# Dataset

## Pre-requisites

* The timeline will be 20 financial years, i.e., from 1st April 2004 to 31st March 2024.
* Per Day Data

## Features

* Open
* Close
* High
* Low
* Volume

## Source

* Data will be scraped from Yahoo Finance using the yfinance python package

## Details

Data will be from 3 economies.

1. USA – S&P 500 index (Data A)
2. UK – FTSE100 index (Data B)
3. India – Nifty50 index (Data C)

## Split

* Train – First 10 years
* Validation – Next 5 Years
* Test – Last 5 Years

# Assumptions

* There will be no shorting. First, “Buy” and then “Sell.”
* We will only look at short-term trading strategies, i.e. within two calendar weeks
* The portfolio starts with an initial 1000 dollar/pound/ruppee investment which is then reinvested for every subsequent trades
* The number of shares are equivalent to the traded amount.
* Brokerage, tax and other expenses are not considered

# Models

## Baseline

1. Moving Averages
2. Highest Profit Scenario
3. Highest Loss Scenario

## DQN

1. RL Model A – DQN on S&P500 Data
2. RL Model B – DQN on FTSE100 Data
3. RL Model C – DQN on NIFTY50 Data

## Output Features

1. Buy
2. Hold
3. Sell

## Evaluation

* Performance of Model A on test Data A
* Performance of Model B on test Data B
* Performance of Model C on test Data C
* Performance of Model A on Data B and C
* Performance of Model B on Data C and A
* Performance of Model C on Data A and B
* Performance of DQN during Covid, 2008 Financial Crisis, etc.
* Performance of the models during country-specific Economic shifts, i.e. Elections, government changes, Brexit, etc.

### Metrics

1. Confusion Matrix
2. F1 Score
3. Sensitivity
4. Recall
5. Accuracy

# Research Questions

1. Is DQN a valid model to use in this scenario? How does the DQN compare against the traditional and highest/lowest profit models?
2. How does a DQN model trained in one market work in the other markets?
3. How does the DQN model fare in different economic conditions?