

## ex58

August 20, 2022

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

     batch_size = 2

     conf = SparkConf().setAppName("ex58")
     sc = SparkContext(conf=conf)
     ssc = StreamingContext(sc, batch_size)
```

```
[2]: inputPath = "data/Ex58/data/"
     outputPath = "out58/"
```

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

```
[4]: fullStations = lines\
     .filter(lambda station: int(station.split(",")[1])==0)\
     .map(lambda station: (fullStations.split(",")[3], fullStations.\
     ↪split(",")[0]))
```

```
[5]: fullStations.pprint()
     fullStations.saveAsTextFiles(outputPath)
```

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

```
[ ]: ssc.awaitTerminationOrTimeout(10)
     ssc.stop(stopSparkContext=False)
```