

## ex59

August 20, 2022

```
[ ]: from pyspark import SparkConf, SparkContext
    from pyspark.streaming import StreamingContext

    batch_size = 2

    conf = SparkConf().setAppName("ex59")
    sc = SparkContext(conf=conf)
    ssc = StreamingContext(sc, batch_size)
```

```
[2]: inputPath = "data/Ex59/data/"
    outputPath = "out59/"
```

```
[3]: lines = ssc.textFileStream(inputPath)
```

```
[4]: fullStations = lines.filter(lambda line : int(line.split(",")[1])==0)
```

```
[5]: numFullStations = fullStations.count()
```

```
[6]: numFullStations.pprint()
    numFullStations.saveAsTextFiles(outputPath)
```

```
[7]: ssc.start()
```

```
[ ]: ssc.awaitTerminationOrTimeout(20)
    ssc.stop(stopSparkContext=False)
```