024-03-11	172.32	173.75	171.43	172.13	60139500	0.0	0.0	2025-01-31 09:40:31	0.011874669331608700	0.011804718661922800	169.73	178.6405	0.0162427584845881	-0.06247976982646140	-0.20988220393984
1 1d	2024-03-12	172.53	173.4	170.39	172.61	59825400	0.0	0.0	2025-01-31 09:40:31	0.002788590019171760	0.0027847091151835500	170.35	177.947	0.007444740288952170	-0.05969117980728970	-0.00522285685780
1 1d	2024-03-13	172.15	172.57	170.14	170.51	52488700	0.0	0.0	2025-01-31 09:40:31	-0.012166154915706100	-0.012240768368537700	170.75	177.254	0.009654573156291530	-0.07185733472299570	-0.1226352017704
1 1d	2024-03-14	172.29	173.68	171.43	172.38	72913500	0.0	0.0	2025-01-31 09:40:31	0.010967098703888300	0.010907396189193000	171.548	176.6985	0.01011527459143800	-0.06089023601910740	0.3891275645996
1 1d	2024-03-15	170.55	172.0	169.68	172.0	121664700	0.0	0.0	2025-01-31 09:40:31	-0.00220443206868548	-0.0022068654058001600	171.926	176.1385	0.00995715342452726	-0.06309466808779290	0.668616922792
1 1d	2024-03-18	174.94	177.07	172.89	173.09	75604200	0.0	0.0	2025-01-31 09:40:31	0.006337209302325510	0.006317213624854300	172.118	175.7105	0.008867631976262970	-0.05675745878546730	-0.37858557165718
1 1d	2024-03-19	173.71	175.97	172.41	175.45	55215200	0.0	0.0	2025-01-31 09:40:31	0.013634525391414700	0.013542411590916600	172.68600000000000	175.4375	0.010539345884682600	-0.0431229333940526	-0.26968078492993
1 1d	2024-03-20	175.09	178.03	174.46	178.03	53423100	0.0	0.0	2025-01-31 09:40:31	0.014705044172128800	0.014597973387547200	174.19	175 258	0.006892722940410280	-0.028417889221923800	-0.03245664237383