## Millennium Data Project

**This is a data project to help us understand how you handle and summarize data, choose prediction model, build and train the model, test the accuracy as well as visualize the result.**

**You are provided 2 years of the daily prices of stock XYZ with the following specification.**

| POSITION | LENGTH | DATA TYPE | DESCRIPTION | REMARKS |
|---|---|---|---|---|
| **1** | 8 | Number | Data Date | yyyymmdd |
| **9** | 15 | String | Ticker | XYZ |
| **24** | 10 | Number | Open | 9999999999 * |
| **34** | 10 | Number | High | 9999999999 * |
| **44** | 10 | Number | Low | 9999999999 * |
| **54** | 10 | Number | Close | 9999999999 * |
| **64** | 10 | Number | Adj Close | 9999999999 * |
| **74** | 10 | Number | Volume | 9999999999 * |
| **84** | 3 | Number | Magnitude:<br>***a positive or negative power of ten by which Open/High/Low/Close/Adj Close must be multiplied to obtain the proper value*** | -5 * |

* ***-999 = not available***

**You can pick your desired language and libraries to properly load the daily price data without errors and summarize the data provided by calculating relevant analytics of the stock performance in tables, summary statistics or charts, etc.**

*Questions:*

1. *List out what kinds of rules you put into the data handling and why?*
2. *What are exceptions?*
3. *How to handle exceptions?*

**You are required to predict the prices for the last 20% of the days using the properly loaded daily price data.**

**You can pick your desired language and libraries to develop THREE approaches with the data, test the accuracy and repeat the tests with different parameters if necessary to improve the result.**

3 approaches can be picked, Moving Average Method, Regression Model and one you choose yourself.

***Moving Average***

*We set the current adjusted closing price as the mean of the adjusted closing price of the previous N days.*

***Linear Regression***

*Linear regression is a linear approach to modeling the relationship between a dependent variable and one or more independent variables. We fit a linear regression model to the previous N values, and use the model to predict the value for the test days..*

**You are required to visualize the price prediction process.**

*Questions:*

4. *Compare the 3 results and comment.*