

ex37

August 12, 2022

```
[ ]: from pyspark import SparkConf, SparkContext  
conf = SparkConf().setAppName("ex37")  
sc = SparkContext(conf=conf)
```

```
[7]: inputPath = "data/Ex37/data/"  
outputPath = "out37/"
```

```
[8]: inputRDD = sc.textFile(inputPath)  
sensorPM10RDD = inputRDD.map(lambda line : (line.split(",")[0],line.  
    ↪split(",")[2]))
```

```
[9]: groupedPairRDD = sensorPM10RDD.reduceByKey(lambda pm1, pm2 : max(pm1, pm2))
```

```
[10]: groupedPairRDD.saveAsTextFile(outputPath)
```

```
[ ]:
```