

ex39

August 12, 2022

```
[ ]: from pyspark import SparkConf, SparkContext  
conf = SparkConf().setAppName("ex39")  
sc = SparkContext(conf=conf)
```

```
[8]: inputPath = "data/Ex39/data/"  
outputPath = "out39/"
```

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

```
[10]: readingsRDD = inputRDD\  
      .map(lambda line : (line.split(",")[0],line.split(",")[1]))\  
      .groupByKey()\  
      .mapValues(lambda dates: list(dates))
```

```
[11]: readingsRDD.saveAsTextFile(outputPath)
```