

# Scalable Data Mining: Assignment 1

## Pankaj Mishra, 17EC35034

### Question 3

This question first needs us to load the CSV data in an RDD which is done similar to the first question. The second and third parts are solved using the concept of joining the RDDs using common keys. The records in the 2 RDDs are compared on the basis of the keys and combined into single element whenever they have common keys. The final part is solved by applying a filter on the combined RDD record's value.