

# Outliers Problem

## Story:

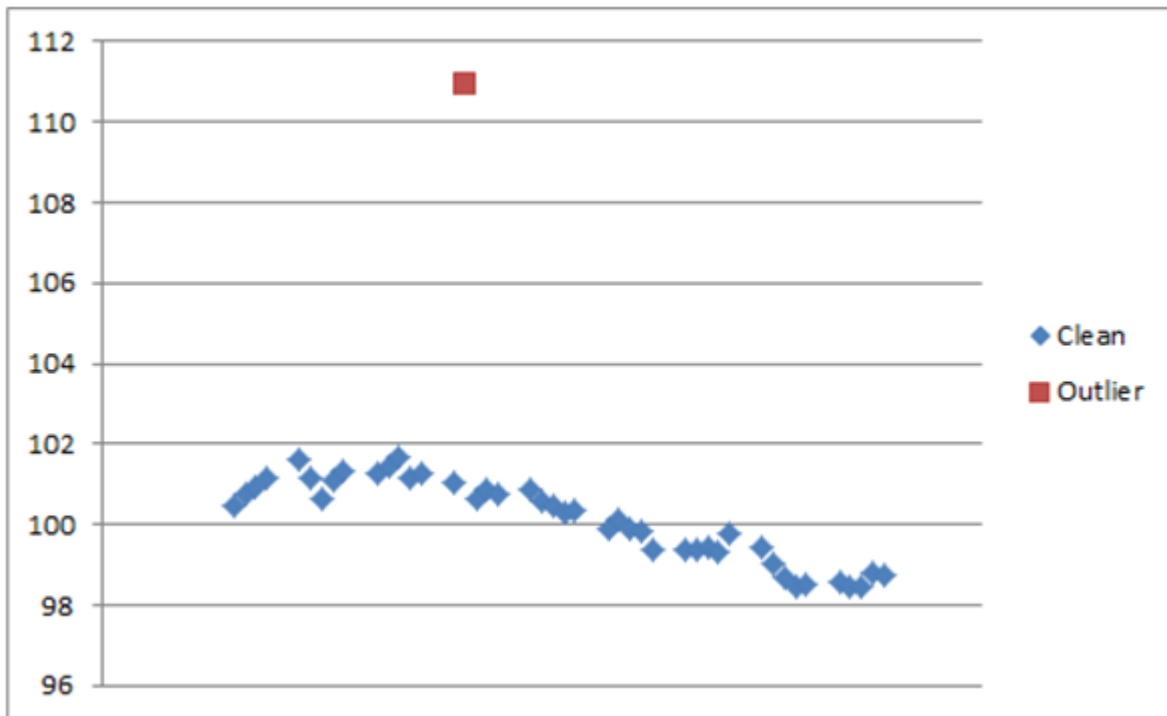
There is an ongoing campaign running on our website. The idea of this campaign is that when a customer buys a specific product he/she gets a discount voucher as a reward.

We buy these vouchers in real-time upon demand (i.e. when we need to perform the fulfillment to the customer) from the third-party vendor automatically. The price for this voucher is fixed during the day but changes at midnight.

Previously we noticed that sometimes the price can jump for one day and then return to the average value on the next day. These days are outliers and on these days we prefer to postpone the fulfillment for one day.

### Technical Description:

Write an application to detect data points which are outliers in the attached CSV file. Please implement the solution in Java.



For this problem, an “Outlier” is defined as a data point with a value that is significantly different from the points immediately around it.

The primary purpose of this exercise is for you to demonstrate your coding style – the way you structure your code and the approach you take. We are interested in your ability to write flexible / maintainable code e.g. interfaces and informative comments.

## Requirements:

- The application should read the CSV file in, inform the user of the outliers found, and then write a clean version of the data with the outlier points removed.
- When analyzing each point the outlier algorithm should only use previous (historical) data points, **as if the data were being received chronologically**.
- **The outlier detection algorithm should not be complex.** It is fine for you to use a basic detection algo of your own design – you don't have to apply a “standard” outlier test as found online (and such tests may not work well for this dataset).
- **You only need to implement a single algorithm against the CSV file data source.** However, you should consider how your code could be extended to support additional functionality. In this case that might mean allowing for alternative outlier detection algorithms or other data sources / outputs such as a database or remote process.