**Lesson 2 – Little Data to Big Data**

**\*\*Instructions:** Please change the text color of your responses to red text. Please organize the endings to each page.

**ACTIVITY 3.2.3 Pirates are the Problem**

**GOALS:**

* Examine the efficiency of an algorithm.
* Separate correlation from causality.
* Analyze an algorithm.

You will be using Google Sheets for this assignment. You will use *four different algorithms* to analyze data. The code is provided for you. Complete the following:

Sometimes sorting the data in different ways will allow different patterns to emerge.

The data sheet contains just data about collisions in Manhattan that involved a cyclist injury or fatality. Use this data to determine the most common cause of these accidents.

What is the last row number that has an entry?

|  |
| --- |
| 7,260 |

In what row number did you enter the formula?

|  |
| --- |
| 7,261 |

According to your data, what was the most common cause of collision?

|  |
| --- |
| Driver Inattention/Distraction |

The insight that you gained from processing the data in this spreadsheet is information. Information is the collection of facts and patterns extracted from data. Use the same techniques to gain information about the most dangerous cross streets for cyclists in Manhattan.

|  |
| --- |
| The most dangerous cross streets for cyclists are 2 avenue, 1 avenue, and Broadway. |

The page linked below has several sort algorithms along with animations illustrating their execution. You will all examine insertion sort and merge sort; your teacher may have you examine additional algorithms if time permits. Use the code, VS Code, and the animations to help you answer the following questions.

How many times was a comparison made?

|  |
| --- |
| 46 comparisons were made in insertion sort, while 44 comparisons were made in merge sort. |

How many times does the algorithm compare two numbers now?

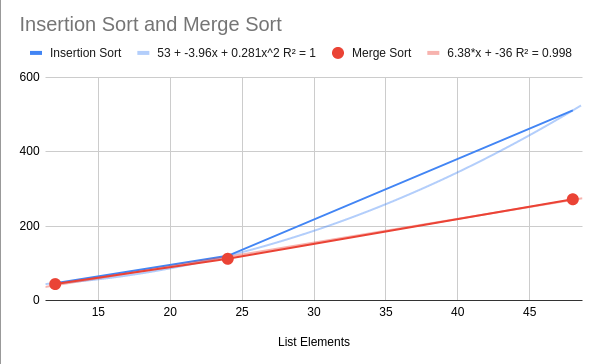
|  |
| --- |
| Insertion sort takes 120 comparisons after the list size is doubled. |

How many times does the algorithm compare two numbers now?

|  |
| --- |
| Insertion sort takes 511 comparisons after the list size is quadrupled. |

Use a spreadsheet to create a line graph and predict how many comparisons would be made by the algorithm if the dataset contained 10,000 items.

|  |
| --- |
| Insertion sort would take around 28 million comparisons. |

Repeat the above step for the merge sort algorithm. How does the graph of insertion sort compare to the graph merge sort?

|  |
| --- |
| Merge sort’s graph seems to be mostly linear, while  insertion sort’s graph seems to be polynomial. |

Is it true that efficiency is the most important factor when writing code?

|  |
| --- |
| No, as sometimes, writing fast code is not the main goal, as some applications require writing code quickly that may not run the fastest, and others may require code to be more maintainable, which may not always be the most fast approach. |