**Understanding of the problem**

This problem requires the use of RDD to manipulate data to understand the workings of RDD, and combining different RDD through intersections and subtracts.

**Approach**

1. Read the file as textFile, remove the header.
2. Split by comma separator.
3. Use operations like filter and map on the RDD to convert to different data.
4. Create tuples, and then reduceByKey to create key-value pairs.
5. Use specialized operations like mapValues and flatMapValues to work with tuple values directly.
6. take() is used instead of collect() to keep in mind the size of the dataset for scalability purposes.

**Limitations**

1. Working with RDD textFile and csv is cumbersome, since the split at a separator will consider the separator in the values as well.
2. The header is always included in the RDD and have to be specifically removed.