HW 2 Guidelines

Input Format for this HW

- For c and c++
  ./a.out input.txt
- Java
  java classfile input.txt
- Python
  python filename.py input.txt

Note : I have mentioned .py file everywhere for easiness. You should create everything as per your prefered language.

Q - 1 Implement the two versions of MergeSort that we discussed in class. Create a table or a plot for the total number of comparisons to sort the data (using data set here) for both cases. Explain.

- Create two file **q1a.py** and **q1b.py** (Please mention in comment which kind of merge sort it is)
- create text file **out.txt** in same directory which contains output in following format
  <No. of Comparison>
  <Sorted array>

  Example Output of out.txt
  345
  1
  2
  3
  4

- Provide analysis

Note : If output is not sorted then you will lose entire marks

Q - 2 Implement Shell sort which reverts to insertion sort. (Use the increment sequence 7, 3, 1). Similar to Q1, create a table or a plot for the total number of comparisons made in the sorting the data for both cases (insertion sort phase and shell sort phase). Explain why Shell short is more effective than Insertion sort in this case.

- Create q2.py
- create text file **out.txt** in same directory which contains output in following format

  <No of Comparison>
  <Sorted Array>

Example Output of out.txt

234
1
2
3
4

- Provide analysis

Note : If output is not sorted then you will lose entire marks

Q - 3 The Kendall Tau distance is a variant of the "number of inversions" we discussed in class. It is defined as the number of pairs that are in different order in two permutations. Write an efficient program that computes the Kendall Tau distance in less than quadratic time on average. Plot your results and discuss. Use the dataset provided here. Note: data0.* for convenience is an ordered set of numbers (in powers of two). data1.* are shuffled data sets of sizes (as given by "*").

- Create q3.py file
- create text file **out.txt** in same directory which contains output in following format <Number of Inversion>

Example Output of out.txt

324

- Provide analysis

Q - 4 Create a data set of 8192 entries which has in the following order: 1024 repeats of 1, 2048 repeats of 11, 4096 repeats of 111 and 1024 repeats of 1111. Write a sort algorithm that you think will sort this set "most" effectively. Explain why you think so.

- Create q4a.py which contains the program for creating text file as per the question and create q4.txt
- Create q4b.py which contains the program for sorting algorithm
- create text file **out.txt** in same directory which contains output in following format

Example Output of out.txt

1
2
3
4

- Provide analysis

Directory Structure

<netid>

-> q1a.py
-> q1b.py
-> q2.py
-> q3.py
-> q4a.py (to create text file)
-> q4.txt
-> q4b.py (algorithm file)

Note -

1. If code will not compile then you will lose 50% grades
2. If code will not produce correct answer then you will lose entire marks
3. If I don't find analysis of any question then you will lose 50% grades
4. If you can't able to produce input / output in provided manner then your HW will not be graded.