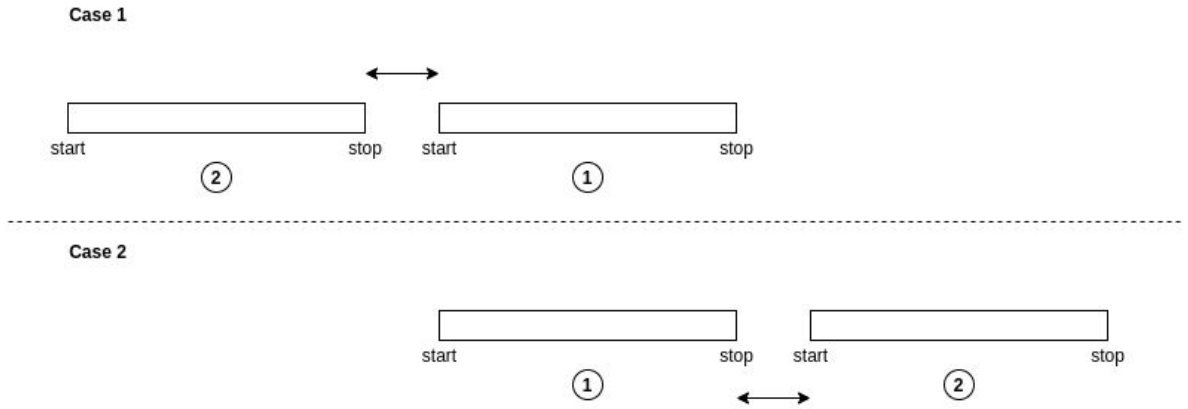
Explanation of the solution for problem 2) a

The conditions that I looked was whether they have mappings in both files on the same chromosome and the mappings are no longer than 1000 base pairs apart. So for the second criteria, I considered the start of one region is within 1000bp or less from the end of another region. So according to that all the overlapping regions and regions with a gap (between one region's end and other region's start) <=1000 will be considered as valid output material.



At first, I came up with a solution containing nested loops. For this problem, when the file is very large since the time complexity was $O(n^2)$ it will take a long time to run and the efficiency will be low. In order to avoid that I came up with a solution that uses binary search using numpy in which the overall complexity will be $O(n(\log n)^2)$ which is well below the previous complexity. In order to use binary search, I had to sort the data frames by start and stop values of the region. Then it will provide the closest 'stop' value from file2 for a given 'start' value from file1. But still, it took a considerable amount of time to complete (~25 mins).