Programming for Data Science

Assignment 3

Group members:

# Introduction

## 1.1 Getting the data

For the purpose of this assignment a group of 20 of datasets pertinent stocks of companies listed in New York Stock Exchange have been selected. These datasets will be combined and will be transformed into one timeseries dataset. main purpose of the newly created dataset is to perform financial analysis on market and stock data that is provided by NYSE.

## 2.0 purpose of dataset

Following names some of the operations that includes in analysis of this dataset:

* Calculating risk and Sharpe ratio
* Calculating return based on prices
* Portfolio allocation
* Price prediction
* Stock comparison
* Technical indicators computation and analysis
* Price chart visualization
* Visualizing technical indicators such as moving averages
* Forecasting future returns using past data with ARIMA model.

Modeling tools like Integrated Autoregressive moving averages (ARIMA) enables the researcher to make prediction, simulate stochastic processes such as prices series and make particular hypothesis.

Furthermore, NYSE stock dataset enables us to predict future prices using machine learning models. To perform such a prediction usually regression is used.

## 2.1 Dataset features

This chapter explains the structure and features of datasets that are being used in this assignment.

NYSE stock dataset hold is a **time series** dataset that holds 5 years of stock price data in NYSE market from 30/11/2012 to 29/11/2017.

Dataset name: NYSE\_Stocks, Type: Time-Series

This dataset is a combination of 20 datasets. Each has 18 columns and 1260 rows (18 x 1260)

|  |  |  |
| --- | --- | --- |
| Attribute | Type | Explanation |
| Date | String | Date in dd/mm/yyy format |
| Financial institution | String | Institution symbol |
| Location | String | Location (city) of institution |
| High | Numeric | Highest price in a day |
| Low | Numeric | Lower price in the day |
| Open | Numeric | The price of the share in the beginning of the trading day |
| Close | Numeric | The price of the share in the closing time of day |
| Volume | Numeric | Number of shares traded in a trading day |
| Aggregated Data 2 Days | Numeric | Mean of open/close/high/low of two trading days |
| Aggregated Data 3 Days | Numeric | Mean of open/close/high/low of two trading days |
| Aggregated Data for 5 Days | Numeric | Mean of open/close/high/low of two trading days |
| Number of employees | Numeric | Number of employees of an institution |
| Net changes 0 (numeric) | Numeric | Net price change of current day |
| Net change 0 (nominal) | Nominal | Determine whether net change is positive or negative |
| Net change 5 (numeric) | Numeric | Net price change of current day of past 5 days |
| Net change 4 (nominal) | Nominal | Determine whether net change is positive or negative |
| Net change 25 (numeric) | Numeric | Net price change of current day of past 25 days |
| Net change 25 (nominal) | Nominal | Determine whether net change is positive or negative |