

## ex47

August 14, 2022

```
[ ]: from pyspark import SparkConf, SparkContext
      from pyspark.sql import SparkSession

      conf = SparkConf().setAppName("ex47")
      sc = SparkContext(conf=conf)
      sparkSQL = SparkSession.builder.getOrCreate()
```

```
[2]: inputPath = "data/Ex47/data/"
      outputPath = "out47/"
```

```
[ ]: #loading data
      df = sparkSQL.read.load(inputPath,
                              format="csv",
                              header=True,
                              inferSchema=True)

      df.show()
```

```
[6]: #using dataframes
      df_filtered = df.filter("gender='male'")
      df_newAge = df_filtered.selectExpr("name", "age+1 AS newAge")
      df_sorted = df_newAge.sort(df_newAge.newAge.desc(), df_newAge.name)
      df_sorted.write.csv(outputPath, header=False)
```

```
[ ]: #using SQL
      df.createOrReplaceTempView("people")
      df = sparkSQL.sql("""
      SELECT name, age+1 AS newAge
      FROM people
      WHERE gender='male'
      SORT BY age desc, name
      """)
```