PROJECT FOR TIME SERIES ANALYSIS COURSE

GOAL: Build models ARIMA and LSTM to predict the price of two following stocks:

HPG: Hoa Phat Group

MWG: Mobile World Investment Corporation

These companies are from different sectors (manufacturing, and retail electronics) and listed on the Ho Chi Minh Stock Exchange (HOSE).

Upload the report with detailed results and code sources into the following folder: https://drive.google.com/drive/folders/1TCdW1Oi3qjYUd-gNmOXUifqBYtaLzS1R?usp=sharing

- Rules for file names: **report_Doan_Hoai_Nhi_BI12-338.doc** for report file, **code_folder_Doan_Hoai_Nhi_BI12-338** for code folder

Deadline: **Before 23h June 02th 2024.**

**Anything uploaded on driver after this time will not be accepted.**

1. **DATA COLLECTION**

Source: Investing.com

Period: **January 2, 2018, to March 31, 2023**

Features: trading date, closing price, opening price, highest price, lowest price, trading volume, and percentage change compared to the previous day.

How to get historical data: Go to investing.com, find a specific stock (e.g. HPG), click on Historical Data button, choose the right period and download data.
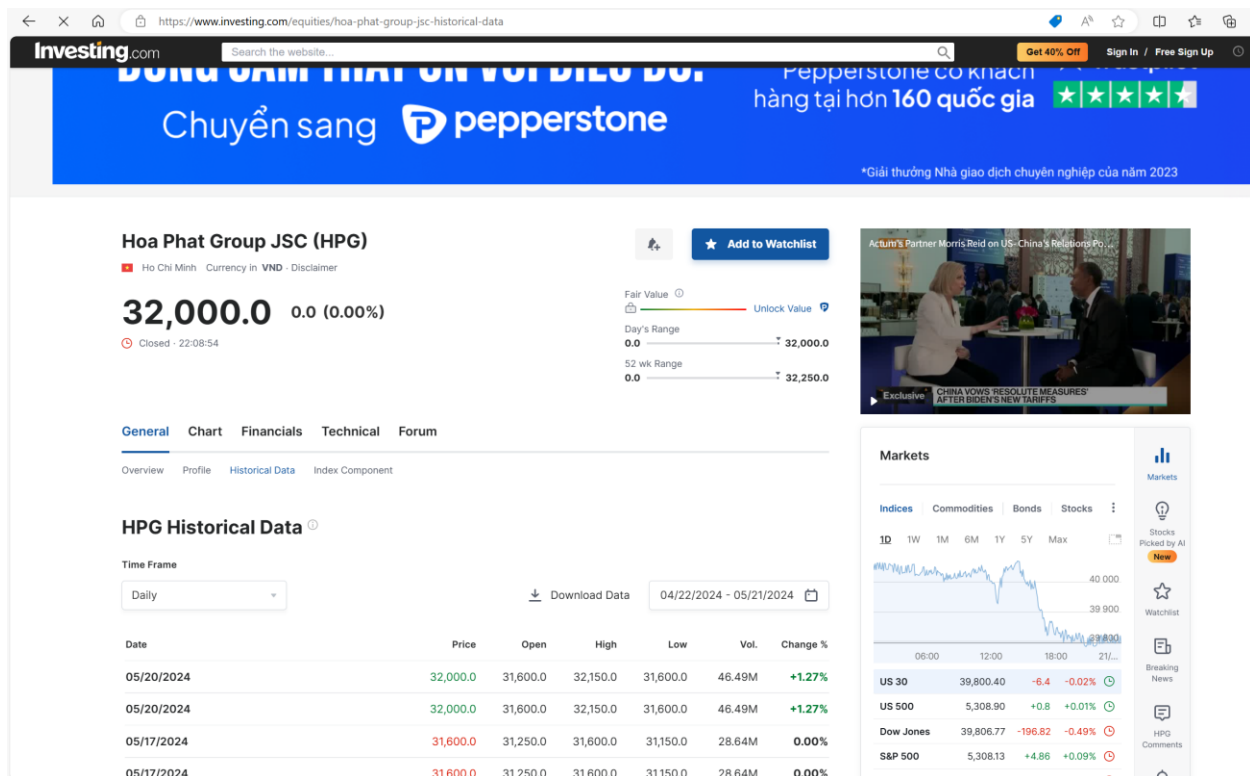
2. REQUIREMENT

- Do some basic descriptive statistics on these time series, plot these curves and give some remarks on the price changes of HPG, MWG during the selected periods.
- Train your models and from January 2nd 2018 to February 28th 2022 and predict what the stock price between March 1st 2022 and March 31st 2023.
- Two models need to be used: ARIMA, LSTM
  - Use some Hyper-parameter tuning process to get the best parameters for each model.
  - Metric results for each model: R square, MAPE, RMSE.
  - Compare the performance between these two models.
  - For each model, need to do certain data processing. With ARIMA, use first-order differencing to make the series stationary. With LSTM, do min-max scaling and standardize the data.

3. GUIDELINE ON MODELS

3.1. For LSTM implementation: see the lecture file in moodle
https://moodle.usth.edu.vn/mod/forum/discuss.php?d=1297

3.2. For ARIMA:

Stationarity Test: Use Dickey-Fuller test.

First-order differencing applied to make the series stationary.

Model Selection: Determine p, d, q values to get best model.

Proceed with forecasts.

4. **REMARK**
- 1 bonus point if you can apply model GRU.
- 2 bonus points if you can apply more the advanced technology for forecasting: Large Language Models for Time Series Forecasting.