

## ex40

August 12, 2022

```
[ ]: from pyspark import SparkConf, SparkContext
     conf = SparkConf().setAppName("ex40")
     sc = SparkContext(conf=conf)
```

```
[5]: inputPath = "data/Ex40/data/"
     outputPath = "out40/"
```

```
[6]: inputRDD = sc.textFile(inputPath)
```

```
[7]: datesRDD = inputRDD.filter(lambda line : float(line.split(",")[2])>50.0)\
     .map(lambda line : (line.split(",")[0], line.split(",")[1]))\
     .groupByKey()\
     .mapValues(lambda dates : len(list(dates)))\
     .map(lambda record : (record[1], record[0]))\
     .saveAsTextFile(outputPath)
```