

ex50

August 14, 2022

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

     conf = SparkConf().setAppName("ex50")
     sc = SparkContext(conf=conf)
     ssql = SparkSession.builder.getOrCreate()
```

```
[2]: inputPath = "data/Ex50/data/"
     outputPath = "out50/"
```

```
[3]: df = ssql.read.load(
        inputPath,
        format="csv",
        header=True,
        inferSchema=True
    )
```

```
[ ]: ssql.udf.register("Concatenation", lambda name, surname : name+" "+surname)
```

```
[5]: final_df = df.selectExpr("Concatenation(name, surname) as name_surname")
```

```
[6]: final_df.write.csv(outputPath, header=False)
```

```
[ ]: #per usare SQL, esattamente come prima, una volta aver definito la UDF:
     df.createOrReplaceTempView("people")
     final_df = ssql.sql("""
     SELECT Concatenation(name, surname) as name_surname
     FROM people
     """)
```