Question 3

1. **Read in the file to an RDD and name it as ‘assignment_2’ and count the number of records.**
   Answer:- 1435
   Method:- We first create an RDD that is a collection of *case classes* that represent a line in the record file and count the number of items in it.

2. **How many records in the log file (used in the last 2 questions) refer to entries in the ‘assignment_2’ file ?**
   Answer:- 62839
   Method:- We now have 2 RDDs, and we remap them as (key, value), where key corresponds to repository name and the value is the corresponding entry. Then we join the 2 remapped RDDs into a single RDD by joining the elements with a common key (repository name). Now count the number of items in the resultant RDD.

3. **Which of the ‘assignment_2’ repositories has the most failed API calls.**
   Answer:- asmagin/sitecore-foundation-codegeneration-composition (5)
   Method:- Use the joined RDD computed above, and filter the entries whose requests have failed (from *rest*). Now, use MapReduce with repository names as key and count the number of entries. And then find the repository with maximum value