

ex38

August 12, 2022

```
[ ]: from pyspark import SparkContext, SparkConf
    conf = SparkConf().setAppName("ex38")
    sc = SparkContext(conf=conf)

[8]: inputPath = "data/Ex38/data/"
    outputPath = "out38/"

[9]: inputRdd = sc.textFile(inputPath).filter(lambda line : float(line.
    ↪split(",")[2])>50.0)

[10]: readingsRdd = inputRdd.map(lambda line : (line.split(",")[0],1))

[11]: counterRdd = readingsRdd.reduceByKey(lambda count1, count2 : count1 + count2)

[12]: finalRdd = counterRdd.filter(lambda line : line[1]>=2)

[13]: finalRdd.saveAsTextFile(outputPath)
```