

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](https://github.com/shubhamdarekar/DSCI560---Shubham/blob/main/DSCI560---Shubham/Lab_3/README.md)

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 \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.

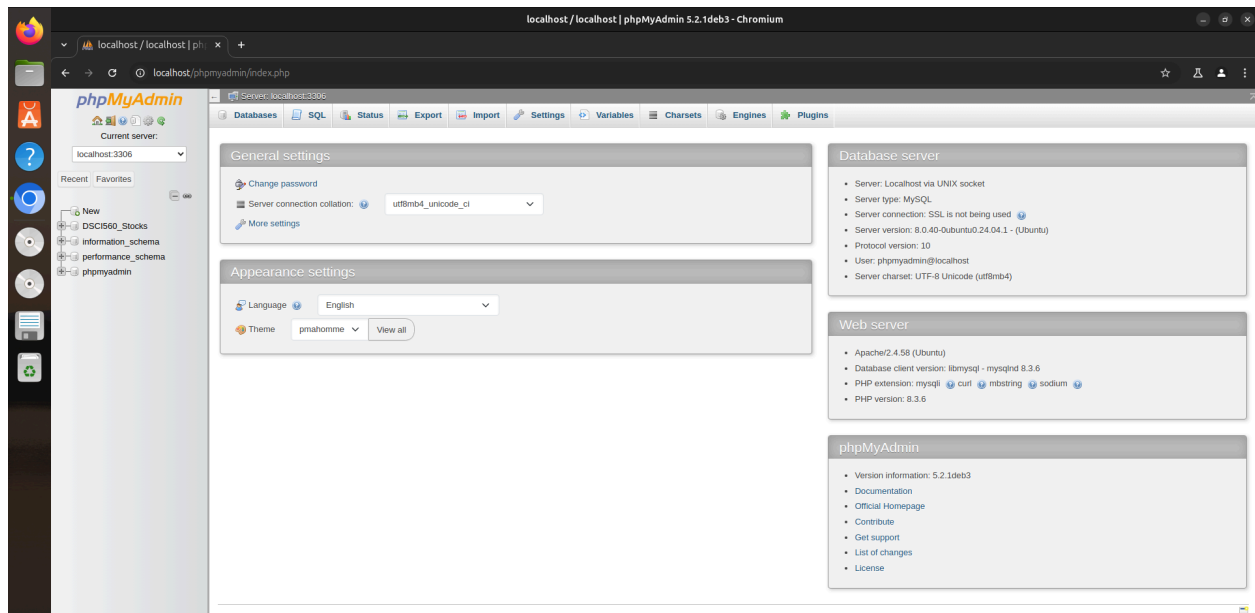
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

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 |
| sys |
+-----+
6 rows in set (0.00 sec)

mysql> █
```

phpMyAdmin:



Task 2: Technologies Used

Database Storage: MySQL

Database Interface: phpMyAdmin

Scraping: Python and yFinancepytho

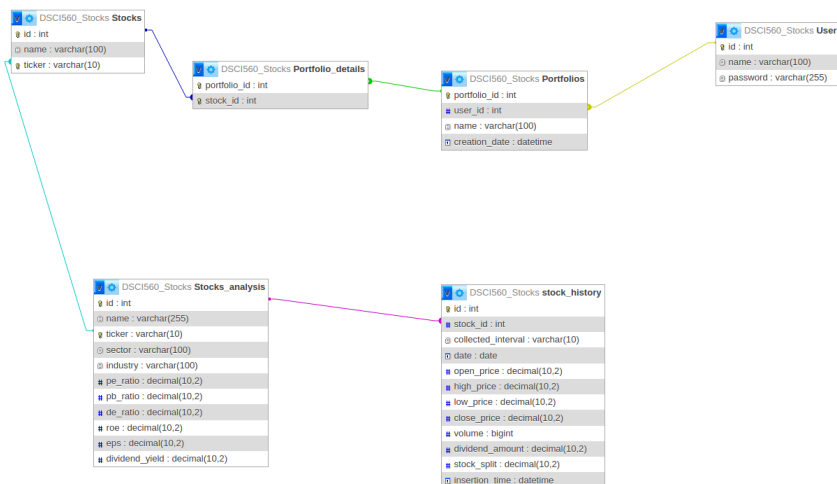
Environment:

```
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              2.2.2
pandas             2.2.3
peewee             3.17.8
pip                24.0
platformdirs       4.3.6
python-dateutil    2.9.0.post0
pytz               2024.2
requests           2.32.3
six                1.17.0
soupsieve          2.6
tzdata             2025.1
urllib3            2.3.0
webencodings       0.5.1
yfinance           0.2.52
```

Task 3: Data Collection / Storage

Database Schema:

```
1  Users:
2      Id
3      Name
4      Password
5
6  Portfolios:
7      portfolio_id
8      user_id
9      name
10     creation_date
11
12 Portfolio_details:
13     portfolio_id
14     stock_id
15
16 Stocks_analysis:
17     id
18     name
19     ticker
20     sector
21     industry
22     pe_ratio
23     pb_ratio
24     de_ratio
25     roe
26     eps
27     dividend_yield
28
29
30 stock_history:
31     stock_id,
32     date,
33     open_price,
34     high_price,
35     low_price,
36     close_price,
37     volume,
38     dividend_amount,
39     stock_split
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
3. 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
3. 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
1. Pfizer, Inc.
2. 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/quote
ail&corsDomain=finance.yahoo.com&formatted=false&symbol=ABEK&crumb=dCeQik02A10
Ticker not recognized
```

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

```
Stocks in Portfolio Big Pharma
1. Pfizer, Inc.
2. Johnson & Johnson
3. AbbVie Inc.

Operations
1. Add Stock
2. 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
      Open      High      Low      Close      Volume      Dividends      Stock Splits
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
      Open      High      Low      Close      Volume      Dividends      Stock Splits
Date
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      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
      Open      High      Low      Close      Volume      Dividends      Stock Splits
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

processed_stock_history														
id	stock_id	collected_interval	date	open_price	high_price	low_price	close_price	volume	dividend_amount	stock_split	insertion_time	daily_return	log_return	moving_avg_5d
1	1	1d	2024-01-31	186.13	186.19	183.45	183.50	55467800	0.0	0.0	2025-01-31 09:40:31			
2	1	1d	2024-02-01	183.09	186.04	182.92	185.95	64885400	0.0	0.0	2025-01-31 09:40:31	0.01303119376454100	0.01303119376454100	
3	1	1d	2024-02-02	178.98	186.42	178.38	184.94	102518000	0.0	0.0	2025-01-31 09:40:31	-0.054146372021598300	-0.054146372021598300	
4	1	1d	2024-02-05	187.23	188.33	184.93	186.77	69668800	0.0	0.0	2025-01-31 09:40:31	0.00895101113874200	0.00895101113874200	
5	1	1d	2024-02-06	185.95	188.39	185.86	188.38	43490800	0.0	0.0	2025-01-31 09:40:31	0.00850220386099155270	0.00850220386099155270	185.908
6	1	1d	2024-02-07	189.71	190.12	187.69	188.49	53439000	0.0	0.0	2025-01-31 09:40:31	0.005803296805454460	0.005803296805454460	186.908
7	1	1d	2024-02-08	188.47	188.62	186.44	187.40	40962000	0.0	0.0	2025-01-31 09:40:31	-0.005795958278643070	-0.005795958278643070	187.196
8	1	1d	2024-02-09	187.97	189.31	187.32	188.17	45155200	0.24	0.0	2025-01-31 09:40:31	0.0041088505763073200	0.0041088505763073200	187.842
9	1	1d	2024-02-12	187.74	187.99	186.12	186.48	41781900	0.0	0.0	2025-01-31 09:40:31	-0.00981240367702610	-0.00981240367702610	187.84000000000000
10	1	1d	2024-02-13	185.10	185.54	182.85	184.37	56529500	0.0	0.0	2025-01-31 09:40:31	-0.011314886314886300	-0.011314886314886300	186.96200000000000
11	1	1d	2024-02-14	184.65	184.86	181.78	183.49	54630500	0.0	0.0	2025-01-31 09:40:31	-0.0074730179303143000	-0.007473017930314300	185.96200000000000
12	1	1d	2024-02-15	182.89	183.82	180.70	183.20	65434500	0.0	0.0	2025-01-31 09:40:31	-0.00407368235618000	-0.00407368235618000	185.142
13	1	1d	2024-02-16	182.76	184.18	181.02	181.65	49701400	0.0	0.0	2025-01-31 09:40:31	-0.0040846698572818860	-0.0040846698572818860	183.838
14	1	1d	2024-02-20	181.13	181.77	179.35	180.91	53665600	0.0	0.0	2025-01-31 09:40:31	-0.00407368235618000	-0.00407368235618000	182.74000000000000
15	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	0.004145707810513600	182.182
16	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	0.011284817791478600	182.28600000000000
17	1	1d	2024-02-23	184.34	184.37	181.57	181.86	45119700	0.0	0.0	2025-01-31 09:40:31	-0.010070219367478100	-0.010070219367478100	181.958
18	1	1d	2024-02-26	181.58	182.1	180.01	180.51	40867400	0.0	0.0	2025-01-31 09:40:31	-0.00745098239739670	-0.00745098239739670	181.73000000000000
19	1	1d	2024-02-27	180.45	183.26	178.91	181.97	54318900	0.0	0.0	2025-01-31 09:40:31	0.0080881455085832	0.0080881455085832	181.94200000000000
20	1	1d	2024-02-28	180.62	181.91	178.88	180.1	136682600	0.0	0.0	2025-01-31 09:40:31	-0.006944939787457000	-0.006944939787457000	184.209
21	1	1d	2024-02-29	180.62	181.91	178.88	180.1	136682600	0.0	0.0	2025-01-31 09:40:31	-0.003706367208948170	-0.003706367208948170	184.20900000000000
22	1	1d	2024-03-01	178.90	179.88	176.74	179.01	73488000	0.0	0.0	2025-01-31 09:40:31	-0.00603192509959500	-0.00603192509959500	183.892
23	1	1d	2024-03-04	175.51	176.26	173.16	174.47	81510100	0.0	0.0	2025-01-31 09:40:31	-0.020381711638272400	-0.020381711638272400	179.26400000000000
24	1	1d	2024-03-05	170.14	171.42	169.01	169.51	95132400	0.0	0.0	2025-01-31 09:40:31	-0.028489849614115500	-0.028489849614115500	176.772
25	1	1d	2024-03-06	170.44	170.62	168.07	169.51	89587700	0.0	0.0	2025-01-31 09:40:31	-0.005895936797090050	-0.005895936797090050	174.32
26	1	1d	2024-03-07	168.54	170.11	167.88	169.38	71765100	0.0	0.0	2025-01-31 09:40:31	-0.007012123969632550	-0.007012123969632550	171.878
27	1	1d	2024-03-08	168.39	173.07	168.33	170.11	76114600	0.0	0.0	2025-01-31 09:40:31	0.010214382736168000	0.010214382736168000	170.198
28	1	1d	2024-03-11	172.32	173.75	171.43	172.13	60139500	0.0	0.0	2025-01-31 09:40:31	0.011894718661922800	0.011894718661922800	169.73
29	1	1d	2024-03-12	172.53	173.4	170.39	172.61	59825400	0.0	0.0	2025-01-31 09:40:31	0.002788500171171780	0.002788500171171780	170.35
30	1	1d	2024-03-13	172.15	172.57	170.14	170.51	52488700	0.0	0.0	2025-01-31 09:40:31	-0.012166154917076100	-0.012166154917076100	170.75
31	1	1d	2024-03-14	172.29	173.68	171.43	172.38	72913500	0.0	0.0	2025-01-31 09:40:31	0.010967088702888300	0.010967088702888300	171.548
32	1	1d	2024-03-15	170.55	172.0	168.68	172.0	121664700	0.0	0.0	2025-01-31 09:40:31	-0.002208854058001600	-0.002208854058001600	171.928
33	1	1d	2024-03-18	174.94	177.07	172.89	173.09	79604200	0.0	0.0	2025-01-31 09:40:31	0.006337200302325510	0.006337200302325510	172.118
34	1	1d	2024-03-19	173.71	175.87	172.41	175.45	55215200	0.0	0.0	2025-01-31 09:40:31	0.0138452391414700	0.0138452391414700	172.68600000000000
35	1	1d	2024-03-20	175.09	178.03	174.48	178.03	53423100	0.0	0.0	2025-01-31 09:40:31	0.014705044172128800	0.014705044172128800	174.19