**Imputing missing data from stock time series**

Members: Khanh Nguyen, Yizhen Zhao, Evgeniya Lagoda, Himanshu Raj, Carlos Owusu-Ansah, Sergei Neznanov

1. **Overview**:

Motivation: The NYSE and NASDAQ average about 252 trading days yearly. What if someone accidentally deleted data from five (or more) trading days? Missing data is often ignored or removed when we analyze data, which could be problematic. One of the problems is that stock returns depend on that "missingness" (S. Bryzgalova, 2022).

Stakeholders: AAPL Inc., investors, financial companies

KPIs:

1. Generate missing data by removing five points from a complete time series data. There are 2 ways to remove these:
   * Case 1: remove five random rows from the data frame.
   * Case 2: remove five consecutive rows randomly from the data frame.
2. Create models to “guess”/ to impute the missing data
   * Baseline model: Linear interpolating
   * Model 1:
   * Model 2:
   * Model 3:
3. Compare the models by computing AIC/ BIC or RMSE.
4. **Data Collection**:

AAPL (Apple Inc.) data from January 1, 2023 until December 31, 2023. There are 250 data points because trading market closes on Saturdays, Sundays, and national holidays. The data includes the following parameters (yahoo finance):

1. Date: starting from January 3, 2023 (Tuesday) to December 29, 2023 (Friday). The dates do not include Saturdays, Sundays, and US national holidays.
2. Open: price when the market opens. The opening time is 9:30 am Eastern Time.
3. High/ Low: highest/ lowest prices on that day.
4. Close: price when the market closes. The closing time is 4:00 pm Eastern Time.
5. Adj Close (Adjusted close): the closing price after adjustments for all applicable splits and dividend distributions.
6. Volume: number of shares that have been traded on a given day.

A daily return is computed from the adjusted close as:

where is the daily return on day , and is the adjusted close on day .