Rocco Vulpis

December 14, 2023

Salting Data using Java and Plotting in Excel

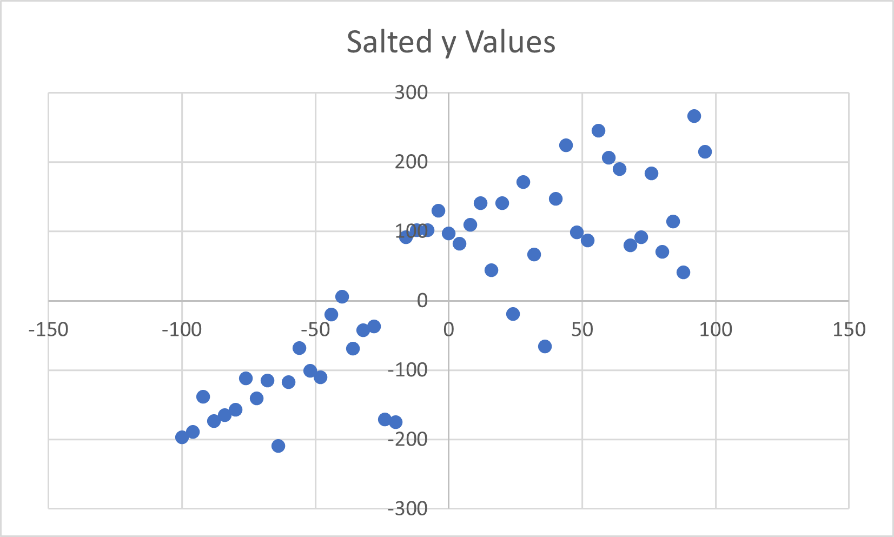
In the salting portion of this project, a mathematical function's y-values were randomly altered using a process known as 'salting'. This involved modifying the data points read from a CSV file, 'data.CSV', and then exporting the adjusted data to a new CSV file, 'saltedData.CSV'. The graph created in Microsoft Excel with this data displays the effects of salting, while still maintaining a resemblance to the original function with noticeable random variations.

The first step involved importing the function's values using the readFile method. This process included the BufferedReader class for line-by-line reading of 'data.CSV', and the parsing of each line's values into doubles. These values were stored in two ArrayLists: one for x-values and one for y-values.

After importing and storing the data, the saltValues method was then called to adjust the y-values. This method sets the range of random adjustments from 0 to 149. Each y-value was randomly increased or decreased within this range, and the adjusted value replaced its original value in the yValues ArrayList.

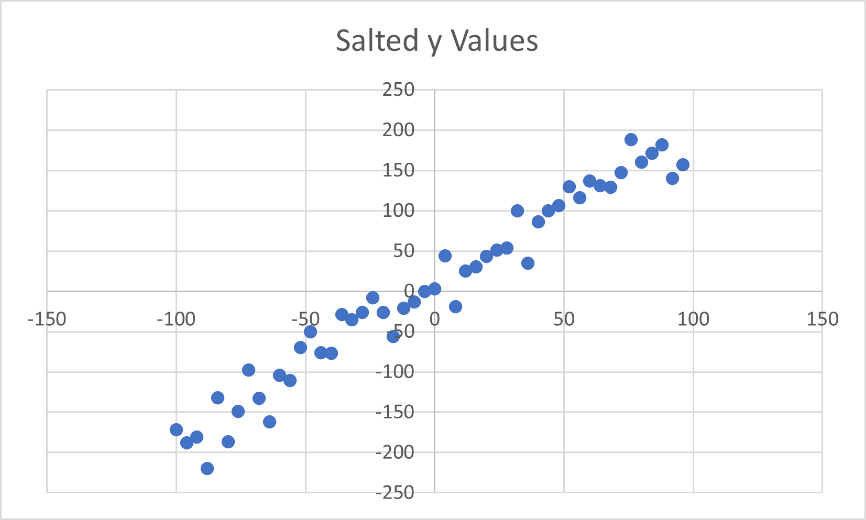
The final step involved exporting the data using the writeToFile method. This method wrote both x and y values to 'saltedData.CSV' using the BufferedWriter class. The generated CSV file was stored into the default workspace directory. In Microsoft Excel, the x and y values were stored into columns A and B of the spreadsheet.

Opening the CSV in Excel and using its charting tools, a scatter plot graph is generated. This graph displays a pattern of points that, while scattered, still hint at the original function's form. The graph's appearance changes based on the 'salting' parameters, showing how the random adjustments to the y-values impact the overall data representation. Below is a screenshot displaying the Excel graph with the salted data points.



The result of y = mx + b after the y values were randomly adjusted between 0 and 150.

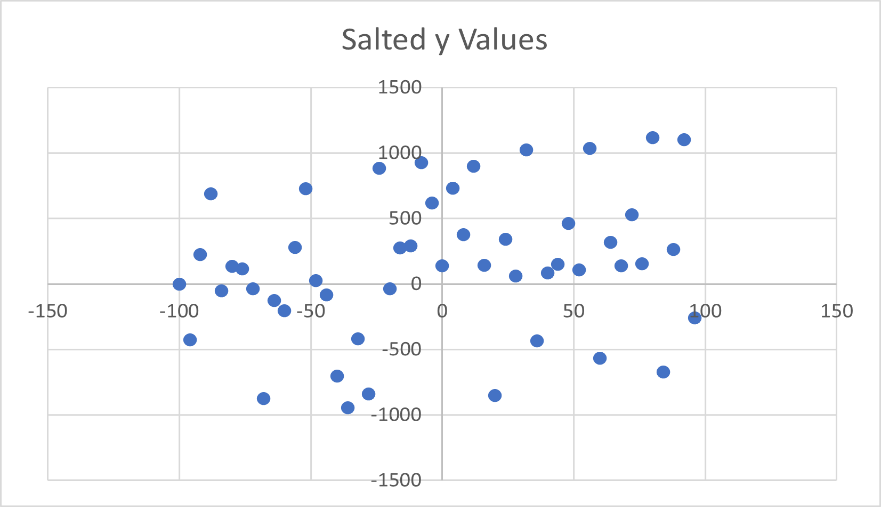
Here are additional examples of graphs created with different ranges of random adjustment values.



Result when adjustment range is between 0 and 50.

A graph with blue dots

Description automatically generated

Result when adjustment range is between 0 and 450.

Result when adjustment range is between 0 and 1000.