11 – STOCK DATA MINING

Data mining is the process of sorting through large amounts of data and picking out relevant information. In this project you will do an analysis of Apple stock prices. Your program will calculate the monthly average prices of Apple stock from 2010 to 2018 and say the best 6 months and the worst 6 months. The CSV file is as follows:

```
Date,Open,High,Low,Close,Adj Close,Volume
2010-10-11,42.105713,42.462856,42.085712,42.194286,28.354528,106938300
2010-10-12,42.201427,42.785713,41.784286,42.648571,28.659813,139636000
2010-10-13,42.885715,43.137142,42.828571,42.877144,28.813412,157523100
2010-10-14,43.098572,43.209999,42.914288,43.187141,29.021738,108824100
```

How to calculate the average price?

Total sales for a day = Adj Close * Volume

The average price for two days=

$$(\text{Adj Close}_1 * \text{Volume}_1 + \text{Adj Close}_2 * \text{Volume}_2) / (\text{Volume}_1 + \text{Volume}_2)$$

To average a whole month, you just add up the total sales (V*C) for each day and divide by the sum of all the volumes ($V_1 + V_2 + \ldots + V_n$).

How can I get the year and month?

The split method will allow you to separate a string by a character. As an example:

```
>>> "2010-10-11".split("-")
['2010', '10', '11']
```

This array can then be indexed using 0 for the year, 1 for the month and 2 for the day.

Decomposing the problem

1. Read the CSV file and loop through each row
2. Split the date into separate parts
3. Check whether you have stored that year and month before, if not create a storage container for it with a 3 index list e.g. [0,0,0]
4. Calculate **Adj Close * Volume** and then add it to **index [0]**
5. Add **Volume** to the previous values stored in **index [1]**
6. Calculate **index [0] / index[1]** and store the result in **index[2]**
7. Find a way to sort the lists and show the top 6 and bottom 6 months.