

## ex41

August 12, 2022

```
[ ]: from pyspark import SparkContext, SparkConf
     conf = SparkConf().setAppName("ex41")
     sc = SparkContext(conf=conf)

[10]: inputPath = "data/Ex41/data/"
     outputPath = "out41/"
     k = 1

[11]: inputRDD = sc.textFile(inputPath)

[12]: pm10ReadingsRDD = inputRDD.filter(lambda line : float(line.split(",")[2])>50.0)\
     .map(lambda line : (line.split(",")[0], 1))\
     .reduceByKey(lambda value1, value2: value1+value2)\
     .top(k, lambda reading : reading[1])

[13]: finalRDD = sc.parallelize(pm10ReadingsRDD).saveAsTextFile(outputPath)
```